Discovery and Characterization of Quadruplex Remodeling Proteins

by

Ewan Kenneth Strome McRae

A Thesis submitted to the Faculty of Graduate Studies of

The University of Manitoba

in partial fulfilment of the requirements of the degree of

DOCTOR OF PHILOSOPHY

Department of Chemistry

University of Manitoba

Winnipeg

ABSTRACT

Guanine quadruplexes are exceptionally stable four stranded structures of nucleic acids that play an important role in regulation of oncogene expression, cell cycle and telomere maintenance. The study of these regulatory pathways is an emergent field and currently little is known about how proteins interact with these structures and control their function. Using a mass spectrometry guided approach, we identified many proteins that can interact with guanine quadruplexes or complexes thereof. From this list we further validated DDX21 can bind directly to RNA guanine quadruplexes through interactions with a unique nucleic acid binding domain in its C-terminus, we did this using a comprehensive suite of biochemical and biophysical techniques. As well as probing binding to quadruplexes we developed a nuclease sensitivity assay that shows that DDX21 binding to quadruplexes destabilizes them sufficiently to allow for strand excision by T1 ribonuclease enzymes, an ability that had only previously been observed by two other proteins. Furthermore, we showed that DDX21 can regulate expression of several important genes in a manner dependent upon its quadruplex binding ability. Together, our results have shed light on the mode by which DDX21 can recognize RNA quadruplex structures and laid the groundwork for understanding the complex network of interactions and enzyme activity which controls quadruplex regulatory function in human pathophysiology.

ACKNOWLEDGEMENT

I would first like to thank my thesis advisor, Sean A. McKenna. Professor McKenna allowed and encouraged me to independently conduct my research and writing of research papers and funding applications, but would always be willing to offer feedback, insights and advise to improve my work.

I would also like to thank the expert technical staff, Dr. Peyman Ezzati & Dr. David E. Davidson, as well as Professor Richard Fahlman of the University of Alberta. Without their expertise the Mass Spectrometry and NMR experiments could not have been successfully conducted. Furthermore, to Dr. Jörg Stetefeld, Dr. Joe O'Neil, Dr. Francis Lin and Dr. Peter Pelka I am grateful for their support and use of their instrumentation for MST, CD, FACS and luciferase experiments, respectively. A special gratitude goes to Dr. Evan Booy for his patience and kindness when teaching me numerous different cell and molecular biology methodologies as well as the many invaluable discussions had during coffee breaks. Also deserving of acknowledgment are the readers of this thesis and specifically those on my committee who provided me with excellent feedback and poignant questions regarding my work.

Finally, I must express my profound gratitude both to my mother as well as to my partner Kaitlyn for providing me with unfailing support and continuous encouragement throughout my years of study and through the process of research and writing this thesis. This accomplishment would not have been possible without them.

Thank you,

Ewan McRae

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1: Introduction

1. On Characterizing the Interactions between Proteins and Guanine Quadruplex Structures of Nucleic Acids

1.1.1 Preface

This introduction is adapted from a review paper, published in the Journal of Nucleic Acids.

Mcrae, E. K. S., Booy, E. P., Padilla-meier, G. P. & Mckenna, S. A. (2017) On characterizing the interactions between proteins and quadruplex structures of nucleic acids. J. Nucleic Acids, 2017**,** 1–11.

Beginning with an introduction on Quadruplexes (G4s) and then focusing on the current understanding of G4 recognition by proteins, the following review should adequately familiarize G4s and contextualize the original research to be presented in the subsequent chapters. In addition to this review paper, there is a general review on helicase proteins and a thorough evaluation of the available literature on the helicase protein DDX21.

1.1.2 Contribution of authors

Ewan McRae wrote the manuscript for this review, Evan Booy and Pauline Meier assisted with editing the manuscript and generating figures.

1.1.3 Introduction to Quadruplexes

Nucleic acid polymers that are rich in guanines have the potential to form a four-stranded structure called a guanine quadruplex (G4). When four guanines are arranged in a plane they are stabilized by hydrogen bonds between the Watson-Crick and Hoogsteen faces of adjacent guanines. These guanine tetrads have extensive surface area for pi-pi base stacking with other tetrads which contributes greatly to the thermodynamic stability of the G4 structure. All guanine G4s share this core feature of stacked tetrads of guanines, which is likely a common feature contributing to the recognition of G4s by proteins. Formation of G4 is dependent upon the presence of monovalent cations that relieve negative electrostatic charge repulsion from the O6 of Guanines concentrated in the center of the tetrad. Positioning of monovalent cations (typically sodium or potassium) in the center of the G4, between two stacked tetrads, relieves the repulsive effect of the negative electrostatic potential[1,2].

G4s can have greatly different topology based on the relative orientation of the phosphodiester backbone connecting the runs of guanines. If all the strands are parallel with respect to the 5' and 3' orientation of the ribose sugars, the loop sections between the runs of guanine must connect from the top to the bottom of the G4 in what is known as a propeller orientation (**Figure 1A)**. The glycosidic bond angles in parallel G4s are typically all in the *anti*-conformation, making this the preferred strand orientation for ribonucleic acids (RNA) G4s as it better accommodates the 2’ hydroxyl group. Similarly, locked nucleic acids (LNA) that use 2’-O,4’-C methylene linkages to force a C3’-*endo* sugar pucker will also strongly favour the all *anti* glycosidic bond conformation afforded by a parallel G4s[3]. For this reason, LNA have been utilized when forcing a parallel G4 topology is desired. For anti-parallel strands the loops are referred to as lateral if connecting adjacent strands and diagonal if connecting strands opposite them; these create a mixture of *syn*

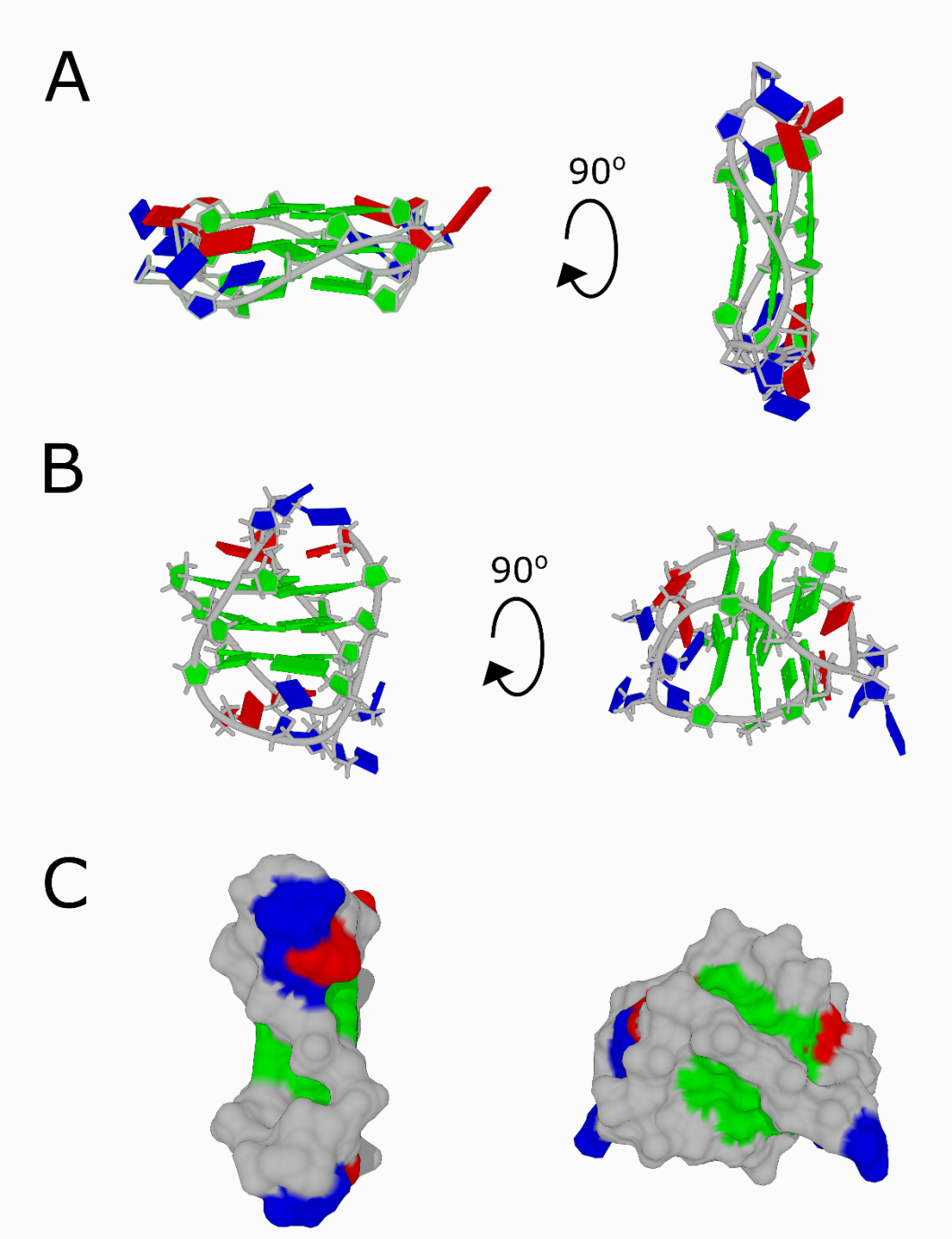


Figure 1-1: Quadruplex Topology

Side view of the human telomere guanine quadruplex in parallel (A) and antiparallel (B) conformations. Space filling models (C) for parallel (left) and antiparallel (right) quadruplexes. PDBs 1KF1 (A) and 143D (B). Guanines shown in green, thymines shown in blue, adenines shown in red and the phospho-ribose backbone shown in grey.

and *anti* glycosidic bond angles and are the structures most often observed in deoxyribonucleic acids (DNA) G4s[4] (**Figure 1-1B)**.

Much like other nucleic acid secondary structures, distinct grooves created by the sugar-phosphate backbone can be observed in the space filling models of G4 structures (**Figure 1-1C**). These grooves as well as the large surface of the tetrads on either end of the G4 and the loops themselves represent the three feasible interaction sites for protein binding partners. Lateral and diagonal loops in anti-parallel G4 leave the grooves accessible but would restrict access to the tetrad face. Propeller-like loops as seen in the parallel G4 pass directly over the grooves, allowing for intramolecular interactions of loop bases with the grooves, thereby obscuring the grooves but leaving the tetrad face more exposed for protein or ligand binding (**Figure 1-1**). Amongst the myriad of small molecule G4-binding ligands that have been synthesized one can find examples of ligands binding grooves, loop sequences and tetrad faces of G4s. In some cases, binding of these ligands has been shown to stabilize, destabilize and even alter the conformation of the G4[5–7]. It would not be unreasonable to hypothesize that as our knowledge of protein-G4 interactions increases we will have examples of proteins that do the same.

Bioinformatic analyses of the human genome identified some ~350,000 unique, putative, G4 forming sites (PGS)[8] and high throughput sequencing methods have identified a further 450,000[9]. Interestingly, there appears to be an evolutionary pressure against PGS in protein coding regions and an enrichment of PGS in other non-coding regions when compared to pseudo-randomly generated DNA sequences. This enrichment of PGS in non-coding regions supports the idea that G4s may be key regulatory elements. Moreover, there is a statistical enrichment of PGS in oncogene promoter regions, some of which have been confirmed to exist in vivo by G4 specific cross-linking studies[10] and other biochemical investigations[11–14]. It has also become clear over the past decade that G4s play an important role in maintenance of telomere length[15–17] and ribosome biogenesis[18], which are dis-regulated in ~80% of cancers[19]. Furthermore, the expansion repeats of guanine rich regions in mRNA have been linked to two common forms of neurological disorders, Fragile X mental retardation[20] and familial Amyotrophic Lateral Sclerosis (ALS)[21].

There exists some controversy as to whether G4 can form *in vivo* and while significant effort has gone towards testing this question experimentally; observing the structure of nucleic acids *in*

vivo is a monumental task that is prone to biasing dependent on the method used to determine structure. One method that is used to visualize unique motifs in cells is immunofluorescence. The Balasubramanian group has developed antibodies to selectively target both DNA[22,23] and RNA

G4[24] and successfully used them to visualize these structures *in vivo*. Furthermore, one could reason that the physiological effects of quadruplex-specific small molecules on cells is evidence of quadruplex formation *in vivo.* In these cases, the arguments can be made that the antibody or small molecule are merely inducing the formation of G4 and that the structures do not exist in the absence of these probes or that the specificity of the antibodies/drugs has changed in a cellular context and that they are no longer binding to G4 but exerting an effect through interactions with other structures. A recent study using high-throughput sequencing and *in vivo* chemical modifications has indicated that RNA G4s are globally unfolded in some mammalian cell lines[25]. A similar critique can be made, that the chemicals added for modification of nucleic acids could be altering their structure *in vivo* or that the concentration of chemical reagent added results in non-specific effects. Regardless of being able to prove the existence of G4s *in vivo,* many proteins have highly conserved G4 specific binding domains that would be hard to rationalize the persistence of were they not for some biological purpose. One way to probe this purpose is by disruption of the protein-G4 interaction.

Targeting G4-protein interactions with small molecule drugs has potential to treat a multitude of diseases[17,26,27]. Currently there is one such compound in clinical trials for cancer treatment, that is CX-3543 or Quarfloxin. While originally selected for its interaction with the MYC G4, CX-3543 has been shown to localize to cell nucleoli where it inhibits Pol I transcription by blocking the interaction between nucleolin and G4s in rDNA[18]. As well as being targets of small molecule therapeutics, G4s themselves have been employed as therapeutics to target specific proteins. One such example are thrombin binding DNA aptamers that inhibit thrombin-induced platelet aggregation and clot-bound thrombin, these have potential for use as anticoagulants during various surgeries and are currently in clinical trials[28]. Although thrombin is not by nature a G4-binding protein, G4 aptamers display remarkable affinity and specificity for regions of thrombin that normally interact with other proteins.

To better understand the mechanism of G4-mediated cellular processes and how small molecules interfere with them, it is imperative to understand how proteins recognize their endogenous G4 binding partners. While many G4-binding proteins have been identified and validated[29], there is relatively little information on how they specifically recognize G4s. Typical approaches to study this include altering some of the features of either the G4 substrate or protein, subsequently assessing how these mutations affect the interaction and inferring from this their role in G4 recognition. This method is relatively simple and inexpensive to perform and can yield useful information, however it is not always trivial to show that such mutations are directly mediating the interaction and not indirectly affecting the interaction by altering the conformation of the mutated species. Small-angle X-ray scattering is a solution-based technique that can be used to determine low-resolution models of protein and nucleic acid. Though it does not provide enough resolution to distinguish individual atoms, it can provide insight into the shape and orientation of species in a protein nucleic-acid complex that can be complementary to other data obtained from mutation analysis and/or other higher-resolution structural methods such as NMR and X-ray crystallography. While there are currently over 250 high-resolution NMR and X-ray structures of G4s deposited on the Research Collaboratory for Structural Bioinformatics (RCSB) website and nucleic acid database (NDB), comparatively few high-resolution structures of protein interacting with G4 have been solved. The most extensively studied interaction is between G4 containing DNA aptamers and thrombin, on which a number of structures have been deposited (PDB: 1HAP, 1HAO, 3QLP, 4DII, 4DIH, 5CMX, 4I7Y, 4LZ1, 4LZ4, 5LUW, 5LUY, 5EW1 & 5EW2). More recently the structure of a peptide from the prion protein bound by an RNA G4 aptamer has been solved (2RSK, 2RU7). Currently, to our knowledge, only 4 high-resolution structures have been solved showing G4-peptide interactions involving proteins that interact with G4 as part of their normal cellular function (2LA5, 5DE5, 2N21, 5VHE). Herein we will discuss the insights gained from these structures as well as some key mutational analysis studies and look at the benefits, as well as drawbacks, of these approaches.

1.1.4 Thrombin-binding by G4 DNA aptamers

The development of thrombin-binding DNA aptamers for their anti-coagulant properties has been ongoing since the early 1990s; although thrombin is not naturally a G4-binding protein the 15 base DNA G4 (TBA), developed through SELEX, can tightly bind thrombin and inhibit fibrin clot formation[30,31]. This nucleic acid sequence forms an antiparallel G4 with two guanine

tetrads connected by three lateral loops, one face of the G4 is covered by two TT loops and the other by a TTA loop. Initial high-resolution structures of TBA (1HAP, 1HAO) showed that the fibrinogen binding site of thrombin (exosite I) was likely interacting with the loops of TBA; however, due to lack of electron density in the loop regions it was unclear whether this interaction was with the TT loops or TGA loop[32]. 15 years later 3 structures (3QLP, 4DII, 4DIH) were produced from the lab of Filomena Sica that definitively show the TT loops acting as a pincer that bind either side of the protruding region of exosite I utilizing multiple polar and hydrophobic contacts[33,34].

Many variants of the aptamer have been made that enhance the stability, affinity for and inhibition of thrombin. One such potent second-generation thrombin aptamer is the HD22-27mer[35] for which a high resolution X-ray structure (4IY7) of the aptamer-thrombin complex has been solved[36] (**Figure 1-2A)**. HD22-27mer forms a mixed duplex-G4 structure where the duplex is directly enchained to the G4. The HD22-27mer forms an unusual pseudo-G4 topology with four loops, where the first loop connects two guanines in the same tetrad and the remaining loops form a more typical lateral antiparallel loop pattern. The 2^{nd} and 4^{th} of these loops interact with each other through Watson-Crick A-T hydrogen bonds and cap one side of the G4, while the intra-tetrad loop and the 3^{rd} loop also form A-T hydrogen bonds and cap the opposite side of the G4. In contrast to the structure of thrombin and TBA, protein recognition of HD22-27mer involves the extended double stranded region as well as a bulged-out thymine and the pseudo-G4 core, and appears to involve exosite II and not exosite I. G20 is the last guanine before the duplex region and forms three polar contacts with Arginine 93, two through ribose hydroxyl groups and one to its phosphate group. Loops 2 and 4 of the G4 contribute to the stability of the complex by forming multiple hydrophobic interactions with thrombin but unlike with TBA, only thymine 9 has multiple polar contacts with exosite II.

Recently a 31 nucleotide third-generation thrombin aptamer, RE31, was crystalized in complex with thrombin[37] (**Figure 1-2B)**. Like the HD22-27mer, RE31 has an extended duplex region as well as a two tetrad G4. Unlike HD22-27mer the duplex region of RE31 exhibits continuous base stacking with the G4 rather and the duplex region does not appear to be important for binding to thrombin. Despite the presence of a G4-duplex junction, the binding of RE31 to

thrombin is much more similar to TBA than HD22-27mer, both molecules use the TT loops to interact with exosite I through hydrogen bonds and hydrophobic interactions.

Despite the persistence of a two-tetrad G4 in these thrombin binding aptamers, interactions with the G4 core is not observed in any of the structures. Instead, the main site of interaction, which is common between all the structures mentioned above, is with the second and fourth loop of the G4. While TBA and RE31 use mainly polar contacts and some ancillary hydrophobic interactions to interact with exosite I of thrombin, the G4 loops of HD22-27mer uses mostly hydrophobic interactions to bind exosite II. This common mechanism of binding to different sites of thrombin suggest that the G4 acts as a scaffold that presents the single stranded loops to the binding pocket on the protein.

1.1.5 Interaction between the RGG motif and G4 structures

Motifs rich in arginine and glycine have been known to play functional roles in numerous physiological processes, usually involving nucleic acid interactions, for many decades[38]. Glycine is the most flexible of the amino acid side chains in terms of its accessible dihedral angles, thus sequences rich in glycine can sample a large variety of conformations allowing them to be very promiscuous binding partners. Arginine is a positively charged amino acid with delocalized pi electrons. This allows for favorable electrostatic interactions with the phosphate backbone of nucleic acids, as well as potential base stacking with the nitrogenous bases and hydrogen bonding in as many as three different directions via its guanidinium moiety. Many of the currently known G4-binding proteins have RGG/RG motifs[29,38] some of which have been shown

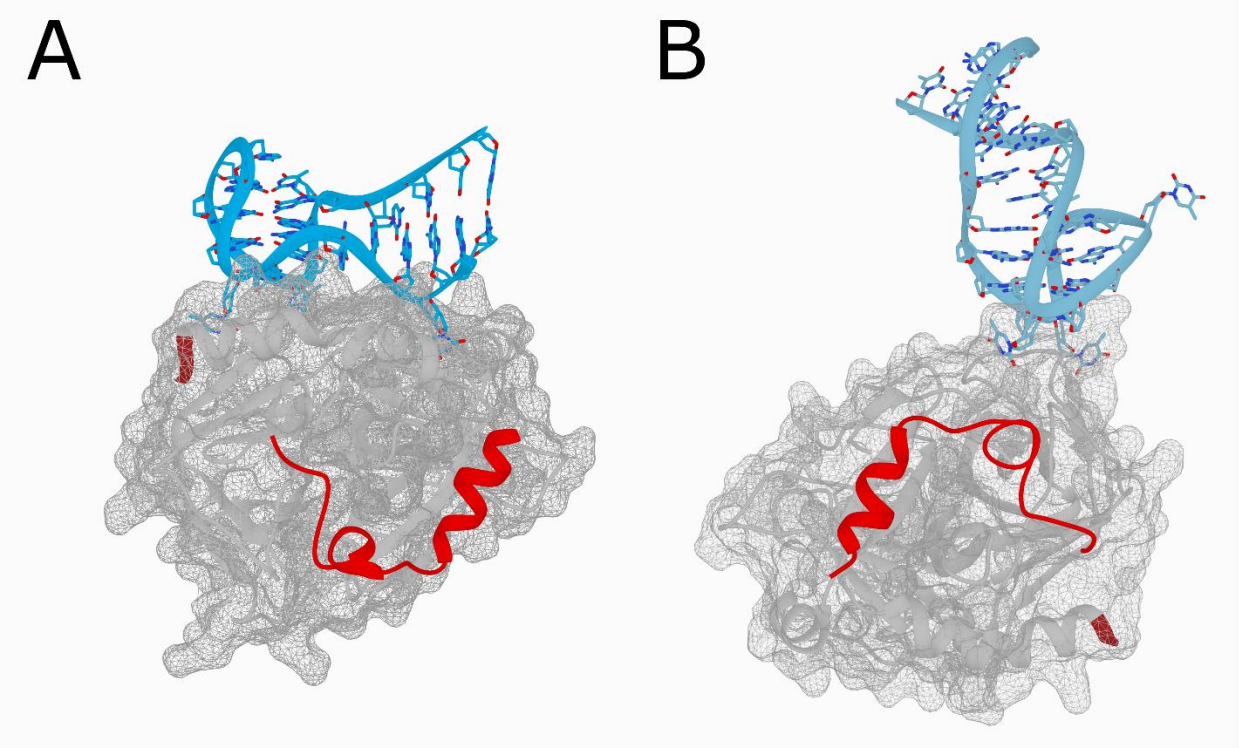


Figure 1-2: Recognition of quadruplexes by thrombin

X-ray structures of HD22-27mer (A) and RE31 (B) guanine quadruplex-containing aptamers bound to thrombin, for ease of differentiating the thrombin binding sites: shown in red ribbon is the light chain of thrombin, shown in grey ribbon is the heavy chain of thrombin with the C-terminal nub of the heavy chain shown in red. DNA aptamers shown in blue. PDBs 4I7Y (left) and 5CMX(right).

to be critical to G4 binding and recognition, including the translated in sarcoma/fused in sarcoma (TLS/FUS)[39], Nucleolin (NCL)[40], Ewings Sarcoma protein (EWS)[41], Epstein Barr virus nuclear antigen 1 EBNA1[42], DDX21[43] and fragile X mental retardation protein (FMRP)[44].

Since the most obvious structurally unique feature of a G4 is the tetrad face, one might expect the arginine of the RGG motifs to contribute to base stacking interactions with the terminal guanine tetrads. However, there is only indirect evidence that the tetrad face plays a role in RGG mediated interactions. This indirect evidence is that most of the small molecule G4-binding ligands have been shown to interact tightly with the tetrad face via pi-pi stacking interactions and two of such molecules, BRACO-19 and CX-3543, have been shown to disrupt interactions between G4-binding proteins (EBNA1)[42] and NCL[18] that contain the RGG motif. While this indirectly implicates competition with the protein for the tetrad face of the G4, these compounds also bind electrostatically to the grooves and loops of the G4[45,46] possibly preventing protein-groove interactions or re-orienting the loop structure to disrupt protein binding in such a manner.

1.1.6 Recognition of phospho-ribose backbone of G4 loops by RGG/RG domains

Biochemical investigations in the lab of Takanori Oyoshi have revealed that the RGG domains of the EWS and FUS/TLS proteins are critical for G4 binding and do so through interactions with the phospho-ribose backbone of the loop regions[47–49]. The EWS protein's RGG3 domain has high affinity for specific RNA and DNA G4 that is dependent on the arginines in this region[41]. Interestingly this binding seems to be dependent on the presence of loops of at least 2 nucleotides and higher affinity is observed with longer loops[49]. Replacing the loop sequences with abasic deoxy-ribose backbones did not affect this affinity, indicating that the interaction was with the backbone itself and not the nucleotide bases. In contrast, both the backbone and bases of the TT loops in TBAs are involved in thrombin binding and thus replacing the TT loop residues of TBAs with abasic backbone reduces the affinity for thrombin as well as the orientation of the aptamer on the surface of thrombin[50–52]. Furthermore, an intermolecular G4 with single stranded overhangs did not demonstrate the same affinity, indicating the structure of the backbone in the G4 loops was important for recognition by the RGG3 domain. While the initial strand orientation of the G4 did not affect binding, RGG3 was shown by circular dichroism to

convert the topology of DNA G4 from antiparallel to parallel. Thus, it seems that the phospho-ribose backbone structure in the propeller type loop conformation is being specifically recognized by the RGG3 domain of EWS[49].

Mutation of the three phenylalanine residues in the RGG domain of FUS/TLS to tyrosine eliminate the binding to the DNA G4 while retaining affinity for the RNA G4[47]. Interestingly, when the DNA G4 loops were replaced with an abasic phospho-ribose backbone the binding capability was restored. This indicates that the tyrosine residue specifically recognizes the 2' hydroxyl group of ribose sugars in the loop region and not the guanine tetrads. RNA G4 with no loops or locked nucleic acid loops (which obscure the 2' hydroxyl) abrogate the binding, consistent with the notion of 2' hydroxyl recognition being important and ruling out possible differences caused by strand orientation in the DNA G4. Truncation of the protein to exclude two tyrosine residues in the native RGG domain of TLS/FUS abolishes the affinity for RNA, while retaining the affinity for DNA, G4[47]. Truncation of the section containing three phenylalanines has the opposite effect, maintaining affinity for RNA, while losing affinity for DNA, G4. Mutations of these aromatic residues to alanine abolishes all affinity for G4, suggesting that their interactions are not only important for specificity to the ribose sugar but are critical for G4 recognition[48].

These mutational studies with EWS and FUS/TLS demonstrate that RGG domains, like in the case of thrombin binding aptamers, can recognize the loops of G4 structures. Both RGG domains preferentially interact with G4-containing longer loops as opposed to those lacking loops. Furthermore, the RGG domains of EWS and FUS/TLS need only the phospho-ribose backbone to be present in the loops for full affinity, whereas thrombin interacts with both the backbone and bases of TBA loops. This indicates that the binding of G4 by EWS and FUS/TLS is not sequence specific and that the position of the backbone of the loops is likely key to specific recognition by EWS and FUS/TLS.

1.1.7 G4 binding by the RGGGGR peptide of the FMRP protein

Currently the only high-resolution structures of an RGG motif binding to G4 are both from the laboratory of Dinshaw Patel (2LA5, 5DE5) where they have studied the interaction between the 26 amino acid RGG box peptide of FMRP and 36 nucleotides of the SC1 (Transcription factor 19) RNA, which forms a mixed G4-duplex structure, by NMR[53] and X-ray

crystallography'[54] **(Figure 1-3)**. This short peptide has been previously shown to retain specificity and affinity for G4 structures and not other RNA that the full length FMRP protein interacts with, indicating it is responsible for G4 recognition by FMRP[55]. Discrepancies between the structures are attributed to flexible regions within the peptide and loops of the RNA **(Figure 1-3A)**. The structure shows a G4 of parallel strand orientation with 3 guanine tetrads and an unusual mixed base tetrad (GUAU) before the duplex junction. While there are no direct interactions with the guanine tetrads, Arg15, which was shown to be indispensable for binding, is seen to exhibit a cation-pi stacking interaction with A17, which is part of the mixed base tetrad. As well as this stacking interaction, Arg15 is observed in a hydrogen bond with the Hoogsteen face of G7, the first base in the double stranded region **(Figure 1-3B)**. Arg10, another indispensable residue, rests in

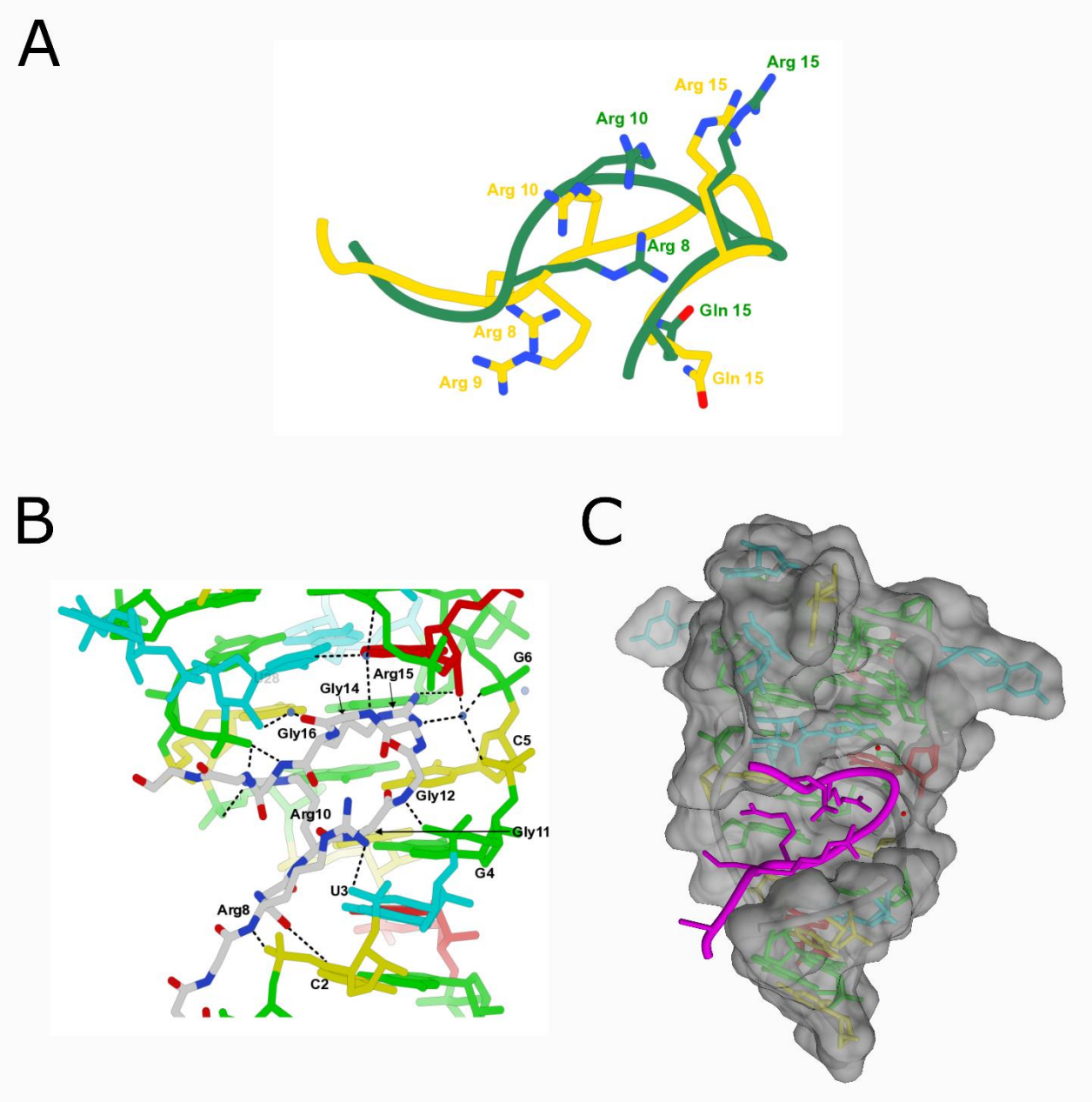


Figure 1-3: Recognition of a quadruplex-duplex junction by FMRP RGG peptide

RGGGGR peptide bound to SC1 RNA quadruplex-duplex junction. . Comparison of peptide conformation between X-ray (green) and NMR (yellow) structures shown in A. Hydrogen bonding pattern between peptide and nucleic acids shown in B. Space filling model shown in C, with peptide shown as a magenta ribbon. PDBs 5DEA (X-ray) and 2LA5 (NMR).

between the first and second bases of the duplex region (C30&G31) and exhibits a cation-pi interaction with C30 and forms bonds with G31 and the phospho-ribose backbone. Mutation of

these first two base pairs has been shown to reduce the affinity of the peptide for SC1 ten-fold[55]. The two arginines that make contact with these bases act as an anchor at the base of the G4 stabilizing the junction from four to two stranded RNA, a feature that was exemplified by RNase digestion experiments.

The four glycines between Arg10 and Arg15 form a tight beta-turn motif and they make many important contacts with the RNA. Gly11 appears to be indispensable (based on mutations made to the peptide), as it coordinates hydrogen bonds from its backbone NH and CO to the carbonyl of G4 and the NH_2 group of C5 in the duplex region. Gly14 makes the second contact with the G4 region via a hydrogen bond from its backbone amide to the ribose sugar of the uridine in the mixed tetrad **(Figure 1-3B)**.

Unlike the thrombin binding aptamers and RGG domains of EWS and FUS/TLS, the SC1 RNA-FMRP interactions do not involve the loop bases at all, the high-resolution structures and mutational analysis show that the important interactions are with the duplex region adjacent to the G4 as well as the mixed base tetrad. From the space filling model **(Figure 1-3C)** it appears that the RGGGGR peptide is occupying a groove created by the tetraplex to duplex junction, thus its' mode of recognition appears to be a mixture of van der Waals space filling and sequence specific interactions mediated by the placement of arginines in this cavity.

1.1.8 G4 recognition of the tetrad face by the DHX36 specific motif

The protein DHX36 is a helicase enzyme that has the capability to bind and unwind RNA G4s with great specificity [56,57]. A region of 13 amino acids that is unique to DHX36 is evolutionarily conserved amongst higher eukaryotes and is both necessary and sufficient for binding of G4s[58].

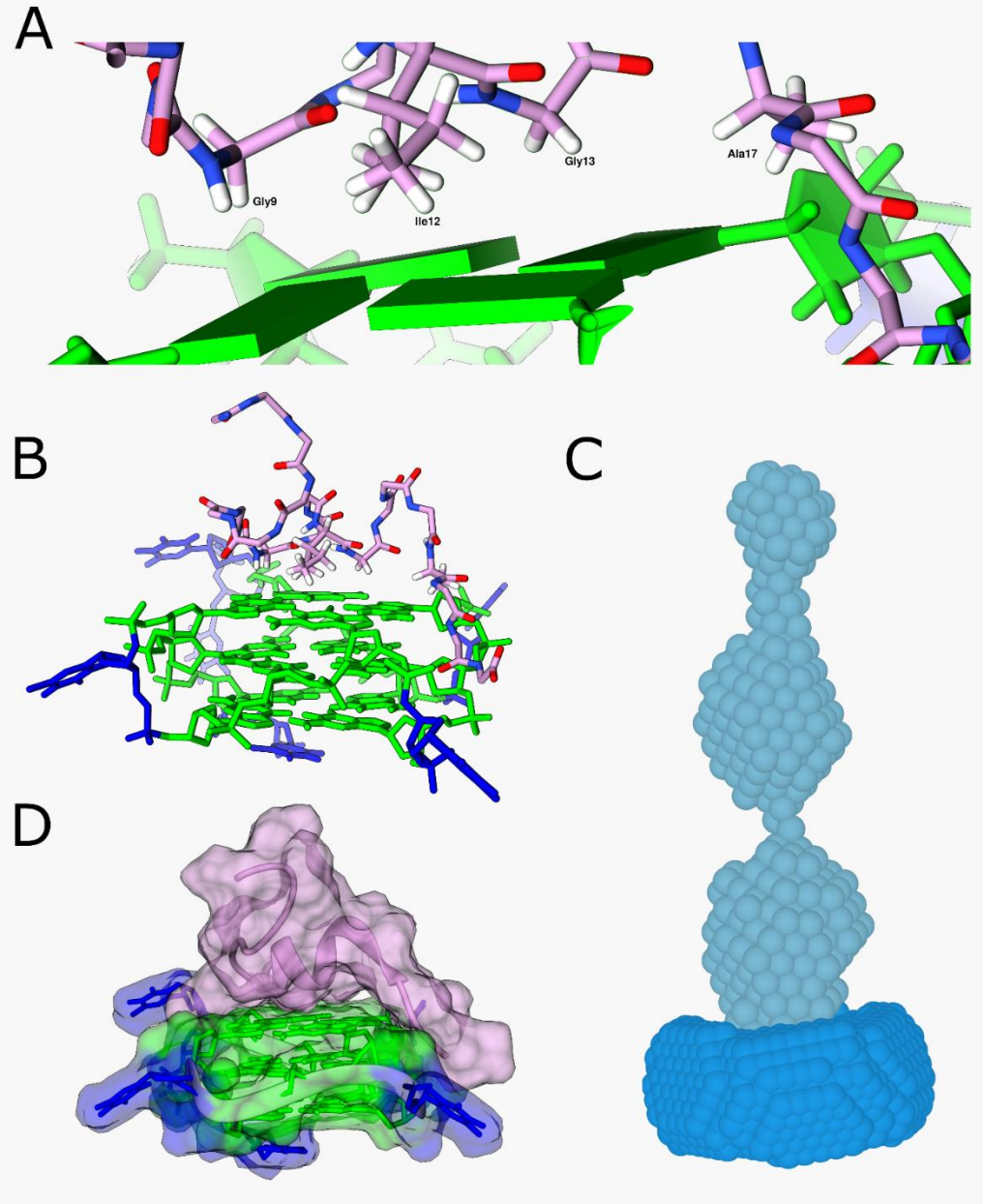


Figure 1-4: Recognition of quadruplex by a DHX36 peptide

NMR structure of 16 amino acid peptide (pink) from DHX36 bound to a DNA guanine

quadruplex (green and blue). Amino acids that stack on the face of the guanine tetrad are labelled in A, charged amino acids that interact with the backbone of the quadruplex are shown in B. Small angle X-ray scattering data of a 52 amino acid fragment of DHX36 (turquoise) bound to an RNA quadruplex (blue) shown in C. Transparent space filling diagram of the 16 amino acid DHX36 peptide bound to DNA quadruplex shown in D. PDB 2N21 (NMR) and SAX structure[59].

Low resolution small angle x-ray scattering has shown that the G4 recognition motif is oriented towards the tetrad face **(Figure 1-4C)**[59,60] and mutations to the loop sequence of an endogenous G4 reduced affinity, indicating that loop conformation may also be important for recognition of G4 by this motif[61]. A high-resolution NMR structure from the lab of Anh Tuân Phan showing a small 18 amino acid peptide bound to a synthetic DNA G4 TT(GGGT)x4 (2N21) supports this mode of binding[62].

The structure shows the propeller type loop arrangement typical of parallel G4s with the DHX36-peptide forming a short alpha-helical region followed by a turn that covers the entire tetrad face. Since DHX36 has significant preference for the parallel versus antiparallel loop orientation, the loop sequences were limited to a single thymine residue which promotes formation of the parallel species. The two aromatic residues (Trp13 and Tyr14) do not make any contacts with the DNA, but are purported to stabilize the L shape of the peptide. Four residues, namely Gly9, Gly13, Ile12 and Ala17 are found at the tetrad-peptide interface close enough to permit CH-pi interactions **(Figure 1-4A)**.

Positively charged residues Lys8, Arg10 and Lys19 are each in close enough proximity to the grooves of the G4 to form electrostatic interactions which could help stabilize the observed interaction **(Figure 1-4B/4D)**; however, previous mutation studies have shown Lys8 and Arg10 to be dispensable in the context of the full length protein[58].

The observed mode of binding explains DHX36's preference for RNA G4s, since RNA G4s form almost exclusively the parallel strand orientation with propeller type loops, they are expected to have exposed tetrad faces. Binding to a large exposed hydrophobic surface like the tetrad face provides a significant increase in entropy through de-solvation and is thus a highly favorable interaction. DNA G4s which have antiparallel or hybrid type loop structures can avoid

this entropy penalty through interactions between the loop bases and tetrad face which bring the polar sugar phosphate back bone over the top of the G4 and shield the hydrophobic tetrad face.

