

Reporter assays for studying quadruplex nucleic acids

Kangkan Halder, Martin Benzler, Jörg S. Hartig*

Department of Chemistry and Konstanz Research School Chemical Biology (KoRS-CB), University of Konstanz, 78457 Konstanz, Germany

ABSTRACT

DNA and RNA G-quadruplexes have gained increasing attention due to their potential role in a wide range of biological functions. The majority of functional studies characterize the influence of quadruplexes in gene expression including transcription and translation. Many of these studies have used reporter assays to elucidate the effect of quadruplexes at certain positions in promoters and untranslated mRNA regions (UTRs). Reporter assays are the preferred method to ascertain the biological function of DNA or RNA G-quadruplexes intracellularly due to their ready availability, fast cloning and experimental setup and reproducibility. Moreover, these reporter assays are also helpful to compare or screen for selectivity and efficacy of small molecules that target DNA and RNA G-quadruplexes in the cellular context. Here we briefly discuss various aspects of reporter assays followed by a review of available studies using reporter assays to understand the role and functions of DNA and RNA quadruplexes in gene expression.

Keywords:

G-quadruplex
Transcription
Translation
5'-UTR
3'-UTR
Luciferase assay

1. Introduction

Guanine-rich nucleic acid sequences are able to form higher order secondary structures, so called guanine-quadruplexes (G-quadruplexes). These non-canonical structures are composed of four or more Guanine-tracts (G-tracts) which are able to form stacking guanine tetrads, where the tetrads stack on top of each other. The guanines within the tetrad interact via Hoogsteen hydrogen bonds between the N1, N7, O6 and N2 and are stabilized by monovalent cationic ions like potassium or sodium [1]. The G-tracts are interspersed by short loop sequences [2]. DNA G-quadruplexes can adopt diverse topologies depending on the length and number of individual G-tracts, the length and sequence context of linker residues, but also of the surrounding conditions [1,3–6]. In contrast to DNA G-quadruplexes which can adopt different topologies, RNA G-quadruplexes seem to be monomorphous and fold only into the all-parallel conformation independent of the sequence and surrounding condition [7]. This monomorphism is due to the C3'-endo type of sugar pucker which leads to the favored anti-conformation and therefore the all-parallel topology [8]. DNA and RNA quadruplexes have been implicated in a wide range of biological functions like telomere regulation and maintenance [9,10], transcription regulation [11,12], translation regulation [13,14], nucleosome positioning [15,16], CpG methylation [17], recombination [18,19] and recently in replication [20]. These studies together have brought more attention towards DNA and RNA G-quadruplexes (for recent review see [21,22]), requiring more tools and experimental setups to scrutinize the various biological functions

of DNA and RNA quadruplexes. Although various biophysical and biochemical methods have been employed to investigate the structure formation for DNA and RNA quadruplexes, (reviewed in [23]), their *in vivo* formation and relevance requires further experiments, preferably under near physiological conditions. In this regard, reporter gene expression assays are a most commonly used tool due to their relatively easy experimental setup, ready availability of the reagents and low cell requirements (~8–10,000 mammalian cells). The reporter constructs are easily prepared using simple cloning or site directed mutagenesis methods and the reporter expression results are sensitive and reproducible. In addition to the *in vivo* reporter assays described for characterizing quadruplex influences on gene expression there is the possibility to utilize *in vitro* transcription/translation assays utilizing cell extracts to express reporter genes in order to investigate the influence of RNA quadruplexes on translation reactions, for a description see [47–49].

2. Description of methods

The reporter assays are normally performed via transient transfection of plasmids carrying the G-quadruplex of interest e.g. in the promoter or 5'-UTR of the reporter gene. Simultaneously, similar constructs with mutation(s) in G-quadruplex forming sequences leading to loss in structure formation are often used for comparison of expression. By doing so, an effect of the quadruplex structure opposed to the mere presence of a G-rich sequence can be attributed. It is also essential to quantify the level of the corresponding mRNA either by northern blot [24] or quantitative real-time PCR [25] in order to shed light on the mechanistic basis of quadruplex action. In case of DNA G-quadruplexes it allows to validate changes

* Corresponding author. Fax: +49 7531 885140.

E-mail address: joerg.hartig@uni-konstanz.de (J.S. Hartig).

in the mRNA levels and thereby confirm a transcription regulation mechanism. For RNA G-quadruplexes it will clarify e.g. whether quadruplex formation is a means of translation regulation or whether it changes mRNA abundance e.g. by modulating mRNA stability.

2.1. Generation of reporter assay plasmids

The reporter gene with the G-quadruplex sequence at a desired site could be inserted via cloning using restriction enzymes or by site directed mutagenesis (Fig. 1). The former method is preferred for relatively long DNA sequences (for example gene promoters of ~100 to a few thousand nucleotides) while the later method is convenient for smaller sequences (~15–40 bp). Insertion of small 5'-UTR RNA G-quadruplex-forming sequences via site directed mutagenesis has been described in detail earlier [26]. There are various plasmid backbones (vectors) and reporter genes for expression available, though luciferase enzymes are most commonly used as reporter genes. Luciferase activity as reporter is easily determined using luminescence readers and the reagents are readily available. In case of the vector backbones, dual luciferase reporter systems (*Renilla* and firefly luciferase) are ideal because they are cloned on a single vector backbone with similar promoter and UTR constructs which rule out any discrepancy due to transfection or transcription discrepancy between the sample and control expression.

2.2. Cell line, propagation, harvest and lysis

Usually a matter of availability and choice, a wide range of mammalian cells has been used by various groups to perform reporter assays, mostly luciferase assays (Tables 1 and 2). It is recommended to let the cells grow for 24–72 h, as necessitated by the experiment, until the final confluence of the well reaches 90–100%. In general practice, ~10,000 cells seeded in a 96 well plate for 24 h incubation, followed by the transfection (plasmid with luciferase as reporter gene, Fig. 1) and another 24 h of incubation leads to 70–90% confluency before the luciferase assay. The recommended growth media for the cells can usually be found at <http://www.atcc.org>.

