Biochemical Characterization of G-Quadruplex Recognition in the PITX1 mRNA by the Specificity Domain of the Helicase RHAU

By

Emmanuel O. Ariyo

A thesis submitted to Faculty of Graduate Studies of

The University of Manitoba

In partial fulfillment of the requirements of the degree of

DOCTOR OF PHILOSOPHY

Department of Chemistry

Faculty of Science

University of Manitoba

Winnipeg, Manitoba, Canada

ABSTRACT

Biologically, G-quadruplexes (G4s) are single-stranded structures that fold back on themselves, and are formed by DNA or RNA molecules in which one guanine base from each chain associate via cyclic Hoogsteen hydrogen bonding to form planar G-tetrad. Two or more of these tetrads can hydrophobically stack on top of one another to form G4 which are further stabilized by the presence of a mandatory monovalent cation centered between the planes. RNA Helicase Associated with AU-rich element (RHAU) is a member of the ATP-dependent DExH/D family of the RNA helicases that can bind and resolve G4 structures. To gain insight into the structural basis of G4 recognition by RHAU, a series of G4s were characterized using biophysical techniques including UV-Visible spectroscopy, electrophoretic mobility shift assays, dynamic light scattering, circular dichroism, small angle x-ray scattering and nuclear magnetic resonance spectroscopy. The mRNA for the Pituitary homeobox 1 (PITX1) protein was chosen as a model system to investigate G4 structures as it possesses three distinct G4 forming sequences in its 3′- untranslated region (UTR) that interact with RHAU (Q1: $PITX1_{1371\text{-}1400}$, Q2: $PITX1_{1901\text{-}1930}$, and Q3: $PITX1_{2044\text{-}2079}$). First, Q2 was studied and demonstrated to adopt a parallel G4 orientation that can interact with the N-terminal domain of RHAU via its G-tetrad face. Interestingly, the DNA counterpart of Q2 adopted a different hybrid-type G4 structure with mixed parallel and antiparallel strands which was unable to interact with RHAU. Next, Q3 from PITX1 mRNA was studied as it contains long loops of 6 nucleotides to investigate the potential impact of interrupting loops on both the G4 structure and interaction with proteins/small molecules. Scrambling of only specific loop sequences impacted the global fold of Q3 while retaining a parallel G4 conformation. Disruption of the global fold through loop scrambling was shown to impact binding affinity of RHAU and a small molecule ligand that selectively interacts with the G-

tetrad faces of parallel G4s. Taken together, the data in this thesis provides a biochemical and structural starting point for understanding the recognition of RNA G4s by binding partners.

ACKNOWLEDGEMENTS

The Almighty God must first be appreciated for His infinite care, impalpable guide, incredulous blessing and unremitting mercy for the successful completion of this program. I am grateful to my graduate program supervisor, Dr. Sean McKenna for his advice, kindness, guidance, creative ideas and sound initiation to recondite research methodologies during my project. His professional initiatives converted my mistakes into lessons, pressure into productivity and skills into strength, working and learning in the research environment of the McKenna lab resulted in several publications that pave the way to various scholarships and awards.

The past and present members of McKenna lab, including Dr. Evan Booy, Dr. Trushar Patel, Dr. Soumya Deo, Dr. Edis Dzananovic, were great to work and collaborate with. I also acknowledge Ryan Howard for his assistance during the early stage PhD program in McKenna lab.

My appreciation equally goes to members of my advisory committee, Dr. Joe O'Neil, Dr. Jorg Stetefeld and Dr. Peter Pelka for their advice and guidance during my program. I sincerely appreciate the time and effort they invested in reviewing my thesis. Special thanks to Dr. Peter Pelka for his kind advise in acquiring more research background knowledge.

Great appreciation goes to the funding organizations including, Manitoba health Research Council (MHRC), Canadian Institutes of Health Research (CIHR) and Faculty of Graduate Studies (FGS), university of Manitoba.

Finally, I would like to appreciate Heri Kavira for her support, love and encouragement during my PhD program at the university of Manitoba.

DEDICATION

This project is dedicated to Almighty God for His mercy and infinite love to me throughout my academic endeavors. I also dedicate the project to my lovely daughters, Mowafoluwa Happiness Ariyo and Ewaoluwa Favour Ariyo for making me a proud father.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AEBSF	4-(2-Aminoethyl)-benzenesulfonyl fluoride
ASO	antisense oligonucleotides
Bcl2	B-cell lymphoma 2
BLM	bloom syndrome protein
BRCA1	breast cancer type 1 susceptibility protein
CD	circular dichroism
CNBP	cellular nucleic-acid-binding protein
CTR	c-terminal region
DLS	dynamic light scattering
D_{max}	maximum particle dimension
Dna2	DNA replication helicase/nuclease 2
dsRNA	double-stranded RNA
EMSA	electrophoretic mobility shift assays
ESR	estrogen receptor α
FANCJ	Fanconi anaemia group J protein
FGF-2	fibroblast growth factor-2
FMR2	fragile X mental retardation 2
G4	G-quadruplex
G4R1	G-quadruplex resolvase 1
HCD	Helicase core domain
hnRNP	heterogeneous nuclear ribonucleoprotein
HSQC	heteronuclear single quantum coherence

hTR	human telomerase RNA
HuR	human antigen R
IGF II	insulin growth factor II
IRES	internal ribosome entry site
MAZ	myc-associated zinc-finger
MCM	minichromosome maintenance protein
NF90	nuclear factor 90
NMM	*N*-methyl mesoporphyrin IX
NMR	nuclear magnetic resonance spectroscopy
NRAS	neuroblastoma rat sarcoma
NS3	nonstructural protein 3
NSD	normalized spatial discrepancy
NTR	N-terminal region
PARP-1	Poly [ADP-ribose] polymerase 1
PIPER	3,4,9,10-Perylenetetracarboxylic Diimide
PITX1	paired-like homeodomain 1
POT1	protection of telomeres 1
PQS	putative quadruplex forming sequences
P(r)	pair distance distribution function
RAP1	repressor activation protein 1
RASAL 1	RAS protein activator like 1
r_H	hydrodynamic radii
RHAU	RNA helicase associated with AU-rich element

RPA	replication protein A
RSM	RHAU-specific motif
SAXS	small angle X-ray scattering
SEC	size exclusion chromatography
Sgs1	small growth suppressor 1
SRSF	serine/arginine-rich splicing factor
TBE	Tris-borate buffer with ethylenediaminetetraacetic acid
TDS	Thermal difference spectra
TEBPβ	telomere end-binding protein-beta
TERT	telomerase reverse transcriptase
TGFβ2	transforming growth factor β2
TIN2	TRF1 interacting protein 2
Topo I	Topoisomerase I
TPP1	tripeptidyl peptidase
TRF-1	telomeric repeat-binding factor 1
TRF-2	telomeric repeat-binding factor 2
T_m	melting temperature
TMPyP4	tetra-(N-methyl-4-pyridyl) porphyrin
UP1	unwinding protein 1
UTR	untranslated region
VEGF	vascular endothelial growth factor
WRN	Werner's syndrome protein
Zic-1	Zinc finger of the cerebellum 1

CHAPTER 1

INTRODUCTION

1.1 Nucleic acids

Nucleic acids are biomolecules important to all living organisms. Although the initial discovery of nucleic acids was made by Friedrich Miescher in 1869 (1) while he was performing an experiment on the chemical composition of leukocytes, however, modern experimental studies have focused on biological and medical research which have provided insights into genomic, forensic, pharmaceuticals and biotechnology studies (2, 3).

The basic components of nucleic acids are the nucleotides, comprised of a nitrogenous base, a pentose sugar (deoxyribose in DNA and ribose in RNA), and at least one phosphate group. The nucleotide bases are the purines, guanine (G) and adenine (A) and the pyrimidines, cytosine (C), thymine (T) and uracil (U) (4). Nucleosides are made from purines and pyrimidines having a sugar bonded to a ring nitrogen and to distinguish atoms of the sugar from those of the heterocyclic bases, a prime is attached to the numbers (e.g., 2′ or 3′) in the sugar. In ribonucleosides, the sugar moiety is *D*-ribose, while it is 2′ -Deoxy-*D*-ribose in deoxyribonuleosides. The sugar is bonded to the heterocyclic base through a β-N-glycosidic bond, to N-1 of a pyrimidine or N-9 of a purine. Steric hindrance by the base restricts rotation about the β-N-glycosidic bond of nucleosides and nucleotides, resulting in specific *syn* and *anti* conformations; the *anti* conformation is more common. Nucleotides are nucleosides containing phosphate groups bonded to hydroxyl groups of the sugar via esterification, with the 5′- and 3′- prefixes representing the linkage of the phosphates groups from the 5′- and 3′-hydroxyl groups of the sugar respectively (5).

The 5′-phosphate group of one nucleotide can be linked to the 3′-hydroxyl group of the pentose sugar of another nucleotide via esterification reaction to form a phosphodiester bond, in which the pentose moieties are connected by a 3′ → 5′ phosphodiester bond that forms the backbone of DNA and RNA. The monomers in DNA are joined to form the polymer by 3′, 5′-phosphodiester bridges resulting into a single strand possessing a polarity with one end denoted as the 5′-hydroxyl or phosphate terminus, and the other

as the 3′-phosphate or hydroxyl terminus (5).

DNA molecules are usually double-stranded, with genetic information contained in the template strand copied during nucleic acid synthesis while the opposite strand matches the RNA transcript that encodes for protein (5). Double-stranded nucleic acids are made up of complementary base pairing from antiparallel strands known as Watson-Crick base pairing, resulting from hydrogen bonds between the bases (adenine base pairs with thymine in DNA or uracil in RNA and guanine base pairs with cytosine). RNA molecules, on the other hand, are single-stranded (6) but in the presence of proper complementary base pairs, single stranded RNA can fold back on itself like a hairpin to form double-stranded structures with altered helical features relative to DNA because of the additional 2'-hydroxyl group on its ribose sugar. Therefore, RNA often forms highly ordered three-dimensional structures arising from several small tracts of intramolecular interactions resulting from Watson-Crick base pairing and other noncanonical base-pairing.

RNA molecules usually have characteristic secondary and tertiary structures that are specific to their biological function. RNA molecules play various roles in cellular contexts and are classified as noncoding and coding RNAs. Noncoding RNA comes in different varieties with the most prominent being transfer RNA (tRNA) and ribosomal RNA (rRNA). Others include; small nuclear RNA (snRNA), found in the nucleus and tightly bound to proteins in complexes known as snRNPs (small nuclear ribonucleo proteins,) and the most abundant of these molecules are the U1, U2, U5, and U4/U6 particles that are involved in splicing of pre-mRNA to give mature mRNA. Small Nucleolar RNA (snoRNA) are found in nucleolar extracts and function in the processing of rRNA, which often results in the methylation and pseudouridylation of immature rRNA. MicroRNAs (miRNAs) are small regulatory RNAs, about 22 to 26 nucleotides long that function in gene regulation while small interfering RNAs (siRNAs) are molecules containing 21 to 25 base pairs, that also play roles in the regulation of gene expression. Ribosomal RNA

(rRNA) contributes to the assembly and function of ribosomes and transfer RNA (tRNA) is an adapter molecule involved in the translation of information from RNA into a specific sequence of amino acids that form the protein. Messenger RNA (mRNA) is the coding RNA that serves as template for the synthesis of protein.

1.1.1 Nucleic acid folding

Early crystallographic studies of DNA unveiled a surprising diversity of related structures, including the A, B, and Z-forms of DNA that differ significantly in their helical features (**Fig. 1.1**). Under physiological conditions (7), the predominant structure is B-form which adopts a right-handed anti-parallel helix with 10.5 base pairs per turn, a narrow minor groove and a wide major groove (8, 9). In this structure all of the nitrogenous bases adopt the *anti* conformation, with the Watson-Crick hydrogen bonding face pointing away from the deoxyribose sugar. A-form DNA is comprised of 11 base pairs per turn with a broader and shallower minor groove, narrower and deeper major groove than that of the B-form, and is formed *in vitro* under dehydrating conditions. Z-form DNA was accidentally discovered during attempts to crystallize DNA duplex with alternating purine-pyrimidine tracts, and adopts a left-handed antiparallel helix, with the bases assuming a *syn* conformation (10).

A characteristic feature of nucleic acid helices is the sugar puckering that occurs as a result of the furanose rings in nucleotides, twisted out of plane in order to reduce steric clashes between atoms on the sugar ring. Puckering is described by the displacement of carbons-2′ and 3′ from the median plane of C1′-O4′-C4′ in the sugar. The *endo* face of the sugar is on the same side as C-5′ and the nucleotide base while the *exo* face is on the opposite side of the base (**Fig. 1.3 C**) (11). A C2′-endo conformation is achieved when the *endo* displacement of C-2′ is greater than the *exo* displacement of C-3′ (characteristic of B-form DNA) and C3′-endo conformation is formed when the *endo* displacement of C-3′ is greater than the *exo*

displacement of C-2′ (prevalent in A-form DNA). A key feature that distinguishes RNA from DNA is the presence of a hydroxyl group at the 2′ position of its ribose sugar, resulting in a predominant C-3′ *endo* (A-form) conformation of biological RNA due to the formation of an intramolecular hydrogen bond from O-2′ in one RNA residue to O-4′ in the next (11). The presence of the 2′-OH group in RNA molecules may lead to chemical attack on the adjacent phosphodiester bond and consequently cleavage of the backbone, making RNA less chemically stable than DNA (12).

Highly ordered structures formed from single-stranded DNA and RNA have been reported to be involved in many important biological processes including replication, transcription, translation, recombination and repair (13, 14). Other highly ordered DNA and RNA structures including DNA triplexes, formed when pyrimidine or purine nucleotide bases occupy the major groove of the DNA double helix to form Hoogsteen base pairs with purines of the Watson-Crick base pairs, and the focus of this thesis G-quadruplexes, formed from guanine rich DNA or RNA sequences have been reported to play significant roles in myriads of cellular processes including natural gene regulation (15).

1.2 Quadruplex Nucleic Acids

Biologically, G-quadruplexes (G4) are single-stranded structures that fold back on themselves, and are formed by DNA or RNA in which one guanine base from each chain associates via cyclic Hoogsteen (16) hydrogen bonding to form a planar G-tetrad. Two or more such tetrads hydrophobically stack on top of each other to form the G4 and are stabilized by the presence of a required monovalent cation (typically K^+) in the center between the planes (17). It was shown as early as 1910 that high concentrations of guanylic acid (GMP) in aqueous solution can form gels (18). Gellert *et al.* (19), demonstrated by X-ray fiber diffraction experiments the structure of a G4 in which each guanine residue in the G-tetrad acts both as an acceptor and a donor of two hydrogen bonds. The topologies of G4s are based on a number of factors

including, glycosidic conformation (anti/syn), monovalent cation (K^+/Na^+), the number of nucleic acid strands in the sequence (intermolecular, intramolecular), strand orientation (parallel/antiparallel) and the number of G-tetrads involved in stacking to form the stable structure (20, 21).

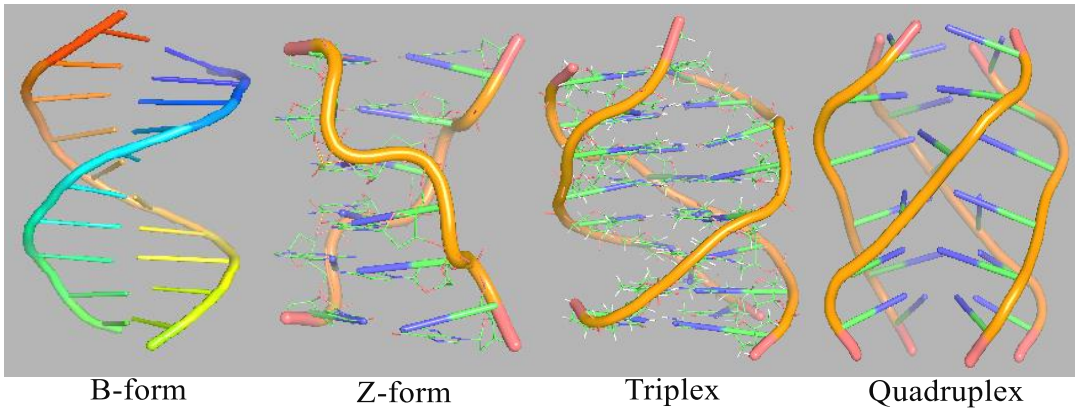

Fig. 1.1. B, Z, triplex and Quadruplex nucleic acid structures. B-form (PDB ID #: 3r86) adopts a right-handed anti-parallel helix with 10.5 base pairs per turn, a narrow minor groove and a wide major groove assuming *anti* conformation; Z-form (PDB ID #: 3fqb) adopts a left-handed anti-parallel helix, with the bases assuming a *syn* conformation; triplex (PDB ID #: 1d3x) nucleic acid is formed when purine or pyrimidine bases occupy the major groove of DNA double helix and form base pairing with purines of the Watson-Crick basepairs; G4 (PDB ID #: 1rau) is formed by DNA or RNA in which one guanine base from each chain associates via cyclic Hoogsteen hydrogen bonds.

Computational studies conducted by the use of consensus sequence ($G_{3+}N_{1-7}G_{3+}N_{1-7}G_{3+}N_{1-7}G_{3+}$) to search for putative quadruplex forming sequences (PQS) have shown that human genome may contain more than 300 000 sequences having the possibility of forming G4s PQS (22). A recent deep sequencing experiment of DNA G4 structures in the human genome revealed 716,310 distinct PQS (23), with 451,646 not predicted by computational methods (22, 24, 25). The localization of (PQS) is non-random; PQS have been reported to co-localize with functional regions of the genome and moreover, they are highly conserved

within different species (26), with the highest conservation in mammalian species and a decrease in non-mammals and other lower organisms (27). Large quantities of PQS occur at telomeres consisting of 5 to 10 000 bp of tandem TTAGGG repeats in humans. Other regions including gene promoters, intron/exon borders, specific immunoglobulin switch targets have been reported to contain high levels of PQS (28) and about 90% of human DNA replication sites have also been shown to contain PQS (**Fig. 1.2**) (29, 30). For RNA, previous findings have suggested that in approximately 3000 human genes, the 5′-untranslated region (UTR) of the encoded mRNA contain PQS which suggest a significant role in translation regulation (31, 32). Collectively, the presence of PQS in key genomic regions suggest that they are capable of forming stable structures that play key regulatory roles in cells.

1.2.1 History of the G-quadruplex

In 1910, it was shown that concentrated solutions of guanylic acid can form a gel and in the 1960s, X-ray diffraction technique was used to analyze the gel which revealed the arrangement of planar tetramers of guanines linked by Hoogsteen hydrogen bonds (19, 33). After about ten years following the discovery of these planar tetramers (named G-tetrad) (**Fig. 1.3 B**) (34, 35), it was shown that several repeats of guanine bases in polyguanylic acid (in RNA) and polydeoxyguanylic acid (in DNA) can form four-stranded structures known as G4 as a result of stacked G-tetrads (36-38). In a study conducted by Walter Gilbert and Dipankar Sen in 1988, the first endogenous nucleic acid sequences that form G4 structures were discovered in the immunoglobulin switch regions, and this led to the suggestion that these structures could facilitate the alignment and recombination of chromosomes during meiosis (39).

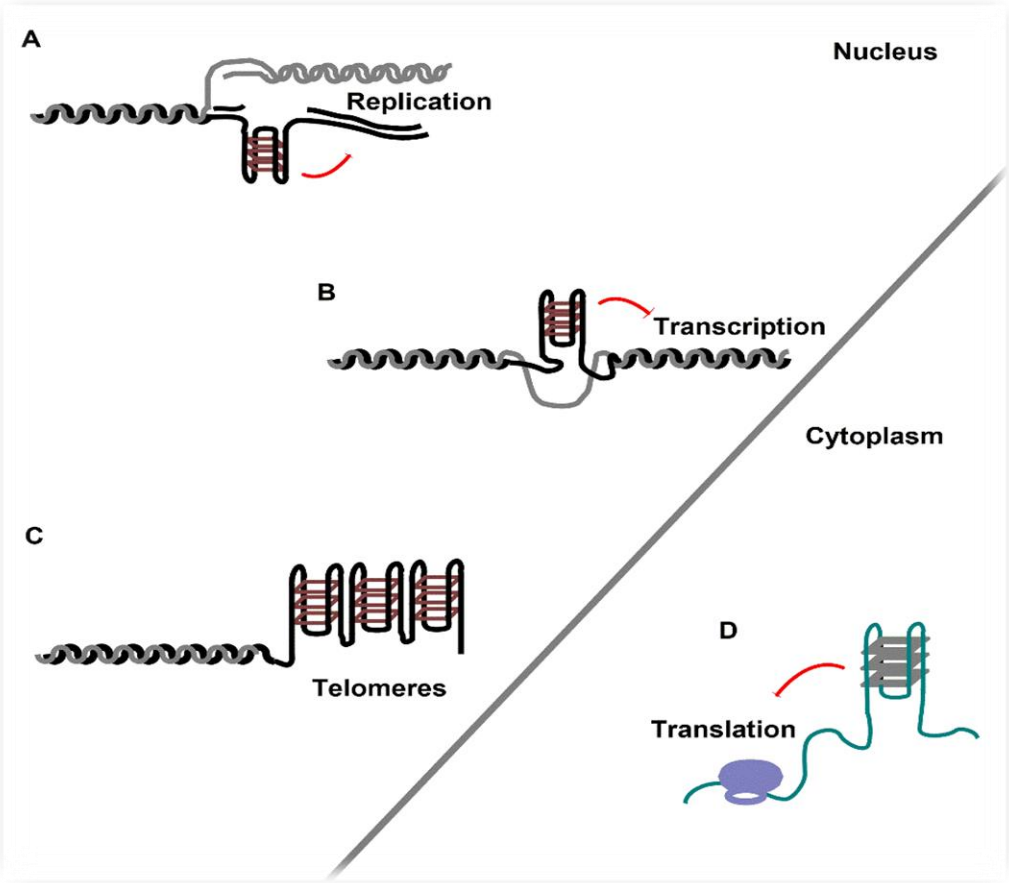

Fig. 1.2. Established locations of G4 structures. (**A**) Replication sites **(B)** transcription sites **(C)** telomeres, single stranded G-rich overhangs **(D)** translation sites (Red-bars represent obstructions to transcription, replication and translation, although G4 structures have been shown in some cases to upregulate translation) (40)

G4 structures have also been shown to be formed at telomeres consisting of many tandem repeats of guanine-rich motifs, including hexameric repeats of TTAGGG/CCCTAA in vertebrates (41). *Tetrahymena* telomeric DNA was proposed in 1987 by Elizabeth Blackburn's laboratory to form a double G-G base paired hairpin (42) and research publications from Thomas Cech and Aaron Klug also suggest the formation of G4 structures by telomeric sequences of *Oxytricha* and *Tetrahymena* (43, 44). In 1998, a G4 forming sequence was found to form G4 structure at the control region of proto-oncogene c-Myc, suggesting a role in the regulation of transcription (45). Although the study of G4 structures has continued to increase over

the years, nonetheless, several challenges have been associated with the assessment of their biological roles *in vivo*. Recent studies have shown that these structures can be visualized throughout the mammalian genome by the use of specific antibodies (46, 47) and evidence of the biological roles mediated by G4s have begun to be unraveled through the investigation of proteins that interact with them (48-50).

1.2.2 G-tetrad

A G-tetrad is a planar association of four guanine molecules (**Fig. 1.3 A**) held together by four central hydrogen bonds between H^1 of nitrogen and O^6 of carbonyl group and also by four external hydrogen bonds between H^2 of amine and N^7 (**Fig. 1.3 B**). G-tetrads have two faces referred to as a "head" (+) and a "tail" (-), where the "head" is defined as the face showing a clockwise orientation of the internal hydrogen bonds from the donor (NH) to the acceptor (C=O) group, and vice versa for the "tail."

The ribose and deoxyribose sugars of the nucleoside can adopt several conformations of which the two most favorable are the C2′-*endo* (mostly found in DNA G4s) and the C3′-*endo* (mostly found in RNA G4s) (**Fig. 1.3 C**) (51). There is a covalent linkage between the guanine bases and the sugar through a glycosidic bond that can exhibit two major different torsion angles namely, *syn* and *anti* (**Fig. 1.3 D**). All G4 structures have four grooves namely; a small, a wide and two medium grooves which are defined as the spaces bounded by the phosphodiester backbones (**Fig. 1.3 D**) (17). Parallel G4s contain all guanine glycosidic angles in an *anti*-conformation, while anti-parallel G4s were shown to have both *syn* and *anti* guanine conformations with at least one of the four strands oriented in an anti-parallel manner to the others (17, 52).

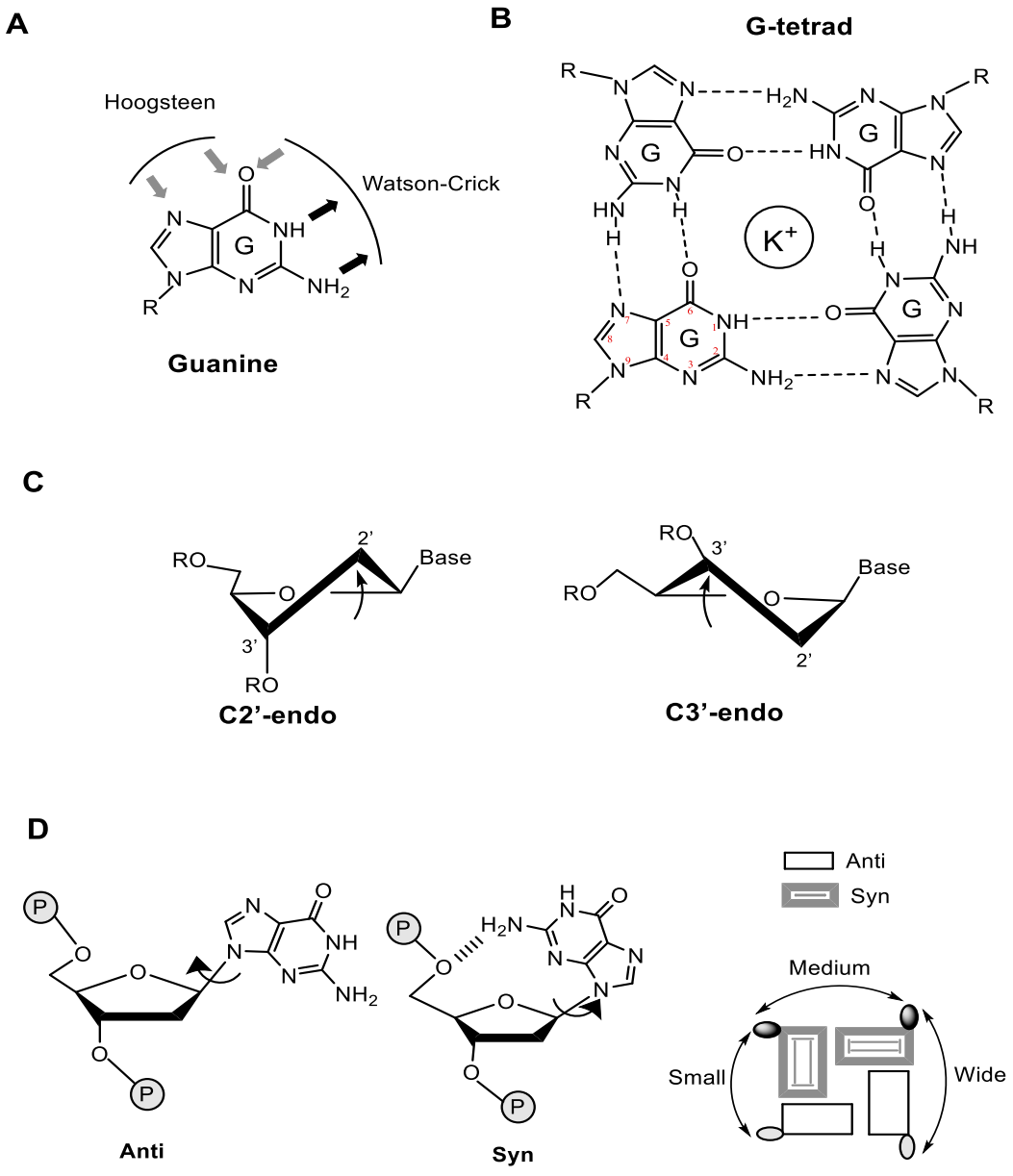

Fig. 1.3. Guanine and G-tetrad (A) Guanine molecule showing two hydrogen bond faces (Watson-Crick and Hoogsteen) involved in G-tetrad formation. *Arrows* indicate H-bond donors (*in black*) and acceptors (*in grey*). **(B)** Structure of G-quartet showing the hydrogen bonds. **(C)** The two most favorable sugar conformations of guanine quartets: C2′-*endo* or C3′-*endo*. **(D)** Torsion angles of guanine glycosidic bond (*syn* and *anti*) that determine groove dimension of G-quartet: wide, medium and small.

1.2.3 The role of cations in the stability of G4 structures

The coordination of cations by the closely spaced carbonyl oxygen atoms of a G-tetrad was postulated long before the first high-resolution structure of a G4 was determined. G4 structures are stabilized by the interaction of a monovalent cation with the negatively charged carbonyl oxygen atoms located at the centre of the G-tetrad stacks (53) and these monovalent cations stabilize the structure in the order of $K^+ > NH_4^+ > Rb^+ > Na^+ > Cs^+ > Li^+$ (17, 54). K^+ and Na^+ cations have been characterized extensively, largely due to their physiological importance and the ability to stabilize G4 structures. A number of other cations have also been shown to promote the formation of G4 and these include monovalent cations Rb^+, Cs^+, (55, 56), NH_4^+ (57, 58), Tl^+ (59) and divalent cations such as Sr^{2+} (56, 60), Ba^{2+} (61), and Pb^{2+} (57, 62) with the order of G4 structure stability as follow: $Sr^{2+} > Ba^{2+} > Ca^{2+} > Mg^{2+}$ (61). Interestingly, Ca^{2+} and Mg^{2+} ions do not enhance the formation of G4 structures in the absence of other cations, however, low concentrations of divalent cations initially stabilize G4s while increasing concentrations destabilize them (57). Cations interact with G4 structures in two ways namely, cation-dipole interactions through 2x4 central oxygen atoms (inner ions) from two stacked G-tetrads which reduces the electronic repulsion (unique for G-quadruples) and electrostatic interactions with the grooves and the phosphate backbone (outer ions) common within most nucleic acid structures.

1.3 Structural polymorphisms of G4 nucleic acids

G4s can be formed from one, two or four separate strands of DNA (or RNA) and show a wide variety of topologies as a result of different factors, including the number of oligonucleotide strands, composition of the nucleotide sequence, the orientation of strands, the nature and composition of loops, glycosidic bond angles and type of cations present. The formation of G4 is largely based on two or more core G-tetrads that are stacked on top of each other with loops connecting guanine strands (**Fig. 1.4**). The

simplest unimolecular DNA or RNA G4 structure (intramolecular) can be depicted as: $G_x W_a G_x W_b G_x W_c G_x$, where $_X$ represents the number of guanine residues in each short G-tract directly involved in G-tetrad formation and W_a, W_b and W_c represent any combination of nucleotide residues (including guanine) that are involved in the formation of loops (17). In this simple form there is an assumption that all the G-tracts existing within a G4 sequence are identical, and while this may be true for telomeric sequences in some vertebrates, it may vary in the case of non-telomeric genomic sequences, as also observed in some lower eukaryotic telomeric sequences (**Table 1.1**) (17).

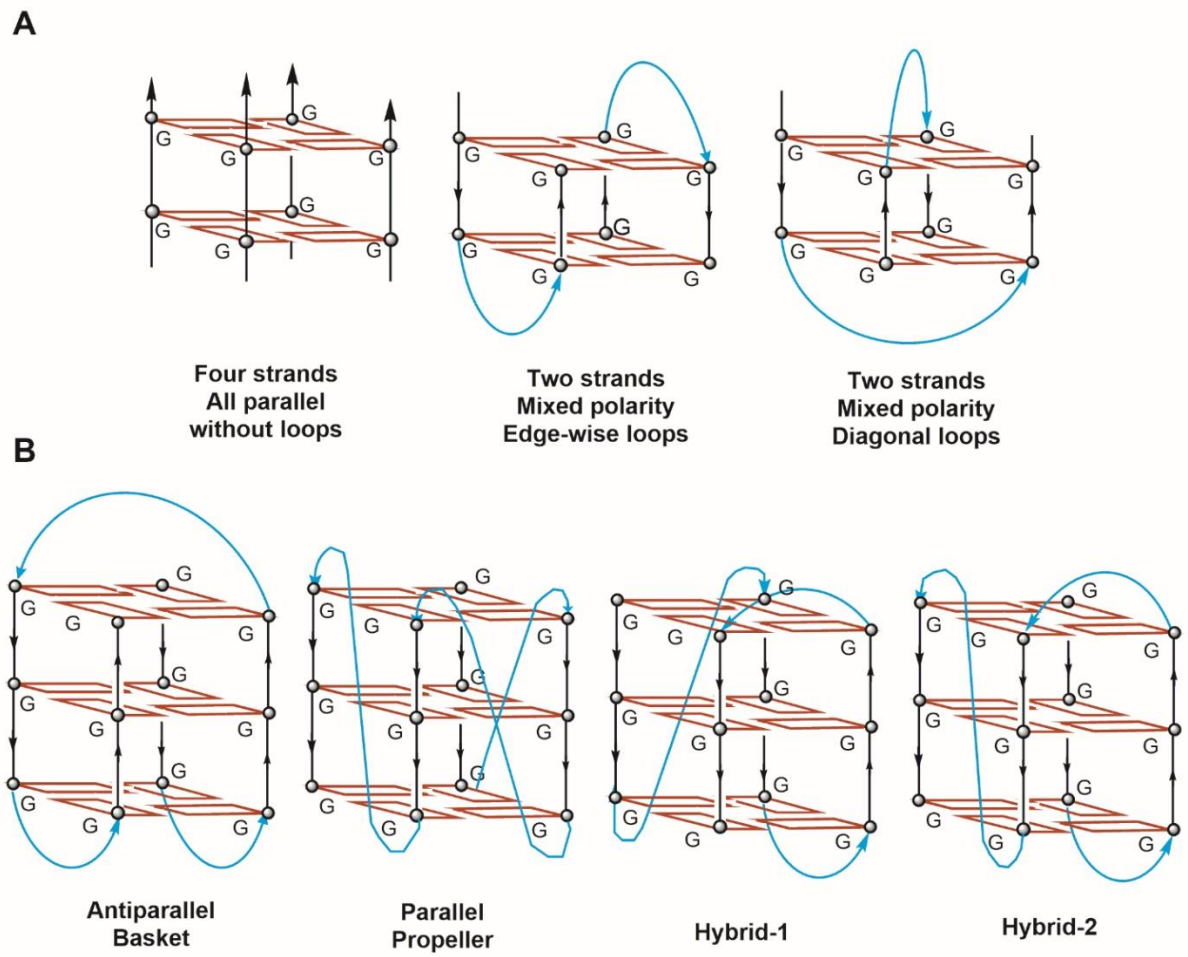

Fig. 1.4. Schematics of G4 topologies. (A) Intermolecular (tetramolecular and bimolecular) quadruplex structures. **(B)** Unimolecular (intramolecular) quadruplex structures. Loops are colored cyan and G-quartets colored red (17, 63)

G-tracts may be of varying lengths depending on the specific quadruplex sequence being considered and in some cases, guanine residues in the longer tracts may be involved in the formation of loops. The classification of G4 structures formed by a single or double strands is based on the polarities of the strands and also on the position of the loops joining the guanine strands (**Fig. 1.4**).

Table 1.1. Some known telomeric DNA sequences. (sequences involved in telomeric G4 formation may vary from vertebrates to lower eukaryotes)

Group	Organism	Telomeric repeat
Vertebrates	Human, mouse, *Xenopus*	TTAGGG
Filamentous fungi	*Neurospora crassa*	TTAGGG
Slime moulds	*Physarum, Didymium,*	TTAGGG
Kinetoplastid protozoa	*Trypanosoma, Crithidia*	TTAGGG
Ciliate protozoa	*Tetrahymena, Glaucoma* *Paramecium Oxytricha* *Stylonychia, Euplotes*	TTGGGG TTGGG(T/G) TTTTGGGG
Apicomplexan protozoa	*Plasmodium*	TTAGGG(T/C)
Higher plants	*Arabidopsis thaliana*	TTTAGGG
Green algae	*Chlamydomonas*	TTTTAGGG
Insects	*Bombyx mori*	TTAGG
Round worms	*Ascaris lumbricoides*	TTAGGC
Fission yeasts	*Schizosaccharomyces pombe*	TTAC(A)(C)G(1–8)
Budding yeasts	*Saccharomyces cerevisiae*	TGTGGGTGTGGTG

1.3.1 RNA and DNA G4s

RNA G4s have been reported to form structures that are thermodynamically more stable, less hydrated and more compact than their DNA G4 counterparts (64, 65). The basic difference between RNA and DNA G4s is the presence of a ribose sugar rather than a deoxyribose sugar. The 2′ hydroxyl group in the ribose sugar gives rise to specific intramolecular interactions within the structure which results in an increased stability of RNA G4 over DNA G4. RNA G4 predominantly adopts the parallel conformation (**Fig. 1.4 B**; parallel propeller) due to the role played by its 2′ hydroxyl group, exerting a constraint on the quadruplex topology, and thereby preventing the *syn*-conformation which is a requirement for the antiparallel topology (**Fig. 1.3 D**). Tang *et al.* demonstrated the effect of the 2′ hydroxyl group in G4 structures by examining the *syn/anti* preferences of individual guanine nucleotides on the overall fold of G4s using site-specific substitution of deoxyribonuleosides with ribonucleosides. Their findings showed that with ribonucleoside substitution, there was orientation restriction of the base about the glycosidic bond to the *anti* conformation, where C3′ *endo* sugar puckering is imparted which favors the parallel conformation of the RNA G4 (66). It has been shown that several factors including loop size, length of the G-tracts and the flanking nucleotide bases contribute to the folding patterns of G-rich sequences to form stable G4 structures (67-70). Monovalent and divalent cations have been shown as well to induce different G4 structures with different stabilities (71-74). The effect of absence or presence of monovalent cations, such as potassium and sodium ions on the thermal stability of G4 structures have been investigated. RNA G4s were shown to exhibit much higher stabilities than corresponding DNA G4s (64). A methodical investigation of chemically modified G4s by thymine/uracil substitution resulted in a moderate or strong stabilization of the structure and modification at the ribose 2′-position showed loss of hydrogen bond capacity that strongly affected the G4 conformation (75).

1.3.2 Loop Conformation in Parallel and Anti-parallel G4 Structures

Intermolecular quadruplexes may be categorized as tetramolecular (formed by the association of four oligonucleotide strands) or bimolecular (formed by the association of two oligonucleotide strands). Although the formation of G4 structures may be promoted under the conditions that revealed single-stranded regions such as during replication, transcription, or within the overhangs of telomeres (48), however, among these single-stranded regions, intramolecular G4s may be more likely formed *in vivo* than intermolecular G4s (48, 76). Intramolecular G4s can be formed by only one molecule of nucleic acid sequence with various loop conformations connecting the G-tracts. The nucleotide sequences existing between the guanines that form G-tetrads can form loops and these play a significant role in determining the nature of folding and stability of G4 structures, therefore, loops of different lengths and sequences can stabilize or disrupt G4s (70, 77). Hydrogen bonding and base stacking occur between loops giving rise to intraloop interactions (78), and recent studies on G4s with non-nucleoside linkers in place of loops revealed preferential formation of parallel G4s which do not depend on the length of such linkers (69, 78). A combination of circular dichroism, UV melting, molecular modeling and simulation techniques have been used to show that a parallel-stranded intramolecular G4 structure is the only fold that is possible when three single residue loops are present. Both parallel and anti-parallel conformations are possible when single thymine loops are combined with longer length loops, or when all loops are longer than two residues (70). Tippana *et al*. (79) showed a diminishing population of parallel G4 folding by replacing thymine residues in the loops with adenine residues indicating loop sequence dependence of G4 folding in some DNA G4s which may be attributed to a different steric hindrance imposed by adenine. G4 folding governed by the loop length and sequence may impart a conformational change that determines binding interactions of small molecules and proteins (79). According to an investigation conducted by Guedin *et al*. (2010), 80 different sequences containing four tracts of three guanines with loops of varied lengths were examined to determine

their effect on G4s; an inverse correlation between total loop length and melting temperature (T_m) was reported in K^+: A decrease of 2º C in T_m was recorded for each nucleotide base added, corresponding to the impact of loop length on the stability of G4 structures (80).

