Molecular recognition in G-quadruplexes : from G-tetrad formation to protein binding

Cheong, Vee Vee

2014

Cheong, V. V. (2014). Molecular recognition in G-quadruplexes : from G-tetrad formation to protein binding. Doctoral thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/65412>

<https://doi.org/10.32657/10356/65412>

NANYANG
TECHNOLOGICAL
UNIVERSITY

**MOLECULAR RECOGNITION IN G-QUADRUPLEXES:
FROM G-TETRAD FORMATION TO PROTEIN BINDING**

CHEONG VEE VEE

SCHOOL OF PHYSICAL AND MATHEMATICAL SCIENCES

2014

**MOLECULAR RECOGNITION IN G-QUADRUPLEXES:
FROM G-TETRAD FORMATION TO PROTEIN BINDING**

CHEONG VEE VEE

SCHOOL OF PHYSICAL AND MATHEMATICAL SCIENCES

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

2014

Acknowledgement

I am grateful toward Dr. Phan Anh Tuấn for the opportunity to study under his supervisor for the four years of post-graduate study. I would like to acknowledge Brendan Orner's research group, Chen Gang's research group, Shao Fangwei's research group, Li Tianhu's research group, and CBC Teaching Lab for their technical support for my research work. I am grateful towards the members of Phan lab for the time we worked together as fellow researchers. I am thankful for my fellow PhD candidate, Wan Jun, for her companionship, patience and listening voice throughout our journey as graduate students, and would like to express my gratitude towards Brahim and Rongli for their guidance in research. Lastly, I would like to express my gratitude to my parents and siblings, for their support, understanding and patience to me.

Table of Contents

Acknowledgement.....	i
List of Figures	v
List of Tables.....	xi
List of Abbreviations	xiii
Abstract	xv
Chapter 1 Introduction.....	1
1.1 Overview on G-quadruplex	2
1.1.1 Formation and structure of G-quadruplex.....	2
1.1.2 Biological implications of G-quadruplex formation <i>in vivo</i>	7
1.2 DNA damage in G-quadruplexes.....	10
1.2.1 Effects of oxidative damage to G-quadruplexes	10
1.2.2 Nitrosative deamination effect on G-quadruplexes.....	11
1.3 G-quadruplex binding proteins	13
1.3.1 RNA helicase associated with AU-rich element.....	13
1.3.2 Bloom syndrome protein, a member of the RecQ family of helicases.....	16
1.3.3 Structural data of G-quadruplex complex with proteins	18
1.4 Aims and Objectives.....	22
Chapter 2 Methodology	23
2.1 Sample Preparation.....	23
2.1.1 DNA synthesis	23
2.1.2 Peptide and protein synthesis.....	25
2.2 NMR Spectroscopy.....	32
2.2.1 Standard NMR experiments for structure elucidation.....	32
2.2.2 NMR Titration	33
2.3 Circular Dichroism spectrophotometry	34
2.4 Ultraviolet spectrophotometry	34
2.5 Gel Electrophoresis.....	35
2.5.1 Agarose gel electrophoresis	35
2.5.2 Sodium dodecyl sulfate Polyacrylamide Gel Electrophoresis	35
2.5.3 Denaturing gel electrophoresis	35
2.5.4 Electromobility Shift Assay.....	35
2.6 Structure calculations	37
2.6.1 Structure computation for human telomeric G-quadruplex modified with a guanine-guanine-xanthine-8-oxoguanine tetrad	37
2.6.2 Structure computation for complex of Rhau peptide with parallel G-quadruplex	40
2.6.3 Structure computation for the RQC domain of BLM protein	41
2.7 Quantum chemical NMR Computations.....	41
2.8 Poisson-Boltzmann Calculations.....	41
Chapter 3 Xanthine and 8-oxoguanine in G-quadruplexes: formation of a G•G•X•O tetrad.....	42
3.1 Introduction	43
3.2 Single substitution of guanine with 8-oxoguanine and xanthine are highly disruptive to G-quadruplexes	45
3.3 Complementary X•O dual substitutions are compatible with G-quadruplexes.....	49
3.4 Protonation states of a G•G•X•O tetrad	52
3.5 <i>HT-17x9o</i> containing a G•G•X•O tetrad maintains the same (3+1) G-quadruplex fold at pH 5.5-8.0.....	55
3.6 Solution structure of <i>HT-17x9o</i> : maintaining the shape and varying the groove charge	59
3.7 Position-specific X•O pair substitution as a versatile tool in engineering the G-quadruplex form ...	61
3.8 Potential applications of a G-quadruplex containing the G•G•X•O tetrad	64
3.9 Conclusion.....	64
Chapter 4 Inverting the G-tetrad polarity of a G-quadruplex using xanthine and 8-oxoguanine.....	66
4.1 Introduction	66
4.2 The non-complementary X•O modification within a G-tetrad resulted in a predominant G-quadruplex form with altered folding topology.....	69

4.3	Elucidation of the folding topology of <i>HT-3x9o</i> via 2-dimensional NMR spectroscopy	71
4.4	Incorporation of non-complementary X•O modification inverted G-tetrad polarity in <i>HT-3x9o</i>	73
4.5	Conclusion	75
Chapter 5 Insights into G-quadruplex specific recognition by the DEAH-box helicase Rhau:		
Solution structure of a peptide-quadruplex structure Structural basis for complex.....		76
5.1	Introduction	77
5.2	Rhau ₅₃ binds preferentially to parallel G-quadruplexes.....	78
5.3	Minimal G-quadruplex binding domain of Rhau	84
5.4	NMR solution structure of Rhau ₁₈	86
5.5	NMR solution structure of Rhau ₁₈ -G-quadruplex complex	87
5.6	Binding of a second peptide molecule.....	89
5.7	Exposed G-tetrad is determinant for G-quadruplex specific recognition by Rhau	91
5.8	Implications for the elucidation of the Rhau ₁₈ -G-quadruplex complex structure in targeting of G-quadruplexes	92
5.9	Conclusion	92
Chapter 6 Structural and binding study of G-quadruplexes with the conserved domains of BLM.....		94
6.1	Introduction	95
6.2	The combined zinc-binding domain with RQC domain exhibits poor solubility while the isolated RQC domain is highly soluble	96
6.3	NMR spectral assignment of BLMRQC.....	101
6.4	The BLMRQC is a globular protein containing the signature winged-helix motif and long unstructured loops	104
6.5	Comparison between our structure and other solution structure and X-ray crystal structure of the BLMRQC domain	106
6.6	The BLMRQC exhibits low DNA binding affinity to G-quadruplexes and duplex DNA.....	107
6.7	The BLM zinc-binding domain is required for tight binding of BLM with DNA.....	110
6.8	Conclusion	112
Chapter 7 Summary and future prospective.....		113
7.1	Summary	113
7.2	Future prospective	116
Appendix		122
Bibliography.....		139

List of Figures

Chapter 1 Introduction

Figure 1.1	: Different forms of DNA structures and its associated base-pairs.	2
Figure 1.2	: Schematic representation of various conformations of human telomeric G-quadruplexes.	3
Figure 1.3	: Different orientations of the DNA glycosidic angle.	4
Figure 1.4	: Three forms of G-tetrad core with different number of layers.	5
Figure 1.5	: Four possible column orientations within a G-quadruplex core.	5
Figure 1.6	: Three types of loops (linkers) observed on a G-quadruplex.	6
Figure 1.7	: Schematic representation of the human telomere, with the capability to form G quadruplex.	7
Figure 1.8	: Schematic representation of cellular processes interrupted by the formation of G-quadruplexes.	8
Figure 1.9	: Formation of DNA-damaged guanines.	10
Figure 1.10	: Schematic representation of conserved domains in Rhau .	15
Figure 1.11	: Schematic representation of conserved domains in RecQ family.	17
Figure 1.12	: Selected structures of G-quadruplex-protein complex.	20

Chapter 2 Methodology

Figure 2.1	: Schematic diagram for the synthesis of ^{15}N , ^{13}C -labelled oligonucleotide via hairpin extension.	24
Figure 2.2	: FPLC purification of ^{15}N , ^{13}C -Rhau ₁₈ .	27
Figure 2.3	: SDS-PAGE of ^{15}N , ^{13}C -Rhau ₁₈ .	28
Figure 2.4	: Schematic diagram of PCR-based fusion of two gene fragments used to construct GSTekBLMZRQC recombinant plasmid.	30
Figure 2.5	: Example gel shift image demonstrating DNA-protein complex formation with DNA.	36

Chapter 3 Xanthine and 8-oxoguanine in G-quadruplexes: formation of a G•G•X•O

tetrad

Figure 3.1	: Chemical drawing guanine analogues used in this study: xanthine (X) and 8-oxoguanine (O) and the schematic diagram of <i>HT</i> G-quadruplex.	43
Figure 3.2	: Characterization of <i>HT</i> G-quadruplex the constructs: <i>HT-9o</i> , <i>HT-17x</i> , <i>HT-17x9o</i> .	44
Figure 3.3	: Characterization of single-modified <i>HT</i> and the constructs <i>HT-9o</i> , <i>HT-16o</i> , and <i>HT-17o</i> .	45
Figure 3.4	: Characterization of <i>HT</i> constructs with point mutation of xanthine.	46
Figure 3.5	: Characterization of <i>HT</i> constructs with point mutation of inosine.	47
Figure 3.6	: Characterization of <i>HT</i> double-modified with X•O substitution.	50
Figure 3.7	: Characterization of <i>HT-21x17o</i> and the constructs: <i>HT-17o</i> , <i>HT-21x</i> and <i>HT-21x17o</i> .	51
Figure 3.8	: Comparison of imino spectra of <i>HT</i> , <i>HT-17x9o</i> and <i>HT-9x3o</i> recorded from pH 5.0, 7.0 and 9.0.	52
Figure 3.9	: NMR titration of <i>HT-17x9o</i> .	53
Figure 3.10	: Comparison of changes in experimental and computed NMR chemical shifts <i>HT-17x9o</i> .	54
Figure 3.11	: Acid dissociation constant (pKa) values of xanthine (X) which undergoes deprotonation at N3.	54
Figure 3.12	: NMR spectral assignment of <i>HT-17x9o</i> at pH 5.5.	55
Figure 3.13	: NMR spectral assignment of <i>HT-17x9o</i> at pH 7.0.	56
Figure 3.14	: NMR spectral assignment of <i>HT-17x9o</i> at pH 8.0.	57
Figure 3.15	: NOESY spectrum of <i>HT-17x9o</i> recorded at pH 5.5, pH 7.0 and pH 8.0, illustrating H1-H1/H7-H1/H1-H7 through space correlations.	58
Figure 3.16	: Spectra of through-bond 2D NMR experiments used for verification of NOE assignment of <i>HT-17x9o</i> .	59
Figure 3.17	: Imino proton spectrum of <i>HT-17x9o</i> at pH 7.0 recorded after 15 min exposure in D ₂ O.	59
Figure 3.18	: Side view and top view of 10 superimposed lowest energy conformations of <i>HT-17x9o</i> .	60
Figure 3.19	: Comparison between double-modified <i>HT-17x9o</i> and unmodified <i>HT</i> G-quadruplex.	60

Figure 3.20	: Comparative electrostatic potential surface representation of <i>HT-17x9o</i> and its wild-type <i>HT</i> G-quadruplex	61
Figure 3.21	: Comparison of different G-quadruplexes <i>HT1</i> and <i>HT2</i> applied in structure engineering experiment.	62
Figure 3.22	: CD spectra of <i>HT1</i> and <i>HT1-10x16o</i> ; <i>HT2</i> and <i>HT2-4x10o</i> .	63

Chapter 4 Inverting the G-tetrad polarity of a G-quadruplex using xanthine and 8-oxoguanine

Figure 4.1	: G-quadruplex and arrangement of G•G•X•O tetrad used in this study.	68
Figure 4.2	: Characterization of <i>HT-3x9o</i> by 2D NMR spectroscopy under basic conditions (pH 9.0)	69
Figure 4.3	: Imino proton spectra of <i>HT</i> , <i>HT-3x9o</i> and <i>HT-9x17o</i> G-quadruplexes recorded at pH 7.0 and pH 9.0.	70
Figure 4.4	: NMR spectral assignment of <i>HT-3x9o</i> at pH 9.0 showing the imino region.	71
Figure 4.5	: NOESY spectrum of <i>HT-3x9o</i> recorded at pH 9.0 (mixing time, 250 ms) showing H1-H1/H7-H1/H1-H7 through space correlations.	72
Figure 4.6	: Sequential 3-5, 10-11, 15-17 and 21-23 walk on NOESY spectrum of <i>HT-3x9o</i> (mixing time, 300 ms) at pH 9.0,	72
Figure 4.7	: Comparison between the <i>HT</i> and <i>HT-3x9o</i> G-quadruplexes.	73
Figure 4.8	: CD spectra of <i>HT</i> , <i>HT-3x9o</i> and <i>HT-9x17o</i> G-quadruplexes at pH 9.0.	74

Chapter 5 Insights into G-quadruplex specific recognition by the DEAH-box helicase Rhau: Solution structure of a peptide-quadruplex complex

Figure 5.1	: Native gel electrophoresis of DNA in the absence and presence of Rhau ₅₃ at 1:10 ratio	78
Figure 5.2	: Comparative binding of Rhau ₅₃ with non-parallel G-quadruplexes	79
Figure 5.3	: EMSA of Rhau ₅₃ between G-quadruplexes of non-parallel form (<i>HT</i>), parallel form in crowded conditions (<i>HT</i> crowded) and parallel form (<i>T95-2T</i>)	79
Figure 5.4	: NMR imino protons of DNA/RNA G-quadruplexes and duplex sequences used in this study	80
Figure 5.5	: Native gel electrophoresis of parallel DNA G-quadruplexes, parallel RNA G-quadruplexes, non-parallel DNA G-quadruplexes and duplex DNA with Rhau ₅₃	80
Figure 5.6	: Demonstration of preferential binding affinity of Rhau ₅₃ towards parallel G-quadruplexes on imino proton spectra	81
Figure 5.7	: ¹ H, ¹⁵ N-HSQC of free Rhau ₅₃ and in complex with <i>T95-2T</i> at	83

	1:1 ratio	
Figure 5.8	: ^1H , ^{15}N -HSQC spectra of Rhau ₅₃ free and bound to <i>T95-2T</i>	83
Figure 5.9	: Native gel electrophoresis of <i>T95-2T</i> with Rhau peptides of varying length	84
Figure 5.10	: NMR imino proton spectra of <i>T95-2T</i> titrated with increasing concentrations of Rhau ₅₃ , Rhau ₁₈ , and Rhau ₁₆	85
Figure 5.11	: ^1H , ^{15}N -HSQC spectra of free Rhau ₁₈ and bound to <i>T95-2T</i>	85
Figure 5.12	: Native gel electrophoresis of DNA in with and without the presence of Rhau ₁₈	85
Figure 5.13	: Cartoon representation of the ten superimposed structures of Rhau ₁₈	86
Figure 5.14	: NMR assignment of free <i>T95-2T</i> and in complex with Rhau ₁₈ at 1:1 ratio	86
Figure 5.15	: Strip plots illustrating assignment of Rhau ₁₈ in complex with <i>T95-2T</i>	87
Figure 5.16	: NMR solution structure of the Rhau ₁₈ - <i>T95-2T</i> complex	88
Figure 5.17	: Comparison between free and bound structures in the Rhau ₁₈ - <i>T95-2T</i> complex	88
Figure 5.18	: Structure representation of Rhau ₁₈ in complex with <i>T95-2T</i> , with NOESY spectrum showing the intermolecular NOEs between the protein and the DNA imino proton	89
Figure 5.19	: Diagram illustrating different complex state of <i>T95-2T</i> with Rhau ₁₈	90

Chapter 6 Structural and binding study of G-quadruplexes with conserved domains of BLM

Figure 6.1	: Schematic representation of conserved domains in human BLM and BLM fragments used in this study	96
Figure 6.2	: SDS-PAGE of selected BLM fragments	97
Figure 6.3	: ^1H , ^{15}N -HSQC spectrum of BLMRQC recorded at 25 °C	101
Figure 6.4	: Sequential connectivity pathway of BLMRQC (residues 1071–1078)	102
Figure 6.5	: Side chain assignment of carbon and proton resonance of Val1180 in BLMRQC	103
Figure 6.6	: Structure of BLMRQC model	104
Figure 6.7	: Comparison of RQC structures from BLM and WRN	105
Figure 6.8	: Comparison of BLMRQC structures	106
Figure 6.9	: Gel shift electrophoresis of BLMRQC with different DNA substrates	108
Figure 6.10	: Comparative ^1H , ^{15}N -HSQC spectra displaying titration of	109

BLMRQC with G-quadruplex and duplex DNA

Figure 6.11	: Comparative gel shift images between BLMZRQC and BLMRQC	111
Figure 6.12	: Superimposed BLM core protein demonstrating different orientations of RQC domain	112
Appendix		
Figure A.1	: Vector map of pNIC28-Bsa4	122
Figure A.2	: Vector map of pET32-Ek/LIC	123
Figure A.3	: Coding region of pNIC-GST vector	123
Figure A.4	: Vector map of pFB-LIC	124
Figure C.1	: SDS-PAGE of BLMZRQC with residues 992–1192 under denaturing conditions	126
Figure D.1	: SDS-PAGE of Trx-tagged BLM proteins	127
Figure D.2	: SDS-PAGE of GST-tagged BLM fragments expressed in <i>E. coli</i>	128
Figure D.3	: SDS-PAGE of helicase domain-containing BLM fragments	129
Figure D.4	: SDS-PAGE of isolated domains of BLM expressed in <i>E. coli</i> .	130
Figure D.5	: SDS-PAGE of BLM fragments expressed in <i>Sf9</i> cells	131
Figure D.6	: SDS-PAGE of helicase2A domain-containing BLM fragments	132
Figure E.1	: MALDI-TOF spectra of Rhau proteins	133
Figure E.2	: MALDI-TOF spectrum of BLM proteins	134
Figure E.3	: MALDI-TOF spectrum of WRNRQC	134
Figure E.4	: Protein analysis via mass spectrometry of sample obtained from TrxBLMZ after TEV cleavage and ion exchange chromatography	135

List of Tables

Table 1.1	: List of selected G-quadruplex-interacting proteins and their available protein-DNA/RNA structure	14
Table 3.1	: Mono-substituted G-quadruplex sequences and their thermal stability	48
Table 3.2	: Dual-substituted G-quadruplex sequences and their thermal stability	49
Table 3.3	: Sequence of ¹⁵ N-labelled <i>HT-17x9o</i> used in this study	58
Table 3.4	: DNA sequence of X•O modified G-quadruplexes used for structure engineering study and their thermal stability	63
Table 4.1	: DNA sequence of <i>HT</i> and X•O modified G-quadruplexes, and their thermal stability.	70
Table 4.2	: Sequence of ¹⁵ N-labelled <i>HT-3x9o</i> used in this study	73
Table 5.1	: DNA/RNA sequences used in binding study of Rhau peptides.	81
Table 5.2	: Peptide sequences used to determine minimal G-quadruplex binding domain of Rhau.	84
Table 6.1	: Conserved domain make-up and solubility of BLM constructs	100
Table 6.2	: Sequence of oligonucleotides and their corresponding structure utilized in titration with BLM proteins	110
Table B.1	: List of primers used in the cloning procedure	125
Table F.1	: Statistics of the computed structures of 8-oxoguanine, xanthine-double modified human telomeric sequence, <i>HT-17x9o</i>	136
Table F.2	: Statistics of the computed structures of free Rhau ₁₈	137
Table F.3	: Statistics of the computed structure of the <i>T95-2T</i> -Rhau ₁₈ complex	138

List of Abbreviations

A	Adenine
BLM	Bloom syndrome protein
CD	Circular dichroism
COSY	Correlation spectroscopy
C	Cytosine
DNA	Deoxyribonucleic acid
EMSA	Electromobility Shift Assay
FAM	6-Carboxyfluorescein
FPLC	Fast protein liquid chromatography
G	Guanine
HPLC	High performance liquid chromatography
HMBC	Heteronuclear Multi Quantum Coherence
HSQC	Heteronuclear Single Quantum Coherence
HRDC	Helicase and RNaseD C-terminal
IPTG	Isopropyl- β -D-1-thiogalactoside
K_D	Dissociation constant
Kpi	$\text{KH}_2\text{PO}_4/\text{K}_2\text{HPO}_4$
MWCO	Molecular weight cutoff
NCBI	National Center for Biotechnology Information
NOESY	Nuclear Overhauser effect spectroscopy
NMR	Nuclear magnetic resonance
O	8-oxoguanine
PDB	Protein Data Bank
Rhau	RNA associated with AU-rich element
RMSD	Root-mean-square deviation
RQC	RecQ C-terminal
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
T	Thymine
TAMRA	5-Carboxytetramethylrhodamine
TOCSY	Total correlated spectroscopy

UV	Ultraviolet
WRN	Werner syndrome protein
X	Xanthine
ZnBD	Zinc-binding domain

Abstract

Under physiological conditions, guanine-rich oligonucleotides may assemble into a four-stranded structure, the G-quadruplex. G-quadruplexes may cause genomic instability in cells, requiring modulation by proteins.

As guanine-rich DNA is susceptible to DNA damage, a G-quadruplex with complementary 8-oxoguanine-xanthine mutation was characterized. Its structure resolved in K^+ highly resembles the wild-type G-quadruplex. In contrast, the non-complementary arrangement of 8-oxoguanine and xanthine within a G-tetrad inverts the polarity of the modified tetrad.

We demonstrated Rhau₅₃₋₁₀₅ to preferentially bind parallel G-quadruplexes, and identified residues 53–70 (Rhau₁₈) as the determinant G-quadruplex binding region. We present the solution structure of a G-quadruplex bound with peptide, and identified hydrophobic, stacking and electrostatic interactions between the two biomolecules.

BLM is a RecQ helicase that selectively resolves G-quadruplexes. The RQC domain (BLMRQC) is highly soluble and the solution structure of BLMRQC adopts a winged-helix motif. BLMRQC binds DNA weakly and requires the zinc-binding domain for tight G-quadruplex binding.

Chapter 1 Introduction

Guanine-rich oligonucleotides may spontaneously assemble into a G-quadruplex under physiological salt conditions¹. This higher-ordered DNA/RNA structure exhibits high thermostability^{2,3}, and is structurally different from the canonical duplex B-DNA structure⁴. A G-quadruplex is constituted by stacked layers of G-tetrads with the guanine bases adopting a square planar array, which is stabilized by Hoogsteen hydrogen bonds¹ and electrostatic interactions with the cation⁵. Over 375,000 potential G-quadruplex-forming sequences have been identified in the human genome^{6,7}, and the formation of G-quadruplexes in the cell was demonstrated via immunostaining to occur within the human telomere and other chromosomal regions⁸. DNA replication and transcription of these G-quadruplexes-forming sequences may potentially result in genomic instability⁹⁻¹⁷, altered gene expression¹⁸⁻²³, and improper functioning of cellular processes²⁴. Correspondingly, there exist various proteins which have been determined to unwind, promote²⁵⁻³⁰ and cleave³¹⁻³⁵ G-quadruplexes for the modulation of G-quadruplex population within the cells. This thesis focuses on the elucidation of structural features of G-quadruplexes and their associated proteins to establish the molecular recognition principles in such nucleic acid structures.

1.1 Overview on G-quadruplex

1.1.1 Formation and structure of G-quadruplex

Since the start of the 20th century, scientists have reported the formation of a thermostable, gel-like substance when concentrated solutions of guanine monophosphate were cooled^{2,3}. However, it took almost half a century before Gellert *et. al.* revealed the X-ray diffraction pattern of these guanine aggregates, which indicated an assembly of guanine monomers in tetrameric arrangement⁴ (Figure 1.1A). This alternate form of guanine base arrangement differs radically from the contemporary adenine-thymine (A-T) and guanine-cytosine (G-C) base-pairing observed in DNA duplex structures, which followed the Watson and Crick model³⁶. As illustrated in Figure 1.1B, for guanine bases involved in the Watson-Crick base-pairing, three hydrogen bonds are formed between the guanine (purine) base and cytosine (pyrimidine) base, while in the guanine tetrad (G-tetrad), a single guanine base forms two hydrogen bonds with two of its neighbouring guanine bases.

Figure 1.1 Different forms of DNA structures and its associated base-pairs. A) Schematic representation of a G-quadruplex formed by consecutive stacking of three G-tetrads, which are stabilized by cations. Each parallelogram represents a guanine base, with the arrows representing the sugar backbone of the oligonucleotide involved in tetramer formation. The chemical structure of a G-tetrad is showed in the insert, where four guanines (G_{α} - G_{δ}) are arranged in a planar fashion, and coordinated to each other by Hoogsteen hydrogen base pairing; B) cartoon representation of a B-DNA duplex with the chemical structure of adenine-thymine (A-T) and guanine-cytosine (G-C) Watson-Crick base pairs illustrated in the insert. The R group in each diagram refers to a DNA.

The formation of a G-tetrad was demonstrated by Sen and Gilbert in guanine-rich oligonucleotides under physiological salt concentrations¹, where guanine (G) nucleotides spontaneously assemble into G-quadruplexes (Figure 1.1A). Due to the unique structural arrangement of this non-canonical DNA form, considerable effort was placed in resolving the G-quadruplex structure. Currently, approximately 130 G-quadruplex structures of both DNA and

RNA origin of various conformations have been resolved by NMR spectroscopy and X-ray crystallography³⁷, which indicate G-quadruplexes to exhibit structure polymorphism (Figure 1.2). However, analysis of these G-quadruplex forms has identified several common structural features: G-tetrad core; types of loops connectivity; and groove widths, which will be discussed in detail.

1.1.1.1 G-tetrad core

The G-tetrad core of a G-quadruplex is made up of stacked layers of G-tetrads, with its assembly aided by the presence of cations⁵. Each G-tetrad contains four units of the guanine base in a tetrameric arrangement that is stabilized by Hoogsteen hydrogen bonds. The square planar configuration of the G-tetrad is a result of its cyclic hydrogen bond connectivity, whereby the imino N(1)-H and amino N(2)-H of G_(i) are coordinated to O(6) and N(7) of G_(i+1)

Figure 1.2 Schematic representation of various conformations of human telomeric G-quadruplexes. A) Tetrameric parallel-stranded G-quadruplex observed for the single-repeat human telomeric sequences d(TTAGGG) and d(TTAGGGT) in K⁺ solution³⁸; B) dimeric parallel-stranded G-quadruplex observed for the two-repeat human telomeric sequence d(TAGGGTTAGGGT) in a K⁺-containing crystal³⁹ and in K⁺ solution⁴⁰; C) dimeric antiparallel-stranded G-quadruplex observed for two-repeat human telomeric sequence d(TAGGGTTAGGGT) in K⁺ solution⁴⁰; D) asymmetric dimeric (3+1) G-quadruplex observed for the three-repeat human telomeric sequence d(GGGTTAGGGTTAGGGT) in Na⁺ solution⁴¹; E) asymmetric dimeric (3+1) G-quadruplex association observed for the three-repeat human telomeric sequence d(GGGTTAGGGTTAGGGT) and the single-repeat human telomeric sequence d(TAGGGT) in Na⁺ solution⁴¹ and in K⁺ solution (unpublished results); F) basket-type form observed for d[A(GGGTTA)₃GGG] in Na⁺ solution⁴²; G) propeller-type form observed for d[A(GGGTTA)₃GGG] in a K⁺-containing crystal³⁹; H) (3+1) form 1 observed for d[TA(GGGTTA)₃GGG] in K⁺ solution⁴³⁻⁴⁹; I) (3+1) form 2 observed for d[TA(GGGTTA)₃GGGTT] in K⁺ solution^{43,48,50}; J) basket-type form observed for d[(GGGTTA)₃GGGT] in K⁺ solution⁵¹. *Anti* guanines are coloured cyan; *syn* guanines are coloured magenta; loops are coloured red. M, N and W represent medium, narrow and wide grooves, respectively. This figure is adapted from ⁵².

respectively⁴, and facilitates the stacking of G-tetrads within the G-quadruplex scaffold. As illustrated in Figure 1.1A, the square planar array orientates the carbonyl O(6) towards the centre of the G-tetrad. Depending on the cation species, the cation could be situated either on, or in-between the plane of the G-tetrad⁵³⁻⁵⁵, with electrostatic interaction between the O(6) and the cation stabilizing the G-tetrad⁵⁶. This stabilizing effect is sensitive to the cationic species used, as exemplified by the formation of G-quadruplex promoted in the presence of K^+ and Na^+ , but not in Li^+ ^{57,58}. Generally, the K^+ -stabilized G-quadruplexes are extensively studied due to its biological relevance as the most abundant intracellular cation⁵⁹. Moreover, it has stronger thermal stabilizing effects for group I elements (in the order $K^+ > Na^+ \gg Li^+$)^{58,60,61}, although G-quadruplex structures stabilized by Rb^+ , Cs^+ and NH_4^+ have also been reported⁶²⁻⁶⁴.

Glycosidic angle

Figure 1.3 Different orientations of the DNA glycosidic angle. Guanine can adopt two glycosidic orientations: *anti* (left) and *syn* (right), through the rotation of the glycosidic bond, which is highlighted in red.

Two possible glycosidic orientations could be adopted by the guanine base within the G-quadruplex scaffold via rotation about the DNA glycosidic bond. As seen in Figure 1.2A and Figure 1.2B, each guanine base within these G-quadruplexes has adopted the *anti* glycosidic orientation, in which the nucleic acid base is orientated away from the pentose ring (Figure 1.3, left). Guanine bases with their nucleic acid base orientated towards their pentose ring in the *syn* glycosidic orientation (Figure 1.3, right). As illustrated in Figure 1.2C–F, and Figure 1.2H–J, guanine bases adopt both *syn* and *anti* orientations in these G-quadruplexes.

Figure 1.4 Three forms of G-tetrad core with different number of layers. A) 2-layer G-tetrad core; B) 3-layer G-tetrad core; C) 4-layer G-tetrad core.

Based on the number of G-tetrads that are stacked upon each other within a G-quadruplex scaffold, G-quadruplexes have been shown to generally possess at least two G-tetrad layers, as illustrated in Figure 1.2J. However, the three-layered G-quadruplexes (Figure 1.2A – I), and four-layered G-quadruplexes *i.e.* from $d(TG_4T)_4$ in Na^{+65} and $d(G_4T_4G_4)_2$ in K^{+55} are more commonly observed (Figure 1.4).

Strand directionality

Figure 1.5 Four possible column orientations within a G-quadruplex core: A) parallel-stranded; B) anti-parallel orientation, with strands adopting a up-up-down-down directionality; C) anti-parallel strand orientation, with up-down-up-down directionality; D) (3+1) orientation with only one strand with opposite directionality. Strands of the same directionality are coloured in black, and those in opposite directionality are showed in red.

Within a G-quadruplex scaffold, the sugar phosphate backbone of the oligonucleotide strand (or column) connects the guanine bases from one G-tetrad to another, with these columns either having an upwards or downwards directionality. In the four-stranded G-quadruplex structure, four possible column orientations could be arranged within a G-tetrad. This is illustrated in Figure 1.5, with the formation of a parallel-stranded G-tetrad core possessing four columns with the same directionality, two anti-parallel stranded G-tetrad cores with either an up-up-down-down or up-down-up-down directionality, or a (3+1) G-tetrad core comprising of three strands with the same directionality and an opposing strand.

1.1.1.2 Loops connectivity

Figure 1.6 Three types of loops (linkers) observed on a G-quadruplex: A) double chain reversal loop; B) edgewise loop; C) diagonal loop.

Structural data available on the G-quadruplexes generally demonstrated three different forms of connectivity pathway that joins one column to another within the G-tetrad core, with these pathways are termed as loop or linkers. In the first type of loop, the double chain reversal loop (also termed as propeller loop) joins two flanking columns with parallel directionality (Figure 1.6A), while two anti-parallel columns adjacent (Figure 1.6B) or opposite to each other in the G-tetrad core (Figure 1.6C) are joined together by the second and third type of loop respectively.

1.1.1.3 Stoichiometry

As the G-quadruplex scaffold is composed of four individual strands aligned in a square planar fashion, the involvement/or absence of the loops joining these strands is related to the stoichiometry of the G-quadruplex. In a unimolecular G-quadruplex, the four columns are connected by three loops, with the type of loop involved determined by the directionality and location of the two participating columns (Figure 1.2F – J). Bimolecular G-quadruplexes are either formed by two pairs of anti-parallel columns each linked by two loops (Figure 1.2B, C), or are assembled by four columns in which only three columns are joined together by two loops (Figure 1.2D, E). In contrast, tetramolecular G-quadruplex does not possess any loops that link the columns together (Figure 1.2A)

1.1.1.4 Groove width

The spaces formed between the sugar phosphate backbones (columns) of the G-quadruplex are referred to as grooves. Within a G-tetrad core, there exist three types of groove width: wide (W), medium (M) and narrow (N). As illustrated in Figure 1.2A, when all strands exhibit parallel directionality, each of the four resulting grooves will be of M width. Examples of

N and W groove can be seen in both anti-parallel G-quadruplex (Figure 1.2C, J) and (3+1) G-quadruplex (Figure 1.2D-I).

1.1.2 Biological implications of G-quadruplex formation *in vivo*

With the completion of the Human Genome Project, sequence analyses of the human genome have uncovered numerous genomic sectors of high GC content across the human chromosome⁶⁶. In particular, the formation of G-quadruplexes of diverse conformations has been established within the telomeric region, micro- and mini-satellites, and promoter regions via *in vitro* studies^{52,67,68}.

In recent years, the development of a G-quadruplex specific antibody has allowed for the identification of G-quadruplexes within the cells^{8,69,70}. Using this technique, G-quadruplexes were visualized at the telomeres⁸, a nucleoprotein complex located at the chromosomal ends, which prevents chromosomal degradation and fusion⁷¹. Similar immunofluorescent techniques have also been applied in the visualization of G-quadruplexes in the telomeric regions of *Stylonychia lemnae*^{72,73}. The telomere is highly associated with the formation of G-quadruplexes due to its single-stranded 3' overhang, which comprises of numerous tandem repeats of (GGGTTA)_n for humans, and has been shown to assemble into G-quadruplexes under physiological conditions^{38-51,58} (Figure 1.7). The formation of G-quadruplexes was previously shown to be an inhibitor of telomeric elongation by the action of telomerase⁷⁴⁻⁷⁶. Moreover, there exists various telomere-associated proteins (*i.e.* TEβPα and TEβPβ⁷², POT1^{77,78}, BLM⁷⁹) that exhibit G-quadruplex binding, unwinding and formation capability. These findings imply G-quadruplexes to serve a significant role in telomere protection and maintenance of genomic stability⁸⁰.

Figure 1.7 Schematic representation of the human telomere, with the capability to form G quadruplex. An example of a G-quadruplex formed from human telomeric sequence d[TAGGG(TTAGGG)₃] is shown⁴⁹.

In silico analyses of the human genome have established up to 375,000 putative quadruplex sites with the potential to assemble as G-quadruplexes^{6,7}. However, with the DNA typically maintained as duplexes in cells, Watson-Crick base pairing between the G-rich DNA and its complementary strand inhibited the assembly of G-quadruplexes from such G-quadruplex motifs. The dissolution of the G-rich duplex DNA into its single-stranded form via cellular processes such as DNA replication and gene transcription is thus essential for the formation of G-quadruplexes within the cells.

Figure 1.8 Schematic representation of cellular processes interrupted by the formation of G-quadruplexes. A) A DNA replication fork stalled by the presence of a G-quadruplex on the leading strand. The grey arrows indicate the direction the daughter strands is synthesized; B) Formation of a G loop during transcription. G-quadruplexes are formed on the non-transcribed strand, with concurrent generation of an RNA:DNA hybrid on the transcribed strand. DNA transcription by the second RNA polymerase II is interrupted by the G loop. This figure is adapted from various sources⁸¹⁻⁸³.

However, the formation of G-quadruplexes during these cellular processes may be undesirable, as the highly thermostable G-quadruplex can act as a physical blockage for DNA/RNA polymerases. This prevents the onward progression of the replication fork^{10,15,30,84-86} for the DNA and halting transcriptional elongation^{19-22,83,87,88} for the RNA (Figure 1.8), which may result in genomic instability and/or altered gene expression.

Formation of G-quadruplexes during cell replication has been demonstrated by Biffi *et al.*, where fluorescent immunostaining of G-quadruplexes bound to a G-quadruplex-specific

antibody (BG4) was found to be present at an elevated level during the S phase of the cell cycle, and shown to be absent when DNA replication was arrested by the application of aphidicolin to the cell cultures⁸. DNA replication fork stalled by G-quadruplexes can be restored by G-quadruplex resolving helicases such as BLM and Pif1, which can resolve G-quadruplexes formed on the parent strand of the leading and lagging strand respectively^{81,89}.

Formation of G-quadruplexes during DNA transcription of G-rich DNA regions arrests the onward progression of the RNA polymerase II and generates a stable RNA:DNA hybrid on the non-coding strand^{19,20}. This G-quadruplex-induced transcriptional blockage is collectively termed as a G loop, which is illustrated in Figure 1.8B. G loops have been visualized via electron microscopy on *E. coli* with impaired RnaseH and RecQ helicase activity following *in vitro* transcription through a G-rich region^{83,87,90,91}. The enrichment of G-quadruplex resolving proteins such as XPD/XPB at transcriptional start sites⁹² implicates the need to regulate the formation of G-quadruplexes for unaltered gene expression.

Overall, the G-quadruplex appears to possess diverse functionalities in the human genome, with its formation highly modulated within the cells. Based on the wide diversity of conformation that the G-quadruplex can adopt, and the localization of G-quadruplex forming sequences within key regulatory regions, the G-quadruplex may potentially serve as a molecular tool for targeting proteins *in vivo*, and also as a potential target for drug therapeutics studies⁹³⁻⁹⁵.

1.2 DNA damage in G-quadruplexes

Within the cells, exposure of DNA to mutagens results in the formation of DNA lesions⁹⁶. The presence of such mutations within the genomic sequence is potentially damaging towards the proper functioning of downstream cellular processes, which may have implications on aging⁹⁷ and/or cancer⁹⁸ if left unregulated. A source of these mutagens includes the endogenous metabolites: reactive oxygen species (ROS) and reactive nitrogen species (RNS), which may cause DNA damage through oxidative damage and nitrosative deamination respectively^{96,99,100} (Figure 1.9). Guanine base lesions are predominantly generated on the DNA strand following the attack by these bioactive molecules¹⁰¹⁻¹⁰³, and the guanine base lesions may potentially be incorporated within G-quadruplexes when DNA damage is afflicted on G-rich genomic regions^{104,105}.

Figure 1.9 Formation of DNA-damaged guanines. Left) Oxidative products: 2,6-diamino-4-hydroxy-5-formaminopyrimidine (FapyG) and 8-oxoguanine (O); right) nitrosative deamination products: oxanosine (Ox) and xanthine (X). The major degradation products of each reaction pathway are underlined.

1.2.1 Effects of oxidative damage to G-quadruplexes

8-oxoguanine (O, Figure 1.9) is a major DNA oxidative damage product arising from the action of ROS-generated $\cdot\text{OH}$ radicals on guanine base¹⁰³. Amongst the four nucleobases, guanines possesses the lowest oxidation potential¹⁰² and thus is readily oxidized, with 100,000 O potentially formed daily per cell^{106,107}. With O formation occurring more prevalently within the guanine-rich telomeric regions over non-telomeric genomic sectors^{105,108}, the telomeric region

may accumulate more oxidative damage¹⁰⁹, and correspondingly increases the likelihood of O incorporated within the G-quadruplex scaffold *in vivo*.

In an *in vitro* study by Szalai and colleagues, site-directed mutagenesis of G → O on an anti-parallel G-quadruplex forming sequence resulted in the disruption of its native G-quadruplex fold, with the mutated G-quadruplex losing its inhibitory effects on telomerase¹¹⁰. Volícková and co-workers also reported the thermal destabilization for all G-to-O point mutations in another human telomeric anti-parallel G-quadruplex, particularly for O substitutions on the middle G-tetrad layer¹¹¹. Single-substitution of O on another human telomeric G-quadruplex with a predominant (3+1) fold demonstrated a drop in thermal stability together with a dramatically altered imino proton pattern or CD profile, indicative of a major structural change after O substitution^{112,113}. Singh *et. al.*, also showed in a parallel G-quadruplex that the complete G-to-O substitutions of a G-tetrad exhibited reduction in thermostability and alteration to the parent G-quadruplex fold¹¹⁴. In a kinetic study performed on a tetramolecular parallel G-quadruplex originating from *Tetrahymena*, Gros *et. al.* have determined that the G-quadruplexes formed after point mutations of O are disfavouredly assembled in comparison to with the wild-type¹¹⁵. Overall, the effect of O substitution within a G-quadruplex appears to be detrimental to its thermostability, and the conservation of its parent G-quadruplex fold.

1.2.2 Nitrosative deamination effect on G-quadruplexes

Guanine bases attacked by RNS undergoes nitrosative deamination to form xanthine (2'-deoxyxanthosine; Figure 1.9)¹¹⁶, a stable DNA lesion with a half-life of 17,700 h at physiological pH¹¹⁷. Other mechanisms contributing to the X population in DNA includes the action of deaminase enzymes on DNA¹¹⁸, nucleobase deamination as a result of hydrolytic action on guanine¹¹⁹, and from defective purine metabolism¹²⁰.

Although the percentage of guanine bases is equal to that of cytosine bases within the cells^{121,122}, Caulfield *et. al.* has determined guanine bases to undergo nitrosative deamination more readily than cytosine bases, with xanthine (X) formed as the major product¹⁰¹. The effect of nitrosative deamination on DNA was demonstrated by Dong *et. al.*, where up to 99 ± 14 of X per 10^6 nucleotides was generated on pUC19 plasmid DNA following 12 h exposure to physiological

concentrations of $\text{NO}\cdot$ RNS¹²³. However, in a subsequent study, only 7 ± 1 of X per 10^6 nucleotides was detected on human TK6 lymphoblastoid cells under similar $\text{NO}\cdot$ exposure conditions¹²⁴. The much lower levels of DNA lesion generated on the TK6 cells in comparison to the plasmid was possibly attributed to the high efficiency of DNA repair processes that acted on the high levels of X generated, or due to limited exposure of nuclear DNA to the $\text{NO}\cdot$ radicals¹²⁴.

