6656–6668 *Nucleic Acids Research, 2017, Vol. 45, No. 11* *Published online 4 May 2017*
doi: 10.1093/nar/gkx380

Human DDX21 binds and unwinds RNA guanine quadruplexes

Ewan K.S. McRae[1], Evan P. Booy[1], Aniel Moya-Torres[1], Peyman Ezzati[3], Jörg Stetefeld[1,2,4] and Sean A. McKenna[1,2,4,*]

[1]Department of Chemistry, University of Manitoba, Winnipeg, Manitoba, Canada, [2]Department of Biochemistry & Medical Genetics, University of Manitoba, Winnipeg, Manitoba, Canada, [3]Manitoba Centre for Proteomics and Systems Biology, Section of Biomedical Proteomics, Department of Internal Medicine, Rady Faculty of Health Sciences, University of Manitoba and Health Sciences Centre, Winnipeg, Manitoba, Canada and [4]Manitoba Institute for Materials, University of Manitoba, Winnipeg, Manitoba, Canada

Received December 19, 2016; Revised March 31, 2017; Editorial Decision April 21, 2017; Accepted April 24, 2017

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.

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 (1–7). 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 (8). 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 (9,10). G4s are very stable structures *in vitro* under conditions mimicking cellular conditions, with RNA G4s being more stable than their DNA counterparts (11,12). 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, ∼12 000 have been identified by high throughput reverse transcription sequencing methods (rG4-seq) (6,13). 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 (4,14,15). 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 (16–19). 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 (13), 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 (20). Subsequent characterization of DDX21 showed that

*To whom correspondence should be addressed. Tel: +1 204 272 1562; Fax: +1 204 474 7608; Email: sean.mckenna@umanitoba.ca

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 (21–23). DDX21 has been implicated in ribosomal RNA biogenesis (24–26), viral RNA sensing as part of the innate immune system (27–29) and cancer progression, where it has been correlated with disease free survival in breast cancer (30) through regulation of c-jun activity and rRNA processing (31,32). 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.

MATERIALS AND METHODS

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 (33). 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: CGGGAA UUAAGUUCAAACGAA, 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 (34). 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 Supplementary Table. Plasmid containing the DDX21 gene was purchased from Genscript Inc. The pSKB- plasmid was a gift from Eric Marois (Addgene plasmid # 62540) (35). 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 ThermoFisher Scientific (Ottawa, ON, Canada).

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 (36). 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 (37). After purification, the S1 RNA was ∼95% in the top band conformation, boiling for 5 min 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 min before snap cooling on ice.

Recombinant protein expression

$DHX36_{(53–105)}$, herein referred to as DHX36 RSM, was purified from *Escherichia coli* as described previously (38). Protein purification of DDX21 constructs and full length DHX36 was performed as described previously (39), substituting commercial transfection formula for (90 μg per 150 mm cell culture dish) polyetheleneimine. Protein concentration was determined spectrophotometrically by absorbance at 280 nm using extinction coefficients calculated with the ProtParam tool on ExPASy servers (40). For the streptavidin pull-down assays with overexpressed DDX21 constructs, per 150 mm 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 h post transfection.

SDS-PAGE

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

Protein thermal shift assays (PTS)

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

Streptavidin pull-down assays and western blotting

To screen for G4 binding proteins, HEK293T cells from confluent 150 mm dishes were resuspended in 10 ml of cold phosphate-buffered saline per plate and pelleted by centrifugation at 1500g for 5 min at 4°C. Per plate of cells, the cell pellet was suspended in 250 μl of cytoplasmic lysis buffer (25 mM Hepes, pH 7.9, 5 mM KCl, 0.5 mM $MgCl_2$, 0.5% (v/v) NP-40) supplemented with Halt protease and phosphate inhibitor cocktail and Ribolock RNase inhibitor (ThermoFisher Scientific). The cells were lysed on ice for 5 min and centrifuged at 5000 rpm in a benchtop microfuge for 5 min at 4°C. The supernatant was set aside on ice and the pellet was suspended in 250 μl nuclear lysis buffer (25 mM Hepes, pH 7.9, 10% (w/v) sucrose, 350 mM 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 30 s followed by passage through a 20-gauge needle. The nuclear and cytoplasmic fractions were combined and cleared of insoluble material by centrifugation at 14 000 rpm in a bench-top microfuge for 10 min at 4°C. 5′-Biotin labelled RNA were added to a final concentration of 500 nM to 500 μl of cell extract and rotated for 30 min at room temperature. Subsequently, 40 μl of Pierce streptavidin magnetic beads (ThermoFisher Scientific) were added to the samples and rotation at room temperature continued for another 30 min. The beads were captured in a magnetic rack and the cell extract aspirated. The beads were then washed three times with 1 ml of a 1:1 mixture of the cytoplasmic and nuclear lysis buffers, followed by three washes each with 1 ml of cytoplasmic lysis buffer then three washes with 1 ml of nuclear lysis buffer. For validation of DDX21 as a hit, 20% of the beads from a pull-down sample and 50 μg of cell extract was loaded onto a 10% SDS-PAGE and western blotting was performed as described previously (33).

