

# Identification and characterization of RNA guanine-quadruplex binding proteins

Annekathrin von Hacht<sup>1</sup>, Oliver Seifert<sup>2</sup>, Marcus Menger<sup>3</sup>, Tatjana Schütze<sup>1</sup>, Amit Arora<sup>4</sup>, Zoltán Konthur<sup>5,6</sup>, Peter Neubauer<sup>7</sup>, Anke Wagner<sup>1</sup>, Christoph Weise<sup>8</sup> and Jens Kurreck<sup>1,\*</sup>

<sup>1</sup>Institute of Biotechnology, Department of Applied Biochemistry, TUB 4/3–2, Technische Universität Berlin, Gustav-Meyer-Allee 25, 13355 Berlin, Germany, <sup>2</sup>Institute for Cell Biology and Immunology, University of Stuttgart, Allmandring 31, 70569 Stuttgart, Germany, <sup>3</sup>RiNA GmbH, Volmerstraße 9, 12489 Berlin, Germany, <sup>4</sup>Institute for Molecular Biosciences, University of Frankfurt, Max-von-Laue-Str. 9, 60438 Frankfurt/Main, Germany, <sup>5</sup>Max Planck Institute for Molecular Genetics, Ihnestr. 63–73, 14195 Berlin, Germany, <sup>6</sup>Max Planck Institute of Colloids and Interfaces, Am Mühlberg 1, 14476 Potsdam, Germany, <sup>7</sup>Institute of Biotechnology, Department of Bioprocess Engineering, ACK-24, Technische Universität Berlin, Ackerstraße 76, 13355 Berlin, Germany and <sup>8</sup>Institute of Chemistry and Biochemistry, Freie Universität Berlin, Thielallee 63, 14195 Berlin, Germany

Received August 14, 2013; Revised March 27, 2014; Accepted March 28, 2014

## ABSTRACT

Guanine quadruplex (G-quadruplex) motifs in the 5' untranslated region (5'-UTR) of mRNAs were recently shown to influence the efficiency of translation. In the present study, we investigate the interaction between cellular proteins and the G-quadruplexes located in two mRNAs (MMP16 and ARPC2). Formation of the G-quadruplexes was confirmed by biophysical characterization and the inhibitory activity on translation was shown by luciferase reporter assays. In experiments with whole cell extracts from different eukaryotic cell lines, G-quadruplex-binding proteins were isolated by pull-down assays and subsequently identified by matrix-assisted laser desorption ionization-time of flight mass spectrometry. The binding partners of the RNA G-quadruplexes we discovered included several heterogenous nuclear ribonucleoproteins, ribosomal proteins, and splicing factors, as well as other proteins that have previously not been described to interact with nucleic acids. While most of the proteins were specific for either of the investigated G-quadruplexes, some of them bound to both motifs. Selected candidate proteins were subsequently produced by recombinant expression and dissociation constants for the interaction between the proteins and RNA G-quadruplexes in the low nanomolar range were determined by surface plasmon resonance spectroscopy. The present study may thus help to increase our understanding of

the mechanisms by which G-quadruplexes regulate translation.

## INTRODUCTION

DNA and RNA molecules rich in guanine have the potential to form unusual secondary structures known as guanine quadruplexes (G-quadruplexes). These motifs comprise a structure of  $\pi$ -stacked tetrads formed by the coplanar arrangement of four Hoogsteen-paired guanines (1). The stability of G-quadruplex structures is significantly affected by the presence of a central cation, with potassium being the most suitable ion to stabilize G-quadruplexes (2). G-quadruplex motifs in DNA have been studied extensively for many years. For example, cellular G-quadruplexes in the telomeres have been shown to be important for the maintenance of eukaryotic chromosomes (3). Furthermore, G-quadruplex motifs in the promoter region of various genes involved in carcinogenesis are considered to control gene expression on the transcriptional level (for a review, see (4)). Recently, the existence of G-quadruplexes in human cells was visualized for the first time with a structure-specific antibody (5). Since RNA is usually single-stranded, the formation of G-quadruplex structures can be expected to occur even more easily than in double-stranded DNA, as base-pairing with complementary sequences will not compete with the folding process. Research in the last few years revealed that G-quadruplexes in RNA molecules fulfill a number of important cellular functions: telomeric repeat-containing RNA (TERRA) was described to regulate telomerase activity and further RNA G-quadruplexes are involved in termination of transcription, polyadenylation, regulation of alternative splicing, subcellular sorting of mRNAs and modulation of translation (2,6–9).

\*To whom correspondence should be addressed. Tel: +49 30 314 27582; Fax: +49 30 314 27 502; Email: jens.kurreck@tu-berlin.de

Of particular interest for the present study is the finding that G-quadruplexes in the 5'-UTRs of mRNAs interfere with translation. In 2007, Balasubramanian and coworkers (10) first reported that the presence of an RNA G-quadruplex in the 5'-UTR of an mRNA exhibits an inhibitory effect on translation. While these experiments were carried out in an *in vitro* translation assay, we were able to demonstrate that the evolutionarily conserved G-quadruplex in the mRNA of the zinc-finger protein of the cerebellum 1 (*Zic-1*) drastically inhibits translation in human cells without altering the mRNA level (11). Thus, the G-quadruplex only regulates the translation process and does not significantly influence transcription. Further experiments confirmed that the *Zic-1* protein was only synthesized in the absence of a 5'-UTR, whereas a 73 nucleotide-long fragment of the 5'-UTR containing the G-quadruplex motif inhibited expression of *Zic-1*.

A bioinformatic analysis carried out by Huppert et al. (12) revealed that around 2000 genes of the human genome contain potential G-quadruplexes in the 5'-UTR and that the G-quadruplex motifs are overrepresented at the 5'-end of the 5'-UTRs. In a systematic approach, Halder et al. analyzed the influence of loop size and number of GGG-repeats and observed a correlation between the inhibitory activity of G-quadruplexes and their stability, which depends on the number of nucleotides in the loop and repeat numbers (13). While the initial studies used isolated G-quadruplex motifs as a 5'-UTR, subsequent experiments confirmed the inhibitory effect of a G-quadruplex in the natural, full-length 5'-UTR of the matrix metalloproteinase *MT3-MMP*, also known as *MMP16* (14). Partial repression of translation by G-quadruplexes in the 5'-UTR of mRNAs has, in the meantime, been reported for various other genes such as the *fragile X mental retardation protein (FMRP)*, *Telomeric repeat-binding factor 2 (TRF2)*, the *human estrogen receptor  $\alpha$*  and *B-cell lymphoma 2 (Bcl-2)* (15–18). In an attempt to investigate the biological relevance of G-quadruplexes, such an element in the 5'-UTR of the cyclin D3 mRNA was shown to affect cell cycle progression (19). Surprisingly, a G-quadruplex in the 5'-UTR of the *Transforming Growth Factor-beta (TGF- $\beta$ 2)* was recently reported to augment translation (20). Furthermore, a G-quadruplex motif in the 3'-UTR of the proto-oncogene *Pim1* was reported to repress translation in a similar manner as the majority of the G-quadruplexes in the 5'-UTR (21). For a complete survey of RNA G-quadruplexes in mRNAs, see (5).

The exact mechanism of the translational inhibition by G-quadruplex motifs in the 5'-UTR has not yet been elucidated. In the absence of evidence for a more complex mechanism, it is reasonable to hypothesize that the stable structure sterically blocks the ribosome. Most of the G-quadruplex motifs in the 5'-UTRs of mRNAs were found to repress translation by approximately 50% and only a few cases have been reported to date, in which a stronger inhibitory effect of up to 80% has been observed (2,22). It is therefore likely that an equilibrium exists between the folded state, in which the G-quadruplex is formed and translation is prevented, and an open state, which allows the ribosome to perform protein synthesis.

A number of small molecular compounds have been characterized that modulate the stability of G-quadruplexes.

The porphyrin TmPyP4 (5,10,15,20-tetra(*N*-methyl-4-pyridyl)porphyrin) was found to stabilize the DNA G-quadruplex of the c-myc promoter, thereby repressing transcription (23). Interestingly, an opposite effect was observed for G-quadruplexes composed of RNA: TmPyP4 unfolds the extremely stable G-quadruplex of *MMP16* and alleviates its repressive effect (24). As a consequence, translation of the coding sequence downstream of the G-quadruplex was enhanced. In a similar manner, the destabilizing effect of TmPyP4 on the G-quadruplex of the *fragile X premutation* mRNA resulted in an increase of translation (25). In contrast, small molecules such as bisquinolinium compounds and a pyridine-2,6-bisquinolino-dicarboxamide derivative named RR110 were found to stabilize G-quadruplexes in the mRNAs, thereby repressing translation (26,27).

Several studies have addressed interactions between proteins and G-quadruplexes in telomeres and promoters. For example, affinity purification assays revealed the multifunctional phosphoprotein nucleolin to interact with the DNA G-quadruplex in the c-myc promoter (28). Binding of nucleolin was found to stabilize the G-quadruplex structure and to reduce the promoter activity. While considerable knowledge has been accumulated about the interaction between DNA G-quadruplex motifs and proteins, only very little is known about protein binding to G-quadruplex-forming RNA sequences (29). On the RNA level, FMRP was described to bind to a G-quadruplex motif in the coding region of its own mRNA, thereby inhibiting translation by a negative feedback loop (30). Furthermore, pull-down assays, followed by matrix-assisted laser desorption ionization-time of flight (MALDI-TOF) mass spectrometry (MS), were used to identify proteins that interact with the G-rich TERRA (31). Among the proteins that bind to the TERRA transcript were members of the heterogeneous nuclear ribonucleoprotein (hnRNP) family. To the best of our knowledge, no studies have been carried out to date to identify proteins that bind to the G-quadruplex motifs in 5'-UTRs.

The aim of the present study was therefore to characterize interactions between cellular proteins and G-quadruplexes located in the 5'-UTRs of mRNAs. We chose two factors whose protein levels were shown to play a role in tumor pathology. *MMP16* is a member of the family of matrix metalloproteinases that are involved in the degradation of extracellular matrix proteins. Upregulation of *MMP16* has been described to be associated with the invasiveness of cancer cell (32) and angiogenesis in tumors (33). The second factor under investigation, the actin-related protein 2 (ARPC2), is one of the seven subunits of the ARP2/3 protein complex, which is involved in actin cytoskeleton branching and thereby contributes to cell migration. While cell migration is essential in several physiological biological processes, its aberrant activation may also lead to the formation of metastases in cancer. Silencing of the ARP2/3 complex was recently shown to disturb pancreatic cancer cell migration (34). To fully understand their relevance in tumor biology, it is thus important to elucidate regulation mechanisms that influence the protein levels of these two factors.

## MATERIALS AND METHODS

### Cell culture

HeLa and HEK293 cells were cultured in low glucose Dulbecco's modified Eagle's medium (PAA Laboratories GmbH, Cölbe, Germany) containing 10% fetal bovine serum, 2 mM glutamine, non-essential amino acids and the antibiotics penicillin and streptomycin. HEK293 medium was supplemented with 4.5 mg/ml glucose. Cells were grown at 37°C in a humidified atmosphere with 5% CO<sub>2</sub>.