The length of loops in RNA G4s has also been shown to have an inverse relationship with their stability, and this was demonstrated by Pandey *et al*. (81), where RNA oligonucleotides sequences containing four tracts of two guanines with uracil loops of varied lengths ranging from one to eleven were examined; a reduction in T_m from 60ºC to 14ºC was observed for the G4 RNA containing the longest loops, while both composition and length of the loops affect the thermal stabilities of naturally occurring G4 RNAs sequences examined in the same study (81). Parallel strands connected adjacently to each other from the 3'-end of the first G-tract to the 5′-end of the second G-tract gives rise to the formation of strand-reversal loops also referred to as propeller loops (**Fig. 1.4 B**, parallel propeller). Mixed parallel/antiparallel topologies (hybrid-1 and hybrid-2) have been reported in crystal structures (**Fig. 1.4 B**, hybrid 1 & 2) (63) and in solution by NMR methods (73, 82) for quadruplexes formed from human telomeric DNA sequences and more recently in a number of non-telomeric quadruplexes. In these structures, three G-tracts were oriented in the same direction while the fourth strand was in the opposite direction (**Fig. 1.4 B**, hybrid 1 & 2). Anti-parallel G4 structures are formed when two of the strands are antiparallel to each other and the majority of this topology are found in bimolecular G4s structures but has also been shown to exist in unimolecular G4 structures published to date. Lateral or edge-wise loops (**Fig. 1.4 A**) may connect adjacent G-strands in an anti-parallel manner as shown in the structures of both asymmetric quadruplexes observed in solution for the sequence d($TG_4T_2G_4T$) by NMR and for a bimolecular quadruplex structure formed by the sequence d($GGGCT_4GGGC$) (83, 84). Anti-parallel diagonal loops can also connect opposite G-strands as previously shown in the structure formed by the *Oxytricha nova* telomeric sequence d($G_4T_4G_4$) (85-87) where the direction of the loops alternate between parallel and antiparallel G-strands.

1.4 Biological roles of G4 nucleic acids

Biologically, genomic DNA exist mainly as double stranded helical structure packaged into chromatin, and the opportunity of G4 structure formation occurs during DNA replication, transcription and repair when the DNA exists as single stranded by the disruption of Watson-Crick base pairing, which in turn allows the formation of an alternative Hoogsteen hydrogen bonding that are found in G4s (53, 88). In contrast to the Watson-Crick base pairing theory, it has been shown that Hoogsteen hydrogen bonding can be formed in canonical double stranded DNA (89), indicating that G4 structures may be formed without necessarily completely melting double stranded DNA.

The initial studies of G4s were based on their roles at the end of telomeres, but recent studies have shifted to other important regions of the genome including the promoters of proto-oncogenes (39, 45, 90, 91), immunoglobulin heavy chain switch regions (39), mutational hotspots, maintenance of chromosomal integrity, regulation of replication, transcription and recombination processes (35, 48). Recent findings have shown that approximately 90% of human DNA sites that are frequently involved in active DNA replication contain G4 forming motifs (29, 30, 92) and in about 3000 human genes explored, PQS were shown to be present in the 5′-UTR and 3′-UTR of the encoded mRNAs that may play roles in translation regulation (**Fig. 1.2 D**) (31, 32).

The single stranded nature of RNA allows it to adopt higher ordered structure compared to its DNA counterpart (93) and the roles played by RNA G4s have also been studied extensively in recent years. In 1991, Kim *et al.* showed that a 19-base oligonucleotide derived from the 5S rRNA of *Escherichia coli*, containing short runs of Gs could tetramerize (form quadruplex) (94). RNA G4s have been shown to exist in the 3′ untranslated region (UTR) of insulin growth factor II (IGF II) mRNA downstream of the endonucleolytic cleavage site (95). The presence and regulatory functions of RNA G4s at telomeric ends, long non-coding RNAs, introns and coding regions of mRNAs have also been investigated (21, 50, 96-99).

A recent publication by Biffi *et al.* (46) demonstrated for the first time the visualization of RNA G4s within the cytoplasm of human cells using a specific G4 antibody, giving an insight into the selective trapping of endogenous RNA G4s by specific G4 ligands, a strategy based on the use of small molecules to differentiate RNA G4s from DNA G4s in cells. Visualizing and quantitating DNA G4s (100), and RNA G4s in human cells provide new research strategies to investigate the functions and dynamics of G4 structures in a biological context.

1.4.1 G4s in translation and transcription

G4 structures have been shown to exist in the 5′-UTR of cellular mRNA and generally play a role in translational repression. Kumari *et al*. used a cell-free translation reporter assay to demonstrate how a thermodynamically stable G4 in the 5′-UTR of human neuroblastoma rat sarcoma (NRAS) mRNA impedes gene expression *in vitro* (98). In another study, a dual luciferase reporter assay was used to investigate the inhibitory role displayed by an RNA G4 structure found at the 5′-UTR of human Zic-1 (Zinc finger of the cerebellum 1) in eukaryotic cells (101). According to genome-wide computational analysis, there are approximately 4,141 5′-UTR G4 motifs in the 5′-UTR of mRNAs (102) and more functional roles of G4 secondary structure in the repression of gene expression at translational level have been confirmed experimentally in several genes including, Exon C of human and bovine estrogen receptor α (ESR) (103, 104), B-cell lymphoma 2 (Bcl2) (105), fibroblast growth factor-2 (FGF-2) (106), vascular endothelial growth factor (VEGF) (107), telomeric repeat-binding factor 2 (TRF-2) (108), and transforming growth factor β2 (TGFβ2) (109).

The addition of a G4 stabilizing ligand, or mutations of the G4s at the promoter region of c-MYC oncogene have been demonstrated to impact gene expression *in vivo* (**Fig. 1.5 A**) (45, 110). RNA capable of forming stable G4 within the 5′-UTR of the human NRAS proto-oncogene has also been demonstrated

to regulate translation *in vitro* (**Fig. 1.5 B**) (98). Bonnal *et al.* (106) showed that both a G4 and two stem-loop structures found within fibroblast growth factor-2 (FGF-2) internal ribosome entry site (IRES) were needed to initiate translation and similarly, the requirement of G4 for cap-independent translation initiation in human vascular endothelial growth factor (VEGF) IRES has been shown (107). In addition, G4s in the 3′-UTR of some mRNAs have been reported to play roles in alternative polyadenylation and shortening of transcripts (32).

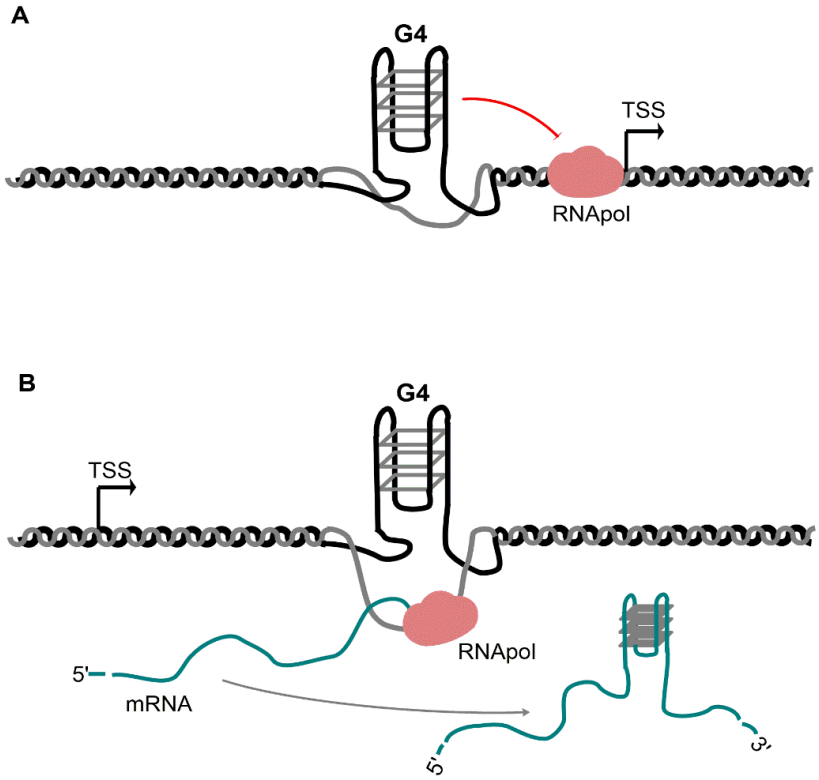

Fig. 1.5. Schematics of the existence of G4 structures in the gene promoters or at the 5′-UTR of mRNAs could **(A)** mediate inhibition of transcription initiation by RNA polymerase **(B)** regulate translation (40)

Genome-wide gene expression studies in yeast and human cells revealed drastic changes in the expression levels of several genes following the addition of G4 stabilizing ligand tetra-(N-methyl-4-pyridyl) porphyrin (TMPyP4) and these genes were enriched in PQS (111, 112). In another study conducted using a single

chain antibody specific for G4 stabilization, a change in the expression level of genes rich in PQS was observed, which was not confined to promoters alone but also was found on PQS at the ends of the genes, suggesting the involvement of G4 structures in transcription regulation (102, 113).

1.4.2 G4s at the telomeres

Telomeres are nucleoprotein complexes that cap and protect the ends of chromosomes against damage during cell division, nonhomologous end-joining of chromosomes and attack from nucleases (114-116) (**Fig. 1.6 A**). The telomeric regions of mammalian chromosomes is made up of tandem repeats of the sequence $d(TTAGGG)_n$, with a G-rich strand between 100 and 200 nucleotides that forms the 3′ overhang known as the G-tail. The length of telomeres differs between species, with 5-10 kb in humans, and up to 150 kb in mice (117-123). Telomeric DNA has been shown to interact with various proteins, including protection of telomeres 1 (POT1), telomere repeat binding factor 1 and 2 (TRF1 and TRF2), tripeptidyl peptidase (TPP1), TRF1 interacting protein 2 (TIN2) and repressor activation protein 1 (RAP1) (119, 124-126). The localization of other proteins such as Bloom syndrome (BLM), Werner syndrome (WRN) and RecQ helicases at the telomeres have been demonstrated to unwind G4 structures *in vitro* and *in vivo*, and are needed for the maintenance of its integrity (127-131). The structure and stability of telomeres is related to cancer and aging (116, 132), as in a normal cell every cell replication gives rise to a loss of 50-200 nucleotides at the telomere (133-135) and at a critical minimum length the cell undergoes apoptosis (133). Telomerase is a ribonucleoprotein (RNP) complex (136-138) that maintains the length of telomere in human stem cells, reproductive cells and cancer cells (139, 140) using an RNA molecule, human telomerase RNA (hTR) as a template to add TTAGGG nucleotide repeats to the end of telomeres, catalyzed by the telomerase reverse transcriptase (TERT). The maintenance of telomere stability is key to long-term proliferation of tumors (141-143), and the immortalization of human cells have been shown to be associated with

reactivation of telomerase, suggesting an escape from cell death (144). As a result of the activation of telomerase reverse transcriptase (TERT) enzyme in immortalized cells, their telomeres do not shorten after replication, an enabled uncontrolled cell division (145).

The formation of G4 structures coupled with their unwinding at the tandem repeats of TTAGGG sequence were reported to play significant roles in chromosomal end maintenance (146) and direct evidence for the existence of G4 structures at the telomeres *in vivo* was first reported in studies involving specific antibodies directed against these structures (147, 148). An intermolecular and antiparallel G4 structure was formed at the telomere of *Stylonychia* and the structure was resolved in the course of DNA replication, indicating that G4 structures may be implicated in capping the ends of telomeres (**Fig. 1.6 A**).

The mechanism involving G4 folding and unfolding to promote the synthesis of telomere ends by telomerase has been studied and in separate investigations conducted by Paeschke and Lipps *et al.* (149, 150) that revealed how the recruitment of telomerase by telomere end-binding protein-beta (TEBPβ) facilitates the unwinding of G4 DNA in ciliates, and also how the telomerase-associated RecQ protein-like helicase resolves telomeric G4 structures during replication. The inhibition of telomerase by G-tetrad DNA structures in human telomeres as reported by Zahler *et al.* was based on the interaction of G4 stabilizing ligands which caused a reduction in telomere length as a result of impaired telomere metabolism (**Fig. 1.6 B**) (151-154).

1.5 G4 stabilizing ligands

The stabilization of G4 structures by small molecule ligands including porphyrins has been investigated (155-157); findings showed that ligands that can interact selectively with G4s and stabilize them, but with no affinity for double-stranded DNA molecules, may provide potential usefulness in pharmaceuticals. G4 stabilizing ligands may interact with G4 structures within the

telomeres, causing a disruption that may inhibit the growth of tumor cells by triggering cellular responses that may lead to apoptosis or senescence (158).

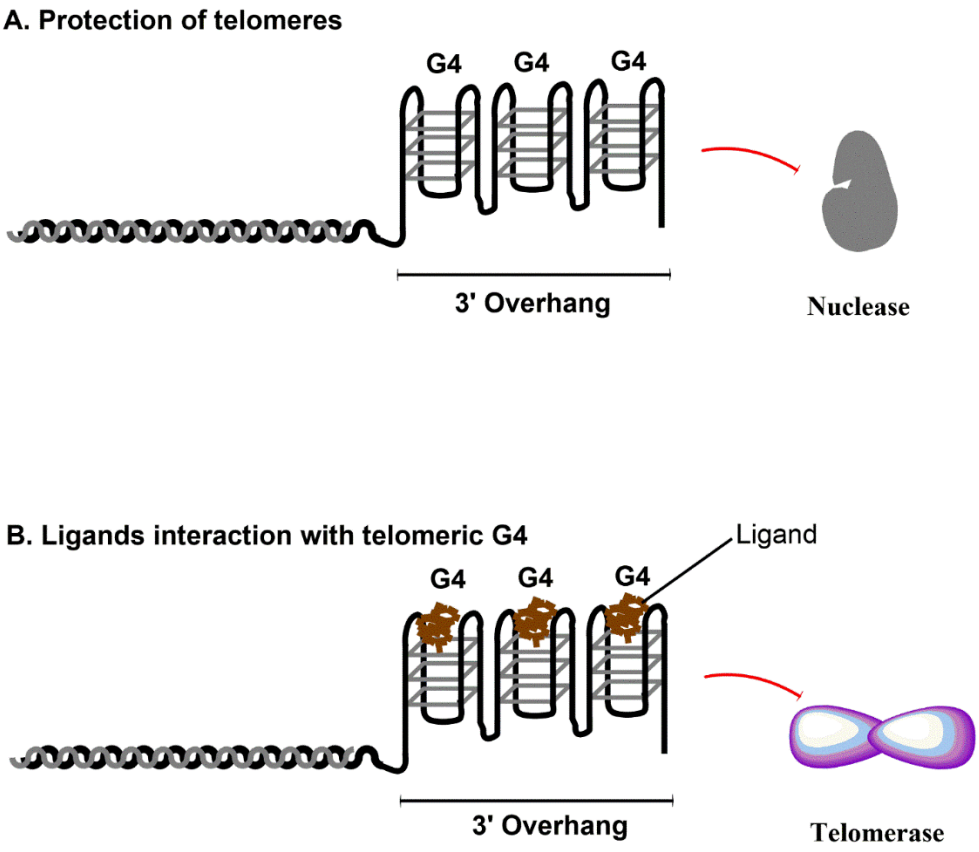

Fig. 1.6. Schematics of G4 structures at the telomeres. (A) The G-rich overhang sequence of human telomere can adopt intramolecular G-quadruplex structures that may regulate the activity of telomerase enzyme, or protect the end from endonucleases. **(B)** A G-quadruplex binding ligand interacts with G-quadruplex structures to inhibit the synthesis of telomere repeats by telomerase, leading to the shortening of telomeres (40)

G4 stabilizing ligands such as *N*-Methyl mesoporphyrin IX (NMM) have been shown to be exceptionally selective for parallel G4s relative to antiparallel G4s and duplex DNA, therefore, NMM has been used in biology and chemistry research to target G4 structures (159). Zahler *et al*. (151) showed that a folded G4 structure within telomeric DNA when used as a primer for *Oxytricha nova* telomerase *in vitro* affected the

extent of telomere elongation, suggesting a negative regulatory effect *in vivo*. Florence *et al*. (160) showed inhibition of telomerase activity *in vitro* when ethidium derivatives were used to target and stabilize telomeric G4s. The initial design of G4 binding ligands was based on acridine compounds (**Fig. 1.7**) which seemed to be valuable for the recognition of G4 as long as its substitution pattern and the protonation ability of the middle ring nitrogen was fully optimized (155). A derivative of anthraquinone was the first small molecule ligand developed to interact with DNA G4 and inhibits the activity of telomerase enzyme (152) and other compounds that bind G4 structures have since been developed (161-165). G4 interacting ligands are designed to exploit two features; π-stacking and electrostatic interactions. Four different classes of ligands have been characterized based on their cationic nature: (a) cationic ligands with locally protonated amine appendages, (b) *N*-methylation of an aromatic component, (c) incorporation of a metal centre and (d) non-cationic ligands.

Cationic ligands containing locally protonated amine appendages have been reported to interact by hydrophobic π-stacking between its flat aromatic core and the accessible guanine residues of the G-tetrad, coupled with ionic interaction between the protonated sidechains of the ligand and the quadruplex-grooves (166). Several examples of this group of ligands have been investigated (**Fig. 1.7**), including 9-[4(N,N-dimethylamino)phenylamino]-3,6-bis(3-pyrrolodinopropionamido) acridine (BRACO-19) (167, 168), 3,4,9,10-Perylenetetracarboxylic Diimide (PIPER) (169) and quinacridines (MMQ) (170, 171).

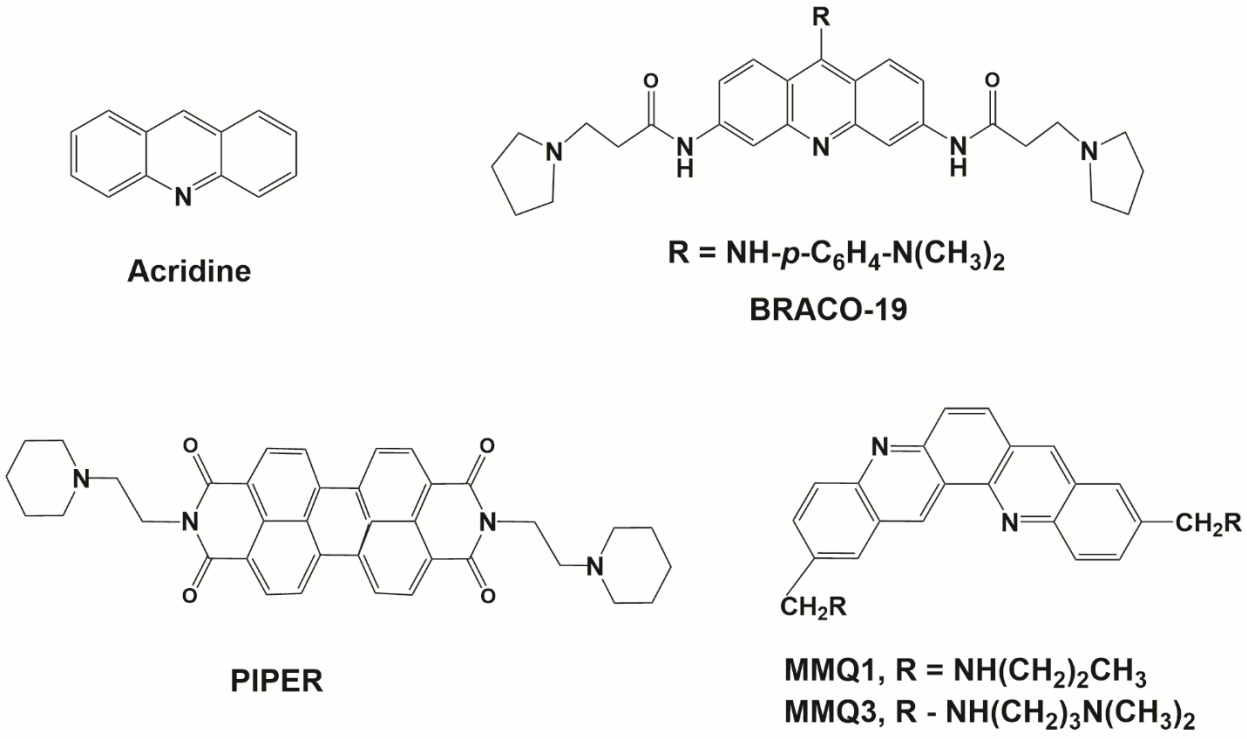

Fig. 1.7. Examples of cationic G4 interacting ligands based upon locally protonated amine appendages (167-169)

The second category of G4 stabilizing molecules involve the use of *N*-methylated ligands that contain quaternized moiety on the aromatic rings (*i.e.* a cation consisting of a central positively charged nitrogen atom with four substituents), with the advantage of water solubility without the need for cationic side-chains and an increase in the π-stacking capability. 5, 10, 15, 20-tetra-(N-methyl-4-pyridyl) porphine (TMPyP4) is an example of this class that has been studied extensively to have a high affinity for G4s and interacts by intercalation between adjacent G-tetrads or onto the exposed G-tetrads (172-175). Others include 5, 10, 15, 20-tetra-[4-hydroxyl-3-(trimethylammonium) methyl-phenyl] porphyrin (TQMP) (176) and 5, 10, 15, 20-[tetra-(N-methyl-3-pyridyl)]-26,28-diselena sapphyrin chloride (Se2SAP) (177) (**Fig. 1.8**) that have been shown to bind selectively and strongly to G4s.

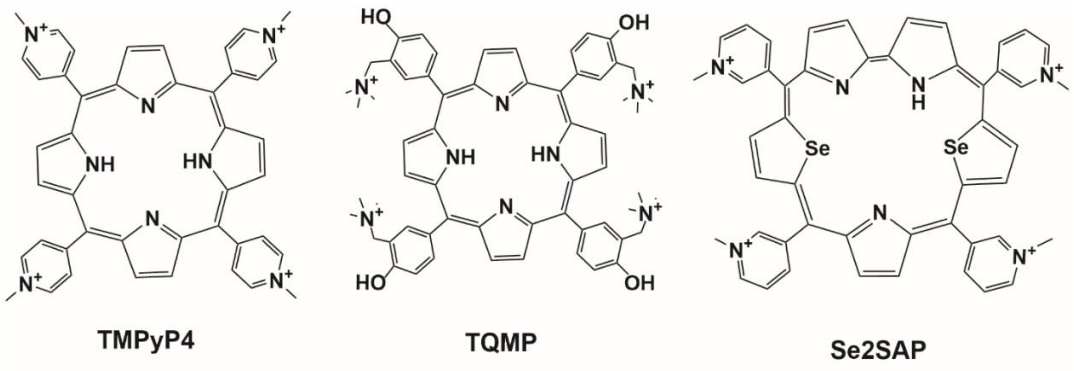

Fig. 1.8. Examples of cationic G4 ligands based on N-alkylation of aromatic linkages (counter-ions; Cl^- for TMPyP4, Se2SAP and I^- for TQMP) (174, 176, 177)

The third class is the metallo-organic G4 ligands synthesized by the insertion of metal ions in the central cavity of the compounds believed to promote the positioning of the central metal core over the cation channel of the quadruplex and as a result, maximizing the stacking interactions between the chelating agent and the accessible G-tetrad (178). The first example of this group includes findings describing the insertion of a copper or manganese metal ion in the middle chamber of TMPy4 (179) (**Fig. 1.9**).

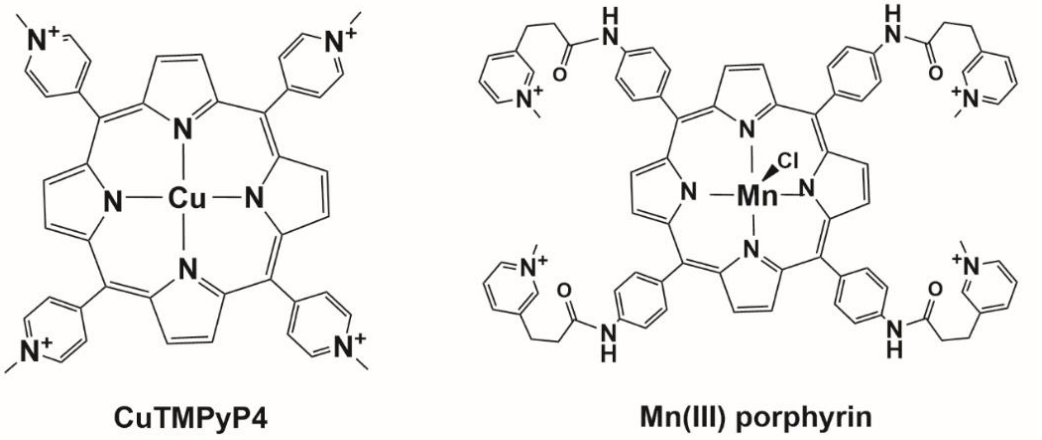

Fig. 1.9. Examples of metallo-organic G4 ligands (counter-ion is Cl^- for CuTMPyP4 and Mn(III) porphyrin (179)

The last class of G4 ligands is the neutral ligands and an example of this is telomestatin, a natural

molecule isolated from *Streptomyces annulatus* in 2001 by Shinya's group (180) (**Fig. 1.10**). This compound is one of the most interesting G4 stabilizers and highly selective. Another neutral macrocyclic G4 ligand is the *N*-methyl mesoporphyrin IX (NMM) which has been studied for its exceptional selectivity for parallel versus antiparallel G4 folds. NMM is also recognized for the ability to adjust its macrocycle geometry to closely fit that of the terminal G-tetrad needed for optimum π- π stacking (159) (**Fig. 1.10**).

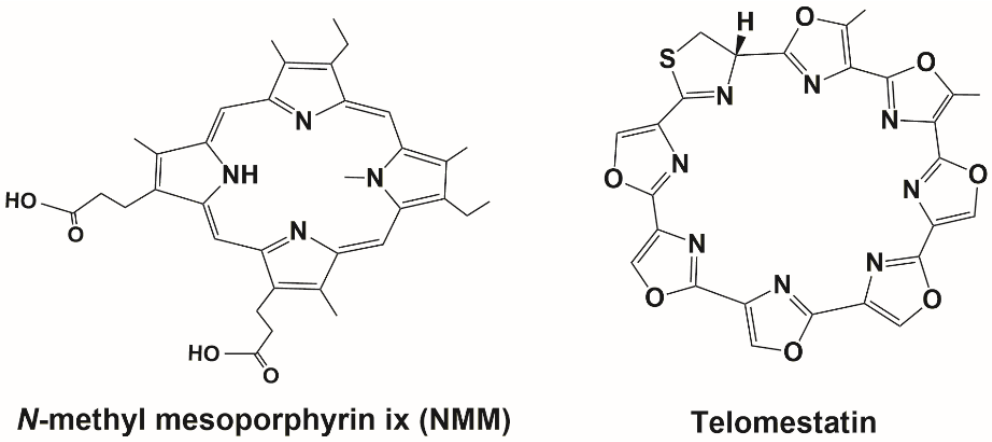

Fig. 1.10. Neutral macrocyclic G4 ligands (159, 180)

Determining structural information about G4s, specific G4 interacting ligands, and knowledge of their mode of interaction with G4s will provide an invaluable insight into the future design of small molecule G4 interacting ligands.

1.6 Proteins that interact with G4s

Proteins can interact with G4s to stabilize their structures (181, 182), or unwind and destabilize G4 structures (183-185), thereby modulating their cellular functions. G4 binding proteins can be classified into those that interact with DNA G4s and those which associate with RNA G4s; a summary of some proteins that recognize, bind and resolve G4 structures is shown in **Table 1.2** and will be introduced in the following sections (186).

Table 1.2. Proteins involved in recognition and binding of G4s.

Location/Function	Gene	Protein
Promoter Regions	*BCL-2*	CNBP
	c-MYC	PARP1
	c-MYC	Nucleolin
	Insulin	Nucleophosmin
	KRAS	IGF-2, insulin
	KRAS	hnRNP A1
	KRAS	MAZ
	MYB	PARP1
	KIT	PARP1
	VEGF	PARP1
		p53 (Mutant)
		MUTS-alpha
		Topo I
Telomere Region		BRCA1
		hnRNP A1
		hnRNP D
		POT1
		RPA
		TEBPs
		TLS/FUS
		Topo I
		TRF2
		UP1
(RNA) G4s		FMR2
		hnRNP A2
		hnRNP A1 (Mutant)
		nucleolin
		Ribosomal proteins
		RHAU
		SRSF 1 and 9
		TRF2
		TLS
G4 resolving helicases		FANCJ
		BLM
		G4R1/RHAU
		Dna2
		WRN
		Sgs 1

CNBP, cellular nucleic-acid-binding protein; PARP-1, Poly [ADP-ribose] polymerase 1; IGF-2, Insulin-like growth factor 2; MAZ, myc-associated zinc-finger; BRCA1, breast cancer type 1 susceptibility protein; hnRNP; heterogeneous nuclear ribonucleoprotein; RPA, replication protein A; TEBP, Telomere End Binding Protein; TLS/FUS, translocated in liposarcoma/fused in sarcoma; Topo I, Topoisomerase I; UP1, unwinding protein 1; FANCJ, Fanconi anemia complementation group J; POT1, protection of telomeres 1; FMR2, fragile X mental retardation 2; RHAU, the RNA helicase associated with AU-rich element; SRSF, serine/arginine-rich splicing factor; BLM, Bloom syndrome protein; TRF2, telomere repeat binding factor 2; WRN, Werner syndrome ATP-dependent helicase; Dna2, DNA replication helicase/nuclease 2; G4R1, G4 Resolvase 1; Sgs1, small growth suppressor 1.

1.6.1 Helicases

Helicases are enzymes that couple the hydrolysis of ATP to the unfolding of double-stranded nucleic acids into their single-strand components (187). Among the most studied helicases are G4 interacting proteins; which play roles in cellular processes involving nucleic acids, including DNA replication and repair, translation, transcription, RNA maturation and splicing, ribosome synthesis, and nuclear export processes (188). The classification of helicases is based on a number of factors, including shared sequence motifs, the structures (helicases not able to form ring structures are grouped under superfamilies 1 and 2 (SF1 and SF2), while ring-forming helicases are grouped under superfamilies 3-6) (189), strand processivity (single-stranded DNA helicase is recognized as an α-helicase while double-stranded DNA helicase is grouped under the β-helicases), and translocation polarity (translocation in the 3′-5′ direction is recognized as type A helicase while translocation in the 5′-3′ direction is recognized as type B helicase) (187).

SF1 consists of (a) SF1A (translocation in the 3′-5′ direction); examples of the best-studied helicases in this group include, Rep and UvrD found in *Escherichia coli*. A crystal structure of Rep helicase revealed two domains that are divided into two subdomains with an ATP-binding site located between two RecA-like (N-and C-core) subdomains, a characteristic observed in many SF1 motifs (190), and (b) SF1B (translocation in the 5′-3′ direction); examples of these are recombination (RecD) and DNA-dependent ATPase (Dda) helicases. The RecD structure revealed two RecA-like core domains and an ATP-binding site in-between, similar to the SF1A helicases (187).

SF2 is the largest helicase superfamily implicated in varieties of cellular processes, distinguished by the presence two tandemly repeated RecA-like domains, with conserved motifs Q, I, Ia, Ib, II, and III in the N-terminal RecA-like domain (NTD) and conserved motifs IV, V, and VI in the C-terminal domain (CTD) (191). Among the most studied subfamilies are DEAD-box RNA helicases (192)

and the RecQ-like family (193). Biochemical information obtained on SF2 proteins suggest that they possess a catalytic core that has ATP-dependent directional movement on single-stranded or double-stranded nucleic acids (187). One of the best studied SF2 helicases that possess the ability to translocate on single-stranded nucleic acid is nonstructural protein 3 (NS3) of hepatitis C virus, which can unwind DNA and RNA duplexes in the presence of 3′-single-stranded overhang (194, 195), and therefore is grouped as a SF2Aα helicase. DEA(D/H) box RNA helicases can translocate on double-stranded nucleic acids and it was suggested by Sengoku *et al*. (191) that members of the DEA(D/H) box RNA helicases having two Rec-like catalytic core domains may incorporate ATP-dependent conformational motion to produce localized RNA remodeling.

SF3 is made up of helicases that were initially found in the genomes of small DNA and RNA viruses (196) and have been reported to be involved with enzymatic activities including, origin recognition and resolving (197). SF3 helicases have four conserved motifs, A, B, B′, and C and have a 3′ -5′ translocation directionality (type A). The most recognized SF3 helicase is the papilloma virus E1 helicase (187).

SF4 members were first isolated and characterized in bacteria and bacteriophages where they act as replicative helicases. They have five conserved sequence motifs; H1, H1a, H2, H3 and H4 (198), and all characterized SF4 helicases have 5′-3′ directionality which corresponds to type B translocation directionality. An example of a well-studied helicase in this family is gp4 protein from bacteriophage T7 (187).

Rho protein is the most studied member of the SF5 helicase superfamily that share some features with SF4 enzymes. It plays a role in the termination of transcription in bacteria by its interaction with a specific sequence at the transcription terminator pause site, that is followed by translocating along the transcript and unwinding the DNA/RNA duplex. It is conserved in three motifs; 1, 1a, and 2 respectively and has a 5′-3′ polarity.

SF6 members include minichromosome maintenance protein (MCM), that plays a role as a replicative helicase in both archaea and eukaryotes (199, 200) with a single copy forming homomeric rings in archaea, while at least six different paralogs (MCM2-7) assemble into a multimeric complex in eukaryotes. Another member of the SF6 family is the RuvB protein which works as dsDNA translocase together with RuvA and RuvC in processing Holiday junctions (201). Structural information on SF6 proteins is still emerging and no crystal structure for the hexameric form is available.

1.6.1.1 Helicase Mechanism of Action

The mechanism of duplex unwinding by helicases will be discussed taking into consideration DEAD- and DExH-box proteins and the two main proposed models by which helicases unwind duplex nucleic acids; the "stepping/inchworm" and the "Brownian motor" models (**Fig. 1.11**) (202, 203). The stepping/inchworm model (**Fig. 1.11 A**) is described as an opened-and-closed conformation of the helicase, where two nucleic acid binding sites bind and release nucleic acids independently in response to the signals received from the NTPase site (204, 205); translocation along one strand of the duplex is achieved by the helicase, resulting into a displacement of hydrogen bonds down the center of the strand as it moves. The Brownian motor model (**Fig. 1.11 B**) involves the co-ordination between the helicase core domains that alternate in the binding affinities for single-stranded/double-stranded nucleic acid and also the binding affinity for NTP (202) hence, the alternating affinity of the helicase for the substrate upon binding and hydrolyzing NTP results in translocation coupled to base pair separation.

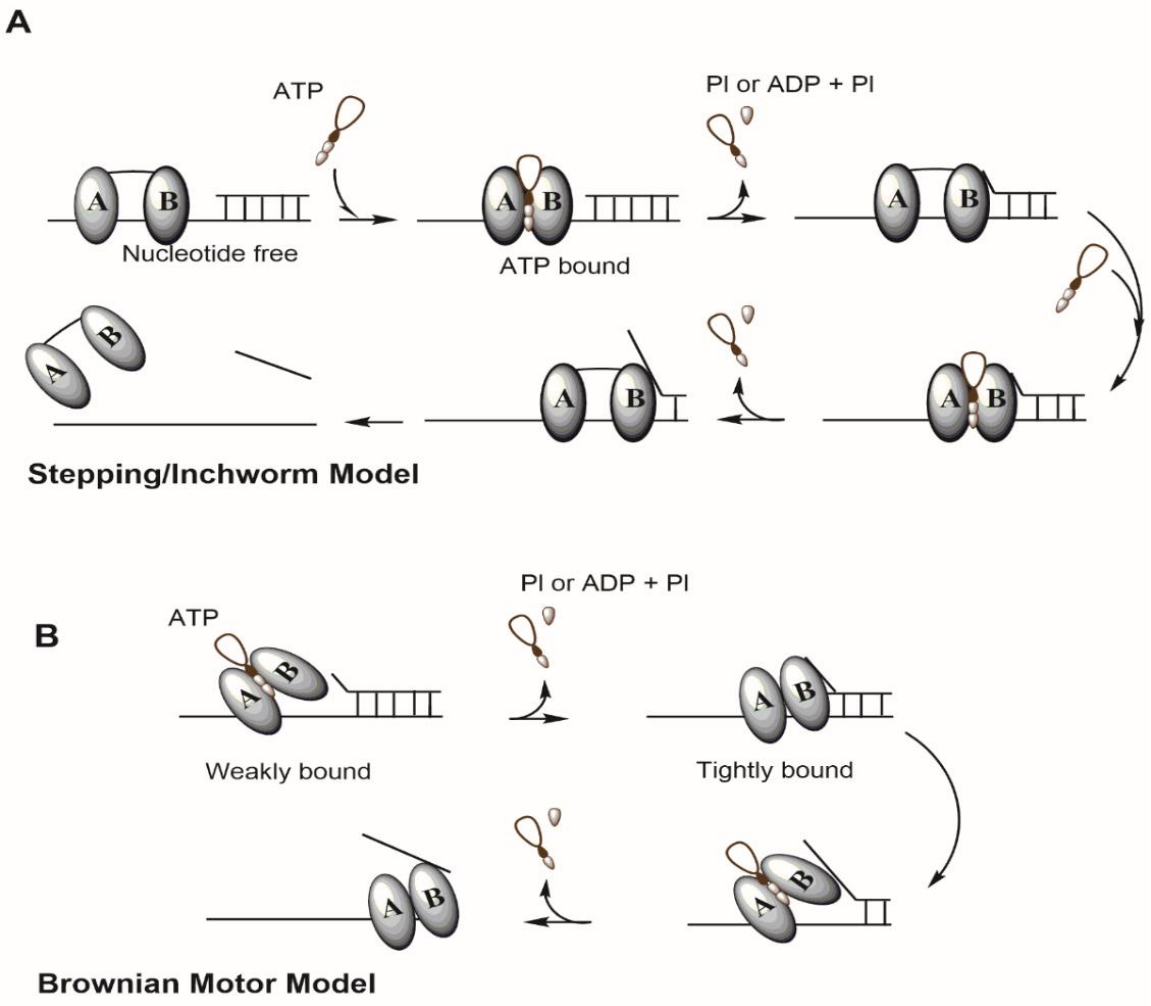

Fig. 1.11. Mechanisms of nucleic acid duplex unwinding by helicases. (A) Stepping/Inchworm model. The helicase binds nucleic acid duplex which increases its affinity for ATP binding, leading to a closed conformation. ATP is hydrolyzed and the helicase adopts an opened conformation, causing translocation of one core domain. **(B)** Brownian motor model. Binding of ATP by the helicase induces a conformational change that results in a weakly bound state between the two domains, and ATP hydrolysis forces the helicase to bind tightly, resulting in a forward translocation along the nucleic acid. The overall opened and closed conformation is repeated for several cycles until the duplex is separated (202, 203).

1.6.2 Telomeric G4-binding proteins

G4 structure forming sequences in telomeres (206, 207) play a significant role in maintaining the integrity of the genome and its stability by offsetting the effect of chromosomal end shortening during replication. Human telomeres are made up of tandem repeats of the sequence $d(TTAGGG)_n$ and a telomere-specific protein complex, shelterin. The shelterin complex is made up of six protein subunits namely: protection of telomeres 1 (POT1), telomere repeat binding factor 1 and 2 (TRF1 and TRF2), tripeptidyl peptidase (TPP1), TRF1 interacting protein 2 (TIN2) and repressor activation protein 1 (RAP1). They regulate the maintenance of telomere length by protecting chromosomal ends from being recognized as damaged DNA (119, 208-210). The POT1/POT1-interacting protein 1 (TPP1) heterodimer has been shown to interact specifically with single-stranded DNA at the 3′-overhang of telomeric repeats (211) to protect the G-overhangs against the binding of replication protein A (RPA) (212, 213). The mode of interaction of POT1/TPP1 with the telomere is similar to that of the telomeric end binding protein (TEBP) described in *Oxytricha nova* (*O. nova*) (125, 126) where the protein complex binds to a 3′-overhang in the form of G-tetrad DNA (214). Replication protein A (RPA) has also been shown to bind and unwind G4 structures (215, 216) with a strand preference in the 5′-3′ direction. In 2014, Ray *et al.* (217) gave an insight into the process by which telomere overhangs are protected by POT1-TPP in comparison to RPA. They revealed that POT1 unwinds parallel G4s while RPA unwinds both parallel and antiparallel conformations. Breast Cancer Type 1 Susceptibility Protein (BRCA1) has been shown to interact directly with human telomeres to regulate the activity of telomerase and hence are implicated in the maintenance of telomeric 3′- overhang length (218, 219). BRCA1's newly recognized roles in telomere regulation and its preferential binding to DNA G4 suggests an important role in the processes involved with G4s in the genome.

1.6.3 G4 binding proteins at promoter regions

The promoter region is a DNA sequence that defines the RNA polymerase start site for transcription of a specific gene, and is usually found directly upstream of the transcription initiation site. Although gene promoter sequences having a continuous stretch of G-tracts can fold into an intramolecular G4 (220), however, other G4 conformations have also been shown to exist in promoter regions including a bimolecular G4 arrangement (221). G4 binding proteins are involved in direct transcriptional regulation, Poly [ADP-ribose] polymerase 1 (PARP-1) interacts with DNA G4s (intramolecular) *in vitro* with high affinity and a stoichiometry corresponding to one G4 for two proteins (222), and according to the results obtained from an enzymatic assay, PARP-1 was activated after interaction with G4s found at the c-Kit promoter (222). The G4-forming element of the murine KRAS oncogene was demonstrated to interact with PARP-1 and MAZ proteins by inducing conformational transformation of duplex to G4 DNA. They have been reported to activate the promoter of KRAS by recognizing a parallel G4 conformation of the quadruplex-forming element (222, 223). Other G4 binding proteins regulate transcription indirectly through chromatin remodeling and DNA repair proteins, including; nucleolin, nucleophosmin, cellular nucleic acid-binding protein (CNBP), heterogeneous nuclear ribonucleoprotein A1 (hnRNP A1) and myc-associated zinc-finger (MAZ) protein. Nucleolin is a phosphoprotein localized in the nucleolus and is highly expressed in proliferating cells with various roles in the biogenesis of the ribosome, transcription, chromatin remodeling, apoptosis and G4 binding (224-228). Nucleolin binds to a quadruplex structure within c-MYC with high selectivity and affinity when compared with other G4s that exist *in vitro*. Nucleolin has also been shown to interact with the c-MYC promoter *in vivo* (229).