Each experimental condition should be performed at least in triplicates. For transfection, the transfecting agent (for example

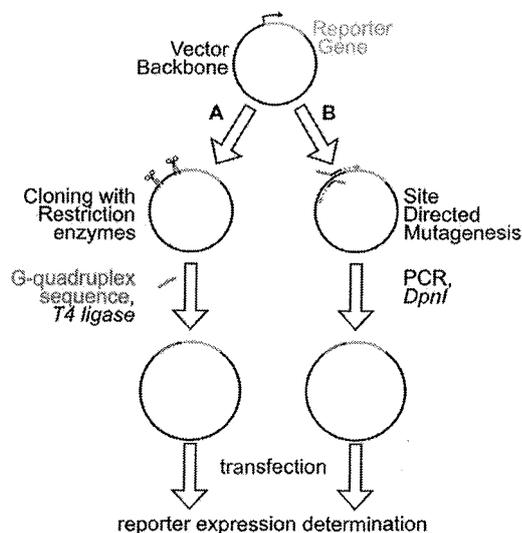


Fig. 1. Methods for inserting G-quadruplex forming sequences in front of a reporter gene by (A) using restriction enzymes or (B) site directed mutagenesis.

Lipofectamine 2000, turbofect, etc.) can be checked for compatibility with the cell line and plasmid size and has to be used according to manufacturer's protocol. However, the authors recommend using minimum antibiotic and serum-free media for the cells (for example 100 μ l for a single well in 96-well plate) during the transfection, which is believed to increase the transfection efficiency of the plasmids due to relatively higher concentration in the media. It is necessary to replenish with full media (media with serum, and antibiotic if used) in 2–3 h after transfection. In case that small molecules specific for G-quadruplexes are to be screened for their effect on expression, it is convenient to add them to the media at desired concentrations during the replenishment of full media. Importantly, such compounds should not be present during the transfection procedure since increased cellular compound uptake due to interaction with the transfection reagent might lead to increased toxicity.

For harvesting the cells, scrapping or using trypsin for adherent cell-lines is recommended only for large surface area, if used. However, for a 96-well plate, washing the cells with cold $1 \times$ PBS, pH 7.5 can be directly followed by cell lysis (usually supplied as a component of the reporter assay kit). The washing has to be carried out by careful and slow pipetting, so that the cells do not detach from the well surface. After addition of lysis buffer, it is important to pipette a couple of times for thorough lysis and homogenization, but also carefully to avoid any frothing. To decrease protein degradation, all lysates are to be kept on ice at all times.

2.3. Assay protocol

The reporter gene assay is usually to be followed as mentioned by the manufacturer's protocol. Usually a minimum of 50 μ l lysate volume is preferred for the assay as smaller volumes can decrease the reproducibility between the replicates. The luciferase activity is measured in relative light units (RLUs) and is required to be normalized by a secondary assay, for example β -gal or firefly/*Renilla* luciferase assay to account for any anomalies due to transfection efficiency between the samples. Usually a dual-luciferase system using a firefly and *Renilla* luciferase gene present on a single plasmid, for example the psiCHECK2 construct from Promega, leaves less ambiguity in luciferase activity arising due to differential transfection in samples. A detailed protocol for the Dual-Luciferase Reporter Assay System (Promega) using the psiCHECK2 plasmid as carried out in our group (see Section 4.3.1 for further information) is described below:

2.3.1. Transfection

- (1) 8000 HEK293 cells per well are seeded in flat-bottom 96 well plates in 200 μ l DMEM and grown to 50% confluency.
- (2) HEK293 cells will be transfected with the following plasmids: Plasmids with the G-quadruplex forming sequences 4G3U, 4G3U2 and 4G3U3 in the 5'-UTR, and plasmids with the corresponding control sequences con4G3U, con4G3U2 and con4G3U3. The experiment is performed in quadruplet as shown in the following scheme (Fig. 2):
- (3) For each well 19 μ l of 37 $^{\circ}$ C DMEM without FCS and antibiotics is mixed with 0.5 μ l of TurboFect in a reaction tube and incubated for 5 min at room temperature. (76 μ l DMEM and 2 μ l TurboFect for the quadruplet)
- (4) Plasmid stocks with 50 ng/ μ l concentration are used for transfection. 1 μ l of plasmid stock is added to the TurboFect mixture and incubate at room temperature for another 10 min. (4 μ l for the quadruplet)

Table 1

Studies using *chloramphenicol acetyltransferase* (CAT)^a and *luciferase*^b as reporter genes for determining the effect of DNA G-quadruplexes on expression.

Insert gene promoter (length)	Small molecule	Cell line	Ref (reporter gene)
<i>c-MYC</i>		DHL-9, Raji	[27] ^a
<i>c-MYC</i>		HeLa S3, A549	[28] ^a
<i>c-MYC</i> (~850 bp)	TmPyP4, TmPyP2	HeLa	[12,32] ^b
<i>c-MYC</i> (~850 bp)		HeLa	[31,33] ^b
<i>c-MYB</i> (~900 bp)		Flp-In 293	[34] ^b
Murine <i>KRAS</i> (300 bp)	TmPyP4, TmPyP2	MIA PaCa-2	[35] ^b
<i>c-MYC</i>	TmPyP4, TmPyP2, Se2SAP, telomestatin	Daoy	[36] ^b
<i>PDGFR-β</i> (2097 bp)		NIH 3T3	[37] ^b
<i>c-KIT</i> (125 bp)	Multiple compounds	HGC-27	[38] ^b

Table 2

Studies using luciferase as reporter genes for determining the effect of RNA G-quadruplexes on gene expression.