### Synthetic RNA Oligonucleotides

The G-quadruplex-forming sequences of the MMP16 and ARPC2 RNAs and their mutated controls were purchased from Purimex GmbH, Grebenstein, Germany:

MMP16 GQ, AACGAGGGAGGGAGGGAGAGGGAG  
AGA-biotin  
MMP16 mut, AACGAGAGAGAGAGAGAGAGAGAG  
AGA-biotin  
ARPC2 GQ, AGCCGGGGGCGUGGGCGGGGACCG  
GGCUUGU-biotin  
ARPC2 mut, AGCCGUAGACUGAGCGAAGACCGA  
GCUUGU-biotin  
ARPC2-2xG, AGCCGUAGACUGGGCGAAGACCGG  
GCUUGU-biotin.

All oligonucleotides were biotinylated at their 3' ends and polyacrylamide gel electrophoresis (PAGE) purified. The underlined nucleotides represent the core quadruplex forming sequences. MMP16 mut and ARPC2 mut were used to exclude proteins with a general RNA-binding capacity, while ARPC2-2xG was included to identify proteins that preferentially bind to G-rich sequences rather than G-quadruplex structures.

### Quadruplex formation

Quadruplex formation was carried out by dissolving biotinylated RNA oligonucleotides in folding buffer containing 10 mM Tris-HCl, pH 7.5, 100 mM KCl and 0.1 mM EDTA. Formation of secondary structures was achieved by heating the RNA to 95°C for 5 min in a thermocycler and cooling it down to 4°C in 2°C per minute steps.

### CD and UV spectroscopy

Prior to CD experiments, quadruplex formation of 1 μM RNA was carried out. Circular Dichroism (CD) experiments were performed at 20°C using a JASCO J-810 spectropolarimeter (JASCO, Gross-Umstadt, Germany) equipped with a Peltier temperature controller. Spectra were recorded in 1 mm quartz cuvettes. CD scans were taken from 220 to 320 nm in duplicates each with four accumulations and their average was calculated. A CD spectrum of the buffer was recorded and subtracted from the spectrum obtained for the RNA-containing solution.

For UV melting and annealing studies, RNA samples were prepared at 1 μM concentration in 10 mM Tris-HCl, pH 7.5 with KCl (10 mM and 100 mM). UV annealing and

melting studies were carried out on a JASCO V-650 UV-visible spectrophotometer equipped with a Peltier temperature controller. Samples were heated to 95°C and cooled down to 20°C at a 0.2°C min<sup>-1</sup> temperature gradient, and absorption data recorded at 295 nm were collected every 0.5 min on both annealing and melting steps. The wavelength of 295 nm was previously shown to yield more precise results than 260 nm when studying G-quadruplexes (35,36). The annealing and melting cycles were performed in duplicate in three independent experiments.

### RT-PCR experiments

Total cellular RNA was extracted from HeLa and HEK293 cells using the RNeasy mini kit (QIAGEN, Hilden, Germany). To remove DNA contaminations, DNase digestion was performed with RQ1 RNase-Free DNase (Promega, Mannheim, Germany). cDNA was synthesized with Revert Aid M-MuLV Reverse Transcriptase (Fermentas, St. Leon-Rot, Germany) and random hexamer primers from 1 μg RNA template. Polymerase chain reaction (PCR) samples contained 1 Unit Taq Polymerase (Rapidozym, Berlin, Germany), 0.2 mM dNTPs, 1 μl cDNA template and 1 μM of each primer. The primer pairs MMP16 fw/rev or ARPC2 fw/rev were used:

MMP16 fw, AAAATGGCAAACGTGATGTGG  
MMP16 rev, GGCTCATCTGAGTCAAATGG  
ARPC2 fw, AGAAGAGGGCAAGGAAGGAG  
ARPC2 rev, TGGCTAAAGAGGACCTGTGG.

Amplification was performed for 30 cycles. PCR products were separated on 1.5% agarose gels and visualized by ethidium bromide staining.

### Dual luciferase assays

To construct the plasmids psiCHECK-2-MMP16-GQ, psiCHECK-2-MMP16-mut, psiCHECK-2-ARPC2-GQ and psiCHECK-2-ARPC2-mut, the following DNA oligonucleotides containing NheI restriction sites were used:

MMP16 GQ sense CTAGAACGAGGGAGGGAGGG  
AGAGGGAGAGA  
MMP16 GQ antisense CTAGTCTCTCCCTCTCCCTC  
CCTCCCTCGTT  
MMP16 mut sense CTAGAACGAGAGAGAGAGAGAG  
AGAGAGAGA  
MMP16 mut antisense CTAGTCTCTCTCTCTCTCTC  
TCTCTCTCGTT  
ARPC2 GQ sense CTAGAGCCGGGGGCTGGGCGGG  
GACCGGGCTTGT  
ARPC2 GQ antisense CTAGACAAGCCCGGTCCCCG  
CCCAGCCCCGGCT  
ARPC2 mut sense CTAGAGCCGTAGACTGAGCGAA  
GACCGAGCTTGT  
ARPC2 mut antisense CTAGACAAGCTCGGTCTTCG  
CTCAGTCTACGGCT.

Each pair of sense and antisense oligonucleotide was annealed and ligated into the unique NheI restriction site of



the psiCHECK-2 plasmid (Promega) upstream of the *Renilla* luciferase start codon.

For the dual luciferase assay, subconfluent HEK293 cells in 24-well plates were transfected with 0.8  $\mu$ g of psiCHECK-2 reporter plasmids described above with Lipofectamine 2000 (Life Technologies GmbH, Darmstadt, Germany) according to the manufacturer's instructions. Activities of firefly and *Renilla* luciferase were measured 24 h after transfection using the Dual-Luciferase Reporter Assay Kit (Promega) on a luminometer (Lumat LB, Berthold, Pforzheim, Germany).

### Pull-down assays and MALDI-TOF mass spectrometry

For the preparation of whole cell extracts, confluent HeLa or HEK293 cell cultures were harvested by trypsination, washed with phosphate buffered saline (PAA Laboratories GmbH) and resuspended in one volume of buffer A (10 mM Tris-HCl, pH 7.5, 10 mM KCl, 1.5 mM MgCl<sub>2</sub>, 0.5 mM dithiothreitol (DTT)). After 15 min incubation on ice, cells were pelleted and resuspended in two volumes buffer A and one volume buffer B (300 mM Tris-HCl, pH 7.5, 1.4 M KCl, 3 mM MgCl<sub>2</sub>) containing protease inhibitor (Complete, Roche, Mannheim, Germany). Cells were disrupted by sonification and insoluble cell matter was removed by centrifugation (10 000 x g, 20 min, 4°C). The resulting protein extract was dialyzed against buffer C (20 mM Tris-HCl, pH 7.5, 100 mM KCl, 20% (v/v) glycerol, 0.5 mM DTT, 0.2 mM Na-EDTA) overnight at 4°C.

The whole cell extracts were used for pull-down assays with the G-quadruplex-forming RNA oligonucleotides. To avoid unspecific binding, streptavidin agarose beads (Sigma-Aldrich Chemie GmbH, Taufkirchen, Germany) were initially incubated in blocking buffer containing 10 mM Tris-HCl, pH 7.5, 100 mM KCl, 0.1 mM Na-EDTA, 1 mM DTT, 0.01% (v/v) Triton X-100, 0.1% (w/v) bovine serum albumin and 0.02% (w/v) tRNA from *Saccharomyces cerevisiae* (Sigma-Aldrich) for 1 h at 4°C on a rotating wheel. Folded 3'-biotinylated RNA oligonucleotides were coupled to 40  $\mu$ l of the solution containing the streptavidin agarose beads at 4°C overnight. 400  $\mu$ l of the protein extract was then incubated with streptavidin agarose beads for 1 h at 30°C. Bound proteins were eluted from the RNA by washing the beads with increasing KCl concentrations (400–2600 mM). Each fraction was concentrated by trichloroacetic acid precipitation and analyzed on a 10% sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE), stained with colloidal Coomassie staining as described in (37).

Peptides were obtained by trypsin in-gel digestion as described previously (38) and peptide masses were analyzed by MALDI-TOF MS using an Ultraflex-II TOF/TOF instrument (Bruker Daltonics, Bremen, Germany) equipped with a 200 Hz solid-state Smart beam™ laser. The mass spectrometer was operated in the positive reflector mode. Mass spectra were acquired over an m/z range of 600–4000.  $\alpha$ -cyano-4-hydroxycinnamic acid was used as the matrix and protein digest samples were spotted using the dried droplet technique. MS/MS spectra of selected peptides were acquired in the LIFT mode (39). Database searches were performed using Mascot (Matrix Science Ltd., <http://www.matrixscience.com>).

Mass tolerance was set at  $\pm$  75 ppm unless otherwise indicated (Supplementary Table S1) and we allowed for one missed cleavage.

### Cloning, expression and purification of recombinant proteins

For expression of recombinant proteins, human cDNA sequences were cloned into the pET28a expression vector (Merck, Darmstadt, Germany) using the BamHI and NdeI restriction sites, so that the expressed proteins contained an N-terminal His-Tag. The following oligonucleotides were used to amplify the coding sequences of nucleolin, EFHD2, SRSF1 and U2AF65:

Nucleolin  $\Delta$ 1–283 fw, GGAATTCATATGGCCAAACAGAAAGCAGC  
 Nucleolin rev, GCGGGATCCCTATTCAAACCTTCGTCTTC  
 EFHD2 fw, GGAATTCATATGGCCACGGACGAGC  
 EFHD2 rev, GTAGGATCCTCAGGCTCGGCCTCC  
 SRSF1 fw, GGAATTCATATGTCGGGAGGTGGTG  
 SRSF1 rev, GCGGGATCCTTATGTACGAGAGCGA  
 U2AF65 fw, GGAATTCATATGTCCGACTTCGACGAG  
 U2AF65 rev, GTAGGATCCCTACCAGAAGTCCCGG.

Nucleolin was expressed as a variant consisting of amino acid residues 284–709 (in the following denoted as  $\Delta$ nucleolin), including all four RNA binding domains (RBDs) and the C-terminal domain.