1.6.4 RNA G4-Binding Proteins

RNA G4s regulate several cellular functions including alternative splicing, telomerase activity,

transcription termination, and translational regulation (93, 230). Far fewer RNA G4-binding proteins have been reported compared to DNA G4-binding proteins (183, 231). Topoisomerase 2-associated proteins (PAT1) have been demonstrated to interact with RNA *in vitro* with a preference for G-motifs, including 5′-UTR of NRAS and when tested in mRNA reporter constructs, they were shown to repress translation as expected for a G4-stabilizing protein (98, 232, 233). The functional role of RNA G4s and their interactions with both FMRP (Fragile X Mental Retardation Protein) and FMR2P (Fragile X Mental Retardation 2 Protein) have been investigated (234, 235); deficiency of these proteins is associated with the Fragile X mental Retardation syndrome (FXS) and the FRAXE-associated mental retardation (FRAXE) (236). Darnell *et al*. showed that FMRP proteins can interact with an RNA target harboring a G4 motif at high affinity and specificity through its arginine-glycine-glycine (RGG) box (235). Von *et al.* (237) revealed a number of proteins that interact with RNA G4s including heterogeneous nuclear ribonucleoprotein (hnRNP), splicing factors, and ribosomal proteins. Others include telomeric repeat-binding factor 2 (TRF2), and DHX36 also known as RNA helicase associated with AU-rich element (RHAU) (238).

1.6.5 G4-Resolving Helicases

The unwinding of DNA and RNA duplexes by helicases is usually needed to allow cellular machineries to act on nucleic acid substrates and so, helicases are needed in practically all cell metabolism including replication, transcription, recombination, repair, and telomere maintenance (239-242). DNA G4-resolving helicases have been studied including RecQ proteins (WRN, BLM and FANCJ) and also RNA G4-resolving helicases DHX9 and RHAU. The irregular metabolism of telomeres in Werner syndrome cells has been linked to the function of WRN helicase in regulating G4 unwinding (243). WRN belongs to the RecQ family and patients with dysfunctional WRN suffer from premature aging. A G4 structure in a DNA template was demonstrated to be unwound by WRN helicase, which promoted replication progression

by polymerase δ *in vitro* (244), and it was shown that WRN requires a 3′-ssDNA tail to unwind a variety of G4 DNAs (245). BLM is another helicase shown to be very active *in vitro* in unfolding G4 DNA structures (246). A mutation in the gene encoding the BLM helicase leads to Bloom syndrome (247) and patients with this syndrome suffer from a rare autosomal recessive disorder characterized by genetic instability and predisposition to cancer development. Recent studies have suggested that BLM unwinds G4 structures in an ATP-independent manner (248) and depending on the G4 structure's environment, BLM helicase may translocate along ssDNA to unwind G4s in an ATP-dependent manner (249). DHX9, also known as RNA helicase A (RHA), has been shown to unwind DNA and RNA duplexes with a 3′ -5′ polarity (250) but preferentially acts on RNA substrates to unwind G4 structures (251). RNA helicase associated with AU-rich element, also known as G4 resolvase (G4R1) and DEAH box protein 36 (DHX36), is an RNA helicase that has been shown to unwind duplex RNA, RNA-DNA hybrids and other non-canonical DNA structures including triplexes (252, 253). This helicase will be discussed further in the following section as it is central to the research done in this thesis.

1.6.5.1 RNA helicase associated with AU-rich element (RHAU)

RHAU, was identified by Tran *et al.* in 2004 as a helicase associated with AU-rich element (ARE) of urokinase plasminogen activator (uPA) mRNA, together with nuclear factor 90 (NF90) and human antigen R (HuR) (254). They showed the involvement of RHAU in ARE-mediated mRNA decay through its RNA-dependent interaction with ARE-binding protein NF90 by recruiting human exosome and a poly(A)-specific exoribonuclease (PARN). RHAU protein accelerated the deadenylation and subsequently, decay of β-globin-ARE^{uPA} (254). Interestingly, RHAU was later identified as a G4 helicase enzyme, capable of binding and unwinding inter- and intra-molecular G4s in DNA and RNA and has subsequently been shown to play regulatory roles in a variety of G4-mediated processes (50, 252, 253). RHAU interacts with

both DNA and RNA G4s with high affinity and specificity through its conserved motif in the N-terminal region, the RHAU-specific motif (RSM) (238, 253). RHAU was shown to play a role in the upregulation of Yin Yang 1 (YY1) gene as demonstrated to interact with the G4 motif of the YY1 promoter *in vitro* and others have demonstrated the roles of RHAU in remodeling of G4s at the 5′-end of the human telomerase RNA (50, 255-257). Booy *et al*. (49) demonstrated a novel role for RHAU in microRNA-mediated translational regulation at a G4-contained in the 3′-UTR of PITX1 mRNA. Similarly, RHAU was found implicated in miRNA-mediated regulation of a gene via miRNA translocation and direct interactions with human argonaute proteins (254, 258-260).

1.6.5.2 RHAU belongs to DExH/D RNA helicases

RHAU belongs to the DExH/D family of proteins that participate in myriad cellular functions including RNA splicing, miRNA processing, RNA trafficking, mRNA stability, ribonucleoprotein remodeling and ribosome assembly (261-264), and some DExH proteins have also been characterized as DNA helicases (including human RecQ and various homologs in other organisms) (265). The DExH/D family, a subset of the SF2 helicase family, are made up of proteins from the DEAD, DEAH and DExH subgroups (266, 267), and they all contain at least eight characteristic sequence motifs, one of which is the ATP-hydrolysis (motif II) from which their names are derived: DEAD, DEAH and DExH (**Fig. 1.12**), signifying single-letter amino acids (188, 268, 269). Biochemically characterized DExH/D proteins have been shown to possess ATP hydrolysis activity which is mostly stimulated by RNA or DNA.

1.6.5.3 The G4 binding motifs of RHAU

According to conservation of domains when compared to other helicases, RHAU possesses a core DEAH-box helicase domain (residues 210-614), representing the conserved ATPase/helicase motifs I-VI

of the DEAD/H-box family, flanked by an N-terminal G4-recognition domain (residues 1-210) and a C-terminal helicase-associated domain (residues 670-1008), that have not been fully characterized (**Fig. 1.13**) (270).

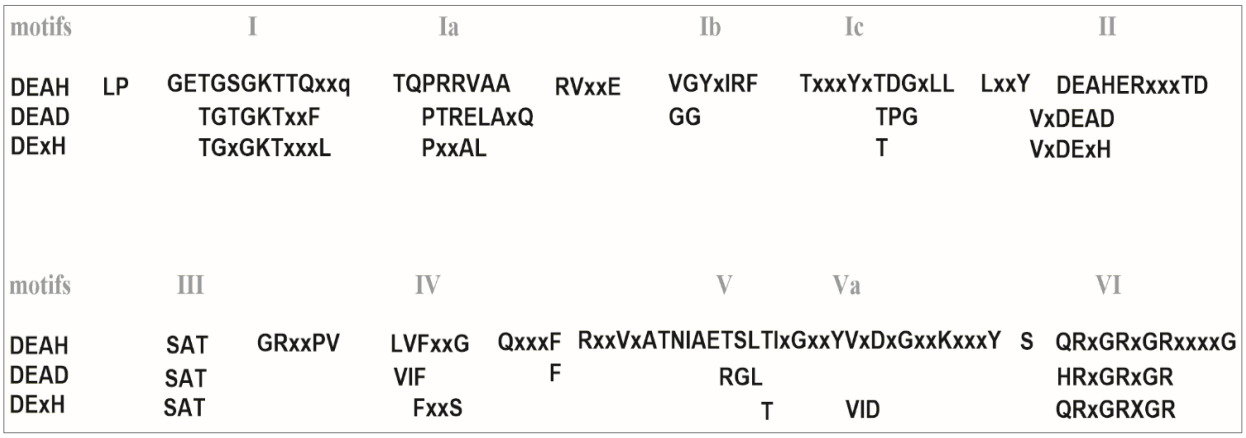

Fig. 1.12. Characteristic sequence motifs of DEAH, DEAD and DExD helicases. Capital letters represent conservative substitutions within each group while x denotes variable residues; amino acids that are aligned from different groups have similar distance and characteristic motifs were identified following the alignments of 29 DEAH, 42 DEAD and 44 DExH proteins respectively (267).

The gene DHX36 is known to generate two variant splices of RHAU, where the first isoform is localized to both the nucleus and the cytoplasm, and the second isoform which lacks 14 amino acids in the helicase domain is predominantly expressed in the cytoplasm (254, 271). The interaction of G4 structures with RHAU is mediated by its short N-terminal specificity motif known as the RHAU-specific motif (RSM), a 13-residue stretch (residues 54-66) in the N-terminal sub-domain (50, 253). Although this segment of RHAU was shown to be sufficient for G4-specific recognition and binding, however, the full-length protein has been shown to possess higher affinity; 100-fold higher affinity for parallel relative to anti-parallel G4s (272-274) and is also required for helicase activity (238, 253, 275). A recent NMR solution structure of a peptide-G4 complex explains how RHAU specifically recognizes parallel G4 structures. This structure

revealed the interaction of a DNA G4 with an 18-amino acid peptide at the N-terminal region of RHAU and demonstrated a stacking interaction between the peptide's amino acid residues glycine (G9), isoleucine (I12), glycine (G13), and alanine (A17) and the DNA guanine of the 5′-end G-tetrad. Clamping of the G4 structure by means of a "three-anchor-point" electrostatic interactions mediated by three positively charged side-chains of peptide lysine (K8), arginine (K10), and lysine (K19) and the DNA phosphate backbone, a binding mode that is similar to that of most ligands selected for specific G4 targeting was also reported (274). The helicase activity of RHAU was reported to be highly sensitive to the thermodynamic stability of G4s, with an inverse relationship between G4 substrates and unwinding by RHAU noticed (272).

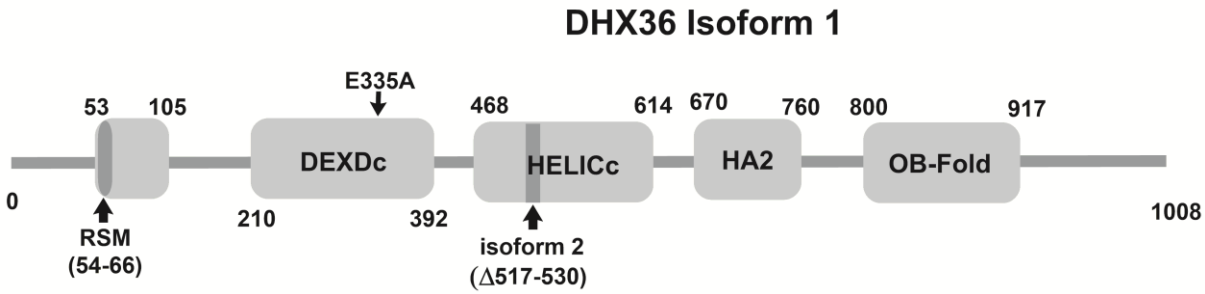

Fig. 1.13. Schematic diagram showing the domain structure of RHAU; predicted by the National Centre for Biotechnology Information (NCBI) conserved domain database (270). RHAU's core DEAH-box helicase domain (residues 210-614), representing the conserved ATPase/helicase motifs I-VI of the DEAD/H-box family, flanked by an N-terminal G4-recognition domain (residues 1-210) and a C-terminal helicase-associated domain (residues 670-1008), that have not been fully characterized are shown.

RHAU has been reported to play critical roles in development and differentiation. Lai *et al*. (276) generated a conditional RHAU gene knockout in mice and showed that a deletion of RHAU in germ line resulted in embryonic lethality. They also performed a statistical analysis of the deregulated genes containing G4 motifs in their promoters, which suggested a role of RHAU in the regulation of gene expression tethered to its G4 resolvase activity (276). The role of RHAU in differentiation was also

demonstrated by Gao *et al*. where they carried out a RHAU knockout in mice germ cells which promoted the formation of G4 DNA structures within the promoter region of tyrosine kinase (C-Kit) gene, leading to a downregulation in its expression. They concluded that RHAU is an essential protein in the differentiation of spermatogonia (cells produced at an early stage during the formation of spermatozoa) (277).

RHAU has gained recognition for interacting with several non-coding RNAs in the regulation of subcellular localization of specific microRNA, and in the remodeling of G4s within the 5′-region of the telomerase RNA (50, 256, 257, 259) and 3′-untranslated region of PITX1 mRNA (49). The helicase activity of RHAU was demonstrated to be sufficient for unwinding a G4 structure within the 5′-region of human telomerase RNA (hTR) which promotes an interaction with 25 internal nucleotides to form a stable P1 helix: a conserved structural element located upstream of the 5′-end of telomerase RNA, that serves as a template boundary limiting reverse transcription to 6-nucleotides (278). RHAU knockdown by siRNA resulted in a significant reduction in average telomere lengths (50).

To gain insight into biologically relevant RNA G4 structure recognition by the helicase RHAU, an RNA co-immunoprecipitation screen was performed in our laboratory and we identified the mRNA for the protein Pituitary homeobox 1 (PITX1). Interestingly, PITX1 mRNA has been shown to possess three distinct G4 forming sequences in the 3′-untranslated region (UTR) of its mRNA (Q1: $PITX1_{1371\text{-}1400}$, Q2: $PITX_{11901\text{-}1930}$, and Q3: $PITX1_{2044\text{-}2079}$), and these G4s, especially Q2 and Q3 have been shown to play roles in the recruitment of RHAU to the PITX1 mRNA, thereby regulating PITX1 protein translation (49, 275).

1.7 The Pituitary Homeobox gene 1 (PITX1, PTX1)

Homeobox, a DNA sequence containing about 180 base pairs, is located within the genes that play regulatory roles in the development of anatomical patterns (morphogenesis) in animals, fungi and plants. Homeobox genes encode homeodomain proteins, which contain an evolutionarily conserved domain found

in many DNA-binding transcription factors that control biological processes, including specification of cell type and pattern formation in embryos (279). Pituitary homeobox 1 gene, also referred to as backfoot (BFT) (280, 281) encodes a member of the bicoid (Bcd) subgroup of paired homeobox proteins. Bcd, a homeodomain-containing protein in *Drosophila* encoded by the maternal gene *bicoid*, is needed to initiate polarity along the posterior axis of the embryo in the early stages of development (282, 283), and is distributed in an anterior to posterior gradient in the embryo (284).

Homeodomain protein is made up of a 60-amino acid helix-turn-helix (HTH) structure consisting of two antiparallel helices at the N-terminus and a third longer helix at the C-terminus, positioned roughly perpendicular to the axes of the first two helices with all three alpha helices connected by short loop regions. The specificity of the DNA-binding motif of homeodomain protein is primarily determined by its third helix, also known as the recognition helix, which inserts itself into the major groove of the recognition site. Interaction is mediated through hydrogen bonds and hydrophobic interactions between specific amino acid side chains, exposed nucleotide bases and thymine methyl groups in the major groove of the DNA (285). The interaction specificity is also achieved through the flexible amino-terminal arm tracking along the minor groove of the DNA, with the second helix contacting the backbone of the DNA. Homeodomain proteins have a common "TAAT" core that determines binding specificity (286). The amino acid at the 9^{th} position of the recognition helix, which corresponds to the 50^{th} position of the homeodomain protein, plays a crucial role in the selective recognition of DNA (287, 288). In general, homeodomains that contain a glutamine (Q50) residue at position 50 binds preferentially to the consensus 'TAATGG' while homeodomains containing a lysine (K50) residue at position 50 interact preferentially with the consensus 'TAATCC'.

PITX1 has been reported to play roles in both development and disease, including a pivotal role in the differentiation of developing pituitary gland, craniofacial structures and hind limbs in early development

of embryos (289-292). According to Klopocki *et al.* (293) improper formation of the lower limbs is likely a result of mutations in the PITX gene. The expression of PITX1 is down-regulated in some tumor cells such as lung, gastric and esophageal cancer (294-296). PITX1 has been reported to bind specifically to the promoter region of the telomerase reverse transcriptase (TERT) gene *in vitro* and *in vivo* to suppress its transcription thereby regulating telomerase activity (297). Furthermore, PITX1 was shown to suppress tumorigenicity by downregulating the RAS pathway through RAS protein activator like 1 (RASAL1) (298) and PITX1 expression is upregulated in response to DNA damage, thereby acting directly as a transcriptional upregulator of protein 53 (p53) which is a crucial element of the DNA damage response frequently mutated in many cancer types (299, 300). Takahito *et al.* demonstrated that a microRNA (miR-19b) regulates the expression of human telomerase reverse transcriptase (hTERT) and cell proliferation by inhibiting PITX1 in melanoma cells (301).

Recently, our laboratory confirmed a specific interaction between endogenously expressed RHAU protein and PITX1 mRNA. Selected regions in the 3′-UTR of PITX1 mRNA were shown to form stable G4 structures that are implicated in RHAU-dependent PITX1 regulation (49). To investigate the functional consequences of the interaction between PITX1 mRNA and RHAU, Booy *et al*. (49) performed RHAU knockdown by siRNA in human cells for 96 hours and monitored protein levels; RHAU knockdown corresponds to an increase in the expression level of PITX1 protein, suggesting RHAU as a negative regulator of PITX1 protein expression. Mutations of the G4-forming sequences at the 3′-UTR of PITX1 mRNA resulted in a reduced upregulation of PITX1 protein expression, in the context of RHAU knock down in a β-galactosidase reporter assay which suggests that G4 structures may be playing regulatory role (49).

1.8 Rationale for the thesis

As mentioned earlier, G4 structures were reported to play significant roles in a number of biological contexts such as regulation of gene transcription and translation among others (**Fig. 1.5**) (302). The helicase RHAU has been shown to interact with G4s, and previous structural/biophysical studies have suggested that this interaction is mediated by the G-tetrad faces of the G4 in both DNA and RNA (238). Furthermore, a recent NMR solution structure of a DNA G4 in a complex with an 18-amino acid peptide at the N-terminal region of RHAU demonstrated a stacking interaction between the peptide's amino acid residues glycine (G9), isoleucine (I12), glycine (G13), and alanine (A17) and the DNA guanine of the 5′-end G-tetrad. Clamping of the G4 structure by means of a "three-anchor-point" electrostatic interactions mediated by three positively charged side-chains of peptide lysine (K8), arginine (K10), and lysine (K19) and the DNA phosphate backbone was also reported (274). This contrasts with another finding which demonstrated that mutations of charged amino acids in the RSM do not have significant impact on binding (253).

Although studies have suggested that RHAU recognizes and interacts with a parallel G4 conformation (**Fig. 1.4B**) and the full-length protein has been shown to possess 100-fold higher affinity for parallel relative to anti-parallel G4s (272-274) however, the mechanism of how RHAU recognizes and resolves these structures has not been fully unraveled. To have further insights into the mode of G4 recognition by RHAU and the impact of loop conformation on G4 structures folding and binding, we have chosen a model for this study, PITX1 mRNA containing sequences capable of forming stable G4 structures in its 3′-UTR and whose interactions with RHAU in a cellular context has been established (49). From this model, three unique G4s will be studied alongside their DNA counterparts to know if their mode of recognition by RHAU is similar to previously characterized G-quadruplexes.

CHAPTER 2

MATERIALS AND METHODS

2.1 Reagents

Molecular biology grade laboratory chemicals used in the course of this research were obtained from the following companies: Integrated DNA Technologies (Coralville, Iowa), Frontier Scientific (Logan, UT, USA), Sigma-Aldrich (Oakville, ON) and Fisher Scientific (Ottawa, ON). Other chemicals used in the preparation of experimental reagents are mentioned in their different sections. Preparation of buffers was carried out using deionized ultra-purified water (MiliQ, EMD Milipore, affiliated to Merck, ON, Canada).

2.2 Plasmid Design and Cloning

The high-copy-number plasmid pUC119 was used as template for all *in vitro* transcription of RNA. A fragment containing 6-nucleotide overhangs at both ends, a HindIII restriction site, the T7 promoter sequence, desired RNA coding sequence, BsaI and EcoRI restriction sites was generated by PCR for each RNA to be transcribed (**Fig. 2.1**).

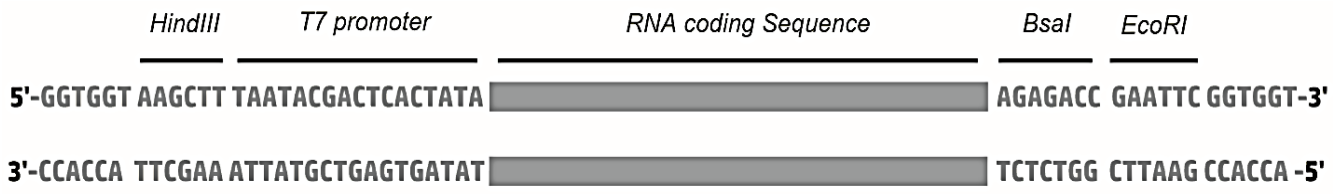

Fig. 2.1. The design of sequence and PCR construction of DNA template for in vitro transcription.

The DNA fragment was digested with HindIII and EcoRI restriction enzymes, purified and sub-cloned into the pUC119 plasmid previously digested with the same restriction enzymes. The preparation of plasmid for

in vitro transcription was carried out in a 2-Litre culture, purified using plasmid Mini or Maxi-preparation kits and linearized with appropriate restriction enzymes.

2.2.1 Cloning process steps

The DNA fragments designed for cloning into pUC119 containing 6-nucleotide overhangs at both ends, a HindIII restriction site, the T7 promoter sequence, the desired RNA coding sequence, BsaI and EcoRI sites are shown in (**Fig. 2.3**).

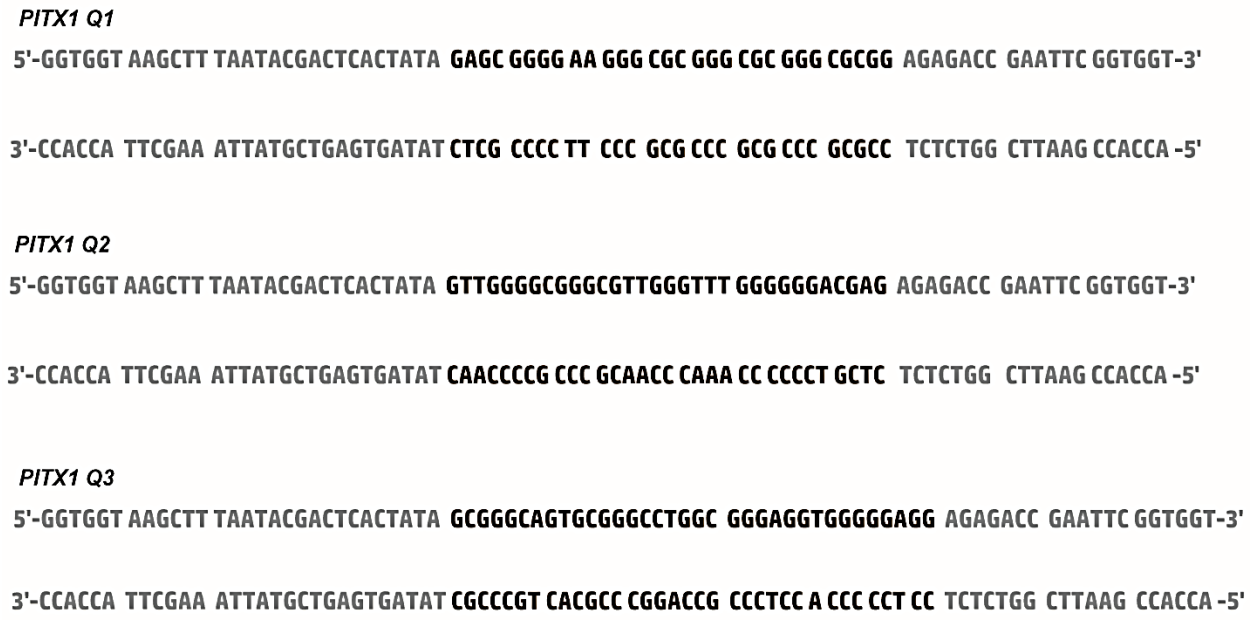

Fig. 2.2. Cloned sequences into pUC119. Fragments containing 6-nucleotide overhangs at both ends, a HindIII restriction site, the T7 promoter sequence, desired RNA coding sequence, BsaI and EcoRI restriction sites generated by PCR for each RNA to be transcribed.

Each fragment was ligated into a pUC119 vector with both the insert and the vector initially treated with EcoRI and HindIII restriction enzymes (Thermo Scientific), performed in 10x FastDigest buffer (Thermo Scientific). The restriction digested insert and vector were purified by using GeneJET PCR cleanup kit

(Thermo Scientific) and purity was confirmed by running the samples on a 1.5% agarose gel at 80 Volts where individual bands were visualized under UV light and excised carefully using a clean razor. The excised DNA bands were purified by GeneJET gel purification kit (Thermo Scientific). Ligation reaction was set up between the purified insert and the vector mixed in 3:1 molar ratio in a 20 µL reaction keeping the total DNA concentration below 10 µg/mL in the presence of T4 DNA ligase (New England Biolabs, Whitby, ON, Canada). The ligation reaction was transformed with MAX Efficiency® DH5α™ competent cells (Life Technologies, Invitrogen) and cells were plated on Lysogeny-broth (LB) agar plates supplemented with 100 µg/mL ampicillin and grown at 37 °C overnight. Individual colonies were selected and grown overnight in LB media supplemented with 100 µg/mL ampicillin in an incubator shaker at 200 rpm (37 °C). Plasmids were purified from overnight culture by miniprep plasmid preparation kit (GeneJET miniprep kit, Thermo Scientific) and the integrity and accuracy of cloned plasmids were verified through sequencing performed by Manitoba Institute of Cell Biology (MICB), Winnipeg.

2.2.2 Plasmid preparation for *in Vitro* Transcription of RNA

Plasmid DNAs were isolated using plasmid extraction and preparatory kits according to the manufacturer's procedures. The following kits were used depending on the amount of plasmid required: Genejet plasmid miniprep kit (#K0503, Thermoscientific, Pittsburg, PA, USA), Genejet plasmid midiprep kit (#K0481, Fermentas life science, Thermoscientific, Pittsburg, PA, USA), Qiagen plasmid maxi kit (#12162) and Qiagen plasmid Giga kit (# 1291, Toronto, ON, Canada). Plasmid DNAs were linearized using BsaI restriction enzyme (NEB) at 37 °C overnight to ensure proper linearization. Linearized plasmids were purified by extraction with phenol/chloroform (1:1), followed by DNA precipitation with 3 M sodium acetate and absolute ethanol. Precipitated DNA was centrifuged at 21000xg on a tabletop Legend Micro 21

R centrifuge (Thermo Scientific) and the pellet was washed with 70% ethanol to remove residual salt. The DNA pellet was dissolved in HPLC-grade water.

2.2.3 *In Vitro* Transcription of RNA using plasmid DNA Templates

Several trial transcription reactions (50 µL reaction volume) were set up to optimize the reaction conditions for large-scale reactions. Each small-scale reaction was set up to vary the concentrations of NTPs, DNA template, $MgCl_2$ and T7 RNA polymerase (produced in-house). An example of a small-scale reaction is demonstrated below;

a. 1 x transcription buffer [40 mM Tris, pH 8.1; 1 mM spermidine, 0.001% (wt/vol) Triton X-100, 10 mM DTT in HPLC water],
b. $MgCl_2$ (5-20 mM),
c. 8 mM NTPs (2 mM each),
d. Linearized DNA template (2.5-10 µg),
e. T7 RNA polymerase (0.5-2.0 µL).

The reactions were incubated at 37 °C for 30-60 minutes and transcripts were analyzed on denaturing TBE-polyacrylamide gel electrophoresis (PAGE) supplemented with 8M Urea to denature any secondary RNA structures. Prior to electrophoresis, RNA samples were mixed with denaturing load buffer (0.02% bromophenol blue, 0.01% xylene cyanol FF, 10% glycerol and 8 M Urea), and heated at 95 °C for 5 minutes and then loaded onto TBE gels. PAGE was carried out at 150 volts in 1 x TBE buffer enclosed in a Mini-Protean 3 cell (Biorad, Mississauga, ON, Canada). Gels were stained with 0.1% toluidine blue solution for about 5 minutes and destained in water to visualize the RNA, and the optimized trial transcription conditions were scaled up to

the desired volume. Pyrophosphate (PPi)-$MgCl_2$ precipitate, a byproduct of the transcription reaction was formed which

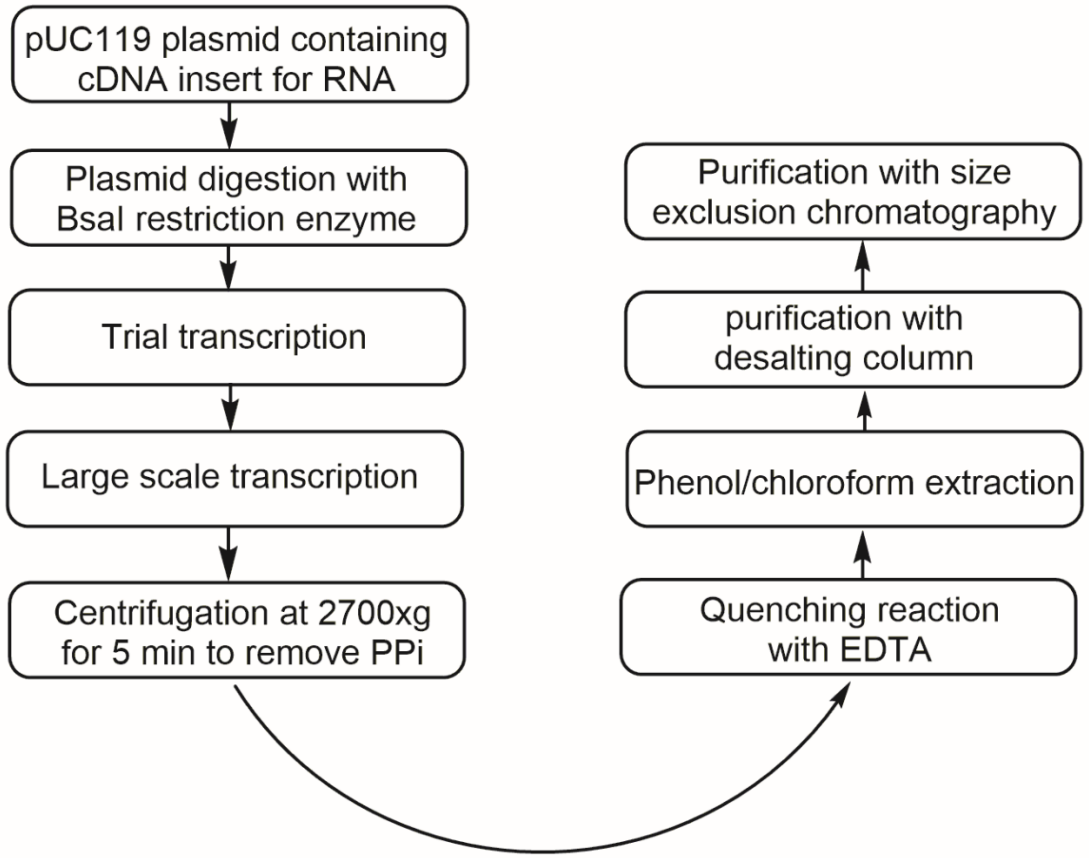

Fig. 2.3. A flow chart of RNA in vitro transcription steps

is removed by centrifugation at 2700xg for 5 minutes as a sediment while the transcription reaction was terminated by the addition of EDTA, pH 8.0 (10% of the reaction volume of 0.5 M EDTA added) to the supernatant up to a concentration of 50 mM with thorough mixing as chelating agent for excess Mg^{2+}.

To denature the T7 RNA polymerase, phenol/chloroform/isoamyl alcohol (25:24:1) extraction was performed, followed by purification of the RNA transcript using a 10-DG desalting column (BioRad) which removes traces of phenol/chloroform and small molecule contaminants

like salts and NTPs. The 10-DG column was equilibrated with 20 mL of RNA buffer (10 mM Tris, pH 7.5; 100 mM KCl, 1 mM EDTA) and 3 mL of transcription reaction was applied to the column and eluted with 5 mM of RNA buffer (10 mM Tris, pH 7.5; 100 mM KCl, 1 mM EDTA). Transcribed RNA was finally purified using size exclusion chromatography (SEC) which separates residual DNA template, aborted transcripts and excess small molecules from the pure RNA. The pure RNA is purified as a single peak and fractions were collected under the peak and analyzed by denaturing TBE-PAGE (**Fig. 2.4**).

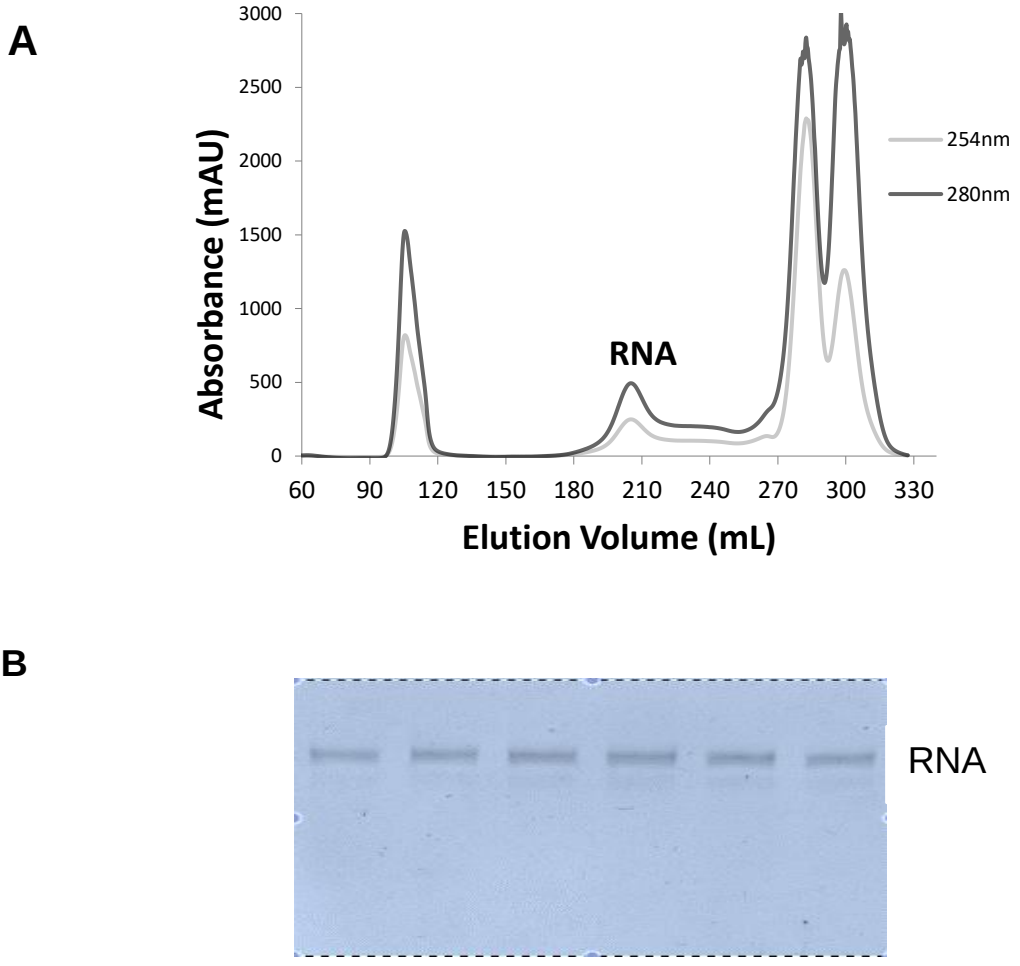

Fig. 2.4. SEC purification of RNA (A) elution profile of *in vitro* transcribed RNA showing the plasmid, RNA and NTPs peaks respectively. **(B)** TBE-PAGE showing RNA in 6 fractions of 5mL each (fractions between 190mL-220mL were collected from the RNA peak)

2.3 G4 preparation

Synthetic RNA and DNA used in this research were ordered from Integrated DNA Technologies (Coralville, Iowa), provided desalted and certified by mass spectrometry from the manufacturer. G4 formation for both *in vitro* transcribed RNA and synthetic RNA were carried out by dissolving the RNA in 10 mM Tris (pH 7.5), 100 mM KCl, 1 mM EDTA at a concentration of 5 µM and heated at 95 °C for 5 minutes and then allowed to cool slowly to room temperature. Nucleic acid in conformations different from G4 were separated by size exclusion chromatography (SEC) on a HiLoad Superdex 75 26/60 column (ÄKTA, GE Healthcare) in 10 mM Tris (pH 7.5), 100 mM KCl, 1 mM EDTA. The purity of the samples was confirmed by running both native and denaturing gel electrophoresis separately (15% native or denaturing TBE-PAGE using urea as a denaturant). The detection of G4-forming RNA and DNA was made by staining the gels with quadruplex specific dye [*N*-methyl mesoporphyrin IX (NMM, Frontier Scientific, Logan, UT, USA)]. Extinction coefficients (260 nm) were calculated from the nucleotide sequence using IDT SciTools (OligoAnalyzer 3.1, Integrated DNA Technologies), corrected for hyperchromicity by the use of absorption spectra at 20 °C and 90 °C (Q2RNA, 253550 $M^{-1}cm^{-1}$; Q2DNA, 264200 $M^{-1}cm^{-1}$; Q3RNA, 347600 $M^{-1}cm^{-1}$).

2.4 Expression and purification of protein

2.4.1 Expression and purification of RHAU53-105 in *E. Coli*

RHAU53-105 fragment containing an N-terminal hexahistidine tag and a thrombin cleavage site for tag removal was generated by PCR from the RHAU-FLAG vector ($RHAU_{1-1008}$ in pIRES-EGFP-FLAG-N1), a kind gift from Dr. Y. Nagamine (FMI, Basel, Switzerland). The

fragment was cloned into the pET15b vector, expressed in *Escherichia coli* strain BL21 (DE3) and purified using a cobalt affinity resin column.

2.4.2 Procedures

2.4.2.1 Expression of RHAU53-105 protein

A 25-mL starter culture of BL21 (DE3) transformed with plasmid in Lysogeny broth (LB) was grown overnight in the presence of 100 µg/mL ampicillin at 37 °C. This was then used to inoculate 1.6 L of LB containing 100 µg/mL ampicillin (in a 4 liter Erlenmeyer flask) and grown at 37 °C in an incubator shaker (200 rpm) until the OD_{600} reached 0.6. After the desired OD was reached, an inducing reagent, isopropyl-β-D-thio-galactopyranoside (IPTG) was added to a final 1 mM concentration and further incubated in an incubator shaker kept at 37 °C (200 rpm) for 3 hours to express the protein.

2.4.2.2 Isolation of RHAU53-105 protein

The culture was transferred into centrifuge bottles and subjected to centrifugation at 6,000xg for 15 min (Thermo Scientific™ Sorvall™ centrifuge) at 4 °C and the cell pellet resuspended in 40 mL of lysis buffer (20 mM sodium phosphate, 2 mM imidazole, 6 M guanidine hydrochloride, pH 7.0). Protease inhibitors including 400 µL of 1mM 4-(2-Aminoethyl)-benzenesulfonyl fluoride (AEBSF), 40 µL of 10 µM pepstatin A in dimethyl sulfoxide and 40 µL of 10 µM bestatin in methanol were added, followed by homogenizing and the cells were lysed by sonication (20s at 50 Hz with pulses at 30s intervals for 10 cycles) on ice. This process was immediately followed by centrifugation at 22,000xg for 30 min at 4 °C (Thermo Scientific™ Sorvall™ centrifuge) to obtain a supernatant which was carefully transferred into a clean bottle.

The clarified cell lysate was loaded onto a pre-equilibrated 10 mL Talon cobalt resin affinity column to remove proteins that do not specifically bind to the column. RHAU53-105 protein was eluted in 5 mL fractions with the elution buffer (150 mM imidazole, 300 mM NaCl, pH 7.0) and fractions were checked for the presence of protein using Bradford reagent. Protein was assessed by a 15% sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) (**Fig. 2.5**).

2.4.2.3 Cleaving of His-Tag from RHAU53-105 protein

Cobalt affinity column fractions containing (RHAU53-105 and His-Tag) were pooled together, concentrated to 750-1000 µg/mL followed by the addition of thrombin (GE Healthcare) at a concentration of 1 µL thrombin/1 mg protein and dialyzed in dialysis buffer (10 mM HEPES, pH 7.5, 154 mM NaCl) at 20 °C overnight to cleave the His-Tag.