Since RNS exhibit reactivity for DNA substrates of the following order: single-stranded DNA > G-quadruplex >> duplex¹⁰¹, the guanine-rich human telomere which contains a 3' overhang may be an attractive target for nitrosative deamination over other DNA regions within the human genome. Under physiological conditions, the damaged DNA telomeric region with X mutations may potentially assemble into X-containing G-quadruplexes. *In silico* studies have demonstrated G-tetrad models formed by X nucleobases¹²⁵ and modified X nucleobases: 9-methylxanthine¹²⁶/3-methylxanthine¹²⁷, to exhibit favourable energetic supporting the formation of a G-tetrad, and are computed to be structurally compatible to the wild-type G-tetrads. Molecular dynamics simulation of G-quadruplexes composed of 3-thioxanthine¹²⁸ and 9-deaza-8-haloxanthine^{129,130} exhibited enhanced ion coordination and stacking interactions, which suggests towards the feasibility of X as a substituent in G-quadruplexes.

1.3 G-quadruplex binding proteins

The potential of G-quadruplex formation in functionally important regions of the human genome suggests that G-quadruplexes may affect critical biological processes¹³¹. Correspondingly, various proteins have been found to interact *in vitro* with G-quadruplexes of ciliate, yeast, and human origin^{132,133}, which may further mediate the effect(s) G-quadruplexes have on these cellular processes. A large portion of these G-quadruplex-interacting proteins are involved in telomeric maintenance^{134,135}, and are critical for genomic stability¹³⁶⁻¹⁴¹. These proteins were also found to demonstrate preferentiality for G-quadruplexes over duplexes⁹⁰. Augmented with their heavy involvement in the protection of genomic integrity and ubiquitous occurrence in the cells, this suggests that the presence of G-quadruplexes may be inherently significant within cellular system^{135,142}.

In a review by Oganessian and Bryan¹³³, these G-quadruplex interacting proteins have been classified into four different categories: G-quadruplex stabilizing proteins, which includes the RGG and RBD domains of nucleolin²⁶⁻²⁸ and MutS α ⁹¹; G-quadruplex destabilizing proteins which comprise of AUF1 (hnRNP D)¹⁴³, UP1 (hnRNP A1)¹⁴⁴⁻¹⁴⁶, Pot1^{77,78,147}, *etc.*^{132,134,135}; G-quadruplex selective nuclease such as GQN1¹⁶; and ATP-dependent G-quadruplex unwinding proteins (helicases) including Rhau¹⁴⁸ and BLM^{79,149} (Table 1.1). The wide diversity of conformation that G-quadruplexes can adopt and the wide array of G-quadruplex interacting proteins suggest a putative role for G-quadruplex as a molecular switch in regulatory mechanism^{150,151}. This signifies G-quadruplex to be an important subject for protein-based recognition studies^{89,135}.

1.3.1 RNA helicase associated with AU-rich element

A ubiquitous DEAH-box protein with specificity towards G-quadruplexes

The RNA helicase associated with AU-rich element (Rhau) is an abundantly expressed protein in the cell nucleus and cytoplasm¹⁵², and is found to be localized within stress granules¹⁵³. Rhau is a DExH protein product of DHX36 gene¹⁵², which critically regulates gene transcription^{154,155} and mRNA stability^{152,156}. Rhau is also highly expressed in testicular tissues¹⁵⁷, and found to be vital for spermatogonia differentiation via regulation of the c-kit gene¹⁵⁸.

Table 1.1 List of selected G-quadruplex-interacting proteins and their available protein-DNA/RNA structure. Data are adapted from^{80,133,159}.

	Protein	Organism	DNA/RNA -protein complex structure		Nucleic acid type&	References
			Resolved by	PDB ID		
G-quadruplex destabilizing proteins	Proteins that disrupts G-quadruplexes					
	UP1/ hnRNP A1	Human	X-ray	2UP1	ssDNA	144
	hnRNP A2	Human	NMR	1X4B	-	160,161
	AUF1/ hnRNP D	Human	NMR	1X0F/1WTB	ssDNA	143
	POT1	Human	X-ray	1XVJ/3KJP	ssDNA	78,162
	RPA	Human	X-ray	2PQA/1L1O/1FGU	-	163-165
G-quadruplex unwinding proteins	Helicases					
	Pif1	<i>Saccharomyces cerevisiae</i>	-	-	-	11,166
	FANCI	Human	-	-	-	12
	Rhau	Human	-	-	-	
	BLM	Human	X-ray	4O3M	dsDNA	167,168
		Chicken		4CGZ		
	WRN	Human	X-ray	3AAF	dsDNA	169
		Chicken				
	RecQ	<i>Escherichia coli</i>	X-ray	1OYW	-	170
	Rtel1	Human	-	-	-	25
Dna2 [#]	Human	-	-	-	171	
G-quadruplex selective nucleases	Nucleases					
	GQN1	Human	-	-	-	35
	Mre11	<i>Saccharomyces cerevisiae</i>	-	-	-	32
		<i>Saccharomyces cerevisiae</i>				
	KEM1/SEP1	<i>Saccharomyces cerevisiae</i>	-	-	-	33,34
G-quadruplex stabilizing proteins	Proteins that bind to G-quadruplexes					
	TEBPβ	<i>Oxytrichia nova</i>	X-ray	1JB7	G4	54
	Rap1	<i>Saccharomyces cerevisiae</i>	X-ray	1IGN	dsDNA	172
	FUS/TLS	Human	NMR	2LCW	-	173,174
	TRF2	Human	X-ray	1W0U/1VFC	dsDNA	175,176
	AFF2/FMR2	Human	-	-	-	177,178
	Est1A/SMG6	Human	-	-	-	179
	Nucleophosmin/ NPM1	Human	NMR	2LLH	-	180
	Nucleolin	Human	NMR	1QWA/1QWB	dsRNA	28,29,181

[&] G4 – G-quadruplex; ssDNA – single-stranded DNA; dsDNA – double-stranded DNA

[#] Dna2 also exhibits nuclease activity towards G-quadruplexes

Rhau has been shown to act as a G-quadruplex resolvase, with high affinity towards both DNA and RNA tetramolecular G-quadruplexes at K_D of 77 and 39 pM respectively¹⁴⁸. Rhau exhibited similar tight G-quadruplex binding affinity for unimolecular DNA G-quadruplex of Zic1 and c-myc origin, with K_D of 3 – 6 pM and 4 pM for the former and latter G-quadruplex respectively¹⁸². In contrast, Rhau was found to exhibit lower selectivity for canonical forms of DNA (forked duplex) and RNA (tRNA) as compared to G-quadruplexes^{148,152}. In addition, competition studies on the resolving activity of Rhau between DNA and RNA G-quadruplexes have demonstrated that Rhau resolves RNA G-quadruplexes more readily than DNA G-quadruplexes¹⁴⁸.

Using RIP-chip analysis, ~ 100 RNA were found to interact with Rhau, particularly for the TERC RNA, which associates with Rhau through a G-quadruplex assembled on its 5' end both *in vitro* and *in vivo*¹⁸³. Immunoprecipitation assay conducted has also shown the Rhau-TERC complex to associate with the telomerase holoenzyme¹⁸³.

Figure 1.10 Schematic representation of conserved domains in Rhau (residue 1–1008). The glycine-rich domain (Gly; residue 10–51), Rhau specific motif domain (RSM; residue 54–66) and helicase domain (residue 201–615) are shown as black, red and white rectangle box respectively. The conserved DExH protein helicase motifs I, Ia, Ib, II, III, IV, V, VI are indicated as grey strips within the helicase domain.

The Rhau specific motif is a G-quadruplex recognition motif

Rhau contains a conserved helicase core with helicase motif I – VI typical of DExH-box proteins. This conserved domain is flanked by its N-terminal region (NTR) containing a glycine-rich domain (Gly) and a Rhau-specific motif (RSM) domain, and a partially conserved C-terminal region (CTR)¹⁵³ (Figure 1.10).

Helicase assays were performed with different mutated Rhau proteins to identify the specific G-quadruplex resolving region within the Rhau protein¹⁸⁴. The G-quadruplex resolving functionality is completely eradicated for a Rhau mutant without its RSM domain (Δ RSM) but not in Δ Gly mutant, which clearly establishes the RSM domain (residues 54–66) as an essential component in the dissolution of a G-quadruplex by Rhau. The importance of the RSM domain in

G-quadruplex interactions was further emphasized in electromobility shift assays of Δ Gly and Δ RSM, where only the former was found to bind G-quadruplexes. As helicase assays conducted on Rhau₁₋₁₀₅, which encompasses the RSM domain, has found this truncated Rhau fragment to be insufficient to dissolve G-quadruplexes, this indicates that the RSM domain behaves purely as a G-quadruplex binding region. Thus, the RSM domain has to work cooperatively with the helicase domain for Rhau to bind and unwind G-quadruplexes effectively. This cooperative effect also affects the binding affinity of Rhau with G-quadruplexes, as exemplified by the weaker binding affinity of the truncated Rhau₁₋₂₀₀ fragment (NTR of Rhau) which exhibited K_D 440 nM to RNA G-quadruplex, in contrast to the stronger binding affinity of K_D 14 nM as displayed by the wild-type Rhau¹⁸⁴. A similar NTR fragment Rhau₅₃₋₁₀₅, was also determined to bind RNA with K_D of 310 nM¹⁸⁵. However, the lack of a high resolution structure of a G-quadruplex-Rhau complex prevents the identification of the binding motif for both G-quadruplex and Rhau RSM domain.

1.3.2 Bloom syndrome protein, a member of the RecQ family of helicases

Bloom syndrome and Bloom syndrome protein

Bloom syndrome (BS), a rare genetic disorders characterized by an increased prevalence of cancer and/or early onset of aging-related diseases has been widely studied in recent years due to chromosomal instability linked with the knockout or alteration of its gene^{141,186,187}. Bloom syndrome protein (BLM), the protein associated with BS, is a 3'–5' DNA helicase belonging to the RecQ family, a subset of the DExH-box group of helicases^{188,189}. Named after its prototypical member¹⁹⁰, RecQ helicases are conserved from *E. coli* and lower eukaryotes to mammals, with highly conserved domains consisting of the helicase domain, the zinc-binding domain (ZnBD), the RecQ C-terminal (RQC) domain, and the Helicase and RNaseD C-terminal (HRDC) domain as illustrated in Figure 1.11. Within the human body, there exist four other BLM homologues: RecQ1 (RecQL1), WRN, RecQ4 and RecQ5 (RecQL5). Mutations in WRN results in the Werner syndrome, with RecQ4 associated with Rothmund-Thomson syndrome, RAPADILLINO syndrome and Baller-Gerold syndrome¹⁹¹. Mutations in RecQ1 and RecQL5 are not associated with any latent disease thus far¹⁹¹. The remaining RecQ proteins are found in *E. coli* (RecQ) and *S. cerevisiae* (Sgs1)¹⁹¹.

Figure 1.11 Schematic representation of conserved domains in RecQ family. Conserved domains of RecQ proteins are illustrated as rectangle boxes. Proteins are aligned by the conserved RQC domain (coloured in blue), which contains the winged-helix motif. Zinc-binding domain is shown in black. Domain alignment was adapted from^{169,191}.

BLM and DNA resolution/repair

With its primary location within the nucleus¹⁹², a key function of the BLM protein in cell is its role in DNA replication and repair¹⁹³: the resolution of B-DNA by BLM in a 3' - 5' direction was first presented by Karow *et. al.*¹⁹⁴, and was later found by Langland and colleagues to be critical for the repair of double-strand breaks¹⁹⁵. Moreover, it acts in tandem with the scaffolding protein involved in DNA repair (SPIDR) during homologous recombination repair¹⁹⁶. The DNA repair functionality of BLM has also been exemplified in stalled DNA replication forks, where BLM and FANC proteins were jointly recruited in the response pathway¹⁹⁷. BLM is able to unravel Holliday junction when complexed with topoisomerase III α , thus suppressing sister chromatid exchange¹⁹⁸. BLM has been implicated to have putative roles in cell replication, double strand break repair and telomere maintenance¹⁹¹. Numerous *in vivo* tests conducted on BLM with WRN have demonstrated BLM to display tumour suppressor-like roles^{199,200}.

Selectivity for G-quadruplexes over canonical DNA forms by the BLM protein

Aside from Rhau²⁰¹, other proteins such as the RecQ helicases²⁰², Pif1 helicase¹⁶⁶, FANCI helicase¹² and XPB/XPD transcriptional helicases⁹² were found to be specific for resolution of G-quadruplex in human cells. Of the five members within the RecQ family, human BLM and WRN have been shown to found to possess exceptional binding and unwinding capabilities for G-quadruplex over single-stranded DNA and duplex DNA^{202,203}. In the same paper, the screening of

various truncated BLM fragments has identified the combined zinc-binding and RQC domain (BLMZRQC) to be the key G-quadruplex binding region (K_D 2.7 nM) within the BLM protein.

Current structures available on RecQ helicases (*E. coli* RecQ, RecQ1, BLM; PDB ID: 1OYW¹⁷⁰, 2WWY/2V1X²⁰⁴, 4O3M¹⁶⁷/4CGZ¹⁶⁸ respectively) has found the topology of ZnRQC to be highly conserved, with a helix-turn-helix motif for the ZnBD, and a wing-helix motif for the RQC domain. Comparison of available DNA-RecQ helicase complexes has shown RecQ1 (PDB ID: 2V1X) and BLM (PDB ID: 4O3M; 4CGZ) to utilize similar DNA-binding mode for binding interaction to a 3' tailed duplex^{167,168,204,205}, with the bulk of the interaction localized on the RQC domain as exemplified by the co-crystal structure of WRNRQC with duplex DNA (PDB ID: 3AAF¹⁶⁹). The ZnRQC domain was similarly identified in *E. coli* RecQ via electromobility shift assay (EMSA) as a core DNA binding region, although the preferential G-quadruplex binding exhibited in BLMZRQC has not been replicated⁹⁰. However, the importance of RQC domain in DNA binding was countered by Gyimesi and coworkers, where a minimal functional module comprising of the helicase domain and zinc-binding domain was determined to be sufficient to perform efficient DNA strand separation, with the excision of the RQC domain having no detrimental effect on its unwinding functionality²⁰⁶.

As a BLM-G-quadruplex complex has yet to be resolved, and current structural and biochemical data could not explain for the preferential G-quadruplex binding capability exhibited by BLMZRQC, it is critical to identify the structural motif involved in the binding of BLM to G-quadruplex, which may provide insight into the working mechanism of BLM (and also WRN) in telomeric maintenance.

1.3.3 Structural data of G-quadruplex complex with proteins

Numerous proteins have been identified to have interactions with G-quadruplex^{135,151}, with several experimental and computational studies highlighting potential G-quadruplex specific recognition motifs^{95,207-209}. However, the specific interactions between G-quadruplexes and proteins have only been documented in a few structures to date.

Horvath and Schultz reported the co-crystallization of G₄T₄G₄ G-quadruplex with *Oxytricha nova* telomere end binding protein (OnTEBP; Figure 1.12A)⁵⁴. However, the protein-DNA contacts were suspected to be induced by crystallization packaging⁵⁴.

A DNA-protein complex comprising of G-quadruplex-based aptamer d(GGTTGGTGTGGTTGG) bound between two thrombin molecule was resolved by both X-ray crystallography and NMR spectroscopy²¹⁰⁻²¹² (Figure 1.12B). However, the focus of the binding interactions by the thrombin molecules was on the flanking TT loops of the aptamer, which was exemplified by the lack of DNA-protein complex formation for the derivative of the aptamer missing its TT flanking ends²¹¹. Moreover, aptamer with mutation of T3 (PDB ID 4LZ4) or T12 (PDB ID 4LZ1) with an abasic spacer still formed a co-crystal with thrombin, with the key DNA binding interface of thrombin, the A-region, targeting the non-mutated T loop²¹³.

Another G-quadruplex-protein complex reported in literature is the co-crystallized pseudo-quadruplex-duplex hybrid with thrombin (PDB ID 4I7Y)²¹⁴. Examination of this structure shows the exosite II of thrombin to have extensive contact with the sugar phosphate backbone, and both edgewise loops of the pseudo-quadruplex motif, with additional anchoring of the DNA structure by its T24 due to its flipped-out orientation. Similar to the previous thrombin-aptamer complex, the thymine bases in these loops appear to be the key anchoring factor for the aptamer onto thrombin, with T9, T18 and T24 all having the closest proximity to the thrombin exosite II.

An arginine-glycine-rich FMRP RGG peptide bound to a duplex-quadruplex hybrid formed from *scI* RNA²¹⁵ (Figure 1.12C) has been reported as a G-quadruplex-peptide complex. However, the binding interaction between the peptide and duplex-quadruplex appeared to be dependent on the duplex strand and the duplex-quadruplex junction instead of G-quadruplex recognition by the peptide.

Katahira and group also reported a solution structure of peptide-RNA complex formed by a short peptide derived from a normal cellular form of prion proteins, with a RNA aptamer that adopts G-quadruplex²¹⁶. However, the peptide appeared to be anchored onto the RNA G-quadruplex by two anchor points, where a tryptophan residue appears to have contribute to the

Figure 1.12 Selected structures of G-quadruplex-protein complex. A) Crystal structure of OnTEBP co-crystallized with DNA G-quadruplex (PDB ID: 1JB7)⁵⁴. B) Crystal structure of G-quadruplex-aptamer co-crystallized with thrombin (PDB ID: 3QLP)²¹¹. C) Solution structure of *scI* RNA G-quadruplex complexed with FMRP RGG peptide (PDB ID: 2LA5)²¹⁵. DNA-protein contact surface are magnified for clarity.

binding via stacking interactions. On the other hand, three lysine residues are involved in electrostatic interaction with the sugar phosphate backbone, in which these binding interactions are non-specific towards G-quadruplexes.

Despite the successful incorporation of G-quadruplexes as part of the above discussed complexes and in other structures^{180,217}, factors other than G-quadruplex-specific binding interactions appeared to have contributed to the formation of the G-quadruplex-protein complex. Thus, the G-quadruplex-specific molecular recognition motifs still remain to be identified and resolved.

1.4 Aims and Objectives

The overall goal of the research presented in this thesis is to elucidate the structural features of G-quadruplexes and G-quadruplex-binding proteins that contribute to the molecular recognition of G-tetrads, and the binding interface of G-quadruplex-protein complex. This goal can be separated into the following objectives:

Objective 1: Formation of G-quadruplexes with DNA-damaged sequences:

- To determine the impact of single and double mutation(s) of 8-oxoguanine/xanthine on G-quadruplexes
- To solve the high resolution solution structure of a double-modified G-quadruplex
- To engineer modified G-tetrads using double modifications of 8-oxoguanine and xanthine

Objective 2: Identification of G-quadruplex structural recognition motif on a Rhau protein fragment

- To resolve the structure of the short Rhau fragment, Rhau₅₃ (residues 53–105)
- To resolve the complex formed by Rhau₅₃, using a well-defined G-quadruplex structure previously determined by our laboratory
- To determine the G-quadruplex molecular recognition principles of proteins by studying DNA-protein interaction via NMR spectroscopy

Objective 3: Elucidation of the structural motifs of BLM protein that determine selectivity for G-quadruplexes over duplexes

- To screen different conserved domains of BLM with a variety of G-quadruplexes of different conformation, and compare these binding affinity with that of duplex DNA.
- Using NMR titration, we intend to map the interactions observed during titration for both G-quadruplex and duplex on the resolved structure of BLM to understand the differences between their binding interfaces.

Chapter 2 Methodology

2.1 Sample Preparation

2.1.1 DNA synthesis

Standard oligonucleotide synthesis

Unlabelled and 4%-¹⁵N-enriched oligonucleotides were synthesized on ABI 394 DNA/RNA synthesizer, with standard DNA phosphoramidites, solid supports and additional reagents purchased from Glen Research Corporation. ¹⁵N-enriched phosphoramidites were purchased from Cambridge Isotope Laboratories. Cleavage of the oligonucleotides from the solid support and deprotection was achieved using the manufacturer's protocol, and purified using PolyPak II cartridges (Glen Research). Purified oligonucleotides were successively dialyzed in deionized water, 25 mM KCl buffer, and again in deionized water, and freeze-dried. Oligonucleotides were dissolved in buffer comprising 20 mM potassium phosphate (Kpi), 70 mM KCl, 10% D₂O, pH 5.0 – 9.0. *T95-2T* samples were prepared similarly except the dialysis was done in 5 mM KCl buffer, and the lyophilized sample resuspended in 5 mM Kpi, 5 mM KCl, pH 7.0, 10 % D₂O buffer. The concentration of each DNA sample is expressed in strand molarity.

Extinction coefficient of the modified sequences was approximated to that of the unmodified sequence.

¹⁵N, ¹³C-oligonucleotide synthesis

Figure 2.1 Schematic diagram for the synthesis of ¹⁵N, ¹³C-labelled oligonucleotide via hairpin extension. All ¹⁵N, ¹³C-substrates are indicated in red.

¹⁵N, ¹³C-labelled T95-2T was prepared according to protocols as described by Zimmer and Crothers²¹⁸ with modifications as illustrated in Figure 2.1. Briefly, a hairpin template with a 5' tail complementary to T95-2T, T95-2T-Tmp, was incubated with 3-fold of ¹⁵N, ¹³C-dGTP/dTTP with Klenow fragment over four days at 37 °C. Denaturing PAGE was used to verify for the full extension of the hairpin. The reaction was stopped via heat denaturation at 95 °C for 10 min. ¹⁵N, ¹³C-labelled T95-2T was subsequently cleaved from the hairpin template by uracil DNA glycosylase (UDG; New England Biolabs) after overnight incubation at 37 °C, and purified by reverse-phase HPLC with the following gradient: A: 0.1 M TEAA p H 6.8; B: 100% acetonitrile; 4 – 8% B (8 min); 8 – 20% B (45 min). Fractions containing ¹⁵N, ¹³C-labelled T95-2T were verified by denaturing gel electrophoresis. These fractions were subsequently pooled, adjusted to pH 11.0, before dialyzing successively against deionized water, 10 mM KCl and water again. The samples were lyophilized and resuspended in 5 mM Kpi, 5 mM KCl, pH 7.0, with 10 % D₂O.

Fluorescent oligonucleotide synthesis

All fluorescent-tagged oligonucleotides were purchased from IDT (Singapore) or AITBiotech. The oligonucleotides were prepared in 100 μ M stocks in buffer containing 20 mM Kpi, 70 mM KCl, pH 7.0, and annealed at 95 °C. For fluorescent-labelled *T95-2T*, the sample is rehydrated and the pH adjusted to 11.0 with 1 M LiOH. Fluorescent-tagged *T95-2T* was dialyzed successively against water, 5 mM KCl and water again, and lyophilized before dissolving in the same buffer as other fluorescent oligonucleotides. The fluorescent oligonucleotides were diluted with the same buffer to the desired concentration when required.

2.1.2 Peptide and protein synthesis

2.1.2.1 Production of Rhau peptide

Preparation of recombinant plasmid

Synthesis of the recombinant plasmid of Rhau₅₃ (residues 53–105) and Rhau₁₄₂ (residues 53–192) was done by NTU/SBS Protein Production Platform (NTU/SBS PPP)²¹⁹. Briefly, the gene encoding for N-terminal region of Rhau was subcloned from IRAT cDNA library into pNIC28-Bsa4 (Appendix A) via ligation independent cloning (LIC). The length of each insertion was adjusted through the use of appropriate primers during polymerase chain reaction (PCR). These recombinant plasmids were transformed into competent cells of BL21 (DE3) T1R and Rosetta T1R features (Rosetta T1R; Protein Production Platform), and verified by DNA sequencing. Each of these constructs encodes for the following: N-His₆-TEV-Rhau-C.

To create the recombinant plasmid of Rhau₁₈, residues 53–70 was subcloned into pET32Ek/LIC (Appendix A; Merck), using recombinant plasmid of Rhau₅₃ as a template. The pET32Ek/LIC vector contains an N-terminal His₆ with thioredoxin (Trx) fusion tag. A forward primer with the sequence: 5'-GACGAC AAGATGGAGAACCTGTACTTGCA-3' with an additional TEV cleavage site (ENLYFQS), and the reverse primer: 5'-GAGGAGAAGCCCGGTTACTGTTTTTCGCGTACCA-3' were used in polymerase chain reaction (PCR) to amplify the target insert. All primers were purchased from IDT (Singapore). Following agarose gel purification of the PCR reaction mixture by QIAQuick Gel Extraction Kit (QIAGEN), the purified insert was treated by T4 DNA polymerease (Promega) to form the LIC

site, and annealed with a similarly treated linearized pET32Ek/LIC vector to form the recombinant plasmid pR18c2T. pR18c2T was subsequently inserted into Rosetta T1R competent cells, and the resulting reaction mixture grown on LB agar plates containing 50 µg/mL carbenicillin. Single colonies obtained after overnight incubation at 37 °C was first screened via PCR for correct insertion of pR18c2T, and verified by sequencing (1stBase, Singapore).

Protein expression and purification

Unlabelled/¹⁵N-labelled/¹⁵N, ¹³C-labelled Rhau₅₃ (residues 53–105) were initially prepared by NTU/SBS PPP. Cells were grown using LEX Bioreactor (Harbinger Biotech) in Terrific Broth (TB; with 8 g/L glycerol) or M9 minimal media supplemented with 100 µg/mL ampicillin at 37 °C until OD₆₀₀ 2.0 was reached. Uniform isotopic-labelling of all proteins were achieved by growing cells in minimal media containing 0.5 g/L ¹⁵NH₄Cl and/or 2 g/L [¹³C₆]-D-glucose. Protein expression was induced through addition of 0.5 mM isopropyl-β-D-thiogalactoside (IPTG), incubating overnight at 18 °C. Cells were harvested by centrifugation at 4,200 rpm at 15 °C for 15 min. Following resuspension of cell pellet in buffer containing 100 mM HEPES, 0.5 M NaCl, 10 mM imidazole, 10 % (w/v) glycerol, 0.5 mM TCEP, pH 8.0, 1.3 U/L Benzonase (Merck), Protease Inhibitor Cocktail (Calbiochem), the cells were lysed by sonication and the suspension clarified through ultracentrifugation at 47,000 x g at 4 °C for 25 min. The His₆-tagged protein was purified by Ni-NTA column (Qiagen) pre-equilibrated with 20 mM HEPES, 0.5 M NaCl, 10 % (w/v) glycerol, 0.5 M TCEP, pH 7.5, washing extensively at 10 mM and 25 mM imidazole concentration before eluting at 500 mM of imidazole. The His₆-tag was cleaved by overnight digestion at 4 °C using Tobacco Envelope Virus (TEV) protease (NTU/SBS PPP). The His₆-tag and TEV protease was removed via application of digested protein to a second round of Ni-NTA chromatography, and the free Rhau₅₃ was pooled. Further purification was carried out with gel filtration chromatography on HiLoad 16/60 Superdex 75 GL pre grade column (GE Healthcare) pre-equilibrated with 20 mM HEPES (pH 7.5), 0.5 M NaCl, 0.5 mM TCEP, 10 % (v/v) glycerol. Rhau₅₃ concentrated down using 3000 MWCO centricon (Sartorius), washed with 20 mM Kpi, 70 mM KCl, 10 % (v/v) D₂O, pH 6.6 and flash frozen in liquid nitrogen before storing at -80 °C. This protein expression and purification method yielded 0.2 – 1.5 mg of labelled

protein from 1 L of culture with ≥ 95 % purity as ascertained by SDS-PAGE and MALDI-TOF mass spectrometry (Appendix D.1).

Figure 2.2 FPLC purification of ¹⁵N,¹³C-Rhau₁₈. FPLC trace obtained from A) affinity chromatography (HistrapHP), with peptide eluting at 0.5 M imidazole; B) cation exchange chromatography (HiTrapSP) eluting at 0.15 M KCl; C) gel filtration chromatography (Superdex 75 10/300 GL). Fractions collected and pooled are marked in red.

Subsequent unlabelled, ¹⁵N-labelled and ¹⁵N,¹³C-labelled Rhau₅₃ protein production was performed in-house, with the following modifications: cells were grown at 37 °C, shaking at 220 rpm till OD₆₀₀ 0.6 – 0.8 before induction with addition of 0.5 mM IPTG. Following 20 h incubation at 18 °C, the cells were pelleted at 10,000 rpm, 4 °C for 15 mins, and resuspended in 20 mM Kpi, 0.5 M KCl, 10 mM imidazole, 10 % (v/v) glycerol, Protease Inhibitor Cocktail (Roche), pH 7.5 buffer and lysed via sonication. The crude lysate was clarified at 30,000 rpm, 4 °C for 15 mins before loading onto HisTrap HP column (GE Healthcare) pre-equilibrated with

20 mM Kpi, 0.5M KCl, 10 % (w/v) glycerol, pH 7.5. The column was washed extensively at 10 mM and 25 mM imidazole before eluting the protein at 100 mM imidazole. The His₆-tag was cleaved off from Rhau₅₃ after overnight digestion at 4 °C by TEV protease, and separated from the peptide pool via cation chromatography using HiTrap SP HP column (GE Healthcare) pre-equilibrated with 20 mM Kpi, 10 % (v/v) glycerol, pH 7.5. Rhau₅₃ was further purified through gel filtration chromatography using Superdex 75 10/300 GL column (GE Healthcare) pre-equilibrated with 20 mM Kpi, 70 mM KCl, 10 % (v/v) glycerol, pH 7.5 to obtain protein of $\geq 95\%$ purity as verified by MALDI-TOF spectroscopy (Appendix D). Fractions containing Rhau₅₃ was combined and concentrated to 0.5 – 1 mL using 3000 Daltons (Da) molecular weight cut-off (MWCO) centricon (Millipore), washing with 20 mM Kpi, 70 mM KCl, 10 % (v/v) D₂O, pH 6.6 buffer before flash freezing with liquid nitrogen. 0.6 – 0.8 mg of unlabelled/labelled protein was produced per litre of media.

Unlabelled, ¹⁵N-Rhau₁₈ and ¹⁵N,¹³C Rhau₁₈ were produced and purified according to the above described procedure with slight modifications: Trx-His₆-Rhau₁₈ was applied to HisTrap HP column (GE Healthcare), washed with stepwise gradient of 10 mM and 20 mM imidazole prior to elution at 200 mM imidazole. Trx-His₆ tag was removed via TEV protease and separated by cation exchange chromatography. The final peptide fragment contains Rhau₁₈, with an N-terminal serine and methionine (SM), a by-product from gene cloning and TEV cleavage. Rhau₁₈ peptide was pooled and loaded onto Superdex 75 10/300 GL column for further purification. Buffer exchange of the peptide was carried out with 20 mM Kpi, 70 mM KCl,

Figure 2.3 SDS-PAGE of ¹⁵N,¹³C-Rhau₁₈, illustrating different purification stages. Lane 1, 5: PageRuler Unstained protein ladder; lane 2: Trx-His₆-Rhau₁₈ after affinity chromatography; lane 3: reference Rhau₁₈ peptide from solid phase synthesis; lane 4: impure Rhau₁₈ after TEV digestion; lane 6: purified Rhau₁₈ after anion exchange and gel filtration chromatography.

10 % D₂O pH 6.6 buffer using 2000 Da MWCO centricon (Sartorius). Rhau₁₈ was quantified by Bicinchoninic acid assay (BCA assay; Pierce), flash frozen with liquid nitrogen and stored at -80 °C. The peptide was characterized by SDS-PAGE and MALDI-TOF, and found to be ≥ 95 % pure (Appendix D). 5 – 10 mg of unlabelled/labelled peptide was obtained per litre of media.

Solid phase synthesis of peptides

All peptides used in this study were purchased from Peptide Synthesis Core Facility (NTU SBS). The lyophilized samples were resuspended in buffer containing 20 mM Kpi, 70 mM KCl, 10 % D₂O, pH 7.0 and quantified by BCA assay.

2.1.2.2 Production of BLM protein

Preparation of recombinant plasmids

Using LIC, recombinant plasmids encoding for BLMh (residues 642–991), BLMhZ (residues 642–1069), BLMhZRQC (residues 642–1192), BLMZ (residues 992–1069), BLMZRQC (residues 992–1192) and BLMRQC (residues 1070–1192) were prepared by NTU/SBS PPP through the insertion of their respective target gene from IRAT cDNA into pNIC28-Bsa4 vector. An unnatural gene where a TEV cleavage site is inserted between the helicase and Zn binding domain (pBLMhTevZRQC) was purchased from AITBiotech (Singapore), and subcloned into pNIC28-Bsa4 vector (Appendix A).

Plasmids encoding for N-TrxBLMZ-C and N-TrxBLMZRQC-C were made through the insertion of their respective target gene into pET32Ek/LIC vector to form recombinant plasmid pBLMzLIC and pBLMzrLIC (Appendix A). Plasmid of BLMZRQC was used as the DNA template, and the primers used in the PCR mixture are listed in Appendix B. Plasmid was sequenced (1stBase, Singapore) and verified to be correct.

The plasmid encoding for N-GSTBLMZRQC-C was prepared with the same LIC technique in pNIC-GST vector (Appendix A; NTU/SBS PPP).

All the above described plasmids of BLM were transformed in *E. coli* Rosetta T1R competent cells.

Figure 2.4 Schematic diagram of PCR-based fusion of two gene fragments used to construct GSTekBLMZRQC recombinant plasmid. GST-tag with a C-terminal Enterokinase cleavage site and the BLM gene fragment are joined together using a two-step amplification PCR strategy²²⁰. Sequences of primers used in this procedure are listed in 5' - 3' directionality.

To prepare the recombinant plasmid for transfection into baculovirus (*Sf9*) cells, an additional GST-tag with an insertion of enterokinase (Ek) cleavage site was cloned into the N-terminal of the gene encoding for BLMZRQC via long PCR-based fusion of two fragments (Figure 2.4)²²⁰, and the resulting gene fragment was inserted into pFB-LIC-Bse (Appendix B). The resulting recombinant plasmid, pGSTekBLMzr, transcribes for N-GST-BLMZRQC-C. N-GST-BLMZ-C was also subcloned from this plasmid.

Sequence alignment

Structure-based multiple sequence alignment was conducted for RQC domain of human BLM, WRN, RecQ1, *Escherichia coli* RecQ and *Saccharomyces cerevisiae* Sgs1 using ClustalX and refined manually. Sequences used were retrieved from NCBI, (protein accession no. *Homo sapiens* BLM: P54132; *Homo sapiens* WRN: Q14191; *Homo sapiens* RecQ1: P46063; *Escherichia coli* RecQ: P15043; *Saccharomyces cerevisiae* Sgs1: P35187), and Pfam (family ID 09382) database.

Protein expression and purification of BLM constructs

All derivatives of BLM except N-TrxBLMZRQC-C and N-TrxBLMZ-C were screened for soluble expression and the high scale protein productions were performed by the Protein Production Platform at SBS/NTU.

For the soluble expression screening, cells that express the desired proteins were grown in 1 mL of Terrific Broth at 37 °C prior to IPTG induction at 18 °C for 20 h. The cells were subsequently harvested and purified by Ni-NTA chromatography as described for the first protocol for purification of Rhau₅₃, prior to analysis by SDS-PAGE.

For large scale protein production, the proteins were expressed and purified by Ni-NTA chromatography and gel filtration chromatography as described earlier. N-TrxBLMZRQC-C and N-TrxBLMZ-C were expressed and purified according to protocols as described for Rhau₅₃. All BLM proteins were desalted into 20 mM Kpi, 70 mM KCl, 10 % (v/v) glycerol, pH 7.5 buffer using centricons of appropriate MWCO (Millipore/Sartorius), and subsequently concentrated down. Small aliquots of the proteins were made and flash frozen in liquid nitrogen before storing at -80 °C. Unlabelled, ¹⁵N-labelled and ¹⁵N,¹³C-labelled were prepared for BLMRQC using the same protocol as described for Rhau₅₃ with the following modification: the final purified BLMRQC was desalted using 5000 Da MWCO centricon (Sartorius) into 20 mM Kpi, 70 mM KCl, pH 7.5 buffer prior to all NMR experiments. 10 – 20 mg of unlabelled/labelled protein was produced per litre of media for BLMhZRQC, BLMhZ, and BLMRQC as quantified by BCA assay. The remaining constructs formed inclusion bodies and were not produced for further characterization.

2.1.2.3 Production of WRNRQC

Preparation of recombinant plasmid

The recombinant plasmid encoding for WRNRQC (residues 949–1079) within a pGEX-5X-1 vector was kindly provided by Dr Ken Kitano. The WRN RQC gene was subsequently subcloned into pNIC28-Bsa4 vector, transformed into Rosetta T1R competent cells, and verified to be correct via DNA sequencing. The final plasmid encodes for N-His₆-WRNRQC-C.

Protein expression and purification of WRNRQC

Unlabelled and ^{15}N , ^{13}C -labelled WRNRQC were expressed and purified by NTU/SBS PPP according to standard protocols as described above. Protein was desalted into 20 mM Kpi, 70 mM KCl, 10 % glycerol, pH 7.5 buffer using 3000 Da MWCO centricon (Millipore), and quantified by BCA assay, with 10 – 20 mg of unlabelled/labelled protein was produced per litre of media.

2.2 NMR Spectroscopy

2.2.1 Standard NMR experiments for structure elucidation

X•O-modified G-quadruplexes

NMR experiments were performed on Bruker AVANCE 400-MHz, 600-MHz and 700-MHz at 25 °C unless stated otherwise. DNA strand concentration was 0.1 – 0.5 mM. All chemical shifts are calibrated against 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS).

For all the NMR experiments in H_2O , jump-and-return (JR) suppression scheme²²¹ have been used to detect the fast-exchangeable imino protons. The delay for the JR has been optimized to maximize the signal at the imino proton region. Signal belonging to the region 10.5 – 12.5 ppm is indicative of formation of $\text{N}(\text{H})\cdots\text{O}$ hydrogen bond pairing the G in the G-quadruplex (Figure 1.1a). ^{15}N -filtered experiment $\{^{15}\text{N}\}$ -JR-HMQC was used to site-specifically assign H1 and H8 resonances of samples containing one guanine nucleotide enriched with 4% ^{15}N isotope label²²². Through-bond correlation experiments such as TOCSY, COSY, $\{^1\text{H}-^{13}\text{C}\}$ -HSQC and $\{^1\text{H}-^{15}\text{N}\}$ -HSQC at natural abundance^{223,224} were used to assign all the proton resonances. Through space correlation experiment NOESY was used to assign/confirm the proton resonances. Distances were extracted using NOESY in H_2O (mixing time, 300 ms) and in D_2O (mixing time, 100 ms and 300 ms). All spectra are processed by Topspin 2.1 and analyzed by SPARKY 3.115²²⁵.

BLMRQC

Spectra were recorded for 0.3 – 1 mM samples at 25 °C for ^{15}N , ^{13}C -BLMRQC samples in 20 mM Kpi, 70 mM KCl, 10 % D_2O , pH 7.5. 2D $\{^1\text{H}-^{15}\text{N}\}$ -HSQC and 3D HNCO were performed as a starting point to monitor the number of H^{N} resonances to determine the optimal sample conditions (pH and temperature) for structure resolution by NMR. Preliminary sequential backbone assignment of BLMRQC was performed by AutoAssign Webserver^{226,227} using 3D HNCACB, CBCA(CO)NH, HNCO and HN(CA)CO. Sequential connectivity of BLMRQC was

also verified by $^1\text{H}\alpha$ and $^1\text{H}\beta$ assignment from HBHA(CO)NH and HBHANH experiments. Side chain assignments were made extensively with 2D $\{^1\text{H}-^{13}\text{C}\}$ -HSQC, and 3D CC(CO)NH, HCC(CO)NH, and HCCH-TOCSY, and further assisted by 3D HNHA, HCCH-COSY, ^{13}C -edited NOESY and ^{15}N -edited NOESY. Side chain assignment of aromatic residues was determined by 2D (HB)CB(CGCD)HD and (HB)CB(CGCDCE)HE experiments. NH_2 side chain assignments were confirmed using NH_2 multiplicity-edited 2D $\{^1\text{H}-^{15}\text{N}\}$ -HSQC experiment.

Rhau₅₃, Rhau₁₈ and Rhau₁₈-T95-2T complex

All NMR experiments were performed on 600 and 700 MHz Bruker spectrometers at 15 or 25 °C. Concentrations of the free Rhau₅₃/Rhau₁₈ and bound Rhau₁₈ samples were typically between 0.1 – 1 mM, in 10 mM KCl, 10 mM potassium phosphate, pH 6.5 and 5 – 10% D₂O. Similar to BLMRQC, NMR experiments for assignments were performed on proteins. For the free and bound Rhau₁₈, 89.3% and 71.4% of backbone ^{15}N , ^1HN , $^1\text{H}\alpha$, $^{13}\text{C}\alpha$, and $^{13}\text{C}'$ resonance assignments have achieved through classical triple resonance NMR. All spectra are processed by Topspin and analyzed by SPARKY 3.115²²⁵.

2.2.2 NMR Titration

NMR pH titration

All DNA samples used in pH titration experiments were prepared at 0.15 – 0.2 mM DNA concentration, with the required pH adjusted using HCl and/or LiOH. Each sample was allowed to equilibrate at room temperature for 5 mins before the NMR spectrum was recorded as described above.

NMR Chemical Shift Perturbation Assay (NMR CSPA)

For 2D $\{^1\text{H}-^{15}\text{N}\}$ -HSQC titration, NMR experiments were performed on 600 MHz and 700 MHz Bruker spectrometer equipped with a CryoProbe. Spectra were recorded at 25 °C unless stated otherwise. For NMR titration of BLMRQC, DNA samples *HT*, *T95-2T*, *CEB25*, and *DXI* were added to 0.2 mM ^{15}N -BLMRQC at protein:DNA ratios 1:0, 1:1, 1:2, 1:4, and the spectra were recorded at each titration point. For Rhau₅₃ and Rhau₁₈, ^{15}N -labelled protein samples were titrated with equimolar or 2-fold excess of DNA and the spectra were recorded at 15 °C. *HT* and

T95-2T were used in NMR titration of Rhau proteins. DSS was used as a reference in all NMR titration experiments. Combined ^{15}N - ^1H chemical perturbation were calculated using the equation:

$$\Delta\delta_{obs} = \sqrt{(\delta^{15}\text{N})^2 + (\delta^1\text{H})^2} \quad 2.1$$

2.3 Circular Dichroism spectrophotometry

For samples in the X/O study in Chapter 3, circular dichroism (CD) spectra were recorded with 5 μM samples in buffer containing 10 mM Kpi, 10 mM KCl, pH 7.0 – 9.0, with a total reaction volume of 450 μL on a JASCO-815 spectropolarimeter at 20 $^{\circ}\text{C}$ over the wavelength range of 220 – 320 nm using a 1-cm pathlength quartz cuvette. Each spectrum was averaged from an accumulation of 3 scans, with the baseline correction and zero-correction at 320 nm. Samples characterized by CD spectroscopy in Chapter 4 were prepared in buffer containing 20 mM Kpi, 70 mM KCl, pH 7.0 – 9.0.