Proteins were expressed in *Escherichia coli* BL21 DE3 using EnPresso® B tablet medium (BioSilta, Oulu, Finland) according to the manufacturer's instructions at 30°C. Expression was induced with 0.2 mM final concentration isopropyl  $\beta$ -D-1-thiogalactopyranoside and cells were harvested 24 h after induction. Cell pellets were harvested by centrifugation (20 min, 4000 x g), resuspended in lysis buffer (50 mM NaH<sub>2</sub>PO<sub>4</sub>, pH 8, 300 mM NaCl, 10 mM imidazole, 2 M NaCl, 2% Triton X-100, 5% glycerol) and disrupted by sonification. Crude extract was cleared by centrifugation for 30 min at 10 000 g. Proteins were purified with Protino® Ni-NTA Agarose (Macherey-Nagel, Düren, Germany) in gravity flow columns. Non-bound proteins were removed by washing with 50 mM NaH<sub>2</sub>PO<sub>4</sub>, pH 8, 300 mM NaCl, 20 mM imidazole, 2 M NaCl, 2% Triton X-100, 5% glycerol. Proteins were eluted in 50 mM NaH<sub>2</sub>PO<sub>4</sub>, pH 8.0, 300 mM NaCl, 250 mM imidazole, 2 M NaCl, 2% Triton X-100, 5% glycerol and dialyzed against 10 mM HEPES, pH 7.5, 150 mM NaCl, 10% glycerol, 1 mM DTT and 0.1% Triton X-100. Protein concentrations were determined using the Bicinchoninic Acid (BCA) Protein Assay Reagent (Thermo Fisher Scientific, Bonn, Germany).

### Surface plasmon resonance

Surface plasmon resonance (SPR) experiments were performed on the Biacore-X or Biacore 2000 platform (GE Healthcare Europe, Freiburg, Germany). All experiments were performed at 25 °C. Proteins were immobilized on carboxymethyl dextran (CMD 500 L) sensor chips (Xantec, Duesseldorf, Germany) using amine coupling. Immobilization was performed in the SPR running buffer

HBS-EP (0.01 M HEPES pH 7.4, 0.15 M NaCl, 3 mM EDTA, 0.005% v/v Surfactant P20; GE Healthcare Europe) using a flow rate of 5  $\mu$ l/min. The surface of the CMD 500 chip was activated with 35  $\mu$ l 0.02 M 1-ethyl-3-(3-dimethylaminopropyl)-carbodiimide and 0.05 M N-hydroxysuccinimide. Each Protein diluted in HBS-EP buffer was injected until around 1000 response units were immobilized, the surface was blocked with 35  $\mu$ l 1 M ethanolamine (pH 8.5). An activated and blocked channel without protein served as reference. Binding analysis was conducted at a flow rate of 20  $\mu$ l/min in a 1:10 mixture of binding buffer (10 mM Tris-HCl, pH 7.5, 150 mM KCl) and HBS-EP to match the buffer in which the oligonucleotides were diluted and to minimize bulk shift effects. Both channels of the flow cell were injected with 30  $\mu$ l of the RNA solutions in a range of 6–1000 nM. After each RNA injection, the chip surface was regenerated by injection of 10  $\mu$ l 1 M KCl. For background correction, each response signal was adjusted by subtraction of the reference cell signal as well as by subtraction of adjusted buffer injection signals. Association and dissociation rates and constants of the RNA-protein complexes were determined using BIAevaluation software (GE Healthcare Europe). The data were globally fitted to a 1:1 interaction model (Langmuir binding).

## RESULTS

### Biophysical characterization of the MMP16 and ARPC2 RNA G-quadruplexes

Two G-quadruplex motifs were chosen to study the interaction between RNA structures and cellular proteins. The G-quadruplex located in the 5'-UTR of MMP16, also known as MT3-MMP, was previously shown to form a very stable structure (14). In contrast, the G-quadruplex in the 5'-UTR of the mRNA of ARPC2 was only predicted by bioinformatics means (12) and has not yet been experimentally verified.

The two RNA G-quadruplex motifs were analyzed by the software quadruplex-forming G-Rich Sequences (QGRS) Mapper (<http://bioinformatics.ramapo.edu/QGRS/index.php>). QGRS Mapper is a software program that generates information about composition and distribution of putative QGRS and provides a G-Score as measurement of G-quadruplex probability (40). Both RNA sequences are able to form very stable G-quadruplex structures with three guanine tetrads. The ARPC2 G-quadruplex can form up to four variants of the stable structure with a G-Score of 39 and 41 and nucleotide length of 19 and 20 nt, while the MMP16 G-quadruplex can form only one variant with a G-Score of 40 and nucleotide length of 17 nt.

For pull-down assays, biotinylated oligonucleotides were used that consisted of the G-quadruplex-forming sequences and flanking nucleotides, in order to simulate the natural environment of the motif and to have a spacer between the biotin and the G-quadruplex motif. In the control oligonucleotides, several guanosines were substituted so that they cannot fold into a G-quadruplex conformation. These controls were used to exclude proteins with a general RNA-binding capacity. An additional control with a G-rich sequence (ARPC2-2xG) was used to identify proteins that

bind to G-stretches but are not specific for G-quadruplex motifs.

The first step was to analyze the formation of the G-quadruplex structure of the respective oligonucleotides by CD and UV spectroscopy. CD spectra of the G-rich MMP16 RNA sequence in 10 mM KCl showed a positive peak at 267 nm and a negative peak at 245 nm (Figure 1A) indicating the presence of a parallel G-quadruplex structure (10,41). The mutated variant lacks these characteristics since it cannot form a G-quadruplex structure. The CD data are in line with a previous report on this G-quadruplex (14). UV melting experiments carried out at 10 mM KCl showed a reversible sigmoidal transition at 295 nm (Figure 1A) indicative of the formation of a G-quadruplex structure (35). The  $T_m$  was found to be  $69.2 \pm 0.2^\circ\text{C}$  at a KCl concentration of 10 mM. At higher salt concentrations (100 mM KCl) the structure could not be unfolded, demonstrating the high stability of the MMP16 RNA G-quadruplex (data not shown). No distinct transition was observed for the control oligonucleotide, as would be expected if it does not form a G-quadruplex.

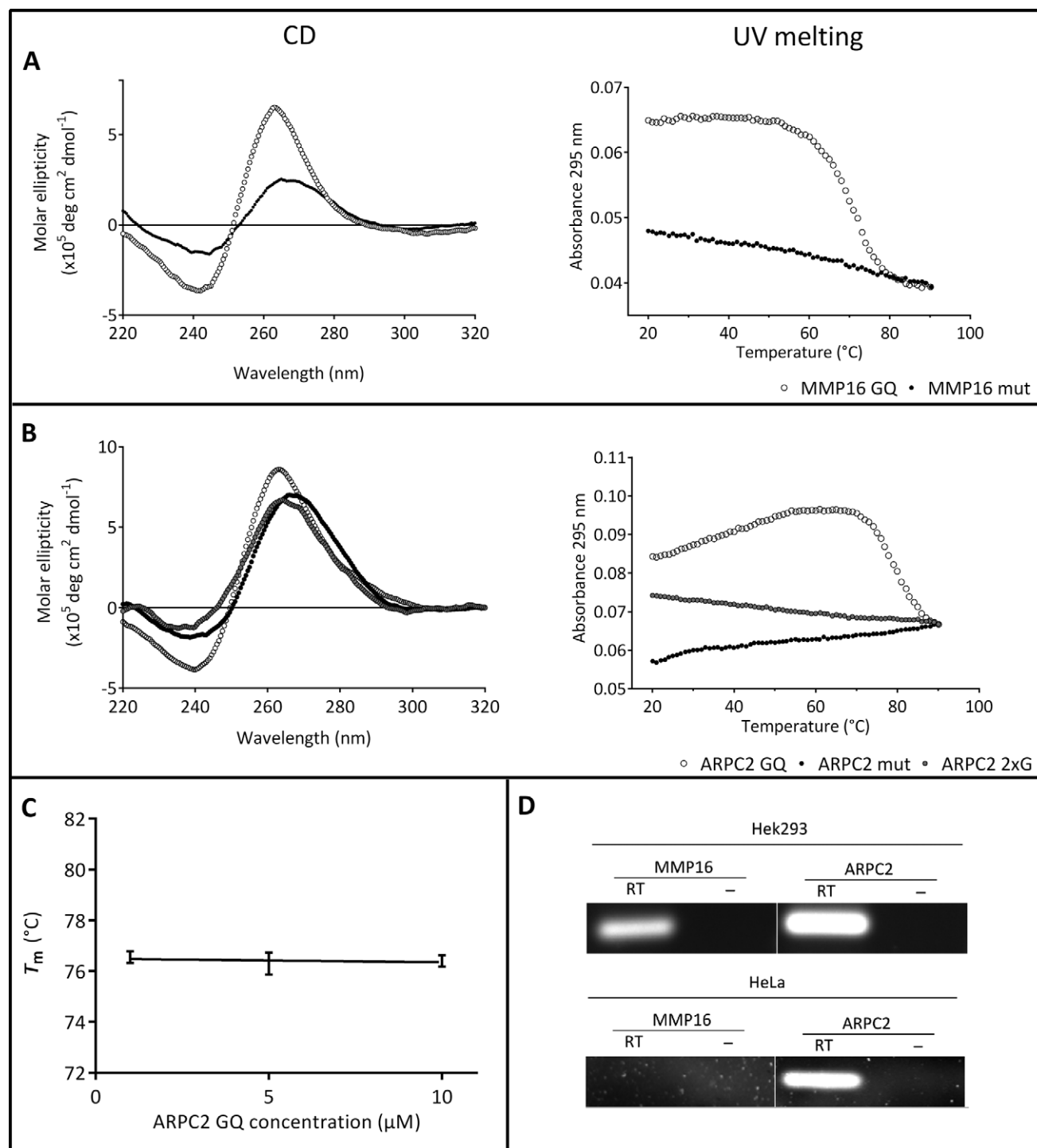
For the G-rich oligonucleotide of the ARPC2 mRNA, the CD spectra showed a positive peak at 267 nm and a negative peak at 238 nm (Figure 1B), which again is indicative of the presence of a parallel G-quadruplex structure. A slight deviation from other reported values (10,11,41) can be due to the presence of flanking bases on either side of the core G-quadruplex-forming sequence that may form alternative structures. The CD spectra of the mutated controls (ARPC2 mut and ARPC2-2xG) showed positive peaks around 270 nm and negative peaks around 230–240 nm. As the spectra of the mutated sequences do not differ significantly from the CD spectrum of the putative G-quadruplex-forming sequence, UV melting analyses at 295 nm in 100 mM KCl were carried out to gain further information (Figure 1D). These experiments showed a sigmoidal transition with a  $T_m$  of  $76.5 \pm 0.2^\circ\text{C}$  for the G-rich sequence of ARPC2, as expected for a G-quadruplex motif. Neither of the mutated controls showed a sigmoidal transition, indicating so that they do not fold into a G-quadruplex structure.

Since the ARPC2 G-quadruplex has not been previously described, we characterized it further by measuring the concentration dependency of  $T_m$ . As can be seen in Figure 1C,  $T_m$  does not vary in the concentration range under investigation. This finding further supports the hypothesis that the sequence forms an intramolecular G-quadruplex.

Taken together, all data obtained from UV melting and CD experiments are in good agreement with well-known characteristics for parallel G-quadruplex motifs. It is therefore reasonable to assume that both RNA oligonucleotides form G-quadruplex structures. The G-quadruplex of the ARPC2 G-rich sequence, however, seems to be less stable than that of the MMP16 G-quadruplex since it has to compete with alternative conformations.