Insert gene UTR (length)	Small molecule	Cell line	Ref
<i>Zic-1</i> (27 bp)		HeLa	[43]
<i>MT3</i> (282 bp)		HeLa	[44]
Synthetic (15–33 bp), <i>MAPK2</i> (15 bp), <i>CHST2</i> (22 bp) and <i>PCGF2</i> (25 bp)		HEK293, LN18, B18	[45]
<i>EBAG9</i> , <i>FZD2</i> , <i>BARHL1</i> , <i>NCAM2</i> , <i>THRA</i> , <i>AASDHPPT</i> , <i>TNFSF12</i>		HEK 293	[46]
<i>ESR1</i> (~500 bp)		MCF-7	[47]
<i>ESR1</i> (~25 bp)		BST	[48]
<i>BCL-2</i> (493 bp)		MCF7, HGC27, MCF10A	[49]
<i>FGF-2</i> (484 bp)		SK-Hep1	[50]
p53		A549, MCF-7, HCT116	[51]
Synthetic (15–33 bp)	360A, PhenDC3, PhenDC6	HEK293	[53]

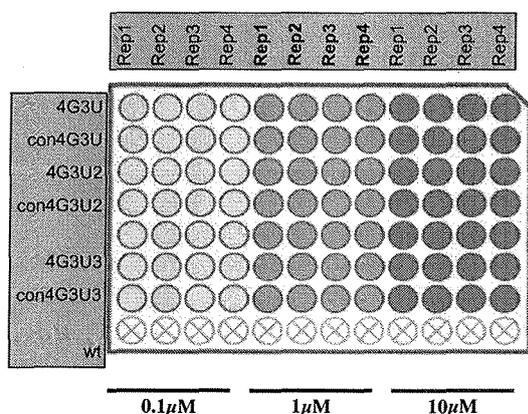


Fig. 2. Pipetting scheme for transfection of HEK293 cells followed by incubation with small molecules: HEK293 cells will be transfected with plasmids containing G-quadruplex forming sequences 4G3U, 4G3U2 and 4G3U3 in the 5'-UTR, and plasmids with the corresponding control sequences con4G3U, con4G3U2 and con4G3U3. Incubation with different concentrations of small molecules as indicated by the bars will be started 5 h after transfection.

- (5) Next 20 μ L of the culture medium is removed from the cells and replaced by 20 μ L of the TurboFect reaction mixture.
- (6) The plates were placed back in the 37 $^{\circ}$ C cell-culture incubator for 5 h.
- (7) Small molecules were added to the cells at the final concentrations of 0.1, 1 and 10 μ M.
- (8) The plates were placed back in the 37 $^{\circ}$ C cell-culture incubator for 19 h (about 70–90% cell confluency) followed by Dual Luciferase Reporter Assay.

2.3.2. Dual-Luciferase Reporter (DLR) Assay

In the Dual-Luciferase Reporter (DLR) Assay System (Promega), the activities of firefly (*Photinus pyralis*) and *Renilla* (*Renilla reniformis*, also known as sea pansy) luciferases are measured sequentially

from a single sample. The firefly luciferase reporter is measured first by adding Luciferase Assay Reagent II (LAR II) to generate a “glow-type” luminescent signal. After quantifying the firefly luminescence, this reaction is quenched, and simultaneously the *Renilla* luciferase reaction is initiated by adding the ‘Stop & Glo’ reagent to the same tube. The ‘Stop & Glo’ reagent also produces a “glow-type” signal from the *Renilla* luciferase, which decays slowly over the course of the measurement (Fig. 3).

- (1) 1 \times PLB is provided in the first two lanes of a clear 96-well plate with conical wells, so that a multichannel pipette can be used to transfer 25 μ L of PLB buffer to each well with transfected cells, as shown in the figure above.
- (2) After carefully removing the medium, the cells are once washed with 1 \times PBS.
- (3) 25 μ L of 1 \times PLB is added to each well. Incubate the cells on a plate-shaker for 15 min, 600 rpm and 37 $^{\circ}$ C.
- (4) In the meantime the luciferase assay reagent-II (LAR II) and 1 \times ‘stop and glo’ reagent are prepared. For each luciferase reaction 25 μ L of each LAR II and ‘stop and glo’ are

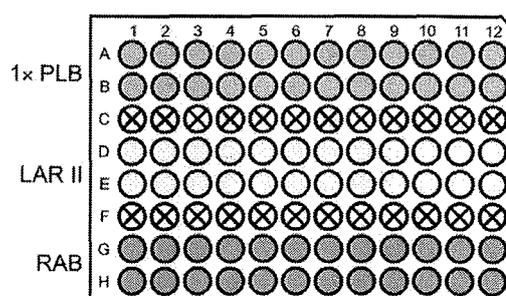


Fig. 3. Pipetting scheme for providing required solutions for the Dual-Luciferase Reporter assay: For easy handling it is recommended to provide the 1 \times PLB (grey), the LAR II (yellow) and the “Stop & Glo” solution (green) in a 96-well plate with conical wells, so that a multichannel pipette can be used for a fast transfer of the required volumes to the 96-well plate containing the cells/cell lysate and multiple samples can be processed simultaneously.

distributed in the clear 96-well plate with conical wells, as detailed in the figure above, so a multichannel pipette can be used for transfer.

- (5) After cell lysis, 20 μ L of lysate are transferred to a black half-surface 96-well microplate.
- (6) For measuring the firefly activity 25 μ L of LAR II reagent is added to each well of lysate and carefully mixed by pipetting 2–3 times. Firefly luminescence is immediately read-out in a luminescence reader. As the luminescence signal decays over time, it is advised not to scan more than four lanes at a time.
- (7) After firefly activity measurement 25 μ L of 'Stop & Glo' reagent is added to each well, carefully mixed by pipetting 2–3 times and *Renilla* luminescence is immediately read-out.