Sample preparation for MS

Beads from each pull-down sample were washed three times with 1 ml 50 mM ammonium bicarbonate. The beads were transferred and resuspended into Siliconized vials (BioPlas, San Rafael, CA, USA) in 50 μL of 50 mM ammonium bicarbonate and the proteins were reduced by 10 mM DTT at 50°C for 30 min. The proteins were alkylated with 30 mM iodoacetamide for 30 min in the dark at room temperature. Unreacted iodoacetamide was quenched by addition of 15 mM DTT. Finally, the protein complexes were digested by 500 ng of sequencing grade Trypsin (Promega) overnight at 37°C using a tube roller with gentle horizontal mixing. The reaction was stopped using 1% trifluoroacetic acid (TFA) by adding 50 μl of 3% TFA to the peptide–beads mixture. The beads were vortexed for 10 min and peptides were extracted. In order to maximize peptide-yield from each sample, we preformed two sequential extractions using 200 μL of 0.1% TFA in acetonitrile, and 20 mM ammonium formate (pH 11) 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 1 ml C18-SD extraction disc cartridge (3M,USA). 2 μg of desalted peptide as determined by NanoDrop 2000 (ThermoFisher) was used for LC–MS/MS analysis.

Nano-RP–LC–MS/MS

Samples were analyzed by nano-RP–LC–MS/MS using an a splitless Ultra 2D Plus (Eksigent, Dublin, CA, USA) system coupled to a high-speed Triple TOF™ 5600+ mass spectrometer (SCIEX, Concord, Canada). Peptides were injected via a PepMap100 trap column [0.3 × 5 mm, 5 μm, 100 Å, (Thermo Scientfic, CA, USA) and a 100 μm × 200 mm analytical column packed with 3 μm Luna C18 (2) (Phenomenex Inc., CA, USA) 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 min LC–MS run time. Eluent B had a gradient from 0% to 35% over 48 min, 35–85% in 1 min and was kept at 85% for 5 min 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 s MS survey scan in the mass range of 380–1250 (m/z) were followed by 20 MS/MS scans of 100 ms 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 >150 counts. Former target ions were excluded for 7 s. A sweeping collision energy setting of 37 ± 15 eV was applied to all precursor ions for collision-induced dissociation.

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 42 147 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 (19,41,42) using the GO Ontology database, released 28 February 2017.

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 0.05 mg/ml bovine serum albumin (BSA). Both the labelled RNA and the protein samples were centrifuged at 21 500g for 10 min to remove aggregates. Binding reactions were prepared in MST buffer supplemented with 0.05 mg/ml

BSA to a total volume of 20 μl. Protein samples were serially diluted to obtain a concentration range spanning ∼0.1 × K_D to ∼10 × 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 quadratic Hill equation:

$$Y = \frac{1}{1 + \left(\frac{K^h}{x^h - 0.5 * \left(L + x^h + K^h - \left(\left(L + x^h + K^h \right)^2 - 4 * \mathrm{L} * x^h \right) 0.5 \right)} \right)} \quad (1)$$

where Y is the thermophoresis signal, K is the dissociation constant K_D, x is the total protein concentration, L is the total RNA concentration and h is the Hill coefficient.

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 (5 mM) 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 1 U/ul. The samples were then mixed again and incubated in a 30°C water bath for 20 min. 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 min, cooled on ice for 1 min 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).

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°C and cooling on ice. Each species (top and bottom band) was then diluted to 10 nM in 30 ul of PBS and treated with either no RNase, 0.5 units of T1, or 10 pg of RNase at room temperature. After 20 min the samples were mixed with an equal volume of denaturing load dye, heated at 95°C for 1 min, 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 KCl and 5% glycerol. ATP concentration was either 0, 1 mM (+) or 5 mM (+++), 25 ng of protein was used and the samples were incubated at 30°C for 20 min. The reaction mixtures were loaded directly onto a 10% polyacrylamide (29:1) TBE gel and resolved by electrophoresis at 75 V for 90 min before visualization with SYBR Gold®.

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 to 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 min 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, 0.2 mg/ml Orange G) and loaded onto a 10% acrylamide (29:1) native TBE gel and electrophoresed at 75 V for 90 min before staining in 10 ml of water with a 25 μM Thioflavin T concentration for 5 min. 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 min and de-stained in water before imaging with an Epson perfection 3170 photo-scanner.

Circular dichroism

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

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.