### Expression of MMP16 and ARPC2 in HEK293 cells

In order to obtain biologically relevant data, pull-down assays were carried out with a cell line that endogenously expresses the two genes under investigation. Reverse



**Figure 1.** Basic characterization of the G-quadruplex motifs under investigation. (A) and (B) CD spectra and melting curves of the G-quadruplex motifs and their mutated control oligonucleotides (A: MMP16; B: ARPC2). (C) Melting temperature at different concentrations of the ARPC2 G-quadruplex. All melting experiments were carried out at 100 mM KCl, except for the MMP16 G-quadruplex-forming sequence, since it did not unfold in the measured temperature range. (D) RT-PCR experiments to confirm expression of MMP16 and ARPC2 in HEK293 cells. Additional experiments were carried out with HeLa cells. Control experiments without reverse transcription are shown (-).

transcriptase-PCR (RT-PCR) experiments confirmed that HEK293 cells express MMP16 as well as ARPC2. This cell line thus represents an orthotopic model to study the interaction between the mRNAs of MMP16 and ARPC2, respectively, and cellular proteins. Further experiments were

carried out with extracts from HeLa cells. In this case, only the mRNA of ARPC2, but not of MMP16, could be detected by RT-PCR (Figure 1D).



### Inhibition of translation by the MMP16 and ARPC2 RNA G-quadruplexes

The next step was to analyze the inhibitory activity of the G-quadruplex motifs. These studies were carried out as described before for Zic-1 (11). Dual luciferase assays were performed with the psiCHECK-2 vector that allows simultaneous expression of *Renilla* and firefly luciferase from a single plasmid. DNA sequences encoding the RNA G-quadruplexes and their mutated controls, respectively, were cloned upstream of the *Renilla* luciferase. The vector constructs were transfected into HEK293 cells and 24 h after transfection cells were harvested. Dual luciferase measurements were carried out and the ratios of luciferase activity for the *Renilla* and firefly luciferase were normalized to the value obtained for the empty psiCHECK-2 vector. As can be seen in Figure 2A, both, the G-quadruplex motifs of MMP16 and ARPC2, repress translation by approximately 50% and 40%, respectively. The control for the MMP16 G-quadruplex also had a minor effect on the translation efficiency. The partial inhibition of translation by the G-quadruplexes is in the range of what was observed for other G-quadruplexes (2). Morris and Basu (14) found a slightly more pronounced repression of translation by the MMP16 G-quadruplex of 60%, which may be due to the fact that they used the full length 5'-UTR, while we only used the isolated G-quadruplex motif with a few flanking nucleotides.

For the less well characterized G-quadruplex of ARPC2, additional experiments were carried out with the full-length 5'-UTR to investigate whether the G-quadruplex also inhibits translation in its natural sequence environment. The results obtained with the full-length 5'-UTR (Figure 2B) were very similar to those described for the isolated G-quadruplex motif: the G-quadruplex sequence inhibited translation by approximately 50%, while a control with point mutations that prevent formation of the G-quadruplex only had a negligible effect. This finding supports the assumption that a G-quadruplex structure is formed in the 5'-UTR of the ARPC2 that inhibits translation.

### Identification of RNA G-quadruplex binding proteins by pull-down assays

The central aim of the present study was to isolate and identify proteins that bind to the RNA G-quadruplex motifs. Therefore, pull-down assays with whole cell extract and the G-quadruplex-forming RNA oligonucleotides or mutated control oligonucleotides were performed. Proteins were eluted with increasing KCl concentrations and further analyzed by SDS-PAGE. For MS, the following protein bands were chosen: (i) proteins that only bound to the G-quadruplex-forming RNA but not to the mutated sequences and (ii) proteins that bound to the mutated sequences of MMP16 and ARPC2.

The experiments were conducted with extracts from HEK293 cells and were repeated three times. Figure 3 shows examples for pull-down assays with the MMP16 (A) and ARPC2 (B) G-quadruplex RNA oligonucleotides. Bands of proteins binding specifically to either of the G-quadruplex-forming RNAs were chosen for further analyses by MS. Bands appearing in both the G-quadruplex oligonucleotide

(GQ) as well as the mutated control (M) lanes are most likely unspecific RNA binders.

To reduce proteins with a general capacity to bind to RNA, in an alternative approach, the cellular extracts were initially incubated with the control oligonucleotides. Subsequently, the depleted extracts were incubated with the G-quadruplex-forming oligonucleotides to pull down the specific binding partners. As can be seen in Figure 3 (lane GQ-M), these experiments mostly confirmed the proteins identified with the initial strategy.

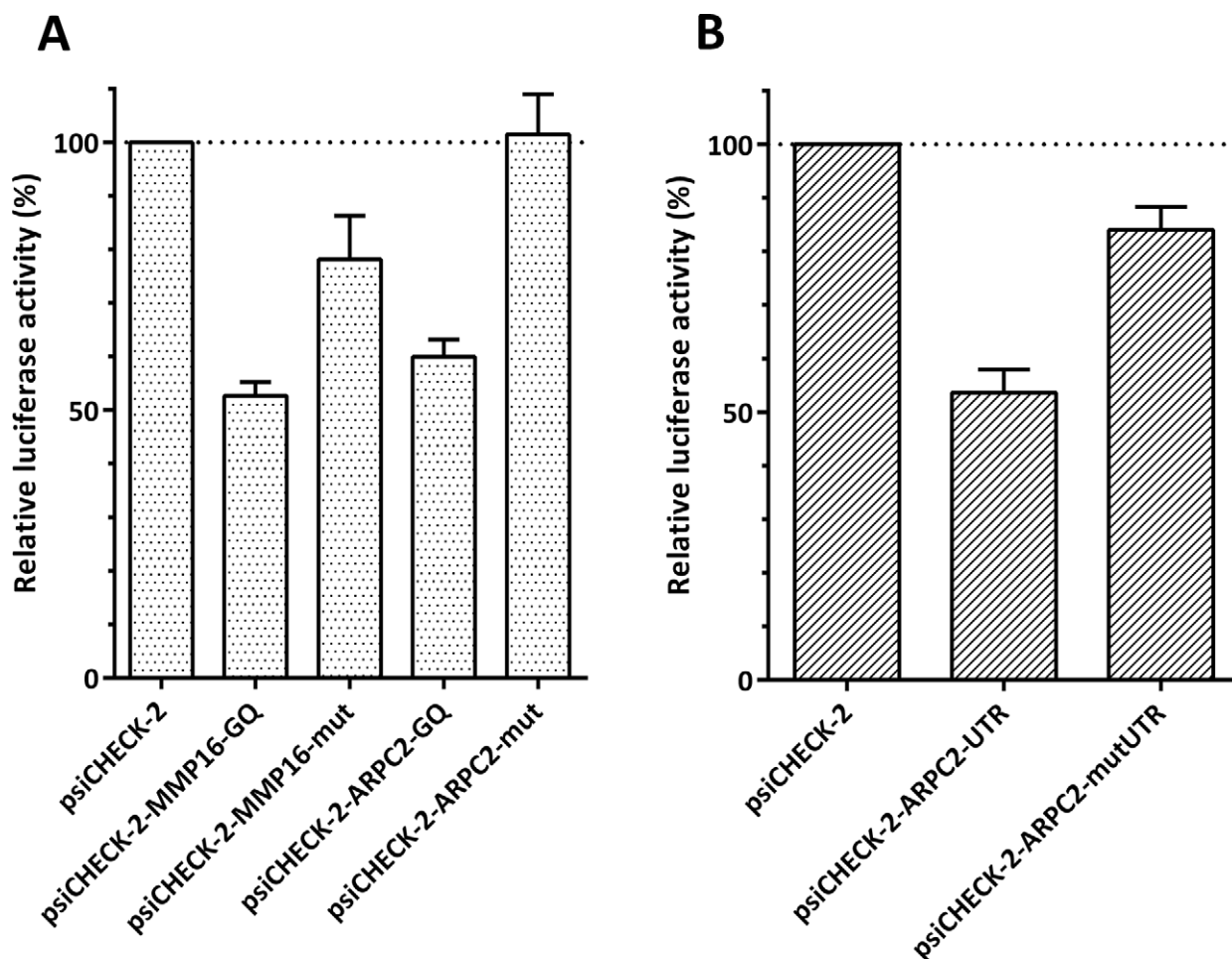
Table 1 summarizes the proteins identified to specifically bind to G-quadruplex structures in the pull-down assays. The detailed MS data are given as supplementary (Supplementary Table S1). The number of times a protein was identified in repeated experiments is given. A much larger number of binding candidates were found for the ARPC2 G-quadruplex than for the MMP16 G-quadruplex. The higher number of identified proteins for the ARPC2 G-quadruplex could be a result of the variety of the G-quadruplex motifs putatively formed by ARPC2 compared to MMP16. As stated before, QGRS mapper analysis revealed that ARPC2 can form up to four variants of the G-quadruplex structure.

Several ribosomal proteins (RPs) were found to interact with either of the two G-quadruplexes. Further interaction partners belong to the class of proteins involved in splicing, including the serine/arginine-rich splicing factors (SRSF) 1 and 9, the splicing factor U2AF65 and the nuclease-sensitive element-binding protein 1 (YB-1). A distinct binder of the MMP16 G-quadruplex was the mitochondrial nicotinamide adenine dinucleotide (NAD)-dependent malic enzyme (ME2). Most of the identified proteins were specific for one of the two G-quadruplex structures. Only the splicing factors U2AF65 and YB-1 bound to both motifs. Further control experiments revealed that the latter candidate (YB-1) has the general capacity of binding to G-rich sequences (see below).

The next aim was to test whether the RNA G-quadruplex-protein interaction is cell-type specific. To this end, the experiments were repeated with HeLa cells. The experiments with HeLa cells were repeated twice. As can be seen in Table 1, several binding proteins such as nucleolin and U2AF65 identified in the experiments with extracts from HEK293 cells were confirmed for the HeLa extracts and only one additional RP (RPS5) was newly identified to bind to the MMP16 G-quadruplex. Thus, similar interactions presumably occur in different cell types.

We also identified the proteins represented by some particularly prominent bands in the control oligonucleotides by MALDI-TOF. Interestingly, all of the identified proteins belong to the class of hnRNPs, and they were similar for both controls: hnRNP A3, A2/B1, A1. None of these proteins bound to the G-quadruplex-forming oligonucleotides. Only actin was identified as a protein that binds to the G-quadruplex motifs as well as to the controls. This finding is likely to reflect the high abundance of this cytoskeleton protein, rather than specific binding.