2.4. Data analysis

The β -gal or secondary protein (luciferase) activity is used to normalize the luciferase activity measurements (RLU_{luc}) by dividing the corresponding β -gal or secondary protein (luciferase) activity from the same well (RLU_{scd}). The ratios for each triplicate are averaged and the standard deviation is calculated, followed by re-normalization of the wild-type (no-insert) luciferase activity to 100. It is important to notice that the β -gal or secondary protein (luciferase) activity should show consistent values between samples during the time course of the experiment (usually 24–72 h), as otherwise the normalization is not useful since the β -gal or secondary protein (luciferase) activity is affected by the experimental conditions.

3. Reporter assays for studying DNA G-quadruplexes

3.1. *c-MYC* promoter DNA G-quadruplex as transcription repressor element

The most studied promoter region with respect to quadruplex functions in gene regulation is the *c-MYC* promoter. A variety of studies have been published with different setups: The human *c-MYC* promoter (919 bp) was cloned upstream of the chloramphenicol acetyltransferase (*CAT*) gene and co-transfected with NM23-H2 expression vector, resulting into 3- to 4-fold relative increase in *CAT* activity [27]. In contrast, a deletion of 56 bp of G-rich sequence (–157 to –108 bp relative to P1) did not show any elevated expression in presence of NM23-H2, suggesting NM23-H2 as a trans-activating factor for *c-MYC* gene via the G-rich sequence. In a similar study, mutations or deletion of *c-MYC* promoter (three constructs, –2402, –148 and –98 to +936 bp) containing the G-rich sequence (–138 to –111 bp) upstream of *CAT* gene in DHL-9 (B cell line) and Raji cells (Burkitt's lymphoma cell line) showed decreased *CAT* activity [28]. To our best knowledge, these two studies are the first observations which used a reporter gene (*CAT*) to elucidate the effect of a promoter G-rich sequence, later shown to form a DNA G-quadruplex motif [29].

Later, a firefly luciferase plasmid named 'Del-4' (pBC-Luc backbone), constructed at the Vogelstein Lab [30] harboring ~850 bp of the *c-MYC* promoter region including the 27 bp G-rich sequence (–142 to –115 bp upstream from the P1 promoter of *c-MYC*), was first used by Siddiqui-Jain et al. [12] to demonstrate that the DNA G-quadruplex motif is a transcription repressor element. The authors used a luciferase reporter assay in HeLa S3 cells for various single- or double-nucleotide mutations in the 27 bp DNA G-quadruplex motif to show the 'chair-form' of the G-quadruplex conformation is biologically relevant and a cationic porphyrin molecule, TmPyP4, can bind and further repress the luciferase expression (~50% inhibition of luciferase activity at 100 μ M treatment for 24 h).

The Del-4 construct was also used to show nucleolin [31] and NM23-H2 [32] as the *c-MYC* promoter DNA G-quadruplex binding proteins. In the first study, the authors used luciferase activity (Del-4 construct) as a measure of nucleolin binding and stabilizing of the *c-MYC* promoter DNA G-quadruplex motif in a concentration dependent manner in MCF-10A cells. While in the second study, the luciferase activity (Del-4 construct) was found to increase in presence of NM23-H2 but challenged by the cationic porphyrin, TmPyP4, in HeLa S3 and A549 cells. Though the mechanism is not fully understood, NM23-H2 appears to transactivate the *c-MYC* expression via the promoter-based DNA G-quadruplex motif.

In another study using a Del-4 construct, the authors showed that the luciferase activity is inducible with increasing concentrations of the polyamines spermidine and spermine (0–15 mM), suggesting that the *c-MYC* activation is governed by quadruplex–polyamine interactions [33].

3.2. Promoter DNA G-quadruplex in *c-MYB* gene is critical for promoter activity

Using a reporter plasmid containing –719 to +200 bp of the *c-MYB* promoter in front of a luciferase coding sequence, Palumbo et al. showed that the deletion of a 42 bp (+17 to +58 bp) G-rich sequence capable of forming three individual G-quadruplex motifs results in only 1–3% of residual promoter activity in three cell lines (CCRF-CEM, K562 and Jurkat), suggesting that the G-quadruplex-forming sequences are critical for transcription regulation [34]. In contrast, deletion of individual G-rich repeats (12 bp) capable of forming G-quadruplex motifs resulted in 3- to 5-fold relative increases in luciferase activities. The authors therefore speculated that at least one G-rich repeat is required for *c-MYB* promoter activity and the three consecutive repeats lead to further control of *c-MYB* expression by binding of proteins such as transcription factors to the G-quadruplex motif – dsDNA equilibrium.

3.3. Promoter DNA G-quadruplex in *PDGF-A* and *PDGFR- β* gene

The *PDGF-A* promoter (–222 to +119 bp) containing a 36 bp G-rich sequence (–82 to –47 bp) shown to form a stable DNA G-quadruplex motif was cloned upstream of a luciferase reporter in the pGL3-basic vector [35]. Transient transfection of the above construct in human pancreatic cancer MIA PaCa-2 cells were shown to decrease the luciferase activity with increasing TmPyP4 concentrations (0–50 μ M), suggesting the *in cellulo* formation of DNA G-quadruplex motifs. The same group later cloned the promoter region of *PDGFR- β* gene (five constructs, –1099, –556, –415, –264 and –140 bp upstream until +116 bp downstream) into the pGL3-basic vector in front of a luciferase gene [36]. The luciferase assay with these constructs in human medulloblastoma Daoy cells suggested an essential regulatory element between –264 and –140 bp which also had a G-quadruplex forming sequence (–165 to –139 bp). The DNA G-quadruplex was shown to be structured using circular dichroism (CD) spectroscopy, DMS footprinting and *Taq* polymerase stop assay. Moreover, using small molecules TmPyP2, TmPyP4, Se2SAP, and telomestatin, the authors showed strongest repression in luciferase activity with telomestatin (0–25 μ M), confirming the efficacy of telomestatin over other small molecules for targeting DNA G-quadruplexes.