1.1.9 Addendum

Shortly after publication of this review, X-ray crystal structures of nearly full-length version of DHX36 bound to a parallel DNA quadruplex was published[63]. This study not only shows unprecedented insights into G4 unwinding by DHX36 but also general insight into the repetitive-unwinding model of other DEAH-box RNA helicase proteins. The structures confirmed what was previously expected, that the DHX36 specific motif (DSM) interacts directly with the tetrad face of the quadruplex, explaining its preference for parallel G4s. The DSM forms an alphahelix that rests on the 5' tetrad of the G4 exposing a hydrophobic core comprised of Ile, Trp, Tyr and Ala, residues that stacks on the exposed nucleobases. Comparatively, the NMR structure of the DHX36 peptide bound to a DNA G4 did not show the Trp and Tyr residues as stacking on the tetrad face, but rather as stabilizing a turn that directed two Gly residues as well as the Ile and Ala in the crystal structure to be stacking on the tetrad. In this case, it is likely that the context of the full-length protein is providing us with a more biologically relevant picture of the interaction than the structure with the peptide.

Assisting the DSM, the oligonucleotide binding (OB) domain, OB-I forms extensive hydrogen bond interactions with the phosphate backbone of the 5' leader sequence. Flanking this set of interactions the OB-II domain contacts the phosphate backbone of the 3' single stranded DNA, feeding it into the Winged Helix (WH) and RecA domains of the helicase core. One of the key findings from this study was that G4 binding alone can induce ratcheting of the Helicase core domains, supporting the repetitive-unwinding model, where a step-wise winching if the 3' ss nucleic acid strand through the enzyme disrupts the 5' structures[64]. The findings reveal that association with the helicase, independent of ATP hydrolysis, re-organizes the G4 and protein structure. After binding to the protein, the 3' most guanine of the G4 has been pulled out of its quartet, shifting the sequence of the G4 such that the 5' most quartet is comprised of a, less stable, non-canonical, A-T-G-G quartet. Pulling of the 3' end of the DNA results in rotation of the C-terminal domain and the RecA2 domain by 28 and 14 degrees, respectively. This step is shown to be reversible and ATP-independent by single-molecule FRET experiments.

The predominant interactions observed are either directly with the nucleobases or with the phosphate backbone, relatively few deoxy-ribose to protein interactions are observed. Combined with the main recognition site on the G4 being the tetrad face, this explains why DHX36 can unwind both RNA and DNA G4 with a strong preference parallel vs antiparallel or hybrid G4. Since the tetrad face is obscured in antiparallel and hybrid type G4s, the DSM has reduced surface area to bind to resulting in less unwinding activity. However, the lack of specific interactions with the ribose sugars allows DHX36 to be indiscriminate towards parallel G4s that are RNA or DNA.

1.1.10 Summary of protein-G4 interactions

Mutational analyses, as well as high-resolution NMR and X-ray studies, have provided invaluable insights into the structures of biological molecules, this in turn has expounded the molecular mechanism of many important biological processes. With these structure function relationships, we can rationally approach how to alter these processes and modify them to fix pathologies. Although the recognition of G4s by proteins is far from being understood, several trends have emerged amongst the data collected to date. It is apparent that the loop structure of G4s as well as the junction regions adjacent to them present a unique molecular landscape for recognition by proteins. So far this appears to be the most common mode of recognition. Thrombin binding aptamers appear to canonically bind thrombin using the thymine bases of two lateral loops like a pincer to latch onto thrombin. The RGG domains of EWS and FUS/TLS also appear to interact mainly with the loops of G4s, the mode of recognition in this case is likely from the placement of the phospho-ribose backbone that connects the strands of the G4 since the bases themselves have been shown to be dispensable. In the case of FMRP the RGGGGR peptide stabilizes the transition from G4 to duplex by filling the junction between them with base stacking and Hoogsteen type hydrogen bonds with the double stranded region.

Mutations made to the FUS/TLS protein and its G4 substrate have shown that preferential recognition of RNA over DNA G4 can be mediated by 2’OH interactions. It will be interesting to see how other G4-binding are affected by such mutations as those performed with FUS/TLS protein. So far DHX36 is the only G4-binding protein that has been shown to directly interact with the open guanine-tetrad face of the G4 and this satisfactorily explains its preference for RNA G4s, since the faces of antiparallel G4s are typically obscured by the loops.

Currently most G4-binding small molecules interact directly with the guanine tetrad face. This would, obviously, directly compete with DHX36 binding but has also been shown to prevent binding of other RGG domain containing G4-binding proteins, making specific targeting and mechanistic elucidation of their effects challenging[18,42,46]. An alternate approach to designing compounds to bind G4s would be to design compounds that mimic their unique loop structure. Since loop recognition seems to be key to the specificity of many G4-binding proteins, creating small molecules to compete with this interaction may result in more specific compounds that could be used in a more targeted manner to affect cellular pathways.

2. Helicase proteins

Recognition of RNA structures and sequences by proteins is a fundamental process in biology. Not only are RNA a key intermediate in the central dogma of molecular biology, leading to protein expression, but emerging roles of non-coding RNA are rapidly being discovered[65,66]. RNA are capable of forming a diverse array of different structures, in some cases an RNA structural element may act as a switch, or a knob, that allows tuning of a biological process through interaction with and or remodeling of the RNA structure[67]. The fate and function of these RNA molecules is dictated by which proteins they encounter and understanding how proteins specifically recognize RNA and alter RNA structure is central to understanding these biological processes.

A class of proteins that bind and alter the structure of nucleic acids are helicase proteins, the majority of helicase proteins fall into the categorization of Superfamily-1 or Superfamily-2 (SF1 or SF2) based on primary and secondary structural conservation. Despite the high level of similarity there exist drastic differences between the activities of the helicases within these families. For example, The efficient separation of complementary strands of DNA is a fundamental requirement for cellular replication and with ~6 billion base pairs of DNA needing to be unwound and replicated every cell cycle it seems to make sense that DNA helicase proteins have evolved into high efficient processive enzymes that translocate in a specific direction (5’-3’ or 3’-5’) as they unwind dsDNA at a rapid pace[68]. Contrariwise, current research suggests that most RNA helicase proteins act in a localized and non-processive manner, succeeding in remodeling short structures but struggling with long dsRNA[69–71]. This adaptation appears to

reflect the genuine function of the helicase proteins, as RNA predominantly forms short structured regions rather than long stretches of double stranded RNA.

The most conserved region of the SF1 and SF2 family helicase proteins is the helicase core, which is comprised of two RecA domains that contain 12 sequence motifs shared between SF1 and SF2. When bound to RNA, the two RecA domains associate and form two clefts presenting the conserved motifs for nucleotide tri-phosphate binding/hydrolysis (motifs Q, I, II, III, VI) and nucleic acid polymer binding (motifs Ia,b,c, IB IBa, V and Vb),respectively[72].

RNA helicase proteins are further divided into two groups identified by the consensus sequence of their motif II, DEAD and DEAH, also referred to as DDX or DHX proteins. Out of the 58 non-redundant, human, RNA helicases identified from a 2011 study[71], 42 were DEAD-box and 16 were DEAH-box RNA helicase proteins. Investigation of the substrate specificities and kinetics of dsRNA unwinding by DEAD-box proteins[69,70,73–75] reveal a common mechanism of unwinding whereby the two helicase core domains interact with the structured RNA and ATP to bring the core domains closer. This “core closure” places significant strain on the RNA structure, resulting in strand separation and a tightly retained complex between the helicase core and a stabilized single stranded, this allows the released strand to explore further conformational opportunities. ATP hydrolysis then triggers a rearrangement of the core domains and efficient dissociation from the temporarily unstructured RNA and the cycle of unfolding is complete. Thus, the DEAD-box helicase protein affords a structural rearrangement of RNA via an ATP-dependent, repetitive, local unwinding process.

DEAH-box helicases have many differences from DEAD-box helicases, they’re named for the consensus sequence difference in motif II but they also lack the consensus Q from the Q-motif. This glutamine residue is involved in adenine base recognition and so DEAH-box helicase proteins lack specificity for ATP as their energy source and can also hydrolyze other NTPs. DEAH-box helicases also have a more conserved C-terminus consisting of a winged helix domain, a ratchet domain and an OB fold, which associate with the helicase core. Rather than binding directly to a structured section of RNA and destabilizing it, DEAH-box helicases require a single stranded 3’ region to bind to[76]. Interestingly, DEAH-box helicases have the ability to disrupt RNA structures many nucleotides 5’ of where they are bound. For example the Prp16 DEAH-box helicase translocates 3’-5’ down mRNA until it reaches the massive spliceosome

complex, at this point it restructures RNA deep inside the spliceosome[77]. X-Ray crystal structures of DEAH-box helicases in varying stages of the ATP cycle[63,78–80] support a model where the C-terminal domains and the RecA2 domain of the helicase ratchet upon binding single stranded nucleic acid, when faced with a 5' structural obstruction the helicase acts like a winch pulling the single stranded region through a cleft between the RecA1 and RecA2 domains. This ATP-independent binding results in a repetitive unwinding activity, similarly to DEAD-box helicases ATP hydrolysis is apparently only necessary for a protein structural change that allows for rapid dissociation from the nucleic acid and re-initiation of the cycle.

3. DDX21

1.3.1 Origins

DDX21 (also known as RNA Helicase II and Gu protein) is a DEAD-box RNA helicase protein that, like many other DEAD box helicase proteins, wears many hats. Since the majority of DDX21 has been observed to localise to the nucleolus, it's function in rRNA biogenesis was the first to be studied. However, many other activities of DDX21 have been described including transcriptional regulation, RNA splicing and modification, viral RNA sensing and translational regulation. Herein I will briefly review the available literature on DDX21.

DDX21 was first described in 1993 by the Hurwitz lab who purified the endogenous protein for its dsRNA unwinding activity. Approximately 60 billion HeLa cells were collected from 60 litres of liquid cell culture for the purification of a dsRNA helicase activity from fractions collected from 4 successive column purifications (anion exchange, phosphocellulose affinity, cation exchange and Heparin affinity). The purified protein was named RNA helicase II. As well having dsRNA helicase activity, it was observed that the protein could also convert non-duplex RNA into a stable secondary structure, an activity referred to as foldase activity[81].

Three years later, a patient, "Gu", with watermelon stomach disease was found to have high antibody titers for a nucleolar protein. The cause and pathogenesis of watermelon stomach disease, also known as gastric antral vascular ectasia (GAVE) disease, are still not known but ~60% of patients present with autoimmune disorders[82]. The antibodies from patient Gu were isolated and screened against a HeLa cDNA expression library to identify the antigen, which was

named Gu protein[83]. The Gu protein turned out to be an RNA helicase protein that also had a foldase activity similar to RNA helicase II.

1.3.2 Structural domains with distinct activities

Over the next 5 years the structural domains responsible for the RNA helicase II/Gu's localization and enzymatic activities were elucidated. A canonical nuclear localization signal (NLS) is found within the N-terminus of DDX21 and nucleolar localisation was located in the C-terminal domain. Interestingly a GFP fusion protein with the 71 C-terminal amino acids of DDX21 was enough to localize it to the nucleolus[84]. This localization was further shown to be dependent on interactions with the disordered protein motifs of nucleolar scaffold protein WDR46[85], as well as interactions with the c-Jun protein[86]. Purification of DDX21 truncations and testing of their enzymatic activity showed that the foldase activity of DDX21 was also dependent upon these 71 C-terminal amino acids and that they could perform the foldase assay without the rest of the protein[87]. Within this region are three repeats of the sequence FRGQR(X_5) (X_5 representing the spacing of the repeats) followed by a PRGQR sequence, that is unique to the DDX21 protein. Mutation of the second FRGQR or the PRGQR to YEGIQ eliminated the foldase activity[88]. As well as DDX21's dsRNA helicase activity and this C-terminal foldase activity, a recent study has shown that DDX21 can resolve R loops, an RNA:DNA heteroduplex structure whose formation can block transcriptional elongation, in an ATP-dependent manner[89]. The authors also present the first evidence for functional post-translational modification of DDX21, a deacetylase enzyme SIRT7 was necessary for this activity and that acetylation of DDX21 by CBP inhibited the activity.

1.3.3 DDX21 and rRNA biogenesis

Shortly after the advent of siRNA as a biotechnology tool, knock-down and overexpression experiments were carried out on DDX21 and focus was put on studying DDX21's role in ribosome biogenesis, since it was primarily thought to localize to the nucleolus. In both Xenopus and Mammalian cells it was observed that knock-down of DDX21 inhibited proper splicing of the ribosomal RNA (rRNA) and resulted in the accumulation of pre-rRNA fragments and decreased production of mature 28S and 18S components [90,91]. Of note, proper processing of the 28S but not the 18S rRNA could be recovered by recombinant expression of a helicase dead DDX21 (i.e. one that cannot hydrolyze ATP), hinting at the possibility that the foldase and not

the helicase activity of DDX21 was needed for proper 28S processing. However, the recovery experiments with just the foldase domain or a foldase dead mutant were never performed.

An important part of ribosome biogenesis is the proper chemical modification of rRNA, this highly regulated process involves the assembly of protein/RNA complexes called snoRNPs (small nucleolar ribonucleic acid protein complexes) which carry out nucleotide specific 2'O-methylation and pseudouridylation[92]. Individual-nucleotide-resolution crosslinking and immunoprecipitation sequencing experiments (iCLIP-seq) revealed DDX21 binding to rRNA and snoRNAs, with the strongest binding overlapping 2'O-methylation sites[93]. DDX21 also co-immunoprecipitated with protein components of the snoRNPs, namely NOP58, fibrillarin and dyskerin and its activity is required for proper rRNA 2'O methylation[94].

1.3.4 DDX21 and transcriptional regulation

The first evidence for transcriptional regulation by DDX21 came from investigation of c-Jun binding partners. C-Jun is a proto-oncogene that forms part of the AP-1 heterodimeric transcription factor and plays roles in G_1/S cell cycle progression and has anti-apoptotic activity. C-Jun is double phosphorylated through the JNK signaling pathway in response to a myriad of cell stresses, this is necessary for its anti-apoptotic activity[95]. DDX21 was shown to directly interact with c-Jun and translocate from the nucleolus to nucleoplasm upon JNK phosphorylation. Furthermore, it was shown that DDX21 was contributing to transcriptional activation of AP-1 target genes in a manner dependent on its helicase activity. Interestingly, the C-terminal 52 Amino acids were shown to be necessary and sufficient for interaction with c-Jun[86]. A further study investigating the relationship between c-Jun and DDX21 found that c-Jun depletion was sufficient to interfere with DDX21 nucleolar compartmentalization and resulted in depletion of 28S and 18S rRNA production as well as inhibition of G_1/S cell cycle progression[96]. From these studies, c-Jun and DDX21 cooperate to perform multiple cellular functions, anti-apoptotic signaling, cell cycle progression and rRNA biogenesis.

More direct evidence for DDX21's role in transcriptional regulation came by the way of chromatin immunoprecipitation coupled to high-throughput DNA sequencing experiments, whereby it was shown that DDX21 was associated with actively transcribed rDNA loci as well as a number of Pol-II promoters with active chromatin marks. *In vitro* assays showed that DDX21 could facilitate release of P-TEFb through remodeling the 7SK RNA, leading to

phosphorylation of Pol-II and triggering transcriptional elongation[93]. DDX21 has also been shown to regulate Pol-I transcription by forming multimeric ring shaped structures surrounding Pol I, suppressing transcription or rRNA. Interaction with a snoRNA-ended long non-coding RNA enhances ribosomal transcription (SLERT) by loosening the DDX21 rings and dissociating it from Pol I and allowing transcription initiation[97]. Supportive of DDX21s role in transcription of ribosomal genes and rRNA processing, re-localization of DDX21 from the nucleolus, results in inhibition of rRNA processing as well as downregulation of ribosomal gene transcription and contributes to the pathogenesis of multiple ribosomopathies (diseases caused by mutations to the ribosome)[98].

1.3.5 DDX21 and cancer

Upregulation of ribosome biogenesis as well as anti-apoptotic signalling pathways are hallmarks of cancer, considering DDX21's role in these processes it is no real surprise that DDX21 has been shown to be aberrantly expressed in a number of cancers. A clinical research study that analyzed breast cancer tissues from 187 patients found that high expression levels of DDX21 in breast cancer tumour cells is correlated with early death, a shorter disease-free survival rate and rapid relapse, it was recommended as a prognostic biomarker for breast cancer[99]. DDX21 depletion in 10 established breast cancer cell lines affected AP-1 activity and induced cell death and cell cycle arrest. Furthermore, DDX21 knock down by lentivirally transduced siRNA essentially eliminated tumorigenicity in a xenograft mouse model and inhibited Ras mediated transformation *in vivo*[99]. Contrary to these findings a separate group analysing 140 patient breast cancer tissues found that higher DDX21 expression correlated positively with survival rates of patients[100]. This study proposed that DDX21 supresses breast cancer metastasis *in vivo* by epigenetically repressing the transcription factor Snail.

Aberrant DDX21 expression has also been observed in numerous other cancer tissue types. A comparison of 90 patients with gastric cancers and normal human gastric cells showed that DDX21 was significantly upregulated in the gastric cancer tissues and positively corelated with tumour size in both patient and mouse zenograft models[101]. DDX21 has also been observed to be overexpressed in colorectal cancers[102] and crucial for proliferation of anaplastic large cell lymphomas[103] and uveal melanoma[104].

1.3.6 DDX21 and the innate immune response

As well as playing key roles in ribosome biogenesis, cell cycle and transcriptional regulation DDX21 also plays a role in the life cycle of multiple viruses. Using a mass-spectrometry guided approach a complex between three RNA helicases (DDX1, DDX21 and DHX36) and an interferon inducing adapter molecule, TRIF, was identified. Interestingly, disruption of this complex by siRNA knockdown of any of the involved proteins significantly reduced the downstream signaling by TRIF, resulting in up to a 60% reduction in interferon production in response to dsRNA simulating a viral infection[105]. A similar DDX21 dependent interferon response was observed when A549 cells were infected with Dengue virus. In this study DDX21 was observed to re-localize from the nucleus to the cytoplasm upon Dengue viral infection, resulting in interferon response and decreased viral replication[106]. Further investigation using an Influenza A virus (IAV) model elucidated a possible mechanism whereby the DDX21-TRIF complex served as an early pattern recognition receptor for foreign RNA and activated expression of a damage associated molecular pattern, a host protein that activates pro-inflammatory signaling through the TLR4/MyD88 pathway[107]. Once again DDX21's role in transcription regulation surfaced when it was observed that DDX21 inhibited early viral transcription of IAV through binding to the viral polymerase protein PB1[108]. Counter to these anti-viral actions of DDX21 one study suggests that DDX21 is assistive in assembly of Rev/RRE complexes, a protein-mRNA interaction that controls HIV-1 mRNA export[109].

4. Thesis outline

The work presented in this thesis expands the current understanding of RNA G4 binding proteins and specifically focuses on the new RNA G4 binding protein DDX21. In Chapter 2, I will describe in detail the discovery and validation of DDX21 as an RNA G4 binding and unwinding protein. Following this, Chapter 3 identifies biological G4 targets of DDX21 and their role in translation regulation. Finally, in Chapter 4 a detailed biophysical examination of the interaction between DDX21's G4 binding domain and G4 structures is made.

Chapter 2: Human DDX21 Binds and Unwinds RNA Guanine Quadruplexes

2.1.1 Preface

The product of my first two years of Ph.D. work, this 2017 Nucleic Acids Research publication:

McRae, E. K. S., Booy, E. P., Moya-Torres, A., Ezzati, P., Stetefeld, J. & McKenna, S. A. (2017) Human DDX21 binds and unwinds RNA guanine quadruplexes. Nucleic Acids Res., 45**,** 6656–6668.

describes how we identified DDX21 as a G4 binding and remodeling protein. We show that the main G4-binding domain is in the C-terminal 209 amino acids of DDX21 and depends on the FRGQR repeat region that is unique to DDX21. We characterize a G4 binding and destabilizing activity of the protein on 3 substrates and demonstrate that mutation to the G4 binding region of DDX21 can affect the expression of genes with G4s in their 3'UTR.

2.1.2 Contribution of authors

Evan P. Booy performed the initial streptavidin pull-down assay that started the investigation. The Mass-spectrometry was performed and analyzed by Peyman Ezzati. Aniel Moya-Torres provided recombinant purified DHX36 protein for nuclease sensitivity assays. The remainder of the experimental work and writing was completed by Ewan K.S. McRae, with editing and guidance from all the listed authors.

2. Abstract

Guanine quadruplexes (G4s) are an important structure of nucleic acids (DNA and RNA) with roles in several cellular processes. RNA G4s require specialized unwinding enzymes, of which only two have been previously identified. We describe the results of a simple and specific mass spectrometry guided method used to screen HEK293T cell lysate for G4 binding proteins. From these results, we validated the RNA helicase protein DDX21. DDX21 is an established RNA helicase but has not yet been validated as a G4 binding protein. Through biochemical techniques, we confirm that DDX21-quadruplex RNA interactions are direct and mediated via a site of interaction at the C-terminus of the protein. Furthermore, through monitoring changes in nuclease sensitivity we show that DDX21 can unwind RNA G4. Finally, as proof of principle, we demonstrate the ability of DDX21 to suppress the expression of a protein with G4s in the 3' UTR of its mRNA.

3. Introduction

Guanine quadruplexes (G4s) are a non-canonical four-stranded structure of nucleic acids that have emerged as key regulatory elements in a variety of cellular processes[8,12,110–114]. G4s are stabilized by pi-pi stacking of planar tetrads consisting of four guanine bases, which are themselves stabilized by hydrogen bonds between Hoogsteen and Watson Crick faces of adjacent guanines[4]. Between the planar tetrads, a cation further stabilizes the G4 by charge screening the electronegativity of the central O6 of the guanines; potassium is the best suited physiologically relevant metal cation for this and other less suited cations can have drastic effects on G4 stability [115,116]. G4s are very stable structures *in vitro* under conditions mimicking cellular conditions, with RNA G4s being more stable than their DNA counterparts[117,118]. This high degree of stability necessitates specialized enzymes for unwinding and sampling of alternative conformations, herein referred to as G4 helicase activity.

In the case of RNA G4s, approximately 12,000 have been identified by high throughput reverse transcription sequencing methods (rG4-seq)[25,113]. Many of these have been experimentally validated and shown to regulate a diverse set of cellular functions such as translational suppression, alternate polyadenylation, transcription termination, 3' end processing, mRNA localization and alternative splicing[111,119,120]. There are currently only two known RNA G4 unwinding proteins (DHX36 and DHX9) but over a dozen such proteins that can act on DNA

G4s, indicating a deficit in our knowledge of RNA G4 unwinding proteins [56,121–123]. Furthermore, recent evidence suggests G4s exist in a predominantly unwound state *in vivo* and that knock-down of DHX36 and ATP depletion have little effect on the globally unwound state of RNA G4[25], this implies a redundancy in the RNA G4 helicase activity. It is therefore crucial to our understanding of RNA G4 mediated regulation of cellular processes to identify other RNA G4 binding and G4 helicase active proteins. To pursue this, we have performed streptavidin pull-down screens with biotinylated RNA G4 from the 3' UTR of PITX1. We analyzed the proteins that bound to the G4 by mass spectrometry and identified many candidate RNA G4 helicase proteins. One of the more abundant hits, DDX21, is an RNA helicase protein that has not yet been validated as a G4 binding protein.

DDX21 was first isolated from nuclear extracts of HeLa cells based on its RNA duplex helicase activity, and at the same time it was observed that the protein possessed an activity termed foldase activity that was able introduce secondary structure to a seemingly single-stranded RNA[81]. Subsequent characterization of DDX21 showed that the two different activities, ATP-dependent helicase and an ATP-independent foldase activity, are performed by two separate domains of the protein, the former by the core helicase domains and the latter by the C-terminus of the protein[84,87,88]. DDX21 has been implicated in ribosomal RNA biogenesis[90,93,94], viral RNA sensing as part of the innate immune system[105,106,108] and cancer progression, where it has been correlated with disease free survival in breast cancer[124] through regulation of c-jun activity and rRNA processing[96,99]. Herein we present the identification and validation of DDX21 as a G4 binding protein as well as evidence that the previously characterized foldase activity is likely an RNA G4 unwinding activity. Furthermore, we show that the region of DDX21 responsible for G4 binding and unwinding can affect the expression of a G4-based reporter assay in HEK293T cells. These results provide the basis for future investigations into novel regulatory pathways mediated by RNA G4 and DDX21.

4. Results

2.4.1 Identification of DDX21, an RNA helicase, as a potential G4 binding protein.

In order to identify new RNA G4 helicase proteins, streptavidin pull-down assays (SPDA) were performed from HEK293T cell lysates with a 5'-biotinylated G4 from the 3' UTR of PITX1 (Q2) that has been extensively characterized[14,60]. We used Nano RP-LC-MS/MS to identify proteins recovered from the SPDA with the Q2 RNA G4 as well as a negative control non-quadruplex RNA Q1mutant. The Q2 SPDA was performed with 4 biological replicates and the 35 proteins with the greatest difference between the number of unique peptides in the G4 and negative control sample are reported with standard error in a bar graph (**Figure 2-1A)**. This was done to filter the list of biotin-binding proteins that were abundant in both G4 and negative control samples, the full list of proteins identified by mass spectrometry can be found in the supplementary information. Of the proteins recovered, RNA G4 binding proteins were 70-fold over-represented **(Figure 2-1B)**, indicating our screen had selectively enriched known RNA G4 binding proteins. The screen also provided many hits that have not previously been reported to interact with G4 and thus will require further validation. DDX21 was among the most abundant RNA helicase proteins recovered from the screen and was chosen for further validation and characterization.

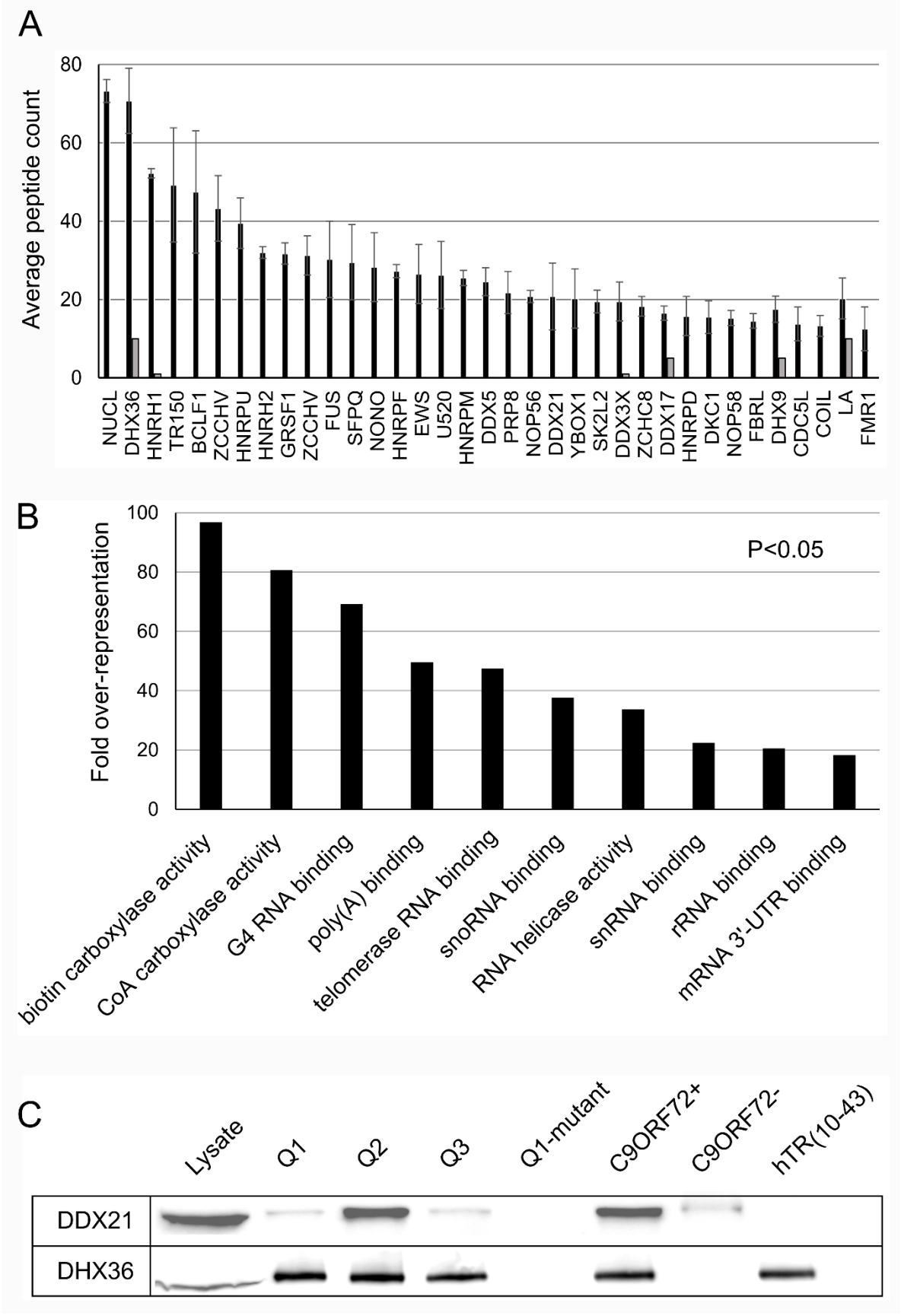

A
Average peptide count
80
60
40
20
0
NUCL
DHX36
HNRH1
TR150
BCLF1
ZCCHV
HNRPU
HNRH2
GRSF1
ZCCHV
FUS
SFPQ
NONO
HNRPF
EWS
U520
HNRPM
DDX5
PRP8
NOP56
DDX21
YBOX1
SK2L2
DDX3X
ZCHC8
DDX17
HNRPD
DKC1
NOP58
FBRL
DHX9
CDC5L
COIL
LA
FMR1
B
Fold over-representation
100
80
60
40
20
0
P<0.05
biotin carboxylase activity
CoA carboxylase activity
G4 RNA binding
poly(A) binding
telomerase RNA binding
snoRNA binding
RNA helicase activity
snRNA binding
rRNA binding
mRNA 3'-UTR binding
C
Lysate
Q1
Q2
Q3
Q1-mutant
C9ORF72+
C9ORF72-
hTR(10-43)
DDX21
DHX36

Figure 2-1: MS Screen for identification of quadruplex binding proteins

DDX21 is an RNA helicase recovered in abundance from pull-down assays with an RNA G4. (A) Proteins identified from streptavidin pull-down assays (SPDA) with RNA G4 by mass spectrometry and sorted by the difference in unique peptide count between G4 and control samples (top 35 hits shown), error bars represent the standard error between four biological replicates. The black bars represent the number of unique peptides from the G4 SPDA and the grey bars represent the number of unique peptides recovered from the negative control samples. (B) Analysis of gene ontology shows a statistically relevant ($P<0.05$) overrepresentation of RNA G4 binding proteins when compared to the expected composition of the cell lysate. (C) SPDA were performed with each of the PITX1-derived G4s, $hTR_{(10\text{-}43)}$, C9ORF72+/- as well as the non quadruplex Q1-mutant. Recovery of DDX21 and DHX36 was detected by western blot. DHX36 was enriched by all the G4 quadruplexes, while DDX21 was only enriched by Q2 and C9ORF72+.

To confirm the mass spectrometry results, pull-down experiments were repeated in HEK293T lysates with a panel of biotinylated RNA quadruplexes and a negative control RNA. The recovered protein was then probed for DDX21 and DHX36 (an established RNA G4 helicase and second highest hit in **Figure 2-1A**) by western blot. We used three G4s from the 3'UTR of the PITX1 messenger RNA (Q1, Q2 and Q3) and a G4 from the 5' region of the human telomerase RNA ($hTR_{(10-43)}$), which have been previously well characterized in our lab[14,57,60,125]. Additionally, we used two other quadruplexes, an i-motif (C9ORF72-) and a G4 (C9ORF72+) from the C9ORF72 expanded repeat that have been extensively characterized by others[126,127]. We used a mutant version of the Q1 quadruplex (Q1mutant), that cannot form a G4, as a negative control.

DHX36 was enriched by pulldown in all samples except the C9ORF72- and the Q1mutant samples, which showed no detectable signal. This is consistent with DHX36's established role as a parallel G4 helicase **(Figure 2-1C)**. Interestingly, DDX21 was enriched only in the Q2 and the C9ORF72+ RNA G4 pull-down samples and only weakly detected in the Q1, Q3 and C9ORF72- samples. No DDX21 was detected from the Q2mutant sample after 5 minutes of exposure but weak signal was observed from $hTR_{(10-43)}$ (data not shown).

2.4.2 The C-terminus of DDX21 contains a unique repeat region that binds G4s.

To determine the region of DDX21 responsible for G4 recognition, pull-downs were repeated with the Q2 G4 after transient recombinant expression of N-terminally FLAG-tagged DDX21 truncations. By truncating at the beginning and end of the conserved helicase core of DDX21, we created constructs lacking the N-terminus ($DDX21_{(217-783)}$), expressing just the N-terminus ($DDX21_{(1-216)}$), just the core helicase domains ($DDX21_{(217-573)}$) and just the C-terminus ($DDX21_{(574-783)}$) **(Figure 2-2A)**. Probing for the FLAG-tag by western blot confirmed that all

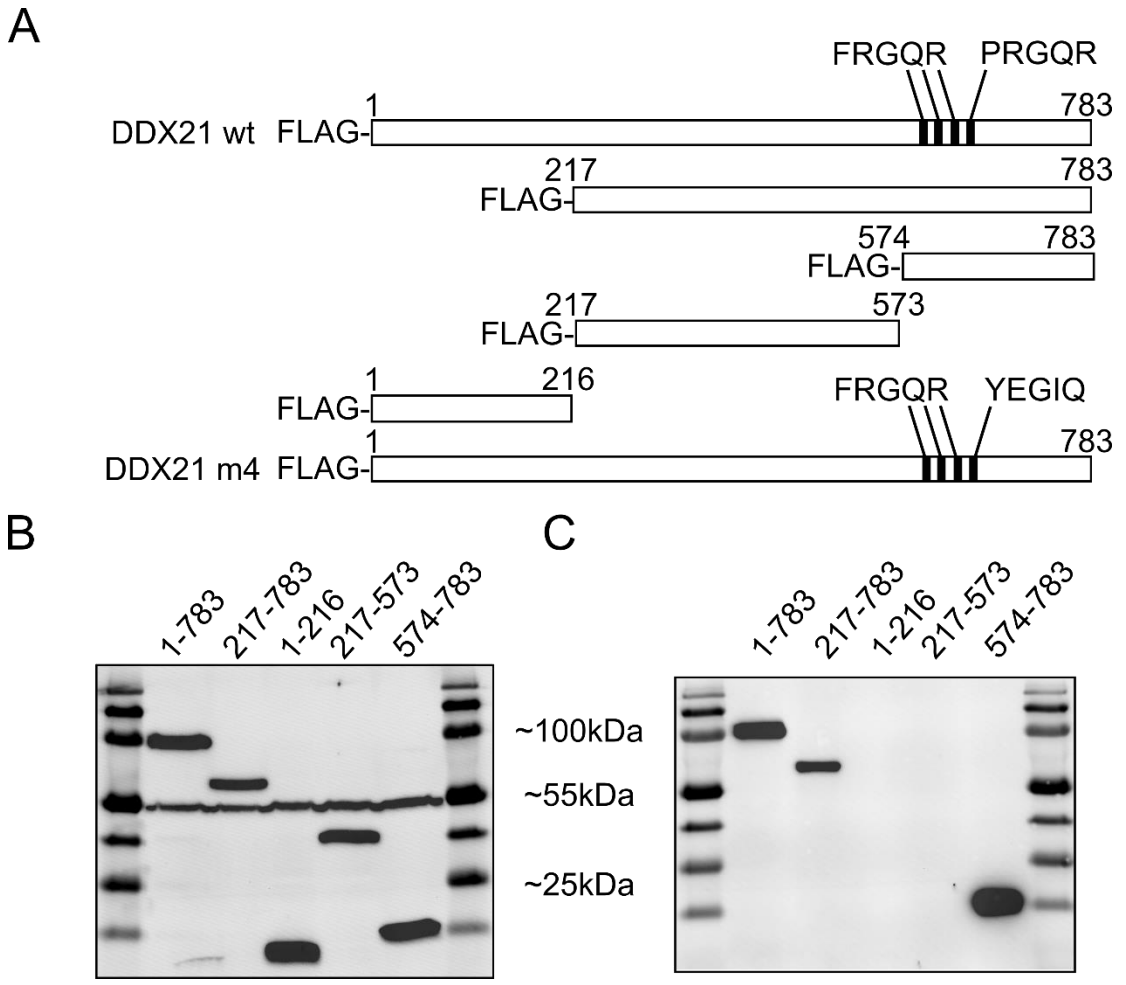


Figure 2-2: Identification of the quadruplex binding domain of DDX21

The C terminus of DDX21 ($DDX21_{(574-783)}$) is both necessary and sufficient for the interaction with the Q2 RNA. (A) Cartoon representation of N-terminally FLAG-tagged truncations of DDX21. (B) Overexpression levels of FLAG-tagged DDX21 truncations in HEK293T cell lysate compared by western blot with anti-FLAG and anti-alpha-tubulin antibodies (~55kDa). (C) Detection of co-precipitated DDX21 truncations from streptavidin pull-down assays with biotinylated Q2 RNA in HEK293T cell lysates. Constructs containing the C-terminal 209 amino acids were all enriched by pull-down with biotinylated Q2, whereas constructs missing the C-terminal 209 amino acids were undetectable.

protein truncations were expressed at roughly equivalent levels in HEK293T cells **(Figure 2-2B)** and revealed that the pull-down had recovered the full-length protein, $DDX21_{(217-783)}$ and $DDX21_{(574-783)}$, but not $DDX21_{(1-216)}$ or $DDX21_{(217-573)}$ **(Figure 2-2-2C).** These results strongly

indicate the C-terminus of DDX21 ($DDX21_{(574-783)}$) is necessary and sufficient for G4 mediated co-precipitation in HEK293T cell lysates.

In order to show a direct interaction between DDX21 and G4 RNA, we purified the full length DDX21 and $DDX21_{(574-783)}$ by FLAG affinity purification **(Figure 2-7A)** and assessed the binding affinity for fluorescently labelled Q2, Q2-mutant and $hTR_{(10-43)}$ RNA by microscale thermophoresis (MST) **(Figure 2-3))**. A high affinity (Kd 10 ± 0.7 nM) interaction was observed between Q2 and the full length protein and an approximately 30 fold weaker (324±7 nM) binding constant was determined for the interaction between $hTR_{(10-43)}$ and full length DDX21. No binding was observed between the Q2-mutant RNA and the full length DDX21 protein.

The specificity for Q2 over $hTR_{(10-43)}$ is diminished in the C-terminal construct of DDX21 indicating that while $DDX21_{(574-783)}$ is both necessary and sufficient for high affinity G4 interaction, another region of the protein is enhancing the affinity for the Q2 RNA. $DDX21_{(574-783)}$ was found to have less affinity (87±7 nM) than the full length protein for Q2 RNA, not dissimilar to what has previously been observed with DHX36, where the full protein has ~10-fold higher affinity for G4 than the G4 binding domain alone[60]. Interestingly, $DDX21_{(574-783)}$ showed higher affinity (149±10 nM compared to 324±7 nM) for the $hTR_{(10-43)}$ G4 than the full length protein. Higher solubility of $DDX21_{(574-783)}$ allowed us to perform binding assays with higher concentrations of protein than the full length DDX21, the upper limit of which revealed a slight affinity for the Q2-mutant that was estimated at ~1000 fold less than the Q2 wild type.