For samples in the characterization of Rhau, the above protocol was utilized with the following modifications: samples were prepared in 20 mM Kpi, 70 mM KCl, pH 6.5 buffer of 500 μL volume, and recorded over the wavelength range of 180 – 250 nm.

2.4 Ultraviolet spectrophotometry

Ultraviolet (UV) melting experiments of oligonucleotides were performed on JASCO V-650 spectrophotometer using a 1-cm pathlength quartz cuvette, monitoring the UV absorbance at 295 nm over a temperature range of 15 – 90 $^{\circ}\text{C}$. Samples were prepared in 450 μL reaction volume, with 5 μM DNA in buffer containing 10 mM Kpi, 10 mM KCl, pH 7.0, unless stated otherwise. The heating and cooling rates were set at 0.2 $^{\circ}\text{C}/\text{min}$. All samples were heated to 90 $^{\circ}\text{C}$ for 15 min prior to the start of these experiments. All spectra were normalized by two baselines corresponding to the completely folded (low temperature) and fully unfolded (high temperature) state. The melting temperature (T_m) was defined as the temperature at which there were equal populations of the folded and unfolded species. T_m values were averaged from both the heating and cooling curve.

2.5 Gel Electrophoresis

2.5.1 Agarose gel electrophoresis

1 – 3.5 % agarose gels were prepared in 1X TAE buffer. DNA samples were mixed at 6:1 ratio with loading dye (Thermo Scientific) and electrophoresed at 120 V for 30 mins in 1X TAE running buffer. GeneRuler 100 bp plus or 1kb DNA ladder (Thermo Scientific) was used as a reference in each gel. Gels were stained with SYBR SAFE DNA gel stain (Life Technologies) prior to visualization (AlphaImager; EM-Green filter).

2.5.2 Sodium dodecyl sulfate Polyacrylamide Gel Electrophoresis

Protein samples were analyzed for their molecular weight based on standard protocol for Laemmli buffer system²²⁸ via sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE). Briefly, protein samples were mixed at 1:1 ratio with SDS reducing buffer, incubated at 95 °C for 5 min to denature the protein sample, before loading into 4 % stacking/12 % resolving gel (15 % for Rhau₁₈). Using tris-tricine-SDS electrophoresis buffer, 240 V was applied to the gel for 25 min before staining with Coomassie Brilliant Blue. Gels were visualized with AlphaImager HP System machine, upper white light. An example of a SDS-PAGE gel is illustrated in Figure 2.3.

2.5.3 Denaturing gel electrophoresis

DNA samples for denaturing gel electrophoresis experiments were mixed 1:1 with DNA loading buffer containing 40 % (w/v) sucrose with 10 M urea and heated at 95 °C for 1 h before loading onto 20 % polyacrylamide gels containing 7 M urea. Constant voltage of 240 V was applied to the gel for 2 h, and the gel was subsequently visualized with either by SYBR GOLD DNA stain (Life Technologies) or UV shadowing.

2.5.4 Electromobility Shift Assay

Interaction between Rhau peptide and DNA/ RNA G-quadruplexes

For the electromobility shift assay (EMSA), in a 20 µL reaction mixture, 100 nM of fluorescent DNA was incubated at room temperature (25 °C) with 0 – 50-fold of Rhau in buffer containing 20 mM Kpi, 70 mM KCl, pH 6.6 for 30 min. The reaction was stopped by the

application of 2 μL of 50 % (v/v) glycerol, and the entire sample was applied to a 20 % polyacrylamide gel containing 10 mM Kpi at pH 7.5 with 1 X TBE with 10 mM Kpi, pH 7.5 as the running buffer. The gel was run at constant voltage of 120 V for 2 h at room temperature prior to visualization on the AlphaImager HP system using UV transilluminator of 302 nm and fluorescein filter. Gel images were processed by the programme ImageJ¹³¹.

Figure 2.5 Example gel shift image demonstrating DNA-protein complex formation with DNA. Free duplex DNA is shown in lane 1, and DNA in complex with WRNRQC in lane 2. DNA and WRNRQC concentration were 0.1 mM. Gel was visualized by UV shadowing.

Interaction between BLM proteins and DNA G-quadruplexes

Reaction mixture of BLM proteins with fluorescent DNA was prepared in the same procedure as described above, with the following modifications: 40-fold of BLM protein was titrated with 50 nM of DNA in pH 7.5 buffer, and allowed to incubate at 4 °C for 2 h. The samples were resolved on 10 % polyacrylamide gels containing 10 mM Kpi, pH 7.5 for 40 min under the constant application of 120 V, with 10 mM Kpi, pH 7.5 used as an electrophoresis buffer. Gel images were visualized and processed as described above. The same protocol was utilized for gel electrophoresis assay of WRNRQC. Gels visualized by UV shadowing contained DNA of $\geq 32.5 \mu\text{M}$ of DNA, and was titrated with either equimolar or 4-fold of BLM protein.

Curve fitting for protein/G-quadruplexes interaction

Interactions between the peptide/protein with G-quadruplexes were assumed to be under an equilibrium as represented in the following equation:

where $[DNA]$ and $[protein]$ represented the concentration of unbound DNA and free protein respectively, and $[DNA-protein]$ represented the concentration of DNA-protein complex formed.

To quantify the fraction of DNA remaining unbound after titration of protein, peak intensity of free DNA band from lane 2 – 5 was divided by the peak intensity of free DNA in lane 1. Thus the fraction of DNA bound (f_b) in the complex was calculated based on the following formula:

$$f_b = \frac{[DNA - protein]}{[DNA]_0} = 1 - \frac{[DNA]}{[DNA]_0} \quad 2.3$$

where $[DNA]_0$ represents the initial DNA concentration. With the inclusion of $[protein]$, initial protein concentration, $[protein]_0$ and dissociation constant, K_D within equation 2.3, equation 2.4 was formulated as:

$$f_b = \frac{(K_D + [DNA]_0 + [protein]_0) - \sqrt{(K_D + [DNA]_0 + [protein]_0)^2 - (4[protein] \cdot [DNA]_0)}}{2[DNA]_0} \quad 2.4$$

with K_D was obtained through non-linear square fitting of experimental data with equation 2.4.

2.6 Structure calculations

2.6.1 Structure computation for human telomeric G-quadruplex modified with a guanine-guanine-xanthine-8-oxoguanine tetrad

Structure calculations were performed with the programme XPLOR-NIH²²⁹ in two general steps: (i) distance geometry simulated annealing; (ii) restrained molecular dynamics refinement. Constraints imposed on the simulation include hydrogen-bond restraints, inter-proton distance restraints, dihedral restraints, planarity restraints and repulsive restraints. Inter-proton distances were deduced from NOESY experiments performed in H₂O (mixing time, 300 ms) and in D₂O (mixing time, 100 ms and 300 ms). The 10 lowest energy structures were selected and displayed with the programme PYMOL²³⁰.

Topology and Parameter files

For the structure calculation of *HT-17x9o*, the topology and parameter files for 8-oxoguanine were obtained from HIC-Up database; topology and parameter files for xanthine were derived from a structure geometry optimization with the Gaussian03 programme²³¹ using the 6-31G* basis set at the Hartree-Fock level of theory.

Non-exchangeable inter-proton distance restraints

Distances between non-exchangeable protons derived from NOESY spectra performed in D₂O at 100 ms and 300 ms mixing time were applied in structure calculations as inter-proton distance restraints. Based on peak intensities from NOESY (100 ms), very strong-to-medium peaks were manually assigned to be very strong (1.50 – 2.50 Å), strong (1.50 – 3.50 Å), medium strong (2.00 – 4.00 Å); medium-to-weak peaks from NOESY spectrum (300 ms) were assigned to be medium weak (2.50 – 4.60 Å), weak (4.00 – 6.50 Å) or very weak (5.75 – 7.75 Å). Boundaries for distance restraints of ambiguous proton assignment were increased by 0.5 Å.

Exchangeable inter-proton distance restraints

Exchangeable proton distance restraints were derived from cross-peaks that exhibit on NOESY spectra performed in H₂O (mixing time, 100 ms), and manually classified to be strong (2.10 – 3.90 Å), medium (2.80 – 5.20 Å) or weak (3.50 – 6.50 Å).

Repulsive restraints

Protons that do not exhibit NOE on NOESY spectra were assigned repulsive distance of 4.00 – 14.00 Å.

Dihedral restraints

All glycosidic χ angle of *anti* guanines are adjusted to $240^\circ \pm 40^\circ$, with experimentally determined *syn* guanines (G3, O_G9, G15, G16, G21) restrained at $60^\circ \pm 40^\circ$.

Hydrogen-bond restraints

All G-tetrads were restrained with O6-N1 (2.95 ± 0.10 Å), N7-N2 (2.90 ± 0.20 Å) distances, with T1•A20 Watson-Crick and T13•A24 reverse Watson-Crick pair constrained by N1 – H3 (2.00 ± 0.20 Å), N1-N3 (3.00 ± 0.20 Å) distances. The force constant for hydrogen bond restraints was maintained at $100 \text{ kcal.mol}^{-1}.\text{Å}^{-2}$ throughout the computation.

Planarity restraints

Planarity of G-tetrads was maintained through application of restraints on their base atoms.

Distance geometry simulated annealing

Initial calculation proceeded with the application of NOE, dihedral angle, hydrogen-bond, planarity, and repulsive restraints on an extended strand of d(TTGGG TTAO₆GG TTAGGX TTAGGG A), which undergoes distance geometry simulated annealing (DGSA) to generate 100 structures with a general fold for *HT-17x9o*.

Restrained molecular dynamics refinement

The 100 structures generated by DGSA were subjected to further refinement with the following procedure: the molecule was heated from 300 K to 1000 K in 5 ps, and allowed to equilibrate for 1 ps with force constants for the distance restraints maintained at 2 kcal.mol⁻¹.Å⁻². For the subsequent 26 ps, force constants for non-exchangeable proton distance restraints were raised to a final value of 50 kcal.mol⁻¹.Å⁻², with both exchangeable proton and repulsive distance restraints increased to 40 kcal.mol⁻¹.Å⁻². The molecule was then slowly cooled to 300 K in a 14 ps interval, and equilibrated for 10 ps. During the last 4.0 ps, coordinates for the molecule were stored every 0.5 ps and averaged. The averaged structure was subsequently subjected to minimization until the gradient of energy is ≤ 0.1 kcal.mol⁻¹.Å⁻². Dihedral (200 kcal.mol⁻¹.Å⁻²) and planarity restraints (G-tetrad: 1 kcal.mol⁻¹.Å⁻²; base pairs: 0.5 kcal.mol⁻¹.Å⁻²) were maintained throughout the refinement procedure. The 10 lowest-energy structures were selected at the end of this procedure.

Molecular Dynamics Refinement

Lowest energy structures from XPLOR calculations underwent further Molecular Dynamics (MD) refinement in explicit solvent using Amber 10 software²³². Systems were neutralized by the introduction of 23 K⁺ ions and solvated using TIP3P water molecule²³³ in a truncated octahedral box. Systems underwent an initial 500 steps of steepest decent minimization followed by 500 steps of conjugated gradient minimization with harmonic potential positions restraints of 25 kcal.mol⁻¹.Å⁻² on solute atoms. Systems were then heated from 100 K to 300 K under constant volume over 10ps while maintaining restraints. Subsequent minimization and equilibration cycles were undertaken in which position restrains on solute atoms were gradually reduced to 5, 4, 3, 2, 1 and 0.5 kcal.mol⁻¹.Å⁻². Final simulations were run in the absence of position restraints at 300K and 1atm for 1ns. Intra-proton distance restraints based upon NMR restraints were imposed throughout these simulations. A large subset of oligomer coordinates

were extracted from MD simulations and briefly minimized in vacuum. The 10 lowest-energy structures were chosen for final presentation. Partial atomic charges used in MD refinement and subsequent PBS calculations (see below) for non-native DNA residues 8-oxoguanine (O) and xanthine (X) were calculated using the RESP procedure at the HF/6-31G(d) level through the RED software²³⁴.

2.6.2 Structure computation for complex of Rhau peptide with parallel G-quadruplex

The structure of free Rhau₁₈ was calculated using the automated procedure from CYANA²³⁵ program. The structure of bound Rhau₁₈ was calculated using XPLOR-NIH²²⁹ program. The computed structures were displayed using PyMOL viewer program²³⁰.

NOE Distance Restraints

NOEs constraints for the free Rhau₁₈ were assigned and given distances using the automated procedure from CYANA. The consistency of NOEs was visually cross-checked on the NOESY spectra. NOEs constrained for *T95-2T* were taken from the previously published structure²³⁶, except for those involving the two first Thymine residues (T1 and T2). Intermolecular NOEs constraints for the bound Rhau₁₈ were manually assigned to be strong (1.50 – 4.00 Å), medium (1.50 – 5.00 Å) and weak (2.50 – 6.00 Å) from ¹³C- and ¹⁵N-edited NOESY spectra. Intermolecular NOEs were manually assigned to be (2.50 – 7.00 Å) using ¹³C-filtered NOESY (in which only the NOEs peaks from the labelled protein and unlabelled DNA were visible).

Dihedral constraints

ϕ/ψ angles constraints of free Rhau₁₈ were derived from by TALOS+²²⁷ based on ¹H_N, ¹H α , ¹³C α , ¹³C β , ¹³C' and ¹⁵N resonances.

An initial extended conformation of *T95-2T* and Rhau₁₈ was generated using the XPLOR program and were separated by 30 Å. The system was then subjected to distance geometry simulated annealing and restrained molecular dynamics refinement by incorporating the restraints with similar protocol as previously described.

2.6.3 Structure computation for the RQC domain of BLM protein

The two-step structure calculation for BLMRQC was performed on XPLOR-NIH with the following constraints imposed: inter-proton distance restraints, dihedral restraints and repulsive restraints. Inter-proton distances were derived from 3D ^{15}N -edited NOESY.

Dihedral constraints

ϕ/ψ angles of BLMRQC were predicted by TALOS+²²⁷ based on $^1\text{H}_\text{N}$, $^1\text{H}_\alpha$, $^{13}\text{C}_\alpha$, $^{13}\text{C}_\beta$, $^{13}\text{C}'$ and ^{15}N resonances.

2.7 Quantum chemical NMR Computations

Computed chemical shifts of G-tetrad systems were obtained using Gaussian 03 Software²³⁷. Model G•G•X•O and G•G•X(-)•O nucleobase tetrads containing two centrally located K^+ ions were minimized at the HF/6-31G(d) and MP2/6-31G(d) levels²³⁸. NMR chemical shifts were then computed for optimized tetrads using the GIAO methods²³⁹ at a HF/6-31G(d) level and calibrated using TMS as a reference. Change in chemical shifts of select protons were presented.

2.8 Poisson-Boltzmann Calculations.

Electrostatic surface potential of *HT-17x9o* and *HT* G-quadruplexes were each calculated using the software APBS²⁴⁰ on the programme PYMOL²³⁰, through the solution of linearized Poisson-Boltzmann equation at 0.15 M ionic strength at 37 °C, with dielectric constant of solvent set at 78.0. Derivation of partial charges used in PBS calculations were the same as those described in the “Molecular Dynamics Refinement” section in Chapter 2.6.1.

Chapter 3 Xanthine and 8-oxoguanine in G-quadruplexes: formation of a G•G•X•O tetrad

Within the human telomere, high susceptibility of guanine-rich DNA sequence to DNA damage¹⁰¹⁻¹⁰³ may result in guanine base lesions such as 8-oxoguanine (O) and xanthine (X) to be incorporated within G-quadruplex forming sequences^{104,105}. We demonstrated the first experimental data for site-directed mutagenesis of X on a G-quadruplex scaffold, which manifested as low thermostability and disruption to the parent G-quadruplex fold, similar to the effect of point mutation of O on G-quadruplexes. A G-tetrad double-modified with O and X (G•G•X•O tetrad) was engineered to retain the cyclic hydrogen bond pattern as observed in wild-type G-quadruplexes through complementary arrangement of the Hoogsteen face of O to the Watson-Crick face of X within the same G-tetrad. The rational incorporation of such G•G•X•O tetrads on a G-quadruplex of (3+1) folding topology demonstrated thermal properties and overall conformation highly similar to that of the wild-type G-quadruplex, while introducing a pH-dependent electronegative charge to the G-quadruplex groove containing the double mutation.

3.1 Introduction

Figure 3.1 A) Chemical structure of guanine analogues used in this study: xanthine (X) and 8-oxoguanine (O) are shaded in pink and blue respectively. The R group in the chemical drawings refer to 2-deoxyribose. B) Top: Sequence of *HT* and its base numbering shown above. Bottom: Schematic representation of human telomeric G-quadruplex (*HT*) utilized in the single-modification study. *Syn* and *anti* guanines are indicated in magenta and cyan respectively.

The effects of substitution of a guanine base by guanine analogs on G-quadruplexes have been the focus of numerous researches^{44,111,112,114,115,241-261}. This chapter deals with two major naturally occurring guanine base lesions: 8-oxoguanine and xanthine. 8-oxoguanine (abbreviated as O, Figure 3.1A) is formed in cells via oxidative damage of the guanine base^{103,262}. Single G-to-O substitutions have been reported to destabilize G-quadruplexes^{110-113,115}. In contrast, little is known about the effects of xanthine, another major guanine analog formed via nitrosative deamination of a guanine base (abbreviated as X, Figure 3.1A)^{116,117} when incorporated within G-quadruplexes.

In theory, a point mutation of either O or X in a within a G-tetrad would result in a loss of hydrogen bond and/or potential issues in steric hindrance (Figure 3.2A, left and middle). However, the disruptive effects of a single modification can be alleviated by concurrent X and O substitution in a proper orientation where the hydrogen donor/acceptor pattern at the Hoogsteen face of O is complementary with that of the Watson-Crick face of X (Figure 3.2A, right). The formation of a G-quadruplex containing an X•O•X•O tetrad was suggested by Hartig and co-workers^{114,263}. However, the formation of such a modified tetrad has not yet been confirmed by a high-resolution structural study. Furthermore, the thermal stability and circular dichroism data of these modified G-quadruplexes indicated the formation of a less stable G-quadruplex or multiple conformations, possibly linked to the fact that an X•O•X•O tetrad often resulted in O adopting an unfavourable *anti* orientation^{264,265}.

Figure 3.2 Characterization of *HT* and the constructs: *HT-9o*, *HT-17x*, *HT-17x9o*. A) Arrangement of the top G-tetrad with modification O at position 9 (left), X at position 17 (middle), and both O and X at position 9 and 17 respectively (right). Hydrogen bonds are represented by dotted lines. Hydrogen bonds connectivity affected by the modification(s) is shaded. B) Imino proton spectra. C) UV melting curves.

In this chapter, the effects of O and systematic X mutations to a modified human telomeric sequence, termed *HT* (Table 3.1) which adopts a well-defined (3+1) G-quadruplex⁴⁹ (Figure 3.1B) were investigated. Using NMR spectroscopy, the solution structure of a G-quadruplex containing a G•G•X•O tetrad was elucidated. The complementary arrangement between X and O was verified to re-establish the hydrogen-bond which would be disrupted by substitution by a single guanine analog. It was demonstrated that O can be rationally incorporated in a *syn* position to achieve high G-quadruplex thermal stability in the G•G•X•O tetrad. The NMR data revealed the X base in the X•O-modified G-quadruplex to exist in both protonated and deprotonated states at physiological pH. The consequences of a deprotonated X is a charge manifested within the G-quadruplex groove, which carries potential implications for drug binding and DNA-protein interactions.

3.2 Single substitution of guanine with 8-oxoguanine and xanthine are highly disruptive to G-quadruplexes

We performed single O substitutions at position G9, G16 and G17 of the (3+1) G-quadruplex formed by the human telomeric sequence *HT* (Figure 3.1 and Table 3.1). Imino proton NMR spectra of these sequences show more than 12 peaks (Figure 3.2B and Figure 3.3), indicating the co-existence of multiple G-quadruplex forms. In agreement with the NMR data, their Circular Dichroism (CD) spectra show notable deviation from that of the original *HT* sequence (Figure 3.3C). The melting temperature (T_m) of these O-substituted sequences is greatly reduced (18–27°C)

Figure 3.3 Characterization of *HT* and the constructs *HT-9o*, *HT-16o*, and *HT-17o*. A) Arrangement of G-tetrad with single substitution of 8-oxoguanine (O). Hydrogen bond connectivity affected by the modification is shaded. B) Imino proton spectra. C) CD spectra.

Figure 3.4 Characterization of *HT* and the constructs with point mutation of xanthine. A) Arrangement of G-tetrad with single substitution of xanthine (X). Hydrogen bond connectivity affected by the modification is shaded. B) Imino proton spectra. C) CD spectra.

Figure 3.5 Characterization of *HT* and the constructs with point mutation of inosine. A) Arrangement of G-tetrad with single substitution of inosine (I). Hydrogen bond connectivity affected by the modification is shaded. B) Imino proton spectra. C) CD spectra.

Table 3.1 Mono-substituted G-quadruplex sequences and their thermal stability

Name	Sequence (5' – 3')				T_m (°C)	ΔT_m (°C)
	<i>S A A</i>	<i>S A A</i>	<i>S S A</i>	<i>S A A</i>		
<i>HT</i>	TT GGG TTA GGG	TTA GGG	TTA GGG	A	58.5 ± 0.4	-
<i>HT-3x</i>	TT X GG TTA GGG	TTA GGG	TTA GGG	A	46.9 ± 0.1	-11.6
<i>HT-4x</i>	TT GX G TTA GGG	TTA GGG	TTA GGG	A	31.5 ± 1.0	-27.0
<i>HT-5x</i>	TT GG X TTA GGG	TTA GGG	TTA GGG	A	40.4 ± 0.3	-18.1
<i>HT-9x</i>	TT GGG TTA X GG	TTA GGG	TTA GGG	A	38.4 ± 0.1	-20.1
<i>HT-10x</i>	TT GGG TTA GX G	TTA GGG	TTA GGG	A	37.6 ± 0.9	-20.9
<i>HT-11x</i>	TT GGG TTA GG X	TTA GGG	TTA GGG	A	41.3 ± 0.2	-17.2
<i>HT-15x</i>	TT GGG TTA GGG	TTA X GG	TTA GGG	A	40.2 ± 0.8	-18.3
<i>HT-16x</i>	TT GGG TTA GGG	TTA GX G	TTA GGG	A	35.2 ± 1.7	-23.3
<i>HT-17x</i>	TT GGG TTA GGG	TTA GG X	TTA GGG	A	46.0 ± 0.5	-12.5
<i>HT-21x</i>	TT GGG TTA GGG	TTA GGG	TTA X GG	A	40.4 ± 0.1	-18.1
<i>HT-22x</i>	TT GGG TTA GGG	TTA GGG	TTA GX G	A	< 25.0*	-33.5
<i>HT-23x</i>	TT GGG TTA GGG	TTA GGG	TTA GG X	A	38.5 ± 0.4	-20.0
<i>HT-3i</i>	TT I GG TTA GGG	TTA GGG	TTA GGG	A	48.8 ± 0.8	-9.7
<i>HT-4i</i>	TT G IG TTA GGG	TTA GGG	TTA GGG	A	45.5 ± 0.7	-13.0
<i>HT-5i</i>	TT GG I TTA GGG	TTA GGG	TTA GGG	A	52.5 ± 0.6	-6.0
<i>HT-9i</i>	TT GGG TTA I GG	TTA GGG	TTA GGG	A	47.8 ± 0.3	-10.7
<i>HT-10i</i>	TT GGG TTA G IG	TTA GGG	TTA GGG	A	45.8 ± 0.5	-12.7
<i>HT-11i</i>	TT GGG TTA GG I	TTA GGG	TTA GGG	A	53.3 ± 0.5	-5.2
<i>HT-15i</i>	TT GGG TTA GGG	TTA I GG	TTA GGG	A	47.0 ± 0.5	-11.5
<i>HT-16i</i>	TT GGG TTA GGG	TTA G IG	TTA GGG	A	50.0 ± 1.5	-8.5
<i>HT-17i</i>	TT GGG TTA GGG	TTA GG I	TTA GGG	A	51.5 ± 0.7	-7.0
<i>HT-21i</i>	TT GGG TTA GGG	TTA GGG	TTA I GG	A	46.0 ± 0.5	-12.5
<i>HT-22i</i>	TT GGG TTA GGG	TTA GGG	TTA G IG	A	48.1 ± 0.4	-10.4
<i>HT-23i</i>	TT GGG TTA GGG	TTA GGG	TTA GG I	A	35.4 ± 0.1	-23.1
<i>HT-9o</i>	TT GGG TTA O GG	TTA GGG	TTA GGG	A	38.5 ± 0.4	-20.0
<i>HT-16o</i>	TT GGG TTA GGG	TTA G OG	TTA GGG	A	< 25.0*	-33.5
<i>HT-17o</i>	TT GGG TTA GGG	TTA GG O	TTA GGG	A	36.5 ± 0.5	-22.0

T_m values of each sample are determined using buffer containing 10 mM potassium phosphate, 10 mM potassium chloride, pH 7.0. The modified residues used are: **X** = xanthine, **O** = 8-oxoguanine, **I** = inosine.

The glycosidic conformation of each (modified) guanine base is positioned in *syn* (*S*) or *anti* (*A*) orientation.

* T_m of sample falls below the detection range of the UV spectrophotometer used in the UV melting experiment.

as compared to that of *HT* (Figure 3.2C and Table 3.1), supporting the previous finding that O substitutions highly destabilize G-quadruplex structures. This destabilization effect may come from a steric clash between the proton H7 of 8-oxoguanine and the amino group of the neighbouring guanine (Figure 3.2A).

Although previous computational studies suggested that tetrads formed entirely by xanthine (**X**) are possible^{128,130}, no experimental studies on the effects of single X substitutions in G-quadruplexes have been reported. Here we systematically performed single X substitutions at all twelve G positions of the *HT* sequence (Table 3.1). NMR and CD spectra of X-substituted

sequences (Figure 3.2B and Figure 3.4B, C) indicate that a G-to-X modification induces the formation of multiple G-quadruplexes in most cases (except for substitutions at G10 and G16). Melting experiments show that single X substitutions also highly destabilize G-quadruplex structures, exhibiting a T_m decrease of over 10 °C as compared to that of *HT* (Figure 3.2C and Table 3.1). This destabilizing effect is most likely due to hydrogen-bond loss and unfavourable electrostatic interaction occurring in the context of a single X substitution (Figure 3.2A, 3.4A). As a comparison, a milder destabilizing effect was observed for inosine (I) substitutions, which appeared to maintain the same G-quadruplex fold in most cases and caused smaller decreases in the melting temperature (Table 3.1 and Figure 3.5).

3.3 Complementary X•O dual substitutions are compatible with G-quadruplexes

The imino proton NMR spectrum of *HT-17x9o* shows a well-resolve spectrum with 10 sharp peak (and a broad peak) indicating the formation of a single G-quadruplex conformation (see further analysis below). The chemical shift pattern (Figure 3.2B) and CD spectrum (Figure 3.6) of *HT-17x9o* are highly similar to that of *HT*, consistent with the conservation of the (3+1) G-quadruplex folding topology.

While individual single substitutions of O at position G9 and X at position G17 (*HT-9o* and *HT-17x*, respectively) were highly disruptive to the *HT* G-quadruplex, simultaneous substitutions of O and X at these positions (*HT-17x9o*) (Table 3.2), allow for the formation of a G•G•X•O tetrad, which restores the hydrogen bonding integrity of the G-tetrad via complementary pairing between X•O. In this tetrad, the X•O pair is linked by two hydrogen bonds: one formed between

Table 3.2 Dual-substituted G-quadruplex sequences and their thermal stability

Name	Sequence (5' – 3')				T_m (°C)	ΔT_m (°C)
	S A A	S A A	S S A	S A A		
<i>HT-3x21o</i>	TT X GG	TTA GGG	TTA GGG	TTA O GG A	53.8 ± 0.8	-4.8
<i>HT-9x3o</i>	TT O GG	TTA X GG	TTA GGG	TTA GGG A	55.4 ± 0.4	-3.1
<i>HT-17x9o</i>	TT GGG	TTA O GG	TTA GG X	TTA GGG A	61.2 ± 1.8	+2.7
<i>HT-10x16o</i>	TT GGG	TTA G X G	TTA G OG	TTA GGG A	59.3 ± 1.0	+0.8
<i>HT-11x15o</i>	TT GGG	TTA GG X	TTA O GG	TTA GGG A	58.6 ± 1.6	+0.1
<i>HT-21x17o*</i>	TT GGG	TTA GGG	TTA GGO	TTA X GG A	45.9 ± 1.4	-12.6

T_m values of each sample are determined using buffer containing 10 mM potassium phosphate, 10 mM potassium chloride, pH 7.0. The modified residues used are: **X** = xanthine; **O** = 8-oxoguanine.

The glycosidic conformation of each (modified) guanine base is positioned in *syn* (S) or *anti* (A) orientation.

* **O** was substituted at the disfavoured *anti* glycosidic position in this construct.

the N1-H1 of xanthine and the C=O group of 8-oxoguanine; the other formed between the N7-H7 of 8-oxoguanine and the C=O group of xanthine (Figure 3.2A).

The complementary X•O dual substitutions in *HT-17x9o* led to recovery of the thermal stability of the G-quadruplex structure (Figure 3.2C), displaying a melting temperature 3 °C higher than that of the unmodified *HT* sequence (Table 3.2), in sharp contrast to the strong destabilization observed for single-modified sequences *HT-9o* and *HT-17x*.

Figure 3.6 Characterization of *HT* double-modified with X•O substitution A) imino proton spectra. B) UV melting curves. C) CD spectra.

Figure 3.7 Characterization of *HT-21x17o* and the constructs: *HT-17o*, *HT-21x* and *HT-21x17o*. A) Imino proton spectra. Imino peaks from the major form indicated by red asterisks (*). B) UV melting curves. C) CD spectra.

As 8-oxoguanine was previously shown to favour a *syn* conformation²⁶⁶, four other X•O-modified sequences were generated with the aim of incorporating a G•G•X•O tetrad in which O were also at the position to adopt a *syn* conformation in the original structure (G3, G15, G16, and G21) (Table 3.2). All these X•O-dual substituted sequences (namely *HT-9x3o*, *HT-11x15o*, *HT-10x16o*, and *HT-3x21o*) were found to exhibit similar characteristics as those of *HT-17x9o* including NMR spectra, CD spectra, and melting temperature, indicating that the same G-quadruplex form had been adopted (Figure 3.6 and Table 3.2)

The sequence *HT-21x17o* was synthesized to probe the effect of substituting O into an *anti* position G17 (Table 3.2 and Figure 3.7). The NMR and CD spectra of *HT-21x17o* were similar to those of *HT*, indicating that this sequence adopted predominantly the same (3+1) G-quadruplex fold despite a minor population being evident in NMR imino proton spectra (Figure 3.7B). However, the melting temperature of *HT-21x17o* with O in an *anti* conformation at position 17 is significantly reduced compared to that of *HT-17x9o* with O substituted into *syn* position 9 (Table 3.2), suggesting that O substitution within a G•G•X•O tetrad exhibits a preference for the *syn* conformation.

3.4 Protonation states of a G•G•X•O tetrad

In the course of this study, an interesting behaviour was observed in the imino proton NMR spectra as a function of pH. This behaviour facilitated the investigation of the protonation state of the G•G•X•O tetrad. At pH 7.0, imino proton spectra of X•O dual substituted G-quadruplexes (Table 3.2) consistently exhibited only 10 major sharp peaks (Figures 3.6, 3.7) out of 13 expected peaks in this region (12 H1 protons of G/X/O and the H7 proton of O involved in base pairing hydrogen bonds, Figure 3.2A). At pH 5.0 and pH 9.0, all 13 sharp peaks could be observed in this region (Figure 3.8). As a comparison, the imino spectrum of the original sequence *HT* shows 12 peaks for the whole range of pH from 5.0 to 9.0 (Figure 3.8).

Figure 3.8 Imino spectra of *HT* (left), *HT-17x9o* (middle) and *HT-9x3o* (right) recorded at pH 5.0, 7.0 and 9.0. Each imino peaks are indicated by a red dot above the spectrum.

We focused our study on the *HT-17x9o* sequence at different pH. NMR imino proton spectra recorded at pH 5.0 – 9.0 showed large chemical shift perturbations which were not observed in the *HT* G-quadruplex (Figure 3.8, 3.9A). Spectral assignment and subsequent analysis

of *HT-17x9o* have revealed that three peaks (O9H7, X17H1 and G21H1) had experienced large chemical shift changes during pH titration. $\{^1\text{H}-^{15}\text{N}\}$ HSQC spectrum of *HT-17x9o* was used to verify the identity of O9H7 and X17H1 (Figure 3.9B) (Further discussions on the identification of these peaks are presented below). A plot of chemical shifts of these protons as a function of pH gave a similar transition point of pH ~ 6.7 for each proton (Figure 3.9C), with the protons O9H7 and G21H1 shifted downfield and 17H1 shifted up field. These observations are consistent with the protonation/deprotonation at position N3 of X17 (Figure 3.9D) and are supported by quantum chemical NMR chemical shift calculations for idealized geometries of a G•G•X•O tetrad in protonated/deprotonated form (Figure 3.10).

Figure 3.9 NMR titration of *HT-17x9o*. A) Imino proton spectra of *HT-17x9o* at pH 5.5, 7.0 and 8.0. Identities of imino protons are indicated above each spectrum. Peak of O9H7 at pH 7.0 was broadened and not seen on spectrum. B) $\{^1\text{H}-^{15}\text{N}\}$ -HSQC spectrum showing N1-H1 cross-peaks, which are labelled with their corresponding nucleotide number for *HT-17x9o* at pH 8.0. C) Plot of NMR chemical shifts observed for O9H7, X17H1 and G21H1 protons of *HT-17x9o* as a function of pH for the range of 5.0 – 9.0, with the mid-transition point represented by a dashed line. Peak shifts of O9H7, X17H1 and G21H1 are coloured in red, blue and grey respectively. D) Ionization of xanthine (X17) on *HT-17x9o*, with the neutral 2,6-diketo form (left) in equilibrium with the anionic form (right) at pH 6.7, corresponding to its acid dissociation constant (pK_a).

Figure 3.10 Comparison of changes in experimental and computed NMR chemical shifts of *HT-17x9o*. Experimental data are NMR chemical shifts observed for O9H7, X17H1 and G21H1 with changes from the protonated (pH 5.0) to the deprotonated (pH 9.0) states (Y-axis). Computational data are computed changes in NMR chemical shifts (X-axis) of these protons in model G•G•X•O tetrads in the protonated and deprotonated state. Model tetrads were optimized at the HF/6-31G(d) level with NMR chemical shifts computed at the HF/6-31G(d) (▲) and MP2/6-31G(d) (●) level.

Figure 3.11 Acid dissociation constant (pK_a) values of xanthine (X) which undergoes deprotonation at N3^{267,268}.

In literature, the protonation states of X in the context of G-quadruplex structure have remained relatively unexplored, although the protonation behaviour of isolated X nucleobases and nucleosides has been extensively studied (Figure 3.11), with a known pK_a of N3 being ~ 7.7 and 5.7 respectively^{269,270}. Through our data, the pK_a of X17N3 in the *HT-17x9o* G-quadruplex was determined via the midpoint transition of NMR chemical shifts (Figure 3.9C), to be $\text{pH} \sim 6.7$. The increase in pK_a of X in the G-quadruplex scaffold from pK_a . This value is the average of pK_a values of the xanthine nucleoside and nucleobase, suggesting the oligonucleotide attached at N9 to confer properties in between that of the nucleoside and nucleobase. However more studies is required to identify the factors influencing this behaviour, which is beyond the scope of this thesis.

3.5 *HT-17x9o* containing a G•G•X•O tetrad maintains the same (3+1) G-quadruplex fold at pH 5.5-8.0

The structural effects of incorporating a G•G•X•O tetrad into a G-quadruplex was investigated by determining the NMR solution structure of the *HT-17x9o* sequence. Guanines were 4% ^{15}N -labelled (Table 3.3) using a low-enrichment site-specific labelling approach²²². Guanine imino (H1) protons of *HT-17x9o* were unambiguously assigned at pH 5.5 (Figure 3.12A), pH 7.0 (Figure 3.13A) and pH 8.0 (Figure 3.14A). Assignment of the H8 protons were carried out at pH 7.0 (Figure 3.13A) and extrapolate for pH 5.5 and 8.0, as no significant chemical shift changes for these protons were observed as a function of pH. Resonance assignments of O9 and X17 were assisted by NOESY (Figure 3.15) and $\{^1\text{H}-^{15}\text{N}\}$ -HSQC experiment at pH 8.0, where chemical shifts for O9H7 and X17H1 are distinct from those of guanines (Figure 3.9B). Aromatic

Figure 3.12 NMR spectral assignment of *HT-17x9o* at pH 5.5. A) Imino region of ^{15}N -filtered spectra recorded on samples with 4%- ^{15}N -enrichment at indicated positions. Assignment of O9H1, O9H7 and X17H1 were based on $\{^1\text{H}-^{15}\text{N}\}$ -HSQC and through-space NOE correlation from NOESY spectrum. B) NOESY spectrum of *HT-17x9o* (mixing time, 300 ms) used for elucidation of its folding topology. H1-H8 cross-peaks are framed and labelled with the number of imino protons followed by that of H8. C) Schematic representation of X•O-modified human telomeric G-quadruplex (*HT-17x9o*). *Syn* and *anti* guanines are indicated in magenta and cyan respectively. Characteristic NOE connectivities for each G-tetrad layer of *HT-17x9o* are shown below, with top, middle and bottom G-tetrad coloured as orange, green and purple respectively.

and sugar protons were assigned in H8/H6-H1' NOE sequential connectivity of *HT-17x9o*, assisted by using through-space experiment (NOESY; Figure 3.15) and through-bond experiments (TOCSY, $\{^1\text{H}, ^{13}\text{C}\}$ -HSQC; Figure 3.16C). The adoption of *syn* glycosidic orientation by G3, G15, G16 and G21 were verified based on the high intensity observed for intra-residue H8-H1' NOE cross-peaks (Figure 3.14C). The remaining guanines and xanthine X17 adopt the *anti* glycosidic conformation.

Figure 3.13 NMR spectra and assignments of *HT-17x9o* at pH 7.0. A) ^{15}N -filtered spectra of the imino region (left) and aromatic region (right) recorded on samples with 4% ^{15}N -enrichment at indicated positions. B) NOESY spectrum of *HT-17x9o* (mixing time, 300 ms) used for elucidation of its folding topology. H1-H8 cross-peaks are framed and labelled with the number of imino protons followed by that of H8. C) Schematic representation of X•O-modified human telomeric G-quadruplex (*HT-17x9o*). *Syn* and *anti* guanines are indicated in magenta and cyan respectively. Characteristic NOE connectivities for each G-tetrad layer of *HT-17x9o* are shown below, with top, middle and bottom G-tetrad coloured as orange, green and purple respectively.

The H1 and H8 assignments of guanine, xanthine, and 8-oxoguanine bases in *HT-17x9o* were applied to determine the alignment of three G-tetrads. At pH 8.0, the characteristic H1-H8 cyclic NOE connectivity pattern recorded for *HT-17x9o* (Figure 3.14B) could be identified for each G-tetrad: top – G3•G21•X17•O9; middle – G4•G10•G16•G22; bottom – G5•G11•G15•G23. Imino protons from the G4•G10•G16•G22 tetrad are most protected against solvent exchange with D₂O (Figure 3.17), thus verifying their positioning as the middle G-tetrad. The hydrogen-bond directionality of the top G-tetrad (except for the hydrogen bond formed between O9H7 and X17O2) is opposite that of the other two G-tetrads, with the glycosidic conformations of residues from the top G-tetrad being *syn•syn•anti•syn*, and those of the middle and bottom G-tetrad being *anti•anti•syn•anti*. The G3-G5, O9-G11, and G21-G23 tracts are oriented in the same direction, while the G15-X17 tract is oriented in the opposite direction. The corresponding three loops are propeller, edgewise, and edgewise respectively.

Figure 3.14 NMR spectra and assignments of *HT-17x9o* at pH 8.0. A) ¹⁵N-filtered spectra recorded on samples with 4% ¹⁵N-enrichment at indicated positions. B) NOESY spectrum of *HT-17x9o* (mixing time, 300 ms) used for elucidation of its folding topology. H1-H8 cross-peaks are framed and labelled with the number of imino protons followed by that of H8. C) Sequential 5'–3' walk o NOESY spectrum of *HT-17x9o* based on H8-H1' NOE as traced by black lines. H8 of O9 is absent and hence the connectivity of G8-O9 and O9-G10 are not shown. Intra-residue NOE correlations are labelled according to their respective nucleotide number. Cross-peaks visible at lower spectral threshold are framed.

Table 3.3 Sequence of ^{15}N -labelled *HT-17x9o* used in this study

Entry	Sequence (5' – 3')					
<i>HT-17x9o</i> [3]	TT	[*G]GG	TTA OGG	TTA GGX	TTA GGG	A
<i>HT-17x9o</i> [4]	TT	G[*G]G	TTA OGG	TTA GGX	TTA GGG	A
<i>HT-17x9o</i> [5]	TT	GG[*G]	TTA OGG	TTA GGX	TTA GGG	A
<i>HT-17x9o</i> [10]	TT	GGG	TTA O[*G]G	TTA GGX	TTA GGG	A
<i>HT-17x9o</i> [11]	TT	GGG	TTA OG[*G]	TTA GGX	TTA GGG	A
<i>HT-17x9o</i> [15]	TT	GGG	TTA OGG	TTA [*G]GX	TTA GGG	A
<i>HT-17x9o</i> [16]	TT	GGG	TTA OGG	TTA G[*G]X	TTA GGG	A
<i>HT-17x9o</i> [21]	TT	GGG	TTA OGG	TTA GGX	TTA [*G]GG	A
<i>HT-17x9o</i> [22]	TT	GGG	TTA OGG	TTA GGX	TTA G[*G]G	A
<i>HT-17x9o</i> [23]	TT	GGG	TTA OGG	TTA GGX	TTA GG[*G]	A

(X = xanthine, O = 8-oxoguanine, [*G] = ^{15}N -labelled guanine)

Figure 3.15 NOESY spectrum of *HT-17x9o* recorded at A) pH 5.5, B) pH 7.0 and C) pH 8.0, with H1-H1/H7-H1/H1-H7 through space correlations labelled as indicated on the spectrum. NOE correlation from minor form are marked by asterisks (*).