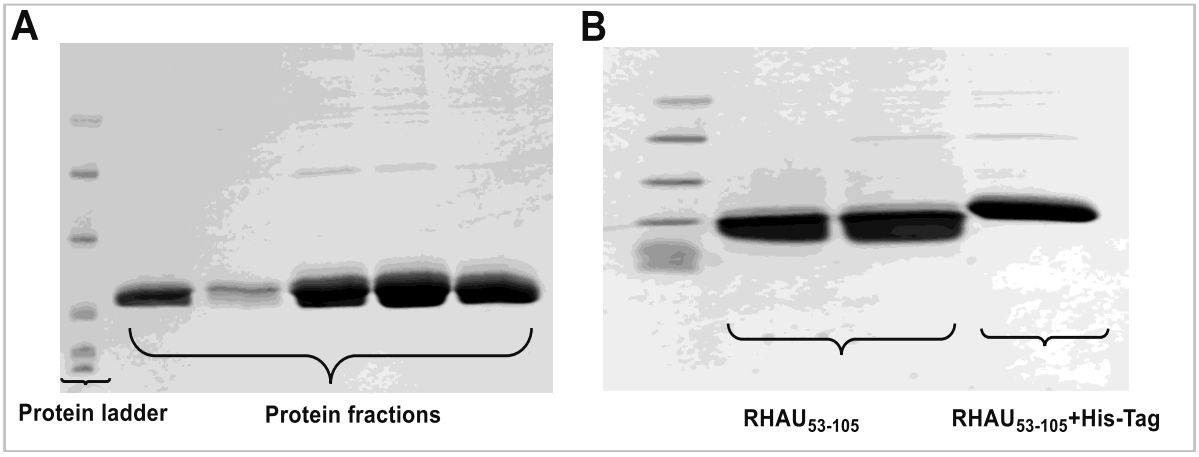

Fig. 2.5. Purification steps of RHAU53-105. (A) SDS-PAGE showing protein fractions after cobalt resin affinity column purification. **(B)** SDS-PAGE showing RHAU53-105 before and after the removal of the His-Tag.

The removal of the His-Tag was performed by loading the cleaved and dialyzed protein on a pre-equilibrated 1 mL HiTrap Benzamidine FF column (GE Healthcare) and eluted with dialysis buffer

(10 mM HEPES, pH 7.5, 154 mM NaCl), followed by the removal of thrombin with low pH elution buffer (50 mM glycine, pH 3.0, 500 mM NaCl). Pure eluted protein was assessed on a 15% sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) (**Fig. 2.5 B**).

2.4.2.4 Preparation of Isotopically enriched ^{15}N-labelled RHAU53-105

Isotopically enriched ^{15}N-labelled RHAU53-105 was overexpressed in M9 minimal media prepared using the following recipes: 5 x M9 media, pH 7.0 (64g $Na_2HPO_4.7H_2O$, 15g KH_2PO_4, 2.5g NaCl, 2g ^{15}N-NH_4Cl) were dissolved in sterilized deionized water and filtered (sterile), followed by the addition of 20 mL of 20% glucose (sterile filtered), 20 mL of 1 M $MgSO_4$ (sterile filtered), 100 µL of 1 M $CaCl_2$ (sterile filtered) and the solution was made up to 1 liter with sterilized deionized water. A 25-mL starter culture of BL21 (DE3) transformed with plasmid in LB was grown overnight in the presence of 100 µg/mL ampicillin at 37 °C, followed by the dilution of starter culture: M9 media in the ratio (1: 100) with 100 µg/mL ampicillin and grown for 5-6 hours at 37 °C to obtain an OD of 0.8. The culture was induced with 1mM IPTG and grown further for 6 hours to express protein and the isotopically enriched protein was purified as described above for RHAU53-105.

2.5 G4-RHAU53-105 complex preparation

G4-RNAs were diluted to 10 µM in (10 mM Tris, pH 7.5, 100 mM KCl, 1 mM EDTA) and heated at 95 °C for 5 minutes and then allowed to cool slowly to room temperature. RHAU53-105 was also diluted to 10 µM in (10 mM HEPES, pH 7.5, 154 mM NaCl). G4 and protein were mixed in an equimolar ratio, and agitated slowly on a rotator for 15 minutes at room temperature and the complexes formed were separated from individual components by SEC on a HiLoad Superdex 75 26/60 column (GE-Healthcare, Mississauga, Canada) in 10 mM Tris, pH 7.5, 100

mM KCl, 1 mM EDTA, and fractions were collected under the peak area. RNA-protein complex in the fractions were tested using an in-line spectrophotometer (260/280 nm simultaneously) and confirmed by native gel electrophoresis following the purification. Fractions containing RNA-protein complex were pooled together and concentrated in 50 mL Millipore concentrators (molecular weight cut-off, 3000; Millipore). The concentration of the complex was determined by UV absorption using the extinction coefficient ε_{260nm} of the RNA, since the nucleic acid dominates the spectrum.

2.6 Instrumentation and Methods

2.6.1 Electrophoretic Mobility Shift Assay (EMSA)

EMSAs were carried out by combining RNA (100-150 nM) with increasing concentrations of RHAU or RHAU53-105 (0-1000 nM) in RHAU buffer [50 mM Tris-acetate (pH 7.8), 100 mM KCl, 10 mM NaCl, 3 mM $MgCl_2$, 70 mM glycine, 10% glycerol], followed by incubating the reaction at room temperature for 15 minutes. Native load dye [0.02% bromophenol blue, 0.01% xylene cyanol FF, 10% glycerol in 1 x Tris/Borate/EDTA (TBE)] was added, mixed thoroughly and RNA-protein complexes were resolved by native 15% polyacrylamide gels (29:1 acrylamide: *bis* ratio) in 0.5 x TBE at 80 V, 4 °C for 2 hours. At the completion of electrophoresis, gel staining was with SYBR Gold fluorescent nucleic acid dye (Invitrogen, Burlington, ON), and imaged on a Fluorchem Q imaging device using Cy2 excitation LEDs and emission filters (ProteinSimple, San Jose, California). The quantification of bands from three independent experiments was achieved using Alpha View-FlorChem Q software provided by the manufacturer.

2.6.2 Microscale thermophoresis (MST)

Binding reactions were prepared in 50 mM Tris-HCl buffer (pH 7.8) with 150 mM NaCl, 10 mM $MgCl_2$, 0.5% glycerol and 0.05% Tween to a total volume of 20 µL. RHAU53-105 was diluted 16 times by 2:1 serial dilution to achieve concentrations ranging from 250–0.6 nM, and mixed with fluorescent 3′-FAM labeled G4-RNA (Integrated DNA Technologies, Coralville, Iowa) was held constant at 25 nM. Premium coated capillaries (NanoTemper Technologies, San Francisco, CA) were used for all measurements. Measurements were performed at an LED power of 90% and MST-IR power of 40% on the Monolith NT.115 instrument under room temperature conditions (21.5 °C). For each run the infrared laser was applied for 35 seconds and the reverse T-Jump data signals of the MST-traces were fitted using the law of mass action (equation 2.6) for 1:1 binding from which the K_d values were obtained (303).

$$FB = \frac{[\mathrm{A}] + [B] + Kd - \sqrt{([A] + [B] + Kd)^2 - 4[AB]}}{2[\mathrm{B}]} \qquad (Equation\ 2.6)$$

K_d: dissociation constant to be determined, [A]: concentration of titrated molecule, RHAU53-105, [B]: constant concentration of fluorescent G4-RNA (303).

$$FB = \frac{F_o - F}{F_o - F_1} \qquad (Equation\ 2.7)$$

Where; F_o: Unbound, F_1: Bound, F: Variable (individual 'y' average) (304)

2.6.3 UV-Visible Studies

UV/VIS spectra were collected on a dual beam Evolution 260 Bio UV-Visible spectrophotometer (Thermo Scientific) using quartz cuvettes (Hellma Analytics, Concord, ON, Canada) with an optical pathlength of 1 cm and absorbance spectra recorded in the 200-350 nm range at a scan speed of 600 nm/min and a data interval of 1 nm. G4-RNA and G4-DNA were prepared at a concentration of 2 μM in 10 mM Tris (pH 7.5), 100 mM KCl, 1 mM EDTA and measured in triplicate with background corrected against spectra of buffer alone. Thermal difference spectra (TDS) were generated by subtracting buffer-corrected spectra at 20°C from those at 90°C. To facilitate direct comparison of spectral shapes between G4-RNAs and G4-DNAs, differences were normalized to the maximum observed absorbance value, as previously suggested by Mergny *et. al* (305). NMM titrations were performed in UV flat bottom microtiter® plates (Thermo Electron Corporation, Weaverville, NC, USA) at 25°C by stepwise addition of 0.2 to 0.3 mM G4 RNA into a solution containing 1.8 μM NMM, ensuring constant porphyrin concentration (306).

2.6.3.1 Binding constant determination

2.6.3.2 Scatchard analysis

The titration data obtained from the interaction between G4 RNA and NMM were used in the construction of Scatchard plots (r/C_f versus r) which allows the determination of binding parameters as shown in the following equation: $r/C_f = K_a (n - r)$. K_a represents the equilibrium binding constant, C_f is the concentration of free ligand, n is the number of ligand molecules bound per G4 RNA, and r is a binding ratio that corresponds to C_b/[G4 RNA]. C_b, representing the concentration of bound porphyrin was determined by first obtaining the fraction of NMM bound (β) at each G4 RNA concentration according to the formula:

$$\beta = \frac{A_{free} - \mathrm{A}}{A_{free} - A_{bound}},$$

where A_{free} and A_{bound} are the absorbances of free and fully bound porphyrins at the Soret maximum for free NMM, 378 nm. The concentrations of bound and free porphyrins were calculated from the equations: ($C_b = C_t$ x β), and ($C_f = C_t - C_b$) where C_t represents the total porphyrin concentration. The binding constant was initially determined by direct fitting of the titration data to a one site binding hyperbola using the equation;

$$Y = \frac{B_{\max} * \mathrm{X}}{(K_d + X)} \quad ,$$

Y: concentration of bound ligand, X: concentration of free ligand, B_{max}: maximum Specificity binding, K_d: the dissociation constant.

A linearization of data from saturation binding isotherms using Scatchard analysis provides the best possible means of obtaining estimates for K_d and B_{max}. All data manipulations were done using GraphPad Prism software 7.0.

2.6.4 Circular dichroism spectropolarimetry (CD)

2.6.4.1 Principle

Biomolecules that possess molecular asymmetry (mirror images not identical) are named to be chiral (307). The interaction of a chiral molecule with polarized light (**Fig. 2.6**) is highly specific and plays a very important role in the characterization of both small molecule and macromolecular structures (308). The measurement commonly made to determine the effects of polarized light on chiral molecules is known as circular dichroism (CD) and is defined as the difference in absorption of left-hand and right-hand circularly polarized light. Although, these

measurements are relatively small, however, they can be measured rapidly and accurately with modern instrumentation to produce structural information on carbohydrates, nucleic acids, proteins, pharmaceuticals, *etc*. (308). Circular dichroism can be used to study conformational changes in the structure of macromolecules and consequently, the unfolding of biomolecules such as proteins, nucleic acids, glycosides can be measured as a change in CD spectra. A G4 stem is made up of stacked guanosines that could adopt an *anti* or a *syn* glycosidic bond angle (GBA) (309) and G4s can be grouped according to their GBA sequence within a G4 stem (310). The three groups of G4 topologies described earlier are sensitive to CD signals, attributed to well-isolated guanine bands in the 240-290 nm region and connected to well characterized π-π^* transitions (311, 312).

CD spectroscopy has numerous advantages: it is a simple, fast and relatively cheap method when compared to NMR and X-ray crystallography. CD is an extremely sensitive technique that requires small amounts of sample and allows the titration of other samples under various conditions of pH, temperature, organic solvents and salt types. The shape of the spectrum, including positive maxima and negative minima can provide diagnostic information about biomolecules.

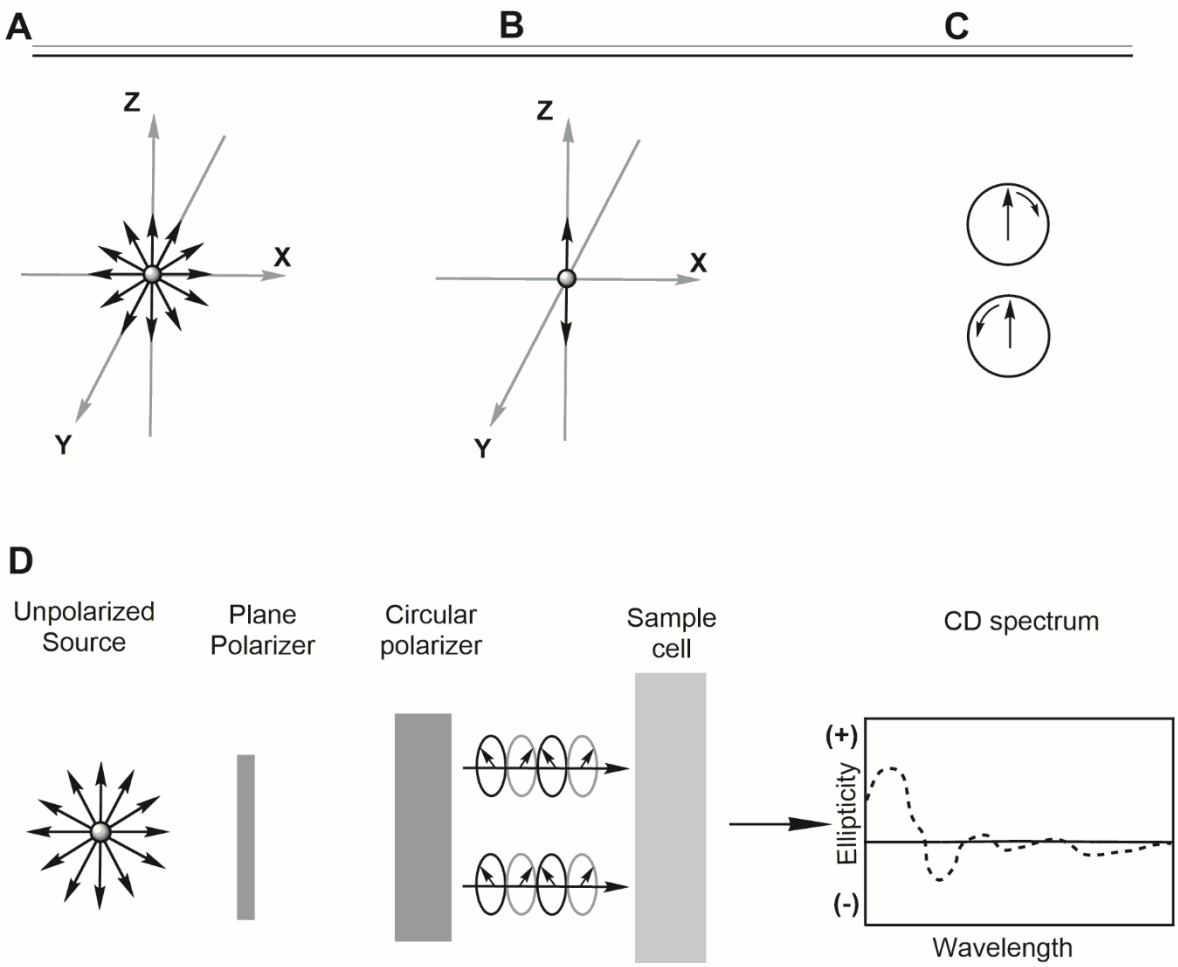

Fig. 2.6. Schematic diagram showing the Circular Dichroism technique. The electric field components **of (A)** unpolarized light; where the electric field oscillates in all directions **(B)** plane or linearly polarized light; where the electric field oscillates in one direction **(C)** circularly polarized light; where the direction of rotation can be clockwise or anticlockwise. **(D)** Schematics representing how a difference in absorption is measured in circular dichroism (307).

2.6.4.2 Sample preparation and experimental specifications

All spectra were collected on a calibrated Alfa Aesar J-810 spectropolarimeter (Jasco Inc., USA) between 200-340 nm in a 1.0 mm pathlength and a 32 s integration time. The concentration of the samples was kept at 20 µM in 10 mM Tris, pH 7.5, 100 mM KCl, 1 mM EDTA and

measurements were performed in triplicate with baseline-corrected by subtraction of the buffer alone. Circular dichroism thermal melting was also performed with curves generated in the same buffer, following the ellipticity at 262 nm with spectra normalized by the number of nucleotides (glycosidic bonds) per unit volume.

2.6.5 Dynamic Light Scattering (DLS)

Dynamic light scattering (DLS) is a technique used to determine the size distribution profile of macromolecules in solution (313). This technique was employed in my study to ensure monodispersity of sample over a range of concentration and DLS data were collected using a Nano-S Dynamic Light Scattering system (Malvern Instruments Ltd., Malvern, UK) as described previously (314). Samples were filtered through a 0.1 μm filter (Millipore) and 5-minute equilibration at 20 °C prior to measurement. In all, 15 measurements were made per sample at different concentrations in 10 mM Tris (pH 7.5, 100 mM KCl, 1 mM EDTA], and three independent samples were tested for each condition. The size of macromolecule determined by DLS corresponds to its hydrodynamic radius (r_H), which depends on both the size/shape of the molecule and its interaction with the solvent.

2.6.6 Small Angle X-ray Scattering (SAXS)

2.6.6.1 Principle

Small Angle X-ray Scattering (SAXS) is a small-angle scattering (SAS) technique in which the elastic scattering of X-rays (wavelength 0.1 to 0.2 nm) by a sample having inhomogeneity in the nm-range, is recorded at a very low angle (typically 0.1 to 10°). This angular range contains information about the average shape and size of macromolecules, characteristic

distances of partially ordered material and pore sizes (315, 316). Because SAXS is a solution technique, some key assumptions are made to analyze the experimental data obtained. One of these assumptions is that the system under consideration is isotropic (sample is uniform in all orientations), which could be as a result of the rotation of the particles in solution. The second assumption is that there is no correlation between two scattering centers (electrons) that are present in two different molecules (316, 317). X-rays are an electromagnetic wave of high energy and very short wavelength and when they hit electrons, the electrons resonate with the frequency of the X-ray and this results in the emission of coherent secondary waves that interfere with each other either constructively (waves in phase) or destructively (waves out of phase). A coherent (elastic) scattering, which contributes almost entirely to the scattering data is produced when the secondary waves generated have the same wavelength as that of the original X-ray (316). Because incoherent (inelastic) scattering is weak and does not interfere with the elastic scattering at small angles, it is often ignored. The detector is used to obtain information on the distance between the two electrons based on their intensities and amplitude.

Phase difference information is obtained from the amplitude which in-turn depends on the path difference of the two rays. The schematic representation of the major steps involved in SAXS data collection, processing and *ab initio* modelling is shown in (**Fig. 2.7**). A software known as PRIMUS (318) is used in the conversion of a scattering pattern into a SAXS scattering curve, both of which are in reciprocal space. The curve shows the log of intensity $I(q)$ of the scattered X-ray plotted against the momentum transfer (s). Momentum transfer (s) is defined by the equation below;

$$(s) = 4\pi\theta/\lambda \qquad \text{(equation 2.8)}$$

Where θ is the scattering angle and λ is the wavelength of the X-ray.

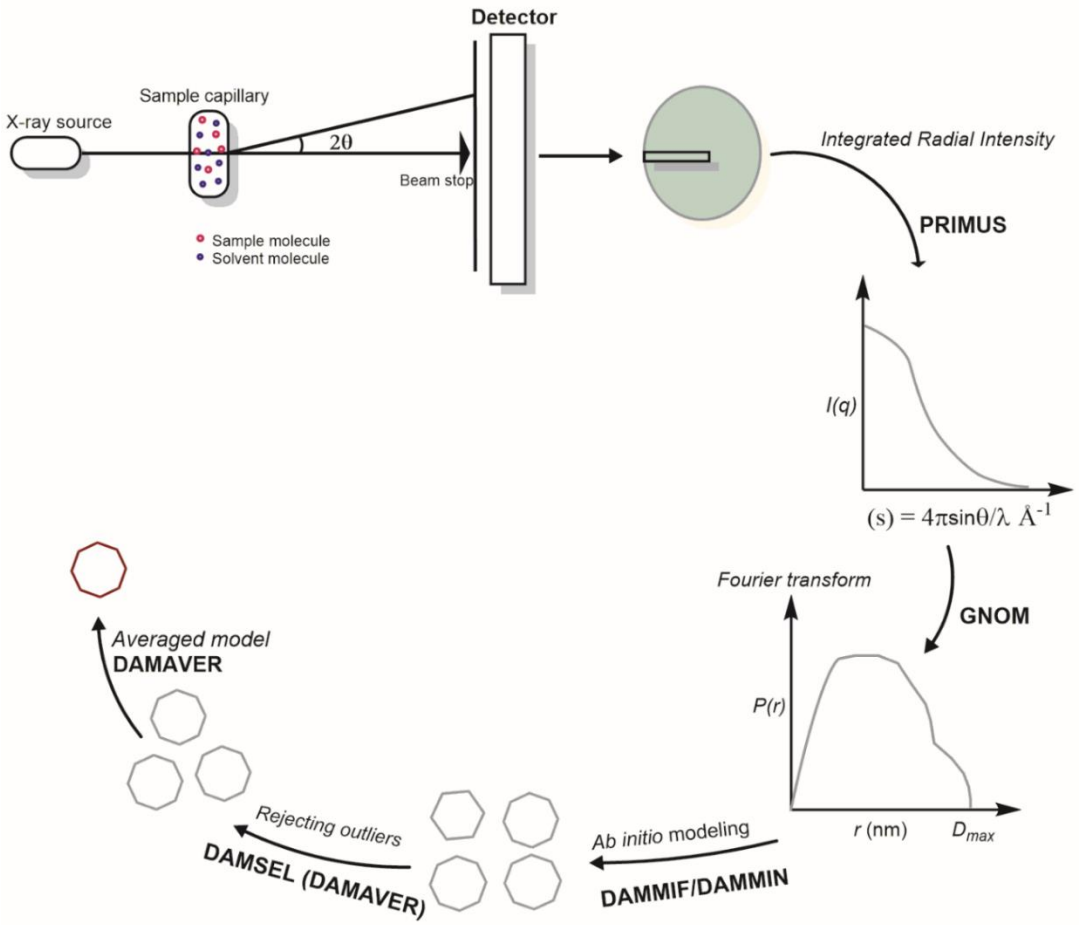

Fig. 2.7. Schematics of small angle X-ray scattering technique. Collimated X-ray beams (Cu Kα radiation) are scattered by the sample, and these rays hit the detector which are registered as a circular scattering pattern given in reciprocal space. The radial intensity of the circular pattern is integrated to obtain the SAXS scattering curve, momentum transfer (s) plotted against log of intensity *I*(*q*). A pair distribution function curve *P*(*r*) in real space is obtained from the scattering curve with the X-axis representing the maximum particle dimension (*Dmax*) and radius of gyration is determined in this step. Building of *ab initio* models is made from the *P*(*r*) plot from which outlier models are rejected and an averaged model is obtained. All the programs used in data analysis and ab initio modelling are in capital letters (315, 319).

When the value of θ is large, the destructive interference (secondary waves emitted from all electrons in all possible phases) will lead to no scattering with directionality of θ. However, at

smaller scattering angles, the phase difference between the scattered waves will be smaller (limiting destructive interference) hence, scattering is enabled in the direction of the small angle to be recorded on the detector. There would have been no interference at zero angle, but data cannot be obtained at that angle because the beam-stop is located at that position to protect the detector from direct exposure to the incident X-ray (316, 320).

The pair distance distribution function, $P(r)$, in real space is generated from the scattering data obtained in reciprocal space. $P(r)$ is a histogram of the pair distances between electrons inside of the biomolecule that can be used to ultimately determine low-resolution structural features (321). The GNOM software program is used to generate the pair distribution function curve, $P(r)$, by an inverse Fourier transform which de-convolutes distance information on electron pairs (321, 322). Considering (**Fig. 2.8**), the probability of finding (j) at a distance (r) from point (i) is $\gamma 0(r_{ij})$, and the number of possible (i) is directly proportional to the volume (v) of the molecule; where the number of possible (j) is directly proportional to r_{ij}^2. The number of electron pairs (i, j) with a distance r_{ij} between them is directly proportional to $\rho^2\ \gamma 0(r_{ij})\ Vr^2$ and the pair distribution is given by;

$$P(r) = r^2\ \gamma 0(r)\ V\rho^2 = r^2\ \gamma(r) \qquad \text{(equation 2.9) (319)}$$

The radius of gyration (r_G), which is defined as "the mean square distances to the centre of mass weighted by the contrast of electron density" (319) an index of non-sphericity, is obtained from GNOM software which uses a Guinier plot to obtain a straight-line fitting from the slope; $-(r_G^2/3)$.

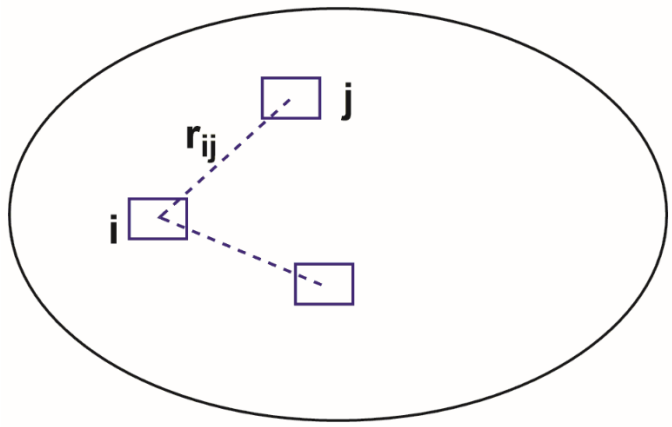

Fig. 2.8. Schematic showing an electron pair used in the explanation of P(r) curve. The electron (j) at a distance (*r*) from the reference electron (i) is within a molecule with volume (*V)*; and the density of electrons in the molecule is given by (*ρ).*

The maximum particle dimension (D_{max}), representing the distance from the origin to the point where the curve intercepts the X-axis again is also obtained by GNOM. The r_G is given by the equation;

$$r_{G^2} = \frac{\int_{Vr} \Delta\rho(r)\, r^2 dVr}{\int_{Vr} \Delta\rho(r)\, dVr} \qquad \text{(equation 2.10) (319)}$$

Where; *V* = *volume* of the particle or sample

$\Delta\rho(r)$ = contrast in electron density between the sample and solvent

(r) = distance from the reference point (i) to a particle or an electron at another point (j)

(r^2) = the square of distance between an electron pair.

The purpose of *ab-initio* modelling of SAXS data is to generates a three-dimensional structure from a one-dimensional scattering pattern, where the particle shape is represented by a parametric model, altered to minimize the difference between the computed scattering of the model and the experimental data (323). In *ab-initio* modelling, a particle, (usually a sphere of radius 'R') having an X-ray scattering pattern similar to that of sample is filled with 'N' number of densely packed small spheres with radius 'r' (dummy atoms). All the beads representing the particles (particles

under investigation) must be interconnected to form a single body; this modelling approach is known as bead-modelling and discussed by Svergun *et al.* (320, 324). The programs use energy minimization through a simulated annealing protocol (325) which involves heating of the system and allowing a random modification, followed by cooling down the system and confirming if the modification gave rise to an energy decrease of the system. DAMMIF and DAMMIN, which are *ab initio* modelling software use r_G and D_{max} to provide constraints for the models generated (320, 324). The conformity of the predicted scattering of the model molecules after the periodic energy minimization step is compared to the experimental scattering profile and this is determined by a parameter known as chi (χ). A program known as DAMAVER is used to align *ab initio* low resolution models generated by DAMMIN, selects the most probable one and builds an averaged model where the level of agreement of the filtered model to the individual *ab initio* model is determined by normalized spatial discrepancy (NSD). NSD is a quantitative measurement of the alignment between two models where each model is represented by a set of dummy atoms (326).

2.6.6.2 Sample concentrations and experimental specifications

SAXS data were collected using a Rigaku 3-pinhole camera (S-MAX3000) equipped with a Rigaku MicroMax + 002 microfocus sealed tube (Cu-Kα radiation at 1.54 Å) and Confocal Max-Flux (CMF) optics operating at 40W as previously reported by Dzananovic *et al.* (327). Scattering data were collected at the following sample concentrations; Q2RNA and Q2DNA (0.8, 1.1, and 1.7 mg/ml), and Q2RNA/RHAU53-105 complex (1.3 and 1.5 mg/mL), Q3RNA and Q3R_M2 (3.1 and 2.8 mg/mL) in 10 mM Tris pH 7.5, 100 mM KCl, 1 mM EDTA. The raw intensity data were integrated with the SAXSGUI software package (JJ X-Ray Systems A/S, LyngBy, Denmark). Buffer subtraction and merging of data at multiple concentrations were performed

using the program PRIMUS (318). The pair distance distribution function plot, root mean square radius of gyration (r_G) and the maximum particle dimension (D_{max}) were obtained using the program GNOM (322). *Ab initio* shape modeling was performed using the program DAMMIF based on a simulated annealing protocol (320, 323). Twenty models for each entity were then generated, rotated, aligned and averaged using the program DAMAVER (328). To calculate the solution hydrodynamic properties of the averaged-filtered models, HYDROPRO (329) was employed following a similar approach outlined by Dzananovic *et al.* (327). Sample quality was confirmed for each sample before and after data collection by gel electrophoresis and DLS.

2.6.7 Nuclear Magnetic Resonance (NMR) Spectroscopy

Nuclear magnetic resonance spectroscopy (NMR) is a useful tool in the study of G4 structures and their complexes with interacting proteins. This technique has been used to determine the atomic resolution structures, stability and dynamics, as well as intermolecular interactions in G4 structures (330). In proton NMR, the formation of a G-tetrad produces characteristic guanine imino protons (H1), which display chemical shifts within the range of 10-12 ppm (331) in contrast to 13-14 ppm produced by Watson-Crick base pairing (332). Heteronuclear single quantum correlation spectroscopy (HSQC), a 2D-NMR frequently used in the field of protein NMR, is an experiment that provides the correlation between nitrogen and amide protons where each amide group produces a peak in the HSQC spectra. Protein interactions with nucleic acids or ligands can be investigated by comparing the HSQC of free protein with the one in bound state.

2.6.7.1 Principle

The physical foundation of NMR is based on the magnetic properties of atomic nuclei. Odd mass nuclei (odd number of protons and neutrons) possess fractional spins: I = ½ and the following are some isotopes that possess magnetic moments (μ); ^{1}H, ^{13}C, ^{14}N, ^{17}O, ^{19}F, ^{31}P. A spinning charge generates a magnetic field and the resulting spin-magnet has a magnetic moment (μ) proportional to the spin (equation 2.11) and in the presence of an external magnetic field (B_o), two spin states are possible: a lower energy orientation having the magnetic moment aligned parallel with the applied field; and a higher energy anti-parallel orientation (333).

$$\mu = \gamma \, I \, h/2\pi \qquad \text{(equation 2.11) (333)}$$

where (μ) is the magnetic moment of the nuclei, (I) is the spin number, (γ) is a constant known as magnetogyric ratio which relates the magnetic moment (μ) and the spin number (I) for any specific nucleus, (*h*) is the Planck's constant.

The energy difference between two spin states depends on the external magnetic field strength and have the same energy when external field is zero. Irradiation of a sample with radio frequency energy that corresponds exactly to the spin state separation of specific set of nuclei will cause an excitation (resonance) which is measured. In an NMR experiment, energy in the form of electromagnetic radiation is used to excite most of the protons to be in opposite alignment with an external magnetic field (B_o). When the external energy is removed, the energized nuclei relax back to the low energy state and the fluctuation involved with this relaxation is called resonance which can be detected and converted to peaks via Fourier transformation. The relaxation pattern of protons differs, depending on the chemical environment: chemical shift is the resonant frequency of a nucleus relative to a standard and the position and number of chemical shifts are diagnostic of the structure of a molecule. In heteronuclear single quantum correlation spectroscopy (HSQC),

the transfer of magnetization from a proton to a heteronuclear atom (^{15}N or ^{13}C) through an insensitive nuclei enhanced by polarization transfer (INEPT) (334) step is made and after a delay time (t_1), the magnetization is transferred back to the proton via retro-INEPT step which produces signal that can be detected.

2.6.7.2 Sample concentrations and experimental specifications

Q3RNA or $Q3M_2$ (100 μM), and Q2RNA (85 μM) in complex with ^{15}N RHAU53-105 in 10 mM Tris, (pH 7.5), 100 mM KCl, 1 mM EDTA, and 10% deuterium oxide (v/v) were prepared in an identical manner to that previously described by Meier *et al.* (238). All spectra were acquired on a Varian Unity INOVA 600 MHz spectrometer. Data processing and spectrum generation were performed using iNMR (http://www.inmr.net).

CHAPTER 3

Biophysical Characterization of G-Quadruplex Recognition in the PITX1 mRNA by the Specificity Domain of the Helicase RHAU

The results presented here have been published in a peer-reviewed journal: E.O. Ariyo, E.P. Booy, T.R. Patel, E. Dzananovic, E.K. McRae, M. Meier, K. McEleney, J. Stetefeld and S.A. McKenna (2015) Biophysical Characterization of G-Quadruplex Recognition in the PITX1 mRNA by the Specificity Domain of the Helicase RHAU. PLoS ONE 10(12): e0144510.

AUTHOR CONTRIBUTIONS

EOA the performed research, analyzed the data, and wrote the manuscript. EPB performed and designed the research. TRP analyzed data and edited the manuscript. EDZ analyzed data. EKM performed research (MST). MM analyzed data and edited the manuscript. KM provided SAXS technical support and sample acquisition. JS designed the research. SAM designed the research and edited manuscript.

3.1 Introduction

G4s found in RNA and DNA have been shown to adopt parallel, anti-parallel, or hybrid (mixture of parallel and antiparallel) strand orientation (17). A combination of biophysical studies and high resolution structures of G4-RNA revealed that they are thermodynamically more stable *in vitro* than their DNA counterparts under near-physiological conditions due to the presence of 2′-OH group on the ribose sugar enhancing the formation of hydrogen bonds, and consequently promoting the parallel conformation of G4-RNA over an antiparallel one (335-337). An evolutionary conservation survey of RNA and DNA motifs showed that G4 motifs are highly conserved in the genomes of living organisms (22, 302, 338, 339), and it was recently demonstrated that G4 formation is regulated dynamically during cell-cycle progression (100, 340). Accumulating evidence suggests an important role of G4 structures in regulating gene expression (302). Genome-wide computational analysis has identified more than 300,000 potential intramolecular G4-forming sequences in the human genome (22, 341) and revealed a higher

prevalence of these sequences in functional genomic regions such as telomeres, promoters (302, 342), untranslated regions (UTRs) (98, 343) and introns (344). Taken together, these observations suggest that G4 structures participate in regulating myriad biological processes.

DNA G4 recognition and remodeling by helicases such as Fanconi anaemia group J protein (FANCJ), Bloom syndrome protein (BLM), DNA repair protein (REV1) and Werner's syndrome protein (WRN) have been reported (345-348). RNA Helicase Associated with AU-rich element (RHAU, DHX36, G4R1) is a member of the human ATP-dependent DEAH-box family of RNA helicases, although DNA G4 helicase activity has also been observed with this enzyme (349, 350). RHAU uses a local, non-processive mechanism to unwind G4s, similar to that of eukaryotic initiation factor 4A on double-stranded substrates (272, 351). RHAU has nanomolar to sub-nanomolar affinity for G4s, and orders of magnitude weaker affinity for other observed nucleic acid conformations (49, 50, 238, 273, 274, 352). Furthermore, RHAU has 100-fold higher affinity for parallel relative to non-parallel G4s (272-274). Helicase activity is highly sensitive to G4 stability, with an inverse correlation observed (272). Based on domain conservation with other helicases, RHAU's core DEAH-box helicase domain (residues 210-614) is flanked by an N-terminal G4-recognition domain (residues 1-210) and C-terminal helicase associated domains (residues 670-1008) that have yet to be fully characterized (270). G4 specificity is mediated by the RHAU-specific motif (RSM), a 13-residue stretch (residues 54-66) in an N-terminal subdomain that is necessary, but not sufficient for full G4 binding affinity (50, 253). A truncation of the full-length protein, RHAU53-105, adopts a defined and extended conformation in solution, orienting the RSM at one end (238). RHAU53-105 retains both nanomolar affinity for G4s *in vitro* and the ability to outcompete endogenous RHAU for G4 targets in a cellular context (49, 238). Investigation of RHAU53-105 in complex with RNA G4 from human telomerase RNA by

NMR, SAXS, and complimentary biophysical approaches suggested interaction with the G-tetrad face (as opposed to the sugar-phosphate backbone) as the recognition surface for the RSM (238). This mode of recognition was partially supported by a recent high-resolution structure of an 18 amino acid peptide ($RHAU_{53-70}$) in complex with a parallel DNA G4 showing 4 amino acids (G59, I62, G63, A67) mediating interaction on the G-tetrad faces (274). While the structure also suggests electrostatic interactions with the phosphate backbone, a previous study demonstrated no significant impact of charged amino acid mutations in the RSM (253). Differences between DNA and RNA G4 binding by RHAU have also previously been reported, likely owing to the conformational restraints imposed by the 2'-OH in RNA (238, 273, 274).

To expand our understanding of biologically relevant RNA G4 recognition by the helicase RHAU, we previously performed an RNA co-immunoprecipitation screen and identified the messenger RNA (mRNA) for the protein Pituitary homeobox 1 (PITX1, P-OTX, backfoot) (49). PITX1 functions as a transcription factor that plays a pivotal role in the differentiation of the developing pituitary gland, craniofacial structures and hind limbs in early embryonal development (289-292). Recently, malformations in the lower limbs could be attributed to mutations in the PITX gene (293). Deletions in PITX1 cause a spectrum of lower-limb malformations including mirror image polydactyly. PITX1 expression is down regulated in a number of tumor types including lung, colorectal, gastric and esophageal cancer and reduced PITX1 expression has been correlated with decreased overall patient survival (294-296). Most interestingly, the PITX1 mRNA possesses three distinct G4 forming sequences in the 3′-untranslated region (UTR) of its mRNA (Q1: $PITX1_{1371-1400}$, Q2: $PITX1_{1901-1930}$, and Q3: $PITX1_{2044-2079}$). These G4s play roles in the recruitment of RHAU to the PITX1 mRNA and ultimately regulate PITX1 protein translation (49). In cell lysates and with purified components, both RHAU and RHAU53-105 can interact with Q1,

Q2, or Q3. Here, we characterize the Q2RNA/RHAU53-105 complex using a combination of electrophoretic mobility shift assays, UV-VIS spectroscopy, circular dichroism, dynamic light scattering, small angle X-ray scattering (SAXS) and nuclear magnetic resonance spectroscopy. Our integrated approach suggests that the RSM recognizes the planar guanine quartet face of parallel RNA G4s. The results presented here are published in, E.O. Ariyo *et al.* (2015) PLoS ONE 10(12): e0144510.

3.2 Results and Discussion

3.2.1 Results

3.2.1.1 Q2RNA and Q2DNA each adopt a single, monomeric conformation

Synthetic Q2RNA was heat denatured, cooled, and purified by size exclusion chromatography (see Materials and Methods). Q2RNA elutes as a compact dominant peak with a shoulder corresponding to larger hydrodynamic volumes (**Fig. 3.1**) from the HiLoad Superdex 75 26/60 column. The nucleic acid sequences used in this study are presented in (**Fig. 3.2 A**). Native gel electrophoresis confirmed that the dominant peak contains a single RNA conformation (**Fig. 3.2 B**). To understand the potential differences between RNA and DNA G4 recognition, we also investigated the DNA equivalent to Q2RNA (Q2DNA). Using an identical procedure, Q2DNA eluted in a single symmetric peak that contains a single conformation as determined by native gel electrophoresis (**Figs. 3.1** and **3.2 B**).

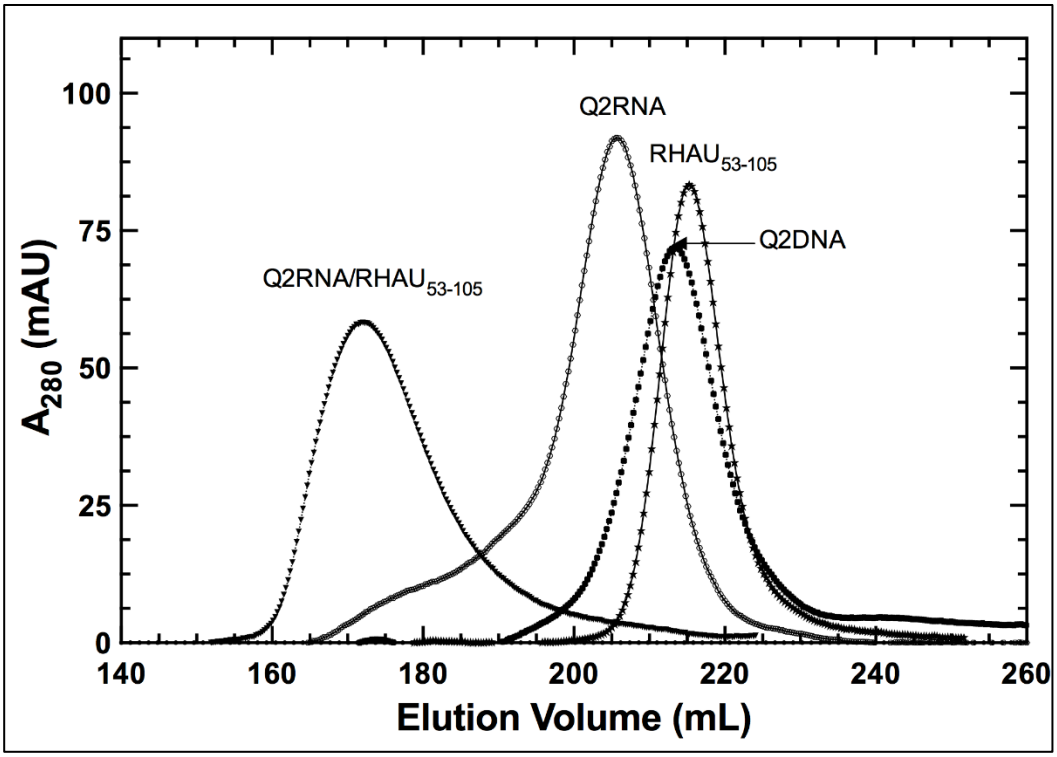

Fig. 3.1. Purification of protein and nucleic acid components. Elution profiles obtained on a HiLoad Superdex 75 26/60 column, with each species labeled.

