

# Predictable suppression of gene expression by 5'-UTR-based RNA quadruplexes

Kangkan Halder<sup>1</sup>, Markus Wieland<sup>1,2,3</sup> and Jörg S. Hartig<sup>1,2,3,\*</sup>

<sup>1</sup>Department of Chemistry, <sup>2</sup>Konstanz Research School Chemical Biology (KoRS-CB) and

<sup>3</sup>Zukunftskolleg, University of Konstanz, Universitätsstr 10, 78457 Konstanz, Germany

Received April 28, 2009; Revised July 15, 2009; Accepted August 6, 2009

## ABSTRACT

Four-stranded DNA and RNA quadruplexes or G4 motifs are non-B DNA conformations that are presumed to form *in vivo*, although only few explicit evidence has been reported. Using bioinformatics the presence of putative DNA G-quadruplexes within critical promoter regions has been demonstrated and a regulatory role in transcription has been suspected. However, in genomic DNA the presence of the complementary strand interferes with the potential to form a quadruplex motif. Contrarily RNA G4 motifs have no such limitation and consequently strong interference with gene expression is suspected. Nevertheless, experimental evidence is scarce. Here we show a well-defined structure–function relationship of synthetic quadruplex sequences in 5'-UTRs in multiple mammalian cell-lines. We establish a universal 'translational suppressor' effect of these motifs on gene expression at the translational level and show for the first time that specific features such as loop-length and the number of 'GGG'-repeats further determine the suppressive impact. Moreover, a consistent and predictable repression of gene expression is observed for naturally occurring RNA G4 motifs, augmenting the functional relevance of these unusual nucleic acid structures.

## INTRODUCTION

DNA and RNA sequences with a minimum four interspersed 'GGG' repeats can potentially fold back to form stable G-quadruplex (G4) motifs, a four-stranded conformation composed of stacked guanine tetrads interacting via Hoogsteen hydrogen bonds which is further stabilized by chelating monovalent metal ions such as sodium or potassium (1,2). Initial computational searches identified putative sequences which can form

G4 motifs throughout the genome in various mammalian (3–5) and bacterial (6) species. More rigorous analyses of the human genome suggested a statistically significant enrichment in gene promoters (7,8) and untranslated regions (UTRs) (9). Based on these findings, regulatory functions of DNA quadruplexes have been suggested in gene expression (10,11) and genomic maintenance (12–14). Although in genetic mechanisms such as replication, recombination (15), and transcription transient single-stranded DNA is observed, the presence of the complementary strand should disfavor quadruplex formation (16,17). Interestingly the single-stranded guanine-rich telomeric overhang, which has no duplex competition against structure formation, has been shown to form a quadruplex motif under *in vitro* conditions, while the *in vivo* existence is still controversial (18). The telomeric repeat has been studied in great detail with respect to its potential to form quadruplex structures (1,2). The mammalian telomeric overhang flips over and invades the telomeric duplex sequence generating an unique conformation called 't-loop' (19). An increasing number of proteins have been identified that specifically interact with DNA quadruplexes, resulting in enhanced or disrupted formation of quadruplex motifs (20–28). Since the induction of quadruplexes seems to interfere with telomere length maintenance and hence with the replicative life-span, various classes of small molecules targeting genomic quadruplexes have been developed, displaying anti-proliferative activities due to inhibition of proto-oncogene expression (29) and telomerase activity (30). In contrast, very little is known about the properties and influence of RNA quadruplex formation *in vivo*. The formation of four-stranded RNA structures is well established *in vitro*, demonstrating that G-rich RNAs form quadruplexes with stabilities equal to their DNA counterparts (31–33). Computational searches for potential RNA quadruplexes in human mRNAs have identified a prevalence in 5'-UTRs (9), but only little experimental evidence concerning its effect on gene expression is available. Two recent studies demonstrate a suppressive influence of individual RNA quadruplexes found in the *NRAS* (34) and *ZIC-1* (35) 5'-UTRs. We recently demonstrated

\*To whom correspondence should be addressed. Tel: +49 7531 882398; Fax: +49 7531 885140; Email: joerg.hartig@uni-konstanz.de

a pronounced effect of artificially introduced quadruplexes on gene expression in bacterial mRNAs (31), however, an extensive study of the impact of RNA quadruplexes in mammalian mRNAs is lacking. In addition to characterizing naturally occurring sequence motifs (34,35), the use of artificial sequences allows for studying how distinct quadruplex features such as loop compositions and the number of G-rich repeats contribute to the suppressive effects of these unusual nucleic acid structures on gene expression (31). Here we systematically characterize the influence of synthetic and natural quadruplex-forming sequences on gene expression in several mammalian cell lines.