Figure 3.16 Spectra of through-bond 2D NMR experiments used for verification of NOE assignment of *HT-17x9o*: A) TOCSY; B) $\{^1\text{H}, ^{13}\text{C}\}$ -HSQC. Through-bond correlation between H1' and C2' are indicated with green, blue, and red lines for T6, T12 and T18 respectively.

Figure 3.17 Imino proton spectrum of *HT-17x9o* at pH 7.0 recorded after 15 min exposure in D_2O .

Overall, the G-quadruplex folding topology of *HT-17x9o* is identical to that of *HT*⁴⁹, indicating that the replacement of two guanines with an X•O pair did not disturb the original (3+1) G-quadruplex scaffold. Furthermore, the H1-H8 NOE patterns recorded at pH 5.5 (Figure 3.12B) and pH 7.0 (Figure 3.13B) are similar to those observed at pH 8.0 (Figure 3.14B), indicating that both the protonated and deprotonated forms of *HT-17x9o* conserve the (3+1) G-quadruplex folding topology of the unmodified sequence (Figures 3.12C and 3.13C).

3.6 Solution structure of *HT-17x9o*: maintaining the shape and varying the groove charge

Structure determination of *HT-17x9o* was performed on the deprotonated form, which was predominantly formed at pH 8.0. The solution structure of *HT-17x9o* (Figure 3.18) was computed on the basis of NMR constraints (Appendix F and Figure 3.14). The G-tetrad core consists of one narrow, one wide, and two medium grooves, with the first TTA linker (propeller loop) traversing across a medium groove, while the T12-A14 and T18-A20 edgewise loops cross a wide groove at the bottom and a narrow groove at the top respectively. Hydrogen bond interaction between 5' and

3' flanking nucleotides with the two edgewise loop adopt a Watson-Crick T1•A20 and a reverse Watson-Crick T13•A24 base pair, with the former and the latter capping the top and the bottom of the G-tetrad core respectively.

Figure 3.18 A) Side view and B) top view of 10 superimposed lowest energy conformations of *HT-17x9o*. Adenines are coloured in green; thymine, orange; backbone and sugars, grey; O4' atoms, yellow and phosphorus atoms, dark red. O9 and X17 are coloured in red and blue respectively. *Anti* and *syn* guanines are coloured in cyan and magenta. C) Side view and D) top view of a representative structure of *HT-17x9o*. Residue 1, 2, 18 and 19 are removed in the top view for clarity.

Figure 3.19 Comparison between double-modified *HT-17x9o* (coloured in blue) and unmodified *HT* G-quadruplex (coloured in red). Structures were aligned based on their G-tetrad core with RMSD 0.649 Å. A) Side view and B) top view of the superimposed structures, with sugar phosphate backbone and bases of G-tetrad core presented. Residues 1, 2, 18 and 19 are removed in the top view for clarity.

To examine the structural differences between the dual substituted and original G-quadruplex, *HT-17x9o* was aligned with *HT* (Figure 3.19). Superimposition of both structures showed a high degree of overlap. Thus, aside from the changes in the chemical structure of two bases (X and O) made to the top G-tetrad, structural motifs of *HT* are highly conserved in *HT-17x9o*.

However, a particular feature of introducing an X•O pair into a G-quadruplex at pH 8.0 is the deprotonated and negatively charged xanthine base. We computed the partial charge of X and O nucleotides and calculated the electrostatic potential at the surface of the molecules (see Chapter 2.8). Comparative electrostatic potential surface representations between the *HT-17x9o* and the *HT* G-quadruplexes show a notable decrease in electrostatic potential on the wide groove of the G-quadruplex (Figure 3.20).

Figure 3.20 Comparative electrostatic potential surface representation of A) *HT-17x9o* and its wild-type B) *HT* G-quadruplex, depicting charge differences at the wide groove. Surface representation of both G-quadruplexes was presented at half transparency, with electrostatic potentials at ± 10 kT/e, whereby white or red colour indicated for neutral or negative charge potential respectively.

3.7 Position-specific X•O pair substitution as a versatile tool in engineering the G-quadruplex form

With the introduction of G•G•X•O tetrads into a human telomeric G-quadruplex forming sequences, the dual X•O-modification has been successfully integrated at five different positions on the *HT* G-quadruplex scaffold (based upon the *syn* position $G \rightarrow O$ conditional) without significantly affecting the fold and melting temperature. Consequently, these two modifications may be potentially applied as structural engineering tool to enforce the assembly of a desired G-quadruplex form within a G-quadruplex sequence that adopts multiple forms, which was

evaluated in two other human telomeric G-quadruplex sequences *HT1* and *HT2* (Table 3.4). Each of this G-quadruplex forming sequence adopts a different major G-quadruplex conformations despite having nearly identical DNA sequences (Figure 3.21A, B)^{48,49}. However, minor G-quadruplex forms could be observed in *HT1* and *HT2* (Figure 3.21C). Using the predetermined structure of *HT1* and *HT2*, X and O could be rationally incorporated within the same G-tetrad, such that the favourable thermodynamics displayed only for X•O pair could be exploited to strongly favour the assembly of the major G-quadruplex form in both *HT1* and *HT2*, with *HT1*-10x16o and *HT2*-4x10o constructed respectively (Table 3.4).

Figure 3.21 Comparison of different G-quadruplexes A) *HT1* and B) *HT2* applied in structure engineering experiment: (Top) DNA sequence of each G-quadruplex and their respective nucleotide numbering is shown directly above.; (Middle) schematic representations of different types of (3+1) hybrid G-quadruplexes used in this study; (Bottom) Comparative NMR imino proton spectra between the unmodified and double-modified constructs of *HT1* and *HT2*, minor forms significantly reduced after incorporation of X or O are indicated by red arrows. Nucleotides modified with X and O are shaded in red and blue respectively

Both of the double-modified constructs were characterized by NMR spectroscopy, and their imino proton spectra displayed 10 sharp peaks corresponding to a single conformation for both *HT1* and *HT2* (Figure 3.21C). CD spectrum (Figure 3.22) and thermal stability (Table 3.4) of both modified constructs were determined to be similar to that of its reference construct. The significant reduction in the minor forms for both engineered constructs (as indicated by the red arrows) on Figure 3.21C, together with the likeliness of the imino proton pattern of the modified

construct as the wild-type G-quadruplex points towards the shift in equilibrium towards the major form of *HT1* and *HT2* for *HT1-10x16o* and *HT2-4x10o* respectively, which is in agreement with both CD (Figure 3.22) and thermostability data (Table 3.4).

The current data suggest X•O pair to be an excellent structure engineering tool for enforcing the formation of a desired G-quadruplex fold, which have potential applications both in the fundamental studies of the G-quadruplex structure, and also for the engineering of G-quadruplex nano-structures. Moreover, the robustness of using the G•G•X•O tetrad as an engineering tool is highlighted in *HT-21x17o*, where a G-quadruplex which maintained the proper folding topology is formed even when O is substituted into an *anti* position guanine but at a cost of $\sim 11^\circ\text{C}$ in thermal stability.

Figure 3.22 CD spectra of *HT1* and *HT1-10x16o* (left); *HT2* and *HT2-4x10o* (right).

Table 3.4 DNA sequence of X•O modified G-quadruplexes used for structure engineering study and their thermal stability

Name	Sequence (5' – 3')				T_m ($^\circ\text{C}$)	ΔT_m ($^\circ\text{C}$)
	S A A	S A A	S S A	S A A		
<i>HT1</i>	TA GGG	TTA GGG	TTA GGG	TTA GGG	56.3 ± 0.2	-
<i>HT1-10x16o</i>	TA GGG	TTA G X G	TTA G O G	TTA GGG	58.0 ± 0.8	+1.6
<i>HT2</i>	TA GGG	TTA GGG	TTA GGG	TTA GGG TT	51.0 ± 0.8	-
<i>HT2-4x10o</i>	TA G X G	TTA G O G	TTA GGG	TTA GGG TT	53.2 ± 0.4	+2.2

T_m values of each sample are determined using buffer containing 10 mM potassium phosphate, 10 mM potassium chloride, pH 7.0. The modified residues used are: **X** = xanthine; **O** = 8-oxoguanine.

3.8 Potential applications of a G-quadruplex containing the G•G•X•O tetrad

Unlike previously studied X•O•X•O tetrads^{114,263}, a G•G•X•O tetrad requires only two residues within the same G-tetrad to be modified, allowing flexibility in the placement of O so as to avoid substitution into an *anti* position guanine. Considering this prerequisite, the X•O pair has been successfully integrated at five positions on the *HT* G-quadruplex scaffold (based on the *syn* position G → O conditional) without significantly affecting the fold and melting temperature. It was demonstrated that the G•G•X•O tetrad can be formed when O is substituted into an *anti* position guanine (*HT-21x17o*) while maintaining proper folding topology but at a cost of 13 °C in thermal stability.

The complementary pairing of X and O in the modified G•G•X•O tetrad is demonstrated to be a versatile design for countering the destabilizing effects of X and O modifications introduced on G-quadruplexes via DNA damage. Although the likelihood of nearby co-occurrence of X and O mutations on a G-rich sequence of the genome is assumed to be small *in vivo*, such nearby mutations could favour one G-quadruplex form over other alternatives.

The X•O modification introduces a negatively charged groove on the molecule, which may have important implications, particularly in DNA nanotechnology. Generation of a charged groove may affect G-quadruplex-protein interactions. From an engineering perspective, G-quadruplexes in therapeutic and sensing applications may have their binding properties enhanced by such modification. Additionally, nano-scale wires built upon G-quadruplex architecture may benefit from the introduction of negatively charged bases in a controlled manner.

3.9 Conclusion

The human telomeric (3+1) G-quadruplex (*HT*) was found to be highly sensitive to single xanthine (X) or 8-oxoguanine (O) substitutions, which induced disruptive effects on the structure and thermal stability of modified sequences. In contrast, complementary pairing of X with O within a single G•G•X•O tetrad facilitates the formation of stable G-quadruplexes. The solution structure of a G•G•X•O-containing G-quadruplex formed by a double mutated sequence *HT-17x9o* was solved, revealing the same folding topology in both protonated and deprotonated form. The X•O pair in the deprotonated state showed an enhanced electronegativity in the modified

groove, with potential application in charge-dependent recognition by other molecules.

Chapter 4 Inverting the G-tetrad polarity of a G-quadruplex using xanthine and 8-oxoguanine

4.1 Introduction

As G-quadruplexes exhibit high thermostability under physiological conditions^{1,68,271,272}, their occurrence in cells may serve as a physical blockage which halts enzyme progression in key biological processes²⁷³. This property of G-quadruplexes can be potentially exploited in therapeutics, with G-quadruplexes being anticancer targets^{274,275}, and applied as antiviral/anticancer agents²⁷⁶⁻²⁷⁸. In DNA nanotechnology, the ability of G-quadruplexes to transport charges between consecutively stacked G-tetrads was exploited to construct conductive nanowires^{279,280} and electronic switches²⁸¹. Such variability leads to a large pool of available G-quadruplex structures for applications in therapeutics²⁷⁶⁻²⁷⁸, and nanotechnology²⁷⁹. For example,

computational and experimental data have shown that different stacking geometries in G-quadruplexes exhibit differences in fluorescence and charge-transfer properties²⁸²⁻²⁸⁴.

Efforts were made to control the G-quadruplex scaffold by adjusting the environment: the type and concentration of cation^{236,285}, or the presence of co-solute such as polyethylene glycol²⁸⁶⁻²⁹¹. Another approach for altering the G-quadruplex scaffold involves the rational incorporation of guanine analogues within the G-quadruplex scaffold, such as the sugar-modified nucleotides^{246,250,251,255,292} (e.g. 2'-deoxy-2'-fluoro-guanosine (^FG), 2'-deoxy-2'-fluoro-arabinoguanosine (^{FANA}G), locked nucleic acid (LNA), peptide nucleic acid (PNA), unlocked nucleic acid (UNA)) and the base-modified guanine analogues^{47,112,115,243} (e.g. inosine, 6-thioguanine, 8-aminoguanine, 8-bromoguanine, 8-methylguanine, 8-oxoguanine).

Using 5'-5' inversion-of-polarity backbone modification, Martino *et. al.*²⁹³ showed that relative strand orientations in the G-tetrad core could be manipulated without altering the loop arrangements. In a new approach for manipulating the G-quadruplex scaffold, Dickerhoff and Weisz²⁹⁴ successfully reversed the hydrogen-bond directionality of a G-tetrad without altering the global fold. Through the rational application of ^FG in syn positions within the modified tetrad, the preference for anti glycosidic orientations drove the inversion of G-tetrad H-bond polarity, while retaining the original folding topology of the G-quadruplex²⁹⁴.

With the ^FG being a sugar-modified guanine analog, we seek to explore the possibility of applying base-modified guanine analogues in G-tetrads to engineer the directionality of hydrogen bond within a G-quadruplex. Two naturally occurring DNA base lesions: 8-oxoguanine (O) and xanthine (X) (Figure 4.1) formed via oxidation¹⁰³ and nitrosative deamination¹¹⁶ of the guanine base respectively, were identified as potential tools for this purpose. The application of these two guanine analogues is previously described in Chapter 3. The complementary arrangement of xanthine and 8-oxoguanine within a G-tetrad (G•G•X•O tetrad) restores the hydrogen bonds pattern disrupted by a single mutation, without altering the original G-quadruplex conformation. This facilitates the selective isolation of a single G-quadruplex form, which is illustrated in Chapter 3.7.

Figure 4.1 G-quadruplex used in this study. A) Schematic representation of a human telomeric G-quadruplex with proposed modifications of X3 and O9. The G-tetrad modified with xanthine and 8-oxoguanine (G•G•X•O tetrad) undergoes a conversion from the B) unfavourable X•O arrangement to the C) favourable X•O arrangement. Hydrogen bonds affected by the modifications are shaded. Guanines in *syn* and *anti* glycosidic conformations are coloured in magenta and cyan respectively. The hydrogen-bond directionality in each tetrad is indicated by the grey arrow.

Recent studies demonstrated that 8-oxoguanine (O) and xanthine (X), two major naturally occurring DNA base lesions formed via oxidation and nitrosative deamination of the guanine base respectively^{103,116}, could be incorporated in a G-quadruplex in a complementary manner to form X•O•X•O^{114,263} or G•G•X•O tetrad (discussed in Chapter 3).

In the current study, we present how introduction of X and O in a non-complementary fashion can drive the reversal of G-tetrad hydrogen-bond polarity (Figure 4.1B, C). Using this approach, the polarity of a G-tetrad can be inverted without affecting the overall fold of the molecule. We modified a human telomeric sequence (*HT*) which has previously been shown to form a well-defined (3+1) G-quadruplex (Chapter 3) (Figure 4.1A) consisting of three G-tetrad layers. The upper G-tetrad of this molecule was modified with G3→X and G9→O substitutions (resulting in the *HT-3x9o* sequence, Table 4.1) introducing X and O in unfavorable orientation in

the original arrangement (Figure 4.1A,B). However, by inverting the glycosidic orientation of the bases, a G•G•X•O tetrad with the favourable X•O pair arrangement could be formed in the reverse orientation (Figure 4.1C), and characterized by NMR, CD and UV spectroscopy to determine the biophysical properties of the modified G-quadruplex.

4.2 The non-complementary X•O modification within a G-tetrad resulted in a predominant G-quadruplex form with altered folding topology

Characterization of *HT-3x9o* by NMR spectroscopy identifies 13 sharp imino protons at 10-12.5 ppm (Figure 4.2A), compared to 12 peaks seen for the unmodified *HT* G-quadruplex⁴⁹. As previously seen in Chapter 3, the additional peak arises from the H7 proton of the 8-oxoguanine in the G•G•X•O tetrad. Upon pH titration, specific imino proton peaks undergo major changes in chemical shift and broadening (Figure 4.3). This has been previously associated with protonation-deprotonation equilibrium of the xanthine base and indicates the involvement of X in G-tetrad formation (discussed in Chapter 3).

Figure 4.2 Characterization of *HT-3x9o* under basic condition (pH 9.0). A) Imino proton spectrum of *HT-3x9o*. Imino protons are labelled by their respective residue numbers. B) NOESY spectrum of *HT-3x9o* (mixing time, 250 ms). H1-H8 cross-peaks utilized for elucidation of folding topology are framed by empty rectangles, and labelled by the imino proton number, followed by that of H8. C) Characteristic H1-H8 NOE connectivity patterns around a G α •G β •G γ •G δ tetrad as indicated with red arrows. D) Characteristic H1-H8 NOE connectivities determined for O9•G17•G21•X3 tetrad (orange); G10•G16•G22•G4 tetrad (green); G11•G15•G23•G5 tetrad (purple).

Table 4.1 DNA sequence of *HT* and X•O modified G-quadruplexes, and their thermal stability.

Name	Sequence (5' – 3')				T_m (°C)	ΔT_m (°C)
	S A A	S A A	S S A	S A A		
<i>HT</i>	TT GGG	TTA GGG	TTA GGG	TTA GGG A	58.5 ± 0.4 ^a 59.4 ± 1.1 ^b	-
<i>HT-3x9o</i>	TT X GG	TTA O GG	TTA GGG	TTA GGG A	44.3 ± 0.7 ^a 45.1 ± 0.7 ^b	-14.2 -14.3
<i>HT-9x17o</i>	TT GGG	TTA X GG	TTA G O	TTA GGG A	47.8 ± 0.1 ^a 42.0 ± 0.0 ^b	-10.7 -17.4

The glycosidic conformation of each (modified) guanine base is positioned in *syn* (S) or *anti* (A) orientation.

^a T_m was determined with the sample at pH 7.0.

^b T_m was determined with the sample at pH 9.0.

Figure 4.3 Imino proton spectra of *HT*, *HT-3x9o* and *HT-9x17o* G-quadruplexes recorded at pH 7.0 (left) and pH 9.0 (right). Imino peaks corresponding to the major form are indicated by red dots.

The X and O lesions were introduced at position 9 and 17 respectively (*HT-9x17o*) (Table 4.1), opposite to the positions in which X and O mutations were introduced on *HT* G-quadruplex in Chapter 3 (*HT-17x9o*). Several imino peaks of *HT-9x17o* exhibit similar imino proton pattern as *HT-3x9o* (marked by red dots), suggesting the formation of the G•G•X•O-containing G-quadruplex. However, the imino proton pattern of *HT-9x17o* also demonstrated above the 13 imino peaks characteristic of X•O-modified G-quadruplexes at pH 9.0 (10 imino peaks at pH 7.0), indicative of the presence of multiple G-quadruplex forms (Figure 4.3). No improvement to the imino proton spectrum was observed for *HT-9x17o* when the pH is lowered to 7.0 (Figure 4.3). As the imino proton pattern of *HT-3x9o* suggests the formation of a predominant G-quadruplex, further structure analysis was carried out on this construct.

4.3 Elucidation of the folding topology of *HT-3x9o* via 2-dimensional NMR spectroscopy

To establish if the G-tetrad polarity has been successfully reversed by the X•O modification, the folding topology of *HT-3x9o* was determined using two-dimensional NMR experiments (Figure 4.2). This can be achieved via tracking the characteristic cyclic H1-H8 NOE patterns observed for *HT-3x9o* (Figure 4.2B, C) obtained using site-specific ^{15}N -labelled samples²²² (Figure 4.4; data not shown; Table 4.2) and cross-verified using NOESY experiments (Figure 4.5, 4.6). Resonance assignment of the imino and aromatic protons of X3 and O9 were confirmed by tracking NOE connectivities obtained from the NOESY experiments (Figure 4.6). Characteristic cyclic H8-H1 NOE patterns observed for *HT-3x9o* at pH 9.0 identifies the top, middle and bottom G-tetrad as O9•G17•G21•X3, G10•G16•G22•G4 and G11•G15•G23•G5 respectively (Figure 4.2D). Three high intensity cross-peaks corresponding to the intra-residue H8-H1' NOEs of G15, G16 and G17 was observed on the NOESY spectrum, which determined these residues to adopt the *syn* glycosidic conformation (Figure 4.6).

Figure 4.4 NMR spectral assignment of *HT-3x9o* at pH 9.0 showing the imino region. ^{15}N -filtered spectra was recorded on samples with 4 %- ^{15}N -enrichment at the indicated positions.

Figure 4.5 NOESY spectrum of *HT-3x9o* recorded at pH 9.0 (mixing time, 250 ms) with H1-H1/H7-H1/H1-H7 through space correlations labelled as indicated on the spectrum. NOE correlation from minor form is marked by asterisks (*).

Figure 4.6 Sequential 3-5, 10-11, 15-17 and 21-23 walk on NOESY spectrum of *HT-3x9o* (mixing time, 300 ms) at pH 9.0, based on H8-H1' NOE traced by red, blue, green and purple lines respectively. Intra-residue NOE correlations are labelled according to its respective nucleotide number. Cross-peaks of strong intensity are framed in black, and implicates residue 15, 16 and 17 to adopt the *syn* glycosidic orientation. Cross-peaks corresponding to intermolecular H8-H1' connections from *syn-syn* steps (each represented by an asterisk) are not seen on the spectrum.

Table 4.2 Sequence of ^{15}N -labelled *HT-3x9o* used in this study

Name	Sequence (5' – 3')					
<i>HT-3x9o</i> [4]	TT X [$^*\text{G}$]G	TTA O GG	TTA GGG	TTA GGG	TTA GGG	A
<i>HT-3x9o</i> [5]	TT X G[$^*\text{G}$]	TTA O GG	TTA GGG	TTA GGG	TTA GGG	A
<i>HT-3x9o</i> [10]	TT X GG	TTA O [$^*\text{G}$]G	TTA GGG	TTA GGG	TTA GGG	A
<i>HT-3x9o</i> [11]	TT X GG	TTA O G[$^*\text{G}$]	TTA GGG	TTA GGG	TTA GGG	A
<i>HT-3x9o</i> [15]	TT X GG	TTA O GG	TTA [$^*\text{G}$]GG	TTA GGG	TTA GGG	A
<i>HT-3x9o</i> [16]	TT X GG	TTA O GG	TTA G[$^*\text{G}$]G	TTA GGG	TTA GGG	A
<i>HT-3x9o</i> [17]	TT X GG	TTA O GG	TTA GG [$^*\text{G}$]	TTA GGG	TTA GGG	A
<i>HT-3x9o</i> [21]	TT X GG	TTA O GG	TTA GGG	TTA [$^*\text{G}$]GG	TTA GGG	A
<i>HT-3x9o</i> [22]	TT X GG	TTA O GG	TTA GGG	TTA G[$^*\text{G}$]G	TTA GGG	A
<i>HT-3x9o</i> [23]	TT X GG	TTA O GG	TTA GGG	TTA GG[$^*\text{G}$]	TTA GGG	A

X = xanthine, **O** = 8-oxoguanine, [$^*\text{G}$] = ^{15}N -labelled guanine

4.4 Incorporation of non-complementary **X•O** modification inverted G-tetrad polarity in *HT-3x9o*

The established hydrogen-bond pattern identifies X and O to be situated within the top G17•G21•X3•O9 tetrad (Figure 4.2D), in which X and O adopt a favourable arrangement as illustrated in Figure 4.1C. Although the overall fold adopted by *HT-3x9o* is unchanged from the (3+1) G-quadruplex fold of the original sequence *HT*⁴⁹, the modified tetrad has undergone a polarity inversion in the hydrogen-bond directionality (Figure 4.7).

Figure 4.7 Comparison between the *HT* and *HT-3x9o* G-quadruplexes. A) Schematic diagram of G-quadruplexes illustrating the wild-type *syn•anti•syn•syn* conformation in *HT* G-quadruplex (left) converting to *anti•syn•anti•anti* in *HT-3x9o* (right). *Syn* and *anti* guanines (analogues) are colored in magenta and cyan respectively. B) Schematic representation of the hydrogen bond directionality of each G-tetrad in *HT* and *HT-3x9o*. Each G-tetrad is simplified as a grey parallelogram, with the hydrogen-bond directionality is represented by black arrows. The reversal in directionality is represented by the inversion in colours.

A consequence of this G-tetrad inversion is a change in the circular dichroism (CD) profile. The CD spectrum of *HT-3x9o* is notably different from the classical (3+1) pattern observed for the *HT* sequence (Figure 4.8A), whereby the CD spectrum of *HT-3x9o* exhibits a negative peak at 240 nm and positive peak at 265 nm similar to the CD signature of a parallel G-quadruplex. The dramatic shift in CD profile of *HT-3x9o* is associated with its non-alternating G-tetrad polarity, where the G-tetrads are aligned and stacked in a manner similar to that observed in parallel G-quadruplexes^{295,296}, and was previously reported by Dickerhoff and Weisz for another G-quadruplex with an inverted G-tetrads²⁹⁴.

Figure 4.8 CD spectra of *HT*, *HT-3x9o* and *HT-9x17o* G-quadruplexes at pH 9.0.

As changing the hydrogen bond directionality of the G-quadruplex is intrinsically linked to a flip in glycosidic torsion angles in the modified tetrad (Figure 4.1B, 4.7), the G17•G21•X3•O9 tetrad adopts a *syn•anti•syn•syn* configuration, consistent with the intensity observed for intra-residue H8-H1' NOEs (Figure 4.6). This is in contrast to the reported structure of *HT*, in which the G17•G9•G3•G21 tetrad adopts an *anti•syn•anti•anti* configuration⁴⁹. The

resulting G-quadruplex scaffold adopted by *HT-3x9o* positioned O in an unfavourable *anti* glycosidic position^{265,297}, which destabilized *HT-3x9o* by 14 °C in comparison to the original structure (Table 4.1).

To understand the effects of positioning O in the favourable *syn* glycosidic position with inversion of G-tetrad polarity in an X•O-modified G-quadruplex, the thermal stability of *HT-9x17o* was determined. However, the incorporation of X and O destabilized *HT-9x17o* by 11°C and 17 °C for pH 7.0 and pH 9.0, respectively in comparison to the unmodified G-quadruplex. The CD spectrum of *HT-9x17o* (Figure 4.8B) appeared to display a mixture of CD profile from that of the (3+1) and parallel G-quadruplex topology, which correlated with the multiple G-quadruplex forms as identified by its imino proton spectrum (Figure 4.1B). Consequently, the strong destabilization observed on *HT-9x17o* may not be a true representation of the thermal stability of a G•G•X•O-containing G-quadruplex. Further studies are required to understand the incomplete conversion from the unfavourable to favourable G•G•X•O tetrad arrangement in *HT-9x17o*, and to determine the thermal stability of such G•G•X•O-containing G-quadruplex.

4.5 Conclusion

In conclusion, we have demonstrated that a judicious placement of two base modified guanine analogues, X and O within a G-tetrad may drive the reversal of G-tetrad polarity (Figure 4.7), while conserving the overall fold of the unmodified G-quadruplex. This approach can thus be applied as a tool to programme the hydrogen-bond directionality within individual G-tetrads. The X•O modification may have potential in drug-based or nanomaterials applications, and facilitate fundamental studies of the G-quadruplex structure without modifying the sugar moiety. The application of X and O within the G-quadruplex scaffold also has the additional benefit of possessing a highly electronegative groove which may modulate its properties, as demonstrated in Chapter 3.

Chapter 5 Insights into G-quadruplex specific recognition by the DEAH-box helicase Rhau: Solution structure of a peptide-quadruplex structure Structural basis for complex

G-quadruplexes are four-stranded nucleic acid structures implicated in important biological processes^{8,298,299}. Interactions between G-quadruplexes and proteins are essential for the cellular functions of G-quadruplexes^{9,12,79,300}. The DEAH-box helicase Rhau (also named as DHX36 or G4R1) specifically binds and unwinds G-quadruplexes, with preferences towards parallel RNA/DNA G-quadruplexes over duplexes and non-parallel G-quadruplex^{148,182-184,201}. We have refined the key G-quadruplex-binding domain of Rhau to an 18 residue peptide located within the N-terminal region of Rhau (hereby termed as Rhau₁₈), and resolved its structure in complex with a parallel G-quadruplex. The peptide almost covers the entire 5'-end top G-tetrad, and it interacts with DNA using CH/ π and CH₃/ π -stacking interactions, and electrostatic interactions. The

structure of the complex explains how Rhau specifically recognizes parallel G-quadruplex over other conformations, and suggests a general strategy for G-quadruplex recognition by proteins.

5.1 Introduction

Potential G-quadruplex-forming sequences are prevalent in many important regions of the human genome⁶. It has been now established that G-quadruplexes are formed in the cell and are associated with essential cellular processes such as transcription, replication and recombination^{8,298,299}. Interaction with proteins is required for the cellular function of G-quadruplexes and proteins that interact with G-quadruplexes have been increasingly identified¹³³. For example, helicases, such as BLM⁷⁹, WRN³⁰⁰, FANCI¹² and Pif1⁹, have been shown to bind and unwind G-quadruplex structures.

Structural studies on protein-G-quadruplex interactions have been attempted by different groups. Computational docking studies showed that a G-quadruplex structure could fit into a protein cavity of HIV-integrase and RecA protein^{95,208}. In an NMR-guided simulation, a Nucleophosmin domain was docked onto a G-quadruplex groove¹⁸⁰. Limited available structures of G-quadruplex-protein complexes did not show direct contacts between the proteins and the G-tetrad core; the binding was either dependent on a loop^{54,211-213} or a duplex-quadruplex junction formation^{214,215}. Recently, NMR and simulation studies reported on a complex, where two anti-prion peptide molecules bound at the two ends of an RNA G-quadruplex stack, driven by water entropy gain^{216,301}. However, from the studies reported to date the structural basis for specific G-quadruplex recognition by proteins has not been fully understood.

Rhau, a helicase of the DEAH-box family, has been shown to bind tightly to and resolve G-quadruplex structures with high specificity^{148,182,201}. Rhau is associated with different functions, including the formation of stress granules, interchromatin granule clusters^{153,156} and the degradation of urokinase plasminogen activator mRNA¹⁵². The N-terminal region is not conserved amongst the DEAH-box family and is responsible for the specificity of the helicase³⁰². Rhau was found to be associated with telomerase through binding to the G-quadruplex formed at the 5'-end of the telomerase RNA¹⁸³. The 53-amino-acid fragment (residues 53–1050 of Rhau (termed Rhau₅₃ in this paper, Table 5.1) binds specifically to parallel G-quadruplexes. A 13-amino-acid

Rhau-Specific Motif (RSM) spanning from residues 53–66 has been identified to be essential for the binding¹⁸⁴. SAXS data suggested that Rhau₅₃ binds at an end of a G-quadruplex³⁰³.

In this chapter, the NMR solution structure of a complex formed between a parallel G-quadruplex and a peptide from the N-terminal region of Rhau is reported, revealing general principles for specific recognition of parallel G-quadruplexes by Rhau.

5.2 Rhau₅₃ binds preferentially to parallel G-quadruplexes

Figure 5.1 Native gel electrophoresis of DNA in the absence and presence of Rhau₅₃ at 1:10 ratio. A) *T95-2T*; B) *HT1*; C) *HT1* under crowding conditions. Concentrations of DNA and Rhau₅₃ was 50-100 nM and 500-1000 nM respectively. Filled and open triangle symbols represent the parallel and non-parallel G-quadruplex form of *HT1*, respectively; D) Gel binding analysis for *T95-2T* (asterisk), *HT1* (open triangle) and *HT1* under crowding conditions (filled triangle). Gels were visualized using fluorescence probe attached to the 5'-end of DNA.

G-quadruplex structures are highly diverse regarding strand orientation and loop configurations⁶⁸. Gel electrophoresis shows that Rhau₅₃ binds to the parallel G-quadruplex formed by the *T95-2T* sequence²³⁶ (Figure 5.1A) but not the non-parallel G-quadruplex formed by the human telomeric sequence *HT1*⁴⁹ (Figure 5.1B). To further consolidate the parallel G-quadruplex specificity of Rhau₅₃, conformational transition via addition of poly(ethylene glycol) 200 (PEG 200) was induced on the *HT1* G-quadruplex for conversion from the native (3+1) form to a parallel conformation, which will be screened against Rhau₅₃.

Such cosolute-induced conformational change has been described by Miyoshi *et. al.* to simulate molecular crowding³⁰⁴, in which the driving force behind the conformational change has been associated to the effect of water depletion by Heddi *et. al.*²⁸⁷, although Miller *et. al.* has implicated the change in water hydration to have induced the change in G-quadruplex forms instead^{286,289,290}. As the addition of PEG was utilized purely as a tool for the induction of parallel

Figure 5.2 Comparative binding of Rhau₅₃ with non-parallel G-quadruplexes (G4s) A) *HT2*; B) *HT3*; C) *HT4* in dilute (left) and crowding condition (right). Native gel electrophoresis was conducted with 100 nM DNA in the absence and presence of Rhau₅₃ at 1:10 ratio.

Figure 5.3 EMSA of DNA with increasing concentration of Rhau₅₃: A) *HT1*; B) *HT1* under crowding conditions; C) *T95-2T*. Concentration of DNA was 100 nM; concentration of Rhau₅₃ concentration was 0, 100, 250, 500, 700, 1000, 2500, 5000, 10000, 25000 nM. The folding topology of each sequence is shown next to the sequence name, *Anti* and *syn* guanines are coloured in cyan and magenta, respectively.

Figure 5.4 NMR imino protons of DNA/RNA G-quadruplexes and duplex sequences used in this study (Table 4.1). Name of the sequences are labelled above each spectrum. The folding topology of G-quadruplexes and duplexes are shown next to each spectrum. *Anti* and *syn* guanines are coloured in cyan and magenta respectively; cytosines are colored in brown; bases in DNA duplexes are coloured in red.

Figure 5.5 Native gel electrophoresis of different DNA and RNA in the absence and presence of Rhau₅₃: A) parallel DNA G-quadruplexes; B) parallel RNA G-quadruplexes; C) non-parallel DNA G-quadruplexes; D) Non-parallel *HT4* G-quadruplex with longer incubation time (2 h); E) DNA duplexes. Names of the sequences are indicated above the lane. Ratio between nucleic acids and Rhau₅₃ is indicated on top of each lane. DNA concentrations were fixed at 100 μ M. The gel was revealed using UV shadowing.

Figure 5.6 Demonstration of preferential binding affinity of Rhau₅₃ towards parallel G-quadruplexes. A) Folding topology of the sequence *HT* and *T95-2T*; B) NMR spectra of a mixture of *HTI* (red open triangle) and *T95-2T* (black dot). The imino region of the DNA mixture in the absence (below) and the presence (above) of Rhau₅₃ are showed. The ratio between DNA and Rhau₅₃ are indicated next to each spectrum.

Table 5.1 DNA/RNA sequences used in binding study of Rhau peptides.

	Name	Sequence (5' - 3')	Structure ^a	Binding ^b
RNA	<i>TERC18</i>	GGGUUGCGGAGGGUGGGC	Parallel ³⁰⁵	+
	<i>T95-2T^c</i>	TTGGGTGGGTGGGTGGGT	Parallel ²³⁶	+
	<i>Pu24T</i>	TGAGGGTGGTGAGGGTGGGA	Parallel ³⁰⁶	+
	<i>CEB1</i>	AGGGGGGAGGGAGGGTGG	Parallel ³⁰⁷	+
	<i>CEB25</i>	AAGGGTGGGTGTAAGTGTGGGTGGGT	Parallel ³⁰⁸	+
DNA	<i>93del</i>	GGGGTGGGAGGAGGGT	Parallel ⁹⁵	+
	<i>HT^d</i>	TTGGGTTAGGGTTAGGGTTAGGGA	Non-Parallel ⁴⁹	—
	<i>HT1^d</i>	TAGGGTTAGGGTTAGGGTTAGGG	Non-Parallel ⁴⁹	—
	<i>HT2^d</i>	TAGGGTTAGGGTTAGGGTTAGGGTT	Non-Parallel ⁴⁸	—
	<i>HT3^e</i>	GGGTTAGGGTTAGGGTTAGGGT	Non-Parallel ⁵¹	—
	<i>HT4^d</i>	AGGGCTAGGGCTAGGGCTAGGG	Non-Parallel ³⁰⁹	—
	<i>DX_a</i>	CGCGAATTCGCG	Duplex ³¹⁰	—
	<i>DX_b</i>	ATCTGAGAATCAGAT	Duplex Hairpin ³¹¹	—

^a Structures formation were monitored by NMR

^b A negative binding is considered when the dissociation constant (K_D) is larger than 10 μ M.

^c TAMRA attached to the 5' end

^d FAM attached to the 5' end

^e FAM attached to the 3' end

G-quadruplex structures from their non-parallel forms, we have termed these G-quadruplexes to be formed under molecular crowding.

Conformational transition resulting from molecular crowding for non-parallel G-quadruplexes was verified for the *HT1* G-quadruplex by CD and NMR spectroscopy²⁸⁷, although such transition was not observed for parallel G-quadruplex structures²⁹¹. Corresponding, under molecular crowding condition, *HT1* shows an upper additional band corresponding to the parallel G-quadruplex form, which binds to Rhau₅₃ (Figure 5.1C). Similar results were observed for different human telomeric G-quadruplex conformations (Figure 5.2). Quantitative gel analysis (Figure 5.3) shows that the binding affinity of Rhau₅₃ to parallel G-quadruplexes formed by either *T95-2T* or *HT1* ($K_D \sim 1 \mu\text{M}$) is about 100-fold higher than to non-parallel G-quadruplex of *HT1* (Figure 5.1D).

Assays on different nucleic acid conformations (Table 5.1, Figure 5.4) show that Rhau₅₃ binds to various parallel DNA and RNA G-quadruplexes, but not non-parallel DNA G-quadruplexes and duplexes (Figure 5.5). Rhau₅₃ also failed to bind with the non-parallel *HT4* DNA G-quadruplex even after extension of the incubation from 30 mins to 2 hours (Figure 5.5D). This conclusion is supported by an NMR titration experiment, which shows that upon addition of Rhau₅₃ in a mixture of a parallel and a non-parallel G-quadruplex only peaks belonging to the parallel form are perturbed (Figure 5.6).

Creacy *et. al.* have previously reported Rhau to manifest 2-fold higher binding affinity for a tetramolecular RNA G-quadruplex over a similar G-quadruplex formed from DNA sequences¹⁴⁸. However, the use of a RNA G-quadruplex for NMR studies will be more challenging as compared to that of a DNA G-quadruplex due to the intrinsic spectral complexity of a RNA³¹². Special precautions need to be taken when handling RNA samples as degradation of the sample is an issue that may arise easily due to contamination with RNase³¹³. Moreover, higher costs are involved in the synthesis of ¹⁵N-labelled and unlabelled RNA sequences, with the usage of a RNA G-quadruplex for structure determination of a G-quadruplex-protein complex to be less cost effective. With the goal of this chapter being the elucidation of the G-quadruplex-Rhau binding motif, it may be assumed that there are negligible differences between the binding motifs formed by DNA/RNA G-quadruplexes, on the basis of the structural similarity between a DNA and a

RNA. This is further supported by the relatively small difference in binding affinity between RNA and DNA G-quadruplexes for Rhau. Consequently, the Rhau-G-quadruplex interaction was next studied using *T95-2T*, a simple model of a parallel DNA G-quadruplex, which forms a propeller-type G-quadruplex with three single residues loops²³⁶.

Figure 5.7 ^1H , ^{15}N -HSQC of free Rhau₅₃ (blue) and in complex with *T95-2T* at 1:1 ratio (red). B) Chemical shift variation for all the assigned residues of Rhau₅₃. Inlet box showing the 10 lowest energy of Rhau₁₈

Figure 5.8 ^1H , ^{15}N -HSQC spectra of Rhau₅₃ free (black) and bound to *T95-2T* (red). Boxes represent peaks observed at a lower threshold. Each dotted arrow shows the changes in chemical shifts of a residue in the free and bound form. Assignments of resonances in the Rhau₅₃-*T95-2T* complex are shown.

Figure 5.9 Native gel electrophoresis binding assay between *T95-2T* and different peptides. Lanes 1 and 13 *T95-2T*; lane 2-12: *T95-2T* in the presence of Rhau_{20m1}, Rhau_{20m2}, Rhau₅, Rhau₉, Rhau₁₂, Rhau₁₄, Rhau₁₆, Rhau₂₀, Rhau₁₈, Rhau₂₉ and Rhau₂₃. DNA concentration was 100 μ M. The DNA/peptide ratio was 1:5. DNA on the gel was revealed using UV shadowing.

Table 5.2 Peptide sequences used to determine the minimal G-quadruplex binding domain of Rhau.

Name	Sequence (N _{term} - C _{term}) ^a	Binding ^b
Rhau ₅₃	<i>SM</i> HPGHLKGREIGMWYAKKQGQKNKEAERQERAVVHMD ERREEQIVQLLSVQAK	+
Rhau ₂₉	HPGHLKGREIGMWYAKKQGQKNKEAERQE	+
Rhau ₂₃	HPGHLKGREIGMWYAKKQGQKNK	+
Rhau ₂₀	HPGHLKGREIGMWYAKKQGQ	+
Rhau _{18'}	HPGHLKGREIGMWYAKKQ	+
Rhau ₁₈	<i>SM</i> HPGHLKGREIGMWYAKKQ	+
Rhau ₁₆	HPGHLKGREIGMWYAK	+
Rhau ₁₄	HPGHLKGREIGMWY	–
Rhau ₁₂	HPGHLKGREIGM	–
Rhau ₅	HPGHL	–
Rhau ₉	KGREIGMWY	–
Rhau _{20m1}	HP <u>L</u> <u>H</u> <u>L</u> <u>K</u> <u>L</u> REILMWYAKKQGQ	–
Rhau _{20m2}	HPGHLKGREIGM <u>AA</u> AKKQGQ	–

^a Residues in grey are a protease cleavage and cloning artefact, point mutation of residues are indicated in red and underlined

^b Non-binding is considered when the dissociation constant (K_D) is larger than 100 μ M

5.3 Minimal G-quadruplex binding domain of Rhau

Backbone resonances were assigned for Rhau₅₃ in the free form and in complex with *T95-2T* (Figure 5.7A). Upon complex formation, the chemical shifts of the first 21 amino acids of Rhau₅₃ were largely perturbed, while no significant changes were observed for the rest of the sequence (Figure 5.7B; Figure 5.8), suggesting that the G-quadruplex-binding of Rhau₅₃ is limited to the first 21 amino acids. Binding assays (Figure 5.9) were performed between *T95-2T* and peptides of different lengths (Table 5.2). The 16-amino-acid and 18-amino-acid peptides termed

Rhau₁₆ and Rhau₁₈ respectively, containing the RSM motif were the shortest peptides capable of binding to *T95-2T*. NMR spectra also show the binding of *T95-2T* with Rhau₁₆, Rhau₁₈ and Rhau₅₃ (Figure 5.10). Rhau₁₈ and Rhau₅₃ exhibit similar spectral changes upon DNA binding (Figure 5.10; Fig 5.11). Like Rhau₅₃, Rhau₁₈ binds specifically to parallel G-quadruplex, but not non-parallel G-quadruplex and duplex DNA, as observed by gel electrophoresis (Figure 5.12).