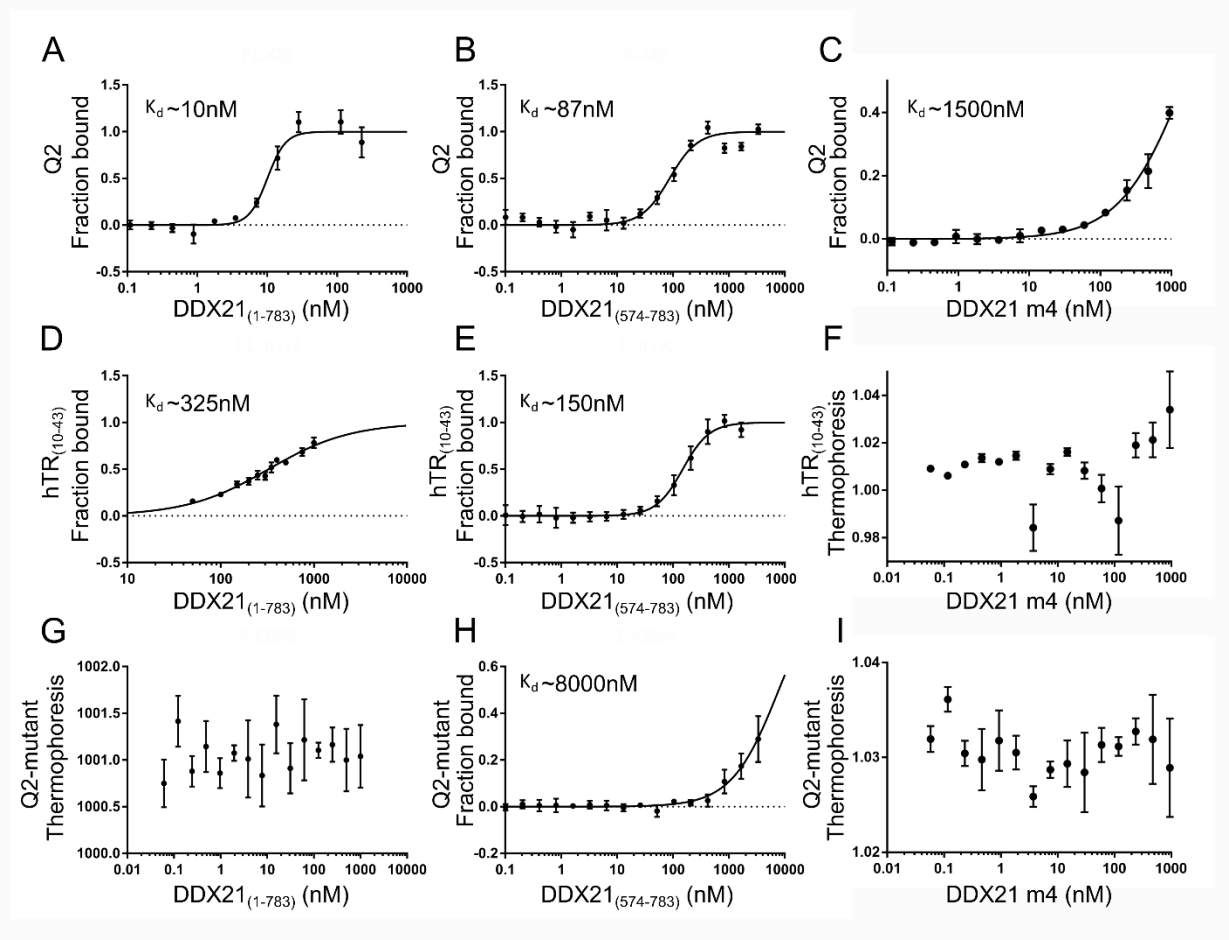


Figure 2-3: Affinity analysis of DDX21 constructs

Dose-response data from microscale thermophoresis experiments were fit to a quadratic binding formula to obtain apparent dissociation constants and Hill coefficients for interactions between Q2 (A,B,C), $hTR_{(10-43)}$ (D,E,F), Q2-mutant (G,H,I) and full length DDX21 (A,D,G), $DDX21_{(574-783)}$ (B,E,H) and DDX21 m4 (C,F,I). Nano-molar affinities are determined between both quadruplex and the wild type DDX21 and DDX21(574-783) constructs, while no, or weak, binding is observed with DDX21 m4. No significant binding was observed between the non-quadruplex Q2-mutant and any DDX21 construct. Error bars represent the standard error between three replicate experiments.

DDX21 has a unique motif in its C-terminal region of three FRGQR repeats and a PRGQR sequence each separated by five amino acids **(Figure 2-2A)**. Previous work on DDX21 has shown that mutation of the PRGQR sequence to YEGIQ essentially abolishes its "foldase" activity[88]. To test whether this sequence might be important for G4 recognition we created a

full-length protein construct with the YEGIQ mutation in the place of the PRGQR sequence (herein referred to as DDX21 m4). Strikingly, purified DDX21 m4 had drastically reduced affinity for the Q2 G4 and no binding was observed with the $hTR_{(10-43)}$ G4 or the Q2 mutant RNA **(Figure 2-3)**. To confirm that the mutation had no drastic effect on the global structure of the folded protein the melting points for the wild type DDX21 and DDX21 m4 were compared by a protein thermal shift assay **(Figure** 2-7**B)**. The melting points were found to be within a degree of each other, indicating that the mutation to the primary sequence had not significantly disrupted the secondary structure of DDX21.

2.4.3 DDX21 and its C-terminal region can unwind RNA G4

Foldase assays that were performed previously on DDX21 by Valdez *et al*[88] used an 81 nucleotide RNA, referred to as S1, that has two predominant conformations which are resolvable on a native gel. The C terminus of DDX21 can convert the faster migrating band into the slower migrating band, which Valdez *et al* interpreted as introduction of secondary structure by DDX21 slowing the migration of the RNA through the gel. This event was referred to as the foldase activity of DDX21. The G-rich sequence composition of the S1 RNA, as well as our discovery that the DDX21 m4 could also no longer bind G4, led us to inquire as to whether the foldase activity could be a misinterpreted G4 helicase activity. After transcription and purification of the S1 RNA we had a single predominant conformation herein called S1 top band. heating at 95° Celsius and cooling on ice could convert the top band to a faster migrating bottom band. Spectroscopic investigation by multiple approaches of both the top and bottom bands gave results consistent with a G4 structure. Addition of either the top or bottom band conformation of S1 to Thioflavin T, a fluorescent dye known to selectively bind G4s[128], caused the dye to exhibit significantly enhanced fluorescence compared to a single stranded RNA control **(Figure 2-4A&B)**. Furthermore, both the circular dichroism **(Figure 2-4C)** and thermal difference spectrum **(Figure 2-4D)** resemble the spectra for previously studied RNA G4[60].

DDX21 and DHX36 both demonstrate the ability to convert the top band to the bottom band in the presence of ATP **(Figure 2-4E)**. Contrarily, the C terminus of DDX21 can hasten the conversion of the bottom band to the top band independent of ATP, but has little to no effect when incubated with the top band alone. To a much lesser extent conversion of the bottom band to the top band can be seen with the full length DDX21 with no ATP present. To determine

whether the difference in migration of the top and bottom bands was due to the use of different guanine tracts in a quadruplex or another non-G4 secondary structure we probed both conformations of S1 with the RNase enzymes T1 and A **(Figure 2-4F)**. Both of these enzymes cut specifically at single stranded regions, with T1 having a preference for cutting after guanines and A having the preference of cutting after cytosines and uracils[129]. If the RNA helicase proteins were unwinding a double stranded structure, we would expect a difference in both the RNase T1 and RNase A cutting patterns between the top and bottom band S1 RNA. However, after treatment of both the top and bottom bands with both RNases we only observed significant differences in the T1 cutting pattern, indicating a change in the accessibility of guanine residues

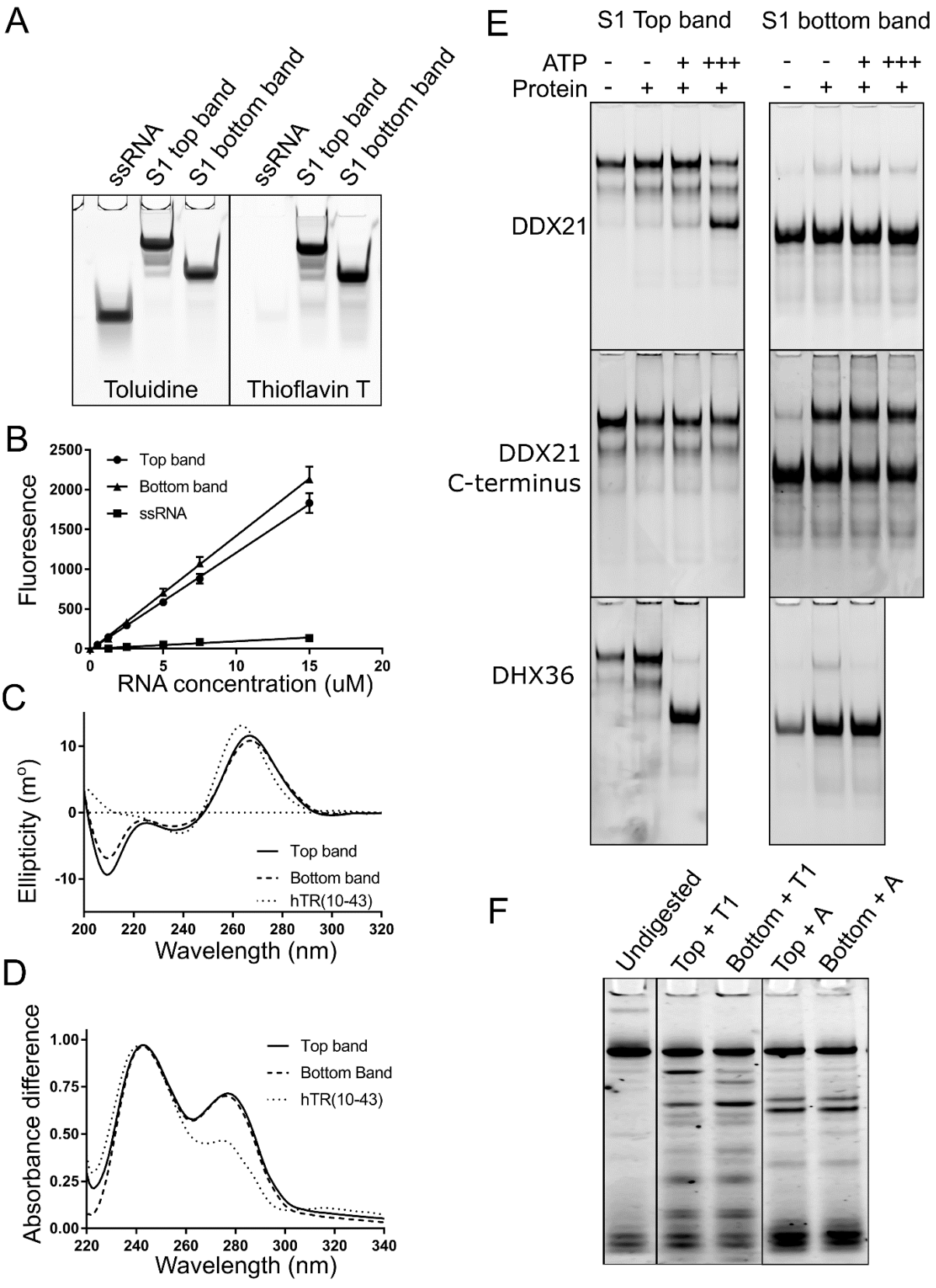

A
ssRNA
S1 top band
S1 bottom band
ssRNA
S1 top band
S1 bottom band
Toluidine
Thioflavin T
B
Top band
Bottom band
ssRNA
Fluoresence
RNA concentration (uM)
C
Ellipticity (m°)
Top band
Bottom band
hTR(10-43)
Wavelength (nm)
D
Absorbance difference
Top band
Bottom Band
hTR(10-43)
Wavelength (nm)
E
S1 Top band
S1 bottom band
ATP
Protein
DDX21
DDX21
C-terminus
DHX36
F
Undigested
Top + T1
Bottom + T1
Top + A
Bottom + A

Figure 2-4: Redefining DDX21's "Foldase" activity

Both the top band and bottom band conformation of the S1 RNA are G4 and DDX21 can unwind these G4 in an ATP dependent manner. Analysis of the two conformations of the S1 RNA (as well as negative control ssRNA and positive control $hTR_{(10-43)}$) by in (A) gel staining with Thioflavin T and Toluidine blue, fluorescence intensity of (B) Thioflavin T, (C) circular dichroism and (D) thermal difference spectra. (E) Unwinding of the S1 RNA by DDX21, DDX21 C-terminus and DHX36. S1 top band (left) or S1 bottom band (right) was incubated with each purified protein without and with 0.5mM (+) or 5mM (+++) ATP for 20 minutes at 30° Celsius before resolution on a native TBE gel and detection by SYBR gold. (F) RNase A and RNase T digestion patterns of 3'-Cy5 labelled S1 top and bottom band species indicate the use of different guanines but not cytosines or uracils in their secondary structure.but not cytosine or uracil. These results indicate that DDX21 can unwind RNA G4, toggling the S1 RNA between two conformations in an ATP dependent manner.

To test if DDX21 can unwind other RNA G4, we performed nuclease sensitivity assays using RNase T1. The G4s tested are resistant to RNase T1 cleavage in their folded state but unwinding of the G4 exposes guanine residues, allowing them to be cleaved and the products resolved by denaturing PAGE. This method is preferred because it allows for very sensitive detection of G4 helicase activity in multiple different G4s without the need for separate single stranded trap RNA. The known G4 helicase protein DHX36 and its quadruplex binding domain (RSM) were used as a positive control and for comparison of activities in these assays. **Figure 2-5** shows gels representative of the nuclease sensitivity of two different RNA (Q2 and $hTR_{(10-43)}$) when incubated with five different protein constructs ($DDX21_{(1-783)}$, $DDX21_{(574-783)}$, DDX21 m4, DHX36 and the RSM. The first two lanes of each gel show the RNA incubated without and with RNase T1 respectively, in the absence of helicase protein; these serve as a baseline for the nuclease sensitivity of the RNA G4. Helicase protein was added in the following 6 lanes as well as ATP or its analogues to study the effect of the presence of ATP or a non hydroyzable analogue on G4 helicase activity.

The enhanced susceptibility to cleavage by RNase T1 (**Figure 2-5**) show that the quadruplex binding motifs (DDX21 C-terminus and DHX36 RSM) on their own are sufficient to distort the G4s significantly enough to enhance nuclease sensitivity. Interestingly, full length DDX21 protein (and DHX36) enhances nuclease sensitivity of $hTR_{(10-43)}$ significantly in the presence of ATP but the RNA is less sensitive to digestion in the presence of the full length protein without ATP or with AMP-PNP. The opposite trend is observed for the Q2 G4, where the full length DDX21 protein (and DHX36) significantly enhance the nuclease sensitivity in the absence of ATP, or with AMP-PNP, but the RNA is less sensitive to digestion when ATP is present with these proteins.

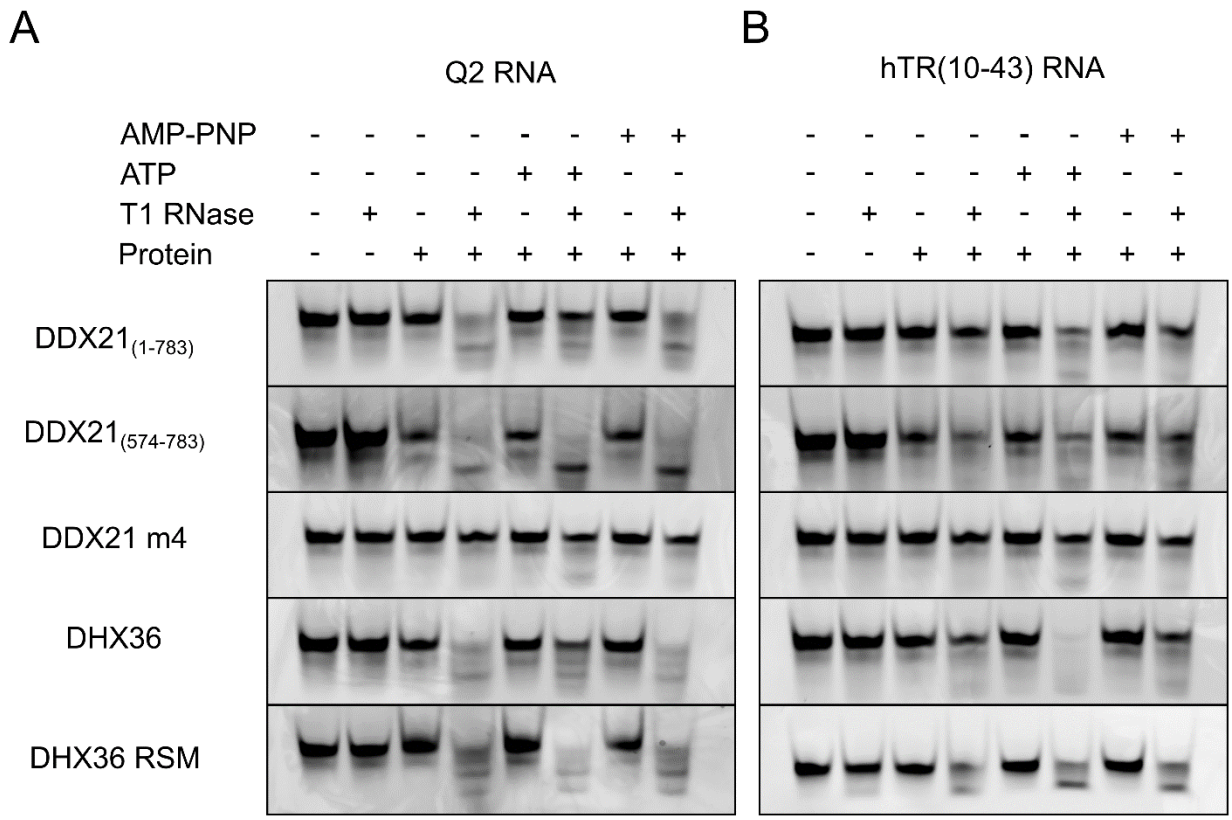


Figure 2-5: Probing quadruplex remodeling through nuclease sensitivity

$DDX21_{(1-783)}$ and $DDX21_{(574-783)}$ significantly enhance the nuclease sensitivity of (A) Q2 and (B) $hTR_{(10-43)}$ RNA G4 while DDX21 m4 does not. Each RNA was incubated with each protein construct with and without RNase T1 and under different ATP conditions, as indicated at the top of the figure, before resolution on by TBE-UREA PAGE and detection by SYBR gold. Disappearance of the main band and appearance of faster migrating bands is indicative of increased accessibility of guanine residues to RNase T1 due to G4 destabilizing activity.

2.4.4 DDX21 affects protein expression in G4 dependent manner

DDX21 has been implicated in regulating protein expression, as have many G4 binding proteins. To test whether gene regulation by DDX21 could be G4 mediated we performed β-galactosidase assays in HEK293T cells using the wild type 3' UTR of PITX1 as well as a version with the three G4s mutated. We used this 3'UTR system because it contains the Q2 G4 that has been shown to interact with DDX21 and because it has been previously used to assess G4 mediated regulation of protein expression via another G4 helicase, DHX36[14],[130]. As a baseline, we measured the activity of β-galactosidase from the two reporter plasmids in cells treated with control siRNA. Simultaneously we measured the expression in cells treated with DDX21 siRNA as well as DDX21 siRNA treated cells that were also treated with pcDNA3 plasmid containing siRNA resistant wild type DDX21, DDX21 m4, $DDX21_{(574\text{-}783)}$ or version of DDX21 with a mutation in motif III that abrogates ATP hydrolysis (DDX21 SAT-mutant) **(Figure 2-6A)**. To compare the protein expression levels of endogenous and recombinant DDX21 constructs western blotting with anti-DDX21 antibody was performed **(Figure 2-6B)**.

The β-galactosidase expression from the wild type UTR containing plasmid was approximately half of that from the plasmid with the mutated G4 region in the samples treated with the control siRNA, consistent with our previous results using this vector[14] as well as other work showing a repressive effect of G4 presence in 3' UTR[131]. No significant difference was observed in the expression levels of β-galactosidase from the plasmid with mutated G4s in the 3'UTR between DDX21 knock-down samples or those expressing recombinant versions of DDX21. When G4 are present in the 3' UTR, expression increases ~2 fold in the DDX21 knock down cells. Re-introduction of the wild type DDX21, $DDX21_{(574\text{-}783)}$ and the DDX21 SAT-mutant but not DDX21 m4 allows expression levels to recover to the level observed with no G4 in the 3'UTR. These results strongly suggest that DDX21 can suppress the expression of genes with G4s in their 3'UTR and that this suppression is dependent upon its ability to bind to and or unwind G4s and does not require ATP hydrolysis.

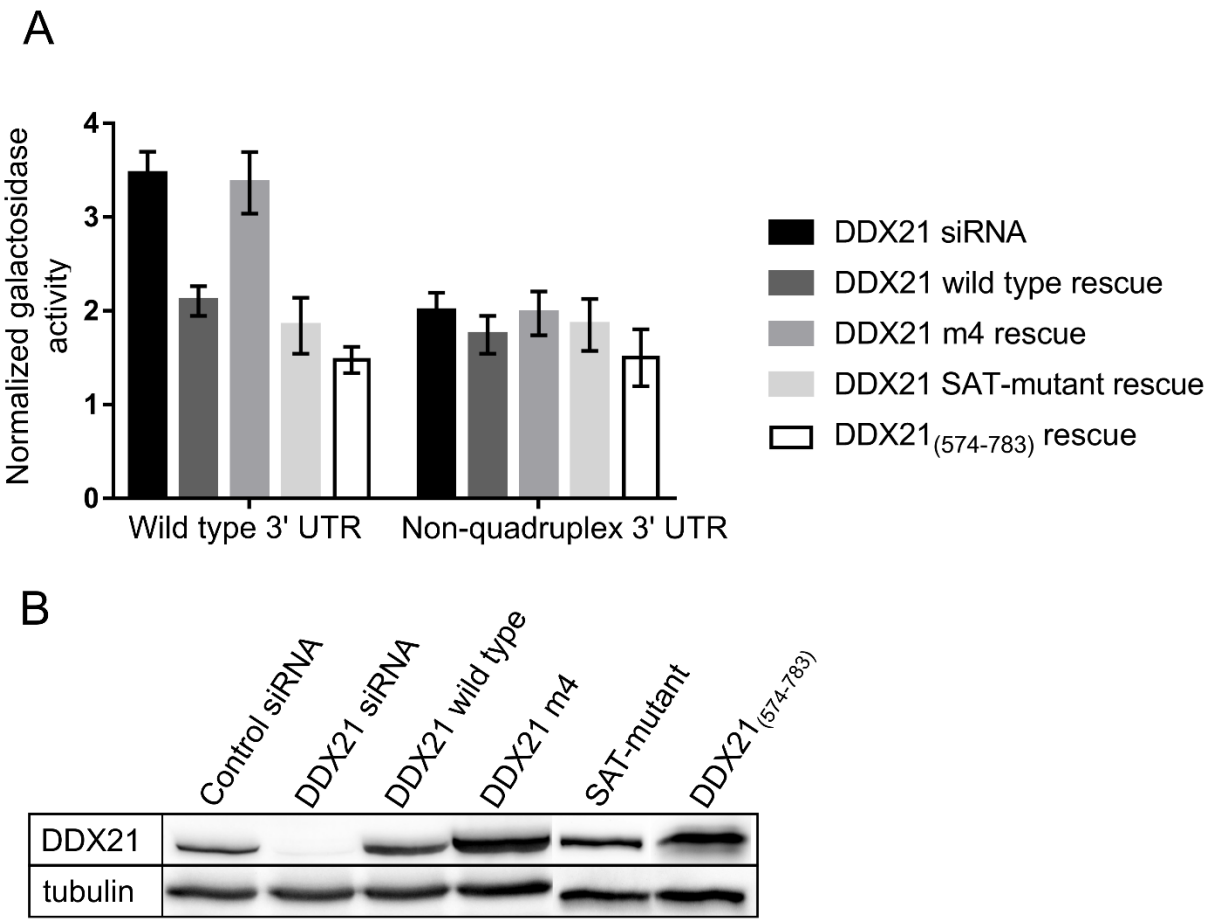


Figure 2-6: DDX21's quadruplex binding ability can affect gene expression

DDX21 wild type but not DDX21 m4 can supress the expression of a beta-galactosidase reporter gene with G4s in its 3'UTR. (A) Expression of beta-galactosidase in HEK293T cells from a transfected plasmid with the PITX1 3'UTR (solid black bars) or with a mutated quadruplex region (white bars) under conditions of DDX21 knock down and rescue with DDX21 wild type or m4 DDX21. Data is the mean of six biological replicates with standard error shown by error bars. Expression was quantified by spectrophotometric detection of ONPG hydrolysis and normalization to protein concentration and the expression level of cells transfected with only control siRNA. (B) Representative western blot of the levels of DDX21 and tubulin in the control, knock down and rescue samples. The DDX21 SAT-mutant and $DDX21_{(574-783)}$ samples were imaged on a separate western blot and cropped into the figure. $DDX21_{(574-783)}$ was

visualized with anti-FLAG primary antibodies because it lacks the binding epitope for the anti-DDX21 antibody.

5. Discussion

In recent years RNA G4 have become recognized as important regulators of gene expression and disease progression[111,112,120,130,132]. Recent studies have identified over 12,000 potential unique RNA G4s in the human transcriptome[25],[113]. Currently only two proteins, DHX36 and DHX9, have been shown to be able to unwind RNA G4, whereas the list of DNA G4 unwinding enzymes has grown much larger[121]. Expansion of our current knowledge of RNA G4 helicase proteins is essential to our understanding of the complex process of regulation by RNA G4 that has evolved in eukaryotes. In a recent landmark publication by Guo and Bartel it is shown that RNA G4 exist *in vivo* in a predominantly unwound state, or at least a distorted state that allows for chemical modification of guanines that would be expected to be protected in a G4 structure; DHX36 depletion and ATP depletion did not affect this globally unwound state[25]. It is clear then that other helicase proteins must be unwinding RNA G4 and we show herein that DDX21 one of what is likely many more proteins capable of doing so. Furthermore, we show that energy input from ATP is not a necessity for all RNA G4 substrates and that significant distortion of RNA G4 can occur with just the quadruplex binding portion of a protein present.

In this study, we used a biotinylated version of an RNA G4 (Q2) from the 3' UTR of PITX1 to isolate proteins from HEK293T cell lysate and identify them by mass spectrometry. The successful validation of DDX21 as a G4 binding and helicase protein confirms that our screening method is appropriate to identify candidate RNA G4 helicases and likely contains more potential G4 binding proteins awaiting validation. Indeed, there is some overlap in identified proteins between our screen and other mass spectrometry guided screens that have used different RNA G4 sequences, including one using the C9ORF72+ G4 that had identified DDX21 as a potential G4 binding protein in 2013[133,134]. Although this group never validated DDX21 via other methods, we have confirmed their finding by repeating the pull-down assay with C9ORF72+ and visualizing DDX21 by western blot. It is our hope that our list of candidate RNA G4 binding proteins will continue to be used as a resource for future identification of G4 interacting proteins, be they directly in contact with the G4 or acting as part of a larger protein complex. A full list of the proteins identified in our screen is available in the supplementary information.

In both the pull-down assays and MST experiments, DDX21 displayed an interesting specificity for different RNA G4 relative to its better studied counterpart DHX36. The Q2 G4 used in the

initial screen interacts strongly with DDX21 in the presence and absence of cell lysate, whereas DDX21 was outcompeted for $hTR_{(10-43)}$ in cell lysate and showed 30 fold less affinity than Q2 with purified components. It is yet unclear the exact cause of this specificity but it seems likely to be either a loop mediated interaction or 3' overhang mediated, as these are the main differences between the two RNAs. This is intriguing as most evidence thus far for G4 recognition by proteins has implicated pi-pi and CH-pi interactions with the tetrad face to be the predominant recognition mode[53,59,62], while electrostatic interactions with nearby phosphate groups play a more supplementary role. However, these studies have all used a short peptide containing the G4 binding motif and not the full protein of interest. Indeed, with the truncated C terminus of DDX21, the difference in affinity for the Q2 and $hTR_{(10-43)}$ is diminished. This indicates that interactions outside of the identified G4 binding domain may be providing the ancillary contacts that confer specificity for different G4 species. This idea is supported by the Hill coefficients from the MST data (Table 2-1) where large amounts of positive cooperativity is observed for the Q2 RNA and the full length DDX21 protein (Hill coefficient of 3.32) suggesting that the initial binding event may be promoting subsequent interactions to occur. No binding cooperativity was observed for the interaction between full length DDX21 and $hTR_{(10-43)}$ (Hill coefficient of 0.98) suggesting that the interaction utilizes only the C-terminal G4 binding domain and not the N-terminal portion of the protein. This hypothesis is also supported by the marginal increase in affinity for $hTR_{(10-43)}$ with $DDX21_{(574-783)}$ and the decrease in affinity observed between $DDX21_{(574-783)}$ and Q2.

Previously we have performed G4 helicase assays with DHX36 and the $hTR_{(10-43)}$ G4 where we used the endogenous binding partner of the G4 forming sequence (25P1) to trap the unwound G4 in a duplex that could be resolved from both the G4 and single stranded 25P1 species[57]. We performed the same assay with DDX21 in place of DHX36 and, while there was apparent remodeling of hTR, we were never able to visualize hybridization to the ssRNA as a distinct band. This could be due to the 5' to 3' dsRNA helicase activity of DDX21. Using the natural binding partner (25P1) for the hTR G4 as a trap in a helicase assay did work very well with DHX36, however, these RNA are more of an exception than a system that is easily transposable to any G4. We found that a T1 RNase assay provided a sensitive and less biased way to visualize the accessibility of guanines that are protected while in a G4 conformation.

Nuclease sensitivity has been used many times to identify G4s[111,120,135] and has also been used to show protection from nucleases by binding partners[54], recently it has also been used by You *et al* to visualize the unwinding of a G4 by DHX36[136]. The use of RNase T1 is appropriate for minimal G4 structures as the majority of guanines will be protected from the ribonuclease. Furthermore, the enzymatic efficiency of RNase T1 allows for this technique to be very sensitive to conformational changes that may have too short a life time to otherwise be resolved by gel electrophoresis methods. Our results show that interaction with the G4 binding domain is sufficient to distort both Q2 and $hTR_{(10-43)}$ enough to allow for digestion by RNase. Data from a previous publication[59] examining the circular dichroism spectrum of G4 in free and protein bound forms shows a decrease in ellipticity and a slight red shift at the ~265nm maximum upon binding of protein to G4, this is indicative of a distortion in the base stacking pattern of the guanine tetrads and reminiscent of what is observed upon partial melting or destabilization of G4 with lithium. Thus, it is likely that the RNA is still in a conformation resembling a G4 but with decreased stacking efficiency, which allows for the increased accessibility of RNase T1.

Interestingly, in the context of the full-length protein we observed differential ATP dependencies between our two RNA G4 substrates that is consistent between both DDX21 and our positive control G4 helicase protein DHX36. Both enzymes reduced the nuclease susceptibility of Q2 in the presence of ATP and enhanced the nuclease susceptibility of $hTR_{(10-43)}$. This is consistent with previous studies on DHX36 that have demonstrated an ATP requirement for the unwinding of hTR G4[57] but not other RNA G4[137].

These results have implications for the utility of the G4 binding domain in G4 helicase mechanism. This is consistent with the repetitive ATP-independent unfolding mechanism proposed by Myong *et al* [137] in their recent study of quadruplex helicase using FRET where a significant change in the FRET efficiency was observed upon binding of DHX36 to the quadruplex. Further investigation with a suite of G4 species rationally designed to discern substrate based ATP dependence would aid in clarifying the sequence dependencies of this effect.

Previous DDX21 foldase assays[88] have used an 81nt RNA known as S1 that DDX21 could unwind as well as a second RNA, termed S2, that could not be unwound. The authors speculated

that heating and cooling of the two RNAs resulted in a single stranded conformation and that incubation with DDX21 unwound the S1 RNA and not the S2 RNA into a double stranded structure but no spectroscopic evidence of this was published. Inspection of the sequence of S1 reveals 7 runs of guanines that could participate in a G4 formation. S2, on the other hand, begins with 3 guanines but has no other guanine rich areas and would likely be unable to form G4. Our results demonstrate that both conformations of S1 are unique G4 that utilize different tracts of guanines, and propose that the previously observed foldase activity of DDX21 is in fact G4 helicase activity. The bottom band is likely a kinetically favorable G4, as it is formed after boiling and snap cooling. Our results indicate that DDX21 can resolve the top band in the presence of ATP sufficiently to allow for refolding of the RNA to the kinetically favored bottom band. Conversely, in the absence of ATP, DDX21 allows for the conversion of small amounts of the bottom band to the top band; $DDX21_{(574-783)}$ does this much more effectively. We hypothesize that the binding of $DDX21_{(574-783)}$ results in a distortion of the G4 conformation, as observed in the nuclease sensitivity assays, that alters the thermodynamic landscape such that it can access the more stable conformation more readily.

The data herein show that DDX21 is capable of binding RNA G4 with high affinity and resolving them in an ATP dependent and independent manner. We have shown that $DDX21_{(574-783)}$ is necessary and sufficient for the G4 interaction and identified a new, unique, G4 binding motif within the C terminus of DDX21; the three FRGQR repeats followed by a final PRGQR sequence, a pattern that does not appear in other human proteins when a BlastP[138,139] search is performed. The sequence in this area is very rich in arginine and glycine and is reminiscent of the RG/RGG binding motif that is observed in many RNA binding proteins[38] as well as the FMRP G4 interaction site[54]. Mutation of the PRGQR residues to YEGIQ has been previously shown to eliminate the foldase activity of DDX21[88]. Interestingly, same mutation, here called DDX21 m4, drastically reduces the affinity of DDX21 for the G4 and renders the protein ineffective in nuclease sensitivity assays as well as our β –galactosidase assays. Considering the specificity for G4 and the relatively low dissociation constants determined by MST along with the high abundance of DDX21 in cells ($\sim 4.2 \times 10^4$ molecules per cell[81]) and the G4 mediated expression regulation of β –galactosidase in HEK293T cells, it is likely that DDX21 is recruited to G4 containing RNA *in vivo* where it can affect their conformational state resulting in a

modulated biological outcome. These results provide the basis for future investigations into novel cellular processes that are regulated by DDX21 RNA G4 interactions.

6. Materials and Methods

2.6.1 Cell culture and reagents

The HEK293T cell line was a gift from Dr. Thomas Klonisch (University of Manitoba). The monoclonal murine anti-RHAU hybridoma was a kind gift from Dr. Yoshikuni Nagamine (Friedrich Miescher Institute for Biomedical Research). Cell culture and monoclonal antibody purification was performed as previously described[57]. The following additional antibodies were used: mouse anti-alpha-tubulin (T6074, Sigma-Aldrich, Oakville, ON, Canada), mouse anti-FLAG® M2 antibody (F3165, Sigma-Aldrich), rabbit anti-DDX21 (NB100-1717, Novus Biologicals, Littleton, CO, USA). DDX21 siRNA were purchased from Life Technologies sequence: CGGGAAUUAAGUUCAAACGAA, all other synthetic RNAs and DNA primers, including the negative control siRNA cat# 51-01-14-04, were purchased from Integrated DNA technologies (Coralville, IA, USA). The $hTR_{(10-43)}$ RNA was purchased and 3' biotin labelled in-house using methods previously described[125]. All other RNA were purchased with a 5' biotin and, separately, with a 5' Cy5, except Q2, which was purchased with a 5' FAM fluorophore. The pCp-Biotin and pCp-Cy5 were purchased from Jena Bioscience (Jena, Germany). Full sequence information for all RNA species used can be found in Table 2-2. Plasmid containing the DDX21 gene was purchased from Genscript Inc. The pSKB- plasmid was a gift from Eric Marois (Addgene plasmid # 62540)[140]. The G4 stain N-methyl-mesoporphyrin IX was purchased from Frontier Scientific (Logan, UT, USA). SYBR Gold and Lipofectamine RNAiMax were purchased from Life Technologies (Burlington, ON, Canada). All standard laboratory chemicals and reagents were purchased from Thermo-Fisher Scientific (Ottawa, ON, Canada).

2.6.2 Preparation of protein constructs and RNA

The cDNA that encodes the human DDX21 protein, isoform 1, was amplified by PCR using DNA primers that were designed to encode an N terminal DYKDDDDK (FLAG) tag and cloned using standard molecular biology techniques. Mutations to the C-terminal FRGQR region and siRNA resistance were accomplished by standard site directed mutagenesis techniques[141]. The S1 RNA for the foldase assays was *in vitro* transcribed from pSKB- (BlueScript II KS) plasmid digested with HindIII and purified as previously described[142]. After purification, the S1 RNA was ~95% in the top band conformation, boiling for 5 minutes and snap cooling on ice was used to convert the RNA to ~95% bottom band conformation. All purchased RNA were ordered

desalted and initially dissolved in TE buffer upon arrival to a stock concentration of 100 µM. From the stock vial, RNA was diluted into the appropriate buffer and heated for 5 minutes before snap cooling on ice.

2.6.3 Recombinant protein expression

$DHX36_{(53-105)}$, herein referred to as DHX36 RSM, was purified from *E.coli* as described previously[59]. Protein purification of DDX21 constructs and full length DHX36 was performed as described previously[143], substituting commercial transfection formula for (90 µg per 150mm cell culture dish) polyetheleneimine. Protein concentration was determined spectrophotometrically by absorbance at 280nm using extinction coefficients calculated with the ProtParam tool on ExPASy servers[144]. For the streptavidin pull-down assays with overexpressed DDX21 constructs, per 150mm cell culture dish, 60 µg of plasmid was used for $DDX21_{(1-783)}$, $DDX21_{(217-783)}$ and $DDX21_{(217-573)}$, but only 10 µg of plasmid and 15 µg of polyetheleneimine was needed to obtain comparable levels of expression for the $DDX21_{(1-216)}$ and $DDX21_{(574-783)}$ constructs 24 hours post transfection.

2.6.4 SDS-PAGE

For SDS-PAGE 5ug of protein was heated at 95 °C for 5 minutes in SDS loading dye (0.5% β-Mercaptoethanol, 0.002% Bromophenol blue, 6% glycerol, 1% Sodium Dodecyl sulfate, 25mM Tris-Cl pH 6.8), centrifuged briefly and loaded onto a 15% acrylamide:bis-acrylamide (39:1). The gel was then electrophoresed at 200V until the bromophenol blue was 0.5cm from the bottom of the gel. The gel was then stained with Coomassie Brilliant Blue for 10 minutes and destained overnight in a glacial acetic acid, methanol and water (10:40:50) solution before imaging with an Epson perfection 3170 photo-scanner.

2.6.5 Protein thermal shift assays (PTS)

Protein thermal shift assays (Thermo-Fisher Scientific Ottawa, ON, Canada) were performed per the manufacturers recommended protocol (1x PTS buffer and 1x PTS dye). Fluorescence was measured on a StepOnePlus real-time PCR system (Thermo-Fisher Scientific Ottawa, ON, Canada) using the existing ROX spectral settings and the slowest temperature increase (1%) to record data for approximately every 1-degree interval. The first derivative of the fluorescence

over temperature was calculated by the StepOnePlus software, normalized to the largest slope value, and used to determine the approximate melting temperature of each protein.

2.6.6 Streptavidin pull-down assays and western blotting

To screen for G4 binding proteins, HEK293T cells from confluent 150mm dishes were resuspended in 10ml of cold phosphate-buffered saline per plate and pelleted by centrifugation at 1500g for 5 minutes at 4°C. Per plate of cells, the cell pellet was suspended in 250µL of cytoplasmic lysis buffer (25mM Hepes, Ph7.9, 5mM KCl, 0.5mM $MgCl_2$, 0.5% (v/v) NP-40) supplemented with Halt protease and phosphate inhibitor cocktail and Ribolock RNase inhibitor (Thermo-Fisher Scientific). The cells were lysed on ice for 5 minutes and centrifuged at 5000rpm in a benchtop microfuge for 5 minutes at 4°C. The supernatant was set aside on ice and the pellet was suspended in 250µL nuclear lysis buffer (25mM Hepes, pH 7.9, 10% (w/v) sucrose, 350mM NaCl, 0.01% (v/v) NP-40) supplemented with Halt protease and phosphate inhibitor cocktail and Ribolock RNase inhibitor, per plate of cells. To lyse the nuclei, the suspension was vortexed for 30s followed by passage through a 20-guage needle. The nuclear and cytoplasmic fractions were combined and cleared of insoluble material by centrifugation at 14000rpm in a bench-top microfuge for 10 minutes at 4°C. 5'-biotin labelled RNA were added to a final concentration of 500nM to 500 µL of cell extract and rotated for 30 minutes at room temperature. Subsequently, 40 µL of Pierce streptavidin magnetic beads (Thermo-Fisher Scientific) were added to the samples and rotation at room temperature continued for another 30 minutes. The beads were captured in a magnetic rack and the cell extract aspirated. The beads were then washed three times with 1ml of a 1:1 mixture of the cytoplasmic and nuclear lysis buffers, followed by three washes each with 1ml of cytoplasmic lysis buffer then three washes with 1ml of nuclear lysis buffer. For validation of DDX21 as a hit, 20% of the beads from a pull-down sample and 50µg of cell extract was loaded onto a 10% SDS-PAGE and western blotting was performed as described previously[57].

2.6.7 Sample preparation for MS

Beads from each pull-down sample were washed three times with 1mL 50mM ammonium bicarbonate. The beads were transferred and resuspended into Siliconized vials (BioPlas, San Rafael, CA) in 50uL of 50mM ammonium bicarbonate and the proteins were reduced by 10mM DTT at 50C for 30 min. The proteins were alkylated with 30mM iodoacetamide for 30min in the

dark at room temperature. Unreacted iodoacetamide was quenched by addition of 15mM DTT. Finally, the protein complexes were digested by 500ng of sequencing grade Trypsin (Promega) overnight at 37C using a tube roller with gentle horizontal mixing. The reaction was stopped using 1% trifluoroacetic acid (TFA) by adding 50uL of 3% TFA to the peptide-beads mixture. The beads were vortexed for 10 minutes and peptides were extracted. In order to maximize peptide-yield from each sample, we preformed two sequential extractions using 200uL of 0.1% TFA in acetonitrile, and 20mM ammonium formate (pH11) in acetonitrile. Peptides from each bead wash were pooled and dried using speed-vac. The dried peptide were dissolved in 100µl of 0.5 % TFA and desalted with 1mL C18-SD extraction disc cartridge (3M,USA). 2ug of desalted peptide as determined by NanoDrop 2000, (ThermoFisher) was used for LC-MS/MS analysis.