The experiments described so far discriminate between proteins binding to sequences that are capable of forming G-quadruplex structures and those with a general tendency to bind to RNA. To exclude proteins that bind to G-rich sequences, additional experiments were performed with a



**Figure 2.** Inhibition of translation by the G-quadruplex motifs of MMP16 and ARPC2, as determined by dual luciferase assays. (A) Dual luciferase assays for the isolated G-quadruplex motifs and (B) for the full-length 5'-UTR of ARPC2. The ratio of *Renilla* and firefly luciferase activity is normalized to the value of the empty psiCHECK-2 vector. The G-quadruplex motif in the 5'-UTR of *Renilla* luciferase reduces enzyme synthesis. Values given are averages  $\pm$  standard deviations of three experiments.

second control sequence with a higher guanine content and two G-stretches (ARPC2-2xG). As can be seen in Figure 3C, only very few proteins bound to this additional control at higher salt concentrations. The proteins of the most prominent bands marked as f and g, respectively, were identified as YB-1 and YB-1 in complex with actin. YB-1 thus does not seem to bind specifically to G-quadruplex structures, but rather recognizes G-rich sequences.

#### Characterization of interactions between RNA G-quadruplexes and proteins

The binding parameters for U2AF65, nucleolin, SRSF1 and EFHD2 were determined by SPR spectroscopy. The proteins were recombinantly expressed with a His-tag in *E. coli* and purified by Ni-NTA-agarose affinity chromatography. Subsequently, the proteins were immobilized on a CMD sensor chip by amine coupling. Binding of the RNA oligonucleotides to the proteins was measured.

Since full-length nucleolin is not soluble when heterologously expressed in *E. coli*, a shortened version was expressed lacking the N-terminal domain (residues 1–283).

This protein, denoted as  $\Delta$ nucleolin, is composed of all four RBDs and the C-terminal domain. SPR measurements were carried out with  $\Delta$ nucleolin and a dissociation constant ( $K_D$ ) of 0.6 nM was determined for the ARPC2 G-quadruplex, while binding to the control oligonucleotide was more than 100-fold weaker, with a  $K_D$  of 83 nM (Table 2 and Figure 4A). Nucleolin was identified as a binding partner of the ARPC2 G-quadruplex, but to test if there is a general interaction with G-quadruplexes, we also examined its affinity for the MMP16 sequences. Surprisingly, here the protein bound tightly to the control as well as to the G-quadruplex oligonucleotide of MMP16 (Figure 4A), indicating that nucleolin is able to recognize diverse nucleic acid motifs.

The splicing factor U2AF65 was next investigated as a candidate for specific binding to G-quadruplexes, as it was found to bind to both G-quadruplexes in the pull-down assays. SPR measurements confirmed that it binds to the ARPC2 and MMP16 G-quadruplexes, while it does not interact with the mutated controls (Figure 4B). Dissociation constants of 1 nM for the ARPC2 G-quadruplex and 5 nM for the MMP16 G-quadruplex were determined (Table 2).



**Table 1.** Proteins identified to bind to the G-quadruplex motifs of the ARPC2 and MMP16 mRNA

	ARPC2 GQ		MMP16 GQ	
	Hek	HeLa	Hek	Hela
<b>Ribosomal Proteins</b>				
60S ribosomal protein L6 (RPL6 <sup>10</sup> )	2 <sup>a</sup>	1		
60S ribosomal protein L7 (RPL7 <sup>11</sup> )	2	1		
60S ribosomal protein L10 (RPL10 <sup>14</sup> )	2			
60S ribosomal protein L12 (RPL12)		1		
60S ribosomal protein L14 (RPL14 <sup>13</sup> )	1			
60S ribosomal protein L19 (RPL19)	1	1		
60S ribosomal protein L26 (RPL26 <sup>15</sup> )	1			
60S ribosomal protein L27a (RPL27a <sup>16</sup> )	1			
40S ribosomal protein S2 (RPS2 <sup>6</sup> )	1			
40S ribosomal protein S5 (RPS5)				1
40S ribosomal protein S6 (RPS6 <sup>12</sup> )	1			
40S ribosomal protein S9 (RPS9 <sup>14, 17</sup> )	2	1		
<b>Splicing Factors</b>				
Heterogeneous nuclear ribonucleoprotein F (hnRNPF <sup>5</sup> )	1			
Heterogeneous nuclear ribonucleoprotein H (hnRNPH <sup>4</sup> )	2			
Heterogeneous nuclear ribonucleoprotein U (hnRNPU)			2	1
Serine/arginine-rich splicing factor 1 (SRSF1 <sup>9, 10</sup> )	3	1		
Serine/arginine-rich splicing factor 9 (SRSF9 <sup>12, 13</sup> )	2	1		
Splicing factor U2AF (U2AF65 <sup>3</sup> )	1	1	2	
<b>Others</b>				
Y-box-binding protein 3	1			
EF-hand domain-containing protein D2 (EFHD2)	2	1		
NAD-dependent malic enzyme, mitochondrial (ME2 <sup>1</sup> )			3	
Nucleolin (Nucl <sup>7</sup> )	3	2		
Ras GTPase-activating protein-binding protein 2 (G3BP2)		1		
RNA binding protein 14 (RBM14 <sup>8</sup> )	2			
Tropomyosin alpha 4 (TPM4)	1			

<sup>a</sup>The number of times a protein was identified is given. Experiments with HEK293 extracts were carried out three times; confirmatory experiments with HeLa extracts were performed twice.

The superscripts refer to the bands marked in Figure 3.

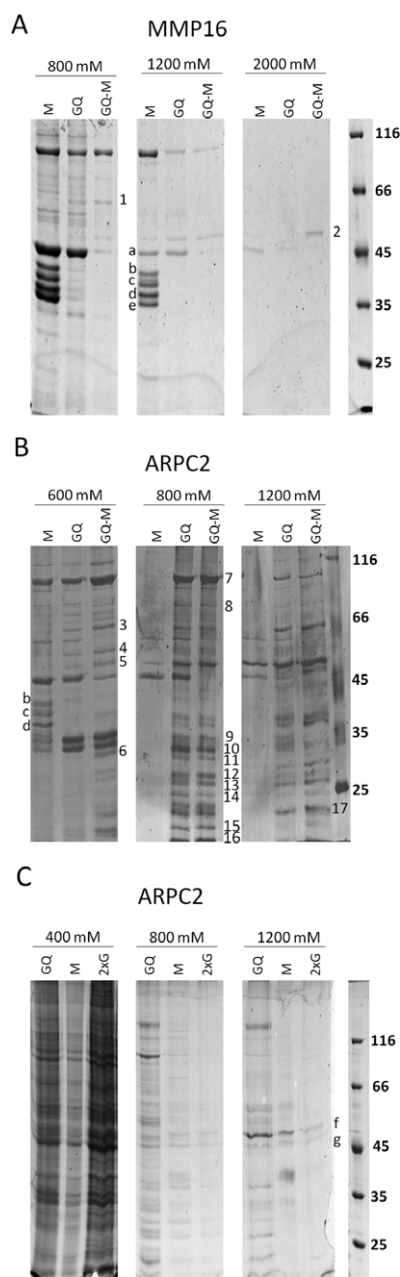
**Table 2.** Association ( $k_a$ ) and dissociation ( $k_d$ ) rates and dissociation constants ( $K_D$ ) determined by SPR spectroscopy; n. d.: not determined due to too weak interactions

Ligand	Analyte	$k_a$ ( $M^{-1}s^{-1}$ ) x $10^6$	$k_d$ ( $s^{-1}$ ) x $10^{-3}$	$K_D$ (nM)	$\chi^2$
$\Delta$ Nucleolin	ARPC2 GQ	8.7	5.1	0.6	0.96
	ARPC2 mut	0.5	41.4	83	1.61
	MMP16 GQ	0.92	1.9	2	0.79
	MMP16 mut	6	3.6	0.6	0.94
U2AF65	ARPC2 GQ	1.1	1.2	1	0.9
	ARPC2 mut			n. d.	
	MMP16 GQ	1.6	8.1	5	0.73
	MMP16 mut			n. d.	

Binding to both mutated control sequences was so weak that it was impossible to determine a  $K_D$ . U2AF65 can thus be considered to bind specifically to RNA G-quadruplex structures, while it does not seem to be specific for a certain G-quadruplex motif.

Additional experiments were carried out with the above-described, guanine-rich control oligonucleotide ARP2–2xG to confirm that the proteins interact specifically with the G-quadruplex structures. As can be seen in Figure 4C, ARPC2 2xG behaves similar to the fully mutated sequence, i.e. nucleolin and U2AF65 only bind to a sequence that is capable of forming a G-quadruplex structure, but not to a G-rich sequence. This finding further supports the hypothesis that the identified proteins bind to the G-quadruplex structure.

According to the pull-down assays, SRSF1 and EFHD2 were two prominent and specific interaction partners of the ARPC2 G-quadruplex. Since both proteins rapidly lost activity when coupled to the sensor chip, only a qualitative comparison of the binding to either of the two G-quadruplexes and the respective control sequences was conducted. In the assay, the mutated oligonucleotides were injected first to counteract the effect of the activity loss. Both proteins bound much stronger to the ARPC2 G-quadruplex than to the MMP16 G-quadruplex or to any of the control sequences (Figure 5). These proteins can therefore be regarded to be highly specific for a certain G-quadruplex motif.



**Figure 3.** Pull-down assays. (A) and (B) G-quadruplex motifs of the MMP16 (A) and ARPC2 (B) mRNAs were coupled to streptavidin agarose beads. Whole-cell extract of HEK293 cells was added to the beads. After several washing steps, proteins were eluted with buffers containing the indicated  $K^+$  concentrations. Proteins were analyzed by SDS-PAGE. Proteins that bind exclusively to the G-quadruplex-sequence (Q) are indicated. GQ-M represents an independent experiment, in which the HEK293 extract was first incubated with the mutated oligonucleotide and then with the G-quadruplex-forming RNA. The indicated proteins were identified by MS-based peptide mass fingerprinting. (C) Pull-down assay with the additional ARPC2 control oligonucleotide with two G-stretches (ARPC2-2xG). 1. ME2 (66 kDa); 2. YB-1 (36 kDa); 3. U2AF65 (54 kDa); 4. hnRNPH (50 kDa); 5. YB-1 (36 kDa) hnRNPF (46 kDa); 6. RPS2 (32 kDa); 7. Nucl (76 kDa); 8. RBM14 (70 kDa); 9. SRSF1 (28 kDa); 10. SRSF1 (28 kDa) RPS6 (29 kDa); 11. RPL7 (29 kDa); 12. SRSF9 (26 kDa), RPS6 (29 kDa); 13. SRSF9 (26 kDa), RPL14 (23.5 kDa); 14. RPL10 (25 kDa) RPS9 (23 kDa); 15. RPL26 (17 kDa); 16. RPL27a (17 kDa); 17. RPS9 (23 kDa); a. Actin (42 kDa); b. hnRNPA3 (39 kDa); c. hnRNPA2B1 (37 kDa) and hnRNPA3 (39 kDa); d. hnRNPA2B1 (37 kDa); e. hnRNPA1 (39 kDa); f. YB-1 (36 kDa); g. YB-1/Actin (36 kDa, 42 kDa).