3.2.1.2 Q2RNA, but not Q2DNA, Stains with a Dye specific for parallel G4

To determine whether the purified nucleic acids adopt a parallel G4 conformation, we employed native gel electrophoresis in combination with N-methyl mesoporphyrin IX (NMM), a dye specific to parallel G4 conformations. The crystal structure of NMM bound to a parallel DNA G4 demonstrates the selectivity for parallel G4s (353). To confirm the validity of the approach, non-G4 double-stranded RNA (dsRNA) and a known parallel RNA G4 (hTR1-17) were included as negative and positive controls, respectively. While both stain efficiently with toluidine blue (nucleic acid stain, pH 2.8), the positive, but not negative control stains with NMM (**Fig. 3.2 B** and **3.2 C**). Q2RNA stains efficiently with NMM, indicating that it adopts a parallel G4

conformation. Interestingly, the Q2DNA does not stain with the dye, suggesting that either it adopts a non-G4 or non-parallel G4 conformation.

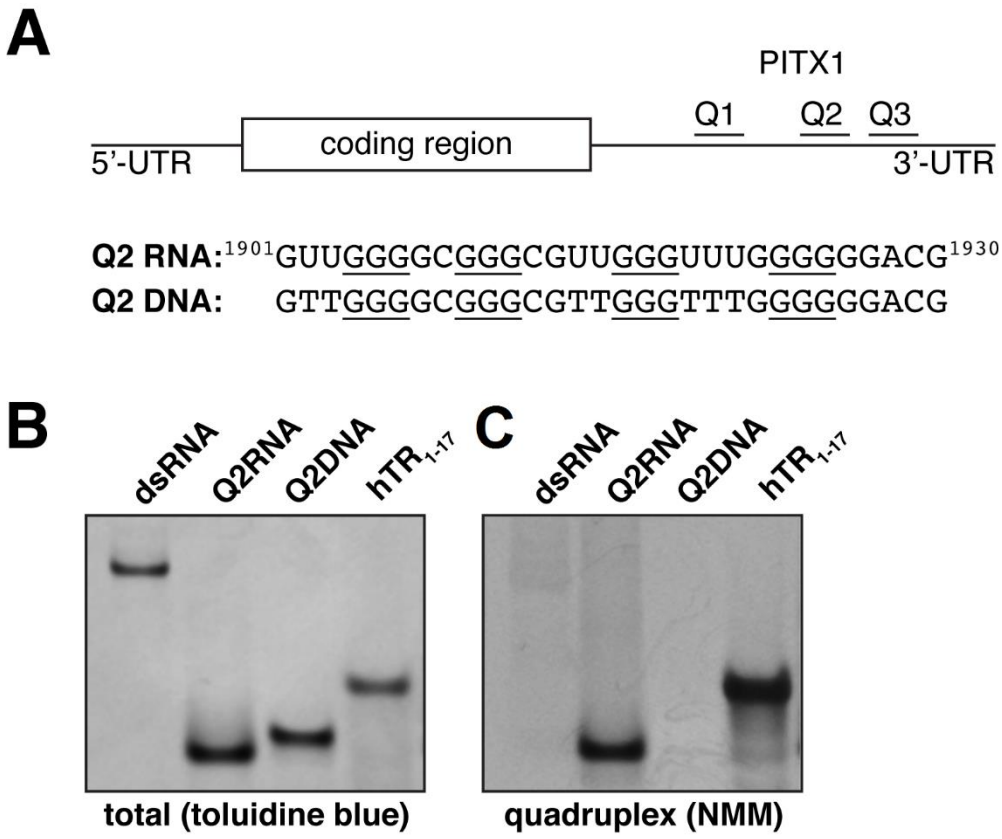

Fig. 3.2. Q2RNA, but not Q2DNA, stains with a parallel G4 dye. (A) Schematic showing the G4 forming regions of PITX1 mRNA; the RNA and DNA equivalent sequences with the guanylate tracts underlined are also shown. 250 pmol of Q2RNA and Q2DNA were separated by Native-Tris-borate EDTA (TBE) polyacrylamide gel electrophoresis, stained with **(B)** toluidine blue, and **(C)** G4-specific dye N-methyl mesoporphyrin IX alongside their positive (G4: hTR_{1-17}) and negative (double stranded RNA) controls.

3.2.1.3 Q2RNA adopts a parallel, while Q2DNA adopts an alternate, G4 conformation

To confirm G4 formation by Q2RNA, thermal difference spectra (TDS) were generated by subtracting the UV absorption spectrum of the folded state (recorded at 20°C) from the

spectrum of the partially denatured state (measured at 90°C). Specific nucleic acid conformations result in specific TDS, generally reflecting the conformational change of the molecule in solution due to a disruption in base-stacking interactions. The TDS obtained for Q2RNA demonstrated features characteristic of G4s (305) with a minimum at 297 nm and two maxima at 240 and 276 nm (**Fig. 3.3**). Interestingly, TDS analysis of Q2DNA showed similar overall features suggestive that it also adopts a G4 structure.

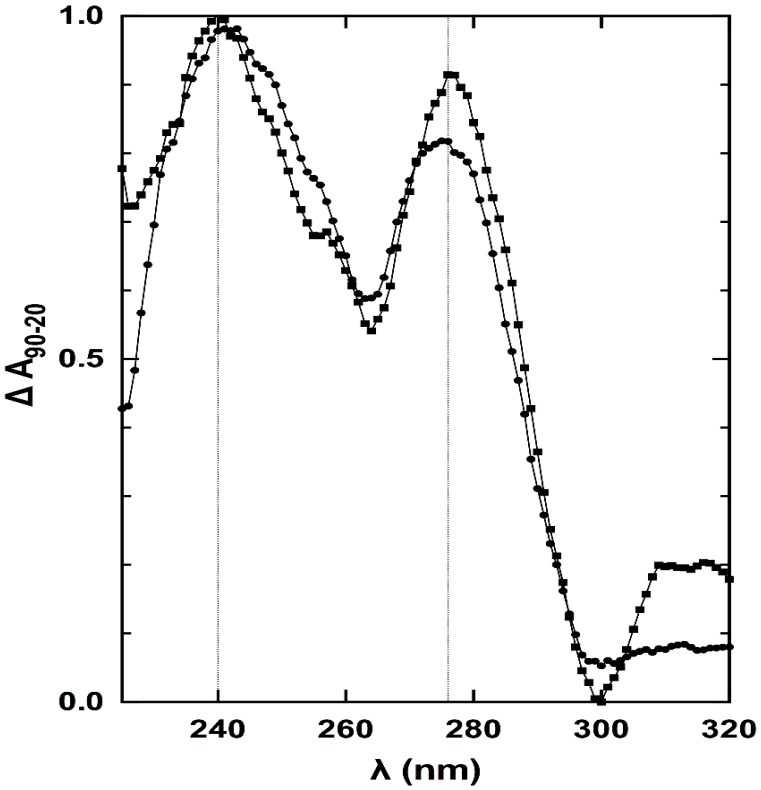

Fig. 3.3. Normalized thermal difference spectra of Q2RNA and DNA counterpart. TDS of Q2RNA (circles) and Q2DNA (squares) in 10 mM Tris, pH 7.5, 100 mM KCl, 1 mM EDTA. *For details of analysis, see Materials and Methods.*

Next, we performed circular dichroism (CD) spectroscopy on the partially denatured and native states at 80°C and 20°C, respectively, in the same buffer as used for TDS analysis. The far-UV

CD spectrum of Q2RNA at 20°C presented features consistent with previously characterized parallel G4s, with an ellipticity minimum at 242 nm and maximum at 264 nm (354) (**Fig. 3.4 A**).

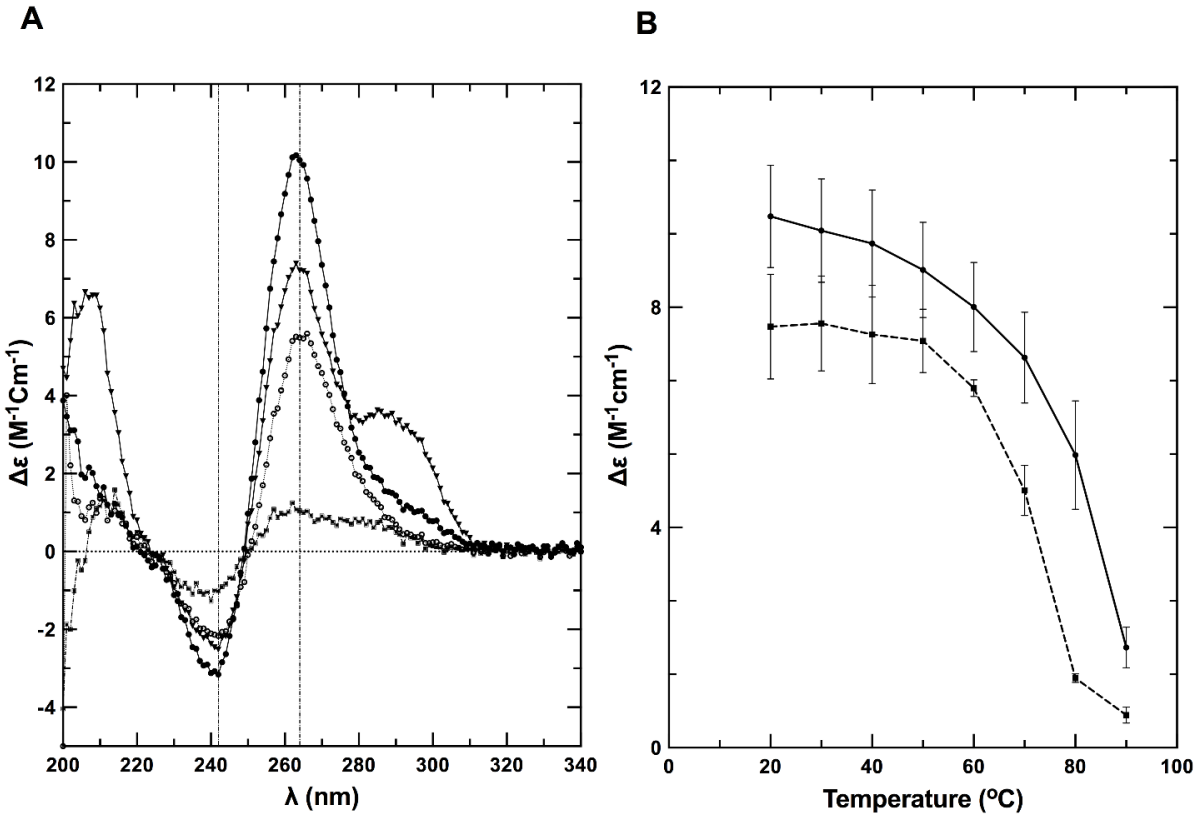

Fig. 3.4. Q2RNA and its DNA counterpart adopt G4 structures. (A) Far-UV CD spectra of Q2RNA and Q2DNA obtained in 10 mM Tris, pH 7.5, 100 mM KCl, 1 mM EDTA buffer of Q2RNA at 20°C (closed circles) and at 80°C (open circles) as well as of Q2DNA at 20°C (closed triangles) and at 80°C (closed squares). **(B)** CD melting curves of Q2RNA (——) and Q2DNA (------) monitored by spectropolarimetry at 264 nm in the same buffer.

At 80°C similar overall spectral features were observed; however, the intensity was modestly muted (approximately 45% at 264 nm), presumably due to partial unstacking of the G4 structure. Q2DNA has similar overall features to Q2RNA at 20°C with the prominent exception of an additional maxima at 290 nm (**Fig. 3.4 A**). The Q2DNA spectrum is consistent with the features of a group II G4 spectrum that has three parallel and one antiparallel strands (73, 148, 355). At

80°C, the Q2DNA is almost completely denatured. To determine the relative stabilities of the RNA and DNA G4s, CD spectra were collected during the process of thermal melting (**Fig. 3.4 B**). Q2RNA was significantly more resistant to denaturation than its DNA counterpart, but the melting profile is similar to that of previously characterized RNA G4s (238). We conclude that Q2RNA adopts, as expected, a parallel G4 conformation, whereas Q2DNA assumes a hybrid-type G4 structure with parallel and antiparallel strands.

3.2.1.4 RHAU interacts with Q2RNA but not its DNA counterpart

Previously, an N-terminal truncation of RHAU (RHAU53-105) containing the RSM has been identified to play a significant role in the recognition of G4s (238, 253). To confirm the original observation, we performed electrophoretic mobility shift assays (EMSA) between Q2RNA and either RHAU53-105 or full-length RHAU (**Fig. 3.5 A**). Both RHAU53-105 and full-length RHAU shift Q2RNA towards a higher molecular weight species in a concentration dependent manner. We observed a higher affinity with full-length RHAU than with the truncated version (as expected). Interestingly, the DNA counterpart, Q2DNA, did not show any appreciable affinity for RHAU53-105. (**Fig. 3.5 A**). Microscale thermophoresis measurements were used to determine a dissociation constant of 1.7±0.3 nM for the RHAU53-105 complex with fluorescently labeled 3'- FAM-Q2RNA (**Fig. 3.5 B**). To further characterize nucleic acid-protein complexes, we prepared pure RHAU53-105 as well as its complex with Q2RNA or Q2DNA, and subjected them to size exclusion chromatography (**Fig. 1**).

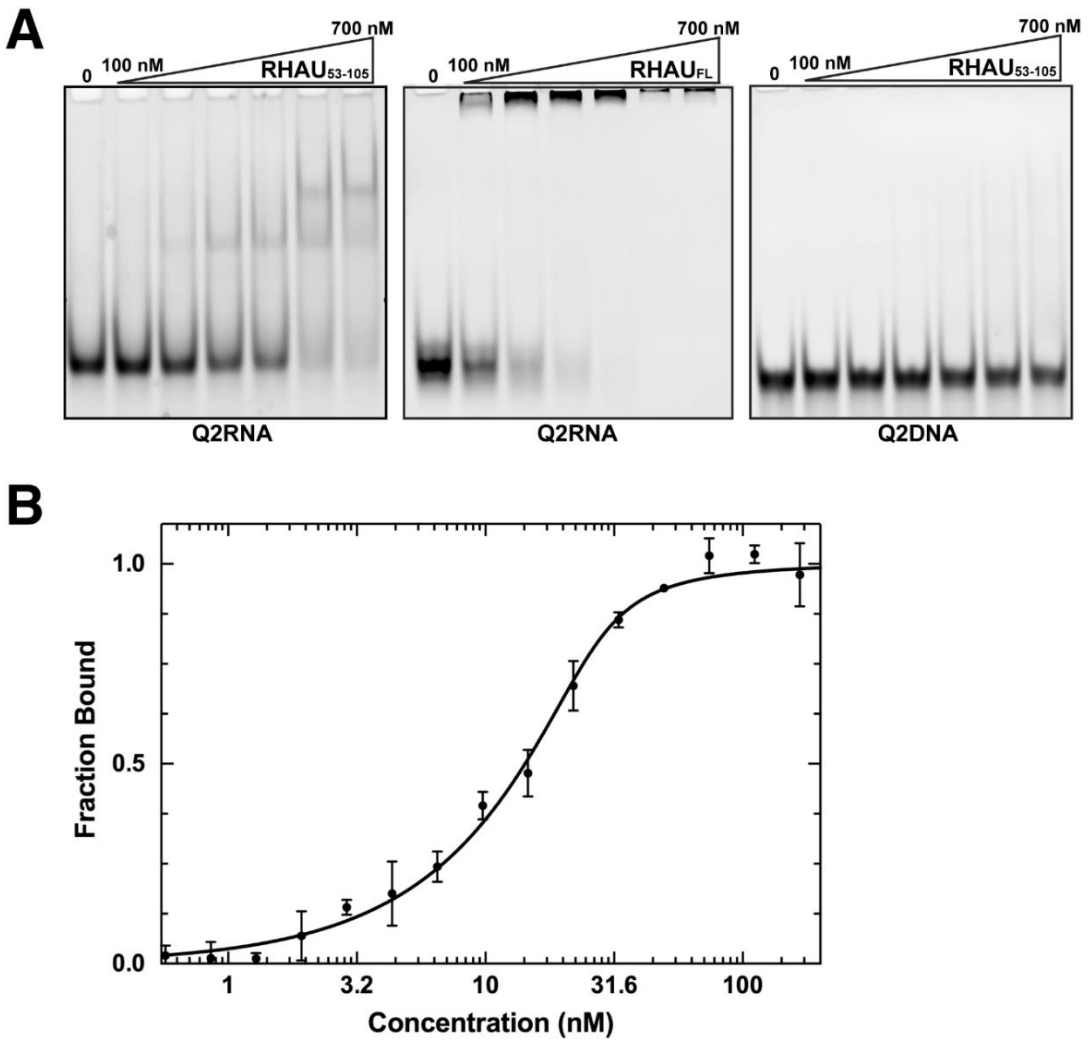

Fig. 3.5. RHAU53-105 forms a complex with Q2RNA. (A) Electrophoretic mobility shift assays (EMSA) were performed using a constant 150 nM concentration of Q2RNA or Q2DNA and a variable concentration from 0–700 nM of RHAU53-105 or full-length RHAU. The 12% native Tris borate-EDTA (TBE) polyacrylamide gels were stained with SYBR Gold for visualization. **(B)** Microscale thermophoresis measurements performed using 3′-FAM Q2RNA (25 nM) in complex with RHAU53-105 at several concentrations (0.6–250 nM). Reverse T-Jump signals from the traces were fit as described in the Materials & Methods.

RHAU53-105 and its complex with Q2RNA eluted as single peaks, with an expected increase in hydrodynamic size accompanying complex formation. Not surprisingly, no complex formation was observed between Q2DNA and RHAU53-105 (data not shown).

3.2.1.5 The association of N-terminal RHAU with Q2RNA does not disrupt G4 Structure

To determine whether RHAU53-105 binding disrupts G4 structure we performed a CD experiment on the purified Q2RNA/RHAU53-105 complex (**Fig. 3.6**). No significant differences were observed between CD spectra from Q2RNA/RHAU53-105 and Q2RNA in the region unique to nucleic acids (~250–320 nm), suggesting that the G4 remains intact upon protein binding.

3.2.1.6 Solution structures of G4s and their complexes with RHAU53-105

To further understand the recognition of G4s by RHAU53-105, we used SAXS to study Q2RNA, Q2DNA, and the Q2RNA/RHAU53-105 complex purified by size exclusion chromatography. DLS was employed as an initial quality control step to ensure sample monodispersity over the range of concentrations used for SAXS acquisition (**Fig. 3.7 A**). Decreasing hydrodynamic radii (r_H) were observed for the molecules in the following order: Q2RNA/RHAU53-105 complex (3.65 nm), Q2RNA (2.01 nm) and Q2DNA (1.65 nm) (**Table 3.1**). Samples did not display any significant self-association in the concentration range subsequently used for SAXS analysis, suggesting suitability for further structural studies (**Fig. 3.7 B**). SAXS data for Q2RNA, Q2DNA AND Q2RNA/RHAU53-105 complex collected at multiple concentrations were merged to obtain a single scattering profile (**Fig. 3.7 C**). The pair distance distribution function, P(r), which represents a histogram of all observed distances between electron pairs in the molecule was obtained from merged data using program GNOM (**Fig. 3.7 D**). Both Q2RNA and Q2DNA demonstrate a P(r) plot consistent with a globular structure, whereas the Q2RNA/RHAU53-105 complex likely adopts an extended conformation based on the elongated tail at longer distances. From this analysis, the radius of gyration (r_G) and maximum particle dimension (D_{max}) were determined (**Table 3.1**), and used as constraints to generate 20 individual

low-resolution models. Individual models, with the chi (χ) values shown in **Table 1**, were rotated and superimposed to obtain an averaged solution conformation (**Fig. 3.8**).

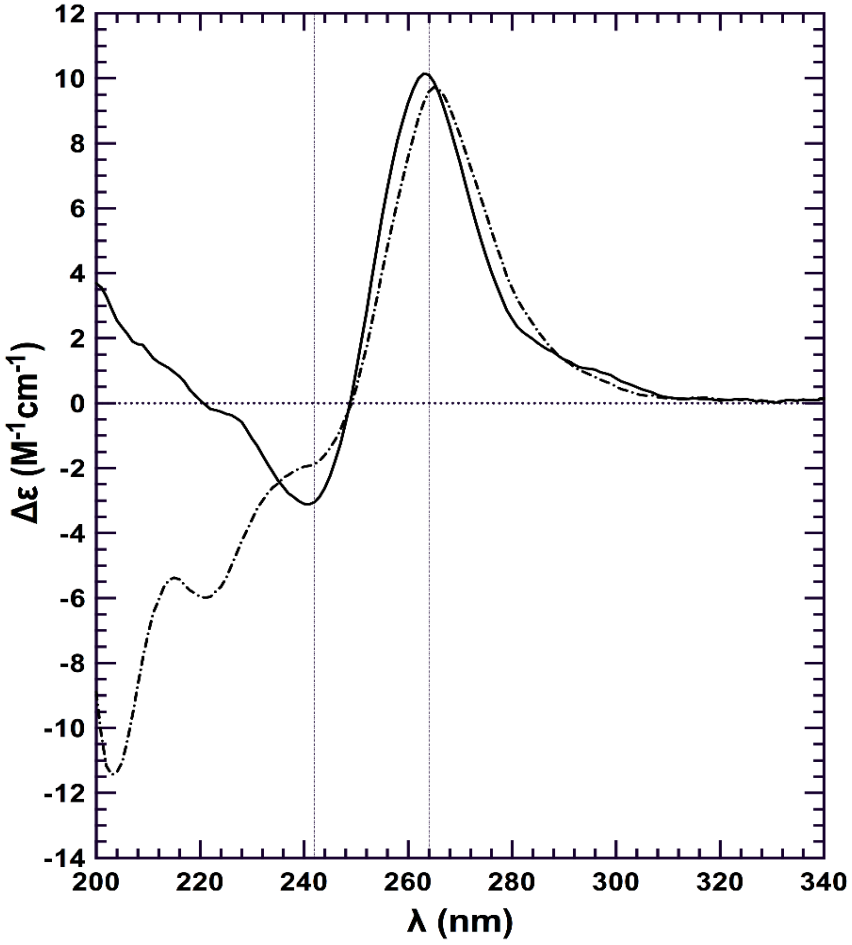

Fig. 3.6. RHAU53-105 binding is not sufficient for G4 unwinding. Far-UV CD spectra of Q2RNA (——); and Q2RNA/RHAU53-105 complex (- - -) at 20 μM and 20°C with all the spectra normalized to the number of nucleotides. The G4 features of the RNA in the context of the complex are observed in the region unique to nucleic acids (~250–320 nm).

Excellent superimposition of individually calculated models were confirmed by the normalized spatial discrepancy (NSD) parameter (≤0.63) for each molecule ensemble. Both RNA and DNA G4s adopt disc-shaped structures with concave bevels at the top and bottom, while the Q2RNA/RHAU53-105 complex adopts an extended shape (**Fig. 3.8**). Superposition of the

previously determined RHAU53-105 solution structure by SAXS onto the Q2RNA/RHAU53-105 complex suggests that G4 recognition is occurring via one of the termini of the protein.

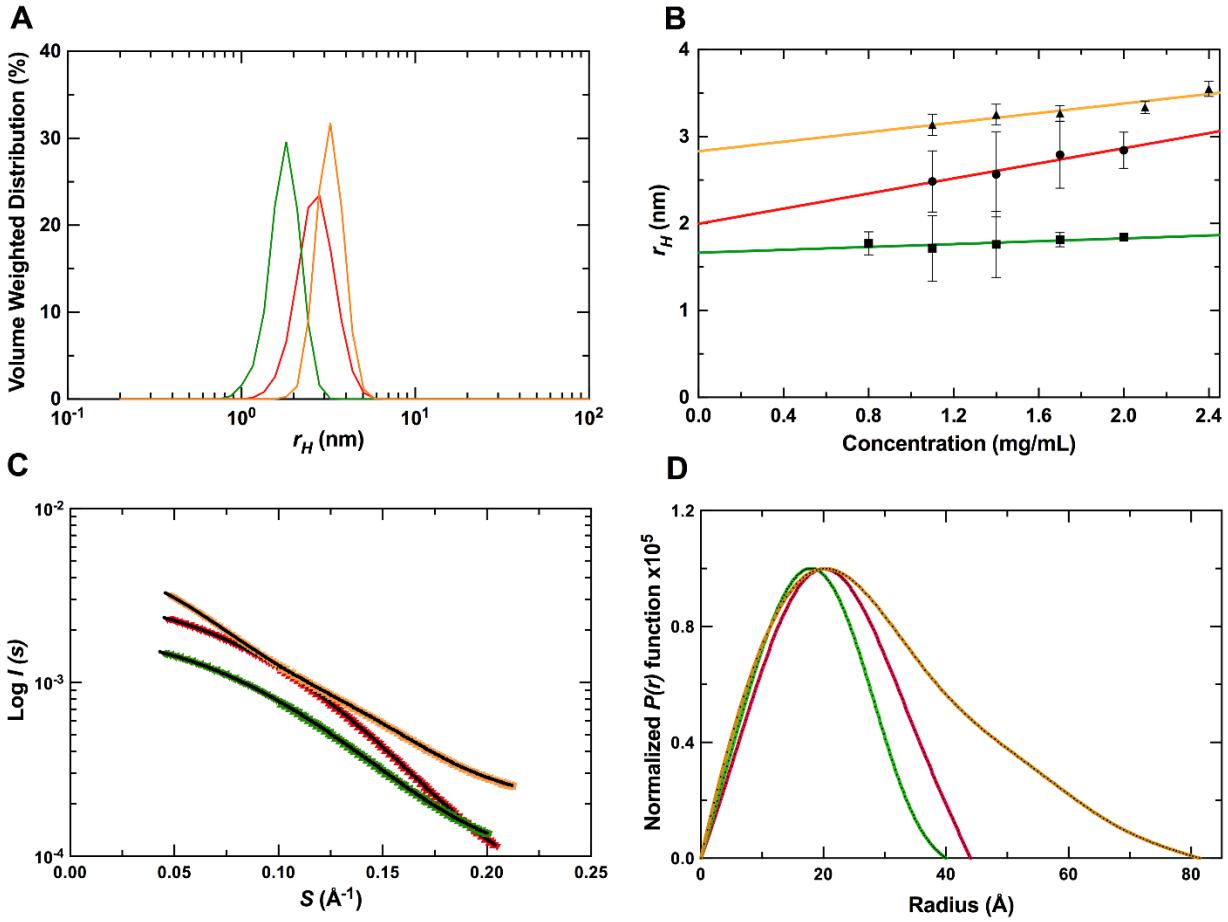

Fig. 3.7. SAXS data collection and analysis. (A) DLS profiles of Q2RNA, Q2DNA and 2RNA/RHAU53-105 complex. **(B)** Concentration dependence of hydrodynamic radii measured by DLS of the molecules in (A). **(C)** Merged SAXS data of Q2RNA, Q2DNA and Q2RNA/RHAU53-105 complex. **(D)** The corresponding pair distance distribution functions. (Color code: red, Q2RNA; green, Q2DNA; golden, Q2RNA/RHAU53-105 complex).

3.2.1.7 Amino acids in and adjacent to the RSM mediate recognition of RNA G4

To determine the region of RHAU53-105 involved in mediating the interaction with G4, we expressed and purified isotopically-enriched ^{15}N-RHAU53-105. After equimolar mixture with Q2RNA, we successfully purified Q2RNA/^{15}N-RHAU53-105 complex, and acquired its

^{15}NHSQC spectrum (**Fig. 3.9**). To determine the region(s) perturbed by G4 binding, we compared the determined HSQC spectrum to those previously determined for free ^{15}N-RHAU53-105 and ^{15}N-RHAU53-105 in complex with another minimal RNA G4 (hTR_{1-20}) (238). We observed nearly identical chemical shift perturbations as previously reported, with significant chemical shift perturbations clustering to the RSM and adjacent α-helix (K58, R60, E61, I62, G63, M64, W65, Y66, A67, K68, K69, N74, K75, A77, and E78). Therefore, we conclude that, as previously observed, residues in the RSM and the adjacent helix are responsible for mediating the G4 interaction.

Table 3.1. Summary of Hydrodynamic Parameters. Standard deviations are indicated in parentheses, and dashes indicate that the values could not be determined.
Published: E.O. Ariyo *et al.* (2015) PLoS ONE 10(12): e0144510.

	Q2RNA		Q2DNA		Q2RNA/RHAU$_{53-105}$	
Parameter	Exp.	Model[d]	Exp.	Model[d]	Exp.	Model[d]
r_H (nm)[a]	2.01 (0.25)	2.28 (0.01)	1.65 (0.04)	2.03 (0.01)	3.65 (0.10)	2.64 (0.02)
r_G (nm)[b]	1.62 (0.01)	1.81 (0.01)	1.42 (0.02)	1.60 (0.01)	2.36 (0.06)	2.47 (0.01)
D_{max} (nm)[c]	4.4	4.6 (0.04)	4.0	4.11 (0.03)	8.1	8.2 (0.06)
χ	0.53	-	0.82	-	0.80	-
NSD	0.5 (0.01)	-	0.5 (0.01)	-	0.63 (0.04)	-

[a] determined from DLS with error from linear regression analysis to infinite dilution from multiple concentrations.
[b] determined from SAXS data with error obtained from P(r) analysis by GNOM.
[c] determined from SAXS data obtained from P(r) analysis by GNOM.
[d] Model-based parameters calculated from HYDROPRO with errors obtained as the S.D. from multiple models.

doi:10.1371/journal.pone.0144510.t001

3.2.2 Discussion

G4s were predicted to play key roles in a number of biological activities including the regulation of gene transcription and translation (302), and evidence for that has accumulated in recent years both in vitro and in vivo (98, 101, 356). Various proteins interact specifically with

G4s, suggesting they fulfill important functions in cellular processes (183). RHAU has been observed in a number of contexts to interact with RNA G4s, but the mechanism of how it recognizes and unwinds these structures is not well characterized. We have chosen as our model system a specific G4 (Q2RNA) found in the 3'-UTR of the PITX1 mRNA, primarily because its interaction with RHAU in a cellular context is established (50). Based on several methods, including staining with an orientation-specific dye, TDS, and spectropolarimetry, Q2RNA is a parallel G4, and adopts a compact, disc-shaped conformation in solution that is consistent with another previously determined RNA G4 structure by SAXS (238). Interestingly, the DNA equivalent (Q2DNA) presented markedly different features, namely in terms of its CD profile and its inability to stain with a parallel G4 dye, despite adopting a similar shape in solution as determined by SAXS and sharing similar hydrodynamic features as Q2RNA. Although a high-resolution structure would unambiguously highlight the differences, our data is consistent with Q2DNA adopting the hybrid-type G4 orientation observed in group II G4s, which has three parallel and one antiparallel strands (73, 355, 357). We anticipate that the ability to accommodate 2′-OH groups in RNA G4s (affecting hydrogen bonding and sugar puckering) is central to the observed conformational differences between Q2RNA and Q2DNA (358).

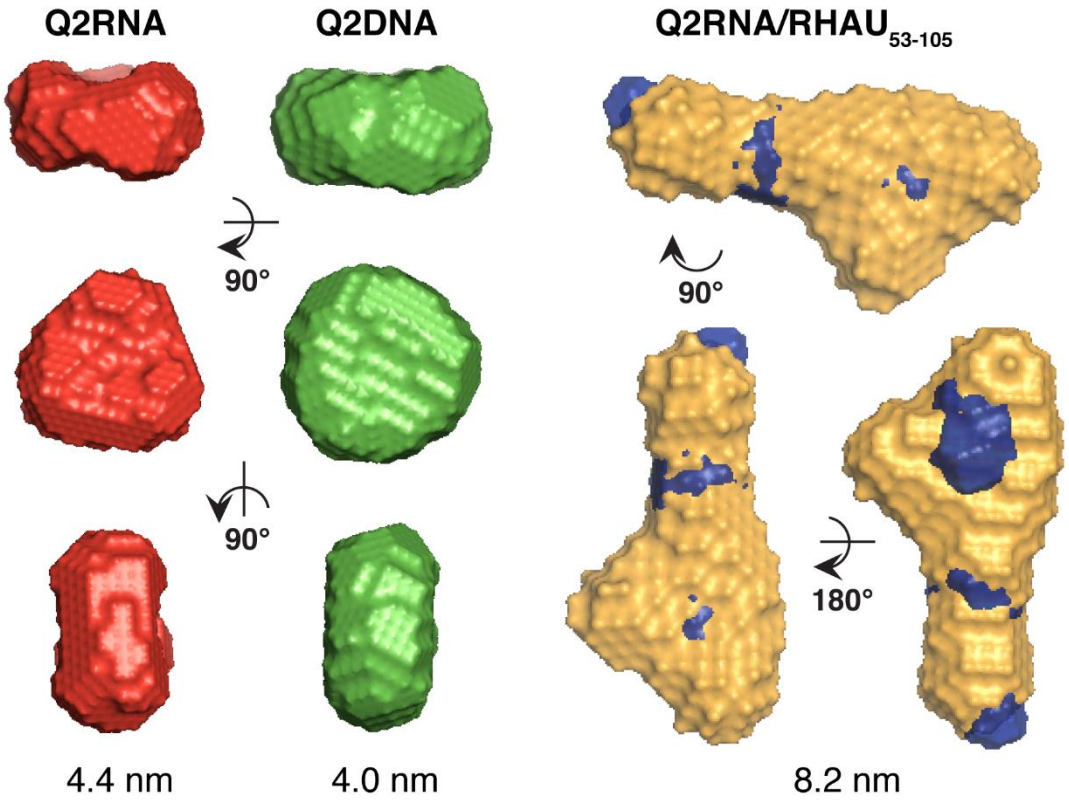

Fig. 3.8. SAXS envelopes of Q2RNA, Q2DNA and Q2RNA/RHAU53-105 complex. Color code: Red, Q2RNA; green, Q2DNA; golden, 2RNA/RHAU53-105 complex. Dmax values are shown beneath each species. The SAXS shape model of RHAU53-105 (blue) was superimposed on the complex (RHAU53-105 SAXS data obtained from previously published findings) (238)

Previous low-resolution structural and biophysical studies suggest that the N-terminal domain of RHAU interacts with the G-quartet face on the top or bottom plane of the G4 for both RNA and DNA (238). Recently, a high-resolution structure of a short N-terminal peptide in complex with a DNA G4 has reinforced this mode of recognition, but also suggested that certain basic amino acid residues mediate specificity through interaction with the phosphodiester backbone (274). Recognition of the Q2RNA G4 by the N-terminal region of RHAU (containing the RSM) uses nearly identical amino acid residues to those previously observed as important with another RNA G4 (238). The elongated solution structure of Q2RNA/RHAU53-105 by SAXS is

consistent with the same protein truncation in complex with another RNA G4 (238), and the superimposition of individual Q2RNA and RHAU53-105 models onto the complex model are consistent with recognition of the G-quartet face as the primary site of recognition.

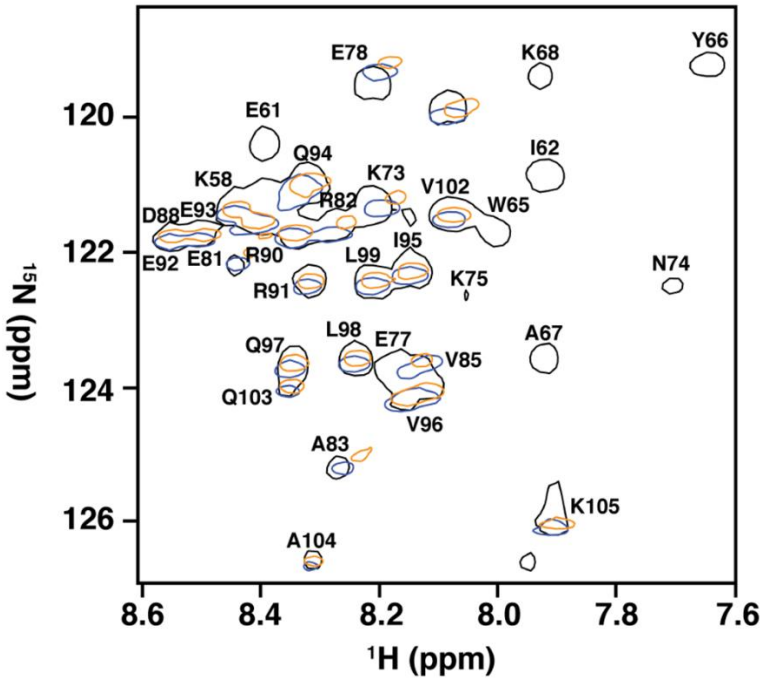

Fig. 3.9. Amino acids in and adjacent to the RSM mediate recognition of RNA G4. ^{15}N-HSQC spectral overlay of RHAU53-105 free (black) and in complex with Q2RNA (orange) or hTR_{1-20} (blue) with a subset of assigned resonances labeled. Data for RHAU53-105 and RHAU53-105/hTR_{1-20} were previously obtained, and used for comparison.

Mechanistic studies have also suggested the importance of a parallel orientation for the recognition by RHAU (272, 274). The parallel G4 specific dye used in this study (NMM) interacts by stacking on the G-tetrad faces (353) and we have observed a significant reduction in the staining intensity of the dye where the G4 is bound to RHAU53-105 as opposed to free G4 (data not shown). This suggests that the protein occupies the tetrad face. A previous study investigating a G4 from the human telomerase RNA (hTR) and its DNA counterpart has revealed that both adopt

a parallel orientation and that both interact with RHAU53-105 by means of the RSM. However, the DNA G4 made additional interactions with RHAU that were not observed in the RNA G4 (238). DNA G4s generally demonstrate lower affinity for RHAU than their RNA counterparts (238, 273, 274), and whether the 2′-OH, a parallel arrangement, or both are important remains to be determined for RNA binding. Given these observations, it was not surprising that different strand orientations adopted by Q2RNA and Q2DNA significantly impact their affinity for RHAU. In the absence of a high-resolution RHAU-RNA G4 structure, our results strongly support the previously observed mode of recognition where strand directionality is key to presenting a parallel G4 face for RHAU binding. High-resolution structural studies of RNA G4s in complex with RHAU will likely confirm the hypothesis that both electrostatic and steric impacts of the 2′-OH also fulfill an important role. While the importance of the N-terminal domain of RHAU has clearly been established, the mechanism whereby full-length protein binds and unwinds G4 structures remains to be elucidated. Binding of truncated RHAU53-105 to RNA or DNA G4 does not attain the full binding affinity observed in full-length RHAU nor does it induce unwinding (50, 238, 273). These features are clearly confirmed again in this study as full-length RHAU has higher affinity than the N-terminal fragment for Q2RNA, and comparison of the CD spectra of Q2RNA free and in complex with RHAU53-105 indicates no G4 unwinding. Therefore, future studies geared towards an understanding of G4 helicase activity in the context of the full-length protein, remain a priority. The work presented here, while focused specifically on the in vitro study of a purified RNA-protein complex, provide the template for an eventual mechanistic understanding of G4 impact on translational regulation of mRNAs, including PITX1.

CHAPTER 4

Impact of G-quadruplex loop conformation in the PITX1 mRNA on protein and small molecule interaction

The results presented here have been published in a peer-reviewed journal: Emmanuel O. Ariyo, Evan P. Booy, Edis Dzananovic, Ewan K. McRae, Markus Meier, Kevin McEleney, Jorg Stetefeld and Sean A. McKenna (2017) Impact of G-quadruplex Loop Conformation in the PITX1 mRNA on protein or small molecule interaction. Biochemical and Biophysical Research Communications (BBRC.2017.04.049).

AUTHOR CONTRIBUTIONS

EOA performed the research, analyzed the data, and wrote the manuscript. EPB performed and designed the research. EDZ analyzed the SAXS data. EKM performed the research (MST). MM analyzed the data and edited the manuscript. KM provided the SAXS technical support and sample acquisition. JS designed the research. SAM designed the research and edited the manuscript.

4.1 Introduction

G-quadruplexes (G4s) are nucleic acid structures that form when guanine-rich tracts fold back on themselves to form π- π stacked planar G-tetrads with guanine bases from each tract associating via a hydrogen bonding network between adjacent guanines in the tetrad (N7:N2H and O6:N1H) (17, 19). Stacked tetrads are further stabilized by guanine O6 coordination by a monovalent cation (typically K^+) centered between the planes (17). In general, RNA G4s adopt a parallel orientation of their G-tracts, owing largely to the presence of the 2′-OH on the ribose sugar (359). G4s have been shown to be very stable *in vitro* under near-physiological conditions, and studies have demonstrated their existence *in vivo* (47, 360). While the stability and kinetics of G4 structures are based on factors including sequence length and G-tract alignment (parallel/antiparallel), recent studies have begun to emphasize the role that G4 loops, the sequence of nucleotides between the guanines that form G-tetrads, play in determining tertiary structure and stability (80, 361, 362). Loops of different lengths and sequences have been shown to stabilize or

disrupt G4s (70, 77). Interestingly, intraloop interactions via hydrogen bonding and base stacking have also been reported, and in some cases these interactions contribute more to G4 stability than π- π stacking of G-tetrads (78, 363). Disruption of G4 formation by targeting loops using antisense oligonucleotides in eukaryotic cells has been reported, impacting translation of reporter genes and endogenously expressed mRNAs (364). A specific G4 loop conformation from the 5′-UTR of multiple human genes has been shown to impair the expression of a reporter gene *in cellula* (365). The folding of G4s driven by the length of loops and nucleotide sequence has been hypothesized to affect binding interactions of small molecules and proteins (79).