2.6.8 Nano RP-LC-MS/MS

Samples were analyzed by nano-RP-LC-MS/MS using an a splitless Ultra 2D Plus (Eksigent, Dublin, CA) system coupled to a high-speed Triple TOF™ 5600+ mass spectrometer (SCIEX, Concord, Canada). Peptides were injected via a PepMap100 trap column [0.3 × 5mm, 5 µm, 100 Å,(Thermo Scientfic, CA) and a 100 µm×200 mm analytical column packed with 3µm Luna C18 (2) (Phenomenex Inc, CA) was used prior to MS/MS analysis. Both eluents A (LC-MS water), and B(LC-MS acetonitrile) contained 0.1% formic acid as an ion-pairing modifier. The tryptic digest was analyzed using 60 minutes LC-MS run time. Eluent B had a gradient from 0% to 35% over 48 minutes, 35% to 85% in 1 minute and was kept at 85% for 5 minutes at a flow rate of 500 nL/min. Key parameter settings for the TripleTOF 5600+ mass spectrometer were as follows: ion spray voltage floating (ISVF) 3000 V, curtain gas (CUR) 25, interface heater temperature (IHT) 150, ion source gas 1 (GS1) 25, declustering potential (DP) 80 V. All data was acquired using information-dependent acquisition (IDA) mode with Analyst TF1.6 software (SCIEX,USA). 0.25 seconds MS survey scan in the mass range of 380–1250(m/z) were followed by 20 MS/MS scans of 100 milliseconds in the mass range of 100–1600 (total cycle time: 2.3 s). Switching criteria were set to ions greater than mass to charge ratio (m/z) 380 and smaller than m/z 1250 with a charge state of +2 to+5 and an abundance threshold of more than 150 counts. Former target ions were excluded for 7 seconds. A sweeping collision energy setting of 37 ± 15 eV was applied to all precursor ions for collision-induced dissociation.

2.6.9 Database search and protein identification

Raw spectra from WIFF files containing MS and MS/MS data were analyzed using Protein Pilot 4.5 software using Paragon algorithm (Sciex). Protein identification parameter of carbamidomethylation of cysteine was selected. All LC-MS samples were searched against the curated UniProt's human proteome release (2017_02) containing 42147 protein entries for unique canonical sequence and splice isoforms. Proteins and peptides identified at >95% confidence (Unused Score >1.3) were used for subsequent comparative quantitative analysis. Gene ontology analysis for molecular function of the mass spectrometry results from the pull-down assay was performed using PANTHER web services[123,145,146] using the GO Ontology database, released 2017-02-28.

2.6.10 Microscale thermophoresis (MST)

To perform MST, DDX21 was diluted in MST buffer (50 mM Tris-Cl pH 7.5 with 50 mM KCl, 5 mM $MgCl_2$, 0.01% NP-40), concentrated to 1 µM and supplemented with 0.05mg/ml bovine serum albumin (BSA). Both the labelled RNA and the protein samples were centrifuged at 21,500xg for 10 minutes to remove aggregates. Binding reactions were prepared in MST buffer supplemented with 0.05mg/ml BSA to a total volume of 20 µL. Protein samples were serially diluted to obtain a concentration range spanning ~0.1x K_D to ~10x K_D, where possible, and mixed with fluorescent RNA which was held constant at a final concentration of 6.125 nM. Premium coated capillaries (NanoTemper Technologies, San Fransisco, CA, USA) were used for all MST measurements. Measurements were performed on the Monolith NT.115 instrument (NanoTemper) at 95% LED power with the blue excitation LED for the FAM-Q2 and FAM-Q2-mutant RNA and 50% LED power for the Cy5-$hTR_{(10-43)}$ RNA. MST-IR power of 40% was used for all the samples. All Samples were prepared and measured three times independently. The resulting signal from thermophoresis was fit to the Hill equation.

2.6.11 Nuclease sensitivity assay

Reactions were prepared in 50 mM Tris-acetate, pH 7.8, 100 mM KCl, 10 mM NaCl, 3 mM $MgCl_2$, 70 mM glycine, 10% glycerol, equal volumes of protein or protein storage buffer (50 mM Tris-Cl pH 7.5, 50 mM KCl, 5 mM DTT, 20% glycerol), the appropriate RNA, ATP/AMP-PNP (5mM) or water and T1 RNase or water were added. RNA concentrations were 150 nM, and 25 ng of protein was used in each sample. ATP and AMP-PNP were added to a final

concentration of 5 mM or the same volume of water was added. The samples were thoroughly mixed before addition of RNase T1 to a concentration of 1U/ul. The samples were then mixed again and incubated in a 30°C water bath for 20 minutes. After incubation the samples were mixed with an equal volume of denaturing load dye (95% Formamide, 0.01% SDS, 0.5 mM EDTA), heated at 95 °C for 5 minutes, cooled on ice for 1 minute and resolved by denaturing Tris-borate-EDTA urea polyacrylamide gel electrophoresis (TBE-UREA PAGE) on 10% acrylamide gels (29:1 acrylamide:*bis* ratio). RNA were detected on a FluorchemQ imager using the appropriate excitation LEDs and emission filters (ProteinSimple, San Jose, CA, USA) after staining with SYBR gold® (Thermo-Fisher Scientific Ottawa, ON, Canada).

2.6.12 Nuclease digestion and DDX21-foldase assay

The S1 RNA was labelled with Cy5 for the digestion assays by ligation of pCp-Cy5 (Jena Bioscience, Jena, Germany) to the 3' end of *in vitro* transcribed S1 using T4 RNA ligase 1 (New England Biolabs, Ipswich, MA, USA). The labelled RNA was then FPLC purified in TE buffer and the predominant peak (S1 top band) was recovered. Half of the S1-Cy5 was converted to the bottom band by heating at 95° Celsius and cooling on ice. Each species (top and bottom band) was then diluted to 10nM in 30ul of PBS and treated with either no RNase, 0.5 units of T1, or 10pg of RNase at room temperature. After 20 minutes the samples were mixed with an equal volume of denaturing load dye, heated at 95 °C for 1 minutes, cooled on ice and resolved by denaturing Tris-borate-EDTA urea polyacrylamide gel electrophoresis (TBE-UREA PAGE) on 15% acrylamide gels (29:1 acrylamide:*bis* ratio). RNA were detected on a FluorchemQ imager using the appropriate excitation LEDs and emission filters (ProteinSimple, San Jose, CA, USA).

The foldase assay was performed using 150 nM S1 RNA concentration in a buffer containing 20 mM HEPES/KOH, pH 7.6, 2 mM DTT, 3 mM $MgCl_2$, 100 mM and KCl, 5% glycerol. ATP concentration was either 0, 1mM (+) or 5mM (+++), 25 ng of protein was used and the samples were incubated at 30 °C for 20 minutes. The reaction mixtures were loaded directly onto a 10% polyacrylamide (29:1) TBE gel and resolved by electrophoresis at 75 V for 90 minutes before visualization with SYBR Gold®.

2.6.13 Fluorescence measurements and in gel staining with Thioflavin T

For the fluorescence measurements, Thioflavin T was first suspended in TEK buffer (10 mM Tris-Cl, pH 8.0, 10 mM EDTA, 50 mM KCl) to a working concentration of 5 µM. S1 RNA and

25P1 RNA (ssRNA) were diluted in TEK buffer into the wells of a 96 well tray to achieve a final concentration range from 0-15 µM. Thioflavin T was then mixed with each of the samples to a final concentration of 2 µM and allowed to incubate for 5 minutes before the fluorescence was measured with a FluorchemQ imager using the Cy2 excitation and emission filters.

For in gel staining of S1 RNA with Thioflavin T, 25 µg of each S1 RNA species as well as a single stranded RNA 25P1 RNA were mixed with a native RNA loading buffer (5% glycerol, 0.2mg/ml Orange G) and loaded onto a 10% acrylamide (29:1) native TBE gel and electrophoresed at 75 V for 90 minutes before staining in 10 mL of water with a 25 µM Thioflavin T concentration for 5 minutes. The gel was then imaged on a FluorochemQ imager using Cy2 excitation and emission filters. The same gel was then stained with 1%(w/v) Toluidine blue for 5 minutes and de-stained in water before imaging with an Epson perfection 3170 photo-scanner.

2.6.14 Circular dichroism

All circular dichroism spectra were obtained using a J-810 spectropolarimeter (Jasco Inc., USA) and a 0.1cm quartz cell (Hellma). Samples and baseline buffer spectra were measured under continuous scanning mode with a 1nm data pitch and a scan speed of 50 nm/min with a response time of 1 second. All spectra presented are from the accumulation of 6 scans of the RNA at 6.5 µM with the buffer signal (TEK) subtracted.

2.6.15 Thermal difference spectra

Thermal difference spectra were collected on a dual beam Evolution 260 Bio UV-Vis spectrophotometer (Thermo Scientific). RNA were diluted in TEK buffer to a final concentration of 1.5 µM and the spectra measured at 20 °C was subtracted from the spectra measured at 90 °C for each sample. Difference spectra were obtained in triplicate and normalized to the maximal absorbance value for a given spectrum.

2.6.16 siRNA transfections and β-galactosidase reporter assays

β –galactosidase reporter assays were performed six times for biological replicates, each data point was normalized to protein concentration and subjected to a two-sided Grubbs' test at a 95% confidence interval to remove any outliers. To perform the reporter assays, $1x10^6$ cells were either mock transfected (with only transfection reagent) or transfected with pcDNA vector

containing siRNA resistant DDX21 wild type or m4 mutant. After 24 hours the cells were split 1:3 and reverse transfected with either DDX21 or a control siRNA using lipofectamine RNAiMAX (Invitrogen) using the protocol supplied by the manufacturer. After a further 48 hours, the cells were transfected with the β-galactosidase reporter vectors containing the PITX1 3'-UTR or a version of the UTR with all three G4s mutated as previously described[14]. After an additional 24 hours, the cells were harvested and levels of β-galactosidase were assessed using the Promega β-galactosidase enzyme assay system (Promega Corporation). Absorbance measurements were made on a Biotech Epoch 96-well spectrophotometer (Fisher Scientific).

7. Supplemental Information

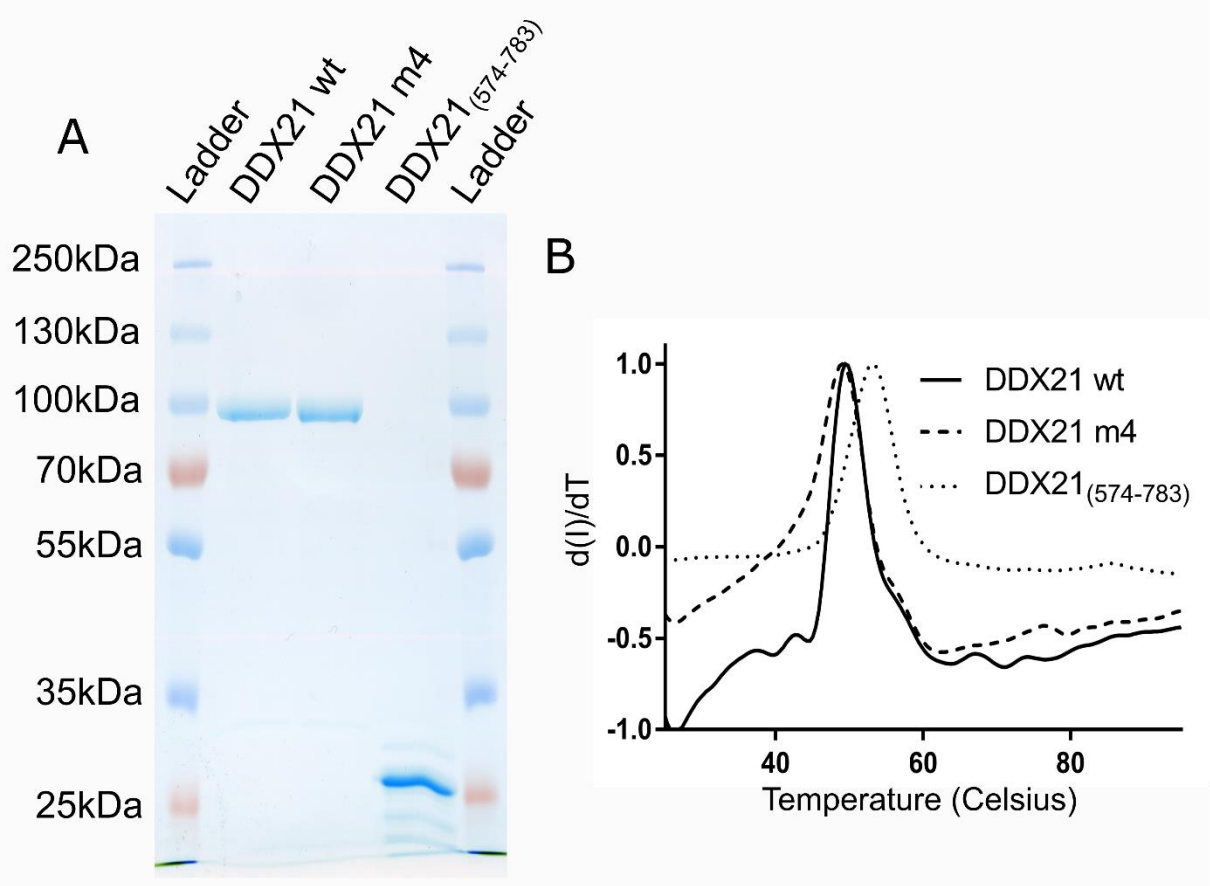


Figure 2-7: Purification of DDX21 constructs

(A) Coomassie brilliant blue stained SDS-PAGE of FLAG-purified protein constructs of DDX21. (B) Melting point analysis of purified proteins using a fluorescent thermal shift assay.

Name	Sequence
Q1	GAGCGGGGAAGGGCGCGGGCGCGGGCGCGG

Table 2-1: Sequences of RNA molecules used

Q2	GUUGGGGCGGGCGUUGGGUUUGGGGGGACG
Q3	AGCGGGCAGUGCGGGCCUGGCGGGAGGUGGGGGAGG
Q1-mutant	GAGCGCGGAAGCGCGCGCGCGCGCGCGG
Q2-mutant	GUUGCGCCGCGCGUUGCGUUUGCGCGCACG
$hTR_{(10-43)}$	AGGGUGGGCCUGGGAGGGGUGGUGGCCAU
C9ORF72+	GGGGCCGGGGCCGGGGCCGGGGCC
C9ORF72-	CCCCGGCCCCGGCCCCGGCCCCGG
S1	GGGCGAAUUGGAGCUCCACCGCGGUGGCGGCCGCU CUAGAACUAGUGGAUCCCCCGGGCUGCAGGAAUUCG AUAUCAAGCU

Table 2-2: MST Fitting Parameters

Best fit values determined from fitting of MicroScale Thermophoresis data to the Hill equation

RNA	DDX21 construct	Dissociation constant (K_d)	Standard Error	Hill coefficient	Standard Error	R square
Q2	(1-783)	10.0 nM	0.7 nM	3.32 nM	0.62	0.9511
	(574-783)	86.9 nM	7.2 nM	1.74 nM	0.22	0.9485
$hTR_{(10-43)}$	(1-783)	324 nM	7 nM	0.980 nM	0.032	0.9594
	(574-783)	149 nM	10 nM	1.82 mnM	0.19	0.9631
Q2-mutant	(1-783)	Not converged	-	-	-	-
	(574-783)	7780 nM	1170 nM	1.03 nM	0.11	0.8694

Chapter 3: An RNA guanine quadruplex-regulated pathway to TRAIL-sensitization by DDX21

3.1.1 Preface

This chapter is available on the pre-print server BioRxiv at the following URL and is currently under review by Nucleic Acids Research:

https://www.biorxiv.org/content/10.1101/588798v1

As stated in the discussion section of the previous chapter, the results from Chapter 2 provide the basis for future investigations into novel cellular processes that are regulated by DDX21 RNA G4 interactions. To identify RNA G4 that interact with DDX21 in a cellular context we used the information gained from the work in Chapter 2 to design the major experiment in Chapter 3. By replacing endogenous DDX21 in human cell culture with either wild-type DDX21 or the M4 mutant we parse out changes in protein expression that are due to the lack of DDX21 G4 binding abilities. We use a top down approach to look at the global effects of DDX21's G4 binding abilities on the whole cell proteome by mass spectrometry. Following this we narrow our scope and focus on the DDX21 affected protein MAGED2. We confirm the presence of RNA G4 in the mRNA of MAGED2 and make a strong case for the regulation of MAGED2 mRNA translation by DDX21. This work identifies numerous promising candidate proteins that are affected by DDX21's G4 interacting abilities that merit further investigation. Furthermore, MAGED2 regulation by DDX21 presents a novel mechanism of sensitizing cells to TRAIL-mediated apoptosis.

3.1.2 Contribution of Authors

Experimental design, writing of the manuscript, figure preparation and all experiments not attributed to others below were performed by Ewan K.S. McRae. Steven J. Dupas performed the cloning and *in vitro* transcription of MAGED2 5'UTRs as well as the EMSA and reverse-transcriptase stalling assays. Evan P. Booy performed the DDX21 immunoprecipitations and created the truncated CMV promoter used in the luciferase assays. Richard P. Fahlman and Ramanaguru Piragasam performed the mass spectrometry experiments and analysis. Sean A. McKenna provided oversight of the project and editorial comments on the manuscript.

2. Abstract

DDX21 is a newly discovered RNA G-quadruplex (rG4) binding protein with no known biological rG4 targets. In this study we identified 26 proteins that are expressed at significantly different levels in cells expressing wild type DDX21 relative to an rG4 binding deficient DDX21 (M4). From this list we validate MAGED2 as a protein that is regulated by DDX21 through rG4 in its 5'UTR. MAGED2 protein levels, but not mRNA levels, are reduced by half in cells expressing only DDX21 M4. The repressive effect of MAGED2 on TRAIL-R2 expression is dampened under these conditions, resulting in elevated TRAIL-R2 mRNA and protein in cells expressing only DDX21 M4 and rendering the previously resistant cells sensitive to TRAIL mediated apoptosis. Our work identifies the role of DDX21 in regulation at the translational level through biologically relevant rG4 and shows that MAGED2 protein levels are regulated, at least in part, by a rG4 forming potential in their 5'UTRs.

3. Introduction

DDX21 is an RNA helicase protein with a diverse set of biological functions. It plays important roles in ribosomal RNA biogenesis[90,91] by coordinating Pol I transcription and association of late acting snoRNAs with pre-ribosomal complexes[93,94]. DDX21 has also been implicated in double stranded RNA (dsRNA) sensing and anti-viral response[105,106,108,109] as well as epigenetic silencing of genes[100]. Abnormal levels of DDX21 protein has been observed in colorectal[102] and breast cancers[124] (where it has been correlated with disease free survival, possibly mediated by regulating c-Jun activity and rRNA biogenesis[96,99]) as well as in schizophrenic patients, making it a potential therapeutic target. Recently, we have shown that DDX21 can bind and unwind guanine-quadruplexes (rG4s) and affect the translation of a reporter construct with a rG4 in its 5' UTR[147] . This activity is mediated by an evolutionarily conserved repeat region (F/PRGQR) in its C-terminus that preferentially interacts with the backbone of RNA rG4s[148]. It is unlikely that such an activity would exist and persist in an enzyme if there were no biological need for it. Recent iCLIP data using anti-DDX21 antibodies showed that ~35% of DDX21 bound RNA is mRNA[98], yet no rG4 targets of DDX21 are currently known.

RNA rG4s are four stranded structures of RNA that form in regions rich in guanines. When four guanines align in a plane they can form an extended hydrogen bonding network called a guanine

tetrad. Successive guanine tetrads can pi stack on top of one another to form a stable rG4 structure. The canonical rG4 forming sequence is GGG(N_{1-7})GGG(N_{1-7})GGG(N_{1-7})GGG, where (N_{1-7}) represents a short loop of 1-7 nucleotides. However, longer loops as well as interrupted G-tracts have also been shown to still allow for rG4 formation. Canonical rG4s[120] as well as irregular rG4s[149] present in mRNA have been shown to modulate translation of the mRNA, either through stalling of scanning pre-initiation complexes on 5' UTRs[120], by alternative splicing mediated by rG4s[150–152] , or direction of miRNA binding to 3' UTRs[14,153]. A transcriptome-wide sequencing approach that exploits the ability of RNA rG4s to cause reverse transcriptase stalling (rG4-seq) has identified thousands of new putative quadruplexes, many of which are non-canonical rG4s that would likely be missed by bioinformatic searches[113].

In our current study we use proteomic mass spectrometry (MS) to compare the levels of proteins in cell populations that are expressing wild type DDX21 or a DDX21 mutant with impaired rG4 binding and unwinding abilities (DDX21 M4). We compare this list of candidate proteins to the rG4-seq database, finding a significant enrichment for rG4 containing mRNA in out list of candidate proteins, and confirm the interaction of DDX21 with the mRNA of candidate proteins by RIP-qPCR experiments. Melanoma associated antigen D2 (MAGED2) protein abundance is significantly reduced in the cells expressing DDX21 M4 and not the WT DDX21. MAGED2 has been implicated in cancer progression and resistance to treatment of human breast cancers, specifically, MAGED2 acts as a transcriptional repressor of tumour necrosis factor-related apoptosis-inducing ligand (TRAIL) death receptor 2 (TRAIL-R2) mRNA[154],[155]. For decades TRAIL has been studied as a potential natural antitumoral protein with the ability to selectively trigger cancer cell death through interactions with surface expressed receptors DR4 and DR5 (TRAIL-R1 and TRAIL-R2)[156]. TRAIL resistance occurs in breast cancer cells when the death receptors DR4 and DR5 (TRAIL-R2) are no longer expressed on the cell surface[157]. TRAIL sensitization is emerging as a novel therapeutic target for treating resistance breast cancers[158,159]. We confirm by qPCR and western blot that regulation of MAGED2 by DDX21 also affects the downstream target of MAGED2, TRAIL-R2, and imparts a sensitivity to TRAIL mediated apoptosis.

MAGED2 has three transcript variants (TV) that differ only in their 5'UTRs, we cloned each unique 5'UTR into Renilla luciferase reporter constructs to determine their affect on translation.

TV1 and TV2 show similar DDX21 dependent effects on translation as was observed by mass spectrometry and treatment with an RNA rG4 stabilizing ligand has a similar effect on the luciferase assays as knock-down of DDX21. Furthermore, mutation of putative rG4 forming stretches of guanines eliminates the difference in translation observed between DDX21 states. Together these results indicate that DDX21 regulates translation of MAGED2 mRNA in a rG4 dependent manner and contributes to sensitization to TRAIL mediated apoptosis.

4. Results

3.4.1 Identifying putative G4 targets of DDX21

To look for potential G4 targets of DDX21 we wanted to compare cell populations expressing a wild-type (WT) and a mutant (M4) DDX21 with impaired G4 binding. The M4 mutant contains a PRGQR to YEGIQ mutation in the C-terminal repeat region of DDX21 that has been previously shown to disrupt rG4 binding and unwinding activities[147]. Differences between the WT DDX21 and M4 DDX21 recovery experiments should be due to differences in DDX21 G4 interactions. Briefly, HEK293T cells were first transfected vector expressing either WT or M4 DDX21. Following this, endogenous DDX21 was depleted by siRNA knock-down. Using label-free whole-cell proteome MS/MS we identified 26 candidate proteins **(Figure 3-1A**Error! eference source not found.**)** whose expression level was significantly different between the M4 and WT samples.

The data in **Figure 3-1A** is represented as the ratio of the amount of a protein detected in the M4 vs WT samples, this allows us to visualize a fold change in protein level in the absence of DDX21's G4 binding. Proteins with large positive values like XPO6 have increased expression in the presence of M4 DDX21, indicating that the G4 interaction may be impeding protein expression. Whereas proteins with large negative values like CNOT3 have significantly decreased expression in the presence of M4 DDX21, indicating that a G4 interaction with DDX21 could be promoting expression of the protein.

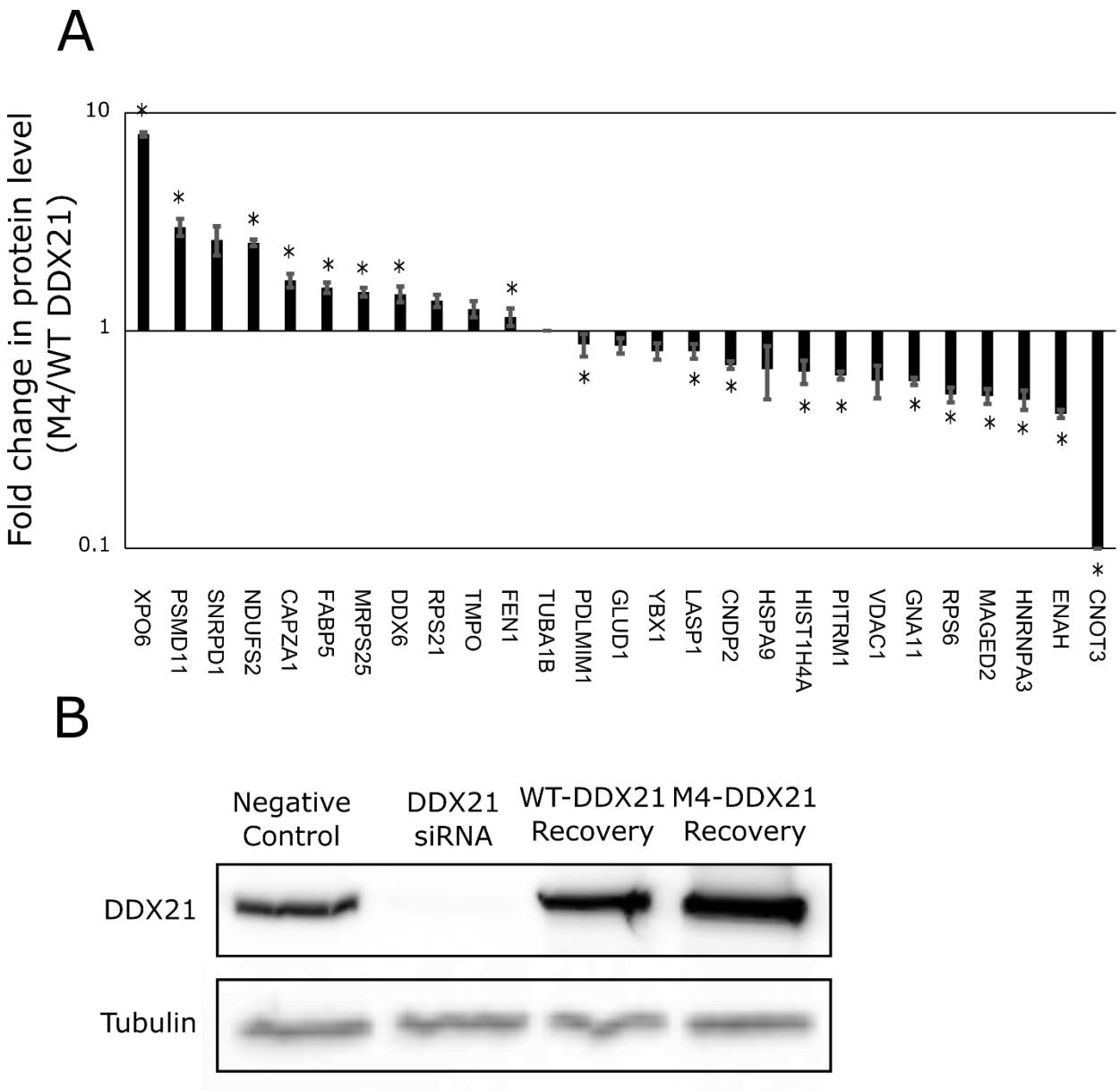


Figure 3-1. Identification of DDX21 G4 targets by whole cell proteome mass spectrometry

(A) Fold difference in protein levels by MS between HEK293T cells expressing M4 DDX21 over WT DDX21. Two-way Anova tests determined the significance of the difference, error bars represent the standard deviation between 3 biological replicates. (B) Representative western blot showing DDX21 levels between samples.

To determine if the change in protein levels could be due to altered mRNA level we performed qPCR on RNA extracts from the samples **(Figure 3-2)**. While some mRNA levels were found to

be significantly different than NC samples (PSMD11, HNRNPA3, TMPO, CAPZA, HIST1H4A & CNOT3), there was no significant difference observed between WT and M4 recovery samples, indicating that the altered mRNA levels are likely artifactual. Since the observed differences between WT and M4 protein levels could not be explained by differences in mRNA abundance, the remaining reasonable explanations are altered translation efficiency or protein stability.

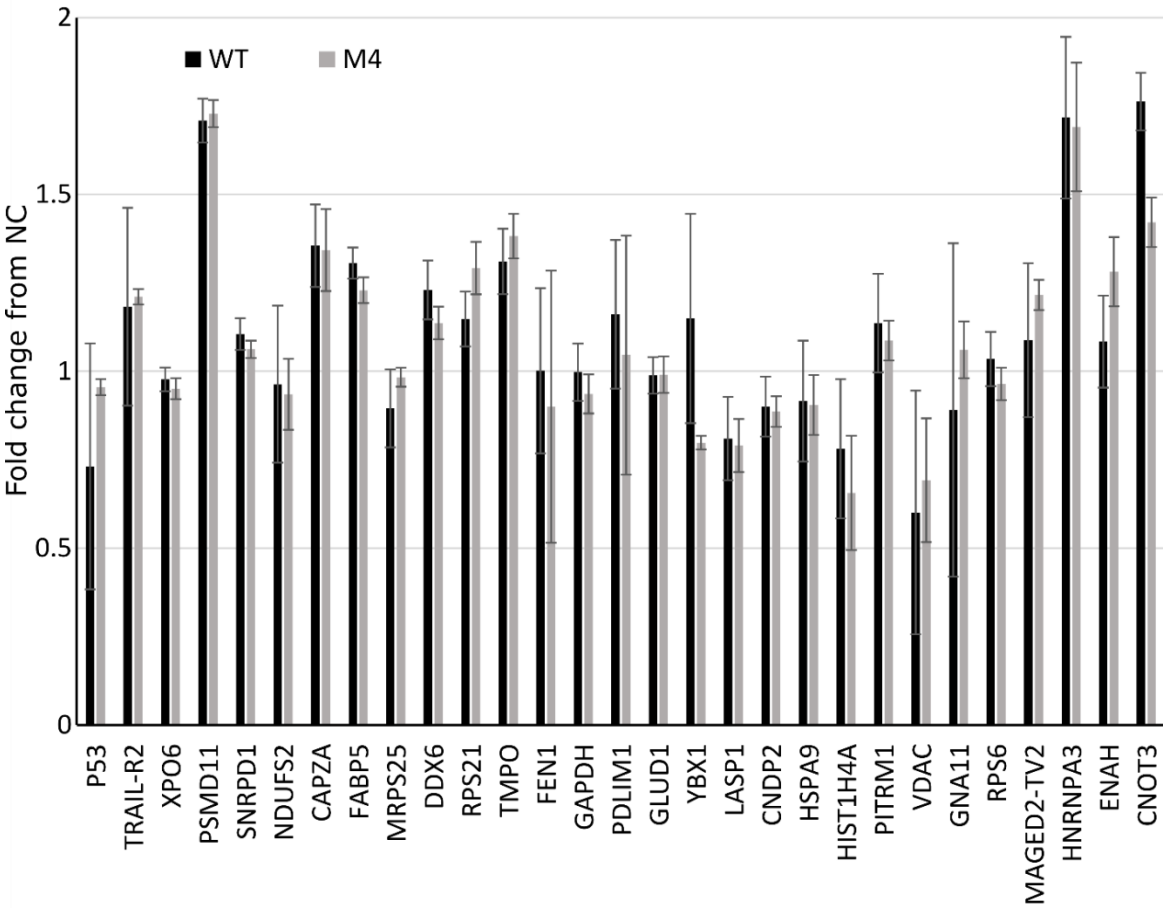


Figure 3-2. Effect of DDX21 status on mRNA levels of candidate proteins

RT-qPCR analysis comparing mRNA levels of target proteins between WT and M4 DDX21 recovery samples. Error bars represent the standard deviation between 3 biological replicates.

From the list of 26 candidate proteins, 20 (76.9%) were shown to have quadruplex forming potential in their mRNA by rG4-seq experiments[113]. Comparatively, of the 17,622 unique transcripts mapped in the rG4-seq experiment only 37.5% showed evidence of RNA rG4 formation. RIP-qPCR experiments using anti DDX21 antibodies showed significant enrichment (3 SD above GAPDH) for 6 of the candidate proteins' mRNA **(Figure 3-3)**. Of these 6 genes, 3

had predicted rG4s in the 5’UTR and 3 had predicted rG4s in the coding sequence (CDS). Interestingly, this distribution is quite the opposite of what was observed in the rG4-seq database in general, they report approximately 62% of the rG4s found to be in the 3’UTR and the remainder to be 16% 5’UTR and 20.6% CDS.

The large enrichment of proteins with rG4 containing mRNA in our list indicates there is good potential for direct regulatory effects to be discovered. The DDX21-IP allows us to further narrow our search by only focusing on the mRNA that are preferentially co-precipitated with DDX21.

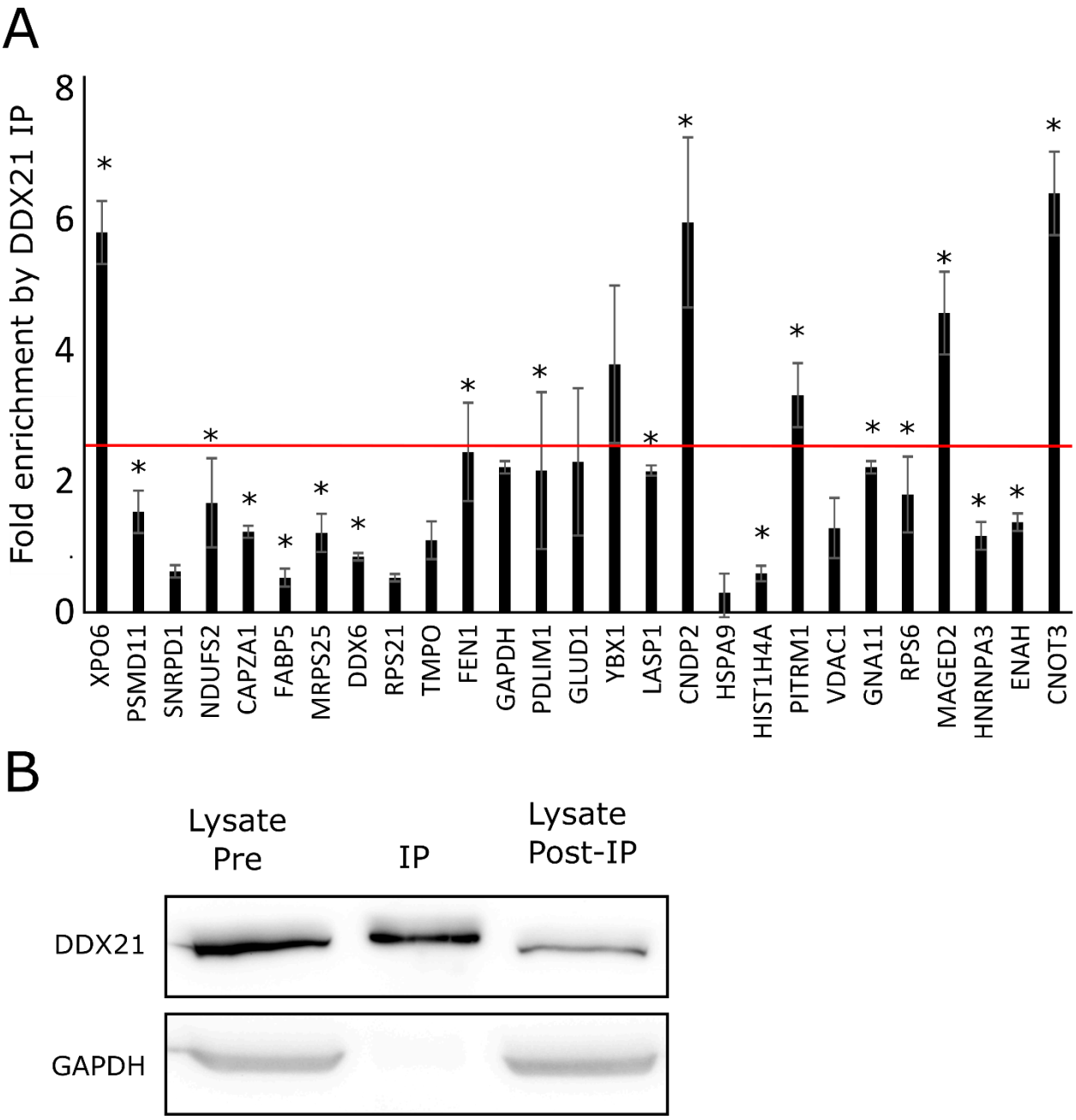


Figure 3-3. RNA immunoprecipitation of DDX21 followed by rt-qPCR of candidate genes

(A) RT-qPCR data for DDX21 RNA IP from HEK293T cell lysate comparing ct values of DDX21 IP sample to the RNA extracted from the pre-IP lysate. The red line represents a 3SD difference in enrichment above GAPDH. (B) Western blot showing DDX21 present in the lysate, IP, and to a lesser extent in the post-IP lysate.

3.4.2 Validating DDX21 targets by Western Blot

We chose to validate the 3 proteins with predicted rG4s in their 5'UTR because of the ease of integration into an existing assay for 5'UTR translational regulation. The two extremes from Figure 3-1, XPO6 and CNOT3 are such due to XPO6 being undetected in two of three WT DDX21 recovery MS samples and CNOT3 being undetected in all M4 DDX21 recovery MS samples. Western blots **(Figure 3-4)** for XPO6 show significant signal in the WT-DDX21 recovery sample and no apparent difference between the other samples. CNOT3 western blots show increased signal from the WT DDX21 samples compared to all other conditions, but no or marginal difference between the NC, KD or M4 samples. MAGED2 western blots have decreased signal in the KD and M4 samples compared to NC and WT, consistent with the effect observed by MS. Based on these results we chose to further investigate the role of DDX21 and it's rG4 binding domain in MAGED2 regulation.

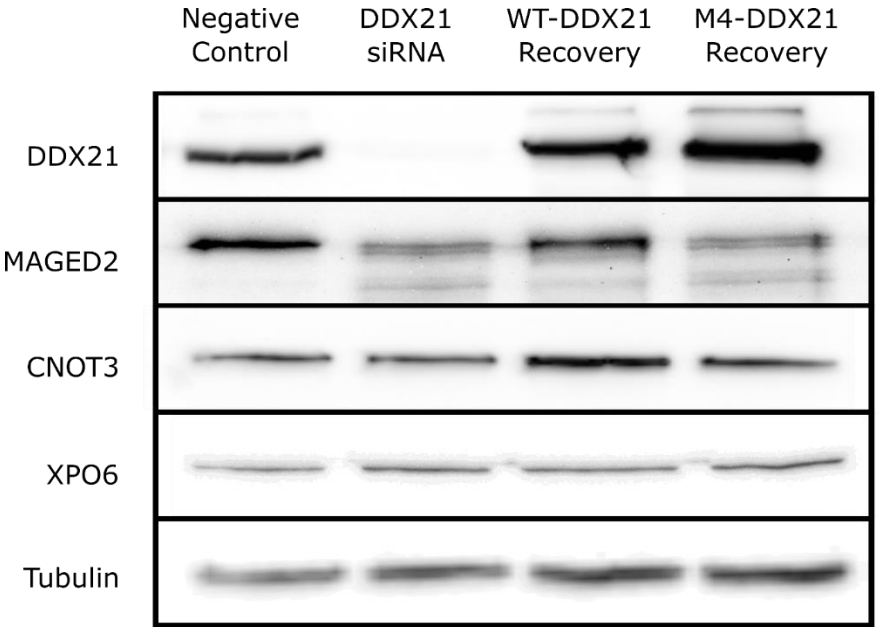


Figure 3-4. Western blots comparing protein levels between negative control, DDX21 knock-down and WT or M4 recovery samples

3.4.3 DDX21 regulates TRAIL-R2 expression and protects cells from TRAIL mediated apoptosis

MAGED2 has been shown to regulate TRAIL-R2 protein and mRNA levels in a p53 dependent manner[154,155,160]. To test whether the changes in MAGED2 level from DDX21 knock-down or M4 mutant could affect TRAIL-R2 we performed qPCR on RNA extracted from our knock-down recovery experiments. The RNA from the HEK293T cells showed no difference in mRNA levels of p53, MAGED2 or TRAIL-R2 **(Figure 3-2)**. Since this pathway has been shown to be p53 dependent, and HEK293T have p53 suppressed by the SV40 large T antigen[161], we repeated the experiment with MCF-7 cells that have been shown to have wild-type p53[162]. In MCF-7 cells we saw an increase in TRAIL-R2 mRNA levels upon DDX21 knock-down, but not p53 or MAGED2 **(Figure 3-5A)**. Furthermore, western blots confirm a change in protein levels of TRAIL-R2 in response to DDX21 activity **(Figure 3-5B)**. Next, we investigated the TRAIL sensitivity of our four conditions, 72 hours after knock-down of DDX21 MCF-7 cells were treated with TRAIL protein and left overnight. To detect TRAIL mediated apoptosis, we used fluorescent Annexin V to stain cells undergoing apoptosis and counted fluorescent cells by flow cytometry **(Figure 3-5C)**. The DDX21 Knock-down and M4-DDX21 recovery samples were significantly sensitized to TRAIL mediated apoptosis with ~20% of the counted cells showed staining by Annexin V. Comparatively the negative control treated cells and the WT-DDX21 recovery had only 0.19 ± 0.02% and 0.52 ± 0.18% of cells stained with Annexin V. These results demonstrate that regulation of MAGED2 by DDX21 is sufficient to affect the TRAIL-sensitization of cells by overexpression of TRAIL-R2.