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, 1×10^6 cells were either mock transfected (with only transfection reagent) or transfected with pcDNA vector containing siRNA resistant DDX21 wild type, m4 mutant, SAT-mutant or $DDX21_{(574\text{-}783)}$. After 24 h, the cells were split 1:3 and reverse transfected with either DDX21 or a control siRNA using lipofectamine RNAiMAX (Invitrogen) following the protocol supplied by the manufacturer. After a further 48 h, 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 (43). After an additional 24 h, 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).

RESULTS

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 (43,44). 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, Q1-mutant. The Q2 SPDA was performed with four 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 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 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.

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 1A) by western blot. We used three

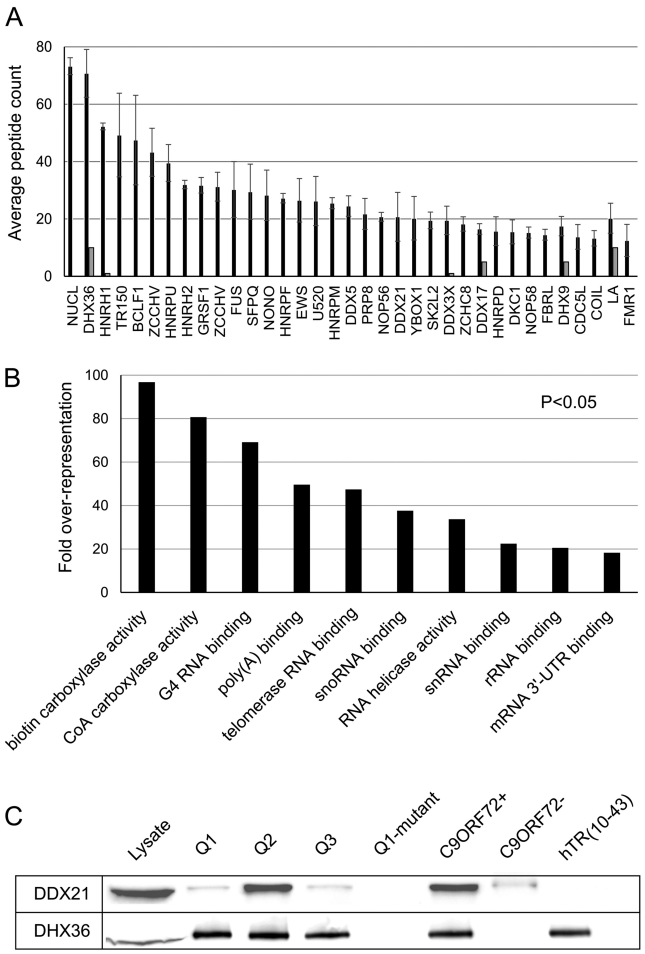

Figure 1. DDX21 is an RNA helicase recovered in abundance from pull-down assays with an RNA G4. (**A**) Proteins identified from streptavidin pull-down assays (SPDA) with RNA G4 by mass spectrometry and sorted by the difference in unique peptide count between G4 and control samples (top 35 hits shown), 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+.

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 ($hTR_{(10\text{–}43)}$), which have been previously well characterized in our lab (33,34,43,44). 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 (45,46). 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 Q1-mutant samples, which showed no detectable signal. This is consistent with DHX36′s es-

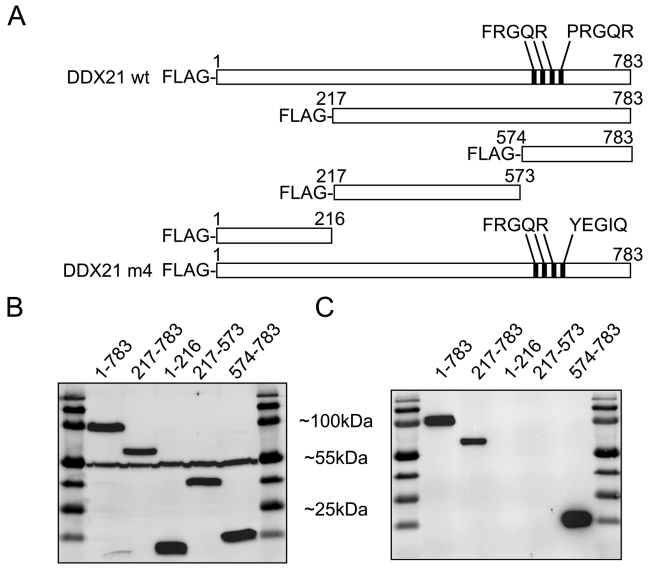

Figure 2. The C terminus of DDX21 ($DDX21_{(574–783)}$) is both necessary and sufficient for the interaction with the Q2 RNA. (**A**) Cartoon representation of N-terminally FLAG-tagged truncations of DDX21. (**B**) Overexpression levels of FLAG-tagged DDX21 truncations in HEK293T cell lysate compared by western blot with anti-FLAG and anti-alpha-tubulin antibodies (~55 kDa). (**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.

tablished role as a parallel G4 helicase (Figure 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 Q1-mutant sample after 5 min of exposure but weak signal was observed from $hTR_{(10–43)}$ (data not shown).