PITX1 is a transcription factor that plays pivotal roles in development including differentiation of the developing pituitary gland, craniofacial structures, and hind limbs in early embryonic development (290, 292). We previously demonstrated that the PITX1 mRNA contains 3 sequences, known as Q1, Q2, and Q3, capable of forming stable G4 structures in its 3′-UTR, and that PITX1 expression is attenuated by interaction with an established RNA G4 helicase (RHAU) (49). The recognition and binding of parallel RNA G4s by RHAU is mediated via its N-terminal RHAU-specific motif (RSM) (238, 253). The structure of a G4 in complex with an 18-amino acid peptide from the RSM demonstrated a stacking interaction between residues (G9, 12, G13, and A17) and the 5′-end G-tetrad, in addition to electrostatic interactions between basic residues (K8, K10, and K19) and phosphate backbone (274). Low-resolution structural studies with an N-terminal subdomain ($RHAU_{53\text{-}105}$) confirm that the N-terminal domain of RHAU interacts with the G-tetrad face on the top or bottom plane of the G4 (238, 275). In this study, we selected $PITX1_{2044\text{-}2079}$ (Q3RNA) (with loops CAGUGC, CCUGGC, and AGGUGG connecting guanine tracts) as a model system to study the impact of loop conformation on G4 structure by comparing it to designed mutants with the same nucleotide composition but scrambled loop

sequences. The impact of loop conformation on both G4 folding and biomolecular interaction was probed using a neutral macrocyclic ligand (*N*-methyl mesoporphyrin IX (NMM)) and a protein binding partner (RHAU). These binding partners were selected based on their established selectivity for parallel G4 structures. Our results demonstrate that loop conformation can impact the global fold of Q3RNA and also impact binding affinity of ligands that interact with the G-tetrad face of G4s.

4.2 Results and Discussion

4.2.1 Results

4.2.1.1 Q3RNA and loop mutants maintain a parallel G4 conformation

The conformation of Q3RNA alongside mutants where the first ($Q3M_1$), second ($Q3M_2$), or first and second loop ($Q3M_{1\&2}$) sequences were scrambled (**Table 4.1**) was investigated to determine the impact of loop conformation on overall G4 structure and recognition. As expected, Q3RNA stained effectively in native gels with NMM, a dye that is specific to the parallel G4 conformation typically observed in RNA G4s confirming a parallel G4 conformation (**Fig. 4.1A**, *bottom panel*). Similarly, Q3M1, Q3M2, and $Q3M_{1\&2}$ stained with NMM, confirming that loop sequence scrambling did not disrupt G4 formation. A non-G4 RNA control (dsRNA) visible by staining with a general nucleic acid stain (toluidine blue) was not stained by NMM (**Fig. 4.1A,** *middle and bottom panels*). Interestingly, while all Q3RNAs are of equivalent size (denaturing gel, **Fig. 4.1A**, *top panel*), mutants with a scrambled loop 2 sequence migrated as a larger size by native gel electrophoresis (**Fig. 4.1A,** *middle panel*). SEC confirmed these observations, where $Q3M_2$ eluted slightly earlier than Q3RNA, suggesting a difference in hydrodynamic size (**Fig. 4.1B**).

Table 4.1
The RNA sequences studied.

Name[a]	5′-3′ sequence[b]
Q3 RNA	AGC GGG CAGUGC GGG CCUGGC GGG AGGU GGGGG AGG
Q3 M_1	AGC GGG **UGCGCA** GGG CCUGGC GGG AGGU GGGGG AGG
Q3 M_2	AGC GGG CAGUGC GGG **CGGCCU** GGG AGGU GGGGG AGG
Q3 $M_{1\&2}$	AGC GGG **UGCGCA** GGG **CGGCCU** GGG AGGU GGGGG AGG

[a] *The name assigned to this sequence in the text, also indicative of loop mutations in the sequence.*
[b] *nucleotides in bold represent rearrangements in loop 1, 2 or both loops*

Thermal difference spectroscopy (TDS) of Q3RNA and the mutants produced spectra consistent with a G4 conformation (305) (**Fig. 4.1C**). In addition, the far-UV CD spectra generated for the RNAs produced spectra with ellipticity minima (242 nm) and maxima (264 nm) that are consistent with a parallel G4 conformation (354) (**Fig. 4.1D**). The relatively high melting temperatures (T_m>60°C) obtained from CD melting experiments (**Fig. 4.1D, E**) also suggest G4 conformation as they are significantly higher than what would be expected from other conformations (including duplex) that these RNA may adopt. Together, we conclude that while the scrambling of loop 2 appears to impact native gel migration, a parallel G4 structure persists.

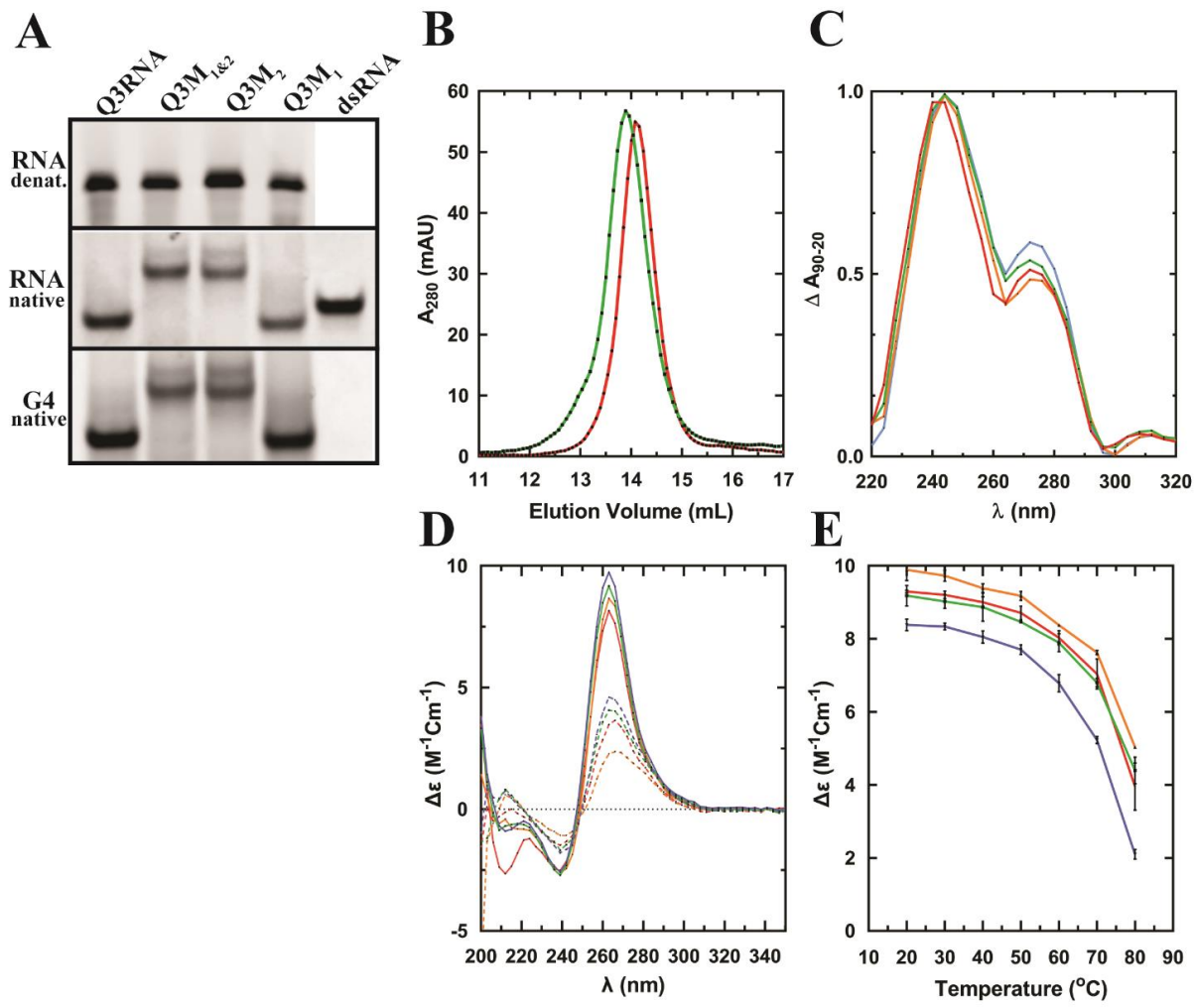

Fig. 4.1. Q3RNA and mutants all adopt parallel G4 conformation (A) Q3RNA and mutants separated by 12% PAGE; *Top*: denaturing TBE-PAGE using urea as a denaturant, *middle*: native TBE-PAGE stained with toluidine blue nucleic acid stain, pH 2.8, and *bottom*: native TBE-PAGE stained with NMM. A non-G4 dsRNA is also shown as a negative control **(B)** Elution profiles of Q3RNA and Q3M2 obtained by SEC **(C)** Normalized thermal difference spectra of Q3RNA and mutants **(D)** far-UV CD spectra generated at 20°C and 80°C. The CD signals at 20°C are shown to be higher than at 80°C **(E)** CD melting curves at 624 nm of samples in (D). All the experiments were performed in 10 mM Tris, (pH 7.5), 100 mM KCl, 1 mM EDTA buffer and legends are shown as follows: Color code: Red, Q3RNA; golden, $Q3M_1$; green, $Q3M_2$; blue, $Q3M_{1\&2}$. Figures are colored as follows: red, Q3RNA; orange, $Q3M_1$; green, $Q3M_2$; blue, $Q3M_{1\&2}$.

4.2.1.2 Loop 2 mutants of Q3RNA adopt an altered G4 solution conformation

To investigate the impact of the potential changes caused by loop composition on the overall structure of Q3RNA, we employed SAXS. As a quality control step, we first performed DLS experiments to ensure the suitability of the samples for SAXS. Consistent with our gel-based observations, Q3RNA and $Q3M_1$ have similar hydrodynamic radii (r_H), with $Q3M_2$ and $Q3M_{1\&2}$ demonstrating slightly larger values (**Fig. 4.2A**, **Table 4.2**). No concentration-dependent change in r_H was observed for any of the RNAs over the range of concentrations subsequently used for SAXS **(Fig. 4.2A, inset)**. SAXS data for Q3RNA and mutants obtained at multiple concentrations were therefore merged and processed to obtain scattering profiles from which the pair distance distribution function (P(r)) were calculated and a histogram of all observed distances between electron pairs in the molecule was generated (**Fig. 4.2B**).

Table 4.2. Summary of Hydrodynamic Parameters

Standard deviations are indicated in parentheses, and dashes indicate that the values could not be determined.

	Q3RNA		$Q3M_1$		$Q3M_2$		$Q3M_{1\&2}$	
Parameter	Exp.	Model[d]	Exp.	Model[d]	Exp.	Model[d]	Exp.	Model[d]
r_H **(nm)**[a]	2.93 (0.25)	2.03 (0.01)	2.90 (0.04)	2.06 (0.01)	3.16 (0.10)	2.07 (0.01)	3.17 (0.02)	2.31 (0.01)
r_G **(nm)**[b]	1.60 (0.01)	1.60 (0.01)	1.65 (0.01)	1.64 (0.01)	1.75 (0.01)	1.74 (0.01)	1.89 (0.01)	1.88 (0.01)
D_{max} **(nm)**[c]	4.1	4.84 (0.01)	4.3	5.15 (0.01)	4.7	5.54 (0.01)	4.80	5.47 (0.01)
χ	0.50	-	0.50	-	0.60	-	0.57	-
NSD	0.47 (0.01)	-	0.48 (0.01)	-	0.63 (0.09)	-	0.54 (0.05)	-

[a] determined from DLS with error from linear regression analysis to infinite dilution from multiple concentrations.
[b] determined from SAXS data with error obtained from P(r) analysis by GNOM.
[c] determined from SAXS data obtained from P(r) analysis by GNOM.
[d] Model-based parameters calculated from HYDROPRO with errors obtained as the standard deviation from multiple models.

From this analysis, the radius of gyration (r_G) and maximum particle dimension (D_{max}) were determined, with results reinforcing that mutants with loop 2 scrambled have larger values than those with the wild type loop 2 (**Table 4.2**). The r_G and D_{max} values were then used as constraints to generate 20 individual low-resolution models that were rotated and superimposed to produce the average solution conformation (**Fig. 4.2C**). Each of the RNAs generally adopt disc-shaped structures, consistent with previously published low-resolution structures of parallel G4s by SAXS (238, 275). Interestingly, $Q3M_2$ and $Q3M_{1\&2}$ present a crescent-shaped structure not observed in Q3RNA or $Q3M_1$ and based on comparison with similar G4s suggests a perturbation to the edges of the G-tetrad face (**Fig. 4.2C**).

4.2.1.3 Stabilization of G4 RNA may involve intramolecular hydrogen bonding

We next performed ^{1}H-imino proton NMR experiments to confirm G4 formation and to determine if intraloop hydrogen bonding could be observed (**Fig. 4.3A**). Hydrogen bonded imino protons (from G and U bases) are protected from solvent exchange, whereas free bases are not observable. The NMR spectrum was dominated by at least 11 unique clustered peaks between 10-12 ppm, which are characteristic of solvent protected G-G hydrogen bonding observed in G4s. Surprisingly, two solvent-protected imino proton peaks were observed distinct from the G-G region; one peak at 13 ppm (consistent with a G-C hydrogen bond) and another at 10 ppm (consistent with a G-U hydrogen bond) (**Fig. 4.3A**).

4.2.1.4 Conformational alteration of loop2 impacts RSM but not RHAU binding

PITX1 Q1, Q2, and Q3RNA have been previously demonstrated to interact with both full length RHAU and an N-terminal subdomain ($RHAU_{53\text{-}105}$) that contains the RSM, with Q3 having

the lowest affinity of the three G4s (49). To determine if the binding affinity of RHAU is affected by the conformational changes induced by loop 2 scrambling, EMSAs were performed between RHAU full length or the isolated $RHAU_{53-105}$ protein and G4 RNAs (**Fig. 4.3B**).

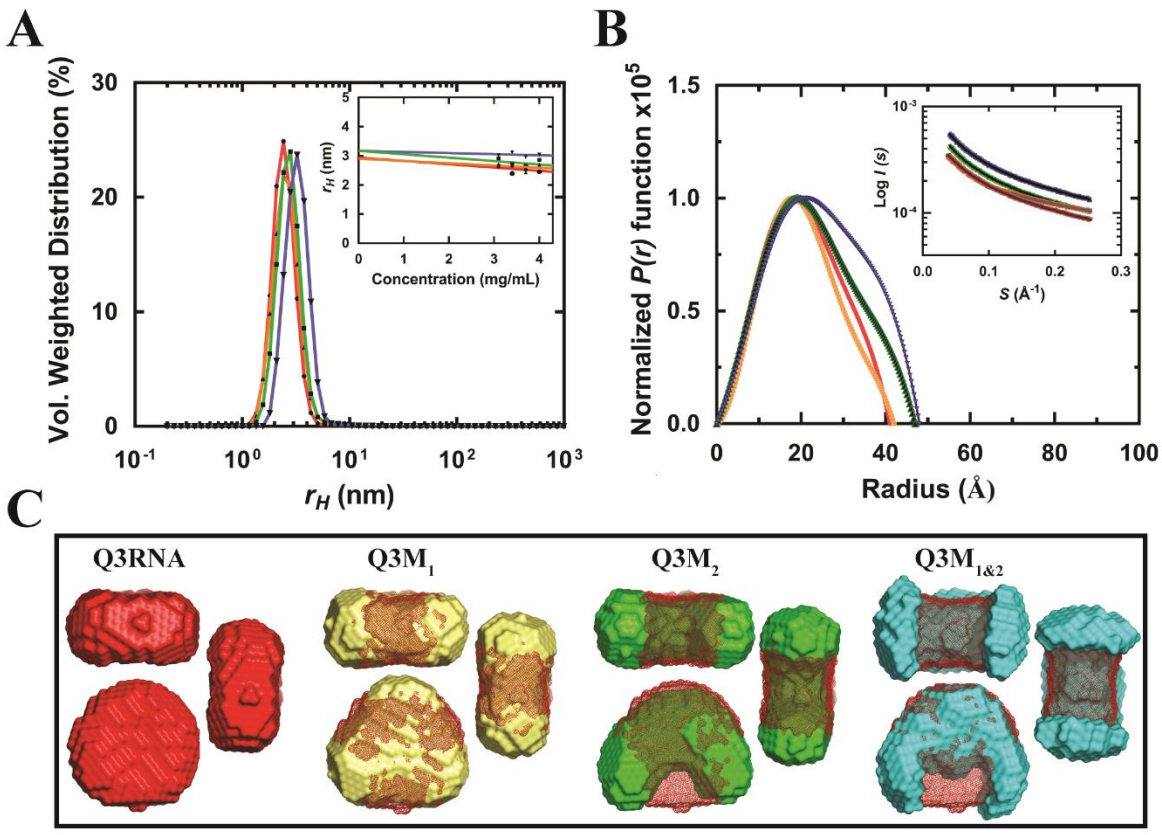

Fig. 4.2. Mutations to loop 2 alters the G4 fold of Q3RNA. Figures are colored as follows: red, Q3RNA; orange, $Q3M_1$; green, $Q3M_2$; blue, $Q3M_{1\&2}$. **(A)** DLS profiles of Q3RNA, $Q3M_1$, $Q3M_2$, $Q3M_{1\&2}$ and concentration dependence of the hydrodynamic radii of the same samples determined by DLS (inset); SAXS data analysis **(B)** The pair distance distribution function of individual G4 RNA and merged SAXS data plots of the same samples obtained from the raw SAXS data (inset) **(C)** Average, filtered SAXS solution structures of G4 RNAs based on 20 individual models, with the overlaid red mesh representing the solution structure of Q3RNA for comparison purposes.

Significantly higher affinity of $RHAU_{53-105}$ is observed for Q3RNA and $Q3M_1$ relative to either $Q3M_2$ or $Q3M_{1\&2}$. Interestingly, the same trend is not observed when EMSAs are performed with full-length RHAU, as both Q3RNA and $Q3M_2$ present similar affinities (**Fig. 4.3C**).

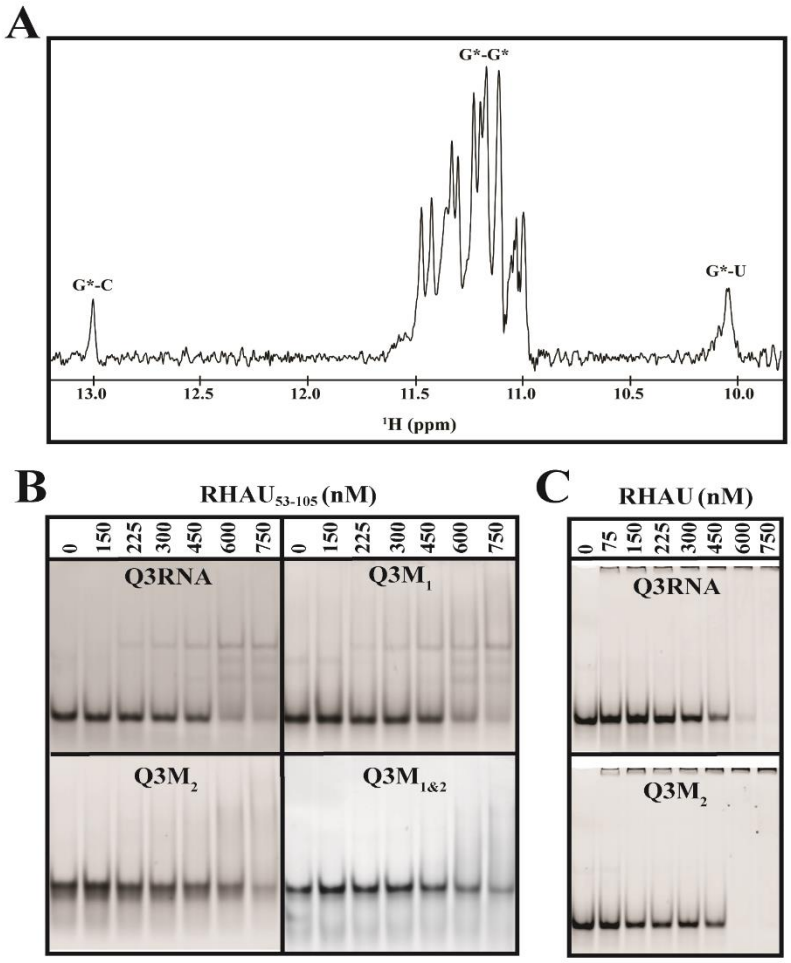

Fig. 4.3. Mutations to loop2 reduces the affinity of a protein binding partner for Q3RNA. (A) Proton NMR spectrum of Q3RNA (100 μM), in 10 mM Tris, (pH 7.5), 100 mM KCl, 1 mM EDTA, and 10% deuterium oxide (v/v). Distinct regions corresponding to G*-C, G*-G*, and G*-U hydrogen binding are highlighted, with the asterisk (*) corresponding to the observable imino proton. **(B and C)** Electrophoretic mobility shift assay carried out in reactions containing a fixed RNA concentration (100 nM) with increasing concentrations of **(B)** $RHAU_{53-105}$ or **(C)** full length RHAU (0-700 nM).

4.2.1.5 G-tetrad face ligand binding is impacted by loop2 scrambling

The interaction between *N*-methyl mesoporphyrin IX (NMM) and G4s occurs via stacking of NMM onto the G-tetrad faces of parallel G4s (159). UV-visible absorbance titration of Q3RNA or $Q3M_2$ into NMM was used to monitor binding by tracking the NMM Soret band intensity and wavelength (379 nm initially) (**Fig. 4.4A, B, C**).

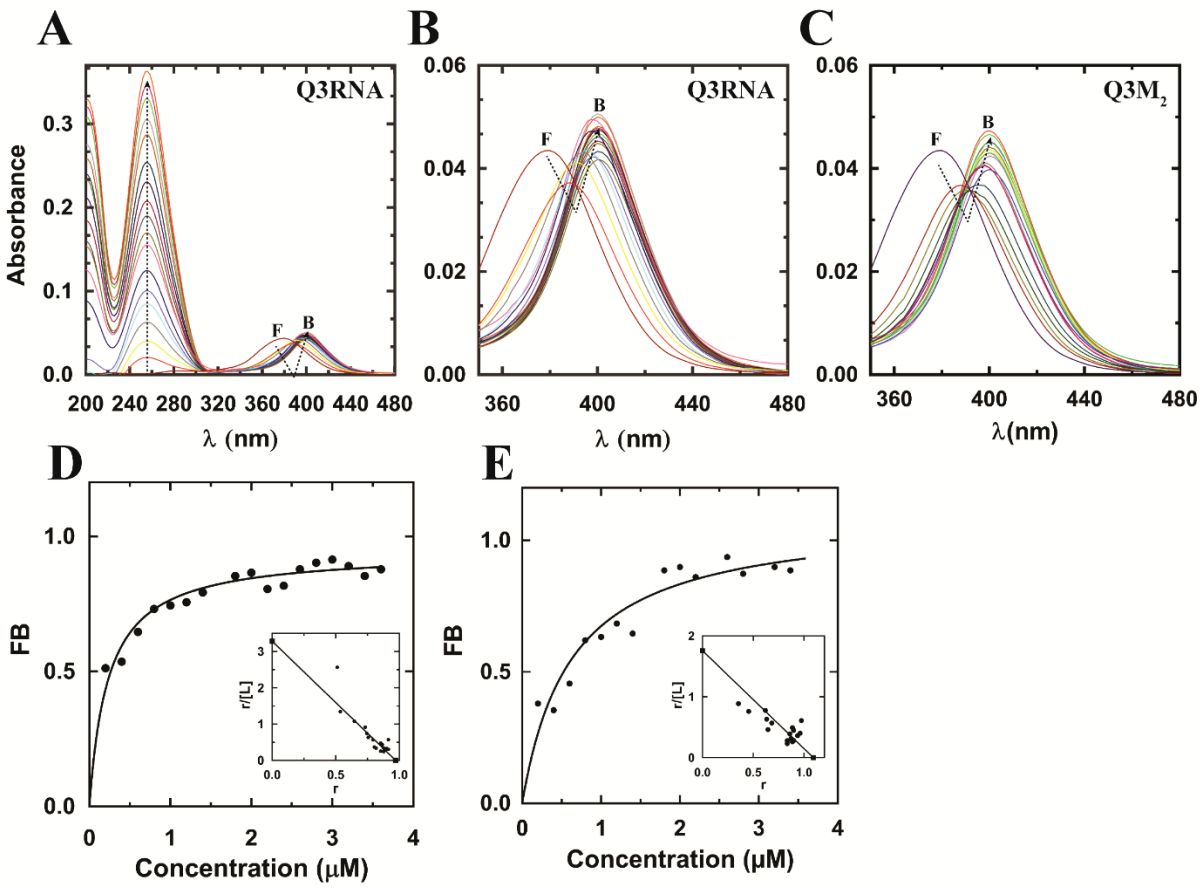

Fig. 4.4. Loop2 sequence composition impacts small molecule binding to Q3RNA (A) UV/Vis spectrum highlighting the impact of increasing concentrations of Q3RNA (increasing peak at 260nM) on NMM binding by monitoring the shift in the NNM Soret band from 379 nm (free; F) to 400 nm (bound; B). Arrows indicate both spectral changes. **(B)** As in (A), but expanded to focus on the NNM shift. **(C)** As in (B) but for the $Q3M_2$ RNA. G4 binding constants were then determined by direct fitting of the titration data to a one site binding hyperbola **(D)** Q3RNA **(E)** $Q3M_2$ and binding stoichiometry was determined by Scatchard analysis (inset) as described in Materials and Methods.

For each RNA, the Soret band of the complex gradually shifted to maxima of 398 and 400 nm respectively, and Scatchard analysis confirmed the expected 1:1 binding stoichiometry (**Fig. 4.4 D, E**). Following the trend observed for $RHAU_{53\text{-}105}$ binding, $Q3M_2$ (621±51 nM) displayed a weaker affinity for NMM than Q3RNA (295±23 nM), suggesting that the loop 2 sequence can impact access to the G-tetrad face.

4.2.2 Discussion

The nucleotide sequence between guanine tracts of G-tetrads can form loops that play an important role in determining the folding and stability of G4 structures (70, 77). We report that scrambling of the sequence in loop 2 but not loop 1 results in a change to the global structure of Q3RNA, suggesting a specific tertiary interaction above and beyond core G4 formation, as observed by RNA migration on native polyacrylamide gels and their solution structures (**Fig. 4.1A, 2C**). Hydrogen bonding and base stacking can occur between loops giving rise to interloop interactions (78). The results obtained from our imino proton NMR experiments of Q3RNA showed distinct G-C and G-U imino proton signals (13 and 10 ppm respectively). Based on these NMR results and that loop 1 scrambling has no disruptive impact on G4 structure, we hypothesize that a loop 2-loop 3 interaction is occurring; G26-C17 and G27-U18 base pairs. It is therefore not surprising that scrambling of loop 2 would result in a G4 structure with an altered solution conformation, as this would likely disrupt the interloop interaction.

The SAXS solution structures obtained for Q3RNA and $Q3M_1$ are disc-shaped, consistent with previously published low-resolution structures of parallel G4s by SAXS (238, 275). However, $Q3M_2$ and $Q3M_{1\&2}$ containing disrupted loops showed crescent-shaped structures that

are not observed in Q3RNA or $Q3M_1$ and when compared to previously published low resolution SAXS structures, perturbations to the edges of the G-tetrad faces were observed (**Fig. 4.2C**).

Previous reports have suggested that loop composition can significantly impact the binding interactions of small molecules and proteins (79), a finding that we have confirmed, in part, with Q3RNA as a model system. Mutation of loop 2 results in a 2-fold reduction in binding affinity for NMM, a parallel G4-specific ligand that stacks on the G-tetrad face (159) **(Fig. 4.4)**. This 2-fold change is reasonable given that while there is a global structural change, Q3RNA is able to retain the parallel G4 conformation despite a scrambled loop 2. The interaction of Q3RNA with $RHAU_{53-105,}$ (an N-terminal subdomain containing the RSM) investigated by EMSA showed a similar trend to the NMM result, with scrambling of loop 2 resulting in a decreased binding affinity for Q3RNA. No significant difference in binding affinity was observed between the full length RHAU and all the G4 RNAs, which is not unexpected given that interactions outside the RSM contribute to additional affinity for G4 structures (253). RNAs containing a scrambled loop 2 showed less interaction with $RHAU_{53-105}$ in a concentration dependent manner (**Fig. 4.3B**), suggesting an impact of the loops on binding of proteins and small molecules by the G-tetrad faces of the G4 structures (**Fig. 4.3B; 4.4D, 4.4E**).

CHAPTER 5

SUMMARY AND FUTURE DIRECTIONS

5.1 RESEARCH SUMMARY

Sequences having the potential to form G4 structures in the human genome have been reported to number approximately 376 000 in the human genome (22), and high-resolution sequencing of DNA G4 structures in the human genome revealed 716,310 distinct G4 structures (23). Although the early studies of G4s were based on their roles at the ends of telomeres, recent studies have shifted unto other important regions of the genome including the promoters of proto-oncogenes (39, 45, 90, 91), immunoglobulin heavy chain switch regions (39), mutational hotspots implicated in the maintenance of chromosomal integrity, and the regulation of replication, transcription and recombination processes (35, 48). The interaction of some proteins with G4 structures to stabilize (181, 182), or unwind and destabilize them, thereby modulating their functions in a biological context have been investigated (183-185). To further expand the understanding of G4 RNA recognition by the helicase RHAU in a biological context, our laboratory recently performed an RNA co-immunoprecipitation screen that led to the identification of messenger RNA (mRNA) for Pituitary homeobox 1 (PITX1, backfoot) protein (49). PITX1 is a transcription factor that plays several roles including, differentiation of developing pituitary gland, craniofacial structures and hind limbs in the early development of embryos (281, 289, 290, 292). Interestingly, PITX1 mRNA possesses three distinct sequences in the 3′-UTR of its mRNA (Q1 RNA: $PITX1_{1371\text{-}1400}$, Q2 RNA: $PITX1_{1901\text{-}1930}$, and Q3 RNA: $PITX_{12044\text{-}2079}$), capable of forming G4 structures and have been demonstrated to be involved in the recruitment of RHAU protein to the PITX1 mRNA to regulate PITX1 protein translation (49).

To understand the structural basis of G4 recognition by helicase proteins, we have purified RHAU protein and characterized a minimal G4-forming sequence in the 3′-UTR of PITX1 mRNA using a variety of biophysical techniques highlighted in Chapter 2. In Chapter 3, based on several

biophysical techniques used including SAXS, we were able to show that a minimal G4-forming sequence in the 3′-UTR of PITX1 mRNA ($PITX1_{1901\text{-}1930}$, Q2 RNA) is a parallel G4 that adopts a compact disc-shaped structure in solution (**Fig. 3.8**) and this finding was consistent with another G4 structure previously determined by SAXS (238). The DNA counterpart of this G4 RNA likely adopts a hybrid-type G4 orientation with three parallel and one parallel strands, consistent with group II G4s (**Fig. 1.4B**) (73, 355, 357). Our findings (reported in Chapter 3) suggests that a G4 structure (Q2RNA) interacts with the N-terminal domain of RHAU protein through its G-tetrad face and this extended our scope into the recognition of parallel G4 structure by the N-terminal region of RHAU.

In Chapter 4, we demonstrate the impact of loop conformation of G4s on protein and small molecule interactions. Previous findings suggested that the stability and kinetics of intramolecular G4 structures are based on a number of factors including hydrogen-bonding network involving N7:N2H and O6:N1H in G-tetrads, central core G-tetrad lone pairs of electrons from GO6 coordination of a monovalent cation (362), and π- π stacking interactions. Loop interaction has also been shown to contribute more to the stability of G4 structures than the stacking of G-tetrads (363). We studied another G4 forming sequence in the 3′-UTR of PITX1 mRNA: ($PITX_{12044\text{-}2079}$, Q3RNA) with three other mutants; $Q3M_1$, containing a nucleotide rearrangement in loop 1; $Q3M_2$, containing a nucleotide rearrangement in loop 2; and $Q3M_{1\&2}$, containing nucleotide rearrangement in both loops 1 and loop 2 (**Chapter 4**)

Our investigation into the roles played by loop composition on the overall G4 structure stability revealed that NMM interacts with Q3RNA and $Q3M_2$ G4 structures via UV-visible absorbance titrations that produced differences in dissociation constants (K_d): Q3RNA (295 ± 23 nM); $Q3M_2$ (621 ± 51 nM) respectively (**Fig. 4.4 D & E**), suggesting that loops in the G4 RNA can

affect the accessibility of proteins or small molecules from interacting with the G-tetrad faces. While the loop conformation can impact the global fold of PITX1 Q3, as observed in their migration pattern on a 12% native polyacrylamide gel (29:1 acrylamide: *bis* ratio) (**Fig. 4.1 A**), the parallel G4 RNA is able to accommodate these conformations to minimize the impact on binding of ligands and proteins that interact with the G-tetrad face of G4 RNA. RNA and DNA sequences that are able to form G4 structures have been found in regions with biological significance (206). This study provides a structural basis for G4 recognition by the N-terminal region of RHAU protein, an interaction mediated via the G-tetrad face of the G4, and also a basis for structural studies of key biologically relevant G4s that interact with proteins and G4-stabilizing ligands via their G-tetrad faces. A more comprehensive structural study involving high resolution structures may provide insights into the design and development of drugs that interact with G4s, to modulate their functions in a cellular context.

5.2 FUTURE DIRECTIONS

The goal of this project was to establish the structural basis of G4 recognition by RHAU protein, and to investigate the impact of loop composition and lengths on the overall G4 structures. Our findings suggest that RHAU recognizes G4 structures via the planar guanine tetrad face of parallel G4s, which is consistent with previous low-resolution structural and biophysical studies (238). In line with the results obtained, this study has certainly induced very interesting questions that could be addressed in future studies. Research in the following key areas are suggested:

(I) We demonstrated that a minimal G4-forming sequence in the 3′-UTR of PITX1 mRNA (Q2RNA) can interact with RHAU53-105 via its G-tetrad face (**Fig. 3.5 A, Fig. 3.9**), while focused

specifically on the *in vitro* study of a purified RNA-protein complex involving a short region of the entire PITX1 mRNA and a segment of RHAU protein, mechanistic understanding of G4 impact on translational regulation of mRNAs, including PITX1 in a cellular context to understand G4 helicase activity should be investigated.

(II) The higher stability of G4 RNA over its G4 DNA counterpart may be due to its ability to accommodate the 2′-OH group that affects hydrogen bonding and sugar puckering, leading to conformational differences between Q2 RNA and Q2 DNA (**Fig. 3.4 A**). We were not able to confirm this observation in this research and high-resolution structures of the RNA and DNA G4s will provide more insight.

(III) We showed that loop conformation can impact the global fold of G4 RNA in the PITX1 mRNA (Q3RNA) and that these conformations can impact the binding of ligands and protein that interact with the G-tetrad face of the G4 RNA (**Fig. 4.3B; Fig. 4.4**). Loops of different lengths and sequences have been shown to stabilize or disrupt G4s (70, 77), and in some cases interactions between loops contribute more to the stability of G4 structures than the stacking of π- π G-tetrads (78, 363). A high resolution structural study of G4s in DNA and RNA with various loop conformations will provide further insights into their roles in the overall G4-protein or G4-small molecule interactions.

(IV) G4 structures can be stabilized by small molecule ligands including porphyrins (155-157). Several aspects regarding G4 binding compounds require clarification, including preferences for tetraplexes over duplexes and binding specificities. Targeting G4 structures in the presence of a

massive excess of double-stranded DNA in a cellular context is an indispensable area to be studied and may provide potential usefulness in pharmaceuticals.

(V) The *in vitro* transcription of G4 RNA in the 3′-UTR of PITX1 mRNA from DNA templates, presented some challenges which could not be addressed during the course of this study (see the appendix). Although, Q2RNA and Q3RNA G4s were transcribed successfully from plasmid templates (**Fig. A.2**), however, the yields were very low and not enough material was obtained to carry out any structural studies. A good study on the factors that may be responsible for the low yield in transcribing G4 RNA from plasmid templates; including the role of helicases in unwinding the double stranded DNA, accessibility of RNA polymerase, G:C content of the template strand and other reaction conditions should be explored. This will be a giant stride into obtaining high yield in G4 RNA transcripts sufficient for structural studies.

(VI) Proving that G4 structures interact with RHAU protein to modulate some biological functions in the cellular context is undoubtedly beyond the scope of this dissertation, however, we believe a good stride in the right direction was laid for future studies.

APPENDIX

RESULTS OBTAINED FROM *IN VITRO* TRANSCRIBED RNA

A.1 Introduction

The *in vitro* transcription of G4 RNA from the 3′-UTR of PITX1 mRNA from DNA templates, followed by characterization using various biophysical methods and hopefully conducting structural studies was one of the initial goals of my research. Although, Q2 and Q3 G4 RNAs were transcribed successfully from plasmid templates, however, several challenges were encountered during the process, with the most being very low yield. Some of these challenges will be discussed in this chapter and results will also be presented.

A.2 Results and Discussion

A.2.1 Visualizing digested plasmid on agarose gel

Mini preparations of the cloned vectors were made after transformation using GeneJet Plasmid Miniprep Kit (see procedures in chapter 2) to produce more copies of each plasmid. The cloned plasmids were checked to ensure accurate sequence insertion after digesting with the appropriate restriction enzymes on 1.8% agarose gels against the mother vector and imaged by FlourChem Q (**Fig. A.1 A**). Individual cloned plasmids containing the sequences corresponding to Q1, Q2 and Q3 were also digested with the BsaI restriction enzyme and purified for *in vitro* transcriptions (**Fig. A.1 B**). Linearized plasmids were purified by extracting with phenol/chloroform (1:1, as described in chapter 2), and used for *in vitro* RNA transcription.

A.2.2 *In vitro* transcription of G4 RNA

The *in vitro* transcription of Q2 and Q3 G4 RNAs from purified linear plasmid templates containing a promoter, ribonucleotide triphosphates, a buffer system that includes DTT and magnesium ions, and an appropriate phage RNA polymerase were performed. Trial transcription

reactions were set up (**Table A.1**) to determine the best reaction conditions for a large scale *in vitro* transcription using linearized plasmid (**Fig. A.2 A**).

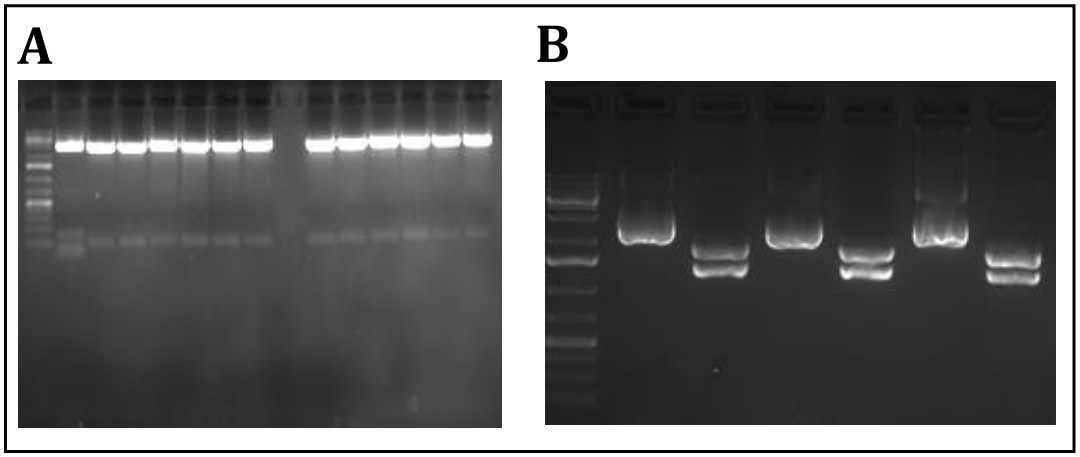

Fig. A.1. Linearized plasmids with HindIII, EcoRI and BsaI restriction enzymes (A) Linearized plasmids (with HindIII and EcoRI restriction enzymes), the insert and the mother vector were seen to migrate differently on the gel and this suggested the right sequence insertion **(B)** Linearized plasmids containing Q1, Q2 and Q3 sequences (Linearized with BsaI restriction enzyme) and used for *in vitro* transcription. The DNA was separated on 1.8% agarose gel in TAE buffer, stained with SYBR DNA. Undigested plasmids and DNA ladder were also shown.

The overall yield of all the G4 RNAs transcribed (Q2 and Q3 RNAs) were very low as shown in the elution profile (**Fig. 2.4 A**), but the initial characterization of the these RNAs using biophysical techniques revealed results that were consistent with G4s.

A.2.3 Assay for purity, Aggregation and Concentration by PAGE

Following the purification of the *in vitro* transcribed RNAs by SEC (See chapter 2), fractions corresponding to the peaks of the RNAs were collected, concentrated and run on 12% native Polyacrylamide Gels in TBE buffer, stained with toluidine blue and viewed as clean bands (**Fig. A.2 B**).

Table A.1. An example of the recipes for a trial *in vitro* transcription reaction

Reagent	Required amount
Plasmid template (Linearized)	2.5 – 10 µg
Nucleotide triphosphates (NTPs)	2 mM each
Magnesium Chloride ($Mgcl_2$)	5 – 20 mM
Transcription buffer (See Chapter 2)	1:10 (10X buffer: Reaction total volume)
T7 RNA polymerase	0.5 – 2.0 µL (in a 50 µL total reaction volume)
RNase free water	Adjusted to make up the total volume

Table A.1. A table showing the recipes required to set up a small scale reaction for *in vitro* transcription. A total reaction volume of 50 µL was set up where some of the reagent concentrations were varied to obtain the best conditions for a large scale transcription reaction.