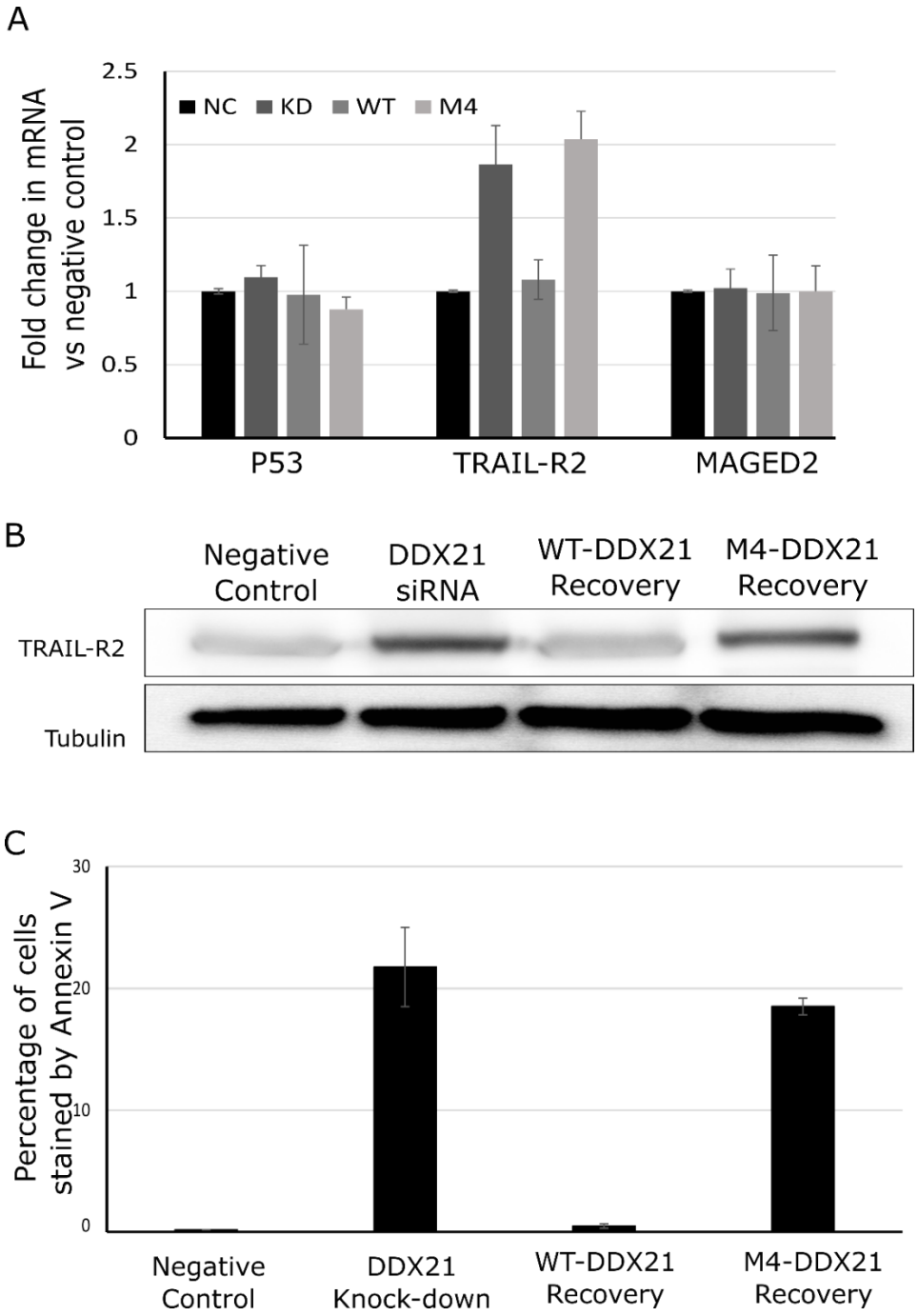


Figure 3-5. Validation of downstream targets of MAGED2

Assessing the effects of DDX21 knock-down, WT and M4 recovery on MCF-7 cells by: (A) RT-qPCR for MAGED2 mRNA and it's down stream targets TRAIL-R2 and P53. (B) Western blot for TRAIL-R2 levels. (C) Flow cytometry analysis of Annexin V staining post TRAIL treatment.

3.4.4 *MAGED2 TV2 is most enriched transcript variant by RIP-rt-qPCR*

The failure of the M4-DDX21 mutant to recover to the wild-type MAGED2 level is indicative of the involvement of rG4s in the regulation. MAGED2 has three alternative splice variants that result in different 5'UTR, but with no change to protein coding sequence. Though the variants share a conserved 31 nucleotides before the start codon, their 5' regions vary in length and sequence from 125nt for TV1, 233nt for TV2 and 193nt for TV3. All transcript variants have high rG4 forming potential, TV1 has 5 runs of 3 consecutive guanines, TV2 has 9 runs of 3 consecutive guanines and TV3 has 8 runs of 3 consecutive guanines. To decipher which of these transcript variants could be contributing to regulation of MAGED2 we first returned to our DDX21-RNA-IP samples and looked for differences in enrichment between primer sets specific for individual transcript variants **(Figure 3-6A)**. While a primer set that doesn't discriminate between MAGED2 transcript variants is enriched 4.6 fold, TV1 and TV3 were only enriched 1.7 and 2.2 fold, respectively. TV2 has the highest fold enrichment with a 5.5 fold change, indicating it is preferentially interacting with DDX21 in cell lysate.

3.4.5 MAGED2 5'UTR Luciferase Assays

As the strongest candidate for DDX21 interaction we site specifically mutated all 9 runs of guanines from TV2 so that there are never more than two consecutive guanines and called this construct, TV2m. TV2 and TV2m were cloned into the dual-luciferase vector, PsiCheck-2, immediately 5' of the hRLuc gene such that they would be transcribed as a 5'UTR. A common criticism of these assays is that the efficient transcription from the promotor on the luciferase vector results in levels of hRLuc mRNA that far exceed the levels of mRNA of the endogenous mRNA. In the context of G4 regulation, the increased number of G4s could exceed the capacity of the cellular G4 helicase proteins to unwind them[163]. In order to mitigate this effect we truncated the CMV promotor in front of our 5'UTR to reduce the level of reporter mRNA being made. HEK293T cells were treated as they were for the MS experiments and at 48 hours post endogenous DDX21 knock-down they were transfected with the dual-luciferase reporter vector.

In the experiments with the wild type TV2 5' UTR, DDX21 siRNA knock-down significantly ($P<0.05$) reduced the luciferase activity from the levels observed in the negative control cells **(Figure 3-6B)**. This effect was rescued by overexpression of the WT DDX21 but not M4. The luciferase activity from the M4 samples was also significantly reduced compared to NC and WT

samples, but not significantly different than the DDX21 KD samples. In the experiments with TV2m or an “empty vector” (no 5’ UTR added), no significant change in luciferase activity was detected between NC, KD, WT or M4 samples.

To begin probing whether the MAGED2 TVs form rG4, we introduced an RNA rG4 stabilizing ligand carboxypyridostatin (cPDS) into the luciferase assays. Cells that were otherwise untreated were treated with cPDS immediately prior to transfection of the reporter vector. Compared to the samples that were not treated with cPDS (or anything else), the treated samples had only 15% of the luciferase activity from the vector with the intact TV2 5’UTR. Whereas, the TV2m reporter construct showed no change in luciferase expression in the presence of cPDS. The empty vector samples were similarly unaffected by cPDS with the exception of one replicate having increased luciferase activity, adding significant error to the mean activity.

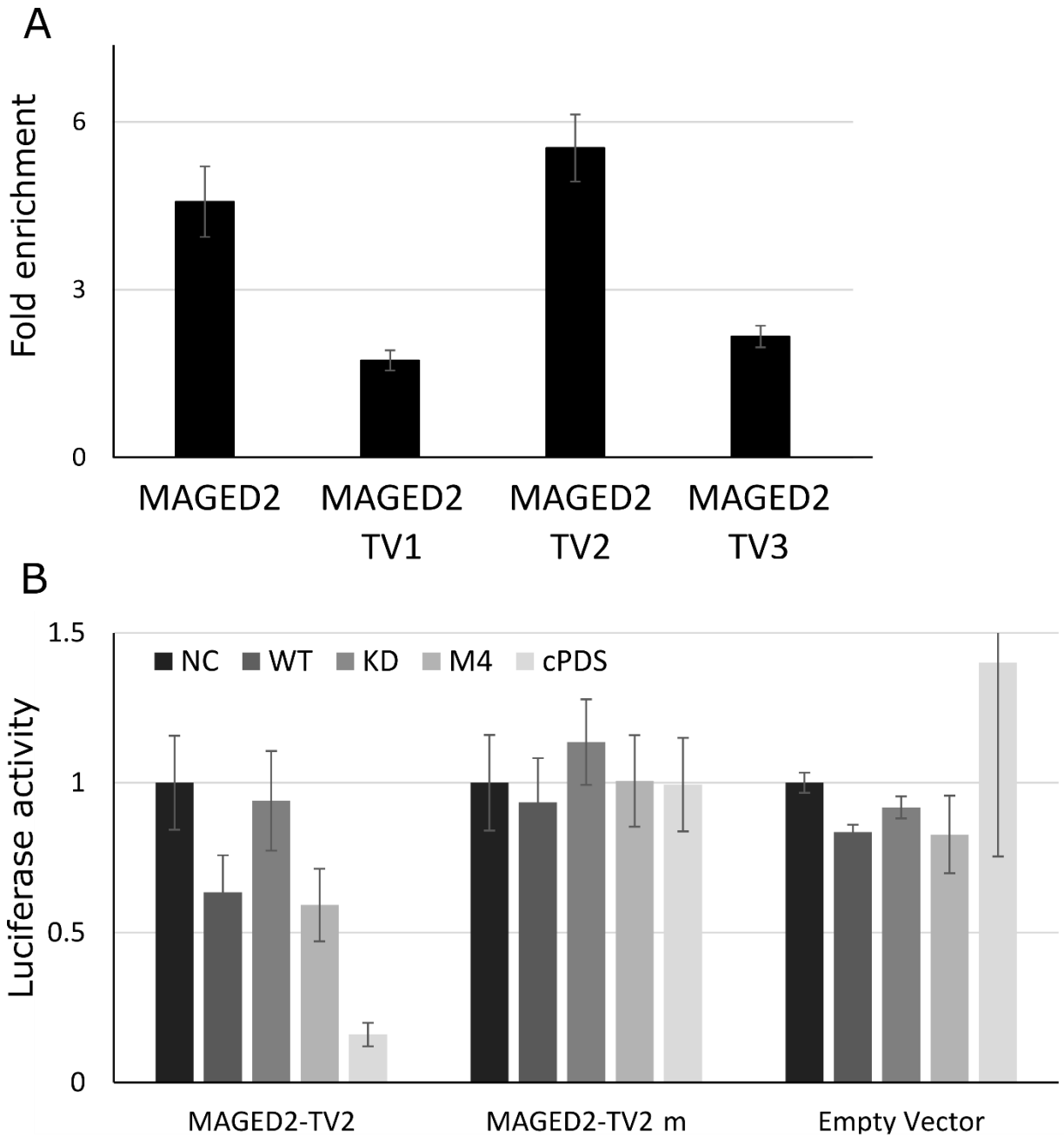
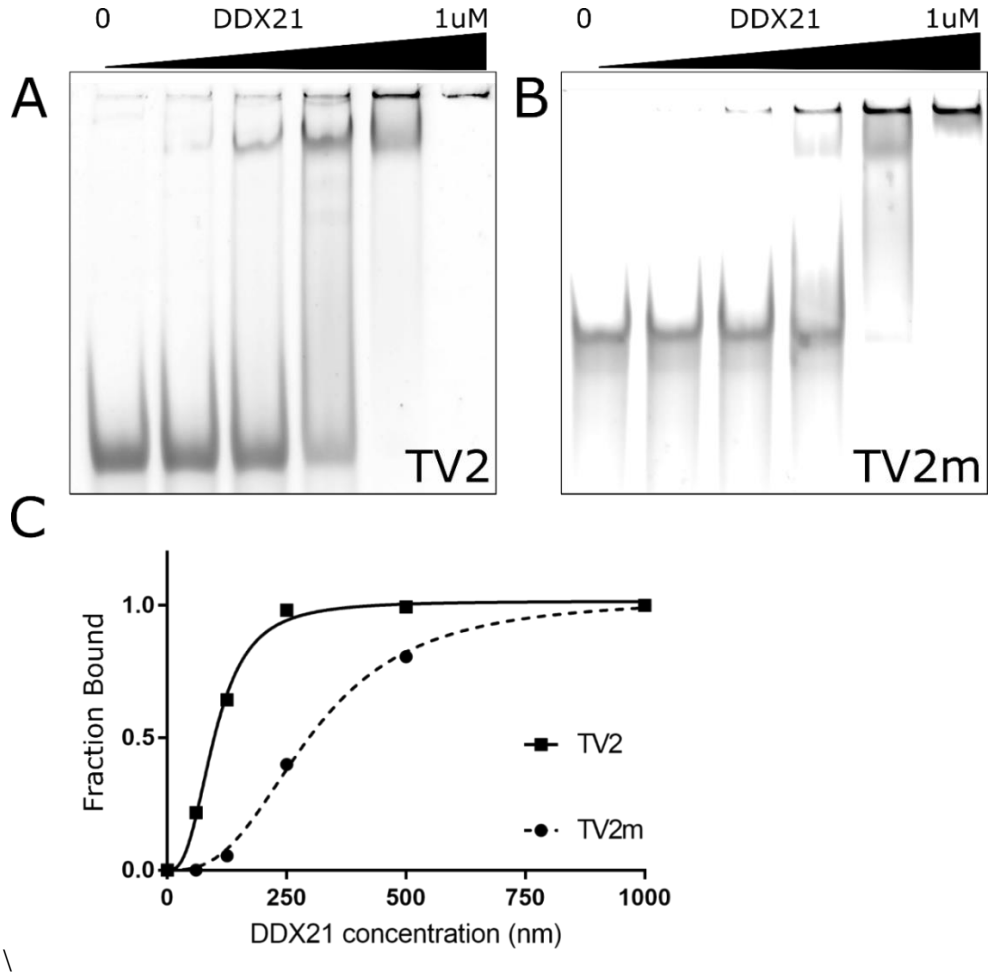


Figure 3-6. Enrichment of MAGED2 TV2 by DDX21 IP and effects on translation

(A) DDX21 RNA IP enriches MAGED2 TV2 preferentially to TV1 and TV3 by RT-qPCR. (B) Looking at the effect of DDX21 status and rG4 stabilizer cPDS on luciferase activity under control of MAGED2 TV2, TV2m and no 5' UTRs.

3.4.6 In Vitro evidence of rG4 formation

To further validate rG4 formation the TV2 and TV2m were *in vitro* transcribed from the PsiCheck-2 vector and purified by size exclusion chromatography. To confirm the possibility of direct interactions with DDX21 and to ascertain any possible specificity of DDX21, electrophoretic mobility shift assays (EMSA) were performed **(Figure 3-7A-B)** with recombinant purified DDX21 from *E.coli*. At the highest DDX21 concentration (1000nM) both RNA were completely shifted to higher molecular weight complexes. Densitometric analysis using the FluorChem Q software allowed for fitting of the EMSA to a binding model (Table 1). Recombinant purified DDX21 showed marginally higher affinity for TV2 vs TV2m (100nm vs 300nm)



\

Figure 3-7. Electrophoretic mobility shift assays between DDX21 and TV2/TV2m

EMSA of MAGED2 5’ UTRs, (A) TV2, (B) TV2m with a serial (1:1) dilution of DDX21. Fitting curves (C) determined in Prism from a Specific binding model with Hill slope from densitometric analysis of EMSA images.

RNA	TV2	TV2m
Kd(nM)	99.4 ± 4.2	302.2 ± 13.0
Hill coefficient	2.8 ± 0.3	2.79 ± 0.26

Table 3-1. Densitometric analysis of EMSA

Scaled to fraction bound 100% by 1uM DDX21. Fitting parameters from One Site specific binding model with Hill slope in Prism GraphPad 7.0.

TV2 and TV2m were assayed for rG4 formation by Thioflavin T (ThT), a fluorescent probe for RNA rG4 formation[128,164]. When ThT is bound to rG4 it exhibits enhanced fluorescence. rG4 conformations are stabilized by monovalent cations, with K+ being the best at stabilizing rG4 and Li+ having destabilizing effects, but alternate conformations of RNA are not greatly affected by these cations[116]. By comparing ThT fluorescence in the presence of rG4 stabilizing and destabilizing cations we strengthen the evidence that enhanced fluorescence is due to rG4 formation. Of the transcript variants TV2 induces the most fluorescence from **ThT (Figure 3-8A)**, the large difference in the fluorescence from the K+ and Li+ containing samples is indicative of rG4 formation being the underlying cause for the fluorescent enhancement. At lower RNA concentrations TV2 has the greatest difference in ThT fluorescence between K+ and Li+ samples but at higher RNA concentrations the difference in fluorescence tapers off, indicating the RNA may be forming rG4 at higher concentrations despite the unfavourable Li+ cation. TV2m has considerably less effect on ThT fluorescence, typically emitting 10-30% the amount of fluorescence compared to TV2 at a given concentration; together with a lack of significant difference between TV2m in K+ or Li+ containing samples this indicates that TV2m does not form a rG4.

Finally, we used a modified reverse-transcriptase stalling assay to demonstrate structural differences between TV2 and TV2m. Briefly, first strand cDNA synthesis was performed on the *in vitro* purified RNA at 37°C for 20 minutes. After purification of cDNA, qPCR was used to amplify the cDNA. We used two primer sets to probe the efficiency of reverse transcription of

TV2 and TV2m RNA, one primer set spanned the entire RNA and the other primer set only amplified the 3’ region of the RNA after the last run of 3 consecutive guanines. On average, the full length product was detected in the TV2m sample 8-10 cycles sooner than TV2 **(Figure 3-8B)**, indicating a huge difference in efficiency. This effect was confirmed by visualizing the PCR products on an agarose gel **(Figure 3-8C)**, both RNA yielded a single band of consistent size and the signal intensity from products on the gel corelates with our qPCR ct values. The control primer set, that does not span a run of more than 3 guanines, showed no significant difference between the CT values for TV2 and TV2m. This confirms that the low efficiency of RT-qPCR for TV2 is due to a structured region of the RNA and not a difference in the amount of RNA input.

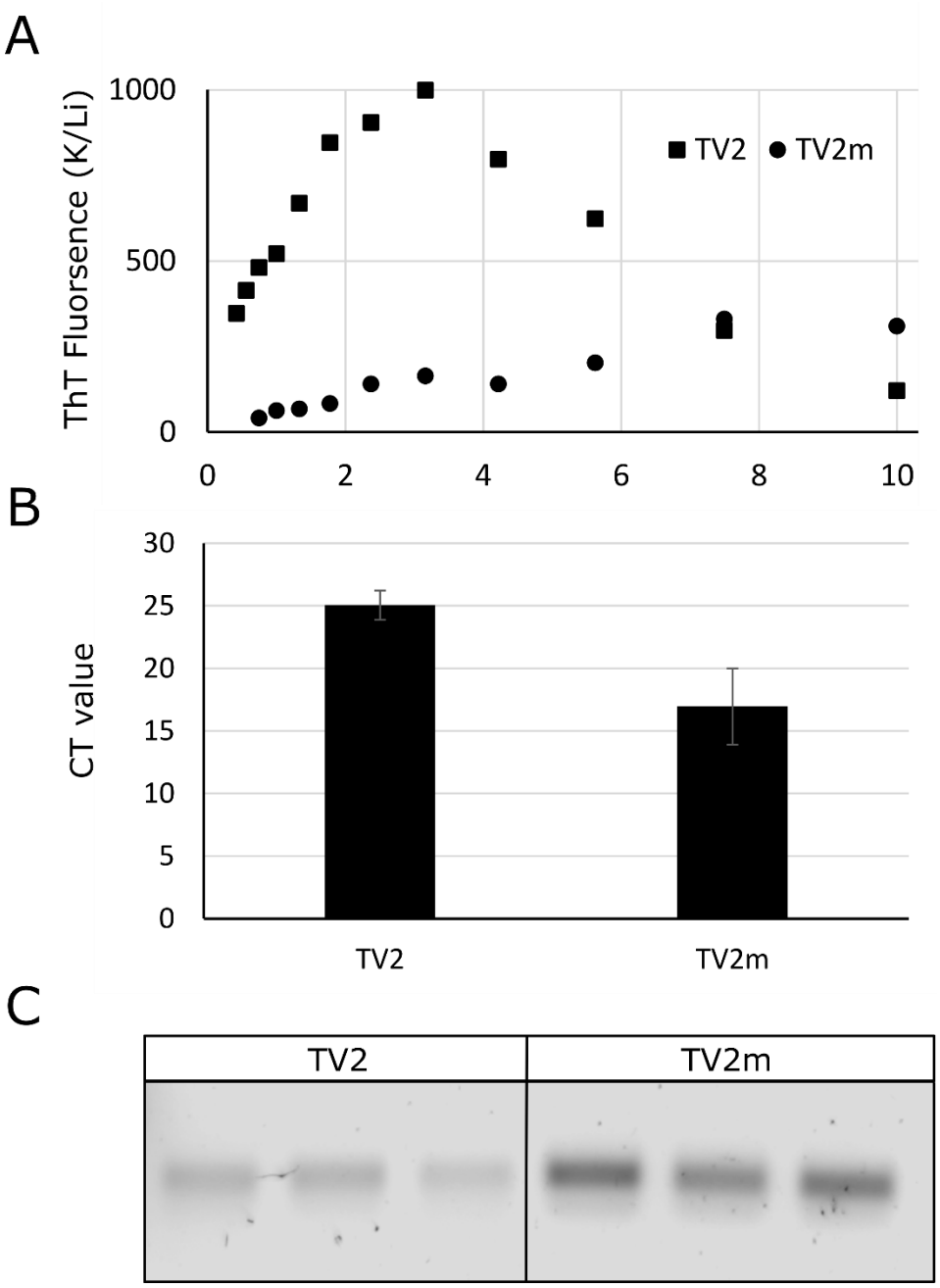


Figure 3-8. In vitro evidence for G4 formation in TV2

In vitro evidence of rG4 formation. (A) Fluorescence enhancement of ThT in the presence of TV2 or TV2m, data represents the ratio of ThT fluorsence in TEK vs TELi buffers. Reverse-transcriptase stalling assay for TV2 and TV2m show differences in CT values by qPCR (B) and confirmed by visualization of the PCR product on an agarose gel (C).

5. Discussion

Recently we discovered DDX21 to be an RNA rG4 binding protein that has the ability to destabilize rG4s[147]. Until now our studies have focused on model RNA systems and there are no rG4s that DDX21 is known to interact with endogenously. Proteomic mass spectrometry guided us to 26 candidate genes that were differentially expressed in WT and M4 DDX21 samples. Gene ontology revealed no immediately apparent pathways that were affected. From this list we validated MAGED2 by western blot, a 2-fold decrease in the M4 DDX21 sample by MS is observed. MAGED2 has been shown to interact with p53 and affect transcription and protein levels of TRAIL-R2. It has been hypothesized that increased MAGED2 expression is a mechanism to selectively inactivate certain P53 targets (i.e. TRAIL-R2). Previous studies have shown that a decrease in MAGED2 protein results in upregulation of P53 and TRAIL-R2 at the mRNA and protein level. [154,160]. We observed a similar effect in MCF-7 cells, revealing a pathway whereby WT DDX21 can facilitate MAGED2 translation which in turn blocks TRAIL-R2 expression. Interestingly, replacement of the WT DDX21 for the M4 mutant results in decreased MAGED2 protein and in turn an increase in TRAIL-R2 expression, indicating that G4 binding and unwinding is important for efficient translation of MAGED2 mRNA. This increase in TRAIL-R2 expression is significant enough to allow for TRAIL induced apoptosis in MCF-7 cells depleted of DDX21 or expressing only the M4 DDX21.

After determining that DDX21 preferentially finds the TV2 5'UTR, by DDX21-IP and rt-qPCR, we investigated a mutant version of TV2 in which all runs of GGG had been changed to GCG (TV2m). The TV2m 5'UTR was not impacted by DDX21 knock-down or recovery, whereas TV2 had decrease luciferase activity with depleted DDX21, M4 DDX21 and in the presence of rG4 stabilizing dye cPDS. Furthermore, *in vitro* transcribed and purified TV2m shows reduced affinity for DDX21 by EMSA and does not demonstrate the K+ sensitive ThT fluorescence response that TV2 does. Together these data provide strong evidence that DDX21 can regulate MAGED2 translation via a rG4 in its' 5'UTR.

MAGE proteins are a diverse set of proteins that contain the MAGE homology domain (MHD), the majority of the MAGE family of proteins are tumor antigens[165]. An interesting exception to this are the MAGED family, these proteins are expressed ubiquitously and contain none of the 20 antigenic peptides known to be recognized by T-lymphocytes[166]. MAGED proteins are

known to regulate cell cycle progression, apoptosis and transcription and are up-regulated in key points of embryogenesis as well as in multiple cancer types[167,168]. Over-expression of MAGED2 has specifically been linked to resistance to treatment with TRAIL[154], making it an interesting and valuable target to validate for rG4 regulation by DDX21.

Targeting of rG4 regulated pathways is an emerging area of therapeutic potential for fighting cancer[169,170] and sensitization of TRAIL-resistant breast carcinoma to TRAIL mediated apoptosis is a current therapeutic target[156,158,159]. Our data underline the potential for the integration of rG4 therapeutics with an existing cancer treatment interest. Current research suggests that DDX21 has a unique specificity for certain rG4s that may be dependent on loop sequence[147,148]. Further work aims to identify the structure of the TV2 rG4 that DDX21 is recognizing, deciphering the specificity of DDX21 may potentially allow for the design of small molecules that specifically block its action.

6. Materials and Methods

3.6.1 Cell Culture and Reagents

The HEK293T cell line was a gift from Dr. Thomas Klonisch and the MCF-7 cell line was a gift from Dr. Spencer Gibson. Both cell lines were grown in Dulbecco's Modified Eagle's Medium from Thermo Fisher Scientific (Ottawa, Canada) supplemented with 10% Fetal Bovine Serum (Thermo Fisher Scientific). Gibco Trypsin/EDTA solution (Thermo Fisher Scientific) was used to detach the MCF-7 cells. Polyethyleneimine (Thermo Fisher Scientific) was used as the transfection reagent for vector DNA transfections

3.6.2 DDX21 siRNA knock-down and recovery

Polyethyleneimine (Thermo Fisher Scientific) was used as the transfection reagent for vector DNA transfections. HEK293T cells were grown to 80% confluence in 150mm cell culture plates and transfected with empty pcDNA 3.1+ (negative control and knock-down samples) or siRNA resistant wt-DDX21 or m4-DDX21 expressing pcDNA3.1+. The following day the cells were split 1:9 and reverse transfected using transfectamine siRNA Max with either negative control siRNA (negative control sample) or DDX21 siRNA (Knock down and wt and m4 recovery samples). 72 hours after reverse transfection the cells were collected in Phosphate Buffered Saline (Thermo Fisher Scientific) and rinsed twice. MCF-7 cells were treated the same way but cultured in 6-well dishes and only split 1:3 when reverse transfecting the siRNA. The remainder were lysed in ice cold RIPA buffer supplemented with 100x HALT proteinase cocktail (Thermo Fisher Scientific). Cells were vortexed and sonicated briefly before pelleting cell debris by centrifugation at 21,000 g. The protein concentration of the soluble portion was determined by Bradford assay and 50 µg of each sample was used for SDS-PAGE and Western blots.

3.6.3 In gel digestion

Protein lanes were visualized by Coomassie Blue staining prior to whole-lane excision. Each lane was subsequently cut into 15 equal bands, with each band corresponding to a region of the gel containing proteins of a distinct molecular weight range. Each of the gel fractions was subjected to in-gel tryptic digestion as previously described[171]; the resulting extracted peptides were then dried and suspended in 60 L of 0.2% formic acid in 5% ACN.

3.6.4 LC-MS/MS Analysis

Digested peptides were analyzed by LC-MS/MS using a ThermoScientific Easy nLC-1000 in tandem with a Q-Exactive Orbitrap mass spectrometer. Five microliters of each sample was subject to a 120 min gradient (0–45% buffer B; buffer A: 0.2% formic acid; buffer B: 0.2% formic acid in ACN) on a 2 cm Acclaim 100 PepMap Nanoviper C18 trapping column in tandem with a New Objective PicoChip reverse-phase analytical LC column. For data-dependent analysis, the top 15 most abundant ions were analyzed for MS/MS analysis while +1 ions were excluded from MS/MS analysis. Additionally, a dynamic exclusion of 30 s was applied to prevent continued reanalysis of abundant peptides. For the analysis, a resolution of 60 000 was used for full scans that ranged from 350 to 2000 m/z and a resolution of 30 000 was used for MS/MS analysis.

For data analysis, raw data files corresponding to samples comprising an entire gel lane were grouped together and searched using Proteome Discoverer 1.4.1.14's SEQUEST search algorithm using the reviewed, nonredundant *H. sapien* complete proteome retrieved from UniprotKB. Search parameters were identical to as previously reported[172]. During data processing, the "Precursor Ion Area Detector" node of Proteome Discoverer 1.4.1.14's SEQUEST workflow editor was implemented to quantify the extracted ion chromatogram for each protein identified from the raw data. Searched results were filtered using a minimum of two medium confidence peptides per protein.

3.6.5 Protein-RNA cop-immunoprecipitation and RT-qPCR

Native RNA immunoprecipitations were performed on one 150 mm dish of HEK293T cells as previously described[173] using anti-DDX21 antibody (ProteinTech 15252-1-AP). RT-qPCR analysis was performed on Applied Biosystems StepOnePlus instrument with the RNA to Ct One-step RT-qPCR kit (Thermo Fisher Scientific) according to manufacturers instructions. To determine fold enrichment, RNA extracted from the total cell lysate using the GeneJET RNA purification kit (Thermo Fisher Scientific) was used to IP to serve as a reference sample. 25 ng of template RNA was used in all RT-qPCR reactions.

3.6.6 Western blots

For each sample, 50 µg of protein in 1X SDS-PAGE load dye was loaded onto a 10% SDS-PAGE gel and ran at a continuous voltage (200V) for 45 minutes. Proteins were transferred to a 0.2 µm PVDF membrane (BioRad Laboratories) using the Trans-Blot Turbo transfer system (BioRad Laboratories). Membranes were blocked for 30 minutes in 5% milk powder dissolved in Tris Buffered Saline + Tween (TBST) (20 mM Tris pH 7.5, 150 mM NaCl, 0.1% Tween). The following antibodies were used: DDX21 (Novus NB100-1716), MAGED2 (ProteinTech 15252-1-AP), TRAIL-R2 (AbCam EPR19310, AB199357), CNOT3 (ProteinTech 11135-1-AP), XPO6 (ProteinTech 11408-1-AP), gapdh, tubulin, goat anti rabbit secondary, goat anti mouse secondary. DDX21, MAGED2 and TRAIL-R2 primary antibodies were incubated overnight at 4 degrees Celsius in TBST+ 5% milk, GAPDH and tubulin primary antibodies were incubated for 1 hour at room temperature in TBST+5% milk. Three sequential 5 minute washes with 5% TBS-T were followed by 1 hour incubation with secondary antibodies (1:10000) in TBST+5% milk and then 4 more 10 minute washes with TBS-T. Prior to imaging with the FluorChem Q system (Protein Simple), 1 ml of Illuminata Forte Western HRP substrate (EMD Millipore) was added to the blots. Due to weak signal from some antibodies a uniform adjustment to the brilliance was made using Inkscape V 0.91.

3.6.7 Annexin V staining & TRAIL treatment

MCF-7 cells were treated as described in the DDX21 siRNA knock-down and recovery section. At 48 hours post knock-down TRAIL protein (SOURCE?) was added to the media to a final concentration of 50 ng/ml. Approximately 18 hours later the cells were washed with 1x trypsin and then incubated with 1x trypsin at 37 °C for 5 minutes. Detached cells were collected, washed with cold PBS and resuspended in 100 µl of cold 1X Annexin-binding buffer supplemented with AlexaFluor 488 annexin V (Thermo Fisher Scientific) and incubated in the dark for 15 minutes. Samples were then diluted to 500 µl with 1X Annexin-binding buffer and kept on ice. Annexin V binding was measured using the BD FACSCalibur platform and analyzed using the BD CellQuest Pro software (BD Biosciences, San Jose, CA), 10,000 cells from each sample were analyzed and each sample performed in biological triplicate.

3.6.8 Luciferase assays

For luciferase assays, HEK293T cells were treated as described in the DDX21 siRNA knock-down and recovery section scaled down to a 6 well dish sample size. For carboxypyridostatin treated cells, cPDS was added to a concentration of 1uM in the DMEM at 48 hours post knock-down. Then the cells were transfected with 1ug of psiCheck-2 vector in 250 µl serum free DMEM with 1.5 µg of polyethyleneimine. After 24 hours the cells were collected in cold PBS. Half of the cells were used for RNA purification using the GeneJET RNA purification kit (Thermo Fisher Scientific) and the other half used in the luciferase assay. The Dual-Glo Luciferase Assay system (Promega, Madison Wi, USA) was used according to manufacturer instructions, luminescence was measured for 10 seconds using the (Pelka's instrument) and each sample performed in at least biological triplicate. To control for transfection efficiency, luminescence from the hRluc protein, under control of the 5'UTR cloned into the NheI site was normalized to the signal from the hluc protein for each sample. Each sample was then normalized to the average value of the negative control samples.

3.6.9 Cloning

The pcDNA3 vector was used to create a CMV promotor lacking nucleotides between 208 and 746, this was then cloned into the BglII and NheI sites of the psiCHECK-2 vector, removing the SV40 promotor. 5'UTRs of MAGED2 TV2 and TV3 were amplified by RT-PCR from RNA isolated from human testes tissue purchased from Takara Bio (Mountain View, CA). The 5'UTR of TV1 was amplified from a DNA fragment purchased from Genscript. Each TV was amplified using primers that contain restriction enzyme sites for NheI and cloned into the psiCHECK-2 vector. The cDNA that encodes the human DDX21 protein, isoform 1, was amplified by PCR using DNA primers that were designed to encode an NheI and NotI restriction enzyme sites and cloned into the pET28b expression vector.

3.6.10 In vitro transcription and protein purification

The 5'UTRs of each transcript variant were transcribed and purified *in vitro* as previously described (Booy, et al., 2012) from linearized recombinant psiCHECK2 vector (Promega) that contained one of the three UTR sequences downstream of a T7 promoter. The vector was linearized with NcoI restriction enzyme and purified by phenol-chloroform extraction. *In vitro* transcription was carried out at 37°C for 3 hours. The transcription reaction was then stopped by

addition of EDTA and T7 polymerase was removed by phenol-chloroform extraction. Residual phenol was removed using a DG10 desalting column (BioRad), from which the RNA was eluted in 20 mM Tris pH 7.5, 1 mM EDTA and either 100 mM KCl (TEK) or 100 mM LiCl (TELi). The RNA was then separated from the plasmid and remaining NTPs by FPLC purification using a Superdex 200 filtration column (GE Healthcare Life Sciences) using TEK or TELi buffer. FPLC fractions were then concentrated and stored at 4°C. RNAs were heated at 95°C and cooled on ice prior to use.

DDX21 was purified from LOBSTR *E. coli* BL21 (DE3) (Kerafast Inc, Boston, MA) cells transformed with DDX21 expressing pET28b. The stop codon from DDX21 cDNA was retained so that the recombinant protein only had the N-terminal 6His tag and thrombin cleavage site and not the C-terminal 6His tag. Cells were grown at 37°C in a shaker incubator to an optical density of 0.4 at A_{600}, then induced with 0.3 mM IPTG and grown overnight at 18^{0}C. Cells were pelleted, then resuspended in 20 mM Tris 7.5, 1 M NaCl, 10% glycerol, 1 mM DTT and 1 mM PMSF, and lysed by sonication. After centrifugation of lysate, recombinant DDX21 was purified from the soluble fraction using HisPur Cobalt Resin (Thermo-Fisher) and eluted using 200 mM imidazole. Elution fractions were dialyzed against 20 mM Tris pH 7.5, 300 mM NaCl, 10% glycerol, 2 mM DTT, and stored at 4^{0}C.

3.6.11 Electrophoretic Mobility Shift Assays (EMSA)

DDX21, or an equal volume of protein storage buffer (20 mM Tris pH 7.5, 300 mM NaCl, 10% glycerol, 2 mM DTT) was added to the binding reactions in a 1:1 serial dilution, with the first condition containing no DDX21. Binding reactions were performed in 50 mM Tris-acetate, pH 7.8, 100 mM KCl, 10 mM NaCl, 3 mM $MgCl_2$, 70 mM glycine, 10% glycerol, 0.05 mg/mL bovine serum albumin (BSA) for 10 min at room temperature and resolved by native Tris-borate EDTA (TBE) 8% polyacrylamide gel electrophoresis (TBE-PAGE) for 150 minutes at 75V. The RNAs were then stained with the fluorescent nucleic acid dye SYBR Gold and imaged on a Fluorchem Q imaging system using the Cy2 excitation LEDs and emission filters (Protein Simple, San Jose CA).

3.6.12 Thioflavin T assays

In a 96-well plate, the *in vitro* transcribed RNA was serially diluted 3:1 with 20 mM Tris pH 7.5, 1 mM EDTA and either 100mM KCl or 100 mM LiCl, from 10 µM to 0.4 µM. Thioflavin T was added with a final concentration of 1.5 µM. Samples incubated for 5 minutes at room temperature and fluorescence was measured with Applied Biosystems Step-One-Plus qPCR machine. Fluorescence data was extracted from the BLUE channel from the raw data output of the Step-One-Plus.

3.6.13 RT-Stop Assay

First strand synthesis of cDNA from 50 ng of *in vitro* transcribed 5'UTR of TV2 and TV2m was performed in a 20 µL reaction using 1 µM reverse primer, 500 µM dNTPs, 4 µL 5X RT Buffer (250 mM Tris-HCl (pH 8.3 at 25 °C), 375 mM KCl, 15 mM MgCl2, 50 mM DTT) (Thermo Fisher), and 200 U Maxima H-Minus RT (Thermo-Fisher). Reactions were performed at 37°C for 20 minutes before heating at 85°C for 5 minutes to inactivate the RT. Forward primer was then added to the reaction at a concentration of 1 µM. The cDNA reaction mixture was diluted 1:10 in nuclease free water, 10 µL of which was mixed with 10 µL of PowerUp Sybr Green Master Mix (Applied Biosystems). 40 cycles of qPCR were performed with 60°C annealing temperature and 1-minute extension time at 72°C using a Step-One-Plus qPCR machine (Applied Biosystems). C_T values were determined using StepOne software. The qPCR products were run on a 1.5% agarose gel stained with Sybr Safe DNA stain (Thermo-Fisher) and imaged with Fluorochem Q imager (Protein Simple, San Jose CA).

Chapter 4: Insights into the RNA quadruplex binding specificity of DDX21

4.1.1 Preface

The work presented in Chapter 4 was completed in while our collaborators were working on the mass spectrometry data collection and analysis that formed the crux of Chapter 3. In this chapter we depart from cellular and molecular biology and perform a detailed biophysical characterization of the interaction between the G4 binding domain of DDX21 and a model RNA G4. Once again, this work builds on the data from Chapter 2, using the information gained about the G4 binding domain boundaries to focus on a minimal interacting portion of DDX21 that can be expressed at high yield, suitable for these sample intensive biophysical experiments. Tangential to finding biological RNA G4:DDX21 interactions, this work focuses on understanding how DDX21 interacts with G4.

This 2018 Biochemica Biophysica Acta publication, presented here in full:

McRae, E. K. S., Davidson, D. E., Dupas, S. J. & McKenna, S. A. (2018) Insights into the RNA quadruplex binding specificity of DDX21. Biochim. Biophys. Acta - Gen. Subj., 1862

provides key insights into the specificity of DDX21 for certain G4 and provides the basis integration of DDX21 into our structural biology pipeline. Previous work had showed that DDX21 may interact with a well characterized G4, TERRA. We confirm that DDX21 can interact with the TERRA G4 and by control variables such as temperature, ionic strength and chemical composition of the loop nucleotides we tease out a mode of interaction that is supported by NMR experiments.

4.1.2 Contribution of Authors

David E. Davidson assisted in collection and processing of NMR data. Steven J. Dupas performed reviewer requested ITC, MST and CD experiments. The remainder of the experimental work and writing was completed by Ewan K.S. McRae, with editing and guidance from all the listed authors.