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 2A). Probing for the FLAG-tag by western blot confirmed that all protein truncations were expressed at roughly equivalent levels in HEK293T cells (Figure 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 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 test for a direct interaction between DDX21 and G4 RNA, we purified the full length DDX21 and $DDX21_{(574–783)}$ by FLAG affinity purification (Supplementary Figure S1A) and assessed the binding affinity for fluorescently labelled Q2, Q2-mutant and $hTR_{(10–43)}$ RNA by microscale thermophoresis (MST) (Figure 3). A high affinity (K_d 10 ± 0.7 nM) interaction was observed between Q2 and the full length protein and an ~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 (44). 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.

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 2A). Previous work on DDX21 has shown that mutation of the PRGQR sequence to YEGIQ essentially abolishes its 'foldase' activity (23). 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 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 (Supplementary Figure S1B). 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.

DDX21 and its C-terminal region can unwind RNA G4

Foldase assays that were performed previously on DDX21 by Valdez *et al.* (23) 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 pu-

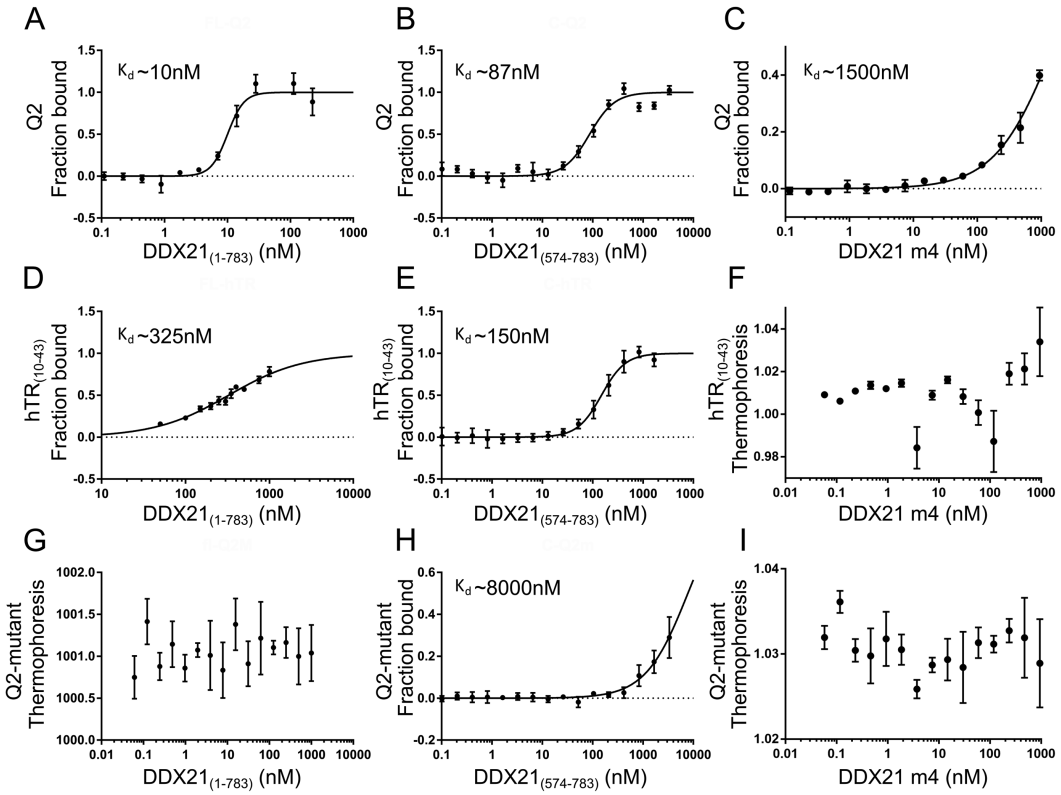

Figure 3. DDX21$_{(1\text{-}783)}$ and DDX21$_{(574\text{-}783)}$ bind quadruplex with high affinity, mutation of four amino acids in a unique repeat region of DDX21 abrogates the interaction. Dose–response data from microscale thermophoresis experiments were fit to a quadratic binding formula to obtain apparent dissociation constants and Hill coefficients for interactions between Q2 (**A**–**C**), hTR$_{(10\text{–}43)}$ (**D**–**F**), Q2-mutant (**G**–**I**) and full length DDX21 (A, D, G), DDX21$_{(574\text{–}783)}$ (B, E, H) and DDX21 m4 (C, F, I). Nano-molar affinities are determined between both quadruplex and the wild type DDX21 and DDX21$_{(574\text{–}783)}$ 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.

rification of the S1 RNA we had a single predominant conformation herein called S1 top band. heating at 95°C 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 (47), caused the dye to exhibit significantly enhanced fluorescence compared to a single stranded RNA control (Figure 4A and B). Furthermore, both the circular dichroism (Figure 4C) and thermal difference spectrum (Figure 4D) resemble the spectra for previously studied RNA G4 (44).