## DISCUSSION

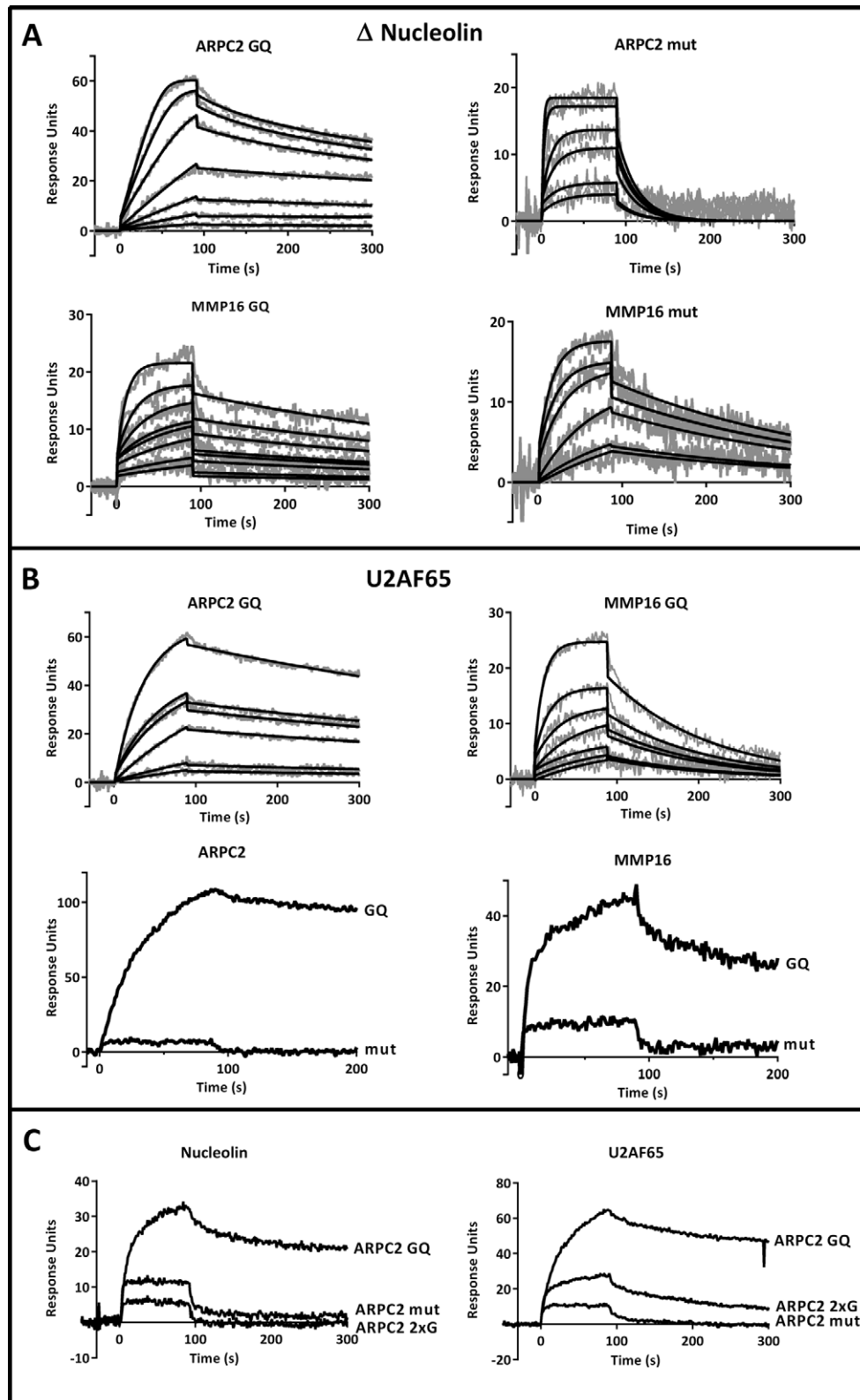
G-quadruplex motifs in RNA have only come into the focus of research in recent years (2,6,7,9). They have been reported to be involved in regulation of translation and to have a function in splicing (9). It is tempting to speculate that the stability and activity of G-quadruplexes might be modulated by cellular factors such as proteins. Indeed, G-quadruplexes have a great potential to bind proteins with high affinity and specificity as shown for aptamers, which often form a G-quadruplex structure. The identification and characterization of interactions between RNA G-quadruplexes and cellular proteins will help increasing our understanding of the physiological functions of these structural elements (6,8). We present the first systematic approach to identify proteins that bind to G-quadruplex elements in 5'-UTRs of mRNAs using pull-down assays and MALDI-TOF MS. Extensive controls were carried out to distinguish between proteins that bind specifically to G-quadruplex structures and those that have a general capacity to interact with RNA or with guanine-rich sequences.

In the present study, we focused on G-quadruplexes located in the 5'-UTR, which have been reported to modulate translation without affecting the mRNA level (8,22). Indeed, several of the identified proteins are involved in translation. A number of RPs like RPS5, RPS6 and RPS9 were found to bind to the ARPC2 or MMP16 G-quadruplex. All three are part of the 43S pre-initiation complex that moves along the mRNA when scanning for the start codon. For G-quadruplexes located in the 5'-UTR it can be assumed that, depending on the position, the quadruplex may inhibit recognition of the cap or start codon or the scanning of the 40S ribosomal subunit for the translational start (42). This could be due to steric blockage of the G-quadruplex structure itself or because of proteins binding to the quadruplex.

Some of the identified proteins are known to be involved in the formation of ribonucleoprotein complexes like the nucleolin-binding ribonucleoprotein complex (43) consisting, among others, of nucleolin, hnRNPs, RPL6, RPL7, RPL10, RPL27a, RPS6 and RPS9, all of them found as binders of the ARPC2 G-quadruplex. Without further analyses, it is difficult to differentiate between proteins interacting directly with the RNA G-quadruplex and proteins which are part of multimeric ribonucleoprotein complexes. In addition, YB-1 also bound to a G-rich sequence that cannot fold into a G-quadruplex structure. It is therefore likely that it recognizes stretches of guanines rather than G-quadruplexes.

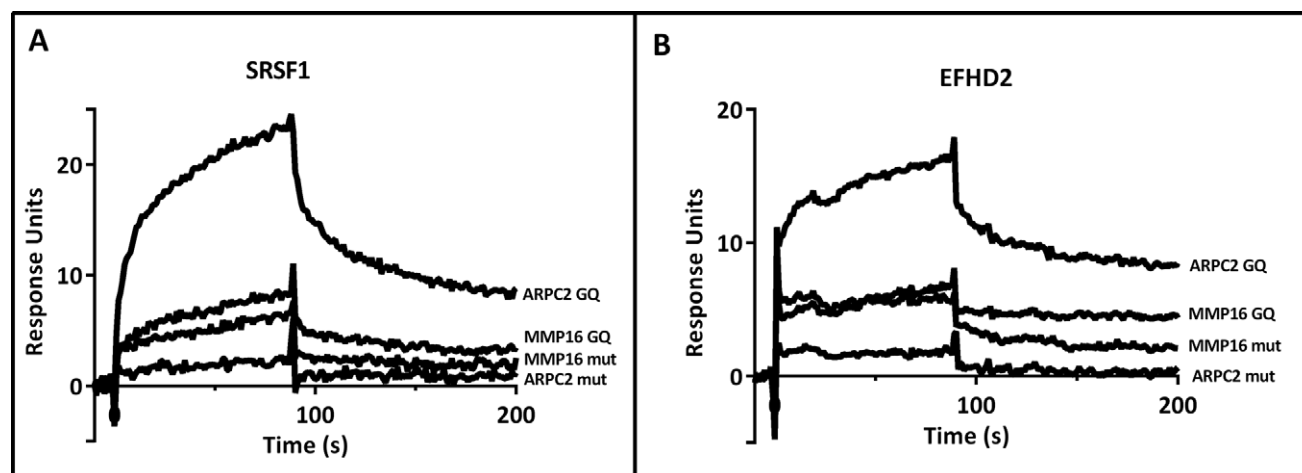
To gain a deeper insight,  $\Delta$ nucleolin was heterologously expressed in *E. coli* and purified via affinity chromatography. Binding constants were determined by SPR to be in a low nanomolar range for the ARPC2 G-quadruplex (Table 2). The strength of binding is thus in the typical order of magnitude of strong biological interactions, indicating that at least nucleolin binds directly to the ARPC2 G-quadruplex.

A comparison of the binding affinities of  $\Delta$ nucleolin to the ARPC2 G-quadruplex and the mutated sequence revealed binding constants of 0.6 nM and 83 nM, respectively (Table 2). Thus, binding of nucleolin to the ARPC2 G-quadruplex is more than 100-fold stronger and



**Figure 4.** Analysis of protein-RNA interactions by SPR spectroscopy. (A) Binding parameters for  $\Delta$ nucleolin to the ARPC2 and MMP16 G-quadruplex and the mutated sequence were analyzed in the concentration range of 1–20 nM and 10–1000 nM, respectively. Data were fitted to 1:1 binding with mass transfer. Sensorgrams are shown in gray and the respective fits in black. (B) Binding of U2AF65 to 100 nM of the ARPC2 and MMP16 G-quadruplex oligonucleotides and experiments with the respective controls. (C) Additional experiments with the G-rich control oligonucleotides ARPC2 2xG in comparison to the G-quadruplex and the exhaustively mutated oligonucleotide.





**Figure 5.** Analysis of protein-RNA interactions by SPR spectroscopy. Binding of SRSF1 (A) and EFHD2 (B) to 100 nM of the G-quadruplex and mutated control sequences of ARPC2 and MMP16.

seems to be dependent on the formation of the quadruplex structure. The binding constant is in good accordance with nanomolar values described for binding of nucleolin to a G-quadruplex structure formed by ribosomal DNA (44). However, a comparison of the binding constants of  $\Delta$ nucleolin to the MMP16 G-quadruplex and its mutated sequence revealed values of 2 and 0.6 nM, respectively (Table 2). An even stronger binding of  $\Delta$ nucleolin to the mutated sequence was determined, although it is not able to form the 3D structure of G-quadruplexes. There are various nucleolin binding motifs described in the literature, such as G-rich consensus motifs, a loop consisting of the nucleotides UCCCGA, AU-rich elements and G-quadruplexes (45–48). This diversity may result from the different RBDs of nucleolin. It is composed of an acidic N-terminus, four central RBDs and a C-terminal RNA-binding motif, the RGG domain (49). The RBDs 3 and 4, as well as the RGG domain were found to be required for interaction between nucleolin and G-quadruplexes (50).