2. Abstract

Guanine quadruplexes can form in both DNA and RNA and influence many biological processes through various protein interactions. The DEAD-box RNA helicase protein DDX21 has been shown to bind and remodel RNA quadruplexes but little is known about its specificity for different quadruplex species. Previous reports have suggested DDX21 may interact with telomeric repeat containing RNA quadruplex (TERRA), an integral component of the telomere that contributes to telomeric heterochromatin formation and telomere length regulation. Here we report that the C-terminus of DDX21 specifically binds to TERRA. We use, for the first time, 2D saturation transfer difference NMR to map the protein binding site on a ribonucleic acid species and show that the quadruplex binding domain of DDX21 interacts primarily with the phosphoribose backbone of quadruplexes. Furthermore, by mutating the 2'OH of loop nucleotides we can drastically reduce DDX21's affinity for quadruplex, indicating that the recognition of quadruplex and specificity for TERRA is mediated by interactions with the 2'OH of loop nucleotides.

3. Introduction

Guanine quadruplexes (G4) are stable four-stranded structures of nucleic acids that form in regions rich in guanine such as human telomeres. Mammalian telomeres comprise of thousands of tandem repeats of the sequence d(TTAGGG), terminated by one to two hundred repeats of just the G-rich sequence, in complex with a number of proteins that serve to protect the telomeres from damage and facilitate replication during cell division[174–176]. The 3' terminal overhang of the human telomere forms an intramolecular (3+1) G4 under physiological conditions, with three of the strands sharing the same polarity and one being anti-parallel to them[177] **(Figure 4-1)**. The telomeric repeat containing RNA (TERRA) is a variable length transcript of the telomeric sequence r(UUAGGG) and forms a propeller loop type parallel G4[178],[179] **(Figure 4-1)**. In concert with its DNA counterpart and telomeric repeat binding proteins, TERRA protects the telomere end[180,181] and regulates telomere length and histone modification[39,182–184].

Along with many other proteins, the DEAD-box RNA helicase protein DDX21 can be co-precipitated with TERRA from nuclear cell extracts[185]. Furthermore DDX21 interacts with telomere maintenance protein WRN, an interaction that is insensitive to RNase A treatment[186].

Recently, we have shown that DDX21 can directly bind and remodel RNA G4 using a domain in its C terminal 209 amino acids (C209)[43]. It is therefore plausible that DDX21 can interact with both the DNA and RNA telomere G4s. DDX21 overexpression has been observed in multiple human cancers[187] and significantly associated with early death and a shorter disease free survival in breast cancer patients[124], making it a potential gene that could be exploited as a cancer drug target. Moreover, disrupting G4-protein interactions in general is an attractive therapeutic approach that has already been met with some success[18]. Despite multiple studies that have been undertaken, the structural basis for specific G4 recognition by proteins is not yet fully understood. The understanding of the molecular basis of G4 recognition is crucial for the development of new therapeutics that inhibit protein binding to G4s as well as our general understanding of the specificity of proteins for individual G4s.

In this report, we tested the affinity of DDX21 for the human telomere DNA G4 (hTel) and TERRA and found DDX21 to have a 20-fold greater affinity for the RNA G4. By comparing the affinities for TERRA with a parallel G4 (c-myc) and a version of TERRA with deoxy-ribose loops (dloop-TERRA) we attribute the higher affinity to interactions with the 2' OH of the loops of the G4. We further deconstruct the thermodynamics of the interaction with TERRA using isothermal titration calorimetry (ITC), which indicate that binding is primarily driven by coulombic interactions. Finally, we use 2D saturation transfer difference (2D-STD) experiments to probe the proximity of DDX21 to different regions of the G4 TERRA.

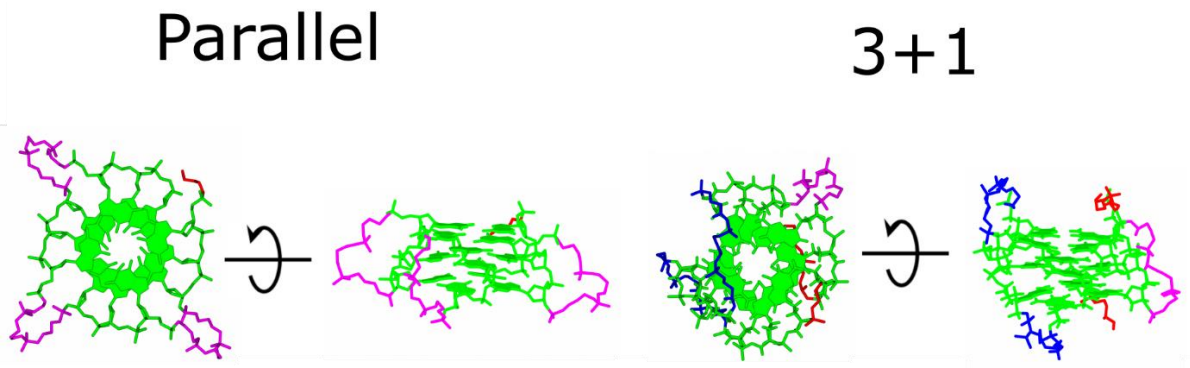


Figure 4-1: Schematic structures comparing G4s of parallel and (3+1) topologies.

The G-quadruplex core is shown in green, propeller loops in pink, lateral loops in blue and extraneous nucleotides in red.

4. Results

4.4.1 DDX21 has specificity for RNA G4 that is driven by 2'OH in the loop sequence

To investigate the specificity of DDX21 for telomeric G4s we used a 22-nucleotide truncation of the human telomere G4 (hTel) and the same length version of its RNA counterpart (TERRA). For the purposes of microscale thermophoresis (MST) measurements the nucleic acids were synthesized with a 5' Cy5 fluorophore. To confirm their folded structure, we performed circular dichroism on each nucleic acid we used **(Figure 4-7)**. As expected hTel shows the positive peak at 295nm, negative peak at 265nm and slightly smaller positive peak at 245nm characteristic of an anti-parallel stranded G4[188], while TERRA exhibits the positive 265nm and negative 240nm peaks of a parallel G4[178].

We purified, from *E.coli,* the recombinantly expressed C-terminal G4 binding domain of DDX21 (C209) and used MST to monitor complex formation between C209 and fluorescently labelled G4s. The thermophoresis data was fit to the Hill equation to determine the apparent dissociation constant (K_d) and Hill coefficient (H). All G4s tested had a H value > 1, indicating cooperative binding of multiple C209 molecules to a G4. Similar to other RNA G4 previously tested by this method[43], C209 has a K_d in the nanomolar range (346$\pm$17nM) for TERRA. Strikingly we observed a 20-fold higher K_d for the hTel **(Figure 4-2A)**. To investigate whether this drastic difference in affinity was due to the different strand topology we utilized a DNA G4 from the promotor region of the c-myc gene (c-myc) that has been shown to form a parallel G4[189]. The c-myc G4 demonstrated a lower K_d than hTel (4.52$\pm$0.22µM vs. 7.28$\pm$0.27µM) but not on the same order of magnitude as TERRA.

To determine whether the 2'OH present in the RNA ribose sugar, but not in the DNA, is responsible for the higher affinity we used a version of TERRA that has 2'OH groups on the guanines that form the G4 core but not the uracils or adenines in the loop sequence that connect the four strands of the G4 (dloop-TERRA). This forces the nucleic acid into the parallel G4 topology that has the same loop sequence as TERRA but without the 2'OH on the loops. Interestingly, the dloop-TERRA had a K_d similar to the c-myc G4 (4.21$\pm$0.07µM). This indicates that the difference in C209 affinity between TERRA and hTel is due to recognition of the 2'OH groups by C209. Interestingly, even a mutant version of TERRA that cannot form a G4

(TERRA-mut) had higher affinity (973±25nM) for C209 than the DNA and deoxy-loop G4s, though still three-fold lower affinity than the G4 version of TERRA.

To investigate the role of ionic interactions in G4 binding by C209 we repeated the MST binding experiments with TERRA and c209 at higher ionic strengths **(Figure 4-2.B)**. Increasing the KCl concentration from 100mM to 375mM resulted in a lowering of the K_d value to 699±57nM. Further increasing the concentration of KCl to 500mM resulted in a further decrease in K_d value to 955±64nM. These results indicate that ionic interactions are, at least in part, responsible for affinity of C209 for G4 RNA.

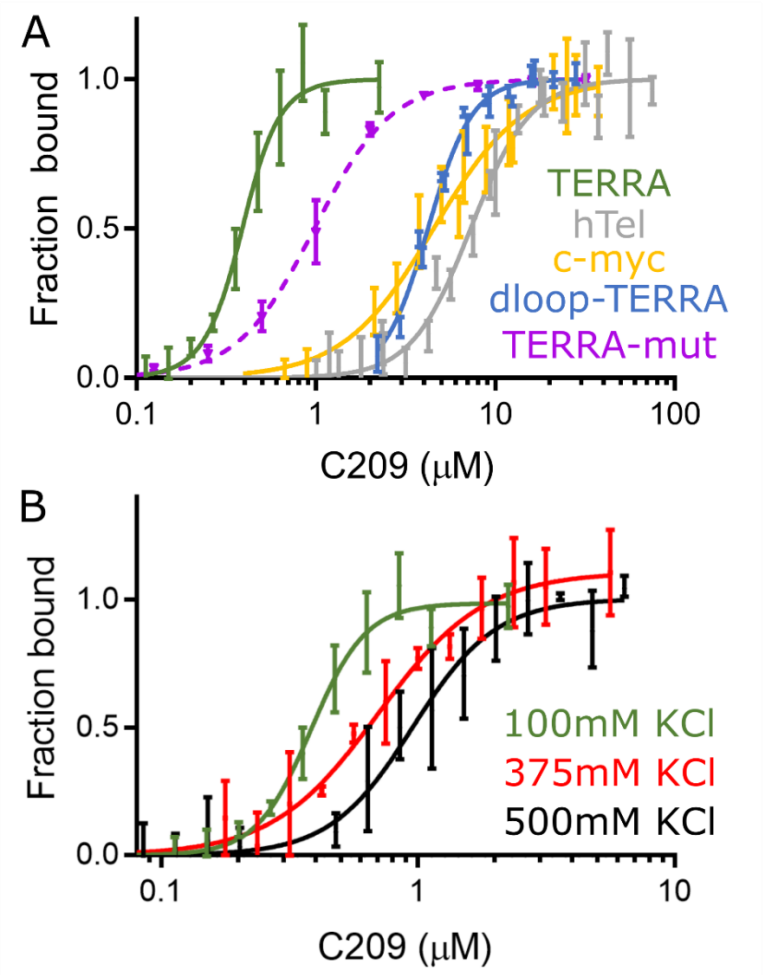


Figure 4-2: Fitted microscale thermophoresis (MST) data

Comparing binding of C209 to (A) different G4s and (B) TERRA G4 at different ionic strengths (100mM, 375mM and 500mM KCl).

4.4.2 Thermodynamic analysis of TERRA binding by DDX21

A well-studied model of TERRA is a dimer consisting of two 10mers with the sequence r(GGGUUAGGGU) (10mer). This version has only two of the three propeller type loops and forms a symmetrical, stacked G4 structure by NMR[190] and crystallography[191]. To date these are the only two structures of TERRA. To better understand the mode of G4 recognition by C209 from a structural perspective we used this RNA 10mer as a model G4 for isothermal titration calorimetry (ITC) and saturation transfer difference NMR (STD-NMR) experiments.

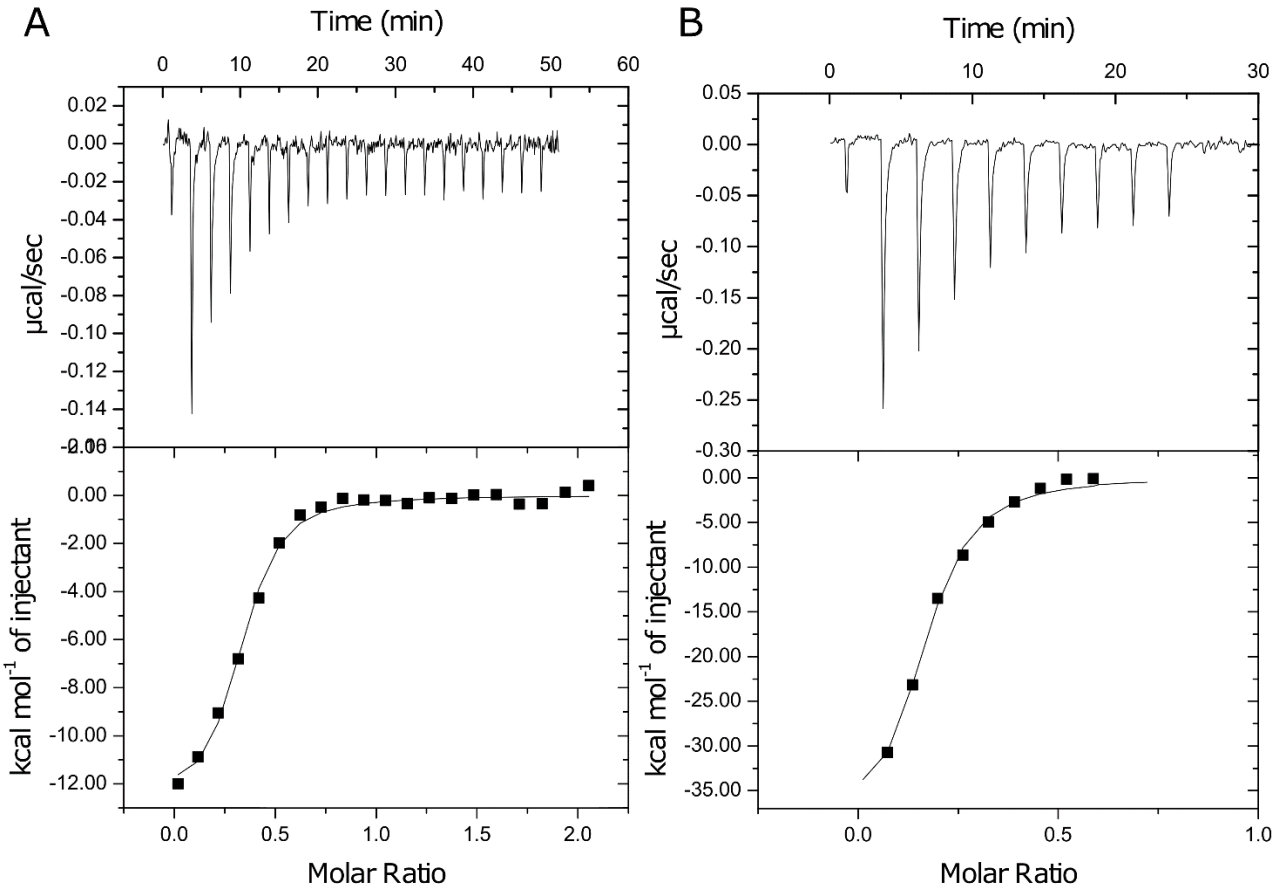


Figure 4-3: Comparing ITC titrations of RNA 10mer (A) and non-labelled TERRA 22mer (B) into C209.

Table 4-1: Thermodynamic parameters for C209 binding to RNA G4s obtained by ITC.

RNA	Temperature	Buffer	N	K_a (M^{-1})	ΔH (kcal/mol)	ΔTS (kcal/mol)	ΔG(kcal/mol)
TERRA	25°C	Tris-Cl	0.21	3.53×10^6 $\pm 5.46 \times 10^5$	-27.58 ±1.41	-18.63 ±2.98	-8.95±1.43
10mer	25°C	Tris-Cl	0.312	3.03×10^6 $\pm 4.80 \times 10^5$	-12.93 ±0.39	-4.08 ±0.66	-8.85 ±1.40
10mer	20°C	PO_4	0.145	4.95×10^6 $\pm 6.72 \times 10^5$	-34.84 ±1.85	-25.81 ±3.76	-9.03 ±1.20
10mer	25°C	PO_4	0.199	1.68×10^7 $\pm 3.91 \times 10^6$	-22.07 ±0.78	-13.17 ±2.96	-8.90 ±2.07
10mer	30°C	PO_4	0.169	3.47×10^6 $\pm 1.14 \times 10^5$	-13.67 ±1.71	-4.60 ±1.61	-9.07 ±2.98

While MST can be used to determine binding constants and stoichiometry, ITC directly measures the enthalpy of binding and the isothermal nature of the experiment allows for the determination of the entropy of binding through use of the Gibbs equation. For these reasons, we used ITC to provide additional insights into G4 binding by C209. Our initial ITC experiments were performed in the same buffer as our MST experiments (20mM Tris-Cl, pH 7.5 with 100mM KCl) **(Figure 4-3)**. When the ITC data was fit to a one site binding model we determined a K_d of 330$\pm$ 52nM between C209 and the 10mer G4 and 280$\pm$44nM between C209 and a non-labeled TERRA 22mer, which is within experimental error of the K_d determined by MST. This gave us confidence that the 10mer of TERRA was behaving similarly to the 22mer (TERRA) and that the full affinity at most required only two of the loop regions.

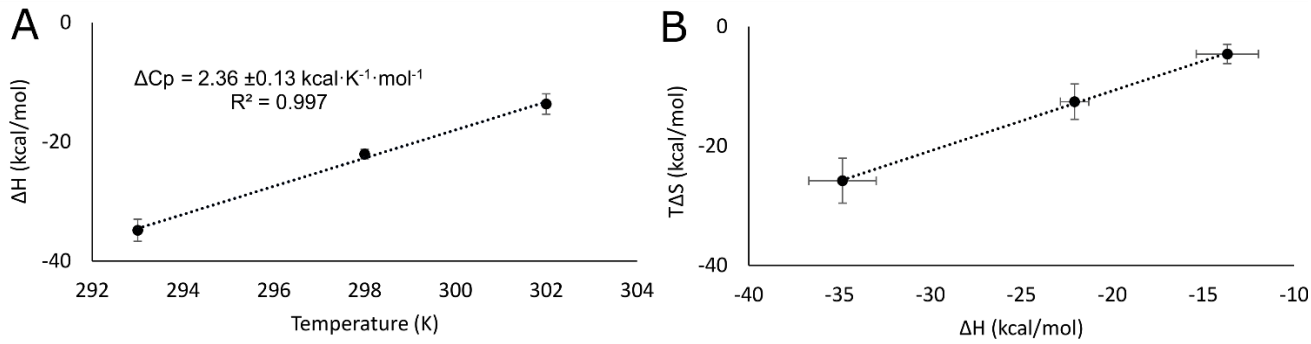


Figure 4-4: Determining the effect of temperature on the thermodynamic properties of C209-G4 interactions.

(A) Temperature dependence of the calorimetric enthalpy of the C209-10mer interaction in phosphate buffer determines the change in heat capacity (ΔCp) of the complex to be large and positive. (B) Enthalpy-entropy compensation plot for the C209-10mer interaction as determined from ITC measurements.

Further to this initial experiment in Tris buffer we also performed ITC in a phosphate buffer **(Figure 4-8)**, which allowed us to take measurements at multiple temperatures without affecting the pH. The interaction was exothermic for all temperatures investigated. Interestingly, a large and positive change in standard heat capacity (ΔCp) was observed across the temperatures investigated **(Figure 4-4A)**, indicating a mostly electrostatically driven interaction. The enthalpy of association decreased from -34.84 ±1.85 kcal/mol at 293K to -13.67 ±1.71 at 303K, while the entropy component of the interaction increased with temperature. The enthalpy-entropy compensation **(Figure 4-4B)** was strong enough that the free energy change corresponding to the association of C209 and 10mer remained essentially constant across the temperatures measured. The two thermodynamic switch temperatures, T_S and T_H, were 303.7K and 307.6K, respectively. Thus, at physiological temperatures (310K) the reaction is entropically driven and endothermic.

4.4.3 2D STD-NMR elucidates the protein binding site on the TERRA G4

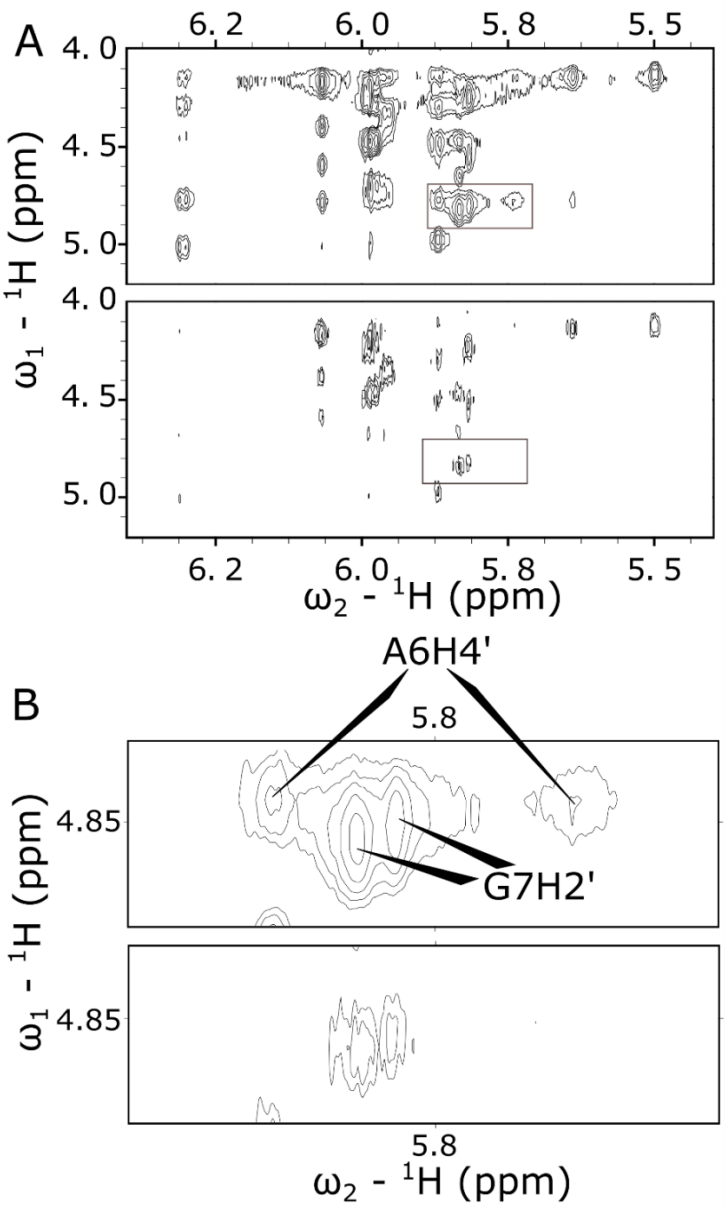


Figure 4-5: 2D NMR of RNA G4 10mer.

(A) Reference NOESY spectrum of RNA G4 10mer (top) and STD NOESY spectrum (bottom)obtained by internal subtraction of the on-resonance saturated spectrum from the reference spectrum by phase cycling. The strength of the cross peaks present in the STD spectrum is correlated to proximity to the protein (C209), while peaks that are absent in STD spectrum but present in the reference spectrum are from protons that are not interacting with the

protein. (B) Enlargement and peak labelled section from the boxed areas from A, reference spectrum (top) and STD spectrum (bottom).

To further investigate the mode of G4 binding by C209 we used STD-NMR to identify the regions of the 10mer G4 that are near C209 upon binding. In the STD experiment, the protein (C209) is selectively saturated with an RF pulse applied to a region with no nucleic acid resonances (-2ppm). This saturation can be transferred to bound RNA via intermolecular ^{1}H-^{1}H cross-relaxation. Over the course of the saturation period a single protein molecule will bind, transfer saturation to the RNA and then dissociate, many times, allowing for sub-stoichiometric protein to be used. This ensures no confounding signals from the protein and no chemical shift perturbation in the RNA spectrum. This “on resonance” spectrum is then subtracted from an “off resonance” spectrum where the saturating RF pulse is applied to a window where neither protein nor RNA resonate (15ppm). In the difference spectrum signals from protons that received no saturation transfer will be cancelled out leaving only signal from protons at the binding interface. Since the protons on the RNA that are closest to the bound protein receive the most saturation transfer, STD-NMR offers significant insights into the binding epitope of the RNA[192].

From our initial 1D H^{1} experiments **(Figure 4-9)** it is clear that the majority of the saturation transfer happened between 3.8 and 5ppm, where the signal from the majority of ribose ring protons resonate. The resolution from our 1D spectrum was insufficient to attribute the signal to specific nucleotides. To increase our resolving power and allow for assignment of the interacting peaks we used a 2D ^{1}H-^{1}H NOESY STD-NMR experiment[193,194] with DPFGSE water suppression[195,196] **(Figure 4-5)**. Using the chemical shifts reported by Phan *et al.*[190] we were able to assign peaks from our NOESY spectrum within 0.02ppm of the values reported in the Biological Magnetic Resonance Bank. To obtain a relative STD amplification factor for each proton we took the ratio of the STD signal intensity to the off-resonance spectrum intensity (I_{STD} / I_0) and scaled them from 0-100%. This allowed us to generate a heat map of the binding epitope and a bar diagram **(Figure 4-6)** depicting the relative STD intensities. A high STD amplification factor means that proton received a significant amount of saturation transfer from the protein and thus was in close proximity to it, whereas a low, or STD amplification factor of 0 indicates little to no saturation transfer was received from the protein and that region is not part of the binding

epitope. A full list of the peak assignments and intensities in the reference and STD spectrum can be found in the supporting information (Table S1).

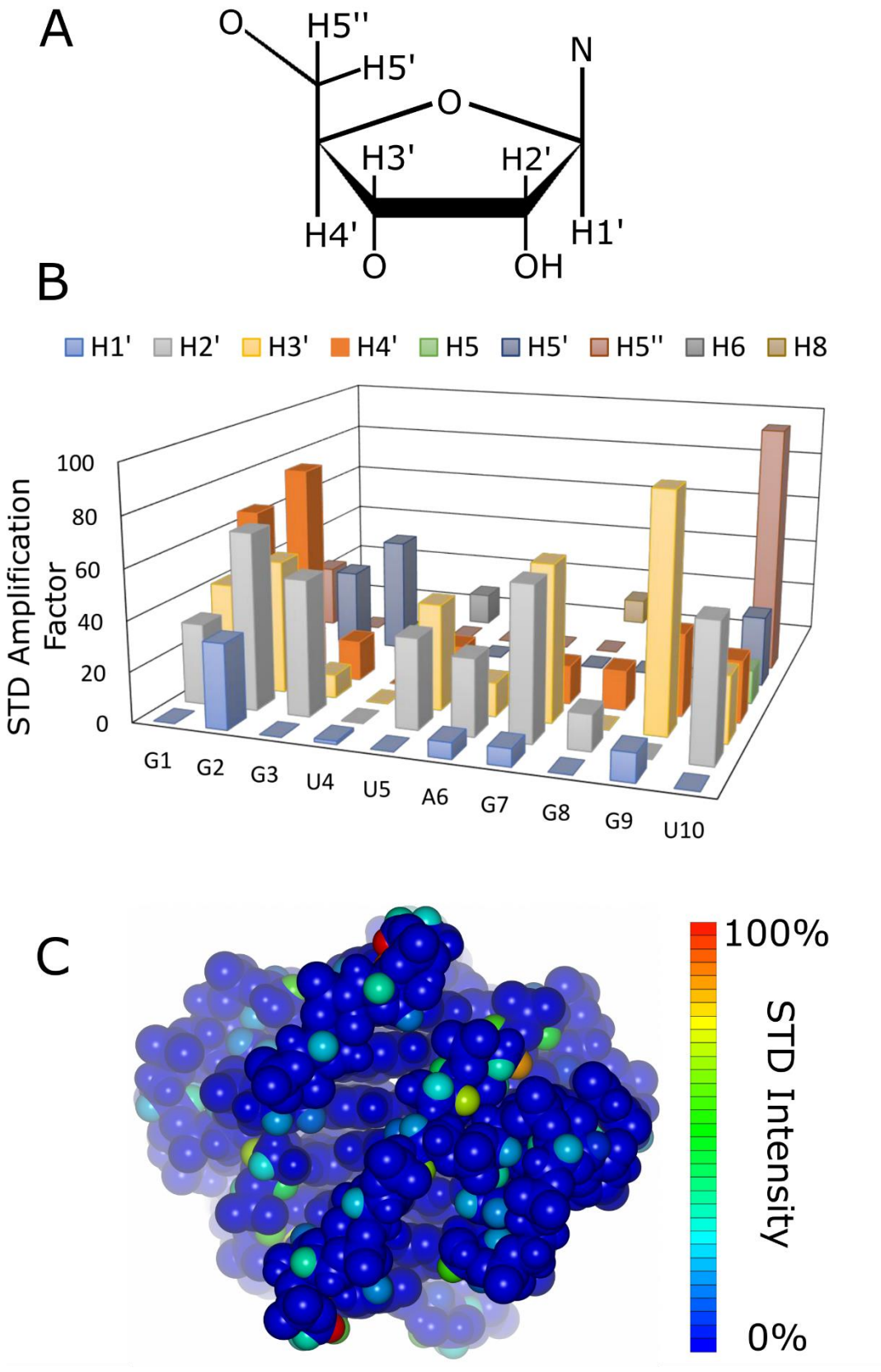

A
O
H5''
H5'
N
O
H3'
H2'
H4'
H1'
O
OH
B
H1'
H2'
H3'
H4'
H5
H5'
H5''
H6
H8
STD Amplification Factor
100
80
60
40
20
0
G1
G2
G3
U4
U5
A6
G7
G8
G9
U10
C
100%
STD Intensity
0%

Figure 4-6: Probing the C209-10mer G4 interaction by 2D-STD-NOESY NMR.

(A) Schematic showing the labeling of protons detected by NMR on the ribose sugar. (B) Bar graph and (C) heat map depicting STD amplification factors for protons from the STD-NOESY spectra of the RNA 10mer G4. Amplification factors represent the signal ratio between the STD spectrum and the reference spectrum (I_{STD} / I_0), high values representing key contact points between RNA and protein and vice versa, and are scaled to the proton with the highest ratio (U10H5’’). The colour scheme in C is scaled from blue, no STD signal, to red, the largest STD Amplification factor (U10H5’’).

5. Discussion

Recognition of G4s by proteins is a topic of great importance for understanding the regulatory roles of G4s. Among the G4 binding proteins studied to date, loop recognition appears to be a common mode of binding. For example, the TT loops of thrombin binding aptamers (TBA and RE31) form a pincer like interaction with exosite I on thrombin[33,37], EWS has a specificity for G4s that is dependent on loop length[49] and specific mutations to the RGG3 domain of FUS/TLS can induce a specificity for RNA G4 that is dependent on loop 2'OH groups[47]. Even the RNA helicase protein DHX36, whose primary mode of interaction is with the guanine tetrad surface, has an affinity that is loop sequence dependent[61].

Using TERRA as a model, our current study reveals DDX21 C209 has a specificity for parallel RNA G4 that is dependent upon 2'OH in the looped nucleotides of the G4. Comparisons between the crystal structures of a TERRA dimer and the DNA counterpart (forced into a parallel structure by crystallization conditions) show that the loop nucleotides share the same sugar puckers. However, modelling of the structure over a 20ns timescale indicate that the propeller loops in the RNA are more rigid than the DNA loops, which un-stack and open up in the simulation. This is attributed to intramolecular hydrogen bond interactions between the 2'OH in the loops and the G4 core contributing to stabilization of the specific structure of the propeller loop[197]. Thus, either specific recognition of the 2'OH or specific recognition of the propeller loop structure, which is stabilized by the 2'OH in the loops, may be the reason for the difference in affinity observed between TERRA and dloop-TERRA with C209.

Interestingly, our ITC experiments reveal a large and positive change in standard heat capacity change ΔCp for the interaction (2.36 ±0.13 $kcal \cdot K^{-1} \cdot mol^{-1}$). The majority of protein-nucleic acid interactions studied to date have a negative ΔCp, for example the interaction of insulin with the ILPR DNA G4[198] or the interaction between Lon protease and the LSPas DNA G4[199]. Typically, large and negative values of ΔCp are attributed to interactions that are hydrophobically driven and result in a reduction in the hydration of nonpolar surface area on the involved species[200–203]. Counter to this, a positive ΔCp is likely representative of an electrostatically driven interaction and de-solvation of polar surfaces, though could also be attributed to partial unfolding and an increase in surface accessible hydrophobic groups. Comparatively few examples exist of large positive ΔCp values exist for intermolecular

interactions. Examples we found in the literature were limited to the formation of a phosphofructokinase tetramer[204], a peptide-heparin interaction[205], cobalt hexamine and spermidine binding to DNA[206] and 7-methyl-GTP binding by eIF4E[207]. Also of note is a study that determined that the unfolding of G4 DNA is associated with a positive change in the standard heat capacity[208]. Our previous work has shown that a group of conserved arginine residues within C209 are essential for G4 binding and that binding distorts G4 enough to allow nuclease digestion of the nucleic acid[43]. It is therefore likely that both the electrostatic interactions between arginines within C209 and the ribose sugars of the G4 as well as partial unfolding of the G4 likely contribute to the large positive ΔCp associated with the interaction.

To confirm the importance of electrostatic interactions in G4 recognition we repeated our MST experiments at high ionic strengths. The lower affinities observed at high ionic strengths are consistent with shielding of the electrostatic charge by the ions in solution. Indeed, if the interaction was predominantly hydrophobic we would expect to see an increased affinity with increased salt concentrations due to the salting-out effects of high concentrations of KCl[209].

RNA-protein interactions are detectable by relatively simple cross-saturation techniques. RNA and DNA binding sites on proteins have been accurately mapped using 2D STD experiments that exploit the 6ppm and 12ppm region to selectively saturate the nucleic acid H5 base resonances and imino protons, respectively[210,211]. It has been shown in principle by Harris *et al* that the reverse experiments can be completed by saturation of the protein backbone CH's at -2ppm and observing the cross-saturation effects on the nucleic acid species[192]. While Harris *et al* only show 1D NMR STD spectra in their publication they allude to the possibility and benefit to using 2D experiments on more complex systems. To the best of our knowledge this is the first publication of such an experiment.

The assignment of peaks from the STD experiments indicate that C209 is in close contact to the ribose backbone of the G4 core, but also makes numerous contacts with the UUA loop nucleotides as well as the terminal U. Notably, the ribose protons H2' and H3' of U5 and A6 as well as H4' of U5 have significant STD signal, while no contacts appear with the ribose sugar of U4. The strong signal from U5 and A6 indicate C209 is in direct contact with the ribose sugars of the loop sequence, while U4 could be providing the structural stability to the loop. Over all, the majority of the large STD amplification factors are observed on H2', H3' and H4' ribose

hydrogens, which is consistent with an electrostatic mode of binding. In comparison the only STD signal we observe from the protons on the nucleic acid bases have a low amplification factor(10-15%) and are only from the looped nucleotides, with the exception of G7H8. This would seem to indicate that C209 is not interacting with the tetrad face.

In conclusion, the C-terminal 209 amino acids of DDX21 preferentially bind the RNA G4 TERRA compared with its DNA counterpart, hTel. Furthermore, this difference in affinity is due to the presence of the 2'OH in the loop sequence. A mode of interaction consisting of electrostatic interactions between polar residues from C209 and phosphoribose backbone and loop sequence is supported by thermodynamic analysis by ITC and contacts observed by STD-NMR. The most common mode of G4 recognition appears to be mediated by loop interactions[212], but our data suggest that DDX21 also contacts with the entire ribose backbone of guanine core. The above described experimental data show that a 2D STD NMR approach may be used for describing the protein binding sites on complex nucleic acid structures for which NMR coordinates have already been elucidated. This method is often used in studies of small molecules that bind nucleic acids but has seen little utility for protein-nucleic acid interactions. In this case, we reveal a potential strategy for the inhibition of DDX21-G4 interactions using G4 groove-binding ligands such as Distamycin A derivatives[213], and further our understanding of potential G4 targets of DDX21. Further investigation into the sequence requirements and effects of loop length G4 recognition by DDX21 would be help clarify whether the specificity comes from an interaction with the 2'OH itself or rather recognition of a loop structure favored by the extra intramolecular contacts afforded by the 2'OH.

6. Materials and Methods

Table 4-2: Oligonucleotide Sequences

Name	Sequence
P1	d(GGTTGGCATATGGCTTCCAGCAAAGATGCCATCAGG)
P2	d(GGTTGGCTCGAGTTATTGACCAAATGCTTTACTGAAACTC)
c-myc	Cy5-d(TGAGGGTGGGTAGGGTGGGT)
hTel	Cy5-d(AGGGTTAGGGTTAGGGTTAGGG)
TERRA	Cy5-r(AGGGUUAGGGUUAGGGUUAGGG)
TERRA-mut	Cy5-r(AGCGUUAGCGUUAGCGUUAGCG)
dloop-TERRA	Cy5-r(GGG)d(UUA)r(GGG)d(UUA)r(GGG)d(UUA)r(GGG)
Non-labelled TERRA	r(AGGGUUAGGGUUAGGGUUAGGG)
10mer	r(GGGUUAGGGU)

4.6.1 Expression and Purification of DDX21-C209

DDX21 cDNA was PCR amplified with primers P1 and P2 and subsequently ligated into the pET-28b(+) expression vector using the Nde I and Xho I restriction sites. Ligation product was transformed into Turbo Competent *E.coli* (New England BioLabs), single colonies were subsequently grown in LB broth (Fischer Scientific) for plasmid purification and sequencing. For protein purification, an overnight culture was started from glycerol stocks of transformed cells and in the morning 10ml was added to 1L of fresh, pre-warmed LB broth. Cells were shaken at 37° and induced to a final IPTG concentration of 1mM when the OD_{600nm} had become 0.4. Cells were grown for a further 5 hours and then collected by centrifugation for 10 minutes at 4000g and resuspended in 30ml of cold lysis buffer (50mM Tris-Cl pH 7.5, 150mM NaCl, 500mK KCl,

1mM PMSF, 5mM DTT). The cell suspension was then lysed by sonication with 10 second pulses and 30 second pauses for a total time of 4 minutes. The resulting lysate was then centrifuged for 45 minutes at 30,000g at 4° C. The supernatant was incubated with 2ml of fresh, pre-equillibrated, Ni-NTA overnight at 4° C. The lysate/bead slurry was then passed through a 5ml gravity flow column and washed with 10ml of lysis buffer, 10ml wash buffer (50mM Tris-Cl pH 7.5, 150mM NaCl) with 2mM imidazole, 25 ml wash buffer with 20mM imidazole and finally eluted with ~50ml wash buffer with 200mM imidazole. The eluent was then diluted and dialyzed as necessary into different buffers for experiments.

4.6.2 Preparation of Nucleic Acid samples

All nucleic acid samples were purchased from IDT (Integrated DNA Technologies) and arrived desalted and lyophilized. All samples were centrifuged, resuspended in H_2O and their concentration confirmed by absorbance at 260nm. Samples were diluted into their experimental buffers and then folded by heating for 5 minutes at 95° C for 5 minutes followed by cooling to 4° C at a rate of 0.1° C per minute.

4.6.3 Circular Dichroism

Circular dichroism spectra were obtained using a Jasco J-810 spectropolarimeter (Jasco Inc., USA) and a 0.1cm quartz cell (Hellma). The samples were prepared in 20mM Tris-Cl pH 7.5 with 50mM KCl and measured under continuous scanning mode with a 1nm data pitch, scan speed of 200nm/minute and a response of 0.25 seconds. The spectra are the average of 6 accumulations. Spectra were converted from the machine units (milli°) to Molar Ellipticity for comparison.

4.6.4 Microscale Thermophoresis (MST)

For MST the DDX21-C209 protein and nucleic acid species were prepared in 20mM Tris-Cl pH 7.5 with 100mM, 375mM or 500mM KCl. Prior to preparing the dilution series both protein and nucleic acid species were centrifuged at 21,500g at 4 ° C for 10 minutes to remove aggregates. Protein samples were serially diluted in 10 µl aliquots to span a final concentration range ~0.1 to 10 times the dissociation constant and then equal volume of Cy5 labelled nucleic acid was added and mixed into each tube so that the final concentration of nucleic acid was 25nM and the final volume 20 µl. Standard capillaries (NanoTemper Technologies, San Francisco , CA, USA) were

used for all MST measurements. Measurements were performed on the Monolith NT.115 instrument (NanoTemper) at 25% LED power and 20% MST power. All samples were prepared and measured at least three times independently. The resulting signal from thermophoresis was fit to the quadratic version of the Hill equation to determine the dissociation constant and Hill coefficient.

4.6.5 Isothermal Titration Calorimetry (ITC)

ITC was performed using a MicroCal ITC200 (GE Healthcare) unit. Samples were prepared in either 20mM Tris-Cl pH 7.5 with 100mM KCl or 20mM potassium phosphate buffer pH 7.0, with a final concentration of 70mM potassium. Protein and nucleic acid samples were centrifuged for 10 minutes at 21,500g prior to the experiments. The binding experiments were performed at 20° C, 25° C and 30° C. A sample syringe with a stirring speed of 1000rpm was used to titrate the RNA 10mer (100µM) into a cell containing ~200 µL of 8µM DDX21-C209. Titrations comprised of 20 injections, an initial 0.4µL injection that was discarded followed by 19 2µL injections. Control experiments were done by injection the RNA 10mer into buffer to obtain the heat effects of dilution of the RNA. The calorimetric enthalpy for each injection was calculated and corrected for the heat of dilution of the RNA. ITC-binding isotherms were analyzed using ORIGIN and the ITC analysis software provided by the manufacturer and fit to a one-site binding model.