3.4. Murine *KRAS* transactivation via promoter DNA G-quadruplex mediated by *MAZ* protein

A 300 bp fragment of the murine *KRAS* promoter cloned into the pGL3-1B basic vector containing a 34-bp G-rich sequence (–322 and –288 bp) upstream of the firefly luciferase gene was used in NIH 3T3 cells showed decreased expression when mutated at

specific sites that lead to disrupted G-quadruplex structures [37]. Moreover, when co-transfected with a MAZ-expressing plasmid, further activation of *KRAS* expression was observed, in line with the transactivation shown by NM23-H2 for the *c-MYC* gene [27,32].

3.5. Screening for DNA G-quadruplex-binding small molecule using luciferase assay

In a recent study, the authors cloned the *c-KIT* promoter (−124 to 0 bp) in the pGL4.10 vector upstream of a luciferase gene transiently transfected in human gastric carcinoma HGC-27 cells followed by incubation with an array of 173 small molecules at 0.2, 1 or 5 μ M concentrations [38]. The promoter region harbored two G-rich sequences shown earlier to form DNA G-quadruplex motif (*c-kit1*, −12 to −33 bp, [39]) and (*c-kit2*, −64 to −83 bp, [40]). The cell-based screening was able to identify two novel small molecules with high affinity and efficacy for *c-KIT* promoter DNA G-quadruplexes.

3.6. Furan-based cyclic homo-oligopeptides acting as selective G-quadruplex binders

The binding, stabilization and selectivity of two furan-based tricyclic homo-oligopeptides to the G-quadruplex of the 22-mer *c-MYC* sequence were characterized by CD spectroscopy, FRET melting assays and ITC experiments [41]. The impact of the two ligands *in vivo* was investigated using Real time PCR. HeLa cells were treated with increasing concentrations of the ligands (0, 50, 100 and 200 μ M) and the level of cMyc transcripts were measured. Almost 50% downregulation of the *c-Myc* transcript could be observed at 200 μ M of ligand 1 and 100 μ M of ligand 2. A cell viability assay revealed only limited toxicity of the ligands and apoptosis measurements showed their potential to induce apoptosis after 24 h.

In a later publication the same authors investigated two further, slightly modified furan-based cyclic homo-oligopeptides for their influence on the cMYC G-quadruplex sequence using the above mentioned *in vitro* methods [42]. The authors were able to improve the binding affinity to the *c-MYC* G-quadruplex sequence for both ligands, compared to the previously described homo-oligopeptides without losing the high selectivity for the G-quadruplex structure. Real time PCR measurements after treatment of HeLa cells with increasing ligand concentrations revealed a downregulation of cMYC mRNA transcript up to 90% at 200 μ M ligand concentration, while one of the ligands already showed 50% downregulation at 50 μ M. To investigate the downregulation of *c-MYC* on the protein level, the authors transfected HeLa cells with the Del 4 plasmid, which contains the 22-mer cMYC sequence upstream of the firefly luciferase promoter. Measurement of reporter gene expression 24 h after transfection revealed a decrease of about 60% for one of the ligands at 100 μ M final concentrations, compared to untreated cells. No significant changes in reporter gene expression were observed using the same plasmid with a mutated 22-mer *c-MYC* sequence which cannot fold into a G-quadruplex structure.

4. Luciferase assay for studying RNA G-quadruplexes

4.1. RNA G-quadruplexes in 5'-UTR

4.1.1. Translational inhibition by a G-quadruplex within the 5'-UTR of the *Zic-1* mRNA

The influence of a G-quadruplex-forming sequence within the 5'-UTR of the *Zic-1* mRNA was investigated using psiCHECK-2 [43]. This vector contains a dual luciferase reporter system, consisting of the *Renilla* and firefly luciferase under control of the

SV40 and HSV-TK promoters, respectively. The 27 nt sequence of the *Zic-1* RNA G-quadruplex was inserted into the 5'-UTR of the *Renilla* luciferase mRNA, 11 nt upstream of the translational start codon. Transient transfection of HeLa cells with this construct showed a decrease of *Renilla* luciferase expression of ~80%, whereas no influence could be observed in case of a mutated control sequence. Evaluation of the mRNA levels of the transfected cells revealed that this inhibition was due to the inhibition of translation.

4.1.2. G-quadruplex within 5'-UTR of *MT3* matrix metalloproteinase mRNA as translational repressor

The entire 5'-UTR (282 nt) of the *MT3* matrix metalloproteinase mRNA, which contains a G-quadruplex forming sequence at positions 211–230 nt with respect to the 5'-end, was inserted into the psiCHECK-2 vector upstream of the *Renilla* luciferase reporter gene [44]. Transient transfection of HeLa cells with this construct revealed an inhibition of reporter gene expression of 55% compared to the luciferase expression of the wild type psiCHECK-2 vector. At the same time no changes in the level of reporter gene expression were observed using a mutated control construct containing specific mutations to prevent G-quadruplex formation. To ensure that this inhibition of reporter gene expression is due to the G-quadruplex forming sequence, only the G-quadruplex forming sequence or the mutated variant were inserted into the psiCHECK-2 vector upstream of the *Renilla* luciferase gene in close proximity to the translation start site. Measurement of the luciferase expression revealed a 60% reduction in case of the G-quadruplex forming sequence compared to the wild type psiCHECK-2 vector. Again, no alterations in reporter gene expression were observed in case of the non-quadruplex, mutated sequence.