DDX21 and DHX36 both demonstrate the ability to convert the top band to the bottom band in the presence of ATP (Figure 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 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 (48). 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 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 5 shows gels representative of the nuclease sensitivity of two different RNA

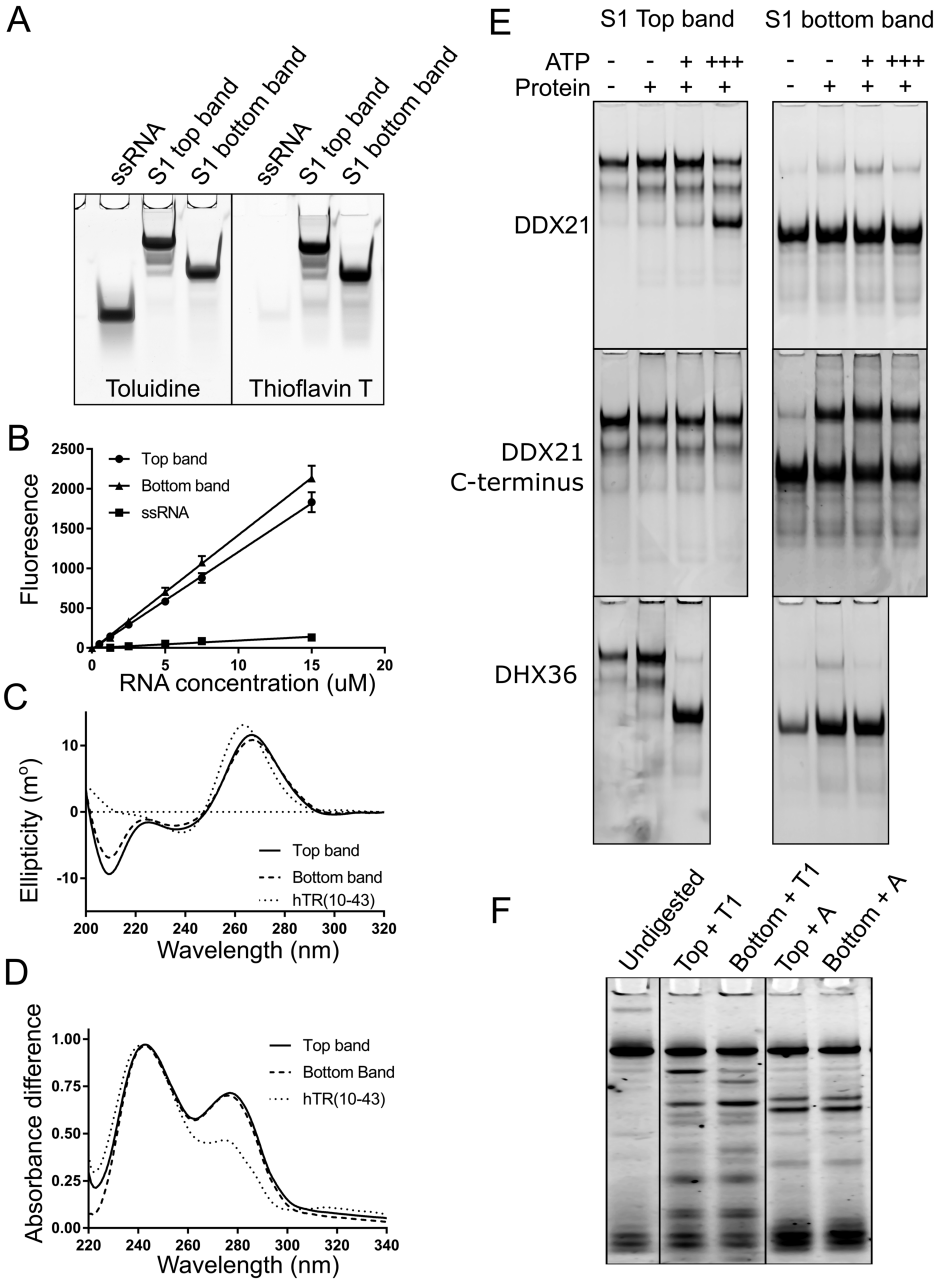

Figure 4. Both the top band and bottom band conformation of the S1 RNA are G4 and DDX21 can unwind these G4 in an ATP dependent manner. Analysis of the two conformations of the S1 RNA (as well as negative control ssRNA and positive control $hTR_{(10-43)}$) by in (**A**) gel staining with Thioflavin T 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 0.5 mM (+) or 5 mM (+++) ATP for 20 min at 30°C 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.