## MATERIALS AND METHODS

### Circular dichroism spectroscopy

All samples were prepared at 5  $\mu$ M strand concentration (metabion GmbH, HPLC purified) in RNase free 10 mM Tris-Cl, pH 7.4, 10 mM KCl, except Figure S3 (1 mM KCl). Samples were heated to 95°C for 5 mins and slowly cooled overnight before scanning in a 1 ml capped-cuvette at 15°C using Jasco 715 spectropolarimeter (Jasco Hachioji, Tokyo, Japan) fitted with a Peltier temperature controller. Circular dichroism (CD) scans were recorded in pentaplets at 50 nm min<sup>-1</sup> with a 2 s response time, 1 nm pitch and 1 nm bandwidth, averaged, buffer subtracted and zero-corrected at 320 nm. For  $T_m$  calculation, samples were annealed from 95 to 15°C and melted back, at a controlled rate of 1°C min<sup>-1</sup>.

### Construction of plasmids

The G-quadruplex sequences (Tables 1 and 2) were introduced into the psiCHECK-2 (Promega, Supplementary Figure 1) by whole plasmid PCR with primers having G-rich 5'-overhangs (Supplementary Figure 5) using the sequences shown in Supplementary Table 1. A detailed protocol for generating the quadruplex insertions can be found in ref. (36).

### Natural 5'-UTR quadruplex sequences

Human genome 5'-UTR sequences (25 193; one FASTA record per refgene) were downloaded from UCSC table browser (<http://genome.ucsc.edu/cgi-bin/hgTables>) as on May 2008. We could retrieve 2013 putative RNA G-quadruplex forming sequences using customized algorithm (6), which is in good accordance with 2334 quadruplex reported using different algorithm and 5'-UTR sequence retrieval option recently (9). We then manually searched for distinct multiple 'GGG'-repeat sequences and the *4MAPK2*, *5CHST2* and *6PCGF2* quadruplex sequences were arbitrarily chosen.

### Cell culture and dual luciferase assay

Cells were grown to ~50% confluency in flat-bottomed 96-well plate under standard conditions of 37°C, 5% CO<sub>2</sub> in Dulbecco's modified Eagles medium (DMEM) supplemented with 10% heat-inactivated fetal calf serum (FCS), 2 mM glutamine, non-essential amino acids

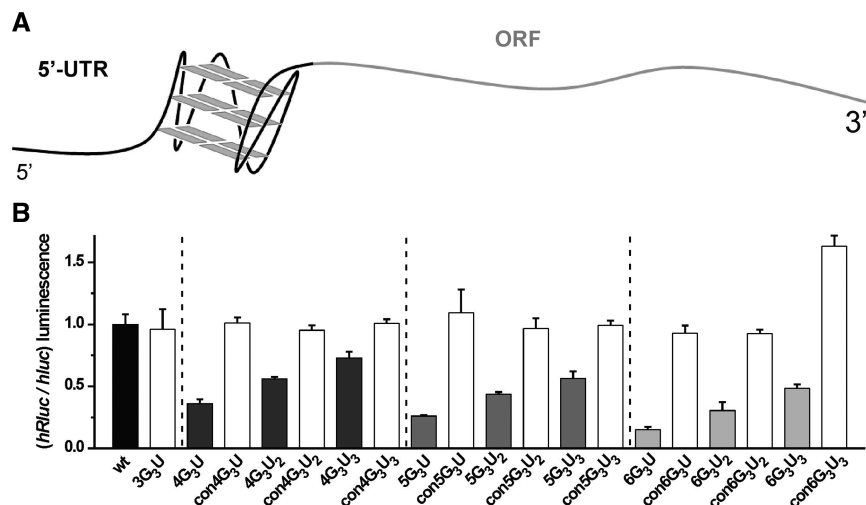
and the antibiotics penicillin and streptomycin, before transient transfection with 50 ng of plasmid constructs (Tables 1 and 2) using Lipofectamine 2000 (Invitrogen) or Turbofect (Fermentas), according to the manufacturer's protocol. Twenty-four hours post-transfection, cells were retrieved and processed using Dual-luciferase Reporter Assay kit (E1910, Promega) according to the manufacturer's protocol. The luciferase activity was determined using a TECAN M200 reader.

### Quantitative real-time PCR assay

HEK293 cells were grown to ~50% confluency in flat-bottomed 6-well plate, under conditions stated above and transfected with Turbofect (Fermentas), according to the manufacturer's protocol. Twenty-four hours later, the total cellular RNA was isolated using RNeasy Plus Mini Kit (74134, Qiagen). For quantification of mRNA levels, 1  $\mu$ g of total RNA was used in a 25  $\mu$ l reaction using Titan One Tube RT-PCR System (11888382001, Roche) and SYBR-green (kindly provided by the Marx group, University of Konstanz) on a Chromo4 Real-Time Detector (Biorad) using primer pairs 5'-CCAACCCTGTTCAGCTTCTTC-3' and 5'-ACCTTGGCCTCGAAGAA TGG-3' for *hluc* mRNA and 5'-TGATCGAGTCCTGGG ACGA-3' and 5'-ACAATCTGGACGACGTCGGG-3' for *hRluc* mRNA. All reactions were done in triplicate.