4.1.3. Predictable influence of gene expression by G-quadruplexes within the 5'-UTR

Artificial G-quadruplex forming sequences that differed in the number of G-tracts as well as in the length of the interspersed loops were cloned 11 nt upstream of the start codon of the *Renilla* luciferase reporter gene located on the psiCHECK-2 vector [45]. Transient transfection of HEK293 cells with these constructs resulted in inhibition of reporter gene expression in a loop-length as well as number of G-tracts dependent manner, while no changes in the level of mRNA could be detected as determined by realtime PCR. In addition, no influence on the level of reporter gene expression could be observed using the same constructs with specific point mutations preventing G-quadruplex formation. Next, the naturally occurring G-quadruplex-forming sequences from *NRAS*, *MAPK2*, *CHST2* and *PCGF2* were inserted into the psiCHECK-2 vector at the same position. Transient transfection of HEK293 cells showed similar inhibition of reporter gene expression in comparison to the artificial G-quadruplex-forming sequences with equal number of G-tracts and equal loop length, indicating a predictable influence of quadruplexes on gene expression with regards to these parameters. Consistent results were obtained by transient transfection of HeLa, human LN18 and mouse B18 cells with the constructs described.

4.1.4. Investigation of computationally searched G-quadruplex-forming sequences in 5'-UTRs

Beaudoin and Perrelaut analyzed databases of 5'-UTRs for potential G-quadruplexes using bioinformatics tools [46]. Sequences of nine different 5'-UTRs were chosen for further investigations based on the importance of the encoded proteins in cellular pathways. Out of these nine candidates, six (*EBAG9*, *FZD2*, *BARHL1*, *NCAM2*, *THRA* and *AASDHPPT*) showed G-quadruplex formation in *in vitro* experiments. These six 5'-UTRs were cloned into the

pRL-TK vector upstream of the *Renilla* luciferase reporter gene. HEK 293 cells were transiently transfected with these constructs or constructs in which the G-quadruplex forming sequence was mutated and cotransfected with the pGL3-control vector (firefly luciferase). For all six mutated G-quadruplex-forming sequences an increase in reporter gene expression ranging from 1.56- to 2.50-fold in comparison to the wild type G-quadruplex sequence was detected. At the same time similar levels of mRNA were observed. In addition, the three remaining G-quadruplex-forming sequences from *TNFSF12*, *MAP3K11* and *DOC2B*, containing many cytosines, did not fold into a G-quadruplex structure. Important cytosines within these sequences, thought to create stem structures with the guanines, were randomly mutated to adenines and the resulting constructs were tested *in vitro* as well as *in cellulo* using the previously described luciferase assay. Thereby the authors were able to show that the specific exchanges from cytosines to adenines can transform these sequences into a G-quadruplex-forming structure. The corresponding transfection experiments in HEK293 cells further revealed the potential of the transformed sequences for their translational repression of reporter gene expression. Conversely, using the same methods a G to C exchange carried out in the G-quadruplex forming sequence of the *AASDHPPT* 5'-UTR was shown to lead to an increase in reporter gene expression, indicative for the presence of a disrupted or at least weakened G-quadruplex structure.

4.1.5. The 5'-UTR transcribed from the exon C of the *ESR1* gene

A fragment of about 500 nt of the human exon C of the *ESR1* gene (positions 1726–2235 nt), including a G-quadruplex-forming sequence was inserted between a T7 promoter and the firefly luciferase reporter gene of the pGL3 plasmid, with the G-quadruplex forming sequence (positions 1881 bp–1904 nt) located 154 nt downstream of the T7 promoter sequence [47]. As a control one of the guanosine tracts was exchanged into adenines to avoid G-quadruplex formation. *In vitro* transcription followed by *in vitro* translation revealed a 6-fold increase of translation for the mutated construct compared to the G-quadruplex forming sequence, indicative of an inhibitory effect of the G-quadruplex in the translation of this mRNA.

In another work the same group investigated a unique G-quadruplex forming sequence located within the exon C sequence of the bovine *ESR1* gene. The 5'-UTR of the bovine *ESR1* gene was inserted into the pGL3-Basic vector upstream of the firefly luciferase reporter gene and 12 nt downstream of the T7 promoter sequence [48]. Next, *in vitro* transcription/translation experiments were performed. A construct in which the G-quadruplex sequence was deleted and a construct in which one G-tract was exchanged against adenines showed a 2-fold and a 15-fold increase in translation efficiency, respectively, in comparison to the G-quadruplex containing construct. For further investigations of this G-quadruplex forming sequence, BST cells were transiently transfected with the described sequences inserted upstream of the firefly luciferase gene on the pGL3-control vector. Measurement of luciferase expression was normalized to the expression of β -galactosidase, originating from a co-transfected control plasmid. In comparison to the G-quadruplex-containing sequence a greater than 15-fold increase of reporter gene expression was observed with the construct encoding the deleted G-quadruplex sequence, while in case of the construct with the mutated G-quadruplex sequence a 7-fold increase was detected.