(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 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 oppo-

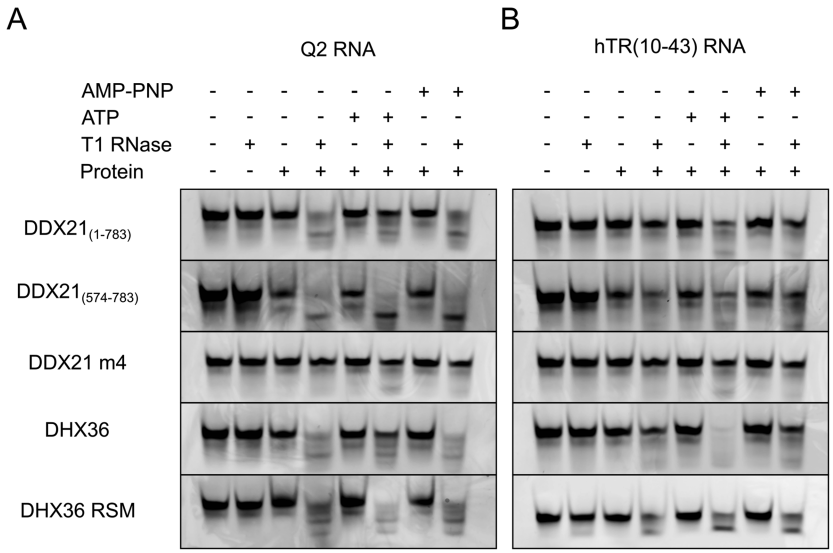

Figure 5. DDX21$_{(1\text{-}783)}$ and DDX21$_{(574\text{-}783)}$ significantly enhance the nuclease sensitivity of (**A**) Q2 and (**B**) hTR$_{(10\text{-}43)}$ RNA G4 while DDX21 m4 does not. Each RNA was incubated with each protein construct with and without RNase T1 and under different ATP conditions, 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.

site 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.

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 (43,49). 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 β-galactosidase activity 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 6A). To compare the protein expression levels of endogenous and recombinant DDX21 constructs western blotting with anti-DDX21 antibody was performed (Figure 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 (43) as well as other work showing a repressive effect of G4 presence in 3′ UTR (50). 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.

DISCUSSION

In recent years, RNA G4 have become recognized as important regulators of gene expression and disease progression (4,5,49,51,52). Recent studies have identified over 12 000 potential unique RNA G4s in the human transcriptome (13,6). Currently only two helicase 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 (16). 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 (13). 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

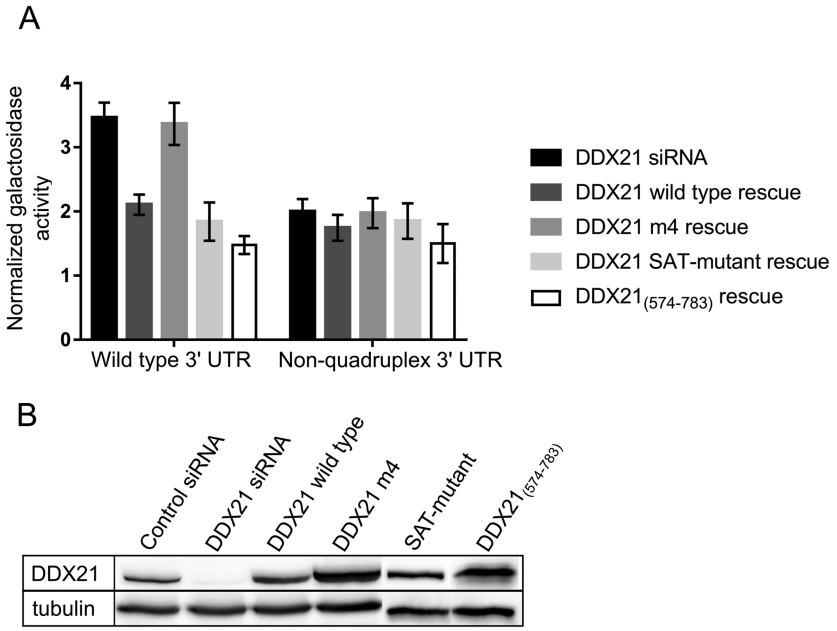

Figure 6. DDX21 wild type but not DDX21 m4 can supress the expression of a beta-galactosidase reporter gene with G4s in its 3′UTR. (**A**) Expression of beta-galactosidase in HEK293T cells from a transfected plasmid with the PITX1 3′UTR or with a mutated quadruplex region under conditions of DDX21 knock down and rescue with DDX21 wild type, DDX21 m4, DDX21 SAT-mutant or $DDX21_{(574\text{-}783)}$. 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 $DDX21_{(574\text{-}783)}$ 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.