## RESULTS AND DISCUSSION

In order to investigate the impact of quadruplex formation on gene expression, we chose to introduce artificially designed G4 motif sequences into a reporter gene setup (Figures 1A and Supplementary Figure 1). The use of such non-natural sequences allows for studying the effect of stability-determining features, circumventing the sequence restrictions that apply when using natural motifs. Since the thermodynamic stability of G4 motifs is known to depend on the loop lengths, we first tested the influence of quadruplexes containing one- (4G<sub>3</sub>U), two- (4G<sub>3</sub>U<sub>2</sub>), or three- (4G<sub>3</sub>U<sub>3</sub>) nucleotide loops (Table 1), inserted eleven nucleotides upstream of the start codon (corresponding to +124 bp with respect to the 5'-end of the mRNA, for a detailed description of the construct see also Supplementary Table 1) of a luciferase reporter. For DNA quadruplexes, the thermodynamic stability increases with decreasing loop lengths, with mononucleotide loops exhibiting the highest stabilities (37,38). We found that the expression of the *Renilla* reporter luciferase in HEK293 cells decreased to residual 36, 56 and 73%, respectively, compared to the parental reporter construct (Figure 1B). To confirm that perturbation of gene expression is solely based on quadruplex structure formation and not due to the mere insertion of the additional G-rich sequence, we investigated a control sequence for each quadruplex studied. For this purpose, we constructed similar G/U-rich control sequences (Table 1) which are unable to form a quadruplex by disrupting the middle tetrads by G to U mutations. Importantly, the expression was restored in all three control sequences (Figure 1B). In addition, we inserted a control sequence containing



**Figure 1.** Effect of 5'-UTR RNA quadruplexes on gene expression. (A) Scheme of an RNA quadruplex located in the 5'-UTR (black) of the reporter gene. A typical RNA quadruplex structure comprising all-parallel strand orientation is shown. (B) The expression of various quadruplex-containing constructs and the respective controls (Table 1) in HEK293 cells, measured using a dual luciferase assay and transient transfection, is shown. Error bars represent standard deviation of three independent experiments.

**Table 1.** Characterized artificial quadruplex sequences

Construct name	Sequence (5'-3')
wt	ACUAUA·GGCUAGCCACC <b>AUG</b> GCUC
3G <sub>3</sub> U	(GGGU) <sub>2</sub> GGG
4G <sub>3</sub> U	(GGGU) <sub>3</sub> GGG
4G <sub>3</sub> U <sub>2</sub>	(GGGUU) <sub>3</sub> GGG
4G <sub>3</sub> U <sub>3</sub>	(GGGUUU) <sub>3</sub> GGG
5G <sub>3</sub> U	GGGU(GGGU) <sub>3</sub> GGG
5G <sub>3</sub> U <sub>2</sub>	GGGUU(GGGUU) <sub>3</sub> GGG
5G <sub>3</sub> U <sub>3</sub>	GGGUUU(GGGUUU) <sub>3</sub> GGG
6G <sub>3</sub> U	GGGUGGGU(GGGU) <sub>3</sub> GGG
6G <sub>3</sub> U <sub>2</sub>	GGGUUGGGUU(GGGUU) <sub>3</sub> GGG
6G <sub>3</sub> U <sub>3</sub>	GGGUUUUGGGUUU(GGGUUU) <sub>3</sub> GGG

RNA quadruplex motif sequences inserted 11 bp upstream (dot) of the start codon of the *hRluc* gene (wt). The start codon is shown in bold. The prefixes 3, 4, 5 or 6 in the construct names represent the number of total 'GGG'-repeat runs in the sequence and were further prefixed with 'con' for the respective control sequences (U instead of underlined G).