Figure 5.10 NMR imino proton spectra of *T95-2T* in the absence and presence of A) Rhau₅₃; B) Rhau₁₈; and C) Rhau₁₆. Imino proton spectra of *T95-2T* in its free and bound form are shown above and below, respectively. DNA concentration was 100 μ M. The amount of peptide in the titration is half that of the DNA, and thus the DNA/peptide ratio is represented as 1:0.5. Imino protons of DNA in the complex are marked by asterisks.

Figure 5.11 ^1H , ^{15}N -HSQC spectra of Rhau₁₈ free (black) and bound to *T95-2T* (red). Dotted arrow shows the change in chemical shifts of a residue in the free and bound form.

Figure 5.12 Native gel electrophoresis of DNA in with and without the presence of Rhau₁₈. Ratio of DNA to protein are indicated on top of each lane. Name of the sequences are indicated on top of each lane. *HT* parallel conformation was obtained using crowding conditions.

5.4 NMR solution structure of Rhau₁₈

The solution structure of Rhau₁₈ (Appendix F for statistics) was determined using standard triple-resonance NMR techniques (described in detail in Chapter 2.2.1). It presents a L-shaped structure (Figure 5.7B, insert), containing an α -helix spanning from Gly5 to Ala17 with a hydrophobic core that comprises Leu7, Ile12 and Trp15 (Figure 5.13).

Figure 5.13 A) Cartoon representation of the ten superimposed structures of Rhau₁₈; B) Side view showing short distances between residues P4, L7, I12 and W15.

Figure 5.14 A) imino proton spectra of free *T95-2T* (bottom spectra) and in complex with Rhau₁₈ at 1:1 ratio. Imino peaks corresponding to the *T95-2T*-Rhau₁₈ complex are marked by asterisks; B) G-tetrad representation and folding topology of *T95-2T*; C) NOESY (mixing time, 700 ms) spectrum centred on the imino proton region showing the exchange between the free and bound form of *T95-2T*, exchange peaks are labelled according to the residue number; D) Strips plots of the filtered $\{^{13}\text{C}, ^1\text{H}\}$ HSQC-NOESY, showing the inter-cross peaks between the peptide and the

DNA.

5.5 NMR solution structure of Rhau₁₈-G-quadruplex complex

Addition of Rhau₁₈ into *T95-2T* resulted in the appearance of a new set of peaks in the DNA imino proton spectrum (Figure 5.14A). Signals belonging to DNA in the Rhau₁₈-*T95-2T* complex were assigned through chemical exchange (Figure 5.14C) with those of the previously assigned free *T95-2T*²³⁶. NMR data (Figure 5.14A, C) indicated that *T95-2T* maintained its three-layered propeller-type parallel G-quadruplex structure upon protein binding (Figure 5.14B).

A complex between ¹⁵N,¹³C-labelled Rhau₁₈ and an excess of unlabelled *T95-2T* was formed. ¹⁵N, ¹³C and ¹H resonances were assigned for Rhau₁₈ in the complex using standard triple resonance experiments (Figure 5.15A). 44 intermolecular NOEs cross-peaks were identified between Rhau₁₈ and *T95-2T* (Figure 5.14D, 5.15B).

Figure 5.15 Strip plots illustrating assignment of Rhau₁₈ in complex with *T95-2T*. A) HNCACB/CBCACONH strips plots showing example of the backbone sequential connectivity of Rhau₁₈ in complex with unlabelled *T95-2T*. B) NOE strip plots showing the intermolecular cross-peaks between the peptide and DNA.

Figure 5.16 NMR solution structure of the Rhau₁₈-T95-2T complex. A) Side- and B) top-view of the ten superimposed lowest-energy structures; C) Side- and D) top-view surface representation of the complex; E – F) Details of peptide-DNA interactions. Rhau₁₈ is coloured in red; guanines, cyan; thymines, orange; DNA backbone, grey; O4' atoms, yellow.

Figure 5.17 Comparison between free and bound structures in the Rhau₁₈-T95-2T complex. A) Superimposition of the free (red, PDB ID 2LK7) and bound (blue) T95-2T; B) Superimposition of

the free (red) and bound (blue) Rhau₁₈.

Figure 5.18 A) Structure representation of Rhau₁₈ in complex with *T95-2T*, dotted lines represent intermolecular NOEs observed between the DNA and the 5' tetrad. B) NOESY spectrum showing the intermolecular NOEs between the protein and the DNA imino proton.

In the solution structure of the Rhau₁₈-*T95-2T* complex (Figure 5.16; Appendix F for statistics), both Rhau₁₈ and *T95-2T* retained their overall initial shapes (Figure 5.17). Rhau₁₈ binds to the 5' end G-tetrad of *T95-2T*, projecting the first two thymine bases outwards (Figure 5.16). The N-terminal residues of Rhau₁₈ are oriented away from *T95-2T*, while the C-terminal residues were found to be close to a medium groove of the DNA (Figure 5.16C). The top G-tetrad is almost entirely covered by the peptide (Figure 5.16D). The CH/ π and CH₃/ π -stacking interactions are observed between peptide residues (Ile12, Gly13 and Ala17) and DNA guanine bases (G3, G11 and G15) of the 5'-end G-tetrad (Figure 5.16E; Figure 5.18). The positively charged side-chains of Lys8, Arg10, and Lys19 are each located above a groove; electrostatic interactions are favoured by the close proximity between these residues and the phosphate groups of G4, G12 and G16, respectively (Figure 5.16F).

5.6 Binding of a second peptide molecule

When Rhau₁₈ was titrated at half the original amount of the DNA (DNA/peptide ratio 1:0.5), the NMR proton spectrum shows two sets of peaks for DNA and only one for the peptide; while at the 1:2 DNA/peptide ratio, there are only one set of peaks for DNA and two for the peptide (Figure 5.19A). These data are consistent with the formation of a first complex containing 1 peptide molecule (at the 1:0.5 DNA/peptide ratio) and the formation of a second complex

Figure 5.19 Diagram illustrating different complex state of *T95-2T* with Rhau_{18'}. A) NMR imino proton spectra of *T95-2T* with different amount of Rhau_{18'}. The amount of peptide added to the DNA sample is represented as the DNA/protein ratio and indicated next to each spectrum. When half portion of Rhau_{18'} was added to the DNA (1:0.5 DNA/peptide ratio), the NMR proton spectrum shows two sets of peaks for DNA and only one for the peptide (imino protons in the complex are marked by asterisks); while at the 1:2 DNA/peptide ratio, there are only one set of peaks for DNA and two for the peptide (signal of Tryptophan HE1 is labelled W15); B) Model for binding between *T95-2T* and Rhau_{18'} showing the formation of a first complex containing 1 peptide molecule (at the 1:0.5 DNA/peptide ratio) and the formation of a second complex containing 2 peptide molecules (at the 1:2 DNA/peptide ratio); C) Imino protons of DNA in the new second complex were unambiguously assigned using site-specific ¹⁵N-labeled samples. Assignments of guanines imino protons of *T95-2T* in the presence of an excess of Rhau_{18'} were obtained from ¹⁵N-filtered experiments, run on samples containing ~4% ¹⁵N-labeled at the indicated position. Reference spectrum is shown on the top; D) NOESY spectrum (mixing time, 350 ms) of *T95-2T* in the presence of an excess of Rhau_{18'}, showing intermolecular NOEs between different protons of Rhau_{18'} and the guanines imino protons of *T95-2T* at the two terminal G-tetrads (labelled in blue for the top 5'-end G-tetrad and red for the bottom 3'-end G-tetrad respectively). NOE data indicate that the second peptide molecule binds at the bottom 3'-end G-tetrad of *T95-2T*. The binding at the 3'-end occurs at a higher peptide concentration than the binding at the 5'-end, indicating a lower affinity for the 3' bottom site.

containing 2 peptide molecules (at the 1:1 DNA/peptide ratio). Imino protons of DNA in the new second complex were unambiguously assigned using site-specific ^{15}N -labeled samples (Figure 5.19C). NOE data (Figure 5.19D) indicate that the second peptide molecule binds at the bottom 3'-end G-tetrad of *T95-2T* (Figure 5.19). The binding at the 3'-end occurs at a higher peptide concentration than the binding at the 5'-end, indicating a lower affinity for the 3' bottom site (Figure 5.19). These differences in affinities might be related to different arrangements of the backbone and orientations of the sugars between the 5'- and 3'-end G-tetrads.

5.7 Exposed G-tetrad is determinant for G-quadruplex specific recognition by Rhau

The size and shape of the Rhau binding motif match well with the unique geometry of a G-tetrad. The peptide covers the exposed hydrophobic G-tetrad and clamps the G-quadruplex structure using 3-anchor-point electrostatic interactions (between three positively charged amino acids and phosphate groups). The structure of the *Rhau*₁₈-*T95-2T* complex explains how Rhau specifically recognizes parallel G-quadruplexes based on their G-tetrad accessibility. In non-parallel G-quadruplexes, G-tetrads are generally covered by different types of loops, which would sterically hinder protein binding. Rhau-G-quadruplex recognition might be more directly associated to an exposed G-tetrad rather than a specific G-quadruplex topology (parallel *vs.* non-parallel). It should be noted that a left-handed parallel G-quadruplex structure with well-covered terminal G-tetrads has been recently reported³¹⁴; on the other hand, the formation of a non-parallel G-quadruplex with exposed G-tetrad might be possible.

However, it is difficult to rationalize the tighter binding affinity for RNA G-quadruplex exhibited by Rhau as reported by Creacy *et. al.*¹⁴⁸ based on the binding mode of a DNA G-quadruplex since the nucleobases and the phosphate backbone from the oligoribonucleotide and oligodeoxynucleotide are structurally similar. Moreover, as the folding topology of the DNA G-quadruplex used in the binding affinity screening by Creacy *et. al.* was not assessed by CD or NMR spectroscopy. As exemplified by another tetramolecular G-quadruplex of the sequence (GGGGCC)₄, the oligodeoxynucleotides were evaluated via CD spectroscopy to assemble into anti-parallel G-quadruplex topology in the presence of K^+ , differing from the parallel form

adopted by the oligoribonucleotides³¹⁵. Thus, it is not known if the 2-fold preference for RNA G-quadruplexes is due to RNA G-quadruplexes being the preferred binding target for Rhau, or a result of preferentiality for parallel G-quadruplex structures (with exposed terminal G-tetrads) since RNA G-quadruplexes predominantly adopts the parallel form³¹⁶. Therefore, further research is required to verify the tighter binding exhibited by Rhau for RNA G-quadruplexes using structurally resolved RNA and DNA G-quadruplexes of the same sequence, which is beyond the scope of this thesis.

5.8 Implications for the elucidation of the Rhau₁₈-G-quadruplex complex structure in targeting of G-quadruplexes

The G-quadruplex binding mode of Rhau is strikingly similar to that of most small molecules selected for specific G-quadruplex binding²⁷⁴, which occur through stacking interactions on the terminal G-tetrads and electrostatic interactions with the phosphate backbone. However, as opposed to Rhau, most small molecules do not distinguish between different G-quadruplex conformations, probably due to their small sizes allowing them to be accommodated on a terminal G-tetrad even in the presence of loops.

Binding to an exposed G-tetrad represents a simple and efficient way to specifically target G-quadruplex structures – a strategy that could also be adopted by other G-quadruplex-binding proteins. Based on this work, the binding domain can be modified to modulate the affinity and specificity to parallel G-quadruplexes, and possibly also other G-quadruplex conformations. The G-quadruplex binding motif can be engineered in other proteins to achieve G-quadruplex-specific functions, such as the regulation of gene expression.

5.9 Conclusion

In this chapter, an 18-amino-acid G-quadruplex-binding domain in the N-terminal region of Rhau was identified, and the solution structure of this peptide bound to a parallel DNA G-quadruplex was resolved. Our structure explains how Rhau specifically recognizes parallel G-quadruplexes. The peptide covers a terminal guanine base tetrad (G-tetrad), and clamps the G-quadruplex using 3-anchor-point electrostatic interactions between three positively charged amino acids and phosphate groups. This binding mode is strikingly similar to that of most ligands

selected for specific G-quadruplex targeting. Binding to an exposed G-tetrad represents a simple and efficient way to specifically target G-quadruplex structures, suggesting a general strategy for G-quadruplex recognition by proteins.

Chapter 6 Structural and binding study of G-quadruplexes with the conserved domains of BLM

BLM is a RecQ family helicase associated with Bloom's syndrome¹⁸⁸, a rare genetic disorder characterized by a predisposition to cancer and chromosomal instability³¹⁷. Within the BLM protein, four highly conserved domains have been identified: helicase domain, zinc-binding domain (ZnBD), RQC domain and HRDC domain (Figure 6.1)¹⁸⁸. BLM has been demonstrated to exhibit selectivity for resolving G-quadruplexes over canonical forms of DNA, and the fragment comprising ZnBD and RQC domains has been previously identified to be a key G-quadruplex binding region⁹⁰. To elucidate the nature of DNA interaction that occurs between the two domains with G-quadruplex, comparative study was conducted on BLM constructs containing the combined ZnBD and RQC domains, and their isolated domains. However, the solubility study (described in detail later) has demonstrated that the protein with combined domains and the isolated ZnBD protein exhibit low solubility, while the RQC domain is highly solubilised. Using NMR spectroscopy, we have obtained the solution structure of BLMRQC, which was found to adopt the signature winged-helix motif within its globular structure. Comparative binding assays

on the combined domain and BLMRQC demonstrated that the tight DNA binding capability exhibited by the G-quadruplex binding motif is lost with the removal of ZnBD in BLMRQC.

6.1 Introduction

Bloom's syndrome is a rare genetic disorder characterized by an increased predisposition of cancer, marked increase in sister chromatid exchanges and gross chromosomal rearrangement³¹⁷. This disease arises as a result of mutation(s) in the BLM protein, a DNA helicase belonging to the RecQ family^{188,317}. As BLM is involved in DNA replication and repair^{141,191,193,195}, and it resolves a wide range of targets, ranging from the single-stranded DNA, duplex DNA to higher order DNA structures such as Holliday junction and G-quadruplex^{79,149,167,202,206,318}, the BLM protein is thus essential for the maintenance of genomic stability^{134,138,140,187,199,200,319-321}.

Comparison between the BLM and its homologues with similar G-quadruplexes resolving capability (WRN²⁰², *E. coli* RecQ⁹⁰ and Sgs1¹⁴⁹) finds only the BLM and WRN to exhibit selectivity for G-quadruplexes over canonical forms of DNA^{132,202,318}. Within the BLM, a key G-quadruplex-binding region containing the zinc-binding domain (ZnBD) and RecQ C-terminal domain (RQC) was identified by Huber *et. al*⁹⁰. Comparison between a truncated BLM fragment containing only these two domains and the full-length BLM has found both constructs to exhibit similar DNA binding capability⁹⁰, with the DNA binding interface determined to be localized within the RQC domain¹⁶⁷.

The structure of WRN protein (RQC domain) in complex with duplex DNA has been resolved, which identified its core DNA-binding interface to be located on the winged-helix motif of the RQC domain¹⁶⁹. Such DNA-binding site was similarly identified for RecQ1^{204,205}, which does not resolve G-quadruplexes, and also in a recent structure of the BLM protein core¹⁶⁷. However, the structure of a BLM-G-quadruplex complex remains elusive, and the lack of a defined G-quadruplex binding site results in poor understanding of the factors contributing to the tight G-quadruplex binding of BLM protein. To address these issues, the protein-DNA interactions between G-quadruplexes and the BLM G-quadruplex binding region will be studied via NMR spectroscopy, and the solution structure of the complex(es) will be attempted for the identification of the G-quadruplex specific recognition motifs on the BLM protein.

6.2 The combined zinc-binding domain with RQC domain exhibits poor solubility while the isolated RQC domain is highly soluble

Figure 6.1 Schematic representation of conserved domains in human BLM and BLM fragments used in this study. All conserved domains are represented by rectangles and coloured in the following order: helicase domain in white; zinc-binding domain (ZnBD) in black; RecQ-C-terminal domain (RQC) in blue; helicase-and-ribonuclease D/C terminal domain (HRDC) in green. Full length BLM spans from residues 1–1417. The original BLM fragment was found to bind tightly with G-quadruplex spans from residues 992–1192 (BLMZRQC). Other constructs used in this study include the isolated ZnBD of BLM (BLMZ) which spans from residues 992–1069; the RQC domain from residues 1070–1192 (BLMRQC); the combined helicase – ZnBD and RQC domain from residues 642–1192 (BLMhZRQC); the combined helicase and ZnBD fragment from residues 642–1069 (BLMhZ); and the isolated helicase domain from residues 642–991 (BLMh).

To determine the individual contributions from each conserved domain in G-quadruplex binding, truncated BLM fragments containing different conserved domain(s) were constructed as follow: the ZnBD fragment (BLMZ), the RQC fragment (BLMRQC), the combined ZnBD and RQC fragment (BLMZRQC), the combined helicase and ZnBD fragment (BLMhZ), and the combined helicase, ZnBD and RQC fragment (BLMhZRQC) (Figure 6.1). The boundaries of each BLM construct were determined via alignment of conserved domains in BLM homologues, using comparisons of truncated BLM proteins reported in literature²⁰⁶.

The production of the BLMZRQC protein was first attempted due to its reported high affinity to G-quadruplexes⁹⁰. As illustrated in Figure 6.2A, BLMZRQC exhibited good protein expression levels, although majority of its expressed protein formed inclusions bodies. Due to the limited quantities of solubilized protein available for native purification, only 40 µg of purified BLMZRQC of ~ 60 % purity was obtained per litre of media after purification via affinity and gel filtration chromatography (Figure 6.2A, lane 13). An attempt was made to purify the remaining

insoluble BLMZRQC under denaturing conditions, where BLMZRQC was solubilised and refolded according to a two-step denaturing and refolding procedure optimized for zinc-finger proteins³²² (Appendix C). However, the solubilised BLMZRQC precipitated during the refolding step despite the use of a stepwise dialysis step that gradually removed the denaturant, and no BLMZRQC was obtained at the end of this procedure (Appendix C).

Figure 6.2 SDS-PAGE of selected BLM constructs. A) SDS-PAGE of BLMZRQC (residues 992–1192) with the following samples: lane 1 and 6 – protein ladder; lane 2 – sample prior to IPTG induction; lane 3 – sample after IPTG induction; lane 4 – insoluble fraction; lane 5 – soluble fraction; lane 7 – BLMZRQC with an N-terminal hexahistidine tag (His₆-BLMZRQC) obtained after affinity chromatography; lane 8 to 12 – undesired protein fractions from gel filtration chromatography (GFC); lane 13 – His₆-BLMZRQC obtained after GFC; B) SDS-PAGE of BLMRQC (residues 1070–1192) with the following samples: lane 1 – protein ladder; lane 2 – sample prior to IPTG induction; lane 3 – sample after IPTG induction; lane 4 – insoluble fractions; lane 5 – soluble fractions; lane 6 to 9 – unbound protein eluted from Ni²⁺ affinity column, lane 10 to 13 – BLMRQC with an N-terminal hexahistidine tag (His₆-BLMRQC) obtained after affinity chromatography, lane 14 – His₆-BLMRQC obtained after GFC. Molecular weight of each band in the protein marker is indicated in kDa.

In an attempt to reduce the formation of inclusion bodies for BLMZRQC, fusion tags containing glutathione S-transferase (GST) or thioredoxin (Trx) were attached N-terminal to the

protein. Both fusion tags have been well documented to improve protein solubility, with the GST tag stabilizing the soluble protein fraction^{323,324}, while the Trx tag capitalized on its native oxidoreductase activity to suppress the production of inclusion bodies^{325,326}. As assessed by SDS-PAGE, the TrxBLMZRQC demonstrated good protein expression levels. However the overexpressed proteins predominantly formed inclusion bodies, indicating that the attachment of the Trx fusion partner failed in improving the solubility of BLMZRQC. The GST-tagged BLMZRQC construct was similarly evaluated by SDS-PAGE, which found the GSTBLMZRC to exhibit similar good protein expression levels but poor solubility to that of the TrxBLMZRQC (Appendix D).

Considering that the target BLM protein originates from human – a higher-order organism, it was speculated that the poor solubility of BLMZRQC is due to the absence of eukaryotic protein processing functionality within the *E. coli* expression system³²⁷. Consequently, both the BLMZRQC and GST-tagged BLMZRQC were expressed in insect cells of *S. frugiperda* (*Sf9*), an eukaryotic expression system typically utilized for gene expression via transfection with baculovirus containing the target gene^{328,329} (Table 6.1). However, both the protein expression and solubility levels of the BLMZRQC protein were determined via SDS-PAGE to be poor in *Sf9* cells, with no improvement in solubility observed even with the attachment of a GST solubility tag (Appendix 9).

In contrast to the BLMZRQC fragment, expression of the helicase domain with the ZnBD and RQC domain in BLMhZRQC demonstrated good protein expression levels and protein solubility (Table 6.1; Appendix D). This resulted in BLMhZRQC to be obtained in high yield, with 2.8 mg BLMhZRQC of 95 % purity obtained per litre of culture. As NMR spectroscopy is typically utilized to resolve proteins with length below 300 residues³³⁰, the BLMhZRQC, with a length of 550 residues is considered a large protein for structure determination via NMR spectroscopy, and is typically resolved via X-ray crystallography.

To obtain a more manageable protein for NMR structure determination, we sought to isolate the BLMZRQC protein from the highly soluble BLMhZRQC construct. The helicase domain will be employed similar to a solubility tag, with a tobacco envelope virus (TEV) protease cleavage site (ENLYFQS) was inserted between the helicase domain and ZnBD (BLMhTEVZRQC) to facilitate removal of the helicase domain. Characterization via SDS-PAGE

finds the engineered BLMhTEVZRQC to have better protein expression levels in comparison to the unmutated protein. However, the solubility of the BLMhTEVZRQC is adversely affected by the insertion of the TEV cleavage site, with no protein obtain after purification (Appendix D; Table 6.1).

A second candidate of BLM protein for the study of interaction via NMR spectroscopy is the zinc-binding domain (ZnBD) – a small molecule comprising of 78 residues. However, the protein expressed more poorly in comparison to the BLMZRQC, and was highly insoluble (Table 6.1; Appendix D). A second-generation BLMZ construct was devised, where the BLMZ was attached C-terminal to a Trx solubility tag (TrxBLMZ). From the SDS-PAGE images, protein expression of the TrxBLMZ is greatly improved from the un-tagged BLMZ construct (Appendix D). However, the solubility of TrxBLMZ remained unchanged from that of BLMZ, and attempts to purify the soluble fraction yielded a wrong protein (Appendix D, E). Poor protein expression levels and low solubility was similarly observed when BLMZ was expressed with the GST-tag in *Sf9* cells (Table 6.1; Appendix D). In contrast, the expression of the helicase domain with ZnBD demonstrated good protein expression and solubility, with a yield of 9 mg of purified protein obtained per litre of culture (Table 6.1; Appendix D). However, the size of the BLMhZ (molecular weight of 51.8 kDa) is beyond the 35 kDa limit of conventional protein NMR spectroscopy³³⁰. A TEV cleavage site insertion as described before for BLMhTEVZRQC was similarly applied in BLMhZ in an attempt to isolate the ZnBD. However, no BLMhTEVZ protein was obtained from this strategy due to poor protein solubility despite good protein expression levels (Table 6.1; Appendix D).

A third candidate of the NMR structural study of the BLM protein is the isolated RQC domain (BLMRQC). With a molecular weight of 16.5 kDa, the BLMRQC is a moderate size molecule. Moreover, the RQC domain has been isolated successfully in its soluble form in the WRN homologue¹⁶⁹, which suggests the BLMRQC to similarly high yielding, and thus suitable for protein production for NMR application.

Using SDS-PAGE, expression of BLMRQC protein was evaluated to be much poorer in comparison to BLMZRQC (Figure 6.2; Appendix D). However, the isolated RQC domain demonstrated excellent protein solubility, with above 60 % of the total expressed proteins soluble

(Figure 6.2B, lane 5; Appendix D). Purification this construct yielded up to 15 mg of purified BLMRQC per litre of culture, in sharp contrast to the other domain – ZnBD with zero yield, and a 375-fold increase from the 40 µg of purified protein obtained from the BLMZRQC construct. Moreover, the yield of BLMRQC remained largely unaffected even when cultured in minimal media (M9) for the growth of isotopically-labelled proteins, yielding 8.8 mg of ¹⁵N, ¹³C-labelled BLMRQC per litre of media, corresponding to a concentration of 0.9 mM sufficient for NMR analysis.

Table 6.1 Conserved domain make-up and solubility of BLM constructs.

Protein name	N-terminal tag	Expressed domain (X) – inserted protease cleavage site	Soluble ^a protein	Expression cell line
BLMhZR	His ₆	Helicase+ZnBD+RQC	Yes	<i>E. coli</i>
BLMhTEVZRQC	His ₆	Helicase(TEV)+ZnBD+RQC	No	<i>E. coli</i>
BLMhZ	His ₆	Helicase+ZnBD	Yes	<i>E. coli</i>
BLMhTEVZ	His ₆	Helicase(TEV)+ZnBD	No	<i>E. coli</i>
BLMZRQC	His ₆	ZnBD+RQC	No	<i>E. coli</i>
BLMZRQCsf9	His ₆	ZnBD+RQC	No	<i>Sf9</i>
GSTBLMZRQC	His ₆ -GST	ZnBD+RQC	No	<i>E. coli</i>
GSTBLMZRQCsf9	His ₆ -GST	ZnBD+RQC	No	<i>Sf9</i>
TrxBLMZRQC	His ₆ -Trx	ZnBD+RQC	No	<i>E. coli</i>
BLMh	His ₆	Helicase	No	<i>E. coli</i>
BLMZ	His ₆	ZnBD	No	<i>E. coli</i>
GSTBLMZsf9	His ₆ -GST	ZnBD	No	<i>Sf9</i>
TrxBLMZ	His ₆ -Trx	ZnBD	No	<i>E. coli</i>
BLMRQC	His ₆	RQC	Yes	<i>E. coli</i>

^a BLM protein with less than 0.5 mg of purified protein obtained per litre of culture is considered insoluble

Constructs with good solubility are framed in black.

Amongst the various BLM construct screened for good protein solubility, only the BLMhZRQC, BLMhZ and BLMRQC were determined to be highly soluble. BLM fragments similar to the BLMhZ and BLMRQC have also been produced in good yield by Gyimesi *et. al.*²⁰⁶ and Kim *et. al.*³³¹. Analysis of the combined solubility data reveals the RQC domain to be the only domain amongst the three conserved domains screened to be soluble in its isolated form; the helicase domain and the ZnDB needed to be expressed within the same protein fragment to be soluble (Table 6.1). With high resolution structures of RecQ helicase core illustrating the ZnBD to be composed of several α -helices connected by highly flexible loops^{170,204}, the close packing

Figure 6.4 Sequential connectivity pathway of BLMRQC (residues 1071–1078). A) Strip plot of CBCA(CO)NH/HNCACB, where the $^{13}\text{C}\alpha$ and $^{13}\text{C}\beta$ cross peaks from the preceding residues (i-1) are joined to the next residue (i) via correlating cross peaks as indicated by the blue and red line respectively; B) Strip plots of HBHA(CO)NH/HBANH with the connectivity for $^1\text{H}\alpha$ and $^1\text{H}\beta$ indicated by red and black line respectively; C) Strip plot of HNCO/HN(CA)CO with connectivity between the $^{13}\text{C}'$ indicated by the black line. Resonance for Thr1078 was not detected due to the low sensitivity of the experiment. Intra-residual cross peaks are indicated with (*).

To obtain the identity of each resonance on the ^{15}N -HSQC spectrum, backbone assignment of each residue was performed using the standard through-bond triple resonance assignment strategy

(described in Chapter 5). We have utilized the sequential connectivity obtained by following resonances of the $C\alpha_i$ and $C\beta_i$ pair to the $C\alpha_{i-1}$ and $C\beta_{i-1}$ pair (3D HNCACB/CBCA(CO)NH), and combining them with the sequential connectivity derived from the paired $H\alpha_i$ and $H\beta_i$ to the $H\alpha_{i-1}$ and $H\beta_{i-1}$ resonances (3D HBHANH/HBHA(CO)NH). This was also further supplemented by resonances from CO_i to CO_{i-1} (3D HNCO/HN(CA)CO). Figure 6.4A illustrates the sequential connectivity established through HNCACB and CBCA(CO)NH for residues 1071–1078, where the $^{13}C\alpha$ and $^{13}C\beta$ resonance of the preceding residue (i-1) obtained from CBCA(CO)NH were matched with their correlating peaks of the current residue (i) obtained by HNCACB. Sequential connectivity of BLMRQC using the $^1H\alpha$ and $^1H\beta$ connectivity determined by HBHANH and HBHA(CO)NH is shown in Figure 6.4B, with the $^{13}C'$ resonance from HNCO and HN(CA)CO experiments similarly applied in the sequential backbone walk in Figure 6.4C. The three sets of backbone connectivity determined from these experiments were cross-verified, which was essential to assign all the backbone resonances conclusively when large number of overlapped resonances occurred on the NMR spectra.

Figure 6.5 Side chain assignment of carbon and proton resonance of Val1180 in BLMRQC. Left) Strip plot of CC(CO)NH viewed from the 1H plane of residue (i-1) – Leu1181; Middle) structure of Val1180. Carbon and proton are coloured in blue and red respectively. α , β , $\gamma1$ and $\gamma2$ identity of each nuclei are indicated by the dotted lines. Right) Strip plot of HCCH-TOCSY viewed from 1H plane of residue (i) – Val1180.

Side chain assignments were made extensively with the 3D CC(CO)NH, HBHANH/HBHA(CO)NH, and HCCH-TOCSY experiments. As illustrated for Val1180 in Figure 6.5, the CC(CO)NH experiment utilized the H_N signal from the next amide bond for the detection of side chain carbon resonances ($^{13}C\alpha$ to $^{13}C\epsilon$) of each residue. The HCC(CO)NH

experiment facilitated the assignment of side chain proton resonances in a similar manner as CC(CO)NH, although the HCC(CO)NH experiment is $^1\text{H}\alpha$ -resonance based. The HBHANH/HBHA(CO)NH experiments were used to distinguish the protons $\text{H}\alpha$ and $\text{H}\beta$ from the rest of the protons side chain. Finally, the HCCH-TOCSY experiment was used to connect all the side-chain protons resonances of each individual residue using the assigned ^{13}C chemical shifts. Further assignments of side chain resonances of BLMRQC were assisted by through-bond and through-space experiments as listed in Chapter 2.2.1, with distance restraints extracted from the ^{15}N -edited, ^{13}C -edited NOESY experiments.

Essentially, 98% of backbone assignment ($^1\text{H}_\text{N}$, ^{15}N , $^{13}\text{C}'$, $^{13}\text{C}\alpha$, $^{13}\text{C}\beta$), and 85% side chain assignment were achieved for all 123 residues of BLMRQC, except for the ^{15}N resonances of Lys1070 and Ser1094.

6.4 The BLMRQC is a globular protein containing the signature winged-helix motif and long unstructured loops

Figure 6.6 Structure of the BLMRQC model, where the preliminary solution structure of BLMRQC is presented in ribbon view. The surface representation of BLMRQC is coloured in grey.

A model of BLMRQC was generated using NMR distance restraints and found to demonstrate good initial convergence, with the structured regions adopting secondary structure motifs of α -helices and β -strands. Figure 6.6A displays the cartoon representation of the BLMRQC model, in which the protein adopts a globular conformation comprising of four α helices and β strands. The secondary structure motifs of BLMRQC are arranged in the following order: an N-terminal β -strand is connected to an α -helix; an extended loop of 22 residues; an α -

helix to an extended loop comprising of 14 residues, before terminating with the fourth α -helix and the C-terminal β -strand (Figure 6.6). The four β -strands formed two sets of anti-parallel β -sheet in the structure of BLMRQC.

From the structure-based sequence alignment of the RQC domain in the RecQ family, the BLMRQC possesses two non-conserved structural elements (Figure 6.7A). The first structural difference arises from the extended loop between $\alpha 1$ and $\alpha 2$ ($\alpha 1$ - $\alpha 2$ loop) in BLMRQC, which is

highlighted by the superimposition of BLMRQC and WRNRQC (Figure 6.7B, insert). This extended loop is not conserved element in the WRN and also in other RecQ helicases. The second structural difference originates from the second extended loop of BLMRQC ($\alpha 2$ – $\alpha 3$ loop), in which a short conserved α -helical segment manifested in place of this loop in the RQC domain of WRN, RecQ1 and *E. coli* RecQ helicases (Figure 6.7B).

6.5 Comparison between our structure and other solution structure and X-ray crystal structure of the BLMRQC domain

Figure 6.8 Comparison of the BLMRQC structures. A) Overlay of structures of isolated RQC domains: our solution structure (blue), the solution structure published by Park *et al.* (red; PDB ID 2MH9)³³³ and the crystal structure reported by Kitano *et al.* (black; PDB ID 3WE2)³³¹, with RMSD 2.76 Å excluding loop residues; B) Alignment of our solution structure, with RQC domain extracted from larger BLM protein structures (pink: PDB ID 4O3M¹⁶⁷; grey: PDB ID 4CDG¹⁶⁸) with RMSD 2.76 Å excluding loop residues. Unique structural features of BLMRQC: $\alpha 1$ – $\alpha 2$ loop and $\alpha 1$ – $\alpha 2$ loop are illustrated in View 1 and View 2 respectively. The solution structure from 2MH9 has an additional α helix ($\alpha 1.5$), and the crystal structure from 3WE2 has an extended $\alpha 3$.

The model of the BLMRQC solution structure was first presented by our group³³⁴, although another solution structure of BLMRQC was published by Park *et al.*³³³ (PDB ID 2MH9) while the manuscript was in preparation. In the past few years, various BLM structures were reported, with the X-ray crystal structure of the free BLMRQC (PDB ID 3WE2) published by Kim *et al.*³³¹, and the protein core of BLM spanning the helicase domain to the HRDC domain (residues 640–

1298) co-crystallized with a nanobody (PDB ID 4CDG)¹⁶⁸ and with a duplex DNA with 3' overhang (PDB ID 4OM3)¹⁶⁷. Excluding the loops, the superimposition of RQC domain from the three X-ray crystal structures (PDB ID 3WE2, 4CDG, 4O3M) demonstrated high degree of overlap (RMSD 0.48 Å), indicating that the RQC domain to be relatively unchanged when isolated from the main protein body.

Our BLMRQC model was first compared with the solution and crystal structure of the isolated BLMRQC (PDB ID 2MH9 and 3WE2 respectively). These three structures were found to be highly overlapped except for the $\alpha 1$ – $\alpha 2$ loop and the $\alpha 2$ – $\alpha 3$ loop (Figure 6.8A). For the former, the solution structure from 2MH9 contains an additional α helix on the $\alpha 1$ – $\alpha 2$ loop (termed as $\alpha 1.5$), which is absent in our solution structure and the 3WE2 crystal structure. Conversely, for the latter, an extended α helix was observed on the $\alpha 3$ of the crystal structure, but such feature was not observed in either the published solution structure of BLMRQC or our model.

6.6 The BLMRQC exhibits low DNA binding affinity to G-quadruplexes and duplex DNA

With the removal of the ZnBD from the previously determined G-quadruplex binding region in BLM⁹⁰, there is a need to reassess if the strong G-quadruplex-binding ability demonstrated by BLM is conserved by the BLMRQC fragment. The binding affinity of BLMRQC to a parallel G-quadruplex and a blunt-end duplex DNA was preliminary evaluated at submillimolar concentration against the RQC fragment of WRN (WRNRQC), which bind the duplex DNA in the submicromolar range¹⁶⁹. Discrete bands corresponding to the formation of DNA-protein complexes were observed for WRNRQC but not for BLMRQC after titration with equimolar amounts of DNA (Figure 6.9A, B). This corresponds to a lack of binding interaction between the BLMRQC and either DNA substrates at submillimolar concentration, in sharp contrast to the WRNRQC, in which the G-quadruplex and duplex DNA screened are mostly bound under the same assay conditions.

To identify if the low binding affinity of BLMRQC is sequence- and/or structure-specific, the BLMRQC was screened against a wider variety of DNA structures: two other parallel G-quadruplexes, a G-quadruplex with (3+1) fold, a G-quadruplex with 3' tail previously shown to be

tightly bound by the ZnBD-RQC fragment (OX1T)⁹⁰, a single-stranded DNA, and second blunt-end duplex DNA (Table 6.2). The amount of DNA probe used was lowered to 32.5 μ M to improve sensitivity in the gel shift assay. However, characterization by gel electrophoresis observed no formation of DNA-protein complex bands even with 4-fold of BLMRQC added, indicating that the BLMRQC does not bind to all the DNA substrate screened, regardless of its DNA sequence or its secondary structure (Figure 6.9C).

Figure 6.9 Gel shift electrophoresis of BLMRQC with different DNA substrates. Comparison of DNA-binding affinity of BLMRQC with WRNRQC. The DNA and protein concentration are used at 0.1 mM (DNA/Protein ratio 1:1). DNA types screened are: A) parallel G-quadruplex *CEB25* and B) duplex DNA *DX1*; C) Gel shift images of BLMRQC screened for binding to different G-quadruplex forms, single-stranded DNA and duplex DNA. 32.5 μ M of DNA was incubated with 4-fold of BLMRQC. Lanes marked with (-) are negative controls containing only DNA. Gel was visualized by UV shadowing.

The DNA binding of the BLMRQC protein was further evaluated by NMR chemical shift perturbation assay (NMR CSPA), an assay which monitors peak perturbation of the protein via $\{^1\text{H}-^{15}\text{N}\}$ -HSQC following the titration of the DNA/ligand. As each resonance observed on the $\{^1\text{H}-^{15}\text{N}\}$ -HSQC spectrum represents each NH bond from the peptide backbone of BLMRQC, subtle shifts to the resonance is associated to interactions with the DNA substrate, and is thus

capable of detecting DNA-protein complexes with weak dissociation constant in the millimolar range³³⁵.

Figure 6.10 Comparative ^1H , ^{15}N -HSQC spectra displaying the titration of BLMRQC with a G-quadruplex and a duplex DNA. 0.2 mM of BLMRQC was titrated with 4-fold of DNA: A) duplex DNA, *DX1*; B) G-quadruplex, *HT*. The resonance that exhibited the most significant peak shift is indicated with a red arrow on the spectra. A magnified view of this resonance superimposed with 2-fold (grey) and 4-fold DNA (red) titration data, is indicated in the insert.

Figure 6.10 illustrates the $\{^1\text{H}\text{-}^{15}\text{N}\}$ -HSQC spectrum of 0.2 mM of ^{15}N -labelled BLMRQC in its DNA-free state (black), and overlaid with the $\{^1\text{H}\text{-}^{15}\text{N}\}$ -HSQC spectrum obtained after 4-fold titration of *HT*, a (3+1) G-quadruplex, and *DX1*, a self-complementary duplex DNA. A cross-peak corresponding to the Ile1168 was observed to be the most perturbed, exhibiting peak shifts of 0.22 ppm and 0.12 ppm recorded for *HT* and *DX1* respectively. The Ile1168 residue is mapped onto the β -wing motif in BLMRQC, which has been determined to be a key DNA binding region in other RecQ helicases^{169,204}. However, the peak perturbation exhibited by BLMRQC is relatively small in comparison to those observed in Chapter 5 (0.5 – 8 ppm) even at high protein concentrations (0.8 mM) of BLMRQC. This suggests the binding affinity of the BLMRQC to DNA to be extremely weak, in the submillimolar range, in agreement with the results obtained from the gel shift assays.

Structural data and mutational analysis of RQC domains in the RecQ family (WRN and RecQ1) have shown the RQC domain to be key region for DNA binding^{169,204,205}. However, the gel shift assays and NMR CSPA indicate BLMRQC to exhibit weak binding with both G-quadruplexes and duplexes when isolated from the main protein body. However, the β -wing is still proven to be involved in DNA binding interactions for BLMRQC, where the residue Ile1168 located on the β -wing exhibited small peak shifts after binding with DNA (Figure 6.10A). This is

in agreement with a recently published X-ray crystal structure of a truncated BLM protein bound to a duplex DNA with 3' overhang (PDB ID 4O3M)¹⁶⁷, which also determined the β -wing to be critical in DNA strand separation. Moreover, site-directed mutagenesis of a nearby residue (Asn1162Ala), resulted in a 10-fold reduction in helicase processivity¹⁶⁷, thus emphasising on the importance of the β -wing in DNA interactions.

Table 6.2 Sequence of oligonucleotides and their corresponding structure utilized in titration with BLM proteins.