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 (53,54). 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 (38,55,56), 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 $hTR_{(10\text{-}43)}$ 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 (Supplementary Table S2) 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 by gel electrophoresis (33). 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 (4,51,57) and has also been used to show protection from nucleases by binding partners (58), recently it has also been used by You *et al* to visualize the unwinding of a G4 by DHX36 (59). 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 (38) 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 DHX36. Both enzymes had 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 (33) but not other RNA G4 (60).

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.* (60) 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 (23) 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 seven runs of guanines that could participate in a G4 conformation. S2, on the other hand, begins with three 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, we 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 (61,62) 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 (63) as well as the FMRP G4 interaction site (58). Mutation of the PRGQR residues to YEGIQ has been previously shown to eliminate the foldase activity of DDX21 (23). 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 (∼4.2×10^4 molecules per cell (20)) 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.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors would like to thank Dr Yoshikuni Nagamine for the monoclonal anti-RHAU 12F33 hybridoma. The au-

thors would also like to thank Dr Thomas Klonisch for the provided cell lines and Dr Eric Marois for the PKSB- plasmid.

FUNDING

Cancer Research Society (Canada) [20085]; Canadian Cancer Society Research Institute [703809]; Research Manitoba Ph.D. studentship (to E.M.); Canada Research Chair program (to J.S.). Funding for open access charge: Canadian Cancer Society Research Institute [703809].
Conflict of interest statement. None declared.

REFERENCES

1. Zahler,A.M., Williamson,J.R., Cech,T.R. and Prescott,D.M. (1991) Inhibition of telomerase by G-quartet DNA structures. *Nature*, **350**, 718–720.
2. 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**, 11593–11598.
3. Huppert,J.L. and Balasubramanian,S. (2005) Prevalence of quadruplexes in the human genome. *Nucleic Acids Res.*, **33**, 2908–2916.
4. 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**, 5898–5911.
5. 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. *et al.* (2014) RNA G-quadruplexes cause eIF4A-dependent oncogene translation in cancer. *Nature*, **513**, 65–70.
6. 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**, 841–844.
7. 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.
8. Burge,S., Parkinson,G.N., Hazel,P., Todd,A.K. and Neidle,S. (2006) Quadruplex DNA: sequence, topology and structure. *Nucleic Acids Res.*, **34**, 5402–5415.
9. Hud,N.V. and Plavec,J. (2006). The role of cations in determining quadruplex structure and stability. *Quadruplex Nucleic Acids*, doi:10.1039/9781847555298-00100.
10. Bhattacharyya,D., Mirihana Arachchilage,G. and Basu,S. (2016) Metal cations in G-quadruplex folding and stability. *Front. Chem.*, **4**, 1–14.
11. Arora,A. and Maiti,S. (2009) Differential biophysical behavior of human telomeric rna and dna quadruplex. *J. Phys. Chem. B*, **113**, 10515–10520.
12. Joachimi,A., Benz,A. and Hartig,J.S. (2009) A comparison of DNA and RNA quadruplex structures and stabilities. *Bioorg. Med. Chem.*, **17**, 6811–6815.
13. Guo,J.U. and Bartel,D.P. (2016) RNA G-quadruplexes are globally unfolded in eukaryotic cells and depleted in bacteria. *Science*, **353**, aaf5371.
14. Agarwala,P., Pandey,S. and Maiti,S. (2015) The tale of RNA G-quadruplex. *Org. Biomol. Chem.*, **13**, 5570–5585.
15. Bugaut,A. and Balasubramanian,S. (2012) 5′-UTR RNA G-quadruplexes: translation regulation and targeting. *Nucleic Acids Res.*, **40**, 4727–4741.
16. Mendoza,O., Bourdoncle,A., Boulé,J.-B., Brosh,R.M. and Mergny,J.-L. (2016) G-quadruplexes and helicases. *Nucleic Acids Res.*, **44**, 1989–2006.
17. 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**, 34626–34634.
18. Chakraborty,P. and Grosse,F. (2011) Human DHX9 helicase preferentially unwinds RNA-containing displacement loops (R-loops) and G-quadruplexes. *DNA Repair (Amst).*, **10**, 654–665.
19. 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**, 645–650.
20. 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**, 21372–21383.
21. 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**, 800–807.
22. 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**, 389–398.
23. Valdez,B.C. (2000) Structural domains involved in the RNA folding activity of RNA helicase II/Gu protein. *Eur. J. Biochem.*, **267**, 6395–6402.
24. 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**, 52307–52314.
25. 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**, 249–253.
26. Sloan,K.E., Leisegang,M.S., Doebele,C., Ramírez,A.S., Simm,S., Safferthal,C., Kretschmer,J., Schorge,T., Markoutsa,S., Haag,S. *et al.* (2015) The association of late-acting snoRNPs with human pre-ribosomal complexes requires the RNA helicase DDX21. *Nucleic Acids Res.*, **43**, 553–564.
27. 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**, 866–878.
28. 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**, 484–493.
29. 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.*, doi:10.1016/j.bbrc.2016.03.120.
30. Cimino,D., Fuso,L., Sfiligoi,C., Biglia,N., Ponzone,R., Maggiorotto,F., Russo,G., Cicatiello,L., Weisz,A., Taverna,D. *et al.* (2008) Identification of new genes associated with breast cancer progression by gene expression analysis of predefined sets of neoplastic tissues. *Int. J. Cancer*, **123**, 1327–1338.
31. Holmstrom,T.H., Mialon,A., Kallio,M., Nymalm,Y., Mannermaa,L., Holm,T., Johansson,H., Black,E., Gillespie,D., Salminen,T.A. *et al.* (2008) c-Jun supports ribosomal RNA processing and nucleolar localization of RNA helicase DDX21. *J. Biol. Chem.*, **283**, 7046–7053.
32. 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**, 449.
33. 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**, 4110–4124.
34. 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**, 5355–5372.
35. Volohonsky,G., Terenzi,O., Soichot,J., Naujoks,D.A., Nolan,T., Windbichler,N., Kapps,D., Smidler,A.L., Vittu,A., Costa,G. *et al.* (2015) Tools for Anopheles gambiae Transgenesis. *G3 (Bethesda).*, **5**, 1151–1163.