4.1.6. Modulation of gene expression by a G-quadruplex within the 5'-UTR of the *BCL-2* mRNA

The entire naturally occurring 5'-UTR of the *BCL-2* mRNA (493 nt) was inserted upstream of the firefly luciferase reporter gene into the pUC18 vector [49]. As a control the G-quadruplex

forming sequence was deleted. The influence of the G-quadruplex forming sequence was investigated by *in vitro* transcription/translation experiments in rabbit reticulocyte lysate, showing the potential of this secondary structure to inhibit mRNA translation. Subsequently these constructs were transferred into a CMV promoter-driven mammalian expression vector upstream of the firefly luciferase reporter gene. MCF10A, MCF7 and HGC27 cells were transiently transfected with these constructs and a quadruplex-mutated construct and co-transfected with a pRL-TK vector for normalization. Expression of luciferase was measured 24 h after transfection and an increase of luciferase expression of 1.9-fold and 2.3-fold was observed in the MCF10A cells for the mutated and the deleted constructs, respectively. Transfection of MCF7 and HGC27 cells revealed qualitatively comparable results, but with a reduced translational suppressive effect in case of the HGC27 cells. No significant changes in the levels of mRNA were detected by quantitative realtime-PCR.

4.1.7. A G-quadruplex motif within the alternative translation region of the *FGF-2* mRNA

The *FGF-2* mRNA contains a 5' alternative translation region (ATR) of 484 nt. This ATR comprises four CUG and one AUG initiation codons [50]. While initiation of translation of the first CUG codon is cap-dependent, the other four initiation codons are initiated by internal entry of ribosomes. Analysis of the secondary structure of the ATR by chemical and enzymatic probing experiments revealed a G-quadruplex-forming motif among other structures. To investigate the elements responsible for internal ribosome entry site (IRES) activation, a bicistronic vector system was used, containing an upstream *Renilla* luciferase open reading frame and a downstream firefly luciferase open reading frame. While translation initiation of the upstream cistron is cap-dependent, the downstream cistron is IRES-dependent. Different constructs with one to four translation initiation codons in the ATR located in the intercistronic space were used for *in vivo* experiments. Upon transient transfection of SK-Hep 1 cells, already shown to harbor *FGF-2* IRES activity, the obtained results revealed that a single IRES module is responsible for initiation of all four initiation codons. Further transfection experiments of SK-Hep 1 cells with a variety of deletion constructs of the ATR revealed a 176 nt region to be sufficient for IRES functionality, consisting of a G-quadruplex and two stem-loop motifs.

4.2. RNA G-quadruplexes in 3'-UTR

4.2.1. A G-quadruplex downstream of the polyadenylation signal modulates p53 pre-mRNA 3'-end processing

In a recent publication the authors inserted the 3'-end region of the p53 pre-mRNA downstream of the *Renilla* luciferase reporter gene into the pRL-CMV vector [51]. This 3'-end region contains a G-quadruplex forming sequence, predicted by bioinformatic analysis, at positions +108–+142 nt with regard to the pre-mRNA 3' end cleavage site. The presence of a G-quadruplex was analyzed by a reverse transcriptase elongation assay, using the avian myeloblastosis virus (AMV) reverse transcriptase. Transient transfection of HeLa cells with this construct revealed a 2.5-fold increase in luciferase gene expression in comparison to a control construct in which the G-quadruplex forming sequence was deleted. A549 cells were transiently transfected with the p53 3'-end region construct or a construct with a mutated G-quadruplex sequence. Subsequently, the cells were exposed to UV light or remained untreated, followed by analysis of 3'-end-processing efficiency by RT-qPCR. A 6- or 4-fold decrease, respectively in 3'-end processing was observed for the mutated construct compared to the wild type construct, indicative for an involvement of the G-quadruplex forming sequence in this processing step.

4.2.2. The influence of a G-quadruplex sequence within the 3'-UTR of PIM1

Using Quadfinder the authors identified a G-rich sequence within the 3'-UTR of the human PIM1 mRNA, 277 nt downstream of the stop codon. Biophysical experiments with a synthetic 22-mer oligonucleotide of the G-rich sequence revealed this sequence to fold into a parallel G-quadruplex structure [52]. For *in vivo* experiments the 22-mer G-quadruplex sequence and a sequence consisting of 100 nt of the natural 3'-UTR of the PIM1 mRNA with the G-quadruplex sequence in the center were inserted into the psi-CHECK 2 plasmid downstream of the *Renilla* luciferase stop codon. For controls specific point mutations were inserted which prevent G-quadruplex formation. HeLa cells were transiently transfected with these constructs. Expression of *Renilla* luciferase was measured 24 h after transfection and normalized to the expression of firefly luciferase. A 2-fold reduction of *Renilla* luciferase expression was observed for both constructs compared to the corresponding mutated control sequences. Analysis of the mRNA levels using RT-qPCR revealed no effect, indicating that the inhibition of reporter gene expression was based on the level of mRNA translation.

4.3. Screening for RNA G-quadruplex binding-small molecule using luciferase assay

4.3.1. Bisquinolinium compounds as potent RNA G-quadruplex stabilizers

In a recent publication our group inserted artificial G-quadruplex forming sequences, previously shown to inhibit gene expression, 11 nt upstream of the start codon of the *Renilla* luciferase reporter gene into the psiCHECK-2 vector [53]. HEK293 cells were transiently transfected with these constructs and subsequently incubated with the bisquinolinium compounds 360A, PhenDC3 and PhenDC6, which have already been shown to be able to stabilize DNA G-quadruplexes, using final compound concentrations between 0.1 and 10 μ M. As controls G-rich constructs with specific point mutations preventing quadruplex formation were used. Measurement of the luciferase activity revealed a further decrease in reporter gene expression in a dose-dependent manner, compared to the cells which were not incubated with the compounds. No influence of the compounds on the level of reporter gene expression was observed using the corresponding control constructs, demonstrating that the addition of these compounds is able to bind to and stabilize the G-quadruplex structure *in vivo*.