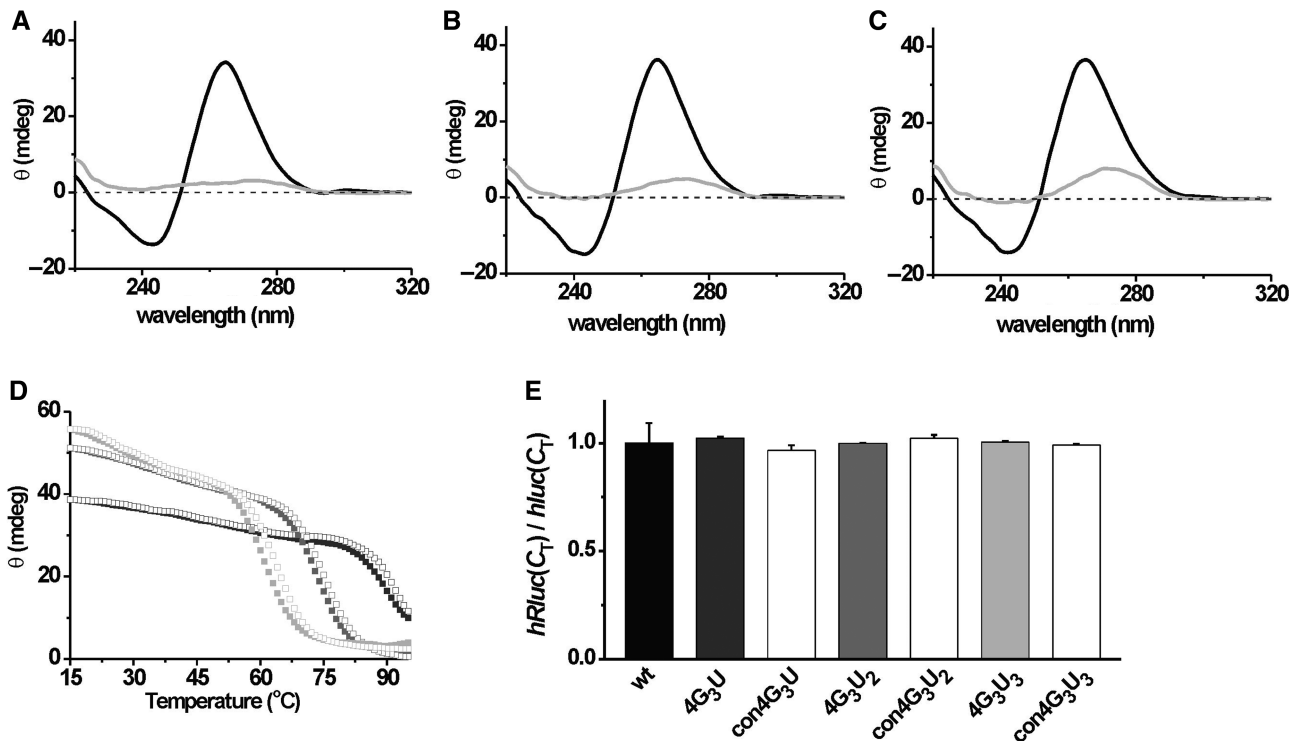
only three 'GGG'-repeats (3G<sub>3</sub>U, Table 1) which by itself is not able to form an intramolecular quadruplex, but could potentially participate in an intermolecular RNA quadruplex. We found no influence on expression for the construct 3G<sub>3</sub>U compared to wt-*hRluc* or the other G/U-rich control sequences (Figure 1B) and hence we conclude that the changes in expression are imminently due to intramolecular RNA quadruplex formation.

In order to comprehend the expression differences observed for the different loop sizes, we decided to structurally and thermodynamically characterize the investigated RNA sequences using CD spectroscopy. For DNA quadruplexes it has been established that the thermodynamic stability is inversely proportional to loop sizes (37,38). For RNA quadruplexes there are only few data available (31), hence we investigated the stability of the quadruplexes 4G<sub>3</sub>U<sub>3</sub>, 4G<sub>3</sub>U<sub>2</sub> and 4G<sub>3</sub>U. We found that the inserted RNA sequences form parallel-stranded

quadruplexes (Figure 2A–C) displaying gradually increasing stability with decreasing loop length from 4G<sub>3</sub>U<sub>3</sub> to 4G<sub>3</sub>U (Figure 2D). The latter finding confirms previous data demonstrating a higher stability of RNA quadruplexes with decreasing loop sizes, behaving similarly to DNA quadruplexes (31). We have also carried out CD spectroscopy with the control RNA sequences that should be unable to adopt quadruplexes. We found no structure formation (Figure 2A–C) which agrees with the absence of effects on gene expression as discussed above. In conclusion, a direct correlation of the quadruplex stability with its influence on gene expression has been demonstrated in mammalian 5'-UTRs.

Next, we investigated at which level of gene expression the quadruplex formation takes effect. For this purpose, we quantified mRNA-levels by real time PCR in order to clarify whether the insertion of the G-rich sequences interfered with mRNA abundance by reducing transcriptional efficiency or mRNA stability. Identical mRNA levels were detected for all tested expression constructs (Figure 2E) in accordance to the results reported recently for the *ZIC-1* quadruplex, pointing at a mechanism acting at the level of translational initiation (35). This result is also supported by a report of quadruplex formation in the *FMR1* mRNA resulting in stalled 40S ribosomal subunits (39,40). Additional experiments are necessary to decipher the exact mechanism of quadruplex interference with translation.