36. Laible,M. and Boonrod,K. (2009) Homemade site directed mutagenesis of whole plasmids. *J. Vis. Exp.*, **27**, 2–4.
37. 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**, 289–294.
38. Meier,M., Patel,T.R., Booy,E.P., Marushchak,O., Okun,N., Deo,S., Howard,R., McEleney,K., Harding,S.E., Stetefeld,J. *et al.* (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**, 35014–35027.
39. 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**, 125–135.
40. 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.
41. 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**, 2129–2141.
42. 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**, 204–210.
43. 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**, 3346–3361.
44. Ariyo,E.O., Booy,E.P., Patel,T.R., Dzananovic,E., McRae,E.K., 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 One*, **10**, e0144510.
45. 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**, 10055–10064.
46. 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**, 17944.
47. 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**, e65.
48. Struhl,K. (2001) Ribonucleases. *Curr. Protoc. Mol. Biol.*, doi:10.1002/0471142727.mb0313s08.
49. 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*, doi:10.1038/onc.2016.282.
50. Arora,A. and Suess,B. (2011) An RNA G-quadruplex in the 3′ UTR of the proto-oncogene PIM1 represses translation. *RNA Biol.*, **8**, 802–805.
51. Bugaut,A. and Balasubramanian,S. (2012) 5′-UTR RNA G-quadruplexes: translation regulation and targeting. *Nucleic Acids Res.*, **40**, 4727–4741.
52. Chen,E. and Joseph,S. (2015) Fragile X mental retardation protein: a paradigm for translational control by RNA-binding proteins. *Biochimie*, **114**, 147–154.
53. 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**, 6630–6644.
54. Mori,K., Lammich,S., Mackenzie,I.R.A., Forné,I., Zilow,S., Kretzschmar,H., Edbauer,D., Janssens,J., Kleinberger,G., Cruts,M. *et al.* (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**, 413–423.
55. Phan,A.T., Kuryavyi,V., Darnell,J.C., Serganov,A., Majumdar,A., Ilin,S., Raslin,T., Polonskaia,A., Chen,C., Clain,D. *et al.* (2011) Structure-function studies of FMRP RGG peptide recognition of an RNA duplex-quadruplex junction. *Nat. Struct. Mol. Biol.* **18**, 796–804.
56. 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. U.S.A.*, **112**, 9608–9613.
57. Houk,K.N., Angeles,L., Hunter,C.A, Krische,M.J., Ley,S.V, Olivucci,M., Thiem,J., Venturi,M., Vogel,P., Kong,H. *et al.* (2013). *Quadruplex Nucleic Acids*. doi:10.1007/978-3-642-34743-6.
58. 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. U.S.A.*, **112**, E5391–E5400.
59. 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**, gkw881.
60. 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. U.S.A.*, **113**, 8448–8453.
61. 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**, 3389–3402.
62. 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**, 220–231.
63. Thandapani,P., O'Connor,T.R., Bailey,T.L. and Richard,S. (2013) Defining the RGG/RG Motif. *Mol. Cell*, **50**, 613–623.