Members of the family of hnRNPs (hnRNP F, H and U) were also identified as potential RNA G-quadruplex-binding proteins. HnRNPs belong to a group of multifunctional proteins involved in all important steps of RNA processing, including maturation and transport of transcripts or translational regulation (51). The hnRNPs F and H were shown to specifically bind to poly(G) tracts (52). In a mechanistic study, Samatanga et al. (53) described that the third RBD of hnRNP F bound exclusively to single-stranded G-tracts and does not recognize the G-quadruplex. However, in our control experiments with the single-stranded G-rich stretch that cannot form a G-quadruplex, we did not observe interactions with hnRNPs. Samatanga et al. propose a model according to which hnRNP F binds co-transcriptionally to mRNA, thereby preventing G-quadruplex formation. A mechanism assuming that hnRNP F and H bind to G-quadruplexes and dissolves the 3D structure, however, is also imaginable. In both cases, the interaction between the proteins and the G-quadruplexes results in enhanced translation. Less is known about the function and properties of hnRNP U, but the pro-

tein was reported to bind AT-rich DNA elements (54), while we found an interaction with a G-quadruplex.

As outlined before, G-quadruplexes located in the 5'-UTR are thought to play a role in the regulation of translation. However, for ARPC2 and MMP16 various additional proteins were identified by MS which are linked to splicing processes (Table 1). While an involvement in the splicing process of G-quadruplexes located in the coding region was described in previous publications (55–57), a G-quadruplex-dependent splicing in the 5'-UTR has not been shown so far. The identification of proteins involved in splicing or alternative splicing like U2AF65, serine/arginine-rich (SR) proteins or hnRNPs also suggests a function of G-quadruplexes in 5'-UTR splicing.

A role of G-quadruplexes in alternative splicing was shown for the tumor suppressor gene *TP53* mRNA, the *BACE1* ( $\beta$ -site amyloid precursor protein cleaving enzyme 1) mRNA and the *FMRI* (fragile X mental retardation) mRNA (55–58). In the mRNA of *TP53* the G-quadruplex structure in intron 3 modulated the splicing of intron 2 (55). Treatment of a lymphoblastoid cell line with a synthetic ligand that binds to single-stranded G-quadruplex structures altered the ratio of two different splice forms. Moreover, a G-quadruplex in an exon of the mRNA of *BACE1* was found to be involved in controlling splice site selection (56). In the *FMRI* mRNA, the G-quadruplex motif functioned as an exonic splicing enhancer that increased the inclusion of an alternatively spliced exon (57). Exonic splicing enhancers are sequences within exons that promote constitutive as well as regulated splicing. It is thought that splicing factors like U2AF65 or SRSF1 are involved in the recognition of these sequences (58).

An interaction between SRSF1 or U2AF65 and G-quadruplexes as demonstrated in the present study has not been shown before. SR proteins are a conserved family of proteins involved in constitutive, as well as in alternative, splicing by binding to exonic and intronic sites to enhance exon recognition (59). SR proteins bind to exonic splicing enhancers with their N-terminal RNA-recognition motifs. Two SR proteins were found to interact with the G-

quadruplex in the 5'-UTR of ARPC2 (Table 1). Relatively little is known about SRSF9, but the protein was described to function as a repressor of 3' splice site-utilization (60).

The *U2AF65* gene encodes the large, 65 kDa subunit of the U2 auxiliary factor (U2AF). Originally, U2AF65 was shown to be involved in recognition of 3' splice sites by binding to the polypyrimidine tract located between the branch point sequence and the AG dinucleotide. It was also identified as a major triplex DNA-binding protein by Nelson and colleagues (61), who speculated that U2AF65 could be able to bind multi-stranded nucleic acid structures like G-quadruplexes, R-loops or D-loops. Our SPR analysis confirmed the strong interactions between U2AF65 and the RNA G-quadruplexes of both ARPC2 and MMP16. We determined dissociation constants of 1 and 5 nM, respectively, while binding to the control sequences was very weak, suggesting a biologically relevant binding to both G-quadruplex structures.

While the capacity to interact with nucleic acids or G-quadruplexes has previously been described for a number of the identified proteins, no such properties have been ascribed to EFHD2 and ME2. EFHD2 is a calcium-binding protein containing three binding sites for SH3 domain-containing proteins, two EF-hand domains, and a coiled-coil domain at the C-terminus. The protein has been shown to be involved in signaling in immature B cells (62,63) and was now found as binding partner of the ARPC2 G-quadruplex (Table 1). ME2 was identified as a pronounced interaction partner of the G-quadruplex of MMP16. It catalyzes the oxidative decarboxylation of malate to pyruvate in the mitochondrion (64). RNA-binding activity was shown neither for EFHD2 nor for ME2 before and might therefore indicate a previously unknown function for both proteins.

For ME2 it may be speculated that interactions with G-quadruplexes may have an impact on the regulation of metabolic pathways. A search for potential G-quadruplex-forming sequences in the 5'-UTR of ME2 revealed four hits. Thus it is conceivable that ME2 regulates its own translation by binding to these G-quadruplex structures. Such a negative feedback loop was shown before for FMRP and the G-quadruplex located in its own mRNA, where it can regulate its splicing (65).

While U2AF65 was binding to both G-quadruplex motifs, all other candidates were specific for either of the two structures, suggesting two different binding mechanisms. The universal binders can be assumed to recognize the G-quadruplex structures, while the specific binders may interact with the loop sequence, which is variable in different G-quadruplexes. Pull-down experiments were carried out with extracts from two cell lines, HEK293 and HeLa cells, and similar results were obtained in both cases. The interactions thus do not seem to be specific for a certain cell type.

In summary, we have isolated proteins that bind to G-quadruplex motifs in the 5'-UTR of mRNAs for the first time using a systematic screening approach. A number of proteins known to regulate translation such as RPs or nucleolin were identified verifying the role of G-quadruplexes in translation. Additionally, proteins involved in splicing or alternative splicing processes have been detected by MS, indi-

cating a role of G-quadruplexes in 5'-UTR splicing. Moreover, we identified proteins that have previously not been associated with interactions to nucleic acids. Most of the identified proteins are specific for either of the two investigated G-quadruplex motifs, while some exceptions seem to have a more general capacity to bind to these structural elements. The determined dissociation constants in the low nanomolar range indicate strong interactions with biological relevance. As an important control, we demonstrate that the identified proteins do not bind to G-rich sequences in general, but only to sequences that are capable of forming a G-quadruplex structure. It will now be necessary to elucidate how binding of the proteins influences stability of the G-quadruplexes and finally modulates translational efficiency or alternative splicing processes to gain deeper insight into the biological relevance of protein-RNA G-quadruplex interactions.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

We thank Soumitra Basu for sending us reporter plasmids encoding the MMP16 G-quadruplex. Furthermore, we are thankful to Beate Koksich and her group for allowing us to perform CD measurements on their instrument. Finally, we want to thank Erik Wade for careful proofreading and valuable comments on the manuscript.

## FUNDING

Deutsche Forschungsgemeinschaft [Ku 1436/7-1]; Technische Universität Berlin [60/54003/00]. Funding for open access charge: Deutsche Forschungsgemeinschaft [Ku 1436/7-1]

*Conflict of interest.* None declared.

## REFERENCES

- Lipps, H.J. and Rhodes, D. (2009) G-quadruplex structures: in vivo evidence and function. *Trends Cell Biol.*, **19**, 414–422.
- Collie, G.W. and Parkinson, G.N. (2011) The application of DNA and RNA G-quadruplexes to therapeutic medicines. *Chem. Soc. Rev.*, **40**, 5867–5892.
- Wu, Y. and Brosh, R.M. Jr. (2010) G-quadruplex nucleic acids and human disease. *FEBS J.*, **277**, 3470–3488.
- Huppert, J.L. (2010) Structure, location and interactions of G-quadruplexes. *FEBS J.*, **277**, 3452–3458.
- Biffi, G., Tannahill, D., McCafferty, J. and Balasubramanian, S. (2013) Quantitative visualization of DNA G-quadruplex structures in human cells. *Nat. Chem.*, **5**, 182–186.
- Halder, K. and Hartig, J.S. (2011) RNA quadruplexes. *Met. Ions Life Sci.*, **9**, 125–139.
- Ji, X., Sun, H., Zhou, H., Xiang, J., Tang, Y. and Zhao, C. (2011) Research progress of RNA quadruplex. *Nucleic Acid Ther.*, **21**, 185–200.
- Bugaut, A. and Balasubramanian, S. (2012) 5'-UTR RNA G-quadruplexes: translation regulation and targeting. *Nucleic Acids Res.*, **40**, 4727–4741.
- Millevoi, S., Moine, H. and Vagner, S. (2012) G-quadruplexes in RNA biology. *Wiley Interdiscip. Rev. RNA*, **3**, 495–507.
- Kumari, S., Bugaut, A., Huppert, J.L. and Balasubramanian, S. (2007) An RNA G-quadruplex in the 5' UTR of the NRAS proto-oncogene modulates translation. *Nat. Chem. Biol.*, **3**, 218–221.