4.6.6 NMR

RNA and protein samples were prepared in 20mM potassium phosphate buffer pH 7.0, with a final concentration of 70mM potassium with 10% D_2O and 90% H_2O. For all NMR experiments the concentration of the G4 (dimer) was 1mM and the concentration of protein was 10µM. All NMR experiments were recorded at 298K on an Inova 600 MHz spectrometer (Agilent / Varian) equipped with a room-temperature 5mm probe optimized for ^1H detection. For saturation transfer difference (STD) experiments, one-dimensional (1D) ^1H spectra were acquired with a spectral width of 9612 Hz, relaxation delay 1.5 s, utilizing 8192 data points for acquisition. On-resonance irradiation at -2 ppm for selective saturation, and off-resonance irradiation at 25ppm for reference spectra. A train of soft Gaussian shaped pulses of 50 ms duration with 1 ms delay between pulses were used, for a total saturation time of 2 s. A Double Pulsed Field Gradient Spin Echo (DPFGSE) was applied to selectively suppress water signals. 1D STD spectra were

obtained by internal subtraction of the saturated spectrum from the reference spectrum by phase cycling. 2D transferred nuclear Overhauser effect spectroscopy (trNOESY) STD experiments were acquired with 128 transients with 100 indirect increments using an NOE mixing time of 150 ms, utilizing DPFGSE for water suppression.

All spectra were processed using NMRPipe/NMRDraw. NMR spectra were zero-filled once and the indirect dimension in 2D experiments was treated with 100 points of forward-backward linear prediction. Both 1D and 2D experiments were apodized with cos-squared line broadening.

7. Supplemental Information

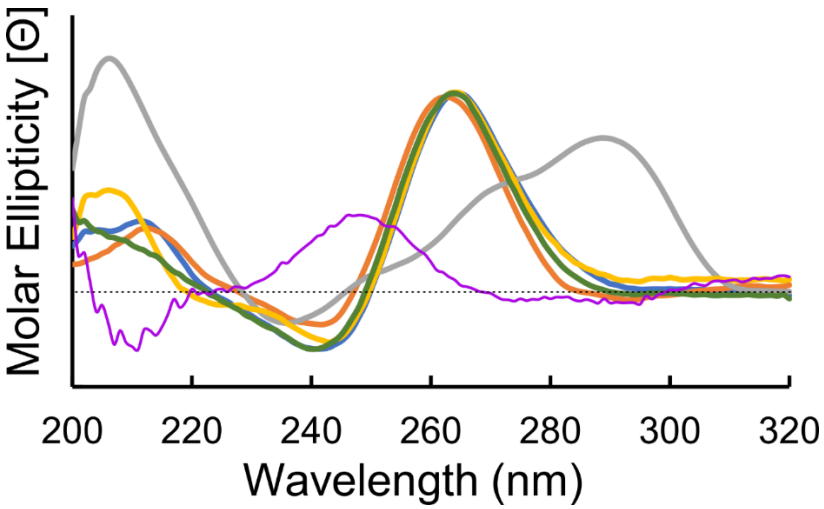


Figure 4-7: Circular dichroism of nucleic acid species used in the study.

Circular dichroism of nucleic acids confirming that the hTel DNA G4 forms a (3+1) non-parallel G4, whereas the rest of the species show the characteristic signal of parallel G4. (hTel-grey, TERRA-mut-majenta TERRA-green, c-myc-yellow, dloop-TERRA-blue, 10mer-orange).

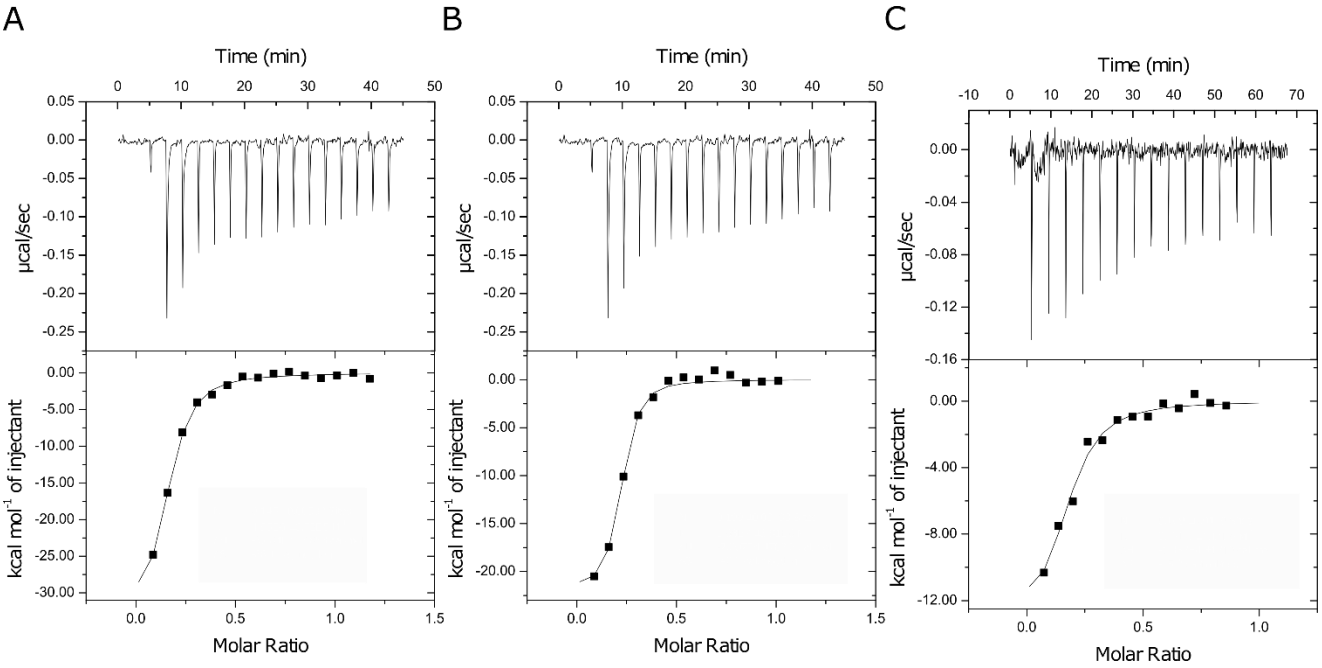


Figure 4-8: ITC titrations of C209 with RNA 10mer G4 in PO4 buffer at (A) 20, (B) 25 and (C) 30 °C.

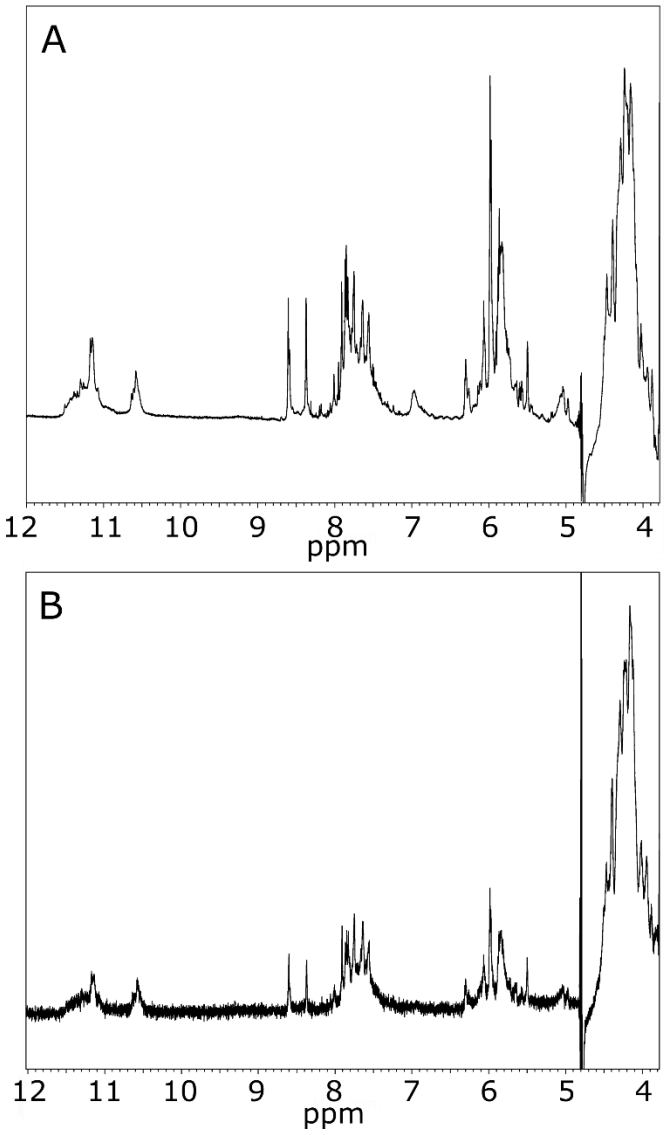


Figure 4-9: 1D STD spectra of 10mer G4 with C209.

(A) 1D ^{1}H reference spectrum of the 10mer RNA G4. (B) STD spectrum depicting specific binding of RNA ribose protons (4-5ppm).

Table 4-3: NOESY assignments of peaks from 2D STD-NMR experiments with C209 and RNA 10mer.

Assignment	w1	w2	Reference S/N	STD S/N
G1H2'-H1'	4.969	5.866	14	16
G1H2'-H4'	4.98	4.291	18	5
G1H2'-H8	4.986	7.912	8	3
G1H2'-A6H2	4.981	8.373	10	4
G1H3'-H1'	4.483	5.866	14	9
G1H3'-H2'	4.497	4.975	8	3
G1H3'-H8	4.488	7.912	16	16
G1H3'-G3H8	4.482	7.762	23	19
G1H4'-H1'	4.298	5.869	16	15
G1H4'-H2'	4.296	4.974	10	11
G1H4'-H8	4.297	7.912	12	9
G1H4'-G2H4'	4.303	3.94	15	10
G1H4'-A6H1	4.279	6.307	7	4
G1H4'-A6H2	4.286	8.371	6	3
G1H5"-H8	4.124	7.911	19	18
G1H5"-A6H2	4.126	8.374	9	7
G2H1'-G1H8	5.655	7.912	6	4
G2H1'-G3H8	5.63	7.749	7	1

G2H1'-A6H2	5.659	8.373	10	4
G2H1'-G7H8	5.648	7.829	9	6
G2H1'-G9H8	5.638	7.647	7	4
G2H2'-G3H8	4.153	7.749	54	52
G2H2'-G9H3'	4.147	4.596	12	13
G2H2'-G9H4'	4.154	4.397	36	65
G2H2'-U10H6	4.155	7.569	32	30
G2H3'-G3H8	4.655	7.749	34	36
G2H3'-G7H1'	4.646	5.832	13	11
G2H3'-G7H8	4.649	7.827	10	7
G2H3'-G8H2'	4.632	4.244	53	54
G2H4'-G1H8	3.946	7.911	19	26
G2H4'-H1'	3.957	5.638	9	7
G2H4'-H5'	3.933	4.109	44	90
G2H4'-U4H5	3.965	5.868	7	6
G2H4'-A6H2	3.952	8.374	8	6
G2H5'-H4'	4.11	3.937	60	60
G2H5'-U10H1'	4.109	5.495	18	20
G3H2'-H5'	4.341	3.829	27	50
G3H2'-H8	4.329	7.748	63	65
G3H2'-U5H1'	4.324	5.961	19	15

G3H3'-H5'	4.702	3.826	8	5
G3H5'-H2'	3.828	4.335	32	54
G3H5'-G9H4'	3.822	4.396	28	41
G4H4'-G7H8	4.437	7.831	47	51
U4H1'-U5H6	5.979	7.868	47	4
U4H1'-A6H8	5.957	8.603	7	1
U4H5-G1H8	5.862	7.911	16	10
U4H5-H6	5.864	7.851	35	7
U4H5-A6H2	5.86	8.374	19	6
U4H6-G1H1'	7.858	5.881	58	27
U4H6-H5	7.86	5.866	58	24
U5H1'-G3H8	5.956	7.748	11	2
U5H2'-U4H6	4.23	7.854	38	29
U5H2'-H6	4.231	7.868	35	28
U5H2'-A6H8	4.253	8.603	16	14
U5H3'-U4H6	4.68	7.851	16	12
U5H3'-H1'	4.695	5.975	12	7
U5H3'-H6	4.675	7.867	15	13
U5H3'-A6H8	4.66	8.602	15	10
U5H4'-G8H1'	4.516	5.817	12	15
U5H6-U4H1'	7.863	5.971	78	38

U5H6-H1'	7.865	5.986	102	39
A6H1-H8	6.305	8.602	14	6
A6H2'-U5H1'	5	5.984	7	8
A6H2'-H1	5.012	6.301	11	5
A6H2'-H8	5.022	8.602	67	41
A6H2'-G9H1'	5.011	6.069	6	0
A6H3'-G1H3'	4.969	4.499	7	1
A6H3'-G9H8	4.973	7.648	34	31
A6H4'-G1H8	4.768	7.912	9	4
A6H4'-G2H1'	4.779	5.641	7	1
A6H4'-G2H4'	4.78	3.939	7	3
A6H4'-U4H5	4.772	5.869	12	2
A6H4'-H1	4.77	6.299	13	2
A6H4'-H2	4.772	8.37	7	3
A6H4'-G9H1'	4.785	6.068	13	2
A6H4'-U10H2'	4.786	3.887	8	2
G7H1'-G8H8	5.831	7.762	15	7
G7H1'-U10H8	5.83	7.637	10	2
G7H2'-A6H8	4.827	8.6	7	5
G7H2'-H1'	4.823	5.833	26	15
G7H2'-H8	4.826	7.829	16	6

G7H2'-G8H1'	4.823	5.816	18	11
G7H2'-G8H2'	4.828	4.241	36	17
G7H2'-G8H8	4.826	7.763	54	47
G7H2'-U10H8	4.809	7.636	18	9
G7H3'-U5H1'	4.463	5.986	44	29
G7H3'-U5H2'	4.469	4.224	35	32
G7H3'-U5H6	4.466	7.867	40	24
G7H3'-A6H8	4.47	8.602	11	6
G7H3'-H1'	4.465	5.835	18	13
G7H3'-G9H8	4.47	7.646	18	14
G7H4'-G3H2'	4.691	4.329	31	32
G7H8-H1'	7.81	5.833	16	11
G8H2'-H1'	4.233	5.817	27	26
G8H4'-G2H4'	4.559	3.942	16	17
G8H4'-A6H2	4.54	8.374	6	2
G9H1'-U10H8	6.051	7.638	8	6
G9H3'-H1'	4.59	6.067	15	14
G9H3'-H4'	4.587	4.396	31	41
G9H3'-H8	4.591	7.637	36	49
G9H3'-U10H2'	4.592	3.884	15	11
G9H3'-U10H4'	4.595	4.136	74	83

G9H3'-U10H6	4.594	7.556	16	14
G9H4'-H1'	4.393	6.068	17	18
G9H4'-U10H6	4.366	7.556	6	7
U10H1'-H6	5.502	7.556	7	2
U10H2'-G9H8	3.878	7.634	10	12
U10H2'-H1'	3.885	5.501	11	14
U10H2'-H6	3.884	7.557	11	13
U10H3'-U5H1'	4.188	5.986	27	25
U10H3'-G9H8	4.197	7.637	36	31
U10H4'-G2H1'	4.143	5.638	18	17
U10H4'-U4H5	4.141	5.867	12	6
U10H4'-A6H1	4.142	6.299	9	2
U10H5-H5"	5.033	4.171	25	11
U10H5-H6	5.044	7.567	30	14
U10H5'-U5H1'	3.984	5.966	6	1
U10H5'-H5"	3.991	4.166	47	90
U10H5"-U5H4'	4.179	4.51	27	31
U10H5"-G7H3'	4.173	4.471	30	36
U10H5"-G9H1'	4.17	6.068	40	35
U10H5"-H5	4.166	5.027	11	5
U10H5"-H5'	4.166	4.016	66	81

Chapter 5: Concluding Remarks

DDX21 is an RNA helicase protein that plays many important cellular roles. My research has, for the first time, demonstrated that DDX21 has high affinity for RNA G4s as well as the ability to destabilize them and allow for conformational switching. At the time of discovery DDX21 was one of three proteins shown to have RNA G4 remodeling activity. Furthering this discovery, I narrowed down the G4 specificity domain to the C-terminus of DDX21 and created site specific mutations that disrupt G4 binding (M4). This allowed us to use the M4 DDX21 as a biological tool to assess the function of G4 binding *in cellulo*. Replacement of endogenous protein with recombinant wild type or M4 DDX21, using siRNA knock-down methodologies, followed by proteomic mass spectrometry allowed us to identify potential G4 targets of DDX21.

One such DDX21/G4 regulated protein we identified is MAGED2, MAGED2 is a P53 interacting protein that has been shown to exert anti-apoptotic effects through regulation of down-stream target TRAIL-R2[154]. RNA G4 detection by rG4-seq[113] has indicated the presence of G4 in the 5' UTR of MAGED2. We confirmed the formation of G4 in transcript variant 2 of MAGED2 mRNA using multiple methods and demonstrate its potential for regulation of translation of MAGED2. DDX21 knock-down or replacement with the M4 mutant results in a two-fold reduction of MAGED2 protein levels and a two-fold increase in TRAIL-R2 protein levels. This increase in TRAIL-R2 levels is significant enough to sensitize MCF-7 cells to TRAIL mediated apoptosis, revealing a new mechanism by which TRAIL sensitization can be achieved for targeted apoptosis in cancerous cells.

To better DDX21's G4 dependent activities we need to understand how DDX21 recognizes G4s. To do this we used an assortment of biophysical techniques to probe the interaction between the G4 specificity domain of DDX21 and a model RNA G4, TERRA. In contrast to the primarily hydrophobic recognition of the G4 tetrad surface by DHX36, these experiments revealed the primary mode of interaction between DDX21 and TERRA to be electrostatic and strongly dependent on loop interactions. Since the loops are the most variable part of the G4, this data helps explain why DDX21 appears to have preference for some RNA G4 over others while DHX36 seems to bind all RNA G4 indiscriminately.

Further projects involving DDX21 are already underway in the McKenna lab. We have had undergraduate honours thesis student Negar Atefi spearheading a project to elucidate the finer details of DDX21's loop specificity. Since nothing is known about the effects of the loop sequence on binding and unwinding activity, the project involves starting with the known binding partner, TERRA, and systematically changing the sequences of the loops. The loop sequence of TERRA being UAA, the first logical step is to test species with UUU and AAA loops to see if there is a difference in binding affinity with DDX21. These are the simplest mutations to make, since the inclusion of guanine in the loop could result in formation of alternate G4 structures and the inclusion of cysteine in the loops could result in competion between a double stranded structure or G4 formation. Once the effect of loop sequence has been ascertained it would be valuable to investigate the effects of loop length on binding and unwinding by DDX21. This could be done in both directions, shortening or lengthening the loops. The results from these studies could aid in the targeted disruption of DDX21 G4 interactions. For example, if a G4 target of DDX21 has a long loop, a short loop and a UUA loop we would be able to rationally guess at the site of interaction based on our prior knowledge of loop specificity.

Another interesting avenue of pursuit would be to look at the effect of G4s on internal ribosome entry site (IRES) formation and initiation of translation from upstream open reading frames (uORFS). Another hit from our whole cell proteome screen for DDX21-G4 affected proteins was XPO6. Upon investigating this hit it was noted that the trend for expression change was the opposite of that which is usually observed. In the presence of DDX21's G4 binding and unwinding activity less of the protein is translated. Usually G4's in the 5' UTR are thought of as roadblocks for the ribosome, but our data would seem to indicate that in this case it is an accelerant. While the presence of G4s in IRES and their ability to promote efficient translation has been reported previously[214,215], the examples of genes regulated in such a way are thus far limited. Preliminary data on the feasibility of such a project would involve the design of a vector that encoded an mRNA with both CAP-dependent and CAP-independent reading frames encoding two different fluorescence proteins. The CAP-dependent gene would be constitutively expressed whereas the CAP-independent gene would come after the stop codon of the first gene and require successful IRES formation for expression. By monitoring the ratio of the two fluorophores one can ascertain the ability of a sequence to act as an IRES.

Another explanation for the effect of DDX21 mutation on XPO6 expression could be ribosome shunting, this is a process where by a structured element in a 5'UTR allows the ribosome to skip over large sections of the mRNA to reach the start codon. This typically involves a uORF immediately prior to the structured section that promotes temporary dissociation of the ribosome and re-attachment downstream of the structured region[216–218]. Methods for evaluating ribosome shunting typically include mutation of the uORFs , which should prevent ribosome dissociation and lead to stalled ribosomes at the structured region of the UTR.

DDX21 has many similarities to the FUS/TLS protein, one of which is the abundance of RGG domains and low-complexity regions. Recent work on FUS has shown that it can form phase-separated liquid-liquid droplets[219]. Phase separation, or liquid-liquid droplet (LLD) formation, is an emerging method of functional compartmentalization of enzymatic activities within cells. A surge of recent studies has highlighted the importance of LLDs in numerous biological processes[220–224], but their basic structural and hydrodynamic properties are poorly understood. Formation of LLDs is driven by multivalent interactions between low-complexity domains of certain proteins under conditions where these interactions become more favorable than solvent interactions[225,226]. This results in the formation of a dense membraneless organelle that has distinct optical properties that allow them to be observed by differential interference contrast (DIC) microscopy. DDX21 contains multiple low-complexity repeats in its C-terminal domain (C209) as well as within its N-terminus. Observations of increased turbidity of C209 at low salt concentrations spurred the investigation into whether it was forming LLDs or amorphous aggregates under these conditions. Preliminary results indicate that LLD formation occurs with C209 at a 5µM concentration in 150mM KCl but not at 300mM KCl. Further experiments should focus on three parameters that drive LLD formation, namely protein concentration, salt concentration and temperature. Each of these parameters should be systematically varied to determine its influence on LLD formation.

Once this initial information has been acquired, mutagenesis to the low complexity region should be performed to see if LLD formation is disrupted. Finally, and of key interest to the field is determining the effect of RNA on LLD formation. DDX21 has been shown to have affinity for a unique structure of RNA known as G-quadruplex. LLD formation should be tested at varying concentrations of, single stranded, double stranded and G-quadruplex RNA. A common method

to observe LLD in cells is to attach a fluorescent marker (i.e. GFP) to the low complexity domain and observe droplet formation by fluorescence and fluorescence recovery experiments in cells expressing this fusion protein. A primer for the methodology mentioned herein would be useful in aiding the experimental design and execution of any student interested in pursuing this project[227].

Combined, this work provides a solid foundation for further investigations into DDX21's role in translation regulation and RNA G4 recognition, both of which are areas of great interest to the scientific community. Furthermore, my work validating DDX21 and MAGED2 from our mass spectrometry screens is only a minute representation of the potential that these datasets have. It is my hope that publishing our list of candidate G4-binding proteins discovered in our initial pull-down assays will aid in the validation of other G4-binding proteins by researchers around the world. Likewise, the effects of DDX21's G4 binding abilities on translation regulation as evaluated by our knock-down recovery mass spectrometry experiments still has much potential for further evaluation. Finally, the biophysical characterization of DDX21's G4 binding domain (C209) will provide the useful constructs and foundational work for a DDX21 structure-function study, an ongoing collaboration between the McKenna and Stetefeld labs.

REFERENCES

1. Howard, F.B., Frazier, J., and Miles, H.T. (1977) Stable and metastable forms of poly(G). *Biopolymers*, **16** (4), 791–809.

2. van Mourik, T., and Dingley, A.J. (2005) Characterization of the monovalent ion position and hydrogen-bond network in guanine quartets by DFT calculations of NMR parameters. *Chemistry*, **11** (20), 6064–79.

3. Pradhan, D., Hansen, L.H., Vester, B., and Petersen, M. (2011) Selection of G-quadruplex folding topology with LNA-modified human telomeric sequences in K+ solution. *Chem. - A Eur. J.*, **17** (8), 2405–2413.

4. Burge, S., Parkinson, G.N., Hazel, P., Todd, A.K., and Neidle, S. (2006) Quadruplex DNA: Sequence, topology and structure. *Nucleic Acids Res.*, **34** (19), 5402–5415.

5. Neidle, S. (2009) The structures of quadruplex nucleic acids and their drug complexes. *Curr. Opin. Struct. Biol.*, **19** (3), 239–250.

6. Zhang, S., Wu, Y., and Zhang, W. (2014) G-quadruplex structures and their interaction diversity with ligands. *ChemMedChem*, **9** (5),

899–911.

7. Ilyinsky, N.S., Varizhuk, a. M., Beniaminov, a. D., Puzanov, M. a., Shchyolkina, a. K., and Kaluzhny, D.N. (2014) G-quadruplex ligands: Mechanisms of anticancer action and target binding. *Mol. Biol.*, **48** (6), 778–794.

8. Huppert, J.L., and Balasubramanian, S. (2005) Prevalence of quadruplexes in the human genome. *Nucleic Acids Res.*, **33** (9), 2908–2916.

9. Chambers, V.S., Marsico, G., Boutell, J.M., Di Antonio, M., Smith, G.P., and Balasubramanian, S. (2015) High-throughput sequencing of DNA G-quadruplex structures in the human genome. *Nat. Biotechnol.*, **33** (8), 877–81.

10. Yuan, L., Tian, T., Chen, Y., Yan, S., Xing, X., Zhang, Z., Zhai, Q., Xu, L., Wang, S., Weng, X., Yuan, B., Feng, Y., and Zhou, X. (2013) Existence of G-quadruplex structures in promoter region of oncogenes confirmed by G-quadruplex DNA cross-linking strategy. *Sci. Rep.*, **3**, 1811.

11. Thakur, R.K., Kumar, P., Halder, K., Verma, A., Kar, A., Parent, J.L., Basundra, R., Kumar, A., and Chowdhury, S. (2009)

Metastases suppressor NM23-H2 interaction with G-quadruplex DNA within c-MYC promoter nuclease hypersensitive element induces c-MYC expression. *Nucleic Acids Res.*, **37** (1), 172–183.

12. Siddiqui-Jain, A., Grand, C.L., Bearss, D.J., and Hurley, L.H. (2002) Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. *Proc. Natl. Acad. Sci. U. S. A.*, **99** (18), 11593–11598.

13. Qin, Y., Rezler, E.M., Gokhale, V., Sun, D., and Hurley, L.H. (2007) Characterization of the G-quadruplexes in the duplex nuclease hypersensitive element of the PDGF-A promoter and modulation of PDGF-A promoter activity by TMPyP4. *Nucleic Acids Res.*, **35** (22), 7698–7713.

14. Booy, E.P., Howard, R., Marushchak, O., Ariyo, E.O., Meier, M., Novakowski, S.K., Deo, S.R., Dzananovic, E., Stetefeld, J., and McKenna, S.A. (2014) The RNA helicase RHAU (DHX36) suppresses expression of the transcription factor PITX1. *Nucleic Acids Res.*, **42** (5), 3346–61.

15. Fletcher, T.M., Sun, D., Salazar, M., and Hurley, L.H. (1998) Effect of DNA Secondary Structure on Human Telomerase Activity. *Biochemistry*, **37**, 5536–5541.

16. Gowan, S.M., Harrison, J.R., Patterson, L., Valenti, M., Read, M. a, Neidle, S., and Kelland, L.R. (2002) A G-quadruplex-interactive potent small-molecule inhibitor of telomerase exhibiting in vitro and in vivo antitumor activity. *Mol. Pharmacol.*, **61** (5), 1154–1162.

17. Bearss, D.J., Hurley, L.H., and Von Hoff, D.D. (2000) Telomere maintenance mechanisms as a target for drug development. *Oncogene*, **19** (56), 6632–6641.

18. Drygin, D., Siddiqui-Jain, A., O'Brien, S., Schwaebe, M., Lin, A., Bliesath, J., Ho, C.B., Proffitt, C., Trent, K., Whitten, J.P., Lim, J.K.C., Von Hoff, D., Anderes, K., and Rice, W.G. (2009) Anticancer activity of CX-3543: A direct inhibitor of rRNA biogenesis. *Cancer Res.*, **69** (19), 7653–7661.

19. Mergny, J.-L., Riou, J.-F., Mailliet, P., Teulade-Fichou, M.-P., and Gilson, E. (2002) Natural and pharmacological regulation of telomerase. *Nucleic Acids Res.*, **30** (4), 839–865.

20. Blice-baum, A.C., and Mihailescu, M. (2013) Biophysical characterization of G-quadruplex forming FMR1 mRNA and of its interactions with different fragile X mental retardation protein isoforms. *RNA*, **20**, 103–114.

21. Haeusler, A.R., Donnelly, C.J., Periz, G., Simko, E.A.J., Shaw, G., Kim, M., Maragakis, N.J., Troncoso, J.C., Pandey, A., Sattler, R., Rothstein, J.D., and Wang, J. (2014) C9orf72 Nucleotide Repeat Structures Initiate Molecular Cascades of Disease. *Nature*, **507** (7491), 195–200.

22. Henderson, A., Wu, Y., Huang, Y.C., Chavez, E.A., Platt, J., Johnson, F.B., Brosh, R.M., Sen, D., and Lansdorp, P.M. (2014) Detection of G-quadruplex DNA in mammalian cells. *Nucleic Acids Res.*, **42** (2), 860–869.

23. Biffi, G., Tannahill, D., McCafferty, J., and Balasubramanian, S. (2013) Quantitative Visualization of DNA G-quadruplex Structures in Human Cells. *Nat. Chem.*, **5** (3), 182–186.

24. Biffi, G., Di Antonio, M., Tannahill, D., and Balasubramanian, S. (2013) Visualization and selective chemical targeting of RNA G-quadruplex structures in the cytoplasm of human cells. *Nat. Chem.*, **6** (1), 75–80.

25. Guo, J.U., and Bartel, D.P. (2016) RNA G-quadruplexes are globally unfolded in eukaryotic cells and depleted in bacteria. *Science*, **353** (6306), 1–8.

26. Keppler, H. (2014) G-quadruplex poses quadruple threat. *Nature*, **507**, 175–177.

27. Ou, T.M., Lu, Y.J., Tan, J.H., Huang, Z.S., Wong, K.Y., and Gu, L.Q. (2008) G-quadruplexes: Targets in anticancer drug design. *ChemMedChem*, **3** (5), 690–713.

28. Parashar, A. (2016) Aptamers in therapeutics. *J. Clin. Diagnostic Res.*, **10** (6), BE01–BE06.

29. Brázda, V., Hároníková, L., Liao, J., and Fojta, M. (2014) DNA and RNA Quadruplex-Binding Proteins. *Int. J. Mol. Sci.*, **15** (10), 17493–17517.

30. Bock, L.C., Griffin, L.C., Latham, J.A., Vermaas, E.H., and Toole, J.J. (1992) Selection of single-stranded DNA molecules that bind and inhibit human thrombin. *Nature*, **355** (6360), 564–566.

31. Macaya, R.F., Schultze, P., Smith, F.W., Roe, J. a, and Feigon, J. (1993) Thrombin-binding DNA aptamer forms a unimolecular quadruplex structure in solution. *Proc. Natl. Acad. Sci. U. S. A.*, **90** (8), 3745–3749.

32. Padmanabhan, K., and Tulinsky, A. (1996) An Ambiguous

Structure of a DNA 15-mer Thrombin Complex. *Acta Crystallogr.*, **52** (2), 272–282.

33. Krauss, I.R., Merlino, A., Giancola, C., Randazzo, A., Mazzarella, L., and Sica, F. (2011) Thrombin-aptamer recognition: A revealed ambiguity. *Nucleic Acids Res.*, **39** (17), 7858–7867.

34. Russo Krauss, I., Merlino, A., Randazzo, A., Novellino, E., Mazzarella, L., and Sica, F. (2012) High-resolution structures of two complexes between thrombin and thrombin-binding aptamer shed light on the role of cations in the aptamer inhibitory activity. *Nucleic Acids Res.*, **40** (16), 8119–8128.

35. Tasset, D.M., Kubik, M.F., and Steiner, W. (1997) Oligonucleotide inhibitors of human thrombin that bind distinct epitopes. *J. Mol. Biol.*, **272** (5), 688–698.

36. Krauss, I.R., Pica, A., Merlino, A., Mazzarella, L., and Sica, F. (2013) Duplex-quadruplex motifs in a peculiar structural organization cooperatively contribute to thrombin binding of a DNA aptamer. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **69** (12), 2403–2411.

37. Russo Krauss, I., Spiridonova, V., Pica, A., Napolitano, V., and

Sica, F. (2016) Different duplex/quadruplex junctions determine the properties of anti-thrombin aptamers with mixed folding. *Nucleic Acids Res.*, **44** (2), 983–991.

38. Thandapani, P., O'Connor, T.R., Bailey, T.L., and Richard, S. (2013) Defining the RGG/RG Motif. *Mol. Cell*, **50** (5), 613–623.

39. Takahama, K., Takada, A., Tada, S., Shimizu, M., Sayama, K., Kurokawa, R., and Oyoshi, T. (2013) Regulation of telomere length by G-quadruplex telomere DNA- and TERRA-binding protein TLS/FUS. *Chem. Biol.*, **20** (3), 341–350.

40. Hurley, V.G. and L.H. (2010) The C-Terminal of Nucleolin Promotes the Formation of the c- MYC G-Quadruplex and Inhibits c-MYC Promoter Activity. *Biochemistry*, **49** (45), 9706–9714.

41. Takahama, K., Kino, K., Arai, S., Kurokawa, R., and Oyoshi, T. (2011) Identification of Ewing's sarcoma protein as a G-quadruplex DNA- and RNA-binding protein. *FEBS J.*, **278** (6), 988–998.

42. Norseen, J., Johnson, F.B., and Lieberman, P.M. (2009) Role for G-quadruplex RNA binding by Epstein-Barr virus nuclear antigen 1 in DNA replication and metaphase chromosome attachment. *J. Virol.*, **83** (20), 10336–10346.

43. McRae, E.K.S., Booy, E.P., Moya-Torres, A., Ezzati, P., Stetefeld, J., and McKenna, S.A. (2017) Human DDX21 binds and unwinds RNA guanine quadruplexes. *Nucleic Acids Res.*, **45** (11), 6656–6668.

44. Zanotti, K.J., Lackey, P.E., Evans, G.L., and Mihailescu, M.-R. (2006) Thermodynamics of the fragile X mental retardation protein RGG box interactions with G quartet forming RNA. *Biochemistry*, **45** (27), 8319–30.

45. Hou, J.-Q., Chen, S.-B., Tan, J.-H., Ou, T.-M., Luo, H.-B., Li, D., Xu, J., Gu, L.-Q., and Huang, Z.-S. (2010) New insights into the structures of ligand-quadruplex complexes from molecular dynamics simulations. *J. Phys. Chem. B*, **114** (46), 15301–15310.

46. Yang, D., and Okamoto, K. (2010) Structural insights into G-quadruplexes: towards new anticancer drugs. *Future Med. Chem.*, **2** (4), 619–646.

47. Takahama, K., and Oyoshi, T. (2013) Specific Binding of Modified RGG Domain in TLS/FUS to G‐Quadruplex RNA: Tyrosines in RGG Domain Recognize 2′-OH of the Riboses of Loops in G‐Quadruplex. *J. Am. Chem. Soc.*, **135** (48), 18016–18019.

48. Takahama, K., Miyawaki, A., Shitara, T., Mitsuya, K., Morikawa, M., Hagihara, M., Kino, K., Yamamoto, A., and Oyoshi, T. (2015) G-Quadruplex DNA- and RNA-Specific-Binding Proteins Engineered from the RGG Domain of TLS/FUS. *ACS Chem. Biol.*, **10** (11), 2564–2569.

49. Takahama, K., Sugimoto, C., Arai, S., Kurokawa, R., and Oyoshi, T. (2011) Loop lengths of G-quadruplex structures affect the G-quadruplex DNA binding selectivity of the RGG motif in Ewing's sarcoma. *Biochemistry*, **50** (23), 5369–5378.

50. Wang, K.Y., Bolton, P.H., Krawczyk, S.H., Bischofberger, N., and Swaminathan, S. (1993) The Tertiary Structure of a DNA Aptamer Which Binds to and Inhibits Thrombin Determines Activity. *Biochemistry*, **32** (42), 11285–11292.

51. Nagatoishi, S., and Sugimoto, N. (2012) Interaction of water with the G-quadruplex loop contributes to the binding energy of G-quadruplex to protein. *Mol. Biosyst.*, **8** (10), 2766.

52. Pica, A., Russo Krauss, I., Merlino, A., Nagatoishi, S., Sugimoto, N., and Sica, F. (2013) Dissecting the contribution of thrombin exosite i in the recognition of thrombin binding aptamer. *FEBS J.*, **280** (24), 6581–6588.

53. Phan, A.T., Kuryavyi, V., Darnell, J.C., Serganov, A., Majumdar, A., Ilin, S., Raslin, T., Polonskaia, A., Chen, C., Clain, D., Darnell, R.B., and Patel, D.J. (2011) Structure-function studies of FMRP RGG peptide recognition of an RNA duplex-quadruplex junction. *Nat. Struct. Mol. Biol.*, **18** (7), 796–804.

54. Vasilyev, N., Polonskaia, A., Darnell, J.C., Darnell, R.B., Patel, D.J., and Serganov, A. (2015) Crystal structure reveals specific recognition of a G-quadruplex RNA by a β-turn in the RGG motif of FMRP. *Proc. Natl. Acad. Sci.*, **112** (39), E5391-400.

55. Darnell, J.C., Jensen, K.B., Jin, P., Brown, V., Warren, S.T., Darnell, R.B., Gadomski, C., Kim, U.J., Korenberg, J.R., Hoogeveen, A.T., and Rudy, B. (2001) Fragile X mental retardation protein targets G quartet mRNAs important for neuronal function. *Cell*, **107** (4), 489–99.

56. Creacy, S.D., Routh, E.D., Iwamoto, F., Nagamine, Y., Akman, S. a., and Vaughn, J.P. (2008) 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.*, **283** (50), 34626–34634.

57. Booy, E.P., Meier, M., Okun, N., Novakowski, S.K., Xiong, S., Stetefeld, J., and McKenna, S. a. (2012) 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.*, **40** (9), 4110–4124.

58. Lattmann, S., Giri, B., Vaughn, J.P., Akman, S.A., and Nagamine, Y. (2010) 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.*, **38** (18), 6219–6233.

59. Meier, M., Patel, T.R., Booy, E.P., Marushchak, O., Okun, N., Deo, S., Howard, R., McEleney, K., Harding, S.E., Stetefeld, J., and McKenna, S.A. (2013) Binding of G-quadruplexes to the N-terminal recognition domain of the RNA helicase associated with AU-rich Element (RHAU). *J. Biol. Chem.*, **288** (49), 35014–35027.

60. Ariyo, E.O., Booy, E.P., Patel, T.R., Dzananovic, E., McRae, E.K.S., Meier, M., McEleney, K., Stetefeld, J., and McKenna, S.A. (2015) Biophysical Characterization of G-Quadruplex Recognition in the PITX1 mRNA by the Specificity Domain of the Helicase RHAU. *PLoS ONEOne*, **10** (12), e0144510.

61. Ariyo, E.O., Booy, E.P., Dzananovic, E., McRae, E.K.S., Meier, M., McEleney, K., Stetefeld, J., and McKenna, S.A. (2017) Impact of G-quadruplex loop conformation in the PITX1 mRNA on protein and small molecule interaction. *Biochem. Biophys. Res. Commun.*, **487** (2), 274–280.

62. Heddi, B., Cheong, V.V., Martadinata, H., and Phan, A.T. (2015) Insights into G-quadruplex specific recognition by the DEAH-box helicase RHAU: Solution structure of a peptide–quadruplex complex. *Proc. Natl. Acad. Sci.*, **112** (31), 9608–9613.

63. Chen, M.C., Tippana, R., Demeshkina, N.A., Murat, P., Balasubramanian, S., Myong, S., and Amaré, A.R.F. (2018) Structural basis of G-quadruplex unfolding by the DEAH/RHA helicase DHX36. *Nature*, **558** (7710), 465–469.

64. Gilman, B., Tijerina, P., and Russell, R. (2017) Nonconventional RNA unwinding mechanisms of DEAD- box and DEAH-box RNA helicase proteins in remodeling structured RNAs and RNPs. *Biochem. Soc. Trans.*, **45** (6), 1313–1321.

65. Ponting, C.P., Oliver, P.L., and Reik, W. (2009) Evolution and Functions of Long Noncoding RNAs. *Cell*, **136** (4), 629–641.

66. Wapinski, O., and Chang, H.Y. (2011) Long noncoding RNAs and human disease. *Trends Cell Biol.*, **21** (6), 354–361.

67. Russell, P. (2012) RNA misfolding and the action of chaperones. *Front Biosci*, **13**, 1–20.

68. Lohman, T.M., Tomko, E.J., and Wu, C.G. (2008) Non-hexameric DNA helicases and translocases: Mechanisms and regulation. *Nat. Rev. Mol. Cell Biol.*, **9** (5), 391–401.

69. Bizebard, T., Ferlenghi, I., Iost, I., and Dreyfus, M. (2004) Studies on three E. coli DEAD-box helicases point to an unwinding mechanism different from that of model DNA helicases. *Biochemistry*, **43** (24), 7857–7866.

70. Rogers, G.W., Lima, W.F., and Merrick, W.C. (2001) Further characterization of the helicase activity of eIF4A. Substrate specificity. *J. Biol. Chem.*, **276** (16), 12598–12608.

71. Umate, P., Tuteja, N., and Tuteja, R. (2011) Genome-wide comprehensive analysis of human helicases. *Commun. Integr. Biol.*, **4** (1), 1–20.