Although four interspersed 'GGG'-repeats are the minimal requirement to form a G4 motif, naturally occurring G4 motifs often contain five, six or even more 'GGG'-repeats (1). In order to investigate the influence of the number of 'GGG'-repeats we examined additional series of 5'-UTR luciferase constructs (Table 1, 5G<sub>3</sub>U<sub>X</sub> and 6G<sub>3</sub>U<sub>X</sub>). Interestingly, we found that gene expression further decreased consecutively from four to six 'GGG'-repeat runs consistently in each loop-length series



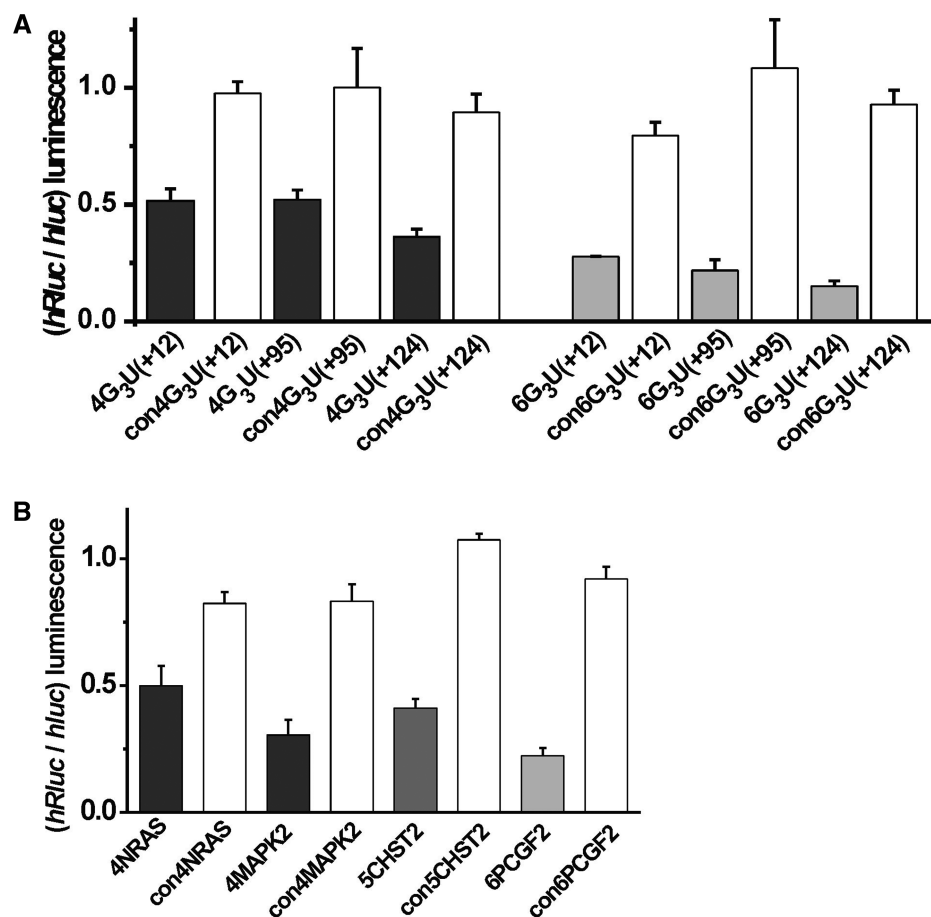
**Figure 2.** *In vitro* stability and *in vivo* abundance of the quadruplex motifs. The CD scans (black) show a positive band around 264 nm and negative maxima around 240 nm for 4G<sub>3</sub>U (A), 4G<sub>3</sub>U<sub>2</sub> (B) and 4G<sub>3</sub>U<sub>3</sub> (C), characteristic for a parallel stranded quadruplex motif, while the respective control sequences; con4G<sub>3</sub>U (A), con4G<sub>3</sub>U<sub>2</sub> (B) and con4G<sub>3</sub>U<sub>3</sub> (C) does not show any quadruplex characteristic peaks (grey). In CD melting experiments (D), denaturing (hollow) and annealing (solid) curves are almost identical and show a  $T_m$  of >95, 73 and 61°C for 4G<sub>3</sub>U (dark grey), 4G<sub>3</sub>U<sub>2</sub> (grey) and 4G<sub>3</sub>U<sub>3</sub> (light grey) RNA-quadruplex motifs, respectively. (E) Relative *hRluc* mRNA levels for different constructs determined by real-time PCR assays [*hRluc*(C<sub>T</sub>)] is normalized to *hLuc* mRNA levels [*hLuc*(C<sub>T</sub>)], as reported earlier (35). Error bars represent standard deviation of three independent experiments.

(Figure 1B and Supplementary Figure 2), with 6G<sub>3</sub>U<sub>x</sub> being more repressive than the five- and four-repeat series, respectively. The most potent suppressor sequence 6G<sub>3</sub>U reduced expression by 85%. Importantly, the control sequences containing G to U mutations as well restored gene expression in the 5- and 6-repeat series, with the clone con6G<sub>3</sub>U<sub>3</sub> showing an unanticipated high expression. In order to investigate whether this unexpected effect of increasing the ‘GGG’ repeat number resulting in even more pronounced suppression, we carried out additional CD melting studies. Surprisingly, we found 5G<sub>3</sub>U and 6G<sub>3</sub>U RNA-quadruplex motifs are thermodynamically more stable than its shorter companion, the 4G<sub>3</sub>U motif (Supplementary Figure 3). This observation is surprising as only one G4 motif with a similar conformation is possible even with five or six ‘GGG’ repeats present and hence is expected to have a similar thermodynamic stability. Although the CD spectra are very similar to the 4G<sub>3</sub>U sequence, pointing at an all-parallel RNA quadruplex (Supplementary Figure 3), additional structural stabilization of the four-stranded structure by the additional ‘GGG’ repeats cannot be ruled out. In light of this finding, it is tempting to speculate that this could be a possible reason why many naturally occurring biologically relevant DNA [*c-myc*, *bcl-2*, *VEGF*, *HIF-1a*) (2) and RNA (*Zic-1* (35), *5CHST2*, and *6PCGF2*, see below) quadruplex sequences contain more than four ‘GGG’-repeat runs.