Name	Sequence (5' – 3')	Structure
Canonical DNA forms		
<i>SS1</i>	AGCTACCATGCCTGCACGAATTAAGCAATTCGTAATC ATGGTCATAGCT	Single-stranded DNA ¹⁶⁹
<i>SS2</i>	AGCTATGACCATGATTACGAATTGCTTAATTCGTGCA GGCATGGTAGCT	Single-stranded DNA ¹⁶⁹
<i>DX1</i>	ACCCTAATTAGGGT	Self complimentary duplex ¹⁶⁹
<i>DX2</i>	<i>SS1</i> + <i>SS2</i>	Duplex ¹⁶⁹
<i>DBY1</i>	TGCTAGTTCTAGCAGGCCCTTGGGCCGGCGCTTGCGC C	Dumbbell duplex ³³⁶
G-quadruplexes		
<i>OX1T</i>	ACTGTCTGACTTGATATTTGGGGTTTGGGGAATGT GA	G-quadruplex [#]
<i>HT</i>	TTGGGTTAGGGTTAGGGTTAGGGA	3+1 ⁴⁹
<i>CEB25</i>	AAGGGTGGGTGTAAGTGTGGGTGGGT	Parallel ³⁰⁸
<i>G41</i>	TTGGGTGGGTTTGGGTGGGT	Parallel ^{&}
<i>T95-2T</i>	TTGGGTGGTGGGTGGGT	Parallel ²³⁶

[#] Structure has not been resolved

[&] Structure was verified via CD spectroscopy

6.7 The BLM zinc-binding domain is required for tight binding of BLM with DNA

Of the two conserved domains in the reported BLM G-quadruplex binding region, we have demonstrated that the C-terminal RQC domain to associate weakly to both G-quadruplexes and canonical forms of DNA. Of the remaining N-terminal domain in this core G-quadruplex binding region of the BLM protein, we were unable to obtain purified ZnBD protein to evaluate its binding affinity for G-quadruplexes and other DNA forms despite various attempts (Table 6.2). However, a comparison between the ZnBD-RQC fragment (BLMZRQC) and BLMRQC in Figure 6.11A shows the former to bind ~60 % of the G-quadruplex probe, whereas no visible depletion in the free G-quadruplex could be observed for the latter. Under the same assay conditions, lesser

complex formation was observed for duplex DNA for BLMZRQC, correlating with the specificity for G-quadruplexes of BLMZRQC as reported in literature⁹⁰. This indicate that the N-terminal ZnBD domain is the major contributor towards the tight DNA binding of the BLM protein. Unfortunately, with only 40 µg of ZnBD-RQC fragment (BLMZRQC) with suitable purity (~60 %) obtained from our protein production attempts (Figure 6.2A), this amount is only sufficient for evaluation by biochemical assays, but not for NMR titration or NMR structural determination, which typical entails protein concentrations in milligram quantities.

Figure 6.11 Comparative gel shift images between BLMZRQC and BLMRQC. 125 nM of DNA: A) *HT* G-quadruplex; B) *DBYI* duplex DNA was titrated with 40-fold of protein. Gel was visualized by fluorescent dye attached on 5' end of each DNA.

The essential nature of the ZnBD domain by the BLM protein in DNA binding interactions sets this RecQ helicase apart from WRN, the other RecQ helicase that demonstrated similar G-quadruplex specificity but is able to bind DNA substrates tightly with the isolated RQC fragment¹⁶⁹.

From the recently available X-ray structures of the BLM protein core in complex with a nanobody¹⁶⁸ or a duplex DNA¹⁶⁷, the RQC domain is connected to the ZnBD via a flexible loop, which pivots the RQC domain ~ 90° about the ZnBD when complexed with different substrate (Figure 6.12). In complex with the nanobody, the RQC domain positions ~ 28 Å away from the main body of the protein, with little surface contact between the ZnBD and RQC¹⁶⁷. However, upon interaction with the duplex DNA, the RQC domain is positioned proximal to the ZnBD, with the sugar backbone of the duplex strand establishing extensive contact with both domains. With these structural data showing the RQC domain to be highly mobile, it is postulated that ZnBD and RQC domains need to act synergistically to manifest strong DNA binding affinity for the BLM

protein. However, more research is required to confirm this hypothesis, and to determine the structural motif within the BLMZRQC fragment that contributes to its specificity towards G-quadruplexes.

Figure 6.12 Superimposed structures of the BLM core protein demonstrating different orientations of RQC domain. BLM protein in complex with a nanobody (PDB ID 4CDG)¹⁶⁸ or with a duplex DNA with 3' overhang (PDB ID 4O3M)¹⁶⁷ were aligned by their helicase domain with RMSD 1.24 Å, with each individual structure illustrated in the insert. The change in orientation of RQC domain indicated by the red arrow, with the RQC domain shaded in red for clarity. The helicase domain is coloured in white, the zinc-binding domain in blue, the RQC domain in red, the 3' overhang DNA in grey and the nanobody in green. Conserved domains except the HRDC domain are illustrated and labelled in the insert. The structure identity is indicated in boldface.

6.8 Conclusion

The solubility study of various BLM proteins containing the conserved helicase/ZnBD /RQC domains has shown the RQC domain exists as a stable and highly soluble protein domain even when isolated from the other protein regions. The high solubility of the BLMhZRQC and BLMhZ but not the BLMZRQC suggests that the ZnBD induces protein misfolding without the stabilizing effects of the helicase domain. The BLMRQC structure was determined to adopt a globular form that retained the signature winged-helix motif as other RQC proteins. Gel shift data and NMR titration have demonstrated that the RQC domain exhibits low binding affinity to both G-quadruplex and canonical forms of DNA, and required the presence of ZnBD to achieve tight binding for G-quadruplexes and duplexes.

Chapter 7 Summary and future prospective

7.1 Summary

G-quadruplexes are four-stranded structures built from stacked G-tetrads ($G\bullet G\bullet G\bullet G$), which are planar cyclical assemblies of four guanine bases interacting through Hoogsteen hydrogen bonds. A G-quadruplex containing a single guanine analogue substitution, such as 8-oxoguanine (O) or xanthine (X) would suffer from the loss of a Hoogsteen hydrogen bond within a G-tetrad and/or potential steric hindrance. We show that a proper arrangement of O and X bases can reestablish the hydrogen bond pattern within a $G\bullet G\bullet X\bullet O$ tetrad. Rational incorporation of $G\bullet G\bullet X\bullet O$ tetrads on a (3+1) G-quadruplex demonstrated similar folding topology and thermal properties to the unmodified G-quadruplex. pH titration conducted on X•O-modified G-quadruplexes monitored by NMR indicated a protonation-deprotonation equilibrium of X with a $pK_a \sim 6.7$. The solution structure of a G-quadruplex containing a $G\bullet G\bullet X\bullet O$ tetrad was determined by NMR spectroscopy, with its folding topology being conserved in both the protonated and deprotonated form. Our data showed that a G-quadruplex containing an X•O pair in the deprotonated state exhibits a more electronegative G-quadruplex groove than the unmodified G-

quadruplex. These differences are likely to manifest in the electronic properties of G-quadruplexes, and may have important implications for quadruplex-protein interactions and drug targeting.

A judicious introduction of X and O in a human telomeric G-quadruplex forming sequence was shown to reverse the hydrogen bond polarity of the modified G-tetrad while preserving its original folding topology. The control exerted over G-tetrad polarity by the X•O modification will be valuable for the design and programming of G-quadruplex structures and properties.

G-quadruplexes have been associated with important cellular processes, which require G-quadruplex-protein interaction. However, the structural basis for specific G-quadruplex recognition by proteins has not been understood. The DEAH-box helicase Rhau (also named DHX36 or G4R1) specifically binds to and resolves parallel-stranded G-quadruplexes. Here we identified an 18-amino-acid G-quadruplex-binding domain of Rhau and determined the structure of this peptide bound to a parallel DNA G-quadruplex. Our structure explains how Rhau specifically recognizes parallel G-quadruplexes. The peptide covers a terminal guanine base tetrad (G-tetrad), and clamps the G-quadruplex using 3-anchor-point electrostatic interactions between three positively charged amino acids and phosphate groups. This binding mode is strikingly similar to that of most ligands selected for specific G-quadruplex targeting. Binding to an exposed G-tetrad represents a simple and efficient way to specifically target G-quadruplex structures.

The ZnBD and ZRQC domain of BLM has also been shown to be highly insoluble and resistant to solubilising attempts through protein engineering, while the isolated RQC domain is highly solubilized. BLMRQC was characterized by NMR spectroscopy, and a structure of BLMRQC generated through the application of NMR restraints was found to retain the winged-helix motif typical of RQC protein, with slight structural alterations at the α -helices proximal to the β -wing. We have demonstrated the truncated BLMRQC protein to exhibit weak binding affinity of in the submillimolar range, which contrasted sharply with the reported nanomolar DNA affinity of the combined ZnBD and RQC protein fragment. Comparison of recently published

BLM X-ray crystal structure has identified the RQC domain to be connected to the ZnBD via a highly flexible loop, indicating the RQC domain to be highly mobile. With crystal structure of the BLM-dsDNA complex showing the ZnBD and RQC domain to exhibit the most interaction with the duplex DNA, the ZnBD and RQC is suggested to act cooperatively to exert strong DNA binding interactions.

7.2 Future prospective

Xanthine, 8-oxoguanine and G-quadruplexes

In the study of a G•G•X•O tetrad, although majority of the double mutants exhibited similar or improved thermal stability in comparison to the wild-type G-quadruplex, two of the mutants, *HT-9x3o* and *HT3x21o* was found to be slightly destabilized by 2 – 4 °C. However, the imino proton spectrum of *HT-9x3o* (and similarly extrapolated to be the same for *HT-3x21o*) demonstrated the unique 13 imino peak pattern characteristic of the G•G•X•O tetrad. This indicates *HT-9x3o* to have assembled as the same type of X•O modified G-quadruplex as *HT-17x9o*. To achieve a better understanding regarding the structural features that resulted in the destabilization of *HT-9x3o*, the solution structure of *HT-9x3o* can be elucidated and compared against *HT-17x9o* and the wild-type *HT* G-quadruplex. Similar assignment strategy using ¹⁵N-filtered experiments and 2D NOESY experiments as described in Chapter 3 can be applied for the structure determination of *HT-9x3o*. Molecular dynamics simulations can also be performed for *HT-9x3o* and *HT-17x9o* to model the unfolding/folding pathways of both mutants for determining factors that affects the thermal properties in the X•O-modified G-quadruplex.

In cells, DNA base lesions O and X are both genotoxic, with the former being the source of O:A mutagen³³⁷, and the latter forming X:T mutations³³⁸. To preserve genomic integrity, there exist DNA response pathway in cells to repair the damage sustained by the DNA³³⁹. The DNA response mechanism is well documented for duplex DNA, with the point mutation of O or X each identified and repaired by base excision repair proteins (hOGG1 for O³⁴⁰, Mpg for X¹¹⁹). Although the findings in Chapter 3 and several literature have identified existing G-quadruplex structures to be strongly destabilized by the formation of such DNA base lesion^{111,112,115}, a proportion of the damaged G-quadruplex still persist under physiological salt conditions as seen for point mutations of O in the external G-tetrads on the human telomeric *HTI* G-quadruplex (T_m range: 43 – 45 °C)¹¹²; tetramolecular TG₄T quadruplex (T_m range: 55 – 65 °C)¹¹⁵; VEGF Pu22 G-quadruplex (T_m range: 57 – 59 °C)¹¹³. This necessitates the recognition and resolving of the affected G-quadruplex prior to the repair of the damaged nucleobase. On duplex DNA, the processivity of hOGG1 and Mpg DNA glycosylases have been evaluated extensively^{119,262,340-342}.

In contrast, the action of DNA glycosylases remains to be investigated for G→X mutation on G-quadruplexes, although recent *in vitro* data on hOGG1 have identified the inability of this DNA glycosylase to repair G→O within a G-quadruplex scaffold^{113,343}.

In order to identify the enzymes which recognize and repair O and X mutations on a G-quadruplex, glycosylase assay can be performed on different libraries of proteins. Using an X/O-modified G-quadruplex labelled with a fluorophore and a dark quencher on the termini, the G-quadruplex will not emit fluorescence when folded. However, glycosylase activity on the G-quadruplex will result in the DNA strand to be cleaved at the mutation site; fluorescence will be emitted from the fluorophore due to the loss of dark quencher from the DNA strand. Such assay can be implemented in microwell array system for high-throughput screening of the protein libraries to facilitate rapid isolation of enzymes which repairs DNA damage in G-quadruplexes.

In addition, little is known about the response mechanism when DNA damage is sustained by G-quadruplexes in cells despite the prevalence of DNA damage occurring on G-rich sequences¹⁰⁸. Moreover, direct evidence of the formation of O within the G-quadruplex scaffold in cells has yet to be shown. Through the use of monoclonal antibodies that are G-quadruplex-specific, Balasubramanian's group and Lansdorp group have independently visualized G-quadruplexes in human cells via immunofluorescent staining.^{8,344} There also exists a commercially-available recombinant monoclonal antibody Fab 166 which facilitates the immunofluorescent imaging of O in DNA *in vivo*³⁴⁵. To study the effect of oxidative damage on G-quadruplexes in cells, it is proposed that double immunofluorescence staining can be performed using the two antibodies each specific for G-quadruplexes and O in cells treatment with hydrogen peroxide to induce oxidative damage. Due to the lack of commercially available anti-X antibody, antibodies that bind with high affinity to X in DNA need to be identified, and subsequently applied in the double immunofluorescence staining approach for cells treated with nitrous oxide to investigate the effect of nitrosative damage on G-quadruplexes in cells.

Although Rhau is not associated with any latent diseases to date, the discovery of the recruitment of Rhau helicases to stress granules as a cellular response to environmental stress¹⁵³ signifies the Rhau helicase to have critical roles in the processing of mRNA, a critical cellular function. Moreover, Rhau are highly expressed in the testis¹⁵⁷, and recently found to participate in spermatogonia differentiation via binding of G-quadruplex motifs on c-kit promoter¹⁵⁸. The G-quadruplex structures appear to act as a molecular switch which Rhau specifically targets, and consequently up-regulates c-kit expression. With the discovery of parallel G-quadruplex specificity exhibited by the Rhau protein as described in Chapter 4, further understanding of the biological functions of Rhau could be made by probing the interactions between parallel G-quadruplex structures made of RNA or derived from the c-kit promoter.

Using a combination of high-throughput gene array assays and computational analysis, interaction between G-quadruplex-forming RNA sequences and the Rhau protein was established, with the human telomerase template RNA (TERC) determined as the primary RNA target of Rhau¹⁸³. A TERC G-quadruplex, *TERC18*, which forms a dimeric parallel RNA G-quadruplex was screened against Rhau₁₈, with full complex formation observed with 2-fold excess of peptide titrated with the RNA G-quadruplex. Since the binding of the Rhau to TERC in living cells is dependent on the formation of the RNA G-quadruplex¹⁸³, it will be advantageous to determine the G-quadruplex recognition motif of *TERC18* with Rhau₁₈ using the same methodology employed for structure determination of the DNA G-quadruplex-Rhau₁₈ complex. Results from the elucidation of a TERC G-quadruplex-Rhau₁₈ complex may lend insight into the functionality of the Rhau helicase in telomeric length regulations in cells.

Comparison of the binding affinity of Rhau₁₈ against Rhau₁₋₂₀₀ and full-length Rhau finds the strength of G-quadruplex binding to increase with the length of the Rhau protein. This suggests that aside from Rhau₅₃ that acts as a key G-quadruplex binding motif, there exist others regions within the wild-type Rhau that interacts strongly with G-quadruplexes. To identify the secondary G-quadruplex-binding regions, it is suggested that longer fragments of Rhau, such as Rhau₁₄₂ (residues 53–195) could be studied for its binding with G-quadruplexes by NMR spectroscopy.

Structure determination of the full length Rhau in complex parallel G-quadruplex, *T95-2T* could be attempted for further understanding of the binding interactions between the Rhau helicase and G-quadruplex. With a length of 1008 residues, elucidation of the full-length Rhau protein via NMR spectroscopy will be very challenging. Spectral analysis of longer proteins will be afflicted by spectral complexity such as line broadening, increased number of NMR-active nuclei and signal degeneracy, which have to be addressed through the application of specialized NMR acquisition techniques (*i.e.* traverse relaxation-optimized spectroscopy, TROSY³⁴⁶), and the use of NMR spectrometers with a higher magnetic field (*i.e.* 700 MHz and above). The large size of Rhau also mandates the triple-labelling of the protein with ¹⁵N, ¹³C and ²H isotopes^{347,348} for improved spectral resolution. The production of Rhau helicase has to be scaled-up as typical NMR sample requires protein concentrations of 0.5 – 1.0 mM³⁴⁹ for NMR analysis. These combined factors greatly increase the time effort and cost involved³³⁰ in deconvoluting the solution structure of Rhau. As such, the alternative structure determination approach – X-ray crystallography will be considered for the structure determination of the full-length Rhau-*T95-2T* complex.

Of the 33 available structures from RecQ and RecQ-like helicases, five DNA-protein complexes has been reported³⁷, with these structures involving complexes formed by a RecQ protein bound either to a single-stranded DNA or a duplex DNA^{167-169,204,350}. However, the structure of a G-quadruplex-RecQ protein complex still remains elusive, and more research is required to facilitate our understanding on the binding interactions between G-quadruplexes and the RecQ helicases.

In Chapter 6, we have evaluated the DNA binding affinity of the RQC domain in BLM, one of the two G-quadruplex-interacting RecQ helicase. The RQC domain was established to poorly interact with DNA in comparison to the previously defined G-quadruplex binding region, BLMZRQC, which spans from the ZnBD to the RQC domain. Consequently, more emphasis will be placed on studying the BLMZRQC for factors causing its improved DNA binding affinity. As the solubility data of various wild-type and engineered BLM fragment suggest that the helicase domain and ZnBD needs to be expressed together, the effect of ZnBD and RQC on G-quadruplex binding could be studied via DNA binding assays on the two soluble BLM fragments: the BLMhZRQC (BLM₆₄₂₋₁₁₉₂) and BLMhZ (BLM₆₄₂₋₁₀₆₉).

The structure of both BLM proteins in complex with G-quadruplexes could be studied to understand the different role of ZnBD and RQC in binding interaction with G-quadruplex. However, the size of the BLMhZ and BLMhZRQC fragment at 450 – 550 residues increases the challenges involved in the elucidation of their solution structure. Thus, the alternative structure determination method: X-ray crystallography, will be attempted concurrently for.

Recently, Chatterjee *et. al.* has reported the production of a BLM fragment encoding from the C-terminal helicase 2A domain to the HRDC domain that can actively unfold G-quadruplexes despite having only half the helicase domain³⁵¹. Thus, it will be advantageous to produce a slightly shorter BLM protein in comparison to the BLMhZ/BLMhZRQC. The shorter construct will reduce the length of the protein by 216 residues, considerably reducing the effort involved in the structure determination. Based on the BLMhZ/BLMhZRQC fragments, BLM proteins containing only the helicase 2A domain were expressed (BLM0.5hZRQC/BLM0.5hZ) and their solubility screened (Appendix D).

Interestingly, the removal of the helicase 1A domain reduces the solubilizing effect afforded by the expression of helicase domain with the ZnBD domain, resulting in BLM0.5hZ to be highly insoluble, with only the BLM0.5hZRQC shown to be soluble (Appendix D).

Appendix

A. Vector maps of cloning vectors

Figure A.1 Vector map of pNIC-28-BsaI³⁵².

Figure A.2 Vector map of pET32-Ek/LIC

Figure A.3 Coding region of pNIC-GST vector. The full vector map of pNIC-GST is unavailable.

Figure A.4 Vector map of pFB-LIC³⁵².

B. List of recombinant plasmids, their associated components

Table B.1 List of primers used in cloning

Construct name	Primer name (F = forward; R = reverse)	Sequence (5' - 3') (<u>XXX</u> = codons of insert)	Vector
pR18c2T	R18c2TF	GAC GAC GAC AAG ATG GAG AAC CTG TAC TTC CAA	pET32 Ek/LIC
	R18c2TR	GAG GAG AAG CCC GGT TA <u>CTG</u> <u>TTT TTT CGC GTA CCA</u>	pET32 Ek/LIC
pBLMzrLIC	blmzrLICF	GAC GAC GAC AAG ATG <u>TTC</u> <u>TATA ACC TATA CAT GAT</u>	pET32 Ek/LIC
	blmzrLICR	GAG GAG AAG CCC GGT TAT <u>GTT</u> <u>TCC ATA AAG TCT AC</u>	pET32 Ek/LIC
pBLMzLIC	hblmzF	GAC GAC GAC AAG ATG GAG AAC CTG TAC TTC CAA TCC ATG <u>TTC TAT ACC TAT</u>	pET32 Ek/LIC
	hblmzR	GAG GAG AAG CCC GGT TAT <u>GTT</u> <u>TTA CAG CAA TTA TCA CAA GAA</u> <u>ACA TCT</u>	pET32 Ek/LIC
pGSTekBLMzr	gstF1	GAC GAC GAC AAG ATG AGC GAT AAA ATT ATT GAG TCC CC	pNIC-GST
	gstekR1	<u>CAT GGA</u> CTT GTC GTC GTC ATC TTT TGG AGG ATG	pNIC-GST
	zekF2	CT CCT CCA AAA GAT GAC GAC GAC AAG TCC ATG <u>TTC TAT ACC</u> <u>TAT CAT GAT</u>	pET32 Ek/LIC

C. Two-step denaturing and refolding procedure for BLMZRQC

BLMZRQC is expressed according to protocols as described in Chapter 2.1.2.2. The cells from 1L of culture were harvested, washed in 150mL washing buffer I (20 mM Tris, pH 8.0), and spin down at 30,000 x g, 4 °C for 30 min. The cells were resuspended with 80mL of lysis buffer (50 mM Tris, pH 8.0), lysed via sonication, and clarified at 30,000 x g, 4 °C for 45 min. The pellet was cleaned by resuspension in 150 mL washing buffer II (50 mM Tris, 50 mM NaCl, 2% Triton X-100, 1.5 mM β -mercaptoethanol, 1.6 M urea, pH 8.0) and mixed for 20 min 4 °C, before it was centrifuged at 30,000 x g, 4 °C for 30 min. Triton X-100 contents was removed from the sample by washing twice with 150 ml of washing buffer I, and centrifuged again at 30,000 x g, 4 °C, 30 min. The cleaned inclusion bodies were dissolved in 5 ml of extraction buffer I (50 mM Tris, 50 mM NaCl, 10 mM β -mercaptoethanol, 7 M Guanidine HCl, pH 8.0), then the solution was centrifuged (30,000 x g, 4 °C, 30 min). The supernatant was diluted with 200 mL of dilution buffer (50 mM Tris, 1 mM EDTA, 50 mM NaCl, 10 mM β -mercaptoethanol, pH 8.0) to rapidly precipitate the denatured protein, followed by centrifugation at 30,000 x g, 4 °C for 45 min. The pellets were collected, and dissolved in 5 mL of extraction buffer II (50 mM Tris-HCl, 50 mM NaCl, 10 mM β -mercaptoethanol, 8 M urea, pH 8.0). The concentration of the solubilized protein was adjusted to 2 mg/mL prior to stepwise dialysis at 4 °C in 40 mM Tris, 150 mM NaCl, 1 mM ZnCl_2 , 10 mM β -mercaptoethanol, 5 M urea, 10 % (v/v) glycerol, pH 7.3, with the urea concentration lowered to 3 M, 1 M, 0 M, at every 12 h intervals to gradually refold the protein. The refolded protein was subsequently purified via Ni-NTA chromatography.

Figure C.1 SDS-PAGE of BLMZRQC with residues 992 – 1192 under denaturing conditions. The following samples are showed in this gel: Lane 1 – BLMZRQC solubilized by 2-step denaturing procedure; 2 – PageRuler Unstained Protein ladder (Fermentas); lane 3 to 9 – fractions obtained from affinity chromatography of refolded BLMZRQC. Molecular weight of each band in the protein marker is indicated in kDa.

D. SDS-PAGE of BLM proteins

Figure D.1 SDS-PAGE of Trx-tagged BLM proteins. A) TrxBLMZRQC with residues 992–1192 and an N-terminal thioredoxin (Trx) tag. The following samples are showed in this gel: Lane 1 – PageRuler Unstained Protein ladder (Fermentas); lane 2 – sample prior to IPTG induction; lane 3 – sample after IPTG induction; lane 4 – insoluble fractions; lane 5 – soluble fractions; lane 6 – undesired proteins (without hexahistidine tag) removed by affinity chromatography; lane 7 – sample after purification by affinity chromatography; B) TrxBLMZ encoding residues 992–1069 attached with an N-terminal Trx tag. In lane 1 – uninduced sample; lane 2 – induced sample; lane 3 – insoluble fraction; lane 4 – soluble fraction; lane 5 to 7 – proteins removed by affinity chromatography; lane 8 and 9 – purified protein after affinity chromatography; lane 10 and 11 – protein marker; lane 12 – purified BLMZ sample after TEV digestion and ion exchange chromatography, which ran higher than the expected molecular weight of 9613 Da. Position of desired bands is indicated by red arrows. Sample in lane 13 was verified by mass spectrometry to be Trx tag instead. Molecular weight of each band in the protein marker is indicated in kDa.

Figure D.2 SDS-PAGE of GST-tagged BLM fragments expressed in *E. coli*. Lanes corresponding of GSTBLMZRC (residues 992 – 1192 and an N-terminal glutathione S-transferase tag) is framed in red, with the desired protein band indicated by a red arrow. The soluble expression refers to the protein obtained after FPLC purification under native conditions. The protein marker is denoted by an M, with the molecular weight of each band in the protein marker is indicated in kDa. The gels were obtained from the NTU/SBS Protein Production Platform.

Figure D.3 SDS-PAGE of helicase domain-containing BLM fragments. (Top) BLMhZ construct which encodes for the helicase domain and zinc-binding domain; (middle) BLMhZRQC which is composed of the helicase domain, zinc-binding domain and RQC domain. Lanes corresponding to respective proteins are framed in red, with the desired protein band indicated by a red arrow. The soluble expression refers to the protein obtained after FPLC purification under native conditions. The protein marker is denoted by an M, with the molecular weight of each band in the protein marker is indicated in kDa. The gels were obtained from the NTU/SBS Protein Production Platform.

Figure D.4 SDS-PAGE of isolated domains of BLM expressed in *E. coli*. BLM fragments comprising of the helicase domain – BLMh (top), the zinc-binding domain – BLMZ (middle), and the RQC domain – BLMRQC (bottom) were expressed in *E. coli* and analysed via SDS-PAGE. Lanes corresponding to respective proteins are framed in red, with the desired protein band indicated by a red arrow. The soluble expression refers to the protein obtained after FPLC purification under native conditions. The protein marker is denoted by an M, with the molecular weight of each band in the protein marker is indicated in kDa. The gels were obtained from the NTU/SBS Protein Production Platform.

Figure D.5 SDS-PAGE of BLM fragments expressed in *Sf9* cells. BLM fragments of BLMZRQC (top), GST-tagged BLMZRQC (middle) and GST-tagged BLMZ (bottom) were expressed in *Sf9* cells and analysed by SDS-PAGE. The lanes corresponding to each respective protein are framed in red, with the desired protein band indicated by a red arrow. The soluble expression refers to the protein obtained after FPLC purification under native conditions. The protein marker is denoted by an M, with the molecular weight of each band in the protein marker is indicated in kDa. The gels were obtained from the NTU/SBS Protein Production Platform.

Figure D.6 SDS-PAGE of helicase2A domain-containing BLM fragments. (Top) BLM0.5hZ construct which encodes for the helicase2A domain and zinc-binding domain; (middle) BLM0.5hZRQC which is composed of the helicase2A domain, zinc-binding domain and RQC domain. Lanes corresponding to respective proteins are framed in red, with the desired protein band indicated by a red arrow. The soluble expression refers to the protein obtained after FPLC purification under native conditions. The protein marker is denoted by an M, with the molecular weight of each band in the protein marker is indicated in kDa. The gels were obtained from the NTU/SBS Protein Production Platform.

E. Mass spectrometry analysis of purified protein

Figure E.1 MALDI-TOF spectra of Rhau proteins. A) ^{15}N -Rhau₁₈, with expected MW of 2387; B) $^{15}\text{N}, ^{13}\text{C}$ -Rhau₅₃, with expected MW of 6850; C) Rhau₁₄₂, with expected MW of 17689.

Figure E.2 MALDI-TOF spectrum of BLM proteins. A) $^{15}\text{N}, ^{13}\text{C}$ -BLMRQC, with expected MW of 14850; B) BLMhZ with expected MW of 51806; C) BLMhZRQC, with expected MW of 65632.

Figure E.3 MALDI-TOF spectrum of WRNRQC, with expected MW of 17654.

Analysis Information

Report Type	Protein-Peptide Summary by Spot	Analysis Type	Combined (MS+MS/MS)
Sample Set Name	130814	Database	SwissProt2
Analysis Name	Lee_Vee	Creation Date	08/22/2013 13:13:49
Reported By	08/22/2013 13:39:24 - sbs	Last Modified	08/22/2013 13:34:44
MS Acq. : Proc. Methods	MengWei\MS Reflector Positive : MengWei\MS Processing Default		
Interpretation Method	MengWei\Interpretation Method-1KV -no keratin		

Gel Idx/Pos Plate [#] Name	17/A17 [1] mw1	Instr./Gel Origin Instrument Sample Name	AK072/130814	Process Status Spectra	Analysis Succeeded 11						
Protein Name	Rank	Species	Accession No.	Protein MW	Protein Score	Protein Score C. I. %	Total Ion Score	Total Ion C. I. %	Confirmed		
Thioredoxin-1 OS=Escherichia coli O6:H1 (strain CFT073 / ATCC 700928 / UPEC) GN=trxA PE=3 SV=2	1		sp P0AA26 THIO_ECOL6	11913.2	81	99.593	51	97.832	.T.		
Protein Group											
Thioredoxin-1 OS=Escherichia coli (strain K12) GN=trxA PE=1 SV=2			sp P0AA25 THIO_ECOLI	11913.2							
Thioredoxin-1 OS=Escherichia coli O157:H7 GN=trxA PE=1 SV=2			sp P0AA27 THIO_ECO57	11913.2							
Thioredoxin-1 OS=Salmonella typhi GN=trxA PE=3 SV=2			sp P0AA29 THIO_SALTI	11913.2							
Thioredoxin-1 OS=Salmonella typhimurium (strain LT2 / SGSC1412 / ATCC 700720) GN=trxA PE=3 SV=2			sp P0AA28 THIO_SALTY	11913.2							
Thioredoxin-1 OS=Shigella flexneri GN=trxA PE=3 SV=2			sp P0AA30 THIO_SHIFL	11913.2							
Peptide Information											
Calc. Mass	Obsrv. Mass	± da	± ppm	Start Seq.	End Seq.	Sequence	Ion Score	C. I. %	Modification	Rank	Result Type
1001.6393	1001.5496	-0.0897	-90	75	83	GIPTLLLFK	36	18.895			Mascot
1001.6393	1001.5496	-0.0897	-90	75	83	GIPTLLLFK					Mascot
1267.6641	1267.5475	-0.1166	-92	59	70	LNIDQNPQTAPK	16	0			Mascot
1267.6641	1267.5475	-0.1166	-92	59	70	LNIDQNPQTAPK					Mascot
1772.0316	1771.8201	-0.2115	-119	75	91	GIPTLLLFKNGEVAATK					Mascot
1779.9963	1779.7632	-0.2331	-131	54	70	LTVAKLNIDQNPQTAPK					Mascot
2094.9412	2094.8188	-0.1224	-58	20	37	ADGAILVDFWAEWCGPC K			Carbamidomethyl (C)[14,17]		Mascot
2126.9309	2126.8699	-0.061	-29	20	37	ADGAILVDFWAEWCGPC K			Carbamidomethyl (C)[14,17], Oxidation (HW)[10,13]		Mascot
Protein mab-21-like OS=Drosophila pseudoobscura pseudoobscura GN=GA18415 PE=3 SV=1	2		sp Q29H55 MB21L_DROPS	43107.8	70	93.977					.T.
Peptide Information											
Calc. Mass	Obsrv. Mass	± da	± ppm	Start Seq.	End Seq.	Sequence	Ion Score	C. I. %	Modification	Rank	Result Type
842.5319	842.4324	-0.0995	-118	166	171	LRIRER					Mascot
1001.5261	1001.5496	0.0235	23	340	349	SPSALEQAAK					Mascot
1001.5261	1001.5496	0.0235	23	340	349	SPSALEQAAK					Mascot
1029.4856	1029.4062	-0.0794	-77	149	156	SMYRDMVK					Mascot
1029.4856	1029.4062	-0.0794	-77	149	156	SMYRDMVK					Mascot
1267.5988	1267.5475	-0.0513	-40	357	367	EMLTNANAFEK					Mascot
1267.5988	1267.5475	-0.0513	-40	357	367	EMLTNANAFEK					Mascot
1277.718	1277.59	-0.128	-100	258	267	RCLSMKLTLR			Carbamidomethyl (C)[2]		Mascot
1574.6064	1574.688	0.0816	52	297	308	DYEWDESCIADR			Carbamidomethyl (C)[8], Oxidation (HW)[4]		Mascot
1634.7236	1634.74	0.0164	10	14	25	MIYQMNRVCAER			Carbamidomethyl (C)[9]		Mascot
1670.7706	1670.6848	-0.0858	-51	325	337	CPHYFLPSLDMFK			Carbamidomethyl (C)[1], Oxidation (HW)[3]		Mascot
1771.9449	1771.8201	-0.1248	-70	338	353	GKSPSALEQAQKQVWR			Oxidation (HW)[15]		Mascot
1792.9486	1792.6805	-0.2681	-150	134	148	IRSRFQTLVAQACDK			Carbamidomethyl (C)[13]		Mascot
1809.0116	1808.7415	-0.2701	-149	43	57	IVQDILKEVELQEPR					Mascot
1809.0116	1808.7415	-0.2701	-149	43	57	IVQDILKEVELQEPR					Mascot
1837.9993	1837.7849	-0.2144	-117	309	323	INGIFLQLISCLQYR			Carbamidomethyl (C)[11]		Mascot
1980.8141	1980.7896	-0.0245	-12	294	308	HPRDYEWDESCIADR			Carbamidomethyl (C)[11], Oxidation (HW)[7]		Mascot
2034.0001	2033.8746	-0.1255	-62	115	132	SMSLWVEFITASGYLSSR					Mascot
2061.0005	2060.8579	-0.1426	-69	136	152	SRFQTLVAQACDKSMYR			Carbamidomethyl (C)[11]		Mascot
2334.1912	2334.0007	-0.1905	-82	113	132	KRSMSLWVEFITASGYLS SR			Oxidation (HW)[7]		Mascot
2334.1912	2334.0007	-0.1905	-82	114	133	RSMSLWVEFITASGYLSS RK			Oxidation (HW)[6]		Mascot

Figure E.4 Protein analysis via mass spectrometry of sample obtained from TrxBMZ after TEV cleavage and ion exchange chromatography. The purified sample is identified to be the Trx tag instead of the expected BLMZ protein.

F. Statistics of computed structures

Table of statistics for *HT-17x9o*

Table F.1 Statistics of the computed structures of 8-oxoguanine, xanthine-double modified human telomeric sequence, *HT-17x9o*

A. NMR restraints		
distance restraints	D ₂ O	H ₂ O
intraresidue	279	1
sequential (i, i+1)	153	11
long-range (i, $\geq i+2$)	40	43
other restraints		
hydrogen bond	64	
dihedral angle	12	
B. Structure statistics		
NOE violations		
number (>0.2 Å)	0.400 $\pm$ 0.490	
maximum violation (Å)	0.174 $\pm$ 0.018	
RMSD of violations (Å)	0.023 $\pm$ 0.002	
deviations from the ideal covalent geometry		
bond lengths (Å)	0.004 $\pm$ 0.000	
bond angles (deg)	0.773 $\pm$ 0.018	
impropers (deg)	0.418 $\pm$ 0.002	
pairwise all heavy atom rmsd values (Å)		
all heavy atoms except G6, G10, G14	0.850 $\pm$ 0.140	
all heavy atoms	1.660 $\pm$ 0.320	

Table F.2 Statistics of the computed structures of free Rhau₁₈

A. NMR restraints		
Distance restraints		
Peptide restraints		
Intra-residue		280
Inter-residue		29
Other Restraints		
Hydrogen-bond restraints		
Torsion angle restraints		
Phi		12
Psi		12
B. Structure statistics		
NOE violations		
Numbers (>0.2 Å)		0.20 ± 0.63
Deviations from standard geometry		
Bond length (Å)		0.005 ± 0.000
Bond angle (°)		0.810 ± 0.017
Impropers (°)		0.933 ± 0.071
Pairwise all heavy atom r.m.s.d. values (Å)		
Backbone heavy atoms (except S1, M2, K19, Q20)		0.621 ± 0.106
All heavy atoms (except S1, M2, K19, Q20)		1.110 ± 0.159
Ramachandran plot analysis (PROCHECK)		
Most favorable region (%)		85.7
Additional allowed region (%)		14.3
Disallowed region (%)		0.0

Table F.3 Statistics of the computed structure of the *T95-2T* – Rhau₁₈ complex

A. NMR restraints	
Distance restraints	
Peptide restraints	
Intra-residue	122
Inter-residue	15
DNA-Peptide restraints	
Intermolecular	44
Other Restraints	
Hydrogen-bond restraints	
Torsion angle restraints	
Phi	10
Psi	10
DNA restraints	287
DNA-peptide restraints	44
B. Structure statistics	
NOE violations	
Numbers (>0.2 Å)	0.10 ± 0.43
Deviations from standard geometry	
Bond length (Å)	0.005 ± 0.000
Bond angle (°)	0.750 ± 0.016
Impropers (°)	0.574 ± 0.082
Pairwise all heavy atom r.m.s.d. values (Å) for the peptide	
Backbone heavy atoms (except S1, M2, K19, Q20)	0.521 ± 0.124
All heavy atoms (except S1, M2, K19, Q20)	1.650 ± 0.217
Ramachandran plot analysis (PROCHECK)	
Most favorable region (%)	65.4
Additional allowed region (%)	33.8
Disallowed region (%)	0.8

Bibliography

1. Sen, D.; Gilbert, W. Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis, *Nature* **1988**, *334*, 364-366.
2. Bang, I. Untersuchungen über die Guanylsäure., *Biochem. Z.* **1910**, *26*, 293-311.
3. Ralph, R. K.; Khorana, H. G.; Connors, W. J. Secondary Structure and Aggregation in Deoxyguanosine Oligonucleotides, *J Am Chem Soc* **1962**, *84*, 2265-&.
4. Gellert, M.; Lipsett, M. N.; Davies, D. R. Helix formation by guanylic acid, *Proc. Natl. Acad. Sci. U. S. A.* **1962**, *48*, 2013-2018.
5. Guschlbauer, W.; Chantot, J. F.; Thiele, D. Four-stranded nucleic acid structures 25 years later: from guanosine gels to telomer DNA, *J. Biomol. Struct. Dyn.* **1990**, *8*, 491-511.
6. Huppert, J. L.; Balasubramanian, S. Prevalence of quadruplexes in the human genome, *Nucleic Acids Res.* **2005**, *33*, 2908-2916.
7. Todd, A. K.; Johnston, M.; Neidle, S. Highly prevalent putative quadruplex sequence motifs in human DNA, *Nucleic Acids Res.* **2005**, *33*, 2901-2907.
8. Biffi, G.; Tannahill, D.; McCafferty, J.; Balasubramanian, S. Quantitative visualization of DNA G-quadruplex structures in human cells, *Nat Chem* **2013**, *5*, 182-186.
9. Ribeyre, C.; Lopes, J.; Boule, J. B.; Piazza, A.; Guedin, A.; Zakian, V. A.; Mergny, J. L.; Nicolas, A. The yeast Pif1 helicase prevents genomic instability caused by G-quadruplex-forming CEB1 sequences in vivo, *PLoS Genet* **2009**, *5*, e1000475.
10. Lopes, J.; Piazza, A.; Bermejo, R.; Kriegsman, B.; Colosio, A.; Teulade-Fichou, M. P.; Foiani, M.; Nicolas, A. G-quadruplex-induced instability during leading-strand replication, *EMBO J.* **2011**, *30*, 4033-4046.
11. Paeschke, K.; Bochman, M. L.; Garcia, P. D.; Cejka, P.; Friedman, K. L.; Kowalczykowski, S. C.; Zakian, V. A. Pif1 family helicases suppress genome instability at G-quadruplex motifs, *Nature* **2013**, *497*, 458-462.
12. Wu, Y.; Shin-ya, K.; Brosh, R. M., Jr. FANCD1 helicase defective in Fanconi anemia and breast cancer unwinds G-quadruplex DNA to defend genomic stability, *Mol Cell Biol* **2008**, *28*, 4116-4128.
13. Bharti, S. K.; Sommers, J. A.; George, F.; Kuper, J.; Hamon, F.; Shin-ya, K.; Teulade-Fichou, M. P.; Kisker, C.; Brosh, R. M., Jr. Specialization among iron-sulfur cluster helicases to resolve G-quadruplex DNA structures that threaten genomic stability, *J Biol Chem* **2013**, *288*, 28217-28229.
14. Katapadi, V. K.; Nambiar, M.; Raghavan, S. C. Potential G-quadruplex formation at breakpoint regions of chromosomal translocations in cancer may explain their fragility, *Genomics* **2012**, *100*, 72-80.
15. Nambiar, M.; Srivastava, M.; Gopalakrishnan, V.; Sankaran, S. K.; Raghavan, S. C. G-quadruplex structures formed at the HOX11 breakpoint region contribute to its fragility during t(10;14) translocation in T-cell leukemia, *Mol Cell Biol* **2013**, *33*, 4266-4281.
16. Koole, W.; van Schendel, R.; Karambelas, A. E.; van Heteren, J. T.; Okihara, K. L.; Tijsterman, M. A Polymerase Theta-dependent repair pathway suppresses extensive genomic instability at endogenous G4 DNA sites, *Nat Commun* **2014**, *5*, 3216.
17. Yadav, P.; Harcy, V.; Argueso, J. L.; Dominska, M.; Jinks-Robertson, S.; Kim, N. Topoisomerase I plays a critical role in suppressing genome instability at a highly transcribed g-quadruplex-forming sequence, *PLoS Genet* **2014**, *10*, e1004839.
18. Johnson, J. E.; Cao, K.; Ryvkin, P.; Wang, L. S.; Johnson, F. B. Altered gene expression in the Werner and Bloom syndromes is associated with sequences having G-quadruplex forming potential, *Nucleic Acids Res.* **2010**, *38*, 1114-1122.
19. Tornaletti, S.; Park-Snyder, S.; Hanawalt, P. C. G4-forming sequences in the non-transcribed DNA strand pose blocks to T7 RNA polymerase and mammalian RNA polymerase II, *J Biol Chem* **2008**, *283*, 12756-12762.
20. Belotserkovskii, B. P.; Liu, R.; Tornaletti, S.; Krasilnikova, M. M.; Mirkin, S. M.; Hanawalt, P. C. Mechanisms and implications of transcription blockage by guanine-rich DNA sequences, *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 12816-12821.