72. Fairman-Williams, M.E., Guenther, U.-P., and Jankowsky, E.

(2010) SF1 and SF2 helicases: family matters. *Curr Opin Struct Biol*, **20** (3), 313–324.

73. Liu, F., Putnam, A., and Jankowsky, E. (2008) ATP hydrolysis is required for DEAD-box protein recycling but not for duplex unwinding. *Proc. Natl. Acad. Sci. U. S. A.*, **105** (51), 20209–20214.

74. Chen, Y., Potratz, J.P., Tijerina, P., Del Campo, M., Lambowitz, A.M., and Russell, R. (2008) DEAD-box proteins can completely separate an RNA duplex using a single ATP. *Proc. Natl. Acad. Sci.*, **105** (51), 20203–20208.

75. Yang, Q., Del Campo, M., Lambowitz, A.M., and Jankowsky, E. (2007) DEAD-Box Proteins Unwind Duplexes by Local Strand Separation. *Mol. Cell*, **28** (2), 253–263.

76. Ozgur, S., Buchwald, G., Falk, S., Chakrabarti, S., Prabu, J.R., and Conti, E. (2015) The conformational plasticity of eukaryotic RNA-dependent ATPases. *FEBS J.*, **282** (5), 850–863.

77. Semlow, D.R., Blanco, M.R., Walter, N.G., and Staley, J.P. (2016) Spliceosomal DEAH-box ATPases remodel pre-mRNA to activate alternative splice sites. *Cell*, **164** (5), 985–998.

78. Walbott, H., Mouffok, S., Capeyrou, R., Lebaron, S., Humbert, O., Van Tilbeurgh, H., Henry, Y., and Leulliot, N. (2010) Prp43p contains a processive helicase structural architecture with a specific regulatory domain. *EMBO J.*, **29** (13), 2194–2204.

79. He, Y., Andersen, G.R., and Nielsen, K.H. (2010) Structural basis for the function of DEAH helicases. *EMBO Rep.*, **11** (3), 180–186.

80. He, Y., Staley, J.P., Andersen, G.R., and Nielsen, K.H. (2017) Structure of the DEAH/RHA ATPase Prp43p bound to RNA implicates a pair of hairpins and motif Va in translocation along RNA. *RNA*, **23** (7), 1110–1124.

81. Flores-Rozas, H., and Hurwitz, J. (1993) Characterization of a new RNA helicase from nuclear extracts of HeLa cells which translocates in the 5’ to 3’ direction. *J. Biol. Chem.*, **268** (28), 21372–21383.

82. Gostout, C.J., Viggiano, T.R., Ahlquist, D.A., Wang, K.K., Larson, M. V, and Balm, R. (1992) The clinical and endoscopic spectrum of the watermelon stomach. *J. Clin. Gastroenterol.*, **15** (3), 256–63.

83. Valdez, B.C., Henning, D., Busch, R.K., Woods, K., Flores-Rozas, H., Hurwitz, J., Perlaky, L., and Busch, H. (1996) A Nucleolar

RNA Helicase Recognized by Autoimmune Antibodies from a Patient With Watermelon Stomach Disease. *Nucleic Acids Res.*, **24** (7), 1220–1224.

84. Ou, Y., Fritzler, M.J., Valdez, B.C., and Rattner, J.B. (1999) Mapping and characterization of the functional domains of the nucleolar protein RNA helicase II/Gu. *Exp. Cell Res.*, **247** (2), 389–398.

85. Hirai, Y., Louvet, E., Oda, T., Kumeta, M., Watanabe, Y., Horigome, T., and Takeyasu, K. (2013) Nucleolar scaffold protein, WDR46, determines the granular compartmental localization of nucleolin and DDX21. *Genes Cells*, **18** (9), 780–97.

86. Westermarck, J., Weiss, C., Saffrich, R., Kast, J., Musti, A.M., Wessely, M., Wilhelm, A., Séraphin, B., Wilm, M., Valdez, B.C., and Dirk, B. (2002) The DEXD/H-box RNA helicase RHII/Gu is a co-factor for c-Jun-activated transcription. *EMBO J.*, **21** (3), 451–460.

87. Valdez, B.C., Henning, D., Perumal, K., and Busch, H. (1997) RNA-unwinding and RNA-folding activities of RNA helicase II/Gu--two activities in separate domains of the same protein. *Eur. J. Biochem.*, **250** (3), 800–807.

88. Valdez, B.C. (2000) Structural domains involved in the RNA folding activity of RNA helicase II/Gu protein. *Eur. J. Biochem.*, **267** (21), 6395–6402.

89. Song, C., Hotz-wagenblatt, A., Voit, R., and Grummt, I. (2017) SIRT7 and the DEAD-box helicase DDX21 cooperate to resolve genomic R loops and safeguard genome stability. *Genes Dev.*, **31**, 1370–1381.

90. Henning, D., So, R.B., Jin, R., Lau, L.F., and Valdez, B.C. (2003) Silencing of RNA Helicase II/Guα Inhibits Mammalian Ribosomal RNA Production. *J. Biol. Chem.*, **278** (52), 52307–52314.

91. Yang, H., Zhou, J., Ochs, R.L., Henning, D., Jin, R., and Valdez, B.C. (2003) Down-regulation of RNA helicase II/Gu results in the depletion of 18 and 28 S rRNAs in Xenopus oocyte. *J. Biol. Chem.*, **278** (40), 38847–38859.

92. Sloan, K.E., Warda, A.S., Sharma, S., Entian, K.D., Lafontaine, D.L.J., and Bohnsack, M.T. (2017) Tuning the ribosome: The influence of rRNA modification on eukaryotic ribosome biogenesis and function. *RNA Biol.*, **14** (9), 1138–1152.

93. Calo, E., Flynn, R. a, Martin, L., Spitale, R.C., Chang, H.Y., and

Wysocka, J. (2014) RNA helicase DDX21 coordinates transcription and ribosomal RNA processing. *Nature*, **518** (7538), 249–253.

94. Sloan, K.E., Leisegang, M.S., Doebele, C., Ramírez, A.S., Simm, S., Safferthal, C., Kretschmer, J., Schorge, T., Markoutsa, S., Haag, S., Karas, M., Ebersberger, I., Schleiff, E., Watkins, N.J., and Bohnsack, M.T. (2015) The association of late-acting snoRNPs with human pre-ribosomal complexes requires the RNA helicase DDX21. *Nucleic Acids Res.*, **43** (1), 553–64.

95. Wisdom, R., Johnson, R.S., and Moore, C. (1999) c-Jun regulates cell cycle progression and apoptosis by distinct mechanisms. *EMBO J.*, **18**, 188–197.

96. Holmstrom, T.H., Mialon, A., Kallio, M., Nymalm, Y., Mannermaa, L., Holm, T., Johansson, H., Black, E., Gillespie, D., Salminen, T.A., Langel, U., Valdez, B.C., and Westermarck, J. (2008) c-Jun Supports Ribosomal RNA Processing and Nucleolar Localization of RNA Helicase DDX21. *J. Biol. Chem.*, **283** (11), 7046–7053.

97. Xing, Y.H., Yao, R.W., Zhang, Y., Guo, C.J., Jiang, S., Xu, G., Dong, R., Yang, L., and Chen, L.L. (2017) SLERT Regulates DDX21 Rings Associated with Pol I Transcription. *Cell*, **169** (4),

664–678.

98. Calo, E., Gu, B., Bowen, M.E., Aryan, F., Zalc, A., Liang, J., Flynn, R.A., Swigut, T., Chang, H.Y., Attardi, L.D., and Wysocka, J. (2018) Tissue-selective effects of nucleolar stress and rDNA damage in developmental disorders. *Nature*, **554** (7690), 112–117.

99. Zhang, Y., Baysac, K.C., Yee, L.-F., Saporita, A.J., and Weber, J.D. (2014) Elevated DDX21 regulates c-Jun activity and rRNA processing in human breast cancers. *Breast Cancer Res.*, **16** (5), 449.

100. Zhang, H., Zhang, Y., Chen, C., Zhu, X., Zhang, C., Xia, Y., Zhao, Y., Andrisani, O.M., and Kong, L. (2018) A double-negative feedback loop between DEAD-box protein DDX21 and Snail regulates epithelial-mesenchymal transition and metastasis in breast cancer. *Cancer Lett.*, **437** (July), 67–78.

101. Cao, J., Wu, N., Han, Y., Hou, Q., Zhao, Y., Pan, Y., Xie, X., and Chen, F. (2018) DDX21 promotes gastric cancer proliferation by regulating cell cycle. *Biochem. Biophys. Res. Commun.*, **505** (4), 1189–1194.

102. Jung, Y., Lee, S., Choi, H.-S., Kim, S.-N., Lee, E., Shin, Y., Seo, J.,

Kim, B., Jung, Y., Kim, W.K., Chun, H.-K., Lee, W.Y., and Kim, J. (2011) Clinical Validation of Colorectal Cancer Biomarkers Identified from Bioinformatics Analysis of Public Expression Data. *Clin Cancer Res*, **17** (4), 700–709.

103. Bonzheim, I., Irmler, M., Klier-Richter, M., Steinhilber, J., Anastasov, N., Schäfer, S., Adam, P., Beckers, J., Raffeld, M., Fend, F., and Quintanilla-Martinez, L. (2013) Identification of C/EBPbeta Target Genes in ALK+ Anaplastic Large Cell Lymphoma (ALCL) by Gene Expression Profiling and Chromatin Immunoprecipitation. *PLoS One*, **8** (5), e64544.

104. Barton Laws, M., Beach, M.C., Lee, Y., Rogers, W.H., Korthius, P.T., Sharp, V., and Wilson, I.B. (2013) Identification of unique MEK-dependent genes in GNAQ mutant uveal melanoma involved in cell growth, tumor cell invasion and MEK-resistance. *Clin Cancer Res*, **17** (1), 148–159.

105. Zhang, Z., Kim, T., Bao, M., Facchinetti, V., Jung, S.Y., Ghaffari, A.A., Qin, J., Cheng, G., and Liu, Y.-J. (2011) DDX1, DDX21, and DHX36 helicases form a complex with the adaptor molecule TRIF to sense dsRNA in dendritic cells. *Immunity*, **34** (6), 866–78.

106. Dong, Y., Ye, W., Yang, J., Han, P., Wang, Y., Ye, C., Weng, D.,

Zhang, F., Xu, Z., and Lei, Y. (2016) DDX21 translocates from nucleus to cytoplasm and stimulates the innate immune response due to dengue virus infection. *Biochem. Biophys. Res. Commun.*, **473** (2), 648–653.

107. Tsai, S.Y., Segovia, J.A., Chang, T.H., Morris, I.R., Berton, M.T., Tessier, P.A., Tardif, M.R., Cesaro, A., and Bose, S. (2014) DAMP Molecule S100A9 Acts as a Molecular Pattern to Enhance Inflammation during Influenza A Virus Infection: Role of DDX21-TRIF-TLR4-MyD88 Pathway. *PLoS Pathog.*, **10** (1), e1003848.

108. Chen, G., Liu, C.-H., Zhou, L., and Krug, R.M. (2014) Cellular DDX21 RNA helicase inhibits influenza A virus replication but is counteracted by the viral NS1 protein. *Cell Host Microbe*, **15** (4), 484–93.

109. Hammond, J.A., Zhou, L., Lamichhane, R., Chu, H.-Y., Millar, D.P., Gerace, L., and Williamson, J.R. (2017) A Survey of DDX21 Activity During Rev/RRE Complex Formation. *J. Mol. Biol.*, **430** (4), 537–553.

110. Zahler, A.M., Williamson, J.R., Cech, T.R., and Prescott, D.M. (1991) Inhibition of telomerase by G-quartet DNA structures. *Nature*, **350** (6320), 718–720.

111. Beaudoin, J.D., and Perreault, J.P. (2013) Exploring mRNA 3'-UTR G-quadruplexes: Evidence of roles in both alternative polyadenylation and mRNA shortening. *Nucleic Acids Res.*, **41** (11), 5898–5911.

112. Wolfe, A.L., Singh, K., Zhong, Y., Drewe, P., Rajasekhar, V.K., Sanghvi, V.R., Mavrakis, K.J., Jiang, M., Roderick, J.E., Van der Meulen, J., Schatz, J.H., Rodrigo, C.M., Zhao, C., Rondou, P., de Stanchina, E., Teruya-Feldstein, J., Kelliher, M.A., Speleman, F., Porco, J.A., Pelletier, J., Rätsch, G., and Wendel, H.-G. (2014) RNA G-quadruplexes cause eIF4A-dependent oncogene translation in cancer. *Nature*, **513** (7516), 65–70.

113. Kwok, C.K., Marsico, G., Sahakyan, A.B., Chambers, V.S., and Balasubramanian, S. (2016) rG4-seq reveals widespread formation of G-quadruplex structures in the human transcriptome. *Nat. Methods*, **13** (10), 841–844.

114. Conlon, E.G., Lu, L., Sharma, A., Yamazaki, T., Tang, T., Shneider, N.A., and Manley, J.L. (2016) The C9ORF72 GGGGCC expansion forms RNA G-quadruplex inclusions and sequesters hnRNP H to disrupt splicing in ALS patient brains. *Elife*, **5**, doi:10.7554/eLife.17820.

115. Hud, N. V., and Plavec, J. (2006) The Role of Cations in Determining Quadruplex Structure and Stability. *Quadruplex Nucleic Acids*, (Chapter 4), 100–130.

116. Bhattacharyya, D., Mirihana Arachchilage, G., and Basu, S. (2016) Metal Cations in G-Quadruplex Folding and Stability. *Front. Chem.*, **4** (38), 1–14.

117. Arora, A., and Maiti, S. (2009) Differential biophysical behavior of human telomeric rna and dna quadruplex. *J. Phys. Chem. B*, **113** (30), 10515–10520.

118. Joachimi, A., Benz, A., and Hartig, J.S. (2009) A comparison of DNA and RNA quadruplex structures and stabilities. *Bioorganic Med. Chem.*, **17** (19), 6811–6815.

119. Agarwala, P., Pandey, S., and Maiti, S. (2015) The tale of RNA G-quadruplex. *Org. Biomol. Chem.*, **13** (20).

120. Bugaut, A., and Balasubramanian, S. (2012) 5’-UTR RNA G-quadruplexes: translation regulation and targeting. *Nucleic Acids Res.*, **40** (11), 4727–4741.

121. Mendoza, O., Bourdoncle, A., Boulé, J.-B., Brosh, R.M., and

Mergny, J.-L. (2016) G-quadruplexes and helicases. *Nucleic Acids Res.*, **44** (5), 1989–2006.

122. Chakraborty, P., and Grosse, F. (2011) Human DHX9 helicase preferentially unwinds RNA-containing displacement loops (R-loops) and G-quadruplexes. *DNA Repair (Amst).*, **10** (6), 654–665.

123. Thomas, P.D., Kejariwal, A., Guo, N., Mi, H., Campbell, M.J., Muruganujan, A., and Lazareva-Ulitsky, B. (2006) Applications for protein sequence-function evolution data: mRNA/protein expression analysis and coding SNP scoring tools. *Nucleic Acids Res.*, **34** (WEB. SERV. ISS.), 645–650.

124. Cimino, D., Fuso, L., Sfiligoi, C., Biglia, N., Ponzone, R., Maggiorotto, F., Russo, G., Cicatiello, L., Weisz, A., Taverna, D., Sismondi, P., and De Bortoli, M. (2008) Identification of new genes associated with breast cancer progression by gene expression analysis of predefined sets of neoplastic tissues. *Int. J. Cancer*, **123** (6), 1327–1338.

125. Booy, E.P., McRae, E.K., Howard, R., Deo, S.R., Ariyo, E.O., Dzananovic, E., Meier, M., Stetefeld, J., and McKenna, S.A. (2016) The RNA Helicase RHAU (DHX36) Interacts with the 3’ tail of the Long Non-coding RNA BC200 (BCYRN1). *J. Biol. Chem.*, **291**

(10), 5355–5372.

126. Zamiri, B., Mirceta, M., Bomsztyk, K., Macgregor, R.B., and Pearson, C.E. (2015) Quadruplex formation by both G-rich and C-rich DNA strands of the C9orf72 (GGGGCC)8•(GGCCCC)8 repeat: Effect of CpG methylation. *Nucleic Acids Res.*, **43** (20), 10055–10064.

127. Kovanda, A., Zalar, M., Sket, P., Plavec, J., and Rogelj, B. (2015) Anti-sense DNA d(GGCCCC)n expansions in C9ORF72 form i-motifs and protonated hairpins. *Sci. Rep.*, **5** (November), 17944.

128. De La Faverie, A.R., Guedin, A., Bedrat, A., Yatsunyk, L.A., and Mergny, J.L. (2014) Thioflavin T as a fluorescence light-up probe for G4 formation. *Nucleic Acids Res.*, **42** (8), e65.

129. Struhl, K. (1989) Ribonucleases. *Curr. Protoc. Mol. Biol.*, **8** (1), 1–3.

130. Matsumura, K., Kawasaki, Y., Miyamoto, M., Kamoshida, Y., Nakamura, J., Negishi, L., Suda, S., and Akiyama, T. (2016) The novel G-quadruplex-containing long non-coding RNA GSEC antagonizes DHX36 and modulates colon cancer cell migration. *Oncogene*, **36** (9), 1191–1199.

131. Arora, A., and Suess, B. (2011) An RNA G-quadruplex in the 3’ UTR of the proto-oncogene PIM1 represses translation. *RNA Biol.*, **8** (February 2015), 802–805.

132. Chen, E., and Joseph, S. (2015) Fragile X mental retardation protein: A paradigm for translational control by RNA-binding proteins. *Biochimie*, **114**, 147–154.

133. Von Hacht, A., Seifert, O., Menger, M., Schütze, T., Arora, A., Konthur, Z., Neubauer, P., Wagner, A., Weise, C., and Kurreck, J. (2014) Identification and characterization of RNA guanine-quadruplex binding proteins. *Nucleic Acids Res.*, **42** (10), 6630–6644.

134. Mori, K., Lammich, S., Mackenzie, I.R.A., Forné, I., Zilow, S., Kretzschmar, H., Edbauer, D., Janssens, J., Kleinberger, G., Cruts, M., Herms, J., Neumann, M., Van Broeckhoven, C., Arzberger, T., and Haass, C. (2013) HnRNP A3 binds to GGGGCC repeats and is a constituent of p62-positive/TDP43-negative inclusions in the hippocampus of patients with C9orf72 mutations. *Acta Neuropathol.*, **125** (3), 413–423.

135. Randazzo, A., Spada, G.P., and Silva, M.. (2013) Circular Dichroism of Quadruplex structures. *Top Curr Chem.*, **330**, 67–86.

136. You, H., Lattmann, S., Rhodes, D., and Yan, J. (2016) RHAU helicase stabilizes G4 in its nucleotide-free state and destabilizes G4 upon ATP hydrolysis. *Nucleic Acids Res.*, **45** (1), gkw881.

137. Tippana, R., Hwang, H., Opresko, P.L., Bohr, V.A., and Myong, S. (2016) Single-molecule imaging reveals a common mechanism shared by G-quadruplex–resolving helicases. *Proc. Natl. Acad. Sci.*, **113** (30), 8448–8453.

138. Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.*, **25** (17), 3389–3402.

139. Cabello, C.M., Bair, W.B., Lamore, S.D., Ley, S., Alexandra, S., Azimian, S., and Wondrak, G.T. (2010) Protein Database Searches Using Compositionally Adjusted Substitution Matrices. *FEBS J.*, **46** (2), 220–231.

140. Volohonsky, G., Terenzi, O., Soichot, J., Naujoks, D.A., Nolan, T., Windbichler, N., Kapps, D., Smidler, A.L., Vittu, A., Costa, G., Steinert, S., Levashina, E.A., Blandin, S.A., and Marois, E. (2015) Tools for Anopheles gambiae Transgenesis. *G3 Genes Genomics Genet.*, **5** (6), 1151–63.

141. Laible, M., and Boonrod, K. (2009) Homemade site directed mutagenesis of whole plasmids. *J. Vis. Exp.*, **27**, 2–4.

142. Kim, I., McKenna, S.A., Viani Puglisi, E., and Puglisi, J.D. (2007) Rapid purification of RNAs using fast performance liquid chromatography (FPLC). *RNA*, **13** (2), 289–94.

143. Booy, E.P., McRae, E.K.S., and McKenna, S.A. (2015) Biochemical characterization of G4 quadruplex telomerase RNA unwinding by the RNA helicase RHAU. *RNA Remodel. Proteins Methods Protoc.*, **1259** (Chapter 9), 125–135.

144. Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D., and Bairoch, A. (2005) Protein Identification and Analysis Tools on the ExPASy Server. *Proteomics Protoc. Handb.*, 571–607.

145. Thomas, P.D., Campbell, M.J., Kejariwal, A., Mi, H., Karlak, B., Daverman, R., Diemer, K., Muruganujan, A., and Narechania, A. (2003) PANTHER: A library of protein families and subfamilies indexed by function. *Genome Res.*, **13** (9), 2129–2141.

146. Mi, H., Dong, Q., Muruganujan, A., Gaudet, P., Lewis, S., and Thomas, P.D. (2009) PANTHER version 7: Improved phylogenetic

trees, orthologs and collaboration with the Gene Ontology Consortium. *Nucleic Acids Res.*, **38** (SUPPL.1), 204–210.

147. McRae, E.K.S., Booy, E.P., Moya-Torres, A., Ezzati, P., Stetefeld, J., and McKenna, S.A. (2017) Human DDX21 binds and unwinds RNA guanine quadruplexes. *Nucleic Acids Res.*, **45** (11).

148. McRae, E.K.S., Davidson, D.E., Dupas, S.J., and McKenna, S.A. (2018) Insights into the RNA quadruplex binding specificity of DDX21. *Biochim. Biophys. Acta - Gen. Subj.*, **1862** (9), 1973–1979.

149. Bolduc, F., Garant, J.M., Allard, F., and Perreault, J.P. (2016) Irregular G-quadruplexes found in the untranslated regions of human mRNAs influence translation. *J. Biol. Chem.*, **291** (41), 21751–21760.

150. Huang, H., Zhang, J., Harvey, S.E., Hu, X., and Cheng, C. (2017) RNA G-quadruplex secondary structure promotes alternative splicing via the RNA-binding protein hnRNPF. *Genes Dev.* , **31** (22), 2296–2309.

151. Ribeiro, M.M., Teixeira, G.S., Martins, L., Marques, M.R., de Souza, A.P., and Line, S.R.P. (2014) G-quadruplex formation enhances splicing efficiency of PAX9 intron 1. *Hum. Genet.*, **134**

(1), 37–44.

152. Weldon, C., Dacanay, J.G., Gokhale, V., Boddupally, P.V.L., Behm-Ansmant, I., Burley, G.A., Branlant, C., Hurley, L.H., Dominguez, C., and Eperon, I.C. (2018) Specific G-quadruplex ligands modulate the alternative splicing of Bcl-X. *Nucleic Acids Res.*, **46** (2), 886–896.

153. Rouleau, S., Glouzon, J.P.S., Brumwell, A., Bisaillon, M., and Perreault, J.P. (2017) 3′ UTR G-quadruplexes regulate miRNA binding. *RNA*, **23** (8), 1172–1179.

154. Tseng, H.Y., Chen, L.H., Ye, Y., Tay, K.H., Jiang, C.C., Guo, S.T., Jin, L., Hersey, P., and Zhang, X.D. (2012) The melanoma-associated antigen MAGE-D2 suppresses TRAIL receptor 2 and protects against TRAIL-induced apoptosis in human melanoma cells. *Carcinogenesis*, **33** (10), 1871–1881.

155. Strekalova, E., Malin, D., Good, D.M., and Cryns, V.L. (2016) Methionine Deprivation Induces a Targetable Vulnerability in Triple-negative Breast Cancer Cells by Enhancing TRAIL Receptor-2 Expression. *Clin Cancer Res*, **21** (12), 2780–2791.

156. Mérino, D., Lalaoui, N., Morizot, A., Solary, E., and Micheau, O.

(2007) TRAIL in cancer therapy: present and future challenges. *Expert Opin Ther Targets*, **11** (10), 1299–1314.

157. Zhang, Y., and Zhang, B. (2008) TRAIL Resistance of Breast Cancer Cells Is Associated with Constitutive Endocytosis of Death Receptors 4 and 5. *Mol Cancer Res*, **6** (12), 1861–1872.

158. Trivedi, R., and Mishra, D.P. (2015) Trailing TRAIL resistance : novel targets for TRAIL sensitization in cancer cells. **5** (April).

159. Manouchehri, J.M., Turner, K.A., and Kalafatis, M. (2018) TRAIL-Induced Apoptosis in TRAIL-Resistant Breast Carcinoma Through Quercetin Cotreatment. *Breast cancer Basic Clin. Res.*, **12**, 31–33.

160. Papageorgio, C., Brachmann, R., Zeng, J.U.E., Culverhouse, R., Zhang, W., and Mcleod, H. (2007) MAGED2 : A novel p53-dissociator. *Int. J. Oncol.*, **31** (5), 1205–1211.

161. Lin, Y., Boone, M., Meuris, L., Lemmens, I., Roy, N. Van, Soete, A., Drmanac, R., Chen, J., Reumers, J., Moisse, M., Lambrechts, D., Peer, Y. Van De, Tavernier, J., and Callewaert, N. (2014) Genome dynamics of the human embryonic kidney 293 lineage in response to cell biology manipulations. *Nat. Commun.*, **5**, 4767.

162. Leroy, B., Girard, L., Hollestelle, A., Minna, J.D., Gazdar, A.F., Soussi, T., and Medical, S. (2014) Analysis of TP53 Mutation Status in Human Cancer Cell Lines: A Reassessment. *Hum. Mutat.*, **35** (6), 756–765.

163. Serikawa, T., Eberle, J., and Kurreck, J. (2017) Effects of Genomic Disruption of a Guanine Quadruplex in the 5’ UTR of the Bcl-2 mRNA in Melanoma Cells. *ARPN J. Eng. Appl. Sci.*, **12** (10), 3218–3221.

164. Xu, S., Li, Q., Xiang, J., Yang, Q., Sun, H., Guan, A., Wang, L., Liu, Y., Yu, L., Shi, Y., Chen, H., and Tang, Y. (2016) Thioflavin T as an efficient fluorescence sensor for selective recognition of RNA G-quadruplexes. *Sci. Rep.*, **6** (April), 24793.

165. Xiao, J., and Chen, H.S. (2005) Biological functions of melanoma-associated antigens (MAGEs) in cell activities. *Chin. J. Cancer*, **24** (1), 124–128.

166. Lucas, S., Brasseur, F., and Boon, T. (1999) A new MAGE gene with ubiquitous expression does not code for known MAGE antigens recognized by T cells. *Cancer Res.*, **59** (16), 4100–4103.

167. Barker, P.A., and Salehi, A. (2002) The MAGE proteins: Emerging

roles in cell cycle progression, apoptosis, and neurogenetic disease. *J. Neurosci. Res.*, **67** (6), 705–712.

168. Kurt, R.A., Urba, W.J., and Schoof, D.D. (2000) Isolation of genes overexpressed in freshly isolated breast cancer specimens. *Breast Cancer Res. Treat.*, **59** (1), 41–48.

169. Rhodes, D., and Lipps, H.J. (2015) Survey and summary G-quadruplexes and their regulatory roles in biology. *Nucleic Acids Res.*, **43** (18), 8627–8637.

170. Fay, M.M., Lyons, S.M., and Ivanov, P. (2017) RNA G-Quadruplexes in Biology: Principles and Molecular Mechanisms. *J. Mol. Biol.*, **429** (14), 2127–2147.

171. Khan, S.R., Baghdasarian, A., Nagar, P.H., Fahlman, R., Jurasz, P., Michail, K., Aljuhani, N., and Siraki, A.G. (2015) Proteomic profile of aminoglutethimide-induced apoptosis in HL-60 cells : Role of myeloperoxidase and arylamine free radicals. *Chem. Biol. Interact.*, **239**, 129–138.

172. Kramer, D.A., Eldeeb, M.A., Wuest, M., Mercer, J., and Fahlman, R.P. (2017) Proteomic characterization of EL4 lymphoma-derived tumors upon chemotherapy treatment reveals potential roles for

lysosomes and caspase-6 during tumor cell death in vivo. *Proteomics*, **17** (12), 1–11.

173. Booy, E.P., McRae, E.K.S., Ezzati, P., Choi, T., Gussakovsky, D., and McKenna, S.A. (2018) Comprehensive analysis of the BC200 ribonucleoprotein reveals a reciprocal regulatory function with CSDE1/UNR. *Nucleic Acids Res.*, **46** (21), 19–24.

174. Blackburn, E.H. (2001) Switching and signaling at the telomere. *Cell*, **106** (6), 661–673.

175. De Lange, T. (2002) Protection of mammalian telomeres. *Oncogene*, **21** (4 REV. ISS. 1), 532–540.

176. De Lange, T. (2005) Shelterin: The protein complex that shapes and safeguards human telomeres. *Genes Dev.*, **19** (18), 2100–2110.

177. Luu, K.N., Phan, A.T., Kuryavyi, V., Lacroix, L., and Patel, D.J. (2006) Structure of the human telomere in K+ solution: An intramolecular (3 + 1) G-quadruplex scaffold. *J. Am. Chem. Soc.*, **128** (30), 9963–9970.

178. Martadinata, H., and Phan, A.T. (2009) Structure of propeller-type parallel-stranded RNA G-quadruplexes, formed by human

telomeric RNA sequences in K+ solution. *J. Am. Chem. Soc.*, **131** (7), 2570–2579.

179. Parkinson, G.N., Lee, M.P.H., and Neidle, S. (2002) Crystal structure of parallel quadruplexes from human telomeric DNA. *Nature*, **417** (6891), 876–880.

180. Azzalin, C.M., Reichenbach, P., Khoriauli, L., Giulotto, E., and Lingner, J. (2007) Telomeric Repeat–Containing RNA and RNA Surveillance Factors at Mammalian Chromosome Ends. *Science*, **318** (3), 798–801.

181. Luke, B., and Lingner, J. (2009) TERRA: Telomeric repeat-containing RNA. *EMBO J.*, **28** (17), 2503–2510.

182. Schoeftner, S., and Blasco, M.A. (2008) Developmentally regulated transcription of mammalian telomeres by DNA-dependent RNA polymerase II. *Nat. Cell Biol.*, **10** (2), 228–236.

183. Deng, Z., Norseen, J., Wiedmer, A., Riethman, H., and Paul, L. (2010) TERRA RNA Binding to TRF2 Facilitates Heterochromatin Formation and ORC Recruitment at Telomeres. *Mol. Cell*, **35** (4), 403–413.

184. Arnoult, N., Van Beneden, A., and Decottignies, A. (2012) Telomere length regulates TERRA levels through increased trimethylation of telomeric H3K9 and HP1α. *Nat. Struct. Mol. Biol.*, **19** (9), 948–956.

185. de Silanes, I.L., d'Alcontres, M.S., and Blasco, M.A. (2010) TERRA transcripts are bound by a complex array of RNA-binding proteins. *Nat. Commun.*, **1** (3), 1–9.

186. Lachapelle, S., Gagné, J.P., Garand, C., Desbiens, M., Coulombe, Y., Bohr, V.A., Hendzel, M.J., Masson, J.Y., Poirier, G.G., and Lebel, M. (2011) Proteome-wide identification of WRN-interacting proteins in untreated and nuclease-treated samples. *J. Proteome Res.*, **10** (3), 1216–1227.

187. Robert, F., and Pelletier, J. (2013) Perturbations of RNA helicases in cancer. *Wiley Interdiscip. Rev. RNA*, **4** (3), 333–349.

188. Ambrus, A., Chen, D., Dai, J., Bialis, T., Jones, R.A., and Yang, D. (2006) Human telomeric sequence forms a hybrid-type intramolecular G-quadruplex structure with mixed parallel/antiparallel strands in potassium solution. *Nucleic Acids Res.*, **34** (9), 2723–2735.

189. Seenisamy, J., Rezler, E.M., Powell, T.J., Tye, D., Gokhale, V., Joshi, C.S., Siddiqui-Jain, A., and Hurley, L.H. (2004) The dynamic character of the G-quadruplex element in the c-MYC promoter and modification by TMPyP4. *J. Am. Chem. Soc.*, **126** (28), 8702–8709.

190. Martadinata, H., and Phan, A.T. (2013) Structure of human telomeric RNA (TERRA): Stacking of two G-quadruplex blocks in K+ solution. *Biochemistry*, **52** (13), 2176–2183.

191. Collie, G.W., Haider, S.M., Neidle, S., and Parkinson, G.N. (2010) A crystallographic and modelling study of a human telomeric RNA (TERRA) quadruplex. *Nucleic Acids Res.*, **38** (16), 5569–80.

192. Harris, K.A., Shekhtman, A., and Agris, P.F. (2013) Specific RNA-protein interactions detected with saturation transfer difference NMR. *RNA Biol.*, **10** (8), 1307–1311.

193. Mayer, M., and Meyer, B. (2001) Group epitope mapping by saturation transfer difference NMR to identify segments of a ligand in direct contact with a protein receptor. *J. Am. Chem. Soc.*, **123** (25), 6108–6117.

194. Kumar, A., Ernst, R.R., and Wüthrich, K. (1980) A two-

dimensional nuclear Overhauser enhancement (2D NOE) experiment for the elucidation of complete proton-proton cross-relaxation networks in biological macromolecules. *Biochem. Biophys. Res. Commun.*, **95** (1), 1–6.

195. Hwang, T.L., and Shaka, A.J. (1995) Water Suppression That Works. Excitation Sculpting Using Arbitrary Wave-Forms and Pulsed-Field Gradients. *J. Magn. Reson.*, **112** (2), 275–279.

196. Dalvit, C. (1998) Efficient multiple-solvent suppression for the study of the interactions of organic solvents with biomolecules. *J. Biomol. NMR*, **11** (4), 437–444.

197. Collie, G.W., Haider, S.M., Neidle, S., and Parkinson, G.N. (2010) A crystallographic and modelling study of a human telomeric RNA (TERRA) quadruplex. *Nucleic Acids Res.*, **38** (16), 5569–5580.

198. Timmer, C.M., Michmerhuizen, N.L., Witte, A.B., Van Winkle, M., Zhou, D., and Sinniah, K. (2014) An isothermal titration and differential scanning calorimetry study of the G-quadruplex DNA-insulin interaction. *J. Phys. Chem. B*, **118** (7), 1784–1790.

199. Chen, S.H., Suzuki, C.K., and Wu, S.H. (2008) Thermodynamic characterization of specific interactions between the human Lon

protease and G-quartet DNA. *Nucleic Acids Res.*, **36** (4), 1273–1287.

200. Sturtevant, J.M. (1977) Heat capacity and entropy changes in processes involving proteins. *Proc. Natl. Acad. Sci.*, **74** (6), 2236–2240.

201. Spolar, R.S., Ha, J.H., and Record, M.T. (1989) Hydrophobic effect in protein folding and other noncovalent processes involving proteins. *Proc. Natl. Acad. Sci. U. S. A.*, **86** (21), 8382–8385.

202. Livingstone, J.R., Spolar, R.S., and Thomas Record, M. (1991) Contribution to the Thermodynamics of Protein Folding from the Reduction in Water-Accessible Nonpolar Surface Area. *Biochemistry*, **30** (17), 4237–4244.

203. Spolar, R., and Record, M. (1994) Coupling of local folding to site-specific binding of proteins to DNA. *Science*, **263** (5148), 777–784.

204. Luther, M.A., Cai, G.Z., and Lee, J.C. (1986) Thermodynamics of Dimer and Tetramer Formations in Rabbit Muscle Phosphofructokinase. *Biochemistry*, **25** (24), 7931–7937.

205. Hileman, R.E., Jennings, R.N., and Linhardt, R.J. (1998)

Thermodynamic analysis of the heparin interaction with a basic cyclic peptide using isothermal titration calorimetry. *Biochemistry*, **37** (43), 15231–15237.

206. Matulis, D., Rouzina, I., and Bloomfield, V.A. (2000) Thermodynamics of DNA binding and condensation: Isothermal titration calorimetry and electrostatic mechanism. *J. Mol. Biol.*, **296** (4), 1053–1063.

207. Niedzwiecka, A., Stepinski, J., Darzynkiewicz, E., Sonenberg, N., and Stolarski, R. (2002) Positive heat capacity change upon specific binding of translation initiation factor eIF4E to mRNA 5′ cap. *Biochemistry*, **41** (40), 12140–12148.

208. Majhi, P., Qi, J., Tang, C.-F., and Shafer, R. (2007) Heat Capacity Changes Associated with Guanine Quadruplex: An Isothermal Titration Calorimetry Study. *Biopolymers*, **89** (4), 302–309.

209. Melander, W., and Horváth, C. (1977) Salt effects on hydrophobic interactions in precipitation and chromatography of proteins: An interpretation of the lyotropic series. *Arch. Biochem. Biophys.*, **183** (1), 200–215.

210. Ramos, A., Kelly, G., Hollingworth, D., Pastore, A., and Frenkiel,

T. (2000) Mapping the interfaces of protein-nucleic acid complexes using cross-saturation. *J. Am. Chem. Soc.*, **122** (46), 11311–11314.

211. Lane, A.N., Kelly, G., Ramos, A., and Frenkiel, T.A. (2001) Determining binding sites in protein-nucleic acid complexes by cross-saturation. *J. Biomol. NMR*, **21** (2), 127–139.

212. Mcrae, E.K.S., Booy, E.P., Padilla-meier, G.P., and Mckenna, S.A. (2017) On characterizing the interactions between proteins and quadruplex structures of nucleic acids. *J. Nucleic Acids*, (9675348), 1–11.

213. Martino, L., Virno, A., Pagano, B., Virgilio, A., Di Micco, S., Galeone, A., Giancola, C., Bifulco, G., Mayol, L., and Randazzo, A. (2007) Structural and thermodynamic studies of the interaction of distamycin A with the parallel quadruplex structure [d(TGGGGT)]4. *J. Am. Chem. Soc.*, **129** (51), 16048–16056.

214. Morris, M.J., Negishi, Y., and Pazsint, C. (2010) An RNA G-Quadruplex Is Essential for Cap-Independent Translation Initiation in Human VEGF IRES. *J. Am. Chem. Soc.*, **132** (50), 17831–17839.

215. Bonnal, S., Schaeffer, C., Créancier, L., Clamens, S., Moine, H., Prats, A.C., and Vagner, S. (2003) A single internal ribosome entry

site containing a G quartet RNA structure drives fibroblast growth factor 2 gene expression at four alternative translation initiation codons. *J. Biol. Chem.*, **278** (41), 39330–39336.

216. Koh, D.C., Edelman, G.M., and Mauro, V.P. (2013) Physical evidence supporting a ribosomal shunting mechanism of translation initiation for *BACE1* mRNA. *Translation*, **1** (1), e24400.

217. Sherrill, K.W., and Lloyd, R.E. (2008) Translation of cIAP2 mRNA Is Mediated Exclusively by a Stress-Modulated Ribosome Shunt. *Mol. Cell. Biol.*, **28** (6), 2011–2022.

218. Rogers, G.W., Edelman, G.M., and Mauro, V.P. (2004) Differential utilization of upstream AUGs in the -secretase mRNA suggests that a shunting mechanism regulates translation. *Proc. Natl. Acad. Sci.*, **101** (9), 2794–2799.

219. Mikhaleva, S., and Lemke, E.A. (2018) Minireview Beyond the Transport Function of Import Receptors : What's All the FUS about ? *Cell*, **173** (3), 549–553.

220. Riaz, N., Wolden, S.L., Gelblum, D.Y., and Eric, J. (2016) Protein Phase Separation: A New Phase in Cell Biology. *Trends Cell Biol.*, **118** (24), 6072–6078.

221. Banani, S.F., Lee, H.O., Hyman, A.A., and Rosen, M.K. (2017) Biomolecular condensates: Organizers of cellular biochemistry. *Nat. Rev. Mol. Cell Biol.*, **18** (5), 285–298.

222. Langdon, E.M., and Gladfelter, A.S. (2018) A New Lens for RNA Localization: Liquid-Liquid Phase Separation. *Annu. Rev. Microbiol.*, **72** (1), 255–271.

223. Plys, A.J., and Kingston, R.E. (2018) Dynamic condensates activate transcription. *Science*, **361** (6400), 329–331.

224. Singh, H.R., and Ostwal, Y.B. (2019) Post-Translational Modification, Phase Separation, and Robust Gene Transcription. *Trends Genet.*, **35** (2), 89–92.

225. Harmon, T.S., Holehouse, A.S., Rosen, M.K., and Pappu, R. V. (2017) Intrinsically disordered linkers determine the interplay between phase separation and gelation in multivalent proteins. *Elife*, **6**, e30294.

226. Ruff, K.M., Roberts, S., Chilkoti, A., and Pappu, R. V. (2018) Advances in Understanding Stimulus-Responsive Phase Behavior of Intrinsically Disordered Protein Polymers. *J. Mol. Biol.*, **430** (23), 4619–4635.

227. Alberti, S., Gladfelter, A., and Mittag, T. (2019) Primer Considerations and Challenges in Studying Liquid-Liquid Phase Separation and Biomolecular Condensates. *Cell*, **176** (3), 419–434.