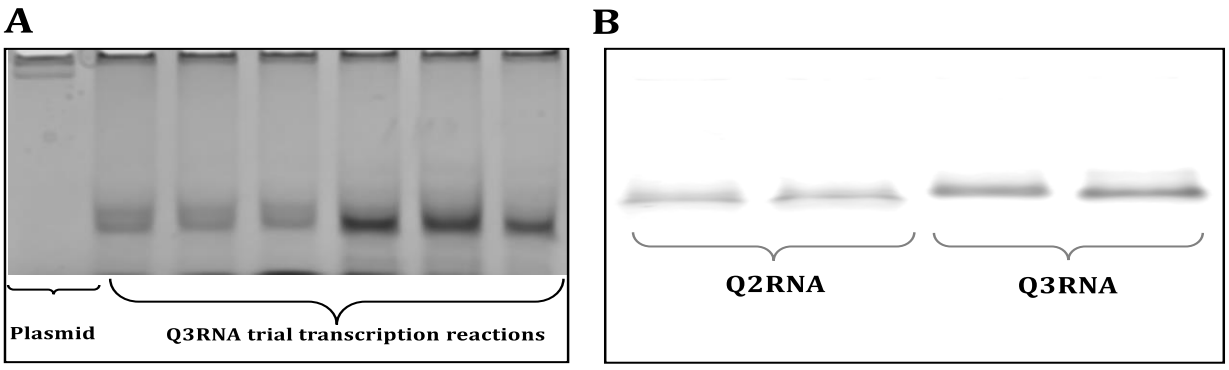

Fig. A.2. Transcribed RNA from linearized plasmid templates (A) Transcribed Q3 RNA from linearized plasmid template with each lane showing RNA yield from different transcription reaction conditions **(B)** Purified Q2 RNA and Q3 RNA by SEC. RNA samples were separated by native-Tris-borate EDTA (TBE) polyacrylamide gel electrophoresis and stained with toluidine blue.

A.2.4 *In vitro* transcribed Q2RNA and Q3RNA were stained by NMM

To test whether the *in vitro* transcribed RNAs are G4s, a reaction each containing 100-300 Pico-moles of Q2, and Q3 RNAs in a total volume of 30μL were set up with both positive and negative controls. The samples were run on 12% Native Polyacrylamide gels and stained with quadruplex specific fluorescent dye *N*-methyl mesoporphyrin IX, **(Fig. A.3).** In all the cases, Q2 RNA, Q3 RNA and hTR_{1-18} (previously characterized G4 RNA) (50) were stained by the quadruplex specific dye while all the negative controls (non-G4 RNA and DNA) did not (**Fig. A.3)**.

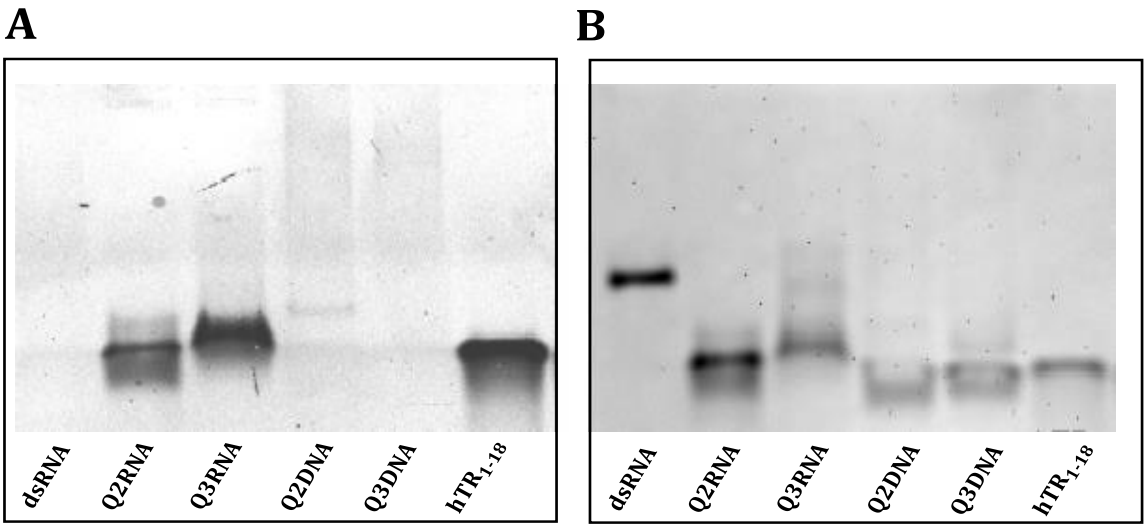

Fig. A.3. Analysis of in vitro transcribed G4 RNA by PAGE. 250 pmol of *in vitro* transcribed Q2RNA, Q3RNA were separated by Native-Tris-borate EDTA (TBE) polyacrylamide gel electrophoresis, stained with **(A)** G4-specific dye, N-methyl mesoporphyrin IX (NMM) **(B)** toluidine blue, and alongside their positive (G4: hTR_{1-18}) and negative (double stranded RNA) controls. The DNA equivalents of both G4 RNAs are also presented.

A.2.5 Spectroscopy

The UV-Visible/melting curves of 10 µM *in vitro* transcribed Q2RNA and its DNA equivalent in 50 mM Tris, pH 7.5, 100 mM KCl, 1 mM EDTA were obtained on a dual beam Evolution 260 Bio UV-Visible spectrophotometer (Thermo Scientific). Absorbance Vs temperature curves provide information on the thermal stabilities of the RNA and DNA in the of course denaturation and renaturation, which were monitored at 260 nm (**Fig. A.4 A**). The CD spectra of *in vitro* transcribed Q2RNA at a concentration of 23 µM in a 0.5 mm cell shows characteristic features at 242 nm (minimum) and 264 nm (maximum) peaks respectively. These features were reduced significantly when the RNA was melted, perhaps due to the disappearance of the base stacking present in G4s (**Fig. A.4 B**).

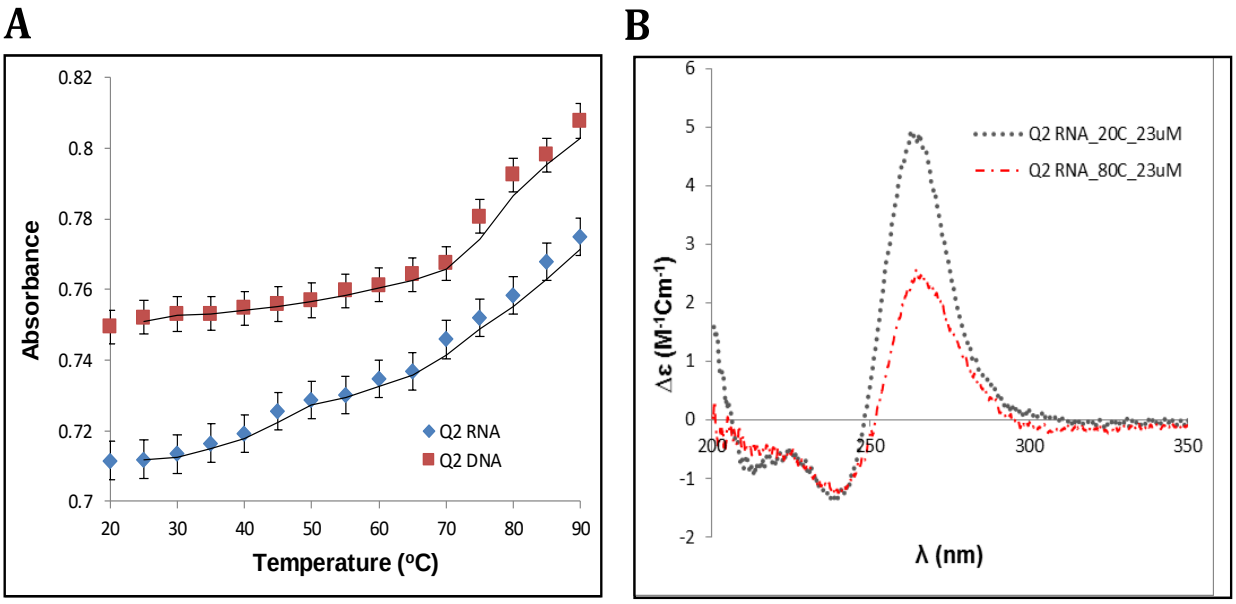

Fig. A.4. UV-Visible/melting curves and CD spectra of in vitro transcribed RNA (A) 10 µM *in vitro* transcribed Q2RNA and its DNA equivalent in 50 mM Tris, pH 7.5, 100 mM KCl **(B)** CD spectra of *in vitro* transcribed Q2RNA at a concentration of 23 µM in 0.5 mm cell showing characteristic features of G4 at 242 nm (minimum) and 264 nm (maximum) peaks respectively.

A.2.6 Dynamic Light Scattering Analysis (DLS)

As a quality control step preceding Small Angle X-ray Scattering (SAXS) analysis in determining the hydrodynamic radius, a dynamic light scattering experiment was performed on Q2 RNA at different concentrations to ensure sample monodispersity. The sample did not display any significant self-association in the concentration range analyzed, however, due to low sample concentrations, DLS data could not be collected for Q3 RNA. Dynamic light scattering data were collected on a Zetasizer Nano S system (Malvern instrument Ltd, Malvern, UK) as described in chapter 2.

A.2.7 *In vitro* transcribed Q2 RNA and Q3 RNA interact with RHAU53-105

To investigate the interaction between *in vitro* transcribed RNAs and RHAU, reactions, each containing 200 nM of either Q2 RNA or Q3 RNA, enriched with 0-1000 nM RHAU53-105 protein in a total volume of 30μL were set up. These were run on a 12% native polyacrylamide gel in TBE buffer alongside positive and negative controls in the cold room for about 5 hours. A shift of the RNAs with N-Terminal RHAU53-105 was observed which revealed an interaction between them. There was no interaction between the non-quadruplex RNA (dsRNA) and RHAU53-105 (**Fig. A.5**).

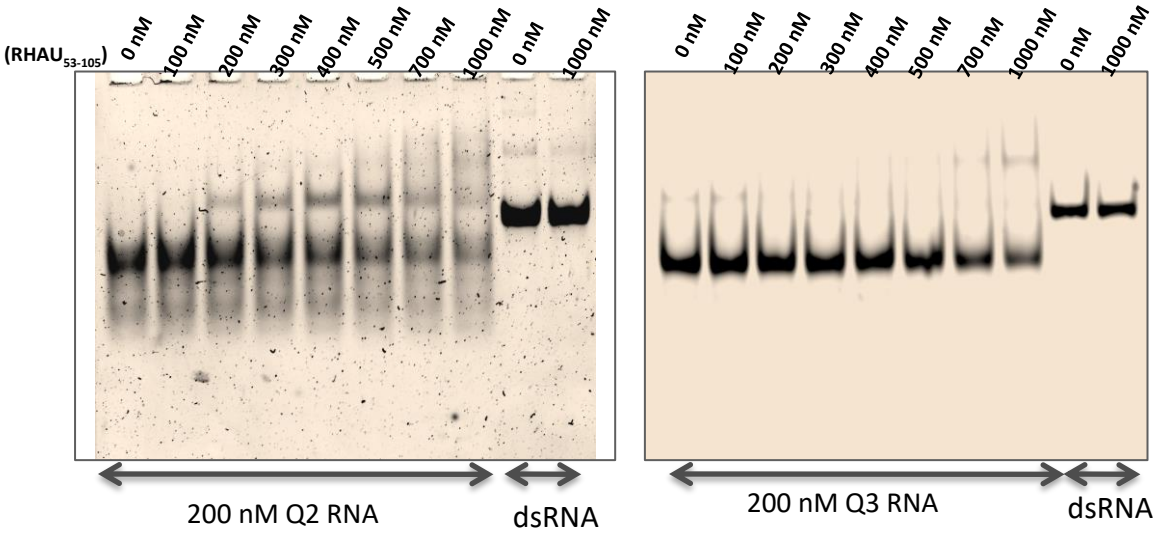

Fig. A.5. RHAU53-105 interacts with in vitro transcribed RNAs. EMSA were performed using a constant 200 nM concentration of Q2 RNA or Q3 RNA and a variable concentration from 0–1000 nM of RHAU53-105. The 12% native TBE polyacrylamide gels were stained with SYBR Gold for visualization

REFERENCES

1. Dahm R. Discovering DNA: Friedrich Miescher and the early years of nucleic acid research. Human genetics. 2008;122(6):565-81.

2. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, et al. The sequence of the human genome. science. 2001;291(5507):1304-51.

3. Budowle B, van Daal A. Extracting evidence from forensic DNA analyses: Future molecular biology directions. Biotechniques. 2009;46(5):339.

4. Nordhoff E, Kirpekar F, Roepstorff P. Mass spectrometry of nucleic acids. Mass Spectrometry Reviews. 1996;15(2):67-138.

5. Robert KM, Daryl KG, Peter AM, Victor WR. Harper's Illustrated Biochemistry. 26th ed: Lange Medical Book/McGraw-Hill; 2003 2003.

6. Todorov TI, Morris MD. Comparison of RNA, single-stranded DNA and double-stranded DNA behavior during capillary electrophoresis in semidilute polymer solutions. Electrophoresis. 2002;23(7-8):1033-44.

7. Watson JD, Crick FH. Molecular structure of nucleic acids. Nature. 1953;171(4356):737-8.

8. Baker ES, Bowers MT. B-DNA helix stability in a solvent-free environment. Journal of the American Society for Mass Spectrometry. 2007;18(7):1188-95.

9. Richmond TJ, Davey CA. The structure of DNA in the nucleosome core. Nature. 2003;423(6936):145-50.

10. Rich A, Wang AH, Nordheim A. Chemistry and Biology of Left-Handed Z DNA. Nucleic acid research: future development. 1983:11.

11. Altona Ct, Sundaralingam M. Conformational analysis of the sugar ring in nucleosides and nucleotides. New description using the concept of pseudorotation. Journal of the American Chemical Society. 1972;94(23):8205-12.

12. Mikkola S, Stenman E, Nurmi K, Yousefi-Salakdeh E, Strömberg R, Lönnberg H. The mechanism of the metal ion promoted cleavage of RNA phosphodiester bonds involves a general acid catalysis by the metal aquo ion on the departure of the leaving group. Journal of the Chemical Society, Perkin Transactions 2. 1999(8):1619-26.

13. SantaLucia Jr J, Hicks D. The thermodynamics of DNA structural motifs. Annu Rev Biophys Biomol Struct. 2004;33:415-40.

14. Coleman WB, Tsongalis GJ. Molecular diagnostics: for the clinical laboratorian: Springer Science & Business Media; 2007.

15. Huppert JL. Hunting G-quadruplexes. Biochimie. 2008;90(8):1140-8.

16. Hoogsteen K. Crystal and Molecular Structure of a Hydrogen-Bonded Complex between 1-Methylthymine and 9-Methyladenine. Acta Crystallographica. 1963;16(9):907.

17. Burge S, Parkinson GN, Hazel P, Todd AK, Neidle S. Quadruplex DNA: sequence, topology and structure. Nucleic Acids Res. 2006;34(19):5402-15.

18. Davis JT. G-quartets 40 years later: from 5'-GMP to molecular biology and supramolecular chemistry. Angew Chem Int Ed Engl. 2004;43(6):668-98.

19. Gellert M, Lipsett MN, Davies DR. Helix formation by guanylic acid. Proc Natl Acad Sci U S A. 1962;48:2013-8.

20. Keniry MA. Quadruplex structures in nucleic acids. Biopolymers. 2000;56(3):123-46.

21. Patel DJ, Phan AT, Kuryavyi V. Human telomere, oncogenic promoter and 5'-UTR G-quadruplexes: diverse higher order DNA and RNA targets for cancer therapeutics. Nucleic Acids Res. 2007;35(22):7429-55.

22. Huppert JL, Balasubramanian S. Prevalence of quadruplexes in the human genome. Nucleic Acids Res. 2005;33(9):2908-16.

23. Chambers VS, Marsico G, Boutell JM, Di Antonio M, Smith GP, Balasubramanian S. High-throughput sequencing of DNA G-quadruplex structures in the human genome. Nat Biotech. 2015;33(8):877-81.

24. Eddy J, Maizels N. Gene function correlates with potential for G4 DNA formation in the human genome. Nucleic acids research. 2006;34(14):3887-96.

25. Kikin O, D'Antonio L, Bagga PS. QGRS Mapper: a web-based server for predicting G-quadruplexes in nucleotide sequences. Nucleic acids research. 2006;34(suppl 2):W676-W82.

26. König SL, Evans AC, Huppert JL. Seven essential questions on G-quadruplexes. Biomolecular concepts. 2010;1(2):197-213.

27. Frees S, Menendez C, Crum M, Bagga PS. QGRS-Conserve: a computational method for discovering evolutionarily conserved G-quadruplex motifs. Human genomics. 2014;8(1):1.

28. Maizels N, Gray LT. The G4 genome. PLoS Genet. 2013;9(4):e1003468.

29. Cayrou C, Grégoire D, Coulombe P, Danis E, Méchali M. Genome-scale identification of active DNA replication origins. Methods. 2012;57(2):158-64.

30. Besnard E, Babled A, Lapasset L, Milhavet O, Parrinello H, Dantec C, et al. Unraveling cell type–specific and reprogrammable human replication origin signatures associated with G-quadruplex consensus motifs. Nature structural & molecular biology. 2012;19(8):837-44.

31. Bugaut A, Balasubramanian S. 5′-UTR RNA G-quadruplexes: translation regulation and targeting. Nucleic acids research. 2012;40(11):4727-41.

32. Beaudoin J-D, Perreault J-P. Exploring mRNA 3′-UTR G-quadruplexes: evidence of roles in both alternative polyadenylation and mRNA shortening. Nucleic acids research. 2013;41(11):5898-911.

33. Bang I. Untersuchungen über die Guanylsäure. Biochemische Zeitschrift. 1910;26:293-311.

34. Balasubramanian S, Hurley LH, Neidle S. Targeting G-quadruplexes in gene promoters: a novel anticancer strategy? Nature reviews Drug discovery. 2011;10(4):261-75.

35. Simonsson T. G-quadruplex DNA structures--variations on a theme. Biol Chem. 2001;382(4):621-8.

36. Arnott S, Chandrasekaran R, Marttila CM. Structures for polyinosinic acid and polyguanylic acid. Biochemical Journal. 1974;141(2):537-43.

37. Zimmerman SB, Cohen GH, Davies DR. X-ray fiber diffraction and model-building study of polyguanylic acid and polyinosinic acid. Journal of molecular biology. 1975;92(2):181-92.

38. Gray DM, Bollum F. A circular dichroism study of poly dG, poly dC, and poly dG: dC. Biopolymers. 1974;13(10):2087-102.

39. Sen D, Gilbert W. Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. 1988.

40. Rhodes D, Lipps HJ. G-quadruplexes and their regulatory roles in biology. Nucleic acids research. 2015:gkv862.

41. Blackburn EH. Telomeres: structure and synthesis. J Biol Chem. 1990;265(11):5919-21.

42. Henderson E, Hardin CC, Walk SK, Tinoco I, Blackburn EH. Telomeric DNA oligonucleotides form novel intramolecular structures containing guanine· guanine base pairs. Cell. 1987;51(6):899-908.

43. Sundquist WI, Klug A. Telomeric DNA dimerizes by formation of guanine tetrads between hairpin loops. 1989.

44. Williamson JR, Raghuraman M, Cech TR. Monovalent cation-induced structure of telomeric DNA: the G-quartet model. Cell. 1989;59(5):871-80.

45. Simonsson T, Kubista M, Pecinka P. DNA tetraplex formation in the control region of c-myc. Nucleic acids research. 1998;26(5):1167-72.

46. Biffi G, Di Antonio M, Tannahill D, Balasubramanian S. Visualization and selective chemical targeting of RNA G-quadruplex structures in the cytoplasm of human cells. Nat Chem. 2014;6(1):75-80.

47. Henderson A, Wu Y, Huang YC, Chavez EA, Platt J, Johnson FB, et al. Detection of G-quadruplex DNA in mammalian cells. Nucleic acids research. 2014;42(2):860-9.

48. Lipps HJ, Rhodes D. G-quadruplex structures: in vivo evidence and function. Trends in cell biology. 2009;19(8):414-22.

49. Booy EP, Howard R, Marushchak O, Ariyo EO, Meier M, Novakowski SK, et al. The RNA helicase RHAU (DHX36) suppresses expression of the transcription factor PITX1. Nucleic Acids Res. 2014;42(5):3346-61.

50. Booy EP, Meier M, Okun N, Novakowski SK, Xiong S, Stetefeld J, et al. 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(9):4110-24.

51. Smith FW, Lau FW, Feigon J. d(G3T4G3) forms an asymmetric diagonally looped dimeric quadruplex with guanosine 5'-syn-syn-anti and 5'-syn-anti-anti N-glycosidic conformations. Proc Natl Acad Sci U S A. 1994;91(22):10546-50.

52. Cang X, Šponer J, Cheatham TE. Explaining the varied glycosidic conformational, G-tract length and sequence preferences for anti-parallel G-quadruplexes. Nucleic acids research. 2011;39(10):4499-512.

53. Hardin CC, Watson T, Corregan M, Bailey C. Cation-dependent transition between the quadruplex and Watson-Crick hairpin forms of d(CGCG3GCG). Biochemistry. 1992;31(3):833-41.

54. Wong A, Wu G. Selective binding of monovalent cations to the stacking G-quartet structure formed by guanosine 5'-monophosphate: a solid-state NMR study. J Am Chem Soc. 2003;125(45):13895-905.

55. Cai MM, Shi XD, Sidorov V, Fabris D, Lam YF, Davis JT. Cation-directed self-assembly of lipophilic nucleosides: the cation's central role in the structure and dynamics of a hydrogen-bonded assembly. Tetrahedron. 2002;58(4):661-71.

56. Wlodarczyk A, Grzybowski P, Patkowski A, Dobek A. Effect of ions on the polymorphism, effective charge, and stability of human telomeric DNA. Photon correlation spectroscopy and circular dichroism studies. J Phys Chem B. 2005;109(8):3594-605.

57. Lee JS. The stability of polypurine tetraplexes in the presence of mono- and divalent cations. Nucleic Acids Res. 1990;18(20):6057-60.

58. Nagesh N, Chatterji D. Ammonium ion at low concentration stabilizes the G-quadruplex formation by telomeric sequence. J Biochem Biophys Methods. 1995;30(1):1-8.

59. Gill ML, Strobel SA, Loria JP. Crystallization and characterization of the thallium form of the Oxytricha nova G-quadruplex. Nucleic Acids Res. 2006;34(16):4506-14.

60. Chen FM. Sr2+ facilitates intermolecular G-quadruplex formation of telomeric sequences. Biochemistry. 1992;31(15):3769-76.

61. Venczel EA, Sen D. Parallel and antiparallel G-DNA structures from a complex telomeric sequence. Biochemistry. 1993;32(24):6220-8.

62. Smirnov I, Shafer RH. Lead is unusually effective in sequence-specific folding of DNA. J Mol Biol. 2000;296(1):1-5.

63. Parkinson GN, Lee MP, Neidle S. Crystal structure of parallel quadruplexes from human telomeric DNA. Nature. 2002;417(6891):876-80.

64. Joachimi A, Benz A, Hartig JS. A comparison of DNA and RNA quadruplex structures and stabilities. Bioorg Med Chem. 2009;17(19):6811-5.

65. Arora A, Maiti S. Differential biophysical behavior of human telomeric RNA and DNA quadruplex. J Phys Chem B. 2009;113(30):10515-20.

66. Tang C-F, Shafer RH. Engineering the quadruplex fold: nucleoside conformation determines both folding topology and molecularity in guanine quadruplexes. Journal of the American Chemical Society. 2006;128(17):5966-73.

67. Rachwal PA, Brown T, Fox KR. Effect of G-tract length on the topology and stability of intramolecular DNA quadruplexes. Biochemistry. 2007;46(11):3036-44.

68. Guo Q, Lu M, Kallenbach NR. Effect of thymine tract length on the structure and stability of model telomeric sequences. Biochemistry. 1993;32(14):3596-603.

69. Risitano A, Fox KR. Influence of loop size on the stability of intramolecular DNA quadruplexes. Nucleic Acids Research. 2004;32(8):2598-606.

70. Hazel P, Huppert J, Balasubramanian S, Neidle S. Loop-length-dependent folding of G-quadruplexes. Journal of the American Chemical Society. 2004;126(50):16405-15.

71. Miyoshi D, Nakao A, Sugimoto N. Structural transition from antiparallel to parallel G-quadruplex of d (G4T4G4) induced by Ca2+. Nucleic acids research. 2003;31(4):1156-63.

72. Xu Y, Noguchi Y, Sugiyama H. The new models of the human telomere d [AGGG (TTAGGG) 3] in K+ solution. Bioorganic & medicinal chemistry. 2006;14(16):5584-91.

73. Ambrus A, Chen D, Dai J, Bialis T, Jones RA, 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(9):2723-35.

74. Phan AT, Luu KN, Patel DJ. Different loop arrangements of intramolecular human telomeric (3+ 1) G-quadruplexes in K+ solution. Nucleic acids research. 2006;34(19):5715-9.

75. Sacca B, Lacroix L, Mergny J-L. The effect of chemical modifications on the thermal stability of different G-quadruplex-forming oligonucleotides. Nucleic acids research. 2005;33(4):1182-92.

76. Tang J, Kan Z-y, Yao Y, Wang Q, Hao Y-h, Tan Z. G-quadruplex preferentially forms at the very 3′ end of vertebrate telomeric DNA. Nucleic acids research. 2008;36(4):1200-8.

77. Smirnov I, Shafer RH. Effect of loop sequence and size on DNA aptamer stability. Biochemistry. 2000;39(6):1462-8.

78. Keniry MA, Owen EA, Shafer RH. The contribution of thymine—thymine interactions to the stability of folded dimeric quadruplexes. Nucleic acids research. 1997;25(21):4389-92.

79. Tippana R, Xiao W, Myong S. G-quadruplex conformation and dynamics are determined by loop length and sequence. Nucleic acids research. 2014:gku464.

80. Guédin A, Gros J, Alberti P, Mergny J-L. How long is too long? Effects of loop size on G-quadruplex stability. Nucleic acids research. 2010;38(21):7858-68.

81. Pandey S, Agarwala P, Maiti S. Effect of loops and G-quartets on the stability of RNA G-quadruplexes. The Journal of Physical Chemistry B. 2013;117(23):6896-905.

82. Dai J, Carver M, Punchihewa C, Jones RA, Yang D. Structure of the Hybrid-2 type intramolecular human telomeric G-quadruplex in K+ solution: insights into structure polymorphism of the human telomeric sequence. Nucleic Acids Res. 2007;35(15):4927-40.

83. Phan AT, Modi YS, Patel DJ. Two-repeat Tetrahymena telomeric d(TGGGGTTGGGGT) Sequence interconverts between asymmetric dimeric G-quadruplexes in solution. J Mol Biol. 2004;338(1):93-102.

84. Kettani A, Bouaziz S, Gorin A, Zhao H, Jones RA, Patel DJ. Solution structure of a Na cation stabilized DNA quadruplex containing G.G.G.G and G.C.G.C tetrads formed by G-G-G-C repeats observed in adeno-associated viral DNA. J Mol Biol. 1998;282(3):619-36.

85. Smith FW, Feigon J. Quadruplex structure of Oxytricha telomeric DNA oligonucleotides. Nature. 1992;356(6365):164-8.

86. Gill ML, Strobel SA, Loria JP. 205Tl NMR methods for the characterization of monovalent cation binding to nucleic acids. J Am Chem Soc. 2005;127(47):16723-32.

87. Haider S, Parkinson GN, Neidle S. Crystal structure of the potassium form of an Oxytricha nova G-quadruplex. J Mol Biol. 2002;320(2):189-200.

88. Li W, Wu P, Ohmichi T, Sugimoto N. Characterization and thermodynamic properties of quadruplex/duplex competition. FEBS letters. 2002;526(1-3):77-81.

89. Nikolova EN, Kim E, Wise AA, O'Brien PJ, Andricioaei I, Al-Hashimi HM. Transient Hoogsteen base pairs in canonical duplex DNA. Nature. 2011;470(7335):498-502.

90. Murchie AI, Lilley DM. Retinoblastoma susceptibility genes contain 5'sequences with a high propensity to form guanine-tetrad structures. Nucleic acids research. 1992;20(1):49-53.

91. Simonsson T, Sjöback R. DNA tetraplex formation studied with fluorescence resonance energy transfer. Journal of Biological Chemistry. 1999;274(24):17379-83.

92. Cayrou C, Coulombe P, Vigneron A, Stanojcik S, Ganier O, Rivals E, et al. Genome-scale analysis of metazoan replication origins reveals their organization in specific but flexible sites defined by conserved features. Genome research. 2011:gr. 121830.111.

93. Ji X, Sun H, Zhou H, Xiang J, Tang Y, Zhao C. Research progress of RNA quadruplex. Nucleic Acid Ther. 2011;21(3):185-200.

94. Kim J, Cheong C, Moore PB. Tetramerization of an RNA oligonucleotide containing a GGGG sequence. Nature. 1991;351(6324):331-2.

95. Christiansen J, Kofod M, Nielsen FC. A guanosine quadruplex and two stable hairpins flank a major cleavage site in insulin-like growth factor II mRNA. Nucleic Acids Res. 1994;22(25):5709-16.

96. Booy EP, McRae EK, Howard R, Deo SR, Ariyo EO, Dzananovic E, et al. RNA Helicase Associated with AU-rich Element (RHAU/DHX36) Interacts with the 3'-Tail of the Long Non-coding RNA BC200 (BCYRN1). J Biol Chem. 2016;291(10):5355-72.

97. Jayaraj GG, Pandey S, Scaria V, Maiti S. Potential G-quadruplexes in the human long non-coding transcriptome. RNA Biol. 2012;9(1):81-6.

98. Kumari S, Bugaut A, Huppert JL, Balasubramanian S. An RNA G-quadruplex in the 5' UTR of the NRAS proto-oncogene modulates translation. Nat Chem Biol. 2007;3(4):218-21.

99. Millevoi S, Moine H, Vagner S. G-quadruplexes in RNA biology. Wiley Interdiscip Rev RNA. 2012;3(4):495-507.

100. Biffi G, Tannahill D, McCafferty J, Balasubramanian S. Quantitative visualization of DNA G-quadruplex structures in human cells. Nat Chem. 2013;5(3):182-6.

101. Arora A, Dutkiewicz M, Scaria V, Hariharan M, Maiti S, Kurreck J. Inhibition of translation in living eukaryotic cells by an RNA G-quadruplex motif. RNA. 2008;14(7):1290-6.

102. Huppert JL, Bugaut A, Kumari S, Balasubramanian S. G-quadruplexes: the beginning and end of UTRs. Nucleic acids research. 2008;36(19):6260-8.

103. Balkwill GD, Derecka K, Garner TP, Hodgman C, Flint AP, Searle MS. Repression of translation of human estrogen receptor α by G-quadruplex formation. Biochemistry. 2009;48(48):11487-95.

104. Derecka K, Balkwill GD, Garner TP, Hodgman C, Flint AP, Searle MS. Occurrence of a quadruplex motif in a unique insert within exon C of the bovine estrogen receptor α gene (ESR1). Biochemistry. 2010;49(35):7625-33.

105. Shahid R, Bugaut A, Balasubramanian S. The BCL-2 5′ untranslated region contains an RNA G-quadruplex-forming motif that modulates protein expression. Biochemistry. 2010;49(38):8300-6.

106. Bonnal S, Schaeffer C, Créancier L, Clamens S, Moine H, Prats A-C, et al. A single IRES containing a G quartet RNA structure drives FGF-2 gene expression at four alternative translation initiation codons. J Biol Chem. 2003;278:39330-6.

107. Morris MJ, Negishi Y, Pazsint C, Schonhoft JD, Basu S. An RNA G-quadruplex is essential for cap-independent translation initiation in human VEGF IRES. Journal of the American Chemical Society. 2010;132(50):17831-9.

108. Gomez D, Guédin A, Mergny J-L, Salles B, Riou J-F, Teulade-Fichou M-P, et al. A G-quadruplex structure within the 5′-UTR of TRF2 mRNA represses translation in human cells. Nucleic acids research. 2010:gkq563.

109. Agarwala P, Pandey S, Mapa K, Maiti S. The G-quadruplex augments translation in the 5′ untranslated region of transforming growth factor β2. Biochemistry. 2013;52(9):1528-38.

110. Siddiqui-Jain A, Grand CL, Bearss DJ, Hurley LH. Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. Proc Natl Acad Sci U S A. 2002;99(18):11593-8.

111. Verma A, Halder K, Halder R, Yadav VK, Rawal P, Thakur RK, et al. Genome-wide computational and expression analyses reveal G-quadruplex DNA motifs as conserved cis-regulatory elements in human and related species. Journal of medicinal chemistry. 2008;51(18):5641-9.

112. Hershman SG, Chen Q, Lee JY, Kozak ML, Yue P, Wang L-S, et al. Genomic distribution and functional analyses of potential G-quadruplex-forming sequences in Saccharomyces cerevisiae. Nucleic acids research. 2008;36(1):144-56.

113. Fernando H, Sewitz S, Darot J, Tavaré S, Huppert JL, Balasubramanian S. Genome-wide analysis of a G-quadruplex-specific single-chain antibody that regulates gene expression. Nucleic acids research. 2009:gkp740.

114. Blackburn EH. Telomere states and cell fates. Nature. 2000;408(6808):53-6.

115. van Steensel B, Smogorzewska A, de Lange T. TRF2 protects human telomeres from end-to-end fusions. Cell. 1998;92(3):401-13.

116. Hackett JA, Feldser DM, Greider CW. Telomere dysfunction increases mutation rate and genomic instability. Cell. 2001;106(3):275-86.

117. De Lange T, Shiue L, Myers R, Cox D, Naylor S, Killery A, et al. Structure and variability of human chromosome ends. Molecular and cellular biology. 1990;10(2):518-27.

118. Kipling D, Cooke HJ. Hypervariable ultra-long telomeres in mice. 1990.

119. de Lange T. Shelterin: the protein complex that shapes and safeguards human telomeres. Genes Dev. 2005;19(18):2100-10.

120. Moyzis RK, Buckingham JM, Cram LS, Dani M, Deaven LL, Jones MD, et al. A highly conserved repetitive DNA sequence, (TTAGGG)n, present at the telomeres of human chromosomes. Proc Natl Acad Sci U S A. 1988;85(18):6622-6.

121. Wright WE, Tesmer VM, Huffman KE, Levene SD, Shay JW. Normal human chromosomes have long G-rich telomeric overhangs at one end. Genes Dev. 1997;11(21):2801-9.

122. Colgin LM, Reddel RR. Telomere maintenance mechanisms and cellular immortalization. Curr Opin Genet Dev. 1999;9(1):97-103.

123. Sfeir AJ, Chai W, Shay JW, Wright WE. Telomere-end processing the terminal nucleotides of human chromosomes. Mol Cell. 2005;18(1):131-8.

124. 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(1):39-45.

125. Xin H, Liu D, Wan M, Safari A, Kim H, Sun W, et al. TPP1 is a homologue of ciliate TEBP-&bgr; and interacts with POT1 to recruit telomerase. nature. 2007;445(7127):559-62.

126. Wang F, Podell ER, Zaug AJ, Yang Y, Baciu P, Cech TR, et al. The POT1–TPP1 telomere complex is a telomerase processivity factor. Nature. 2007;445(7127):506-10.

127. Ding H, Schertzer M, Wu X, Gertsenstein M, Selig S, Kammori M, et al. Regulation of murine telomere length by Rtel: an essential gene encoding a helicase-like protein. Cell. 2004;117(7):873-86.

128. Lin W, Sampathi S, Dai H, Liu C, Zhou M, Hu J, et al. Mammalian DNA2 helicase/nuclease cleaves G-quadruplex DNA and is required for telomere integrity. The EMBO journal. 2013;32(10):1425-39.

129. Opresko PL, von Kobbe C, Laine J-P, Harrigan J, Hickson ID, Bohr VA. Telomere-binding protein TRF2 binds to and stimulates the Werner and Bloom syndrome helicases. Journal of Biological Chemistry. 2002;277(43):41110-9.

130. Opresko PL, Mason PA, Podell ER, Lei M, Hickson ID, Cech TR, et al. POT1 stimulates RecQ helicases WRN and BLM to unwind telomeric DNA substrates. Journal of Biological Chemistry. 2005;280(37):32069-80.

131. Brosh Jr RM. DNA helicases involved in DNA repair and their roles in cancer. Nature Reviews Cancer. 2013;13(8):542-58.

132. Sun D, Lopez-Guajardo CC, Quada J, Hurley LH, Von Hoff DD. Regulation of catalytic activity and processivity of human telomerase. Biochemistry. 1999;38(13):4037-44.

133. Harley CB, Futcher AB, Greider CW. Telomeres shorten during ageing of human fibroblasts. Nature. 1990;345(6274):458-60.

134. Hastie ND, Dempster M, Dunlop MG, Thompson AM, Green DK, Allshire RC. Telomere reduction in human colorectal carcinoma and with ageing. Nature. 1990;346(6287):866-8.

135. Mehle C, Ljungberg B, Roos G. Telomere shortening in renal cell carcinoma. Cancer Res. 1994;54(1):236-41.

136. Grimes BR, Kipling D, McGill NI, Teschke C, Cross SH, Malloy P, et al. Regulation of Telomere Length in Mammalian Cells. Genomic Instability and Immortality in Cancer: Springer; 1997. p. 133-47.

137. Nugent CI, Lundblad V. The telomerase reverse transcriptase: components and regulation. Genes & development. 1998;12(8):1073-85.

138. Bryan TM, Cech TR. Telomerase and the maintenance of chromosome ends. Current opinion in cell biology. 1999;11(3):318-24.

139. Wright WE. Telemerase Activity in Human Germiine and Embryanie Tissues and Cells. Developmental genetics. 1996;18:173-9.

140. Kim NW, Piatyszek MA, Prowse KR, Harley CB, West MD, Ho PdL, et al. Specific association of human telomerase activity with immortal cells and cancer. Science. 1994;266(5193):2011-5.

141. Shay JW, Wright WE. Telomerase activity in human cancer. Current opinion in oncology. 1996;8(1):66-71.

142. Holt SE, Shay JW. Role of telomerase in cellular proliferation and cancer. Journal of cellular physiology. 1999;180(1):10-8.

143. Wright WE, Shay JW. Cellular senescence as a tumor-protection mechanism: the essential role of counting. Current opinion in genetics & development. 2001;11(1):98-103.

144. Bryan T, Englezou A, Gupta J, Bacchetti S, Reddel R. Telomere elongation in immortal human cells without detectable telomerase activity. The EMBO journal. 1995;14(17):4240.

145. Greider CW, Blackburn EH. Identification of a specific telomere terminal transferase activity in Tetrahymena extracts. Cell. 1985;43(2 Pt 1):405-13.

146. Salazar M, Thompson BD, Kerwin SM, Hurley LH. Thermally Induced DNA⊙ RNA Hybrid to G-Quadruplex Transitions: Possible Implications for Telomere Synthesis by Telomerase. Biochemistry. 1996;35(50):16110-5.

147. Schaffitzel C, Berger I, Postberg J, Hanes J, Lipps HJ, Plückthun A. In vitro generated antibodies specific for telomeric guanine-quadruplex DNA react with Stylonychia lemnae macronuclei. Proceedings of the National Academy of Sciences. 2001;98(15):8572-7.

148. Lam EYN, Beraldi D, Tannahill D, Balasubramanian S. G-quadruplex structures are stable and detectable in human genomic DNA. Nature communications. 2013;4:1796.

149. Paeschke K, Juranek S, Simonsson T, Hempel A, Rhodes D, Lipps HJ. Telomerase recruitment by the telomere end binding protein-β facilitates G-quadruplex DNA unfolding in ciliates. Nature structural & molecular biology. 2008;15(6):598-604.

150. Postberg J, Tsytlonok M, Sparvoli D, Rhodes D, Lipps HJ. A telomerase-associated RecQ protein-like helicase resolves telomeric G-quadruplex structures during replication. Gene. 2012;497(2):147-54.

151. Zahler AM, Williamson JR, Cech TR, Prescott DM. Inhibition of telomerase by G-quartet DMA structures. nature. 1991;350(6320):718-20.

152. Sun D, Thompson B, Cathers BE, Salazar M, Kerwin SM, Trent JO, et al. Inhibition of human telomerase by a G-quadruplex-interactive compound. Journal of medicinal chemistry. 1997;40(14):2113-6.

153. Read M, Harrison RJ, Romagnoli B, Tanious FA, Gowan SH, Reszka AP, et al. Structure-based design of selective and potent G quadruplex-mediated telomerase inhibitors. Proceedings of the National Academy of Sciences. 2001;98(9):4844-9.

154. Burger AM, Dai F, Schultes CM, Reszka AP, Moore MJ, Double JA, et al. The G-quadruplex-interactive molecule BRACO-19 inhibits tumor growth, consistent with telomere targeting and interference with telomerase function. Cancer Research. 2005;65(4):1489-96.

155. Monchaud D, Teulade-Fichou M-P. A hitchhiker's guide to G-quadruplex ligands. Organic & biomolecular chemistry. 2008;6(4):627-36.

156. Georgiades SN, Abd Karim NH, Suntharalingam K, Vilar R. Interaction of metal complexes with G-quadruplex DNA. Angewandte Chemie International Edition. 2010;49(24):4020-34.