Since Balasubramanian and co-workers have recently shown that pronounced quadruplex-based repression was evidenced only when the quadruplex motif was located within the first 50 nt of the mRNA 5′-end (41) compared to the position +124 nt as described earlier, we investigated whether moving the quadruplex upstream towards the 5′-end of the mRNA would have any effect on the suppression of gene expression. For this purpose, we inserted 4G<sub>3</sub>U as well as 6G<sub>3</sub>U along with their control sequences at two additional positions (+12 and +95, Supplementary Table 1) in our *hRluc* reporter construct to evaluate any positional effects on expression. We found a gradual but minor decrease in suppression for both 4G<sub>3</sub>U and 6G<sub>3</sub>U, but not in control constructs, when we moved the motif towards the 5′-end (Figure 3A). In conclusion, we observed pronounced suppression of gene expression in all characterized positions in contrast to the study carried out by Balasubramanian and co-workers. This discrepancy may be due to a different experimental setup with respect to the plasmid backbone, 5′-UTR sequence or read-out [a combined *in-vitro* transcription / translation assay was used before to quantify interference with gene expression (41)].

Next, in order to clarify whether the results obtained with our artificially engineered sequences are transferable to naturally occurring quadruplex motifs in mRNAs, we reinvestigated the *NRAS* G4 motif sequence characterized





**Figure 3.** Positional effect and naturally occurring RNA quadruplexes. (A) Normalized luminescence of *hRluc* luciferase of 4G<sub>3</sub>U, 6G<sub>3</sub>U, con4G<sub>3</sub>U or con6G<sub>3</sub>U construct at different positions in the 5'-UTR. Numbers in parentheses indicate the insertion position from the 5'-end. (B) Normalized luminescence of *hRluc* luciferase of naturally occurring RNA G4 motifs and respective control sequences (Supplementary Table 2). Error bars are standard deviations of three independent experiments.

**Table 2.** Characterized natural quadruplex sequences

Construct name	Sequence (5'-3')
4NRAS	GGGAGGGGGCGGUCUGGG
4MAPK2	GGGAGGGAGGGAGGG
5CHST2	GGGCGGGUUGGGUCGGUCGGG
6PCGF2	GGGAUGGGAGGGAGGGAGGGGAAGGG

Sequences of naturally occurring 5'-UTR G-quadruplex motifs were named according to the gene they belong to and prefixed with total number of 'GGG'-repeat runs (4, 5 or 6) in the sequence. Additional prefix of 'con' is used to describe the respective motif control wherein U is inserted instead of G (underlined).

by Balasubramanian and co-workers together with three additional natural 5'-UTR quadruplex sequences (termed 4NRAS, 5CHST2, 4MAPK2 and 6PCGF2, Supplementary Table 2). For this purpose, we introduced the natural quadruplex motifs into the synthetic luciferase 5'-UTR (+124 from transcription start site / -11 from start codon, Supplementary Table 1). We were especially interested whether the general tendencies observed with our synthetic quadruplex sequences with respect to loop lengths and number of GGG repeats would allow for

predicting the influence of the natural motifs on gene expression. Reliably, for the 4NRAS RNA G4 motif (with respective loop-lengths of 1, 2 and 3 nucleotides) the repression (50%, Figure 3B) closely matched the synthetic analogue 4G<sub>3</sub>U<sub>2</sub> (44%, Figure 3B). In case of the other natural motifs 4MAPK2, 5CHST2 and 6PCGF2, the repression (70, 59 and 78%, Figure 3B) also compared well to the synthetic motifs closely resembling loop lengths and 'GGG'-repeat numbers (4G<sub>3</sub>U, 5G<sub>3</sub>U<sub>2</sub> and 6G<sub>3</sub>U<sub>2</sub> with reductions of 64, 57 and 70%, respectively, Figure 1B). Given the overall consistency of repression of naturally occurring 5'-UTR G4 motifs to corresponding synthetic RNA G4 motifs under similar conditions, our study presents a rationale to estimate the effect of naturally occurring RNA G4 motifs on gene expression. Small variations are expected to arise from the different nucleotide composition in loops, which are known to affect the stability of the G4 motif and consequently gene expression (37). Finally, to determine whether the observed quadruplex-mediated repression was specific to the adenovirus-transformed HEK293 cell line, we repeated the experiments using the poorly differentiated human glioblastoma cell line LN18, the human cervical

cancer line HeLa, and the mouse fibroblast line B18. Importantly, consistent but in some cases less pronounced results were obtained (Supplementary Figure 4). In general, the suppressive effects seem to be slightly less prominent using these fundamentally different cell lines with the effect of different loop-lengths being less pronounced as well. Overall, the effects of the different quadruplexes are observed in these alternate cell lines as well. This finding further strengthens the evidence that quadruplex formation in 5'-UTRs affect expressions by a rather unspecific mechanism correlating with thermodynamic stability.