21. Broxson, C.; Beckett, J.; Tornaletti, S. Transcription arrest by a G quadruplex forming-trinucleotide repeat sequence from the human c-myc gene, *Biochemistry* **2011**, *50*, 4162-4172.
22. Levy, M. A.; Kernohan, K. D.; Jiang, Y.; Berube, N. G. ATRX promotes gene expression by facilitating transcriptional elongation through guanine-rich coding regions, *Hum Mol Genet* **2014**.
23. Nguyen, G. H.; Tang, W.; Robles, A. I.; Beyer, R. P.; Gray, L. T.; Welsh, J. A.; Schetter, A. J.; Kumamoto, K.; Wang, X. W.; Hickson, I. D.; Maizels, N.; Monnat, R. J., Jr.; Harris, C. C. Regulation of gene expression by the BLM helicase correlates with the presence of G-quadruplex DNA motifs, *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 9905-9910.
24. Schiavone, D.; Guilbaud, G.; Murat, P.; Papadopoulou, C.; Sarkies, P.; Prioleau, M. N.; Balasubramanian, S.; Sale, J. E. Determinants of G quadruplex-induced epigenetic instability in REV1-deficient cells, *EMBO J.* **2014**, *33*, 2507-2520.
25. Vannier, J. B.; Pavicic-Kaltenbrunner, V.; Petalcorin, M. I.; Ding, H.; Boulton, S. J. RTEL1 dismantles T loops and counteracts telomeric G4-DNA to maintain telomere integrity, *Cell* **2012**, *149*, 795-806.
26. Dempsey, L. A.; Sun, H.; Hanakahi, L. A.; Maizels, N. G4 DNA binding by LR1 and its subunits, nucleolin and hnRNP D, A role for G-G pairing in immunoglobulin switch recombination, *J Biol Chem* **1999**, *274*, 1066-1071.
27. Hanakahi, L. A.; Sun, H.; Maizels, N. High affinity interactions of nucleolin with G-G-paired rDNA, *J Biol Chem* **1999**, *274*, 15908-15912.
28. Gonzalez, V.; Guo, K.; Hurley, L.; Sun, D. Identification and characterization of nucleolin as a c-myc G-quadruplex-binding protein, *J Biol Chem* **2009**, *284*, 23622-23635.
29. Gonzalez, V.; Hurley, L. H. The C-terminus of nucleolin promotes the formation of the c-MYC G-quadruplex and inhibits c-MYC promoter activity, *Biochemistry* **2010**, *49*, 9706-9714.
30. Edwards, D. N.; Machwe, A.; Wang, Z.; Orren, D. K. Intramolecular telomeric G-quadruplexes dramatically inhibit DNA synthesis by replicative and translesion polymerases, revealing their potential to lead to genetic change, *PLoS One* **2014**, *9*, e80664.
31. Masuda-Sasa, T.; Polaczek, P.; Peng, X. P.; Chen, L.; Campbell, J. L. Processing of G4 DNA by Dna2 helicase/nuclease and replication protein A (RPA) provides insights into the mechanism of Dna2/RPA substrate recognition, *J Biol Chem* **2008**, *283*, 24359-24373.
32. Ghosal, G.; Muniyappa, K. *Saccharomyces cerevisiae* Mre11 is a high-affinity G4 DNA-binding protein and a G-rich DNA-specific endonuclease: implications for replication of telomeric DNA, *Nucleic Acids Res.* **2005**, *33*, 4692-4703.
33. Liu, Z.; Frantz, J. D.; Gilbert, W.; Tye, B. K. Identification and characterization of a nuclease activity specific for G4 tetrastranded DNA, *Proc. Natl. Acad. Sci. U. S. A.* **1993**, *90*, 3157-3161.
34. Liu, Z.; Gilbert, W. The yeast KEM1 gene encodes a nuclease specific for G4 tetraplex DNA: implication of in vivo functions for this novel DNA structure, *Cell* **1994**, *77*, 1083-1092.
35. Sun, H.; Yabuki, A.; Maizels, N. A human nuclease specific for G4 DNA, *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 12444-12449.
36. Watson, J. D.; Crick, F. H. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid, *Nature* **1953**, *171*, 737-738.
37. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank, *Nucleic Acids Res.* **2000**, *28*, 235-242.
38. Wang, Y.; Patel, D. J. Guanine residues in d(T2AG3) and d(T2G4) form parallel-stranded potassium cation stabilized G-quadruplexes with anti glycosidic torsion angles in solution, *Biochemistry* **1992**, *31*, 8112-8119.
39. Parkinson, G. N.; Lee, M. P.; Neidle, S. Crystal structure of parallel quadruplexes from human telomeric DNA, *Nature* **2002**, *417*, 876-880.
40. Phan, A. T.; Patel, D. J. Two-repeat human telomeric d(TAGGGTTAGGGT) sequence forms interconverting parallel and antiparallel G-quadruplexes in solution: distinct topologies, thermodynamic properties, and folding/unfolding kinetics, *J Am Chem Soc* **2003**, *125*, 15021-15027.
41. Zhang, N.; Phan, A. T.; Patel, D. J. (3 + 1) Assembly of three human telomeric repeats into an asymmetric dimeric G-quadruplex, *J Am Chem Soc* **2005**, *127*, 17277-17285.
42. Wang, Y.; Patel, D. J. Solution structure of the human telomeric repeat d[AG3(T2AG3)3] G-tetraplex, *Structure* **1993**, *1*, 263-282.
43. Phan, A. T.; Kuryavii, V.; Luu, K. N.; Patel, D. J. Structure of two intramolecular G-quadruplexes formed by natural human telomere sequences in K⁺ solution, *Nucleic Acids Res.* **2007**, *35*, 6517-6525.
44. Matsugami, A.; Xu, Y.; Noguchi, Y.; Sugiyama, H.; Katahira, M. Structure of a human telomeric DNA sequence stabilized by 8-bromoguanosine substitutions, as determined by NMR in a K⁺ solution, *FEBS J.* **2007**, *274*, 3545-3556.

45. Dai, J.; Punchihewa, C.; Ambrus, A.; Chen, D.; Jones, R. A.; Yang, D. Structure of the intramolecular human telomeric G-quadruplex in potassium solution: a novel adenine triple formation, *Nucleic Acids Res.* **2007**, *35*, 2440-2450.
46. Ambrus, A.; Chen, D.; Dai, J.; Bialis, T.; Jones, R. A.; Yang, D. Human telomeric sequence forms a hybrid-type intramolecular G-quadruplex structure with mixed parallel/antiparallel strands in potassium solution, *Nucleic Acids Res.* **2006**, *34*, 2723-2735.
47. Xu, Y.; Noguchi, Y.; Sugiyama, H. The new models of the human telomere d[AGGG(TTAGGG)3] in K⁺ solution, *Bioorg Med Chem* **2006**, *14*, 5584-5591.
48. Phan, A. T.; Luu, K. N.; Patel, D. J. Different loop arrangements of intramolecular human telomeric (3+1) G-quadruplexes in K⁺ solution, *Nucleic Acids Res.* **2006**, *34*, 5715-5719.
49. Luu, K. N.; Phan, A. T.; Kuryavyi, V.; Lacroix, L.; Patel, D. J. Structure of the human telomere in K⁺ solution: An intramolecular (3+1) G-quadruplex scaffold, *J Am Chem Soc* **2006**, *128*, 9963-9970.
50. Dai, J.; Carver, M.; Punchihewa, C.; Jones, R. A.; Yang, D. Structure of the Hybrid-2 type intramolecular human telomeric G-quadruplex in K⁺ solution: insights into structure polymorphism of the human telomeric sequence, *Nucleic Acids Res.* **2007**, *35*, 4927-4940.
51. Lim, K. W.; Amrane, S.; Bouaziz, S.; Xu, W.; Mu, Y.; Patel, D. J.; Luu, K. N.; Phan, A. T. Structure of the human telomere in K⁺ solution: a stable basket-type G-quadruplex with only two G-tetrad layers, *J Am Chem Soc* **2009**, *131*, 4301-4309.
52. Phan, A. T. Human telomeric G-quadruplex: structures of DNA and RNA sequences, *FEBS J.* **2010**, *277*, 1107-1117.
53. Pinnavaia, T. J.; Marshall, C. L.; Mettler, C. M.; Fisk, C. L.; Miles, H. T.; Becker, E. D. Alkali metal ion specificity in the solution ordering of a nucleotide, 5'-guanosine monophosphate, *J Am Chem Soc* **1978**, *100*, 3625-3627.
54. Horvath, M. P.; Schultz, S. C. DNA G-quartets in a 1.86 Å resolution structure of an *Oxytricha nova* telomeric protein-DNA complex, *J Mol Biol* **2001**, *310*, 367-377.
55. Haider, S.; Parkinson, G. N.; Neidle, S. Crystal structure of the potassium form of an *Oxytricha nova* G-quadruplex, *J Mol Biol* **2002**, *320*, 189-200.
56. Arnott, S.; Chandrasekaran, R.; Marttila, C. M. Structures for polyinosinic acid and polyguanylic acid, *Biochem J* **1974**, *141*, 537-543.
57. Dao, N. T.; Haselsberger, R.; Michel-Beyerle, M. E.; Phan, A. T. Following G-quadruplex formation by its intrinsic fluorescence, *FEBS Lett* **2011**, *585*, 3969-3977.
58. Williamson, J. R.; Raghuraman, M. K.; Cech, T. R. Monovalent cation-induced structure of telomeric DNA: the G-quartet model, *Cell* **1989**, *59*, 871-880.
59. Rastegar, A.; 2011/01/21 ed.; Walker, H. K., Hall, H. D., Hurst, J. W., Eds.; Butterworths Publishers: Boston, 1990.
60. Xu, Q.; Deng, H.; Braunlin, W. H. Selective localization and rotational immobilization of univalent cations on quadruplex DNA, *Biochemistry* **1993**, *32*, 13130-13137.
61. Deng, H.; Braunlin, W. H. Kinetics of sodium ion binding to DNA quadruplexes, *J Mol Biol* **1996**, *255*, 476-483.
62. Ida, R.; Wu, G. Direct NMR detection of alkali metal ions bound to G-quadruplex DNA, *J Am Chem Soc* **2008**, *130*, 3590-3602.
63. Lee, J. S. The stability of polypurine tetraplexes in the presence of mono- and divalent cations, *Nucleic Acids Res.* **1990**, *18*, 6057-6060.
64. Smargiasso, N.; Rosu, F.; Hsia, W.; Colson, P.; Baker, E. S.; Bowers, M. T.; De Pauw, E.; Gabelica, V. G-quadruplex DNA assemblies: loop length, cation identity, and multimer formation, *J Am Chem Soc* **2008**, *130*, 10208-10216.
65. Phillips, K.; Dauter, Z.; Murchie, A. I.; Lilley, D. M.; Luisi, B. The crystal structure of a parallel-stranded guanine tetraplex at 0.95 Å resolution, *J Mol Biol* **1997**, *273*, 171-182.
66. Lander, E. S.; Linton, L. M.; Birren, B.; Nusbaum, C.; Zody, M. C.; Baldwin, J.; Devon, K.; Dewar, K.; Doyle, M.; FitzHugh, W.; Funke, R.; Gage, D.; Harris, K.; Heaford, A.; Howland, J.; Kann, L.; Lehoczky, J.; LeVine, R.; McEwan, P.; McKernan, K.; Meldrim, J.; Mesirov, J. P.; Miranda, C.; Morris, W.; Naylor, J.; Raymond, C.; Rosetti, M.; Santos, R.; Sheridan, A.; Sougnez, C.; Stange-Thomann, N.; Stojanovic, N.; Subramanian, A.; Wyman, D.; Rogers, J.; Sulston, J.; Ainscough, R.; Beck, S.; Bentley, D.; Burton, J.; Clee, C.; Carter, N.; Coulson, A.; Deadman, R.; Deloukas, P.; Dunham, A.; Dunham, I.; Durbin, R.; French, L.; Grafham, D.; Gregory, S.; Hubbard, T.; Humphray, S.; Hunt, A.; Jones, M.; Lloyd, C.; McMurray, A.; Matthews, L.; Mercer, S.; Milne, S.; Mullikin, J. C.; Mungall, A.; Plumb, R.; Ross, M.; Shownkeen, R.; Sims, S.; Waterston, R. H.; Wilson, R. K.; Hillier, L. W.; McPherson, J. D.; Marra, M. A.; Mardis, E. R.; Fulton, L. A.; Chinwalla, A. T.; Pepin, K. H.; Gish, W. R.; Chissoe, S. L.; Wendl, M. C.; Delehaunty, K. D.; Miner, T. L.; Delehaunty, A.; Kramer, J. B.; Cook, L. L.; Fulton, R. S.; Johnson, D. L.; Minx, P. J.; Clifton, S. W.; Hawkins, T.; Branscomb, E.; Predki, P.; Richardson, P.; Wenning, S.; Slezak, T.

- Doggett, N.; Cheng, J. F.; Olsen, A.; Lucas, S.; Elkin, C.; Uberbacher, E.; Frazier, M. Initial sequencing and analysis of the human genome, *Nature* **2001**, *409*, 860-921.
67. Neidle, S.; Parkinson, G. N. Quadruplex DNA crystal structures and drug design, *Biochimie* **2008**, *90*, 1184-1196.
68. Patel, D. J.; Phan, A. T.; Kuryavyi, V. Human telomere, oncogenic promoter and 5'-UTR G-quadruplexes: diverse higher order DNA and RNA targets for cancer therapeutics, *Nucleic Acids Res.* **2007**, *35*, 7429-7455.
69. Fernando, H.; Rodriguez, R.; Balasubramanian, S. Selective recognition of a DNA G-quadruplex by an engineered antibody, *Biochemistry* **2008**, *47*, 9365-9371.
70. Lam, E. Y.; Beraldi, D.; Tannahill, D.; Balasubramanian, S. G-quadruplex structures are stable and detectable in human genomic DNA, *Nat Commun* **2013**, *4*, 1796.
71. Blackburn, E. H. Telomeres and telomerase: the means to the end (Nobel lecture), *Angew. Chem. Int. Ed. Engl.* **2010**, *49*, 7405-7421.
72. Paeschke, K.; Simonsson, T.; Postberg, J.; Rhodes, D.; Lipps, H. J. Telomere end-binding proteins control the formation of G-quadruplex DNA structures in vivo, *Nat Struct Mol Biol* **2005**, *12*, 847-854.
73. Schaffitzel, C.; Berger, I.; Postberg, J.; Hanes, J.; Lipps, H. J.; Pluckthun, A. In vitro generated antibodies specific for telomeric guanine-quadruplex DNA react with *Stylochyia lemnae* macronuclei, *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 8572-8577.
74. Temime-Smaali, N.; Guittat, L.; Sidibe, A.; Shin-ya, K.; Trentesaux, C.; Riou, J. F. The G-quadruplex ligand telomestatin impairs binding of topoisomerase IIIalpha to G-quadruplex-forming oligonucleotides and uncaps telomeres in ALT cells, *PLoS One* **2009**, *4*, e6919.
75. Kim, M. Y.; Vankayalapati, H.; Shin-Ya, K.; Wierzb, K.; Hurley, L. H. Telomestatin, a potent telomerase inhibitor that interacts quite specifically with the human telomeric intramolecular g-quadruplex, *J Am Chem Soc* **2002**, *124*, 2098-2099.
76. Shin-ya, K.; Wierzb, K.; Matsuo, K.; Ohtani, T.; Yamada, Y.; Furihata, K.; Hayakawa, Y.; Seto, H. Telomestatin, a novel telomerase inhibitor from *Streptomyces anulatus*, *J Am Chem Soc* **2001**, *123*, 1262-1263.
77. Baumann, P.; Cech, T. R. Pot1, the putative telomere end-binding protein in fission yeast and humans, *Science* **2001**, *292*, 1171-1175.
78. Zaug, A. J.; Podell, E. R.; Cech, T. R. Human POT1 disrupts telomeric G-quadruplexes allowing telomerase extension in vitro, *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 10864-10869.
79. Sun, H.; Karow, J. K.; Hickson, I. D.; Maizels, N. The Bloom's syndrome helicase unwinds G4 DNA, *J Biol Chem* **1998**, *273*, 27587-27592.
80. Murat, P.; Balasubramanian, S. Existence and consequences of G-quadruplex structures in DNA, *Curr Opin Genet Dev* **2014**, *25*, 22-29.
81. Mirkin, S. M. DNA replication: driving past four-stranded snags, *Nature* **2013**, *497*, 449-450.
82. Tornaletti, S. Transcriptional processing of G4 DNA, *Mol Carcinog* **2009**, *48*, 326-335.
83. Duquette, M. L.; Handa, P.; Vincent, J. A.; Taylor, A. F.; Maizels, N. Intracellular transcription of G-rich DNAs induces formation of G-loops, novel structures containing G4 DNA, *Genes Dev* **2004**, *18*, 1618-1629.
84. Castillo Bosch, P.; Segura-Bayona, S.; Koole, W.; van Heteren, J. T.; Dewar, J. M.; Tijsterman, M.; Knipscheer, P. FANCD1 promotes DNA synthesis through G-quadruplex structures, *EMBO J.* **2014**, *33*, 2521-2533.
85. Paeschke, K.; Capra, J. A.; Zakian, V. A. DNA replication through G-quadruplex motifs is promoted by the *Saccharomyces cerevisiae* Pif1 DNA helicase, *Cell* **2011**, *145*, 678-691.
86. Postberg, J.; Tsytlonok, M.; Sparvoli, D.; Rhodes, D.; Lipps, H. J. A telomerase-associated RecQ protein-like helicase resolves telomeric G-quadruplex structures during replication, *Gene* **2012**, *497*, 147-154.
87. Duquette, M. L.; Pham, P.; Goodman, M. F.; Maizels, N. AID binds to transcription-induced structures in c-MYC that map to regions associated with translocation and hypermutation, *Oncogene* **2005**, *24*, 5791-5798.
88. Xiao, S.; Zhang, J. Y.; Wu, J.; Wu, R. Y.; Xia, Y.; Zheng, K. W.; Hao, Y. H.; Zhou, X.; Tan, Z. Formation of DNA:RNA hybrid G-quadruplexes of two G-quartet layers in transcription: expansion of the prevalence and diversity of G-quadruplexes in genomes, *Angew. Chem. Int. Ed. Engl.* **2014**, *53*, 13110-13114.
89. Maizels, N. Dynamic roles for G4 DNA in the biology of eukaryotic cells, *Nat Struct Mol Biol* **2006**, *13*, 1055-1059.
90. Huber, M. D.; Duquette, M. L.; Shiels, J. C.; Maizels, N. A conserved G4 DNA binding domain in RecQ family helicases, *J. Mol. Biol.* **2006**, *358*, 1071-1080.
91. Larson, E. D.; Duquette, M. L.; Cummings, W. J.; Streiff, R. J.; Maizels, N. MutS alpha binds to and promotes synapsis of transcriptionally activated immunoglobulin switch regions, *Curr. Biol.* **2005**, *15*, 470-474.

92. Gray, L. T.; Vallur, A. C.; Eddy, J.; Maizels, N. G quadruplexes are genomewide targets of transcriptional helicases XPB and XPD, *Nat Chem Biol* **2014**, *10*, 313-318.
93. Chung, W. J.; Heddi, B.; Hamon, F.; Teulade-Fichou, M. P.; Phan, A. T. Solution structure of a G-quadruplex bound to the bisquinolinium compound Phen-DC(3), *Angew. Chem. Int. Ed. Engl.* **2014**, *53*, 999-1002.
94. Chung, W. J.; Heddi, B.; Tera, M.; Iida, K.; Nagasawa, K.; Phan, A. T. Solution structure of an intramolecular (3 + 1) human telomeric G-quadruplex bound to a telomestatin derivative, *J Am Chem Soc* **2013**, *135*, 13495-13501.
95. Phan, A. T.; Kuryavyi, V.; Ma, J. B.; Faure, A.; Andreola, M. L.; Patel, D. J. An interlocked dimeric parallel-stranded DNA quadruplex: a potent inhibitor of HIV-1 integrase, *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 634-639.
96. Friedberg, E. C. DNA damage and repair, *Nature* **2003**, *421*, 436-440.
97. Rai, P.; Onder, T. T.; Young, J. J.; McFaline, J. L.; Pang, B.; Dedon, P. C.; Weinberg, R. A. Continuous elimination of oxidized nucleotides is necessary to prevent rapid onset of cellular senescence, *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 169-174.
98. Meira, L. B.; Bugni, J. M.; Green, S. L.; Lee, C. W.; Pang, B.; Borenshtein, D.; Rickman, B. H.; Rogers, A. B.; Moroski-Erkul, C. A.; McFaline, J. L.; Schauer, D. B.; Dedon, P. C.; Fox, J. G.; Samson, L. D. DNA damage induced by chronic inflammation contributes to colon carcinogenesis in mice, *J Clin Invest* **2008**, *118*, 2516-2525.
99. Wiseman, H.; Halliwell, B. Damage to DNA by reactive oxygen and nitrogen species: role in inflammatory disease and progression to cancer, *Biochem. J.* **1996**, *313* (Pt 1), 17-29.
100. Kamiya, H. Mutagenic potentials of damaged nucleic acids produced by reactive oxygen/nitrogen species: approaches using synthetic oligonucleotides and nucleotides: survey and summary, *Nucleic Acids Res.* **2003**, *31*, 517-531.
101. Caulfield, J. L.; Wishnok, J. S.; Tannenbaum, S. R. Nitric oxide-induced deamination of cytosine and guanine in deoxynucleosides and oligonucleotides, *J Biol Chem* **1998**, *273*, 12689-12695.
102. Steenzen, S.; Jovanovic, S. V. How Easily Oxidizable Is DNA? One-Electron Reduction Potentials of Adenosine and Guanosine Radicals in Aqueous Solution, *J Am Chem Soc* **1997**, *119*, 617-618.
103. Misiaszek, R.; Crean, C.; Joffe, A.; Geacintov, N. E.; Shafirovich, V. Oxidative DNA damage associated with combination of guanine and superoxide radicals and repair mechanisms via radical trapping, *J Biol Chem* **2004**, *279*, 32106-32115.
104. Oikawa, S.; Kawanishi, S. Site-specific DNA damage at GGG sequence by oxidative stress may accelerate telomere shortening, *FEBS Lett* **1999**, *453*, 365-368.
105. Kawai, K.; Fujitsuka, M.; Majima, T. Selective guanine oxidation by UVB-irradiation in telomeric DNA, *Chem. Commun.* **2005**, 1476-1477.
106. Gedik, C. M.; Collins, A. Establishing the background level of base oxidation in human lymphocyte DNA: results of an interlaboratory validation study, *FASEB J* **2005**, *19*, 82-84.
107. Ohno, M.; Miura, T.; Furuichi, M.; Tominaga, Y.; Tsuchimoto, D.; Sakumi, K.; Nakabeppu, Y. A genome-wide distribution of 8-oxoguanine correlates with the preferred regions for recombination and single nucleotide polymorphism in the human genome, *Genome Res* **2006**, *16*, 567-575.
108. Oikawa, S.; Tada-Oikawa, S.; Kawanishi, S. Site-specific DNA damage at the GGG sequence by UVA involves acceleration of telomere shortening, *Biochemistry* **2001**, *40*, 4763-4768.
109. Wang, Z.; Rhee, D. B.; Lu, J.; Bohr, C. T.; Zhou, F.; Vallabhaneni, H.; de Souza-Pinto, N. C.; Liu, Y. Characterization of oxidative guanine damage and repair in mammalian telomeres, *PLoS Genet* **2010**, *6*, e1000951.
110. Szalai, V. A.; Singer, M. J.; Thorp, H. H. Site-specific probing of oxidative reactivity and telomerase function using 7,8-dihydro-8-oxoguanine in telomeric DNA, *J Am Chem Soc* **2002**, *124*, 1625-1631.
111. Vorlickova, M.; Tomasko, M.; Sagi, A. J.; Bednarova, K.; Sagi, J. 8-oxoguanine in a quadruplex of the human telomere DNA sequence, *FEBS J.* **2012**, *279*, 29-39.
112. Lech, C. J.; Cheow Lim, J. K.; Wen Lim, J. M.; Amrane, S.; Heddi, B.; Phan, A. T. Effects of site-specific guanine C8-modifications on an intramolecular DNA G-quadruplex, *Biophys. J.* **2011**, *101*, 1987-1998.
113. Zhou, J.; Fleming, A. M.; Averill, A. M.; Burrows, C. J.; Wallace, S. S. The NEIL glycosylases remove oxidized guanine lesions from telomeric and promoter quadruplex DNA structures, *Nucleic Acids Res.* **2015**.
114. Singh, V.; Benz, A.; Hartig, J. S. G quadruplexes stabilised by 8-oxo-2'-deoxyguanosine, *Chemistry* **2011**, *17*, 10838-10843.
115. Gros, J.; Rosu, F.; Amrane, S.; De Cian, A.; Gabelica, V.; Lacroix, L.; Mergny, J. L. Guanines are a quartet's best friend: impact of base substitutions on the kinetics and stability of tetramolecular quadruplexes, *Nucleic Acids Res.* **2007**, *35*, 3064-3075.

116. Suzuki, T.; Ide, H.; Yamada, M.; Endo, N.; Kanaori, K.; Tajima, K.; Morii, T.; Makino, K. Formation of 2'-deoxyoxanosine from 2'-deoxyguanosine and nitrous acid: mechanism and intermediates, *Nucleic Acids Res.* **2000**, *28*, 544-551.
117. Vongchampa, V.; Dong, M.; Gingipalli, L.; Dedon, P. Stability of 2'-deoxyxanthosine in DNA, *Nucleic Acids Res.* **2003**, *31*, 1045-1051.
118. Bass, B. L. RNA editing by adenosine deaminases that act on RNA, *Annu Rev Biochem* **2002**, *71*, 817-846.
119. Wuenschell, G. E.; O'Connor, T. R.; Termini, J. Stability, miscoding potential, and repair of 2'-deoxyxanthosine in DNA: implications for nitric oxide-induced mutagenesis, *Biochemistry* **2003**, *42*, 3608-3616.
120. Pang, B.; McFaline, J. L.; Burgis, N. E.; Dong, M.; Taghizadeh, K.; Sullivan, M. R.; Elmquist, C. E.; Cunningham, R. P.; Dedon, P. C. Defects in purine nucleotide metabolism lead to substantial incorporation of xanthine and hypoxanthine into DNA and RNA, *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 2319-2324.
121. Rudner, R.; Karkas, J. D.; Chargaff, E. Separation of *B. subtilis* DNA into complementary strands. 3. Direct analysis, *Proc. Natl. Acad. Sci. U. S. A.* **1968**, *60*, 921-922.
122. Chargaff, E.; Lipshitz, R.; Green, C. Composition of the desoxypentose nucleic acids of four genera of sea-urchin, *J Biol Chem* **1952**, *195*, 155-160.
123. Dong, M.; Wang, C.; Deen, W. M.; Dedon, P. C. Absence of 2'-deoxyoxanosine and presence of abasic sites in DNA exposed to nitric oxide at controlled physiological concentrations, *Chem. Res. Toxicol.* **2003**, *16*, 1044-1055.
124. Dong, M.; Dedon, P. C. Relatively small increases in the steady-state levels of nucleobase deamination products in DNA from human TK6 cells exposed to toxic levels of nitric oxide, *Chem. Res. Toxicol.* **2006**, *19*, 50-57.
125. Yurenko, Y. P.; Novotny, J.; Sklenar, V.; Marek, R. Exploring non-covalent interactions in guanine- and xanthine-based model DNA quadruplex structures: a comprehensive quantum chemical approach, *Phys Chem Chem Phys* **2014**, *16*, 2072-2084.
126. Paragi, G.; Kovacs, L.; Kupihar, Z.; Szolomajer, J.; Penke, B.; Fonseca Guerra, C.; Bickelhaupt, F. M. Neutral and positively charged new purine tetramer structures: a computational study of xanthine and uric acid derivatives, *New J. Chem* **2011**, *35*, 119-126.
127. Szolomajer, J.; Paragi, G.; Batta, G.; Guerra, C. F.; Bickelhaupt, F. M.; Kele, Z.; Padar, P.; Kupihar, Z.; Kovacs, L. 3-Substituted xanthines as promising candidates for quadruplex formation: computational, synthetic and analytical studies, *New J. Chem.* **2011**, *35*, 476-482.
128. Novotný, J.; Kulhánek, P.; Marek, R. Biocompatible Xanthine-Quadruplex Scaffold for Ion-Transporting DNA Channels, *J. Phys. Chem. Lett.* **2012**, *3*, 1788-1792.
129. Yurenko, Y. P.; Novotný, J.; Mitoraj, M. P.; Sklenář, V.; Michalak, A.; Marek, R. Nucleic Acid Quadruplexes Based on 8-Halo-9-deazaxanthines: Energetics and Noncovalent Interactions in Quadruplex Stems, *J. Chem.Theory Comput.* **2014**.
130. Novotny, J.; Yurenko, Y. P.; Kulhanek, P.; Marek, R. Tailoring the properties of quadruplex nucleobases for biological and nanomaterial applications, *Phys Chem Chem Phys* **2014**, *16*, 15241-15248.
131. Di Antonio, M.; Rodriguez, R.; Balasubramanian, S. Experimental approaches to identify cellular G-quadruplex structures and functions, *Methods*.
132. Fry, M. Tetraplex DNA and its interacting proteins, *Front Biosci* **2007**, *12*, 4336-4351.
133. Oganessian, L.; Bryan, T. M. Physiological relevance of telomeric G-quadruplex formation: a potential drug target, *Bioessays* **2007**, *29*, 155-165.
134. Wu, Y.; Brosh, R. M., Jr. G-quadruplex nucleic acids and human disease, *FEBS J.* **2010**, *277*, 3470-3488.
135. Sissi, C.; Gatto, B.; Palumbo, M. The evolving world of protein-G-quadruplex recognition: A medicinal chemist's perspective, *Biochimie* **2011**, *93*, 1219-1230.
136. Larsen, N. B.; Hickson, I. D. RecQ Helicases: Conserved Guardians of Genomic Integrity, *Adv. Exp. Med. Biol.* **2013**, *767*, 161-184.
137. Opreko, P. L.; Cheng, W. H.; Bohr, V. A. Junction of RecQ helicase biochemistry and human disease, *J Biol Chem* **2004**, *279*, 18099-18102.
138. Rezazadeh, S. RecQ helicases; at the crossroad of genome replication, repair, and recombination, *Mol Biol Rep* **2012**, *39*, 4527-4543.
139. Rossi, M. L.; Ghosh, A. K.; Bohr, V. A. Roles of Werner syndrome protein in protection of genome integrity, *DNA Repair* **2010**, *9*, 331-344.
140. Vindigni, A.; Hickson, I. D. RecQ helicases: multiple structures for multiple functions?, *HFSP J.* **2009**, *3*, 153-164.
141. Wu, L.; Hickson, I. D. RecQ helicases and cellular responses to DNA damage, *Mutat Res* **2002**, *509*, 35-47.

142. Bryan, T. M.; Baumann, P. G-quadruplexes: from guanine gels to chemotherapeutics, *Mol Biotechnol* **2011**, *49*, 198-208.
143. Enokizono, Y.; Konishi, Y.; Nagata, K.; Ouhashi, K.; Uesugi, S.; Ishikawa, F.; Katahira, M. Structure of hnRNP D complexed with single-stranded telomere DNA and unfolding of the quadruplex by heterogeneous nuclear ribonucleoprotein D, *J Biol Chem* **2005**, *280*, 18862-18870.
144. Ding, J.; Hayashi, M. K.; Zhang, Y.; Manche, L.; Krainer, A. R.; Xu, R. M. Crystal structure of the two-RRM domain of hnRNP A1 (UP1) complexed with single-stranded telomeric DNA, *Genes Dev.* **1999**, *13*, 1102-1115.
145. Fukuda, H.; Katahira, M.; Tanaka, E.; Enokizono, Y.; Tsuchiya, N.; Higuchi, K.; Nagao, M.; Nakagama, H. Unfolding of higher DNA structures formed by the d(CGG) triplet repeat by UP1 protein, *Genes Cells* **2005**, *10*, 953-962.
146. Fukuda, H.; Katahira, M.; Tsuchiya, N.; Enokizono, Y.; Sugimura, T.; Nagao, M.; Nakagama, H. Unfolding of quadruplex structure in the G-rich strand of the minisatellite repeat by the binding protein UP1, *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 12685-12690.
147. Wang, H.; Nora, G. J.; Ghodke, H.; Opresko, P. L. Single molecule studies of physiologically relevant telomeric tails reveal POT1 mechanism for promoting G-quadruplex unfolding, *J Biol Chem* **2011**, *286*, 7479-7489.
148. Creacy, S. D.; Routh, E. D.; Iwamoto, F.; Nagamine, Y.; Akman, S. A.; Vaughn, J. P. G4 resolvase 1 binds both DNA and RNA tetramolecular quadruplex with high affinity and is the major source of tetramolecular quadruplex G4-DNA and G4-RNA resolving activity in HeLa cell lysates, *J Biol Chem* **2008**, *283*, 34626-34634.
149. Huber, M. D.; Lee, D. C.; Maizels, N. G4 DNA unwinding by BLM and Sgs1p: substrate specificity and substrate-specific inhibition, *Nucleic Acids Res.* **2002**, *30*, 3954-3961.
150. De Cian, A.; Lacroix, L.; Douarre, C.; Temime-Smaali, N.; Trentesaux, C.; Riou, J. F.; Mergny, J. L. Targeting telomeres and telomerase, *Biochimie* **2008**, *90*, 131-155.
151. Bryan, T. M.; Jarstfer, M. B. Interrogation of G-quadruplex-protein interactions, *Methods* **2007**, *43*, 332-339.
152. Tran, H.; Schilling, M.; Wirbelauer, C.; Hess, D.; Nagamine, Y. Facilitation of mRNA deadenylation and decay by the exosome-bound, DEXH protein RHAU, *Mol Cell* **2004**, *13*, 101-111.
153. Chalupnikova, K.; Lattmann, S.; Selak, N.; Iwamoto, F.; Fujiki, Y.; Nagamine, Y. Recruitment of the RNA helicase RHAU to stress granules via a unique RNA-binding domain, *J Biol Chem* **2008**, *283*, 35186-35198.
154. Booy, E. P.; Howard, R.; Marushchak, O.; Ariyo, E. O.; Meier, M.; Novakowski, S. K.; Deo, S. R.; Dzananovic, E.; Stetefeld, J.; McKenna, S. A. The RNA helicase RHAU (DHX36) suppresses expression of the transcription factor PITX1, *Nucleic Acids Res.* **2014**, *42*, 3346-3361.
155. Huang, W.; Smaldino, P. J.; Zhang, Q.; Miller, L. D.; Cao, P.; Stadelman, K.; Wan, M.; Giri, B.; Lei, M.; Nagamine, Y.; Vaughn, J. P.; Akman, S. A.; Sui, G. Yin Yang 1 contains G-quadruplex structures in its promoter and 5'-UTR and its expression is modulated by G4 resolvase 1, *Nucleic Acids Res.* **2012**, *40*, 1033-1049.
156. Iwamoto, F.; Stadler, M.; Chalupnikova, K.; Oakeley, E.; Nagamine, Y. Transcription-dependent nucleolar cap localization and possible nuclear function of DEXH RNA helicase RHAU, *Exp. Cell Res.* **2008**, *314*, 1378-1391.
157. Fu, J. J.; Li, L. Y.; Lu, G. X. Molecular cloning and characterization of human DDX36 and mouse Ddx36 genes, new members of the DEAD/H box superfamily, *Sheng Wu Hua Xue Yu Sheng Wu Wu Li Xue Bao (Shanghai)* **2002**, *34*, 655-661.
158. Gao, X.; Ma, W.; Nie, J.; Zhang, C.; Zhang, J.; Yao, G.; Han, J.; Xu, J.; Hu, B.; Du, Y.; Shi, Q.; Yang, Z.; Huang, X.; Zhang, Y. A G-quadruplex DNA structure resolvase, RHAU, is essential for spermatogonia differentiation, *Cell Death Dis* **2015**, *6*, e1610.
159. Brazda, V.; Haronikova, L.; Liao, J. C.; Fojta, M. DNA and RNA quadruplex-binding proteins, *Int J Mol Sci* **2014**, *15*, 17493-17517.
160. He, F.; Muto, Y.; Inoue, M.; Kigawa, T.; Shirouzu, M.; Terada, T.; Yokoyama, S. Solution structure of rrm domain in heterogeneous nuclear ribonucleoproteins a2/b1, *Unpublished work* **2005**.
161. Moran-Jones, K.; Wayman, L.; Kennedy, D. D.; Reddel, R. R.; Sara, S.; Snee, M. J.; Smith, R. hnRNP A2, a potential ssDNA/RNA molecular adapter at the telomere, *Nucleic Acids Res.* **2005**, *33*, 486-496.
162. Nandakumar, J.; Podell, E. R.; Cech, T. R. How telomeric protein POT1 avoids RNA to achieve specificity for single-stranded DNA, *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 651-656.
163. Deng, X.; Habel, J. E.; Kabaleeswaran, V.; Snell, E. H.; Wold, M. S.; Borgstahl, G. E. Structure of the full-length human RPA14/32 complex gives insights into the mechanism of DNA binding and complex formation, *J Mol Biol* **2007**, *374*, 865-876.

164. Bochkareva, E.; Korolev, S.; Lees-Miller, S. P.; Bochkarev, A. Structure of the RPA trimerization core and its role in the multistep DNA-binding mechanism of RPA, *EMBO J.* **2002**, *21*, 1855-1863.
165. Bochkareva, E.; Belegu, V.; Korolev, S.; Bochkarev, A. Structure of the major single-stranded DNA-binding domain of replication protein A suggests a dynamic mechanism for DNA binding, *EMBO J.* **2001**, *20*, 612-618.
166. Sanders, C. M. Human Pif1 helicase is a G-quadruplex DNA-binding protein with G-quadruplex DNA-unwinding activity, *Biochem. J.* **2010**, *430*, 119-128.
167. Swan, M. K.; Legris, V.; Tanner, A.; Reaper, P. M.; Vial, S.; Bordas, R.; Pollard, J. R.; Charlton, P. A.; Golec, J. M.; Bertrand, J. A. Structure of human Bloom's syndrome helicase in complex with ADP and duplex DNA, *Acta Crystallogr D Biol Crystallogr* **2014**, *70*, 1465-1475.
168. Newman, J. A.; Savitsky, P.; Allerston, C. K.; Pike, A. C. W.; Pardon, E.; Steyaert, J.; Arrowsmith, C. H.; Von Delft, F.; Bountra, C.; Edwards, A.; Gileadi, O. Crystal structure of the Bloom's syndrome helicase BLM in complex with Nanobody, *Unpublished work* **2014**.
169. Kitano, K.; Kim, S.-Y.; Hakoshima, T. Structural Basis for DNA Strand Separation by the Unconventional Winged-Helix Domain of RecQ Helicase WRN, *Structure* **2010**, *18*, 177-187.
170. Bernstein, D. A.; Zittel, M. C.; Keck, J. L. High-resolution structure of the E.coli RecQ helicase catalytic core, *EMBO J.* **2003**, *22*, 4910-4921.
171. Lin, W.; Sampathi, S.; Dai, H.; Liu, C.; Zhou, M.; Hu, J.; Huang, Q.; Campbell, J.; Shin-Ya, K.; Zheng, L.; Chai, W.; Shen, B. Mammalian DNA2 helicase/nuclease cleaves G-quadruplex DNA and is required for telomere integrity, *EMBO J.* **2013**, *32*, 1425-1439.
172. Konig, P.; Giraldo, R.; Chapman, L.; Rhodes, D. The crystal structure of the DNA-binding domain of yeast RAP1 in complex with telomeric DNA, *Cell* **1996**, *85*, 125-136.
173. Liu, X.; Niu, C.; Ren, J.; Zhang, J.; Xie, X.; Zhu, H.; Feng, W.; Gong, W. The RRM domain of human fused in sarcoma protein reveals a non-canonical nucleic acid binding site, *Biochim Biophys Acta* **2013**, *1832*, 375-385.
174. Takahama, K.; Oyoshi, T. Specific binding of modified RGG domain in TLS/FUS to G-quadruplex RNA: tyrosines in RGG domain recognize 2'-OH of the riboses of loops in G-quadruplex, *J Am Chem Soc* **2013**, *135*, 18016-18019.
175. Court, R.; Chapman, L.; Fairall, L.; Rhodes, D. How the human telomeric proteins TRF1 and TRF2 recognize telomeric DNA: a view from high-resolution crystal structures, *EMBO Rep.* **2005**, *6*, 39-45.
176. Hanaoka, S.; Nagadoi, A.; Nishimura, Y. Comparison between TRF2 and TRF1 of their telomeric DNA-bound structures and DNA-binding activities, *Protein Sci.* **2005**, *14*, 119-130.
177. Bensaid, M.; Melko, M.; Bechara, E. G.; Davidovic, L.; Berretta, A.; Catania, M. V.; Gecz, J.; Lalli, E.; Bardoni, B. FRAXE-associated mental retardation protein (FMR2) is an RNA-binding protein with high affinity for G-quartet RNA forming structure, *Nucleic Acids Res.* **2009**, *37*, 1269-1279.
178. Melko, M.; Douguet, D.; Bensaid, M.; Zongaro, S.; Verheggen, C.; Gecz, J.; Bardoni, B. Functional characterization of the AFF (AF4/FMR2) family of RNA-binding proteins: insights into the molecular pathology of FRAXE intellectual disability, *Hum Mol Genet* **2011**, *20*, 1873-1885.
179. Li, Q. J.; Tong, X. J.; Duan, Y. M.; Zhou, J. Q. Characterization of the intramolecular G-quadruplex promoting activity of Est1, *FEBS Lett* **2013**, *587*, 659-665.
180. Gallo, A.; Lo Sterzo, C.; Mori, M.; Di Matteo, A.; Bertini, I.; Banci, L.; Brunori, M.; Federici, L. Structure of nucleophosmin DNA-binding domain and analysis of its complex with a G-quadruplex sequence from the c-MYC promoter, *J Biol Chem* **2012**, *287*, 26539-26548.
181. Finger, L. D.; Trantirek, L.; Johansson, C.; Feigon, J. Solution structures of stem-loop RNAs that bind to the two N-terminal RNA-binding domains of nucleolin, *Nucleic Acids Res.* **2003**, *31*, 6461-6472.
182. Giri, B.; Smaldino, P. J.; Thys, R. G.; Creacy, S. D.; Routh, E. D.; Hantgan, R. R.; Lattmann, S.; Nagamine, Y.; Akman, S. A.; Vaughn, J. P. G4 resolvase 1 tightly binds and unwinds unimolecular G4-DNA, *Nucleic Acids Res.* **2011**, *39*, 7161-7178.
183. Lattmann, S.; Stadler, M. B.; Vaughn, J. P.; Akman, S. A.; Nagamine, Y. The DEAH-box RNA helicase RHAU binds an intramolecular RNA G-quadruplex in TERC and associates with telomerase holoenzyme, *Nucleic Acids Res.* **2011**, *39*, 9390-9404.
184. Lattmann, S.; Giri, B.; Vaughn, J. P.; Akman, S. A.; Nagamine, Y. Role of the amino terminal RHAU-specific motif in the recognition and resolution of guanine quadruplex-RNA by the DEAH-box RNA helicase RHAU, *Nucleic Acids Res.* **2010**, *38*, 6219-6233.
185. Booy, E. P.; Meier, M.; Okun, N.; Novakowski, S. K.; Xiong, S.; Stetefeld, J.; McKenna, S. A. The RNA helicase RHAU (DHX36) unwinds a G4-quadruplex in human telomerase RNA and promotes the formation of the P1 helix template boundary, *Nucleic Acids Res.* **2012**, *40*, 4110-4124.
186. Goss, K. H.; Risinger, M. A.; Kordich, J. J.; Sanz, M. M.; Straughen, J. E.; Slovek, L. E.; Capobianco, A. J.; German, J.; Boivin, G. P.; Groden, J. Enhanced tumor formation in mice heterozygous for Blm mutation, *Science* **2002**, *297*, 2051-2053.