157. Shi D-F, Wheelhouse RT, Sun D, Hurley LH. Quadruplex-interactive agents as telomerase inhibitors: synthesis of porphyrins and structure-activity relationship for the inhibition of telomerase. Journal of medicinal chemistry. 2001;44(26):4509-23.

158. De Cian A, Lacroix L, Douarre C, Temime-Smaali N, Trentesaux C, Riou J-F, et al. Targeting telomeres and telomerase. Biochimie. 2008;90(1):131-55.

159. Nicoludis JM, Miller ST, Jeffrey PD, Barrett SP, Rablen PR, Lawton TJ, et al. Optimized end-stacking provides specificity of N-methyl mesoporphyrin IX for human telomeric G-quadruplex DNA. Journal of the American Chemical Society. 2012;134(50):20446-56.

160. Koeppel F, Riou J-F, Laoui A, Mailliet P, Arimondo PB, Labit D, et al. Ethidium derivatives bind to G-quartets, inhibit telomerase and act as fluorescent probes for quadruplexes. Nucleic acids research. 2001;29(5):1087-96.

161. Parkinson GN, Cuenca F, Neidle S. Topology conservation and loop flexibility in quadruplex–drug recognition: crystal structures of inter-and intramolecular telomeric DNA quadruplex–drug complexes. Journal of molecular biology. 2008;381(5):1145-56.

162. Collie GW, Promontorio R, Hampel SM, Micco M, Neidle S, Parkinson GN. Structural basis for telomeric G-quadruplex targeting by naphthalene diimide ligands. Journal of the American Chemical Society. 2012;134(5):2723-31.

163. Campbell NH, Karim NHA, Parkinson GN, Gunaratnam M, Petrucci V, Todd AK, et al. Molecular basis of structure–activity relationships between salphen metal complexes and human telomeric DNA quadruplexes. Journal of medicinal chemistry. 2011;55(1):209-22.

164. Collie GW, Sparapani S, Parkinson GN, Neidle S. Structural basis of telomeric RNA quadruplex− acridine ligand recognition. Journal of the American Chemical Society. 2011;133(8):2721-8.

165. Campbell NH, Parkinson GN, Reszka AP, Neidle S. Structural basis of DNA quadruplex recognition by an acridine drug. Journal of the American Chemical Society. 2008;130(21):6722-4.

166. Haider SM, Parkinson GN, Neidle S. Structure of a G-quadruplex–ligand complex. Journal of molecular biology. 2003;326(1):117-25.

167. Moore MJ, Schultes CM, Cuesta J, Cuenca F, Gunaratnam M, Tanious FA, et al. Trisubstituted acridines as G-quadruplex telomere targeting agents. Effects of extensions of the 3, 6-and 9-side chains on quadruplex binding, telomerase activity, and cell proliferation. Journal of medicinal chemistry. 2006;49(2):582-99.

168. Schultes CM, Guyen B, Cuesta J, Neidle S. Synthesis, biophysical and biological evaluation of 3, 6-bis-amidoacridines with extended 9-anilino substituents as potent G-quadruplex-binding telomerase inhibitors. Bioorganic & medicinal chemistry letters. 2004;14(16):4347-51.

169. Fedoroff OY, Salazar M, Han H, Chemeris VV, Kerwin SM, Hurley LH. NMR-based model of a telomerase-inhibiting compound bound to G-quadruplex DNA. Biochemistry. 1998;37(36):12367-74.

170. Mergny J-L, Lacroix L, Teulade-Fichou M-P, Hounsou C, Guittat L, Hoarau M, et al. Telomerase inhibitors based on quadruplex ligands selected by a fluorescence assay. Proceedings of the National Academy of Sciences. 2001;98(6):3062-7.

171. Gabelica V, Shammel Baker E, Teulade-Fichou M-P, De Pauw E, Bowers MT. Stabilization and structure of telomeric and c-myc region intramolecular G-quadruplexes: the role of central cations and small planar ligands. Journal of the American Chemical Society. 2007;129(4):895-904.

172. Sun D, Guo K, Rusche JJ, Hurley LH. Facilitation of a structural transition in the polypurine/polypyrimidine tract within the proximal promoter region of the human VEGF gene by the presence of potassium and G-quadruplex-interactive agents. Nucleic Acids Research. 2005;33(18):6070-80.

173. Guo K, Pourpak A, Beetz-Rogers K, Gokhale V, Sun D, Hurley LH. Formation of pseudosymmetrical G-quadruplex and i-motif structures in the proximal promoter region of the RET oncogene. Journal of the American Chemical Society. 2007;129(33):10220-8.

174. Han FX, Wheelhouse RT, Hurley LH. Interactions of TMPyP4 and TMPyP2 with quadruplex DNA. Structural basis for the differential effects on telomerase inhibition. Journal of the American Chemical Society. 1999;121(15):3561-70.

175. Haq I, Trent JO, Chowdhry BZ, Jenkins TC. Intercalative G-tetraplex stabilization of telomeric DNA by a cationic porphyrin1. Journal of the American Chemical Society. 1999;121(9):1768-79.

176. Wang P, Ren L, He H, Liang F, Zhou X, Tan Z. A phenol quaternary ammonium porphyrin as a potent telomerase inhibitor by selective interaction with quadruplex DNA. ChemBioChem. 2006;7(8):1155-9.

177. Seenisamy J, Bashyam S, Gokhale V, Vankayalapati H, Sun D, Siddiqui-Jain A, et al. Design and synthesis of an expanded porphyrin that has selectivity for the c-MYC G-quadruplex structure. Journal of the American Chemical Society. 2005;127(9):2944-59.

178. Reed JE, Arnal AA, Neidle S, Vilar R. Stabilization of G-quadruplex DNA and inhibition of telomerase activity by square-planar nickel (II) complexes. Journal of the American Chemical Society. 2006;128(18):5992-3.

179. Vialas C, Pratviel G, Meunier B. Oxidative damage generated by an oxo-metalloporphyrin onto the human telomeric sequence. Biochemistry. 2000;39(31):9514-22.

180. Shin-ya K, Wierzba K, Matsuo K-i, Ohtani T, Yamada Y, Furihata K, et al. Telomestatin, a novel telomerase inhibitor from Streptomyces anulatus. Journal of the American Chemical Society. 2001;123(6):1262-3.

181. Laporte L, Thomas GJ. Structural basis of DNA recognition and mechanism of quadruplex formation by the β subunit of the Oxytricha telomere binding protein. Biochemistry. 1998;37(5):1327-35.

182. Fang G, Cech TR. The β subunit of Oxytricha telomere-binding protein promotes G-quartet formation by telomeric DNA. Cell. 1993;74(5):875-85.

183. Fry M. Tetraplex DNA and its interacting proteins. Front Biosci. 2007;12:4336-51.

184. Sun H, Bennett RJ, Maizels N. The Saccharomyces cerevisiae Sgs1 helicase efficiently unwinds GG paired DNAs. Nucleic acids research. 1999;27(9):1978-84.

185. Huber MD, Lee DC, Maizels N. G4 DNA unwinding by BLM and Sgs1p: substrate specificity and substrate-specific inhibition. Nucleic acids research. 2002;30(18):3954-61.

186. Brázda V, Hároníková L, Liao JC, Fojta M. DNA and RNA quadruplex-binding proteins. International journal of molecular sciences. 2014;15(10):17493-517.

187. Singleton MR, Dillingham MS, Wigley DB. Structure and mechanism of helicases and nucleic acid translocases. Annu Rev Biochem. 2007;76:23-50.

188. Gorbalenya AE, Koonin EV. Helicases: amino acid sequence comparisons and structure-function relationships. Current opinion in structural biology. 1993;3(3):419-29.

189. Fairman-Williams ME, Guenther UP, Jankowsky E. SF1 and SF2 helicases: family matters. Curr Opin Struct Biol. 2010;20(3):313-24.

190. Korolev S, Hsieh J, Gauss GH, Lohman TM, Waksman G. Major domain swiveling revealed by the crystal structures of complexes of E. coli Rep helicase bound to single-stranded DNA and ADP. Cell. 1997;90(4):635-47.

191. Sengoku T, Nureki O, Nakamura A, Kobayashi S, Yokoyama S. Structural basis for RNA unwinding by the DEAD-box protein Drosophila Vasa. Cell. 2006;125(2):287-300.

192. Cordin O, Banroques J, Tanner NK, Linder P. The DEAD-box protein family of RNA helicases. Gene. 2006;367:17-37.

193. Bennett RJ, Keck JL. Structure and function of RecQ DNA helicases. Critical Reviews in Biochemistry and Molecular Biology. 2010.

194. Pang PS, Jankowsky E, Planet PJ, Pyle AM. The hepatitis C viral NS3 protein is a processive DNA helicase with cofactor enhanced RNA unwinding. The EMBO Journal. 2002;21(5):1168-76.

195. Tai C-L, Chi W-K, Chen D-S, Hwang L-H. The helicase activity associated with hepatitis C virus nonstructural protein 3 (NS3). Journal of virology. 1996;70(12):8477-84.

196. Gorbalenya AE, Koonin EV, Wolf YI. A new superfamily of putative NTP-binding domains encoded by genomes of small DNA and RNA viruses. FEBS letters. 1990;262(1):145-8.

197. Hickman AB, Dyda F. Binding and unwinding: SF3 viral helicases. Current opinion in structural biology. 2005;15(1):77-85.

198. Ilyina TV, Gorbalenya AE, Koonin EV. Organization and evolution of bacterial and bacteriophage primase-helicase systems. Journal of molecular evolution. 1992;34(4):351-7.

199. Boos D, Frigola J, Diffley JF. Activation of the replicative DNA helicase: breaking up is hard to do. Current opinion in cell biology. 2012;24(3):423-30.

200. Sakakibara N, Kelman LM, Kelman Z. Unwinding the structure and function of the archaeal MCM helicase. Molecular microbiology. 2009;72(2):286-96.

201. Sharples GJ, Ingleston SM, Lloyd RG. Holliday Junction Processing in Bacteria: Insights from the Evolutionary Conservation of RuvABC, RecG, and RusA. J Bacteriol. 1999;181(18):5543-50.

202. Levin MK, Gurjar M, Patel SS. A Brownian motor mechanism of translocation and strand separation by hepatitis C virus helicase. Nature structural & molecular biology. 2005;12(5):429-35.

203. Patel SS, Donmez I. Mechanisms of helicases. Journal of Biological Chemistry. 2006;281(27):18265-8.

204. Wong I, Lohman TM. Allosteric Effects of Nucleotide Cofactors on Escherichia coli Rep Helicase--DNA Binding. Science. 1992;256(5055):350.

205. Yarranton GT, Gefter ML. Enzyme-catalyzed DNA unwinding: studies on Escherichia coli rep protein. Proceedings of the National Academy of Sciences. 1979;76(4):1658-62.

206. Yang D, Okamoto K. Structural insights into G-quadruplexes: towards new anticancer drugs. Future medicinal chemistry. 2010;2(4):619-46.

207. Phan AT. Human telomeric G-quadruplex: structures of DNA and RNA sequences. Febs Journal. 2010;277(5):1107-17.

208. de Lange T. How telomeres solve the end-protection problem. Science. 2009;326(5955):948-52.

209. Palm W, de Lange T. How shelterin protects mammalian telomeres. Annual review of genetics. 2008;42:301-34.

210. Walker JR, Zhu X-D. Post-translational modifications of TRF1 and TRF2 and their roles in telomere maintenance. Mechanisms of ageing and development. 2012;133(6):421-34.

211. Baumann P, Cech TR. Pot1, the putative telomere end-binding protein in fission yeast and humans. Science. 2001;292(5519):1171-5.

212. Lei M, Baumann P, Cech TR. Cooperative binding of single-stranded telomeric DNA by the Pot1 protein of Schizosaccharomyces pombe. Biochemistry. 2002;41(49):14560-8.

213. Loayza D, Parsons H, Donigian J, Hoke K, de Lange T. DNA Binding Features of Human POT1 A NONAMER 5′-TAGGGTTAG-3′ MINIMAL BINDING SITE, SEQUENCE SPECIFICITY, AND INTERNAL BINDING TO MULTIMERIC SITES. Journal of Biological Chemistry. 2004;279(13):13241-8.

214. Horvath MP, Schultz SC. DNA G-quartets in a 1.86 Å resolution structure of an Oxytricha nova telomeric protein-DNA complex. Journal of molecular biology. 2001;310(2):367-77.

215. Qureshi MH, Ray S, Sewell AL, Basu S, Balci H. Replication protein A unfolds G-quadruplex structures with varying degrees of efficiency. The Journal of Physical Chemistry B. 2012;116(19):5588-94.

216. Oakley GG, Patrick SM. Replication protein A: directing traffic at the intersection of replication and repair. Frontiers in bioscience: a journal and virtual library. 2010;15:883.

217. Ray S, Bandaria JN, Qureshi MH, Yildiz A, Balci H. G-quadruplex formation in telomeres enhances POT1/TPP1 protection against RPA binding. Proceedings of the National Academy of Sciences. 2014;111(8):2990-5.

218. Ballal RD, Saha T, Fan S, Haddad BR, Rosen EM. BRCA1 localization to the telomere and its loss from the telomere in response to DNA damage. Journal of Biological Chemistry. 2009;284(52):36083-98.

219. Min J, Choi ES, Hwang K, Kim J, Sampath S, Venkitaraman AR, et al. The breast cancer susceptibility gene BRCA2 is required for the maintenance of telomere homeostasis. Journal of Biological Chemistry. 2012;287(7):5091-101.

220. Neidle S. The structures of quadruplex nucleic acids and their drug complexes. Curr Opin Struct Biol. 2009;19(3):239-50.

221. Shklover J, Weisman-Shomer P, Yafe A, Fry M. Quadruplex structures of muscle gene promoter sequences enhance in vivo MyoD-dependent gene expression. Nucleic Acids Res. 2010;38(7):2369-77.

222. Soldatenkov V, Vetcher A, Duka T, Ladame S. First evidence of a functional interaction between DNA quadruplexes and poly (ADP-ribose) polymerase-1. ACS Chemical Biology. 2008;3(4):214-9.

223. Cogoi S, Zorzet S, Rapozzi V, Géci I, Pedersen EB, Xodo LE. MAZ-binding G4-decoy with locked nucleic acid and twisted intercalating nucleic acid modifications suppresses KRAS in pancreatic cancer cells and delays tumor growth in mice. Nucleic acids research. 2013;41(7):4049-64.

224. Ginisty H, Sicard H, Roger B, Bouvet P. Structure and functions of nucleolin. Journal of cell science. 1999;112(6):761-72.

225. Grinstein E, Du Y, Santourlidis S, Christ J, Uhrberg M, Wernet P. Nucleolin regulates gene expression in CD34-positive hematopoietic cells. Journal of Biological Chemistry. 2007;282(17):12439-49.

226. Angelov D, Bondarenko VA, Almagro S, Menoni H, Mongelard F, Hans F, et al. Nucleolin is a histone chaperone with FACT-like activity and assists remodeling of nucleosomes. The EMBO journal. 2006;25(8):1669-79.

227. Sparks AB, He T-C. Identification of c-MYC as a target of the APC pathway. Science. 1998;281(5382):1509-12.

228. Hanakahi L, Sun H, Maizels N. High affinity interactions of nucleolin with GG-paired rDNA. Journal of Biological Chemistry. 1999;274(22):15908-12.

229. Cogoi S, Paramasivam M, Membrino A, Yokoyama KK, Xodo LE. The KRAS promoter responds to Myc-associated zinc finger and poly (ADP-ribose) polymerase 1 proteins, which recognize a critical quadruplex-forming GA-element. Journal of Biological Chemistry. 2010;285(29):22003-16.

230. Xu Y. Human telomere RNA: A potential target for ligand recognition. Current pharmaceutical design. 2012;18(14):2096-101.

231. Sissi C, Gatto B, Palumbo M. The evolving world of protein-G-quadruplex recognition: A medicinal chemist's perspective. Biochimie. 2011;93(8):1219-30.

232. Marnef A, Standart N. Pat1 proteins: a life in translation, translation repression and mRNA decay. Biochemical Society Transactions. 2010;38(6):1602-7.

233. Marnef A, Maldonado M, Bugaut A, Balasubramanian S, Kress M, Weil D, et al. Distinct functions of maternal and somatic Pat1 protein paralogs. Rna. 2010;16(11):2094-107.

234. Brown V, Jin P, Ceman S, Darnell JC, O'Donnell WT, Tenenbaum SA, et al. Microarray identification of FMRP-associated brain mRNAs and altered mRNA translational profiles in fragile X syndrome. Cell. 2001;107(4):477-87.

235. Darnell JC, Jensen KB, Jin P, Brown V, Warren ST, Darnell RB. Fragile X mental retardation protein targets G quartet mRNAs important for neuronal function. Cell. 2001;107(4):489-99.

236. Melko M, Bardoni B. The role of G-quadruplex in RNA metabolism: involvement of FMRP and FMR2P. Biochimie. 2010;92(8):919-26.

237. von Hacht A, Seifert O, Menger M, Schütze T, Arora A, Konthur Z, et al. Identification and characterization of RNA guanine-quadruplex binding proteins. Nucleic acids research. 2014;42(10):6630-44.

238. Meier M, Patel TR, Booy EP, Marushchak O, Okun N, Deo S, et al. 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(49):35014-27.

239. Bernstein KA, Gangloff S, Rothstein R. The RecQ DNA helicases in DNA repair. Annual review of genetics. 2010;44:393.

240. Dillingham MS. Superfamily I helicases as modular components of DNA-processing machines. Biochemical Society Transactions. 2011;39(2):413-23.

241. Jankowsky E. RNA helicases at work: binding and rearranging. Trends in biochemical sciences. 2011;36(1):19-29.

242. Brosh RM, Bohr VA. Human premature aging, DNA repair and RecQ helicases. Nucleic acids research. 2007;35(22):7527-44.

243. Crabbe L, Verdun RE, Haggblom CI, Karlseder J. Defective telomere lagging strand synthesis in cells lacking WRN helicase activity. Science. 2004;306(5703):1951-3.

244. Kamath-Loeb AS, Loeb LA, Johansson E, Burgers PM, Fry M. Interactions between the Werner syndrome helicase and DNA polymerase δ specifically facilitate copying of tetraplex and hairpin structures of the d (CGG) n trinucleotide repeat sequence. Journal of Biological Chemistry. 2001;276(19):16439-46.

245. Mohaghegh P, Karow JK, Brosh Jr RM, Bohr VA, Hickson ID. The Bloom's and Werner's syndrome proteins are DNA structure-specific helicases. Nucleic acids research. 2001;29(13):2843-9.

246. Sun H, Karow JK, Hickson ID, Maizels N. The Bloom's syndrome helicase unwinds G4 DNA. Journal of Biological Chemistry. 1998;273(42):27587-92.

247. Ellis NA, Groden J, Ye T-Z, Straughen J, Lennon DJ, Ciocci S, et al. The Bloom's syndrome gene product is homologous to RecQ helicases. Cell. 1995;83(4):655-66.

248. Budhathoki JB, Ray S, Urban V, Janscak P, Yodh JG, Balci H. RecQ-core of BLM unfolds telomeric G-quadruplex in the absence of ATP. Nucleic acids research. 2014;42(18):11528-45.

249. Wu W-Q, Hou X-M, Li M, Dou S-X, Xi X-G. BLM unfolds G-quadruplexes in different structural environments through different mechanisms. Nucleic acids research. 2015:gkv361.

250. Zhang S, Grosse F. Multiple functions of nuclear DNA helicase II (RNA helicase A) in nucleic acid metabolism. Acta biochimica et biophysica Sinica. 2004;36(3):177-83.

251. Chakraborty P, Grosse F. Human DHX9 helicase preferentially unwinds RNA-containing displacement loops (R-loops) and G-quadruplexes. DNA repair. 2011;10(6):654-65.

252. Vaughn JP, Creacy SD, Routh ED, Joyner-Butt C, Jenkins GS, Pauli S, et al. 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(46):38117-20.

253. Lattmann S, Giri B, Vaughn JP, Akman SA, 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(18):6219-33.

254. Tran H, Schilling M, Wirbelauer C, Hess D, Nagamine Y. Facilitation of mRNA deadenylation and decay by the exosome-bound, DExH protein RHAU. Molecular cell. 2004;13(1):101-11.

255. Huang W, Smaldino PJ, Zhang Q, Miller LD, Cao P, Stadelman K, et al. Yin Yang 1 contains G-quadruplex structures in its promoter and 5′-UTR and its expression is modulated by G4 resolvase 1. Nucleic acids research. 2012;40(3):1033-49.

256. Lattmann S, Stadler MB, Vaughn JP, Akman SA, Nagamine Y. The DEAH-box RNA helicase RHAU binds an intramolecular RNA G-quadruplex in TERC and associates with telomerase holoenzyme. Nucleic acids research. 2011;39(21):9390-404.

257. Sexton AN, Collins K. The 5′ guanosine tracts of human telomerase RNA are recognized by the G-quadruplex binding domain of the RNA helicase DHX36 and function to increase RNA accumulation. Molecular and cellular biology. 2011;31(4):736-43.

258. Höck J, Weinmann L, Ender C, Rüdel S, Kremmer E, Raabe M, et al. Proteomic and functional analysis of Argonaute-containing mRNA–protein complexes in human cells. EMBO reports. 2007;8(11):1052-60.

259. Bicker S, Khudayberdiev S, Weiß K, Zocher K, Baumeister S, Schratt G. The DEAH-box helicase DHX36 mediates dendritic localization of the neuronal precursor-microRNA-134. Genes & development. 2013;27(9):991-6.

260. Frohn A, Eberl HC, Stöhr J, Glasmacher E, Rüdel S, Heissmeyer V, et al. Dicer-dependent and-independent Argonaute2 protein interaction networks in mammalian cells. Molecular & Cellular Proteomics. 2012;11(11):1442-56.

261. Fuller-Pace FV. DExD/H box RNA helicases: multifunctional proteins with important roles in transcriptional regulation. Nucleic acids research. 2006;34(15):4206-15.

262. Linder P. Dead-box proteins: a family affair—active and passive players in RNP-remodeling. Nucleic acids research. 2006;34(15):4168-80.

263. Fu J, Li L, Liu S, Xing X, Liu G, Lu G. Expression research for human DDX36 and mouse Ddx36 gene in the adult testis. Yi chuan xue bao= Acta genetica Sinica. 2003;30(3):201-8.

264. 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. ACTA BIOCHIMICA ET BIOPHYSICA SINICA-CHINESE EDITION-. 2002;34(5):655-61.

265. Chakraverty RK, Hickson ID. Defending genome integrity during DNA replication: a proposed role for RecQ family helicases. Bioessays. 1999;21(4):286-94.

266. Staley JP, Guthrie C. Mechanical devices of the spliceosome: motors, clocks, springs, and things. Cell. 1998;92(3):315-26.

267. Jankowsky E, Jankowsky A. The DExH/D protein family database. Nucleic acids research. 2000;28(1):333-4.

268. Linder P. Birth of the DEAD box. Nature. 1989;337:121-2.

269. Fuller-Pace FV. RNA helicases: modulators of RNA structure. Trends in cell biology. 1994;4(8):271-4.

270. Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY, et al. CDD: NCBI's conserved domain database. Nucleic Acids Res. 2015;43(Database issue):D222-6.

271. Iwamoto F, Stadler M, Chalupníková K, Oakeley E, Nagamine Y. Transcription-dependent nucleolar cap localization and possible nuclear function of DExH RNA helicase RHAU. Experimental cell research. 2008;314(6):1378-91.

272. Chen MC, Murat P, Abecassis K, Ferre-D'Amare AR, Balasubramanian S. Insights into the mechanism of a G-quadruplex-unwinding DEAH-box helicase. Nucleic Acids Res. 2015;43(4):2223-31.

273. Creacy SD, Routh ED, Iwamoto F, Nagamine Y, Akman SA, Vaughn JP. 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(50):34626-34.

274. Heddi B, Cheong VV, Martadinata H, Phan AT. Insights into G-quadruplex specific recognition by the DEAH-box helicase RHAU: Solution structure of a peptide-quadruplex complex. Proc Natl Acad Sci U S A. 2015;112(31):9608-13.

275. Ariyo EO, Booy EP, Patel TR, Dzananovic E, McRae EK, Meier M, et al. Biophysical Characterization of G-Quadruplex Recognition in the PITX1 mRNA by the Specificity Domain of the Helicase RHAU. PLoS ONE. 2015;10(12):e0144510.

276. Lai JC, Ponti S, Pan D, Kohler H, Skoda RC, Matthias P, et al. The DEAH-box helicase RHAU is an essential gene and critical for mouse hematopoiesis. Blood. 2012;119(18):4291-300.

277. Gao X, Ma W, Nie J, Zhang C, Zhang J, Yao G, et al. A G-quadruplex DNA structure resolvase, RHAU, is essential for spermatogonia differentiation. Cell death & disease. 2015;6(1):e1610.

278. Chen J-L, Greider CW. Template boundary definition in mammalian telomerase. Genes & development. 2003;17(22):2747-52.

279. Gehring WJ, Affolter M, Burglin T. Homeodomain proteins. Annual review of biochemistry. 1994;63(1):487-526.

280. Shang J, Luo Y, Clayton DA. Backfoot is a novel homeobox gene expressed in the mesenchyme of developing hind limb. Developmental dynamics. 1997;209(2):242-53.

281. Shang J, Li X, Ring HZ, Clayton DA, Francke U. Backfoot, a novel homeobox gene, maps to human chromosome 5 (BFT) and mouse chromosome 13 (Bft). Genomics. 1997;40(1):108-13.

282. Driever W. 45 The Bicoid Morphogen: Concentration-dependent Transcriptional Activation of Zygotic Target Genes during Early Drosophila Development. Cold Spring Harbor Monograph Archive. 1992;22:1221-50.

283. Berleth T, Burri M, Thoma G, Bopp D, Richstein S, Frigerio G, et al. The role of localization of bicoid RNA in organizing the anterior pattern of the Drosophila embryo. The EMBO journal. 1988;7(6):1749.

284. Driever W, Nüsslein-Volhard C. A gradient of bicoid protein in Drosophila embryos. Cell. 1988;54(1):83-93.

285. Schofield PN. Patterns, puzzles and paradigms: the riddle of the homeobox. Trends in Neurosciences. 1987;10(1):3-6.

286. Treisman J, Harris E, Wilson D, Desplan C. The homeodomain: A new face for the helix-turn-helix? Bioessays. 1992;14(3):145-50.

287. Hanes SD, Brent R. DNA specificity of the bicoid activator protein is determined by homeodomain recognition helix residue 9. Cell. 1989;57(7):1275-83.

288. Trelsman J, Gönczy P, Vashishtha M, Harris E, Desplan C. A single amino acid can determine the DNA binding specificity of homeodomain proteins. Cell. 1989;59(3):553-62.

289. Lamonerie T, Tremblay JJ, Lanctot C, Therrien M, Gauthier Y, Drouin J. Ptx1, a bicoid-related homeo box transcription factor involved in transcription of the pro-opiomelanocortin gene. Genes Dev. 1996;10(10):1284-95.

290. Lanctot C, Lamolet B, Drouin J. The bicoid-related homeoprotein Ptx1 defines the most anterior domain of the embryo and differentiates posterior from anterior lateral mesoderm. Development. 1997;124(14):2807-17.

291. Shang J, Li X, Ring HZ, Clayton DA, Francke U. Backfoot, a novel homeobox gene, maps to human chromosome 5 (BFT) and mouse chromosome 13 (Bft). Genomics. 1997;40(1):108-13.

292. Szeto DP, Ryan AK, O'Connell SM, Rosenfeld MG. P-OTX: a PIT-1-interacting homeodomain factor expressed during anterior pituitary gland development. Proc Natl Acad Sci U S A. 1996;93(15):7706-10.

293. Klopocki E, Kahler C, Foulds N, Shah H, Joseph B, Vogel H, et al. Deletions in PITX1 cause a spectrum of lower-limb malformations including mirror-image polydactyly. Eur J Hum Genet. 2012;20(6):705-8.

294. Chen Y, Knosel T, Ye F, Pacyna-Gengelbach M, Deutschmann N, Petersen I. Decreased PITX1 homeobox gene expression in human lung cancer. Lung Cancer. 2007;55(3):287-94.

295. Knosel T, Chen Y, Hotovy S, Settmacher U, Altendorf-Hofmann A, Petersen I. Loss of desmocollin 1-3 and homeobox genes PITX1 and CDX2 are associated with tumor progression and survival in colorectal carcinoma. Int J Colorectal Dis. 2012;27(11):1391-9.

296. Lord RV, Brabender J, Wickramasinghe K, DeMeester SR, Holscher A, Schneider PM, et al. Increased CDX2 and decreased PITX1 homeobox gene expression in Barrett's esophagus and Barrett's-associated adenocarcinoma. Surgery. 2005;138(5):924-31.

297. Qi D-L, Ohhira T, Fujisaki C, Inoue T, Ohta T, Osaki M, et al. Identification of PITX1 as a TERT suppressor gene located on human chromosome 5. Molecular and cellular biology. 2011;31(8):1624-36.

298. Kolfschoten IG, van Leeuwen B, Berns K, Mullenders J, Beijersbergen RL, Bernards R, et al. A genetic screen identifies PITX1 as a suppressor of RAS activity and tumorigenicity. Cell. 2005;121(6):849-58.

299. Liu D, Lobie P. Transcriptional activation of p53 by Pitx1. Cell Death & Differentiation. 2007;14(11):1893-907.

300. Yamaguchi T, Miki Y, Yoshida K. The c-Abl tyrosine kinase stabilizes Pitx1 in the apoptotic response to DNA damage. Apoptosis. 2010;15(8):927-35.

301. Ohira T, Naohiro S, Nakayama Y, Osaki M, Okada F, Oshimura M, et al. miR-19b regulates hTERT mRNA expression through targeting PITX1 mRNA in melanoma cells. Scientific reports. 2015;5.

302. Huppert JL, Balasubramanian S. G-quadruplexes in promoters throughout the human genome. Nucleic Acids Res. 2007;35(2):406-13.

303. Breitsprecher D, Schlinck N, Witte D, Duhr S, Baaske P, Schubert T. Aptamer Binding Studies Using MicroScale Thermophoresis. Nucleic Acid Aptamers: Selection, Characterization, and Application. 2016:99-111.

304. Bhattacharjee AJ, Ahluwalia K, Taylor S, Jin O, Nicoludis JM, Buscaglia R, et al. Induction of G-quadruplex DNA structure by Zn (II) 5, 10, 15, 20-tetrakis (N-methyl-4-pyridyl) porphyrin. Biochimie. 2011;93(8):1297-309.

305. Mergny JL, Li J, Lacroix L, Amrane S, Chaires JB. Thermal difference spectra: a specific signature for nucleic acid structures. Nucleic Acids Res. 2005;33(16):e138.

306. Nicoludis JM, Barrett SP, Mergny J-L, Yatsunyk LA. Interaction of human telomeric DNA with N-methyl mesoporphyrin IX. Nucleic acids research. 2012;40(12):5432-47.

307. Hammes GG. Spectroscopy for the biological sciences: John Wiley & Sons; 2005.

308. Fasman GD. Circular dichroism and the conformational analysis of biomolecules: Springer Science & Business Media; 2013.

309. Webba da Silva M. Geometric formalism for DNA quadruplex folding. Chemistry–A European Journal. 2007;13(35):9738-45.

310. Webba da Silva M, Trajkovski M, Sannohe Y, Ma'ani Hessari N, Sugiyama H, Plavec J. Design of a G-Quadruplex Topology through Glycosidic Bond Angles. Angewandte Chemie International Edition. 2009;48(48):9167-70.

311. Clark LB. Electronic spectra of crystalline guanosine: transition moment directions of the guanine chromophore. Journal of the American Chemical Society. 1994;116(12):5265-70.

312. Fülscher MP, Serrano-Andres L, Roos BO. A theoretical study of the electronic spectra of adenine and guanine. Journal of the American Chemical Society. 1997;119(26):6168-76.

313. Berne BJ, Pecora R. Dynamic light scattering: with applications to chemistry, biology, and physics: Courier Corporation; 1976.

314. Patel TR, Morris GA, Zwolanek D, Keene DR, Li J, Harding SE, et al. Nano-structure of the laminin gamma-1 short arm reveals an extended and curved multidomain assembly. Matrix Biol. 2010;29(7):565-72.

315. Hura GL, Menon AL, Hammel M, Rambo RP, Poole Ii FL, Tsutakawa SE, et al. Robust, high-throughput solution structural analyses by small angle X-ray scattering (SAXS). Nature methods. 2009;6(8):606-12.

316. Glatter OK O. Small Angle X-ray Scattering: Academic Press Inc. (London) LTD: 14. ; 1982.

317. Oliveira CLP. Investigating Macromolecular Complexes in Solution by Small Angle X-Ray Scattering: INTECH Open Access Publisher; 2011.

318. Konarev PV, Volkov VV, Sokolova AV, Koch MH, Svergun DI. PRIMUS: a Windows PC-based system for small-angle scattering data analysis. Journal of applied crystallography. 2003;36(5):1277-82.

319. Vachette P, Koch MH, Svergun DI. Looking behind the beamstop: X-ray solution scattering studies of structure and conformational changes of biological macromolecules. Methods in enzymology. 2003;374:584-615.

320. Svergun DI. Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. Biophys J. 1999;76(6):2879-86.

321. Koch MH, Vachette P, Svergun DI. Small-angle scattering: a view on the properties, structures and structural changes of biological macromolecules in solution. Quarterly reviews of biophysics. 2003;36(02):147-227.

322. Svergun DI. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. J Appl Crystallogr. 1992;25:495-503.

323. Franke D, Svergun DI. DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. Journal of applied crystallography. 2009;42(2):342-6.

324. Franke D, Svergun DI. DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. J Appl Crystallogr. 2009;42:342-6.

325. Kirkpatrick S, Gelatt CD, Vecchi MP. Optimization by simmulated annealing. science. 1983;220(4598):671-80.

326. Kozin MB, Svergun DI. Automated matching of high-and low-resolution structural models. Journal of applied crystallography. 2001;34(1):33-41.

327. Dzananovic E, Patel TR, Deo S, McEleney K, Stetefeld J, McKenna SA. Recognition of viral RNA stem-loops by the tandem double-stranded RNA binding domains of PKR. RNA. 2013;19(3):333-44.

328. Volkov VV, Svergun DI. Uniqueness of ab initio shape determination in small-angle scattering. J Appl Crystallogr. 2003;36:860-4.

329. de la Torre JG, Huertas ML, Carrasco B. Calculation of hydrodynamic properties of globular proteins from their atomic-level structure. Biophysical Journal. 2000;78(2):719-30.

330. da Silva MW. NMR methods for studying quadruplex nucleic acids. Methods. 2007;43(4):264-77.

331. Feigon J, Koshlap KM, Smith FW. [10] 1 H NMR spectroscopy of DNA triplexes and quadruplexes. Methods in enzymology. 1995;261:225-55.

332. Patel DJ, Tonelli AE. Assignment of the proton nmr chemical shifts of the T□ N3H and G□ N1H proton resonances in isolated AT and GC Watson-Crick base pairs in double-stranded deoxy oligonucleotides in aqueous solution. Biopolymers. 1974;13(10):1943-64.

333. Günther H. NMR spectroscopy: basic principles, concepts and applications in chemistry: John Wiley & Sons; 2013.

334. Bodenhausen G, Ruben DJ. Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. Chemical Physics Letters. 1980;69(1):185-9.

335. Sacca B, Lacroix L, Mergny JL. The effect of chemical modifications on the thermal stability of different G-quadruplex-forming oligonucleotides. Nucleic Acids Res. 2005;33(4):1182-92.

336. Zhang AY, Bugaut A, Balasubramanian S. A sequence-independent analysis of the loop length dependence of intramolecular RNA G-quadruplex stability and topology. Biochemistry. 2011;50(33):7251-8.

337. Collie GW, Haider SM, Neidle S, Parkinson GN. A crystallographic and modelling study of a human telomeric RNA (TERRA) quadruplex. Nucleic Acids Res. 2010;38(16):5569-80.

338. Frees S, Menendez C, Crum M, Bagga PS. QGRS-Conserve: a computational method for discovering evolutionarily conserved G-quadruplex motifs. Hum Genomics. 2014;8:8.

339. Verma A, Halder K, Halder R, Yadav VK, Rawal P, Thakur RK, et al. Genome-wide computational and expression analyses reveal G-quadruplex DNA motifs as conserved cis-regulatory elements in human and related species. J Med Chem. 2008;51(18):5641-9.

340. Brooks TA, Kendrick S, Hurley L. Making sense of G-quadruplex and i-motif functions in oncogene promoters. FEBS J. 2010;277(17):3459-69.

341. Todd AK, Johnston M, Neidle S. Highly prevalent putative quadruplex sequence motifs in human DNA. Nucleic Acids Res. 2005;33(9):2901-7.

342. Hershman SG, Chen Q, Lee JY, Kozak ML, Yue P, Wang LS, et al. Genomic distribution and functional analyses of potential G-quadruplex-forming sequences in Saccharomyces cerevisiae. Nucleic Acids Res. 2008;36(1):144-56.

343. Huppert JL, Bugaut A, Kumari S, Balasubramanian S. G-quadruplexes: the beginning and end of UTRs. Nucleic Acids Res. 2008;36(19):6260-8.

344. Eddy J, Maizels N. Conserved elements with potential to form polymorphic G-quadruplex structures in the first intron of human genes. Nucleic Acids Res. 2008;36(4):1321-33.

345. Crabbe L, Verdun RE, Haggblom CI, Karlseder J. Defective telomere lagging strand synthesis in cells lacking WRN helicase activity. Science. 2004;306(5703):1951-3.

346. Sarkies P, Reams C, Simpson LJ, Sale JE. Epigenetic instability due to defective replication of structured DNA. Mol Cell. 2010;40(5):703-13.

347. Sun H, Karow JK, Hickson ID, Maizels N. The Bloom's syndrome helicase unwinds G4 DNA. J Biol Chem. 1998;273(42):27587-92.

348. Wu Y, Shin-ya K, Brosh RM, Jr. FANCJ helicase defective in Fanconia anemia and breast cancer unwinds G-quadruplex DNA to defend genomic stability. Mol Cell Biol. 2008;28(12):4116-28.

349. Brazda V, Haronikova L, Liao JC, Fojta M. DNA and RNA quadruplex-binding proteins. Int J Mol Sci. 2014;15(10):17493-517.

350. Gao X, Ma W, Nie J, Zhang C, Zhang J, Yao G, et al. A G-quadruplex DNA structure resolvase, RHAU, is essential for spermatogonia differentiation. Cell Death Dis. 2015;6:e1610.

351. Rogers GW, Jr., Richter NJ, Merrick WC. Biochemical and kinetic characterization of the RNA helicase activity of eukaryotic initiation factor 4A. J Biol Chem. 1999;274(18):12236-44.

352. Giri B, Smaldino PJ, Thys RG, Creacy SD, Routh ED, Hantgan RR, et al. G4 resolvase 1 tightly binds and unwinds unimolecular G4-DNA. Nucleic Acids Res. 2011;39(16):7161-78.

353. Nicoludis JM, Miller ST, Jeffrey PD, Barrett SP, Rablen PR, Lawton TJ, et al. Optimized end-stacking provides specificity of N-methyl mesoporphyrin IX for human telomeric G-quadruplex DNA. J Am Chem Soc. 2012;134(50):20446-56.

354. Kypr J, Kejnovska I, Renciuk D, Vorlickova M. Circular dichroism and conformational polymorphism of DNA. Nucleic Acids Res. 2009;37(6):1713-25.

355. Randazzo A, Spada GP, da Silva MW. Circular dichroism of quadruplex structures. Top Curr Chem. 2013;330:67-86.

356. Wieland M, Hartig JS. RNA quadruplex-based modulation of gene expression. Chem Biol. 2007;14(7):757-63.

357. Lam EY, Beraldi D, Tannahill D, Balasubramanian S. G-quadruplex structures are stable and detectable in human genomic DNA. Nat Commun. 2013;4:1796.

358. Egli M, Portmann S, Usman N. RNA hydration: a detailed look. Biochemistry. 1996;35(26):8489-94.

359. Collie GW, Haider SM, Neidle S, Parkinson GN. A crystallographic and modelling study of a human telomeric RNA (TERRA) quadruplex. Nucleic acids research. 2010:gkq259.

360. Balasubramanian S, Neidle S. G-quadruplex nucleic acids as therapeutic targets. Current opinion in chemical biology. 2009;13(3):345-53.

361. Bugaut A, Balasubramanian S. A sequence-independent study of the influence of short loop lengths on the stability and topology of intramolecular DNA G-quadruplexes. Biochemistry. 2008;47(2):689-97.

362. Lane AN, Chaires JB, Gray RD, Trent JO. Stability and kinetics of G-quadruplex structures. Nucleic acids research. 2008;36(17):5482-515.

363. Ghimire C, Park S, Iida K, Yangyuoru P, Otomo H, Yu Z, et al. Direct quantification of loop interaction and π–π stacking for G-quadruplex stability at the submolecular level. Journal of the American Chemical Society. 2014;136(44):15537-44.

364. Rouleau SG, Beaudoin J-D, Bisaillon M, Perreault J-P. Small antisense oligonucleotides against G-quadruplexes: specific mRNA translational switches. Nucleic acids research. 2014:gku1311.

365. Jodoin R, Bauer L, Garant J-M, Laaref AM, Phaneuf F, Perreault J-P. The folding of 5′-UTR human G-quadruplexes possessing a long central loop. Rna. 2014;20(7):1129-41.