Taken together, the presented work represents the most systematic *in vivo* study of synthetic and naturally occurring quadruplex motifs located in 5'-UTRs in mammalian cell lines. The results demonstrate a coherent impact of RNA quadruplex formation on gene expression and suggest a universal 'translational-suppressor' effect for these non-canonical RNA conformations. We found a predictable correlation of the suppressive effects of RNA quadruplexes with its authoritative features such as loop length and repeat numbers, representing an outstanding example of a facile structure—function relationship of unusual nucleic acid conformations in cellular mechanisms. The present study should allow for predicting the suppressive effects of similar RNA quadruplexes in 5'-UTRs discovered in natural sequence contexts. Moreover, the present system should prove useful for a setup to characterize small molecules selectively targeting RNA quadruplexes.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

The authors thank the group of Prof. M. Leist for the LN18 cell line, the group of Prof. M. Groettrup for HeLa and B18 cell lines and Astrid Joachimi for excellent technical assistance.

## FUNDING

Lichtenberg-Professorship by *VolkswagenStiftung* (to J.S.H.); Alexander von Humboldt-Foundation fellowship (to K.H.). Funding for open access charge: University of Konstanz.

*Conflict of interest statement.* None declared.

## REFERENCES

- 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.
- Patel,D.J., Phan,A.T. and Kuryavyi,V. (2007) Human telomere, oncogenic promoter and 5'-UTR G-quadruplexes: diverse higher order DNA and RNA targets for cancer therapeutics. *Nucleic Acids Res.*, **35**, 7429.
- Huppert,J.L. and Balasubramanian,S. (2005) Prevalence of quadruplexes in the human genome. *Nucleic Acids Res.*, **33**, 2908–2916.
- Todd,A.K., Johnston,M. and Neidle,S. (2005) Highly prevalent putative quadruplex sequence motifs in human DNA. *Nucleic Acids Res.*, **33**, 2901–2907.
- Verma,A., Halder,K., Halder,R., Yadav,V., Rawal,P., Thakur,R., Mohd,F., Sharma,A. and Chowdhury,S. (2008) Genome-wide computational and expression analyses reveal G-quadruplex DNA motifs as conserved cis-regulatory elements in human and related species. *J. Med. Chem.*, **51**, 5641–5649.
- Rawal,P., Kummarasetti,V.B.R., Ravindran,J., Kumar,N., Halder,K., Sharma,R., Mukerji,M., Das,S.K. and Chowdhury,S. (2006) Genome-wide prediction of G4 DNA as regulatory motifs: role in *Escherichia coli* global regulation. *Genome Res.*, **16**, 644–655.
- Huppert,J.L. and Balasubramanian,S. (2007) G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Res.*, **35**, 406–413.
- Zhao,Y., Du,Z. and Li,N. (2007) Extensive selection for the enrichment of G4 DNA motifs in transcriptional regulatory regions of warm blooded animals. *FEBS Lett.*, **581**, 1951–1956.
- 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.
- Du,Z., Zhao,Y. and Li,N. (2008) Genome-wide analysis reveals regulatory role of G4 DNA in gene transcription. *Genome Res.*, **18**, 233–241.
- Qin,Y. and Hurley,L. (2008) Structures, folding patterns, and functions of intramolecular DNA G-quadruplexes found in eukaryotic promoter regions. *Biochimie*, **90**, 1149–1171.
- Halder,K., Halder,R. and Chowdhury,S. (2009) Genome-wide analysis predicts DNA structural motifs as nucleosome exclusion signals. *Mol. Biosyst.* (in press).
- Maizels,N. (2006) Dynamic roles for G4 DNA in the biology of eukaryotic cells. *Nat. Struct. Mol. Biol.*, **13**, 1055–1059.
- Wong,H.M. and Huppert,J.L. (2009) Stable G-quadruplexes are found outside nucleosome-bound regions. *Mol. Biosyst.* (in press).
- Mani,P., Yadav,V., Das,S. and Chowdhury,S. (2009) Genome-wide analyses of recombination prone regions predict role of DNA structural motif in recombination. *PLoS ONE*, **4**, e4399.
- Kumar,N. and Maiti,S. (2008) A thermodynamic overview of naturally occurring intramolecular DNA quadruplexes. *Nucleic Acids Res.*, **36**, 5610–5622.
- Risitano,A. and Fox,K.R. (2003) Stability of intramolecular DNA quadruplexes: comparison with DNA duplexes. *Biochemistry*, **42**, 6507–6513.
- Parkinson,G., Lee,M. and Neidle,S. (2002) Crystal structure of parallel quadruplexes from human telomeric DNA. *Nature*, **417**, 876–880.
- Griffith,J., Comeau,L., Rosenfield,S., Stansel,R., Bianchi,A., Moss,H. and de Lange,T. (1999) Mammalian telomeres end in a large duplex loop. *Cell*, **97**, 503–514.
- Fry,M. (2007) Tetraplex DNA and its interacting proteins. *Front Biosci.*, **12**, 4336–4351.
- Giraldo,R. and Rhodes,D. (1994) The yeast telomere-binding protein RAP1 binds to and promotes the formation of DNA quadruplexes in telomeric DNA. *EMBO J.*, **13**, 2411–2420.
- Froelich-Ammon,S.J., Dickinson,B.A., Bevilacqua,J.M., Schultz,S.C. and Cech,T.R. (1998) Modulation of telomerase activity by telomere DNA-binding proteins in *Oxytricha*. *Genes Dev.*, **12**, 1504–1514.
- Sun,H., Karow,J.K., Hickson,I.D. and Maizels,N. (1998) The Bloom's Syndrome Helicase Unwinds G4 DNA. *J. Biol. Chem.*, **273**, 27587–27592.
- de Lange,T. (2005) Shelterin: the protein complex that shapes and safeguards human telomeres. *Genes Dev.*, **19**, 2100–2110.
- De Boeck,G., Forsyth,R.G., Praet,M. and Hogendoorn,P.C. (2009) Telomere-associated proteins: cross-talk between telomere maintenance and telomere-lengthening mechanisms. *J. Pathol.*, **217**, 327–344.
- Cech,T.R. (2000) Life at the end of the chromosome: telomeres and telomerase. *Angew. Chem. Int. Ed.*, **39**, 34–43.
- Palm,W. and de Lange,T. (2008) How shelterin protects mammalian telomeres. *Annu. Rev. Genet.*, **42**, 301–334.