187. Khakhar, R. R.; Cobb, J. A.; Bjergbaek, L.; Hickson, I. D.; Gasser, S. M. RecQ helicases: multiple roles in genome maintenance, *Trends Cell Biol.* **2003**, *13*, 493-501.
188. Ellis, N. A.; Groden, J.; Ye, T. Z.; Straughen, J.; Lennon, D. J.; Ciocchi, S.; Proytcheva, M.; German, J. The Bloom's syndrome gene product is homologous to RecQ helicases, *Cell* **1995**, *83*, 655-666.
189. Jankowsky, E.; Jankowsky, A. The DEXH/D protein family database, *Nucleic Acids Res.* **2000**, *28*, 333-334.
190. Nakayama, H.; Nakayama, K.; Nakayama, R.; Irino, N.; Nakayama, Y.; Hanawalt, P. C. Isolation and genetic characterization of a thymineless death-resistant mutant of *Escherichia coli* K12: identification of a new mutation (recQ1) that blocks the RecF recombination pathway, *Mol Gen Genet* **1984**, *195*, 474-480.
191. Bernstein, K. A.; Gangloff, S.; Rothstein, R. The RecQ DNA helicases in DNA repair, *Annu. Rev. Genet.* **2010**, *44*, 393-417.
192. Kaneko, H.; Orii, K. O.; Matsui, E.; Shimozawa, N.; Fukao, T.; Matsumoto, T.; Shimamoto, A.; Furuichi, Y.; Hayakawa, S.; Kasahara, K.; Kondo, N. BLM (the causative gene of Bloom syndrome) protein translocation into the nucleus by a nuclear localization signal, *Biochem. Biophys. Res. Commun.* **1997**, *240*, 348-353.
193. Manthei, K. A.; Keck, J. L. The BLM dissolvosome in DNA replication and repair, *Cell Mol Life Sci* **2013**, *70*, 4067-4084.
194. Karow, J. K.; Chakraverty, R. K.; Hickson, I. D. The Bloom's syndrome gene product is a 3'-5' DNA helicase, *J Biol Chem* **1997**, *272*, 30611-30614.
195. Langland, G.; Elliott, J.; Li, Y.; Creaney, J.; Dixon, K.; Groden, J. The BLM helicase is necessary for normal DNA double-strand break repair, *Cancer Res* **2002**, *62*, 2766-2770.
196. Wan, L.; Han, J.; Liu, T.; Dong, S.; Xie, F.; Chen, H.; Huang, J. Scaffolding protein SPIDR/KIAA0146 connects the Bloom syndrome helicase with homologous recombination repair, *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 10646-10651.
197. Pichierri, P.; Franchitto, A.; Rosselli, F. BLM and the FANC proteins collaborate in a common pathway in response to stalled replication forks, *EMBO J.* **2004**, *23*, 3154-3163.
198. Bachrati, C. Z.; Hickson, I. D. Dissolution of double Holliday junctions by the concerted action of BLM and topoisomerase III α , *Methods Mol Biol* **2009**, *582*, 91-102.
199. Chu, W. K.; Hickson, I. D. RecQ helicases: multifunctional genome caretakers, *Nat. Rev. Cancer* **2009**, *9*, 644-654.
200. Hickson, I. D. RecQ helicases: Caretakers of the genome, *Nat. Rev. Cancer* **2003**, *3*, 169-178.
201. Vaughn, J. P.; Creacy, S. D.; Routh, E. D.; Joyner-Butt, C.; Jenkins, G. S.; Pauli, S.; Nagamine, Y.; Akman, S. A. The DEXH protein product of the DHX36 gene is the major source of tetramolecular quadruplex G4-DNA resolving activity in HeLa cell lysates, *J Biol Chem* **2005**, *280*, 38117-38120.
202. Mohaghegh, P.; Karow, J. K.; Brosh, R. M.; Bohr, V. A.; Hickson, I. D. The Bloom's and Werner's syndrome proteins are DNA structure-specific helicases, *Nucleic Acids Res.* **2001**, *29*, 2843-2849.
203. Kamath-Loeb, A.; Loeb, L. A.; Fry, M. The Werner syndrome protein is distinguished from the Bloom syndrome protein by its capacity to tightly bind diverse DNA structures, *PLoS One* **2012**, *7*, e30189.
204. Pike, A. C.; Shrestha, B.; Popuri, V.; Burgess-Brown, N.; Muzzolini, L.; Costantini, S.; Vindigni, A.; Gileadi, O. Structure of the human RECQ1 helicase reveals a putative strand-separation pin, *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 1039-1044.
205. Lucic, B.; Zhang, Y.; King, O.; Mendoza-Maldonado, R.; Berti, M.; Niesen, F. H.; Burgess-Brown, N. A.; Pike, A. C. W.; Cooper, C. D. O.; Gileadi, O.; Vindigni, A. A prominent beta-hairpin structure in the winged-helix domain of RECQ1 is required for DNA unwinding and oligomer formation, *Nucleic Acids Res.* **2011**, *39*, 1703-1717.
206. Gyimesi, M.; Harami, G. M.; Sarlos, K.; Hazai, E.; Bikadi, Z.; Kovacs, M. Complex activities of the human Bloom's syndrome helicase are encoded in a core region comprising the RecA and Zn-binding domains, *Nucleic Acids Res.* **2012**, *40*, 3952-3963.
207. Tanaka, A.; Choi, J.; Kim, S. K.; Majima, T. Interaction of G-quadruplex with RecA protein studied in bulk phase and at the single-molecule level, *J Phys Chem B* **2013**, *117*, 6711-6717.
208. Kuryavyi, V.; Cahoon, L. A.; Seifert, H. S.; Patel, D. J. RecA-binding pile G4 sequence essential for pilin antigenic variation forms monomeric and 5' end-stacked dimeric parallel G-quadruplexes, *Structure* **2012**, *20*, 2090-2102.
209. Samatanga, B.; Dominguez, C.; Jelesarov, I.; Allain, F. H. The high kinetic stability of a G-quadruplex limits hnRNP F qRRM3 binding to G-tract RNA, *Nucleic Acids Res.* **2013**, *41*, 2505-2516.
210. Padmanabhan, K.; Tulinsky, A. An ambiguous structure of a DNA 15-mer thrombin complex, *Acta Crystallogr D Biol Crystallogr* **1996**, *52*, 272-282.
211. Russo Krauss, I.; Merlino, A.; Giancola, C.; Randazzo, A.; Mazzarella, L.; Sica, F. Thrombin-aptamer recognition: a revealed ambiguity, *Nucleic Acids Res.* **2011**, *39*, 7858-7867.

212. Russo Krauss, I.; Merlino, A.; Randazzo, A.; Novellino, E.; Mazzarella, L.; Sica, F. High-resolution structures of two complexes between thrombin and thrombin-binding aptamer shed light on the role of cations in the aptamer inhibitory activity, *Nucleic Acids Res.* **2012**, *40*, 8119-8128.
213. Pica, A.; Russo Krauss, I.; Merlino, A.; Nagatoishi, S.; Sugimoto, N.; Sica, F. Dissecting the contribution of the thrombin exosite I in the recognition of thrombin binding aptamer, *FEBS J.* **2013**, *280*, 6581-6588.
214. Russo Krauss, I.; Pica, A.; Merlino, A.; Mazzarella, L.; Sica, F. Duplex-quadruplex motifs in a peculiar structural organization cooperatively contribute to thrombin binding of a DNA aptamer, *Acta Crystallogr D Biol Crystallogr* **2013**, *69*, 2403-2411.
215. Phan, A. T.; Kuryavyy, V.; Darnell, J. C.; Serganov, A.; Majumdar, A.; Ilin, S.; Raslin, T.; Polonskaia, A.; Chen, C.; Clain, D.; Darnell, R. B.; Patel, D. J. Structure-function studies of FMRP RGG peptide recognition of an RNA duplex-quadruplex junction, *Nat Struct Mol Biol* **2011**, *18*, 796-804.
216. Mashima, T.; Nishikawa, F.; Kamatari, Y. O.; Fujiwara, H.; Saimura, M.; Nagata, T.; Kodaki, T.; Nishikawa, S.; Kuwata, K.; Katahira, M. Anti-prion activity of an RNA aptamer and its structural basis, *Nucleic Acids Res.* **2013**, *41*, 1355-1362.
217. Arcovito, A.; Chiarella, S.; Della Longa, S.; Di Matteo, A.; Lo Sterzo, C.; Scaglione, G. L.; Federici, L. Synergic Role of Nucleophosmin Three-helix Bundle and a Flanking Unstructured Tail in the Interaction with G-quadruplex DNA, *J Biol Chem* **2014**, *289*, 21230-21241.
218. Zimmer, D. P.; Crothers, D. M. NMR of enzymatically synthesized uniformly ¹³C/¹⁵N-labeled DNA oligonucleotides, *Proc. Natl. Acad. Sci. U. S. A.* **1995**, *92*, 3091-3095.
219. www.proteins.sg; NTU/SBS Protein Production Platform.
220. Shevchuk, N. A.; Bryksin, A. V.; Nusinovich, Y. A.; Cabello, F. C.; Sutherland, M.; Ladisch, S. Construction of long DNA molecules using long PCR-based fusion of several fragments simultaneously, *Nucleic Acids Res.* **2004**, *32*, e19.
221. Plateau, P.; Gueron, M. Exchangeable proton NMR without base-line distortion, using new strong-pulse sequences, *J Am Chem Soc* **1982**, *104*, 7310-7311.
222. Phan, A. T.; Patel, D. J. A site-specific low-enrichment (¹⁵N),(¹³C) isotope-labeling approach to unambiguous NMR spectral assignments in nucleic acids, *J Am Chem Soc* **2002**, *124*, 1160-1161.
223. Phan, A. T. Long-range imino proton-¹³C J-couplings and the through-bond correlation of imino and non-exchangeable protons in unlabeled DNA, *J Biomol NMR* **2000**, *16*, 175-178.
224. Phan, A. T.; Gueron, M.; Leroy, J. L. Investigation of unusual DNA motifs, *Methods Enzymol* **2001**, *338*, 341-371.
225. Goddard, T. D.; Kneller, D. G. SPARKY 3.115, *SPARKY 3.115* **2007**.
226. Zimmerman, D. E.; Kulikowski, C. A.; Huang, Y.; Feng, W.; Tashiro, M.; Shimotakahara, S.; Chien, C.; Powers, R.; Montelione, G. T. Automated analysis of protein NMR assignments using methods from artificial intelligence, *J Mol Biol* **1997**, *269*, 592-610.
227. Shen, Y.; Delaglio, F.; Cornilescu, G.; Bax, A. TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts, *J Biomol NMR* **2009**, *44*, 213-223.
228. Gallagher, S. R. One-dimensional SDS gel electrophoresis of proteins, *Curr Protoc Protein Sci* **2012**, *Chapter 10*, Unit 10 11 11-44.
229. Schwieters, C. D.; Kuszewski, J. J.; Tjandra, N.; Clore, G. M. The Xplor-NIH NMR molecular structure determination package, *J Magn Reson* **2003**, *160*, 65-73.
230. Schrodinger, L. The PyMOL Molecular Graphics System, Version 1.3r1, *The PyMOL Molecular Graphics System, Version 1.3r1* **2010**.
231. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V. C., Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A.; C.02 ed.; Wallingford CT: 2004.
232. Case, D. A.; Cheatham, T. E., 3rd; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M., Jr.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber biomolecular simulation programs, *J. Comput. Chem.* **2005**, *26*, 1668-1688.
233. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water, *J Chem Phys* **1983**, *79*, 926-935.

234. Dupradeau, F. Y.; Pigache, A.; Zaffran, T.; Savineau, C.; Lelong, R.; Grivel, N.; Lelong, D.; Rosanski, W.; Cieplak, P. The R.E.D. tools: advances in RESP and ESP charge derivation and force field library building, *Phys Chem Chem Phys* **2010**, *12*, 7821-7839.
235. Guntert, P.; Mumenthaler, C.; Wuthrich, K. Torsion angle dynamics for NMR structure calculation with the new program DYANA, *J Mol Biol* **1997**, *273*, 283-298.
236. Do, N. Q.; Phan, A. T. Monomer-dimer equilibrium for the 5'-5' stacking of propeller-type parallel-stranded G-quadruplexes: NMR structural study, *Chemistry* **2012**, *18*, 14752-14759.
237. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A. J.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. In *Gaussian, Inc*; Revision E.01. ed. Pittsburgh, PA, 2003.
238. Hehre, W. J.; Radom, L.; Schleyer, P. v. R.; Pople, J. *Ab initio molecular orbital theory*; Wiley-Interscience: New York, NY, 1986.
239. Ditchfield, R. Self-consistent perturbation theory of diamagnetism, *Mol. Phys.* **1974**, *27*, 789-807.
240. Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A. Electrostatics of nanosystems: application to microtubules and the ribosome, *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 10037-10041.
241. Datta, B.; Schmitt, C.; Armitage, B. A. Formation of a PNA2-DNA2 hybrid quadruplex, *J Am Chem Soc* **2003**, *125*, 4111-4118.
242. Esposito, V.; Randazzo, A.; Piccialli, G.; Petraccone, L.; Giancola, C.; Mayol, L. Effects of an 8-bromodeoxyguanosine incorporation on the parallel quadruplex structure [d(TGGGT)]₄, *Org. Biomol. Chem.* **2004**, *2*, 313-318.
243. He, G. X.; Krawczyk, S. H.; Swaminathan, S.; Shea, R. G.; Dougherty, J. P.; Terhorst, T.; Law, V. S.; Griffin, L. C.; Coutre, S.; Bischofberger, N. N2- and C8-substituted oligodeoxynucleotides with enhanced thrombin inhibitory activity in vitro and in vivo, *J. Med. Chem.* **1998**, *41*, 2234-2242.
244. Koizumi, M.; Koga, R.; Hotoda, H.; Ohmine, T.; Furukawa, H.; Agatsuma, T.; Nishigaki, T.; Abe, K.; Kosaka, T.; Tsutsumi, S.; Sone, J.; Kaneko, M.; Kimura, S.; Shimada, K. Biologically active oligodeoxyribonucleotides. Part 11: The least phosphate-modification of quadruplex-forming hexadeoxyribonucleotide TGGGAG, bearing 3-and 5-end-modification, with anti-HIV-1 activity, *Bioorg. Med. Chem.* **1998**, *6*, 2469-2475.
245. Kuryavyi, V.; Kettani, A.; Wang, W.; Jones, R.; Patel, D. J. A diamond-shaped zipper-like DNA architecture containing triads sandwiched between mismatches and tetrads, *J. Mol. Biol.* **2000**, *295*, 455-469.
246. Li, Z.; Lech, C. J.; Phan, A. T. Sugar-modified G-quadruplexes: effects of LNA-, 2'F-RNA- and 2'F-ANA-guanosine chemistries on G-quadruplex structure and stability, *Nucleic Acids Res.* **2014**, *42*, 4068-4079.
247. Liu, X.; Kwan, I. C.; Wang, S.; Wu, G. G-quartet formation from an N2-modified guanosine derivative, *Org. Lett.* **2006**, *8*, 3685-3688.
248. Lopez de la Osa, J.; Gonzalez, C.; Gargallo, R.; Rueda, M.; Cubero, E.; Orozco, M.; Avino, A.; Eritja, R. Destabilization of quadruplex DNA by 8-aminoguanine, *Chembiochem* **2006**, *7*, 46-48.
249. Marathias, V. M.; Sawicki, M. J.; Bolton, P. H. 6-Thioguanine alters the structure and stability of duplex DNA and inhibits quadruplex DNA formation, *Nucleic Acids Res.* **1999**, *27*, 2860-2867.
250. Martin-Pintado, N.; Yahyaee-Anzahaee, M.; Deleavey, G. F.; Portella, G.; Orozco, M.; Damha, M. J.; Gonzalez, C. Dramatic effect of furanose C2' substitution on structure and stability: directing the folding of the human telomeric quadruplex with a single fluorine atom, *J Am Chem Soc* **2013**, *135*, 5344-5347.
251. Marusic, M.; Veedu, R. N.; Wengel, J.; Plavec, J. G-rich VEGF aptamer with locked and unlocked nucleic acid modifications exhibits a unique G-quadruplex fold, *Nucleic Acids Res.* **2013**, *41*, 9524-9536.
252. Mekmaysy, C. S.; Petraccone, L.; Garbett, N. C.; Ragazzon, P. A.; Gray, R.; Trent, J. O.; Chaires, J. B. Effect of O6-methylguanine on the stability of G-quadruplex DNA, *J Am Chem Soc* **2008**, *130*, 6710-6711.

253. Pasternak, A.; Hernandez, F. J.; Rasmussen, L. M.; Vester, B.; Wengel, J. Improved thrombin binding aptamer by incorporation of a single unlocked nucleic acid monomer, *Nucleic Acids Res.* **2011**, *39*, 1155-1164.
254. Peng, C. G.; Damha, M. J. G-quadruplex induced stabilization by 2'-deoxy-2'-fluoro-D-arabinonucleic acids (2'F-ANA), *Nucleic Acids Res.* **2007**, *35*, 4977-4988.
255. Randazzo, A.; Esposito, V.; Ohlenschlager, O.; Ramachandran, R.; Mayola, L. NMR solution structure of a parallel LNA quadruplex, *Nucleic Acids Res.* **2004**, *32*, 3083-3092.
256. Risitano, A.; Fox, K. R. Inosine substitutions demonstrate that intramolecular DNA quadruplexes adopt different conformations in the presence of sodium and potassium, *Bioorg. Med. Chem. Lett.* **2005**, *15*, 2047-2050.
257. Russo Krauss, I.; Parkinson, G. N.; Merlino, A.; Mattia, C. A.; Randazzo, A.; Novellino, E.; Mazzarella, L.; Sica, F. A regular thymine tetrad and a peculiar supramolecular assembly in the first crystal structure of an all-LNA G-quadruplex, *Acta Crystallogr D Biol Crystallogr* **2014**, *70*, 362-370.
258. Sagi, J. G-quadruplexes incorporating modified constituents: a review, *J. Biomol. Struct. Dyn.* **2014**, *32*, 477-511.
259. Suzuki, J.; Miyano-Kurosaki, N.; Kuwasaki, T.; Takeuchi, H.; Kawai, G.; Takaku, H. Inhibition of human immunodeficiency virus type 1 activity in vitro by a new self-stabilized oligonucleotide with guanosine-thymidine quadruplex motifs, *J. Virol.* **2002**, *76*, 3015-3022.
260. Virgilio, A.; Esposito, V.; Randazzo, A.; Mayol, L.; Galeone, A. 8-methyl-2'-deoxyguanosine incorporation into parallel DNA quadruplex structures, *Nucleic Acids Res.* **2005**, *33*, 6188-6195.
261. Zhou, J.; Abramov, M.; Liu, F.; Amrane, S.; Bourdoncle, A.; Herdewijn, P.; Mergny, J. L. Effects of six-membered carbohydrate rings on structure, stability, and kinetics of G-quadruplexes, *Chemistry* **2013**, *19*, 14719-14725.
262. Wilson, D. M., 3rd; Sofinowski, T. M.; McNeill, D. R. Repair mechanisms for oxidative DNA damage, *Front Biosci* **2003**, *8*, d963-981.
263. Benz, A.; Hartig, J. S. Redesigned tetrads with altered hydrogen bonding patterns enable programming of quadruplex topologies, *Chem. Commun.* **2008**, 4010-4012.
264. Cheng, X.; Kelso, C.; Hornak, V.; de los Santos, C.; Grollman, A. P.; Simmerling, C. Dynamic behavior of DNA base pairs containing 8-oxoguanine, *J Am Chem Soc* **2005**, *127*, 13906-13918.
265. Hsu, G. W.; Ober, M.; Carell, T.; Beese, L. S. Error-prone replication of oxidatively damaged DNA by a high-fidelity DNA polymerase, *Nature* **2004**, *431*, 217-221.
266. Kouchakdjian, M.; Bodepudi, V.; Shibutani, S.; Eisenberg, M.; Johnson, F.; Grollman, A. P.; Patel, D. J. NMR structural studies of the ionizing radiation adduct 7-hydro-8-oxodeoxyguanosine (8-oxo-7H-dG) opposite deoxyadenosine in a DNA duplex. 8-Oxo-7H-dG(syn).dA(anti) alignment at lesion site, *Biochemistry* **1991**, *30*, 1403-1412.
267. Lichtenberg, D.; Bergmann, F.; Neiman, Z. Tautomerism and ionisation processes in 6-methylthio-2-oxopurines, *J. Chem. Soc. Perkin 1* **1973**, *21*, 2445-2448.
268. Roy, K. B.; Frazier, J.; Miles, H. T. Interaction of poly(X) with poly(U): alkali metal ion specificity suggests a four-stranded helix, *Biopolymers* **1983**, *22*, 2021-2034.
269. Poznanski, J.; Kierdaszuk, B.; Shugar, D. Structural properties of the neutral and monoanionic forms of xanthosine, highly relevant to their substrate properties with various enzyme systems, *Nucleos. nucleot. nucl.* **2003**, *22*, 249-263.
270. Roy, K. B.; Miles, H. T. Tauto Merism and Ionization of Xanthosine, *Nucleos. Nucleot.* **1983**, *2*, 231-242.
271. Davis, J. T. G-quartets 40 years later: from 5'-GMP to molecular biology and supramolecular chemistry, *Angew. Chem. Int. Ed. Engl.* **2004**, *43*, 668-698.
272. Neidle, S.; Balasubramanian, S. *Quadruplex nucleic acids*; Royal Society of Chemistry: Cambridge, 2006.
273. Lipps, H. J.; Rhodes, D. G-quadruplex structures: in vivo evidence and function, *Trends Cell Biol.* **2009**, *19*, 414-422.
274. Balasubramanian, S.; Hurley, L. H.; Neidle, S. Targeting G-quadruplexes in gene promoters: a novel anticancer strategy?, *Nat. Rev. Drug Discov.* **2011**, *10*, 261-275.
275. Monchaud, D.; Teulade-Fichou, M. P. A hitchhiker's guide to G-quadruplex ligands, *Org. Biomol. Chem.* **2008**, *6*, 627-636.
276. Bates, P. J.; Laber, D. A.; Miller, D. M.; Thomas, S. D.; Trent, J. O. Discovery and development of the G-rich oligonucleotide AS1411 as a novel treatment for cancer, *Exp. Mol. Pathol.* **2009**, *86*, 151-164.
277. Collie, G. W.; Parkinson, G. N. The application of DNA and RNA G-quadruplexes to therapeutic medicines, *Chem. Soc. Rev.* **2011**, *40*, 5867-5892.
278. Metifiot, M.; Amrane, S.; Litvak, S.; Andreola, M. L. G-quadruplexes in viruses: function and potential therapeutic applications, *Nucleic Acids Res.* **2014**, *42*, 12352-12366.

279. Livshits, G. I.; Ghabboun, J.; Borovok, N.; Kotlyar, A. B.; Porath, D. Comparative electrostatic force microscopy of tetra- and intra-molecular G4-DNA, *Adv. Mat.* **2014**, *26*, 4981-4985.
280. Livshits, G. I.; Stern, A.; Rotem, D.; Borovok, N.; Eidelshtein, G.; Migliore, A.; Penzo, E.; Wind, S. J.; Di Felice, R.; Skourtis, S. S.; Cuevas, J. C.; Gurevich, L.; Kotlyar, A. B.; Porath, D. Long-range charge transport in single G-quadruplex DNA molecules, *Nat. Nanotechnol.* **2014**, *9*, 1040-1046.
281. Ge, B.; Huang, Y. C.; Sen, D.; Yu, H. Z. A robust electronic switch made of immobilized duplex/quadruplex DNA, *Angew. Chem. Int. Ed. Engl.* **2010**, *49*, 9965-9967.
282. Dao, N. T.; Haselsberger, R.; Michel-Beyerle, M. E.; Phan, A. T. Excimer formation by stacking G-quadruplex blocks, *Chemphyschem* **2013**, *14*, 2667-2671.
283. Lech, C. J.; Phan, A. T.; Michel-Beyerle, M. E.; Voityuk, A. A. Electron-hole transfer in G-quadruplexes with different tetrad stacking geometries: a combined QM and MD study, *J. Phys. Chem. B* **2013**, *117*, 9851-9856.
284. Lech, C. J.; Phan, A. T.; Michel-Beyerle, M. E.; Voityuk, A. A. Influence of base stacking geometry on the nature of excited states in G-quadruplexes: a time-dependent DFT study, *J. Phys. Chem. B* **2015**, *119*, 3697-3705.
285. Campbell, N. H.; Neidle, S. G-quadruplexes and metal ions, *Met. Ions Life Sci.* **2012**, *10*, 119-134.
286. Miyoshi, D.; Karimata, H.; Sugimoto, N. Hydration regulates thermodynamics of G-quadruplex formation under molecular crowding conditions, *J Am Chem Soc* **2006**, *128*, 7957-7963.
287. Heddi, B.; Phan, A. T. Structure of human telomeric DNA in crowded solution, *J Am Chem Soc* **2011**, *133*, 9824-9833.
288. Hansel, R.; Lohr, F.; Trantirek, L.; Dotsch, V. High-resolution insight into G-overhang architecture, *J Am Chem Soc* **2013**, *135*, 2816-2824.
289. Miller, M. C.; Buscaglia, R.; Chaires, J. B.; Lane, A. N.; Trent, J. O. Hydration is a major determinant of the G-quadruplex stability and conformation of the human telomere 3' sequence of d(AG3(TTAG3)3), *J Am Chem Soc* **2010**, *132*, 17105-17107.
290. Vorlickova, M.; Bednarova, K.; Kejnovska, I.; Kypr, J. Intramolecular and intermolecular guanine quadruplexes of DNA in aqueous salt and ethanol solutions, *Biopolymers* **2007**, *86*, 1-10.
291. Xue, Y.; Kan, Z. Y.; Wang, Q.; Yao, Y.; Liu, J.; Hao, Y. H.; Tan, Z. Human telomeric DNA forms parallel-stranded intramolecular G-quadruplex in K⁺ solution under molecular crowding condition, *J Am Chem Soc* **2007**, *129*, 11185-11191.
292. Lusvardi, S.; Murphy, C. T.; Roy, S.; Tanious, F. A.; Sacui, I.; Wilson, W. D.; Ly, D. H.; Armitage, B. A. Loop and backbone modifications of peptide nucleic acid improve g-quadruplex binding selectivity, *J Am Chem Soc* **2009**, *131*, 18415-18424.
293. Martino, L.; Virno, A.; Randazzo, A.; Virgilio, A.; Esposito, V.; Giancola, C.; Bucci, M.; Cirino, G.; Mayol, L. A new modified thrombin binding aptamer containing a 5'-5' inversion of polarity site, *Nucleic Acids Res.* **2006**, *34*, 6653-6662.
294. Dickerhoff, J.; Weisz, K. Flipping a G-Tetrad in a Unimolecular Quadruplex Without Affecting its Global Fold, *Angew. Chem. Int. Ed. Engl.* **2015**, *ASAP*.
295. Gray, D. M.; Wen, J. D.; Gray, C. W.; Repges, R.; Repges, C.; Raabe, G.; Fleischhauer, J. Measured and calculated CD spectra of G-quartets stacked with the same or opposite polarities, *Chirality* **2008**, *20*, 431-440.
296. Randazzo, A.; Spada, G. P.; da Silva, M. W. Circular dichroism of quadruplex structures, *Top Curr. Chem.* **2013**, *330*, 67-86.
297. Grollman, A. P.; Moriya, M. Mutagenesis by 8-oxoguanine: an enemy within, *Trends Genet* **1993**, *9*, 246-249.
298. Piazza, A.; Boule, J. B.; Lopes, J.; Mingo, K.; Largy, E.; Teulade-Fichou, M. P.; Nicolas, A. Genetic instability triggered by G-quadruplex interacting Phen-DC compounds in *Saccharomyces cerevisiae*, *Nucleic Acids Res.* **2010**, *38*, 4337-4348.
299. Rodriguez, R.; Miller, K. M.; Forment, J. V.; Bradshaw, C. R.; Nikan, M.; Britton, S.; Oelschlaegel, T.; Xhemalce, B.; Balasubramanian, S.; Jackson, S. P. Small-molecule-induced DNA damage identifies alternative DNA structures in human genes, *Nat Chem Biol* **2012**, *8*, 301-310.
300. Fry, M.; Loeb, L. A. Human werner syndrome DNA helicase unwinds tetrahelical structures of the fragile X syndrome repeat sequence d(CGG)_n, *J Biol Chem* **1999**, *274*, 12797-12802.
301. Hayashi, T.; Oshima, H.; Mashima, T.; Nagata, T.; Katahira, M.; Kinoshita, M. Binding of an RNA aptamer and a partial peptide of a prion protein: crucial importance of water entropy in molecular recognition, *Nucleic Acids Res.* **2014**, *42*, 6861-6875.
302. Cordin, O.; Banroques, J.; Tanner, N. K.; Linder, P. The DEAD-box protein family of RNA helicases, *Gene* **2006**, *367*, 17-37.
303. Meier, M.; Patel, T. R.; Booy, E. P.; Marushchak, O.; Okun, N.; Deo, S.; Howard, R.; McEleney, K.; Harding, S. E.; Stetefeld, J.; McKenna, S. A. Binding of G-quadruplexes to the N-terminal

- recognition domain of the RNA helicase associated with AU-rich element (RHAU), *J Biol Chem* **2013**, 288, 35014-35027.
304. Miyoshi, D.; Nakao, A.; Sugimoto, N. Molecular crowding regulates the structural switch of the DNA G-quadruplex, *Biochemistry* **2002**, 41, 15017-15024.
 305. Martadinata, H.; Phan, A. T. Formation of a stacked dimeric G-quadruplex containing bulges by the 5'-terminal region of human telomerase RNA (hTERC), *Biochemistry* **2014**, 53, 1595-1600.
 306. Phan, A. T.; Kuryavyi, V.; Gaw, H. Y.; Patel, D. J. Small-molecule interaction with a five-guanine-tract G-quadruplex structure from the human MYC promoter, *Nat Chem Biol* **2005**, 1, 167-173.
 307. Adrian, M.; Ang, D. J.; Lech, C. J.; Heddi, B.; Nicolas, A.; Phan, A. T. Structure and conformational dynamics of a stacked dimeric G-quadruplex formed by the human CEB1 minisatellite, *J Am Chem Soc* **2014**, 136, 6297-6305.
 308. Amrane, S.; Adrian, M.; Heddi, B.; Serero, A.; Nicolas, A.; Mergny, J. L.; Phan, A. T. Formation of pearl-necklace monomorphic G-quadruplexes in the human CEB25 minisatellite, *J Am Chem Soc* **2012**, 134, 5807-5816.
 309. Lim, K. W.; Alberti, P.; Guedin, A.; Lacroix, L.; Riou, J. F.; Royle, N. J.; Mergny, J. L.; Phan, A. T. Sequence variant (CTAGGG)_n in the human telomere favors a G-quadruplex structure containing a G.C.G.C tetrad, *Nucleic Acids Res.* **2009**, 37, 6239-6248.
 310. Drew, H. R.; Wing, R. M.; Takano, T.; Broka, C.; Tanaka, S.; Itakura, K.; Dickerson, R. E. Structure of a B-DNA dodecamer: conformation and dynamics, *Proc. Natl. Acad. Sci. U. S. A.* **1981**, 78, 2179-2183.
 311. Lim, K. W.; Phan, A. T. Structural basis of DNA quadruplex-duplex junction formation, *Angew. Chem. Int. Ed. Engl.* **2013**, 52, 8566-8569.
 312. Wu, H.; Finger, L. D.; Feigon, J. Structure determination of protein/RNA complexes by NMR, *Methods Enzymol* **2005**, 394, 525-545.
 313. Gilman, M. Preparation of cytoplasmic RNA from tissue culture cells, *Curr. Protoc. Mol. Biol.* **2002**, Chapter 4, Unit 4 1.
 314. Chung, W. J.; Heddi, B.; Schmitt, E.; Lim, K. W.; Mechulam, Y.; Phan, A. T. Structure of a left-handed DNA G-quadruplex, *Proc. Natl. Acad. Sci. U. S. A.* **2015**, 112, 2729-2733.
 315. Haeusler, A. R.; Donnelly, C. J.; Periz, G.; Simko, E. A.; Shaw, P. G.; Kim, M. S.; Maragakis, N. J.; Troncoso, J. C.; Pandey, A.; Sattler, R.; Rothstein, J. D.; Wang, J. C9orf72 nucleotide repeat structures initiate molecular cascades of disease, *Nature* **2014**, 507, 195-200.
 316. Halder, K.; Hartig, J. S. RNA quadruplexes, *Met. Ions Life Sci.* **2011**, 9, 125-139.
 317. German, J.; Sanz, M. M.; Ciocci, S.; Ye, T. Z.; Ellis, N. A. Syndrome-causing mutations of the BLM gene in persons in the Bloom's Syndrome Registry, *Hum. Mutat.* **2007**, 28, 743-753.
 318. Guo, R. B.; Rigolet, P.; Zargarian, L.; Fermandjian, S.; Xi, X. G. Structural and functional characterizations reveal the importance of a zinc binding domain in Bloom's syndrome helicase, *Nucleic Acids Res.* **2005**, 33, 3109-3124.
 319. Bachrati, C. Z.; Hickson, I. D. RecQ helicases: suppressors of tumorigenesis and premature aging, *Biochem. J.* **2003**, 374, 577-606.
 320. Brosh, R. M.; Bohr, V. A. Human premature aging, DNA repair and RecQ helicases, *Nucleic Acids Res.* **2007**, 35, 7527-7544.
 321. Karow, J. K.; Wu, L.; Hickson, I. D. RecQ family helicases: roles in cancer and aging, *Curr. Opin. Genet. Dev.* **2000**, 10, 32-38.
 322. Yang, Z.; Zhang, L.; Zhang, Y.; Zhang, T.; Feng, Y.; Lu, X.; Lan, W.; Wang, J.; Wu, H.; Cao, C.; Wang, X. Highly efficient production of soluble proteins from insoluble inclusion bodies by a two-step-denaturing and refolding method, *PLoS One* **2011**, 6, e22981.
 323. Smith, D. B.; Johnson, K. S. Single-step purification of polypeptides expressed in Escherichia coli as fusions with glutathione S-transferase, *Gene* **1988**, 67, 31-40.
 324. Kaplan, W.; Husler, P.; Klump, H.; Erhardt, J.; Sluis-Cremer, N.; Dirr, H. Conformational stability of pGEX-expressed Schistosoma japonicum glutathione S-transferase: a detoxification enzyme and fusion-protein affinity tag, *Protein Sci.* **1997**, 6, 399-406.
 325. LaVallie, E. R.; DiBlasio, E. A.; Kovacic, S.; Grant, K. L.; Schendel, P. F.; McCoy, J. M. A thioredoxin gene fusion expression system that circumvents inclusion body formation in the E. coli cytoplasm, *Biotechnology* **1993**, 11, 187-193.
 326. Yasukawa, T.; Kanei-Ishii, C.; Maekawa, T.; Fujimoto, J.; Yamamoto, T.; Ishii, S. Increase of solubility of foreign proteins in Escherichia coli by coproduction of the bacterial thioredoxin, *J Biol Chem* **1995**, 270, 25328-25331.
 327. Brondyk, W. H. Selecting an appropriate method for expressing a recombinant protein, *Methods Enzymol* **2009**, 463, 131-147.
 328. Kost, T. A.; Condreay, J. P.; Jarvis, D. L. Baculovirus as versatile vectors for protein expression in insect and mammalian cells, *Nat Biotechnol* **2005**, 23, 567-575.

329. Monteiro, F.; Carinhas, N.; Carrondo, M. J.; Bernal, V.; Alves, P. M. Toward system-level understanding of baculovirus-host cell interactions: from molecular fundamental studies to large-scale proteomics approaches, *Front Microbiol* **2012**, *3*, 391.
330. Frueh, D. P. Practical aspects of NMR signal assignment in larger and challenging proteins, *Prog Nucl Magn Reson Spectrosc* **2014**, *78*, 47-75.
331. Kim, S. Y.; Hakoshima, T.; Kitano, K. Structure of the RecQ C-terminal domain of human Bloom syndrome protein, *Sci. Rep.* **2013**, *3*, 3294.
332. Breukels, V.; Konijnenberg, A.; Nabuurs, S. M.; Doreleijers, J. F.; Kovalevskaya, N. V.; Vuister, G. W. Overview on the use of NMR to examine protein structure, *Curr Protoc Protein Sci* **2011**, Chapter 17, Unit 17 15.
333. Park, C. J.; Ko, J.; Ryu, K. S.; Choi, B. S. Solution structure of the RecQ C-terminal domain of human Bloom syndrome protein, *J Biomol NMR* **2014**, *58*, 141-147.
334. Cheong, V. V.; Heddi, B.; Phan, A. T. NMR study of RQC domain of BLM protein, *J Biomol Struct Dyn* **2013**, *31*, 109-109.
335. Sam, M. D.; Clubb, R. T. Preparation and optimization of protein-DNA complexes suitable for detailed NMR studies, *Methods Mol Biol* **2012**, *831*, 219-232.
336. Chow, S. A.; Brown, P. O. Substrate features important for recognition and catalysis by human immunodeficiency virus type 1 integrase identified by using novel DNA substrates, *J. Virol.* **1994**, *68*, 3896-3907.
337. Fromme, J. C.; Banerjee, A.; Huang, S. J.; Verdine, G. L. Structural basis for removal of adenine mispaired with 8-oxoguanine by MutY adenine DNA glycosylase, *Nature* **2004**, *427*, 652-656.
338. Routledge, M. N.; Wink, D. A.; Keefer, L. K.; Dipple, A. Mutations induced by saturated aqueous nitric oxide in the pSP189 supF gene in human Ad293 and E. coli MBM7070 cells, *Carcinogenesis* **1993**, *14*, 1251-1254.
339. Jackson, S. P.; Bartek, J. The DNA-damage response in human biology and disease, *Nature* **2009**, *461*, 1071-1078.
340. Bruner, S. D.; Norman, D. P.; Verdine, G. L. Structural basis for recognition and repair of the endogenous mutagen 8-oxoguanine in DNA, *Nature* **2000**, *403*, 859-866.
341. Kulikowska, E.; Kierdaszuk, B.; Shugar, D. Xanthine, xanthosine and its nucleotides: solution structures of neutral and ionic forms, and relevance to substrate properties in various enzyme systems and metabolic pathways, *Acta Biochim. Pol.* **2004**, *51*, 493-531.
342. Schermerhorn, K. M.; Delaney, S. A chemical and kinetic perspective on base excision repair of DNA, *Acc. Chem. Res.* **2014**, *47*, 1238-1246.
343. Zhou, J.; Liu, M.; Fleming, A. M.; Burrows, C. J.; Wallace, S. S. Neil3 and NEIL1 DNA glycosylases remove oxidative damages from quadruplex DNA and exhibit preferences for lesions in the telomeric sequence context, *J Biol Chem* **2013**, *288*, 27263-27272.
344. Henderson, A.; Wu, Y.; Huang, Y. C.; Chavez, E. A.; Platt, J.; Johnson, F. B.; Brosh, R. M., Jr.; Sen, D.; Lansdorp, P. M. Detection of G-quadruplex DNA in mammalian cells, *Nucleic Acids Res.* **2014**, *42*, 860-869.
345. Persinger, R. L.; Melamed, R.; Bepalov, I.; Wallace, S.; Taatjes, D. J.; Janssen-Heininger, Y. Imaging techniques used for the detection of 8-oxoguanine adducts and DNA repair proteins in cells and tissues, *Exp. Gerontol.* **2001**, *36*, 1483-1494.
346. Pervushin, K.; Riek, R.; Wider, G.; Wuthrich, K. Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution, *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94*, 12366-12371.
347. LeMaster, D. M. Deuterium labelling in NMR structural analysis of larger proteins, *Q Rev Biophys* **1990**, *23*, 133-174.
348. Grzesiek, S.; Anglister, J.; Ren, H.; Bax, A. Carbon-13 line narrowing by deuterium decoupling in deuterium/carbon-13/nitrogen-15 enriched proteins. Application to triple resonance 4D J connectivity of sequential amides, *J Am Chem Soc* **1993**, *115*, 4369-4370.
349. Rehm, T.; Huber, R.; Holak, T. A. Application of NMR in structural proteomics: screening for proteins amenable to structural analysis, *Structure* **2002**, *10*, 1613-1618.
350. Kim, Y. M.; Choi, B. S. Structure and function of the regulatory HRDC domain from human Bloom syndrome protein, *Nucleic Acids Res.* **2010**, *38*, 7764-7777.
351. Chatterjee, S.; Zagelbaum, J.; Savitsky, P.; Sturzenegger, A.; Huttner, D.; Jancsak, P.; Hickson, I. D.; Gileadi, O.; Rothenberg, E. Mechanistic insight into the interaction of BLM helicase with intra-strand G-quadruplex structures, *Nat Commun* **2014**, *5*, 5556.
352. Savitsky, P.; Bray, J.; Cooper, C. D.; Marsden, B. D.; Mahajan, P.; Burgess-Brown, N. A.; Gileadi, O. High-throughput production of human proteins for crystallization: the SGC experience, *J Struct Biol* **2010**, *172*, 3-13.