11. Arora, A., Dutkiewicz, M., Scaria, V., Hariharan, M., Maiti, S. and Kurreck, J. (2008) Inhibition of translation in living eukaryotic cells by an RNA G-quadruplex motif. *RNA*, **14**, 1290–1296.
12. Huppert, J.L., Bugaut, A., Kumari, S. and Balasubramanian, S. (2008) G-quadruplexes: the beginning and end of UTRs. *Nucleic Acids Res.*, **36**, 6260–6268.
13. Halder, K., Wieland, M. and Hartig, J.S. (2009) Predictable suppression of gene expression by 5'-UTR-based RNA quadruplexes. *Nucleic Acids Res.*, **37**, 6811–6817.
14. Morris, M.J. and Basu, S. (2009) An unusually stable G-quadruplex within the 5'-UTR of the MT3 matrix metalloproteinase mRNA represses translation in eukaryotic cells. *Biochemistry*, **48**, 5313–5319.
15. Balkwill, G.D., Derecka, K., Garner, T.P., Hodgman, C., Flint, A.P. and Searle, M.S. (2009) Repression of translation of human estrogen receptor alpha by G-quadruplex formation. *Biochemistry*, **48**, 11487–11495.
16. Gomez, D., Guedin, A., Mergny, J.L., Salles, B., Riou, J.F., Teulade-Fichou, M.P. and Calsou, P. (2010) A G-quadruplex structure within the 5'-UTR of TRF2 mRNA represses translation in human cells. *Nucleic Acids Res.*, **38**, 7187–7198.
17. Shahid, R., Bugaut, A. and Balasubramanian, S. (2010) The BCL-2 5' untranslated region contains an RNA G-quadruplex-forming motif that modulates protein expression. *Biochemistry*, **49**, 8300–8306.
18. Khateb, S., Weisman-Shomer, P., Hershco-Shani, I., Ludwig, A.L. and Fry, M. (2007) The tetraplex (CGG)<sub>n</sub> destabilizing proteins hnRNP A2 and CBF-A enhance the in vivo translation of fragile X premutation mRNA. *Nucleic Acids Res.*, **35**, 5775–5788.
19. Weng, H.Y., Huang, H.L., Zhao, P.P., Zhou, H. and Qu, L.H. (2012) Translational repression of cyclin D3 by a stable G-quadruplex in its 5' UTR: implications for cell cycle regulation. *RNA Biol.*, **9**, 1099–1109.
20. Agarwala, P., Pandey, S., Mapa, K. and Maiti, S. (2013) The G-quadruplex augments translation in the 5' untranslated region of transforming growth factor beta2. *Biochemistry*, **52**, 1528–1538.
21. 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.
22. Beaudoin, J.D. and Perreault, J.P. (2010) 5'-UTR G-quadruplex structures acting as translational repressors. *Nucleic Acids Res.*, **38**, 7022–7036.
23. 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.
24. Morris, M.J., Wingate, K.L., Silwal, J., Leeper, T.C. and Basu, S. (2012) The porphyrin TmPyP4 unfolds the extremely stable G-quadruplex in MT3-MMP mRNA and alleviates its repressive effect to enhance translation in eukaryotic cells. *Nucleic Acids Res.*, **40**, 4137–4145.
25. Ofer, N., Weisman-Shomer, P., Shklover, J. and Fry, M. (2009) The quadruplex r(CGG)<sub>n</sub> destabilizing cationic porphyrin TMPyP4 cooperates with hnRNPs to increase the translation efficiency of fragile X premutation mRNA. *Nucleic Acids Res.*, **37**, 2712–2722.
26. Halder, K., Lary, E., Benzler, M., Teulade-Fichou, M.P. and Hartig, J.S. (2011) Efficient suppression of gene expression by targeting 5'-UTR-based RNA quadruplexes with bisquinolinium compounds. *Chembiochem*, **12**, 1663–1668.
27. Bugaut, A., Rodriguez, R., Kumari, S., Hsu, S.T. and Balasubramanian, S. (2010) Small molecule-mediated inhibition of translation by targeting a native RNA G-quadruplex. *Org. Biomol. Chem.*, **8**, 2771–2776.
28. Gonzalez, V., Guo, K., Hurley, L. and Sun, D. (2009) Identification and characterization of nucleolin as a c-myc G-quadruplex-binding protein. *J. Biol. Chem.*, **284**, 23622–23635.
29. Sissi, C., Gatto, B. and Palumbo, M. (2011) The evolving world of protein-G-quadruplex recognition: a medicinal chemist's perspective. *Biochimie*, **93**, 1219–1230.
30. Schaeffer, C., Bardoni, B., Mandel, J.L., Ehresmann, B., Ehresmann, C. and Moine, H. (2001) The fragile X mental retardation protein binds specifically to its mRNA via a purine quartet motif. *EMBO J.*, **20**, 4803–4813.
31. Lopez de Silanes, I., Stagno d'Alcontres, M. and Blasco, M.A. (2010) TERRA transcripts are bound by a complex array of RNA-binding proteins. *Nat. Commun.*, **1**, 33.
32. Hotary, K., Li, X.Y., Allen, E., Stevens, S.L. and Weiss, S.J. (2006) A cancer cell metalloprotease triad regulates the basement membrane transmigration program. *Genes Dev.*, **20**, 2673–2686.
33. Plaisier, M., Kapiteijn, K., Koolwijk, P., Fijten, C., Hanemaaijer, R., Grimbergen, J.M., Mulder-Stapel, A., Quax, P.H., Helmerhorst, F.M. and van Hinsbergh, V.W. (2004) Involvement of membrane-type matrix metalloproteinases (MT-MMPs) in capillary tube formation by human endometrial microvascular endothelial cells: role of MT3-MMP. *J. Clin. Endocrinol. Metab.*, **89**, 5828–5836.
34. Rauhala, H.E., Teppo, S., Niemela, S. and Kallioniem, A. (2013) Silencing of the ARP2/3 complex disturbs pancreatic cancer cell migration. *Anticancer Res.*, **33**, 45–52.
35. Mergny, J.L., Phan, A.T. and Lacroix, L. (1998) Following G-quartet formation by UV-spectroscopy. *FEBS Lett.*, **435**, 74–78.
36. Rachwal, P.A. and Fox, K.R. (2007) Quadruplex melting. *Methods*, **43**, 291–301.
37. Dyballa, N. and Metzger, S. (2009) Fast and sensitive colloidal coomassie G-250 staining for proteins in polyacrylamide gels. *J. Vis. Exp.*, **30**, 1431.
38. Shevchenko, A., Wilm, M., Vorm, O. and Mann, M. (1996) Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Anal. Chem.*, **68**, 850–858.
39. Suckau, D., Resemann, A., Schuerenberg, M., Hufnagel, P., Franzen, J. and Holle, A. (2003) A novel MALDI LIFT-TOF/TOF mass spectrometer for proteomics. *Anal. Bioanal. Chem.*, **376**, 952–965.
40. Menendez, C., Frees, S. and Bagga, P.S. (2012) QGRS-H Predictor: a web server for predicting homologous quadruplex forming G-rich sequence motifs in nucleotide sequences. *Nucleic Acids Res.*, **40**, W96–W103.
41. Wieland, M. and Hartig, J.S. (2007) RNA quadruplex-based modulation of gene expression. *Chem. Biol.*, **14**, 757–763.
42. Gebauer, F. and Hentze, M.W. (2004) Molecular mechanisms of translational control. *Nat. Rev. Mol. Cell Bio.*, **5**, 827–835.
43. Yanagida, M., Shimamoto, A., Nishikawa, K., Furuichi, Y., Isobe, T. and Takahashi, N. (2001) Isolation and proteomic characterization of the major proteins of the nucleolin-binding ribonucleoprotein complexes. *Proteomics*, **1**, 1390–1404.
44. Hanakahi, L.A., Sun, H. and Maizels, N. (1999) High affinity interactions of nucleolin with G-G-paired rDNA. *J. Biol. Chem.*, **274**, 15908–15912.
45. Abdelmohsen, K., Tominaga, K., Lee, E.K., Srikantan, S., Kang, M.J., Kim, M.M., Selimyan, R., Martindale, J.L., Yang, X.L., Carrier, F. et al. (2011) Enhanced translation by Nucleolin via G-rich elements in coding and non-coding regions of target mRNAs. *Nucleic Acids Res.*, **39**, 8513–8530.
46. Ghisolfi-Nieto, L., Joseph, G., Puvion-Dutilleul, F., Amalric, F. and Bouvet, P. (1996) Nucleolin is a sequence-specific RNA-binding protein: Characterization of targets on pre-ribosomal RNA. *J. Mol. Biol.*, **260**, 34–53.
47. Sengupta, T.K., Bandyopadhyay, S., Fernandes, D.J. and Spicer, E.K. (2004) Identification of nucleolin as an AU-rich element binding protein involved in bcl-2 mRNA stabilization. *J. Biol. Chem.*, **279**, 10855–10863.
48. Gonzalez, V., Guo, K.X., Hurley, L. and Sun, D. (2009) Identification and characterization of nucleolin as a c-myc G-quadruplex-binding protein. *J. Biol. Chem.*, **284**, 23622–23635.
49. Lapeyre, B., Bourbon, H. and Amalric, F. (1987) Nucleolin, the major nucleolar protein of growing eukaryotic cells: an unusual protein structure revealed by the nucleotide sequence. *Proc. Natl. Acad. Sci. U.S.A.*, **84**, 1472–1476.
50. Gonzalez, V. and Hurley, L.H. (2010) The C-terminus of nucleolin promotes the formation of the c-MYC G-quadruplex and inhibits c-MYC promoter activity. *Biochemistry*, **49**, 9706–9714.
51. Han, S.P., Tang, Y.H. and Smith, R. (2010) Functional diversity of the hnRNPs: past, present and perspectives. *Biochem. J.*, **430**, 379–392.
52. Matunis, M.J., Xing, F. and Dreyfuss, G. (1994) The hnRNP F protein: unique primary structure, nucleic acid-binding properties, and subcellular localization. *Nucleic Acids Res.*, **22**, 1059–1067.
53. Samatanga, B., Dominguez, C., Jelesarov, I. and Allain, F.H. (2013) The high kinetic stability of a G-quadruplex limits hnRNP F qRRM3 binding to G-tract RNA. *Nucleic Acids Res.*, **41**, 2505–2516.
54. Fackelmayer, F.O. and Richter, A. (1994) Purification of two isoforms of hnRNP-U and characterization of their nucleic acid binding activity. *Biochemistry*, **33**, 10416–10422.



55. Marcel, V., Tran, P.L., Sagne, C., Martel-Planche, G., Vaslin, L., Teulade-Fichou, M.P., Hall, J., Mergny, J.L., Hainaut, P. and Van Dyck, E. (2011) G-quadruplex structures in TP53 intron 3: role in alternative splicing and in production of p53 mRNA isoforms. *Carcinogenesis*, **32**, 271–278.
56. Fiset, J.F., Montagna, D.R., Mihailescu, M.R. and Wolfe, M.S. (2012) A G-rich element forms a G-quadruplex and regulates BACE1 mRNA alternative splicing. *J. Neurochem.*, **121**, 763–773.
57. Didiot, M.C., Tian, Z., Schaeffer, C., Subramanian, M., Mandel, J.L. and Moine, H. (2008) The G-quartet containing FMRP binding site in FMR1 mRNA is a potent exonic splicing enhancer. *Nucleic Acids Res.*, **36**, 4902–4912.
58. Blencowe, B.J. (2000) Exonic splicing enhancers: mechanism of action, diversity and role in human genetic diseases. *Trends Biochem. Sci.*, **25**, 106–110.
59. Long, J.C. and Caceres, J.F. (2009) The SR protein family of splicing factors: master regulators of gene expression. *Biochem. J.*, **417**, 15–27.
60. Simard, M.J. and Chabot, B. (2002) SRp30c is a repressor of 3' splice site utilization. *Mol. Cell. Biol.*, **22**, 4001–4010.
61. Nelson, L.D., Bender, C., Mannsperger, H., Buerge, D., Kambakamba, P., Mudduluru, G., Korf, U., Hughes, D., Van Dyke, M.W. and Allgayer, H. (2012) Triplex DNA-binding proteins are associated with clinical outcomes revealed by proteomic measurements in patients with colorectal cancer. *Mol. Cancer*, **11**, 38.
62. Avramidou, A., Kroczyk, C., Lang, C., Schuh, W., Jack, H.M. and Mielenz, D. (2007) The novel adaptor protein Swiprosin-1 enhances BCR signals and contributes to BCR-induced apoptosis. *Cell Death Differ.*, **14**, 1936–1947.
63. Dutting, S., Brachs, S. and Mielenz, D. (2011) Fraternal twins: Swiprosin-1/EFhd2 and Swiprosin-2/EFhd1, two homologous EF-hand containing calcium binding adaptor proteins with distinct functions. *Cell Commun. Signal.*, **9**, 2.
64. Sauer, L.A. (1973) Mitochondrial Nad-dependent malic enzyme - new regulatory enzyme. *FEBS Lett.*, **33**, 251–255.
65. Didiot, M.C., Tian, Z.X., Schaeffer, C., Subramanian, M., Mandel, J.L. and Moine, H. (2008) The G-quartet containing FMRP binding site in FMR1 mRNA is a potent exonic splicing enhancer. *Nucleic Acids Res.*, **36**, 4902–4912.