28. Zaug,A.J., Podell,E.R. and Cech,T.R. (2005) Human POT1 disrupts telomeric G-quadruplexes allowing telomerase extension in vitro. *Proc. Natl Acad. Sci. USA*, **102**, 10864–10869.
29. Doherty,K., Sharma,S., Gupta,R. and Brosh,R. (2006) Tetraplex binding molecules as anti-cancer agents. *Recent Pat. Anticancer Drug Discov.*, **1**, 185–200.
30. Mergny,J.-L., Riou,J.-F., Mailliet,P., Teulade-Fichou,M.-P. and Gilson,E. (2002) Natural and pharmacological regulation of telomerase. *Nucleic Acids Res.*, **30**, 839–865.
31. Wieland,M. and Hartig,J.S. (2007) RNA quadruplex-based modulation of gene expression. *Chem. Biol.*, **14**, 757–763.
32. Shafer,R.H. and Smirnov,I. (2000) Biological aspects of DNA/RNA quadruplexes. *Biopolymers*, **56**, 209–227.
33. Pan,B., Shi,K. and Sundaralingam,M. (2006) Base-tetrad swapping results in dimerization of RNA quadruplexes: implications for formation of the i-motif RNA octaplex. *Proc. Natl Acad. Sci. USA*, **103**, 3130–3134.
34. 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.*, **25**, 25.
35. 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.
36. Wieland,M. and Hartig,J.S. (2009) Investigation of mRNA quadruplex formation in Escherichia coli. *Nature Protocols*, in press.
37. Bugaut,A. and Balasubramanian,S. (2008) A sequence-independent study of the influence of short loop lengths on the stability and topology of intramolecular DNA G-quadruplexes. *Biochemistry*, **47**, 689–697.
38. Hazel,P., Huppert,J., Balasubramanian,S. and Neidle,S. (2004) Loop-length-dependent folding of G-quadruplexes. *J. Am. Chem. Soc.*, **126**, 16405–16415.
39. Heitz,D., Rousseau,F., Devys,D., Saccone,S., Abderrahim,H., Le Paslier,D., Cohen,D., Vincent,A., Toniolo,D., la Valle,G. *et al.* (1991) Isolation of sequences that span the fragile X and identification of a fragile X-related CpG island. *Science*, **251**, 1236–1239.
40. Ofer,N., Weisman-Shomer,P., Shklover,J. and Fry,M. (2009) The quadruplex r(CG)n destabilizing cationic porphyrin TMPyP4 cooperates with hnRNPs to increase the translation efficiency of fragile X premutation mRNA. *Nucleic Acids Res.*, **37**, 2712–2722.
41. Kumari,S., Bugaut,A. and Balasubramanian,S. (2008) Position and stability are determining factors for translation repression by an RNA G-quadruplex-forming sequence within the 5' UTR of the NRAS proto-oncogene. *Biochemistry*, **47**, 12664–12669.