5. Concluding remarks

Here we have described various assays being used to determine and target the biological function of DNA and RNA G-quadruplexes. These assays are useful tools in *in vitro* as well as *in vivo* applications to study the influence of known G-quadruplexes or potential G-quadruplex-forming sequences on gene expression. As described, a variety of different systems have been studied utilizing reporter assays for gene expression. Especially in cell culture systems, dual luciferase assays offer easy handling and preparation to examine the biological function of DNA or RNA G-quadruplexes under physiological conditions. Taken together, the results indicate specific roles of quadruplexes that are dependent on the respective position and sequence context of the studied quadruplexes.

Acknowledgements

J.S.H. thanks the *VolkswagenStiftung* for funding a Lichtenberg-Professorship. K.H. thanks the Alexander von Humboldt-Foundation for a fellowship.

References

- [1] Y. Qin, L. Hurley, *Biochimie* 90 (2008) 1149–1171.
- [2] Y. Sannohe, H. Sugiyama, *Curr Protoc Nucleic Acid Chem Chapter 17* (2010) Unit 17.12.11–17.
- [3] P. Hazel, J. Huppert, S. Balasubramanian, S. Neidle, *J. Am. Chem. Soc.* 126 (2004) 16405–16415.
- [4] A. Bugaut, S. Balasubramanian, *Biochemistry* 47 (2008) 689–697.
- [5] A. Zhang, A. Bugaut, S. Balasubramanian, *Biochemistry* 50 (2011) 7251–7258.
- [6] P. Rachwal, T. Brown, K. Fox, *Biochemistry* 46 (2007) 3036–3044.
- [7] D. Zhang, T. Fujimoto, S. Saxena, H. Yu, D. Miyoshi, N. Sugimoto, *Biochemistry* 49 (2010) 4554–4563.
- [8] C. Tang, R. Shafer, *J. Am. Chem. Soc.* 128 (2006) 5966–5973.
- [9] S. Neidle, *FEBS J.* 277 (2010) 1118–1125.
- [10] A.J. Zaig, E.R. Podell, T.R. Cech, *PNAS* 102 (2005) 10864–10869.
- [11] A. Verma, K. Halder, R. Halder, V. Yadav, P. Rawal, R. Thakur, F. Mohd, A. Sharma, S. Chowdhury, *J. Med. Chem.* 51 (2008) 5641–5649.
- [12] A. Siddiqui-Jain, C.L. Grand, D.J. Bearss, L.H. Hurley, *PNAS* 99 (2002) 11593–11598.
- [13] S. Kumari, A. Bugaut, J. Huppert, S. Balasubramanian, *Nat. Chem. Biol.* 3 (2007) 218–221.
- [14] D.J. Patel, A.T. Phan, V. Kuryavyi, *Nucleic Acids Res.* 35 (2007) 7429–7455.
- [15] K. Halder, R. Halder, S. Chowdhury, *Mol. Biosyst.* 5 (2009) 1703–1712.
- [16] H. Wong, J. Huppert, *Mol. Biosyst.* 5 (2009) 1713–1719.
- [17] R. Halder, K. Halder, P. Sharma, G. Garg, S. Sengupta, S. Chowdhury, *Mol. Biosyst.* 6 (2010) 2439–2447.
- [18] F. Boan, J. Gomez-Marquez, *Chem. Biochem.* 11 (2010) 331–334.
- [19] P. Mani, V. Yadav, S. Das, S. Chowdhury, *PLoS One* 4 (2009) e4399.
- [20] K. Paeschke, J. Capra, V. Zakian, *Cell* 145 (2011) 678–691.
- [21] S. Balasubramanian, L.H. Hurley, S. Neidle, *Nat. Rev. Drug Discov.* 10 (2011) 261–275.
- [22] K. Halder, J.S. Hartig, *Met. Ions Life Sci.* 9 (2011) 125–139.
- [23] S. Burge, G.N. Parkinson, P. Hazel, A.K. Todd, S. Neidle, *Nucleic Acids Res.* 34 (2006) 5402–5415.
- [24] S. Streit, C. Michalski, M. Erkan, J. Kleeff, H. Friess, *Nat. Protoc.* 4 (2009) 37–43.
- [25] T. Nolan, R. Hands, S. Bustin, *Nat. Protoc.* 1 (2006) 1559–1582.
- [26] M. Wieland, J.S. Hartig, *Nat. Protoc.* 4 (2009) 1632–1640.
- [27] S. Berberich, E. Postel, *Oncogene* 10 (1995) 2343–2347.
- [28] L. Ji, M. Arcinas, L.M. Boxer, *J. Biol. Chem.* 270 (1995) 13392–13398.
- [29] T. Simonsson, P. Pecinka, M. Kubista, *Nucleic Acids Res.* 26 (1998) 1167–1172.
- [30] T.-C. He, A.B. Sparks, C. Rago, H. Hermeking, L. Zawel, L.T. da Costa, P.J. Morin, B. Vogelstein, K.W. Kinzler, *Science* 281 (1998) 1509–1512.
- [31] V. Gonzalez, K. Guo, L. Hurley, D. Sun, *J. Biol. Chem.* 284 (2009) 23622–23635.
- [32] R.K. Thakur, P. Kumar, K. Halder, A. Verma, A. Kar, J.-L. Parent, R. Basundra, A. Kumar, S. Chowdhury, *Nucleic Acids Res.* 37 (2009) 172–183.
- [33] N. Kumar, R. Basundra, S. Maiti, *Nucleic Acids Res.* 37 (2009) 3321–3331.
- [34] S.L. Palumbo, R.M. Memmott, D.J. Uribe, Y. Krotova-Khan, L.H. Hurley, S.W. Ellinghaus, *Nucleic Acids Res.* 36 (2008) 1755–1769.
- [35] Y. Qin, E.M. Rezler, V. Gokhale, D. Sun, L.H. Hurley, *Nucleic Acids Res.* 35 (2007) 7698–7713.
- [36] Y. Qin, J. Fortin, D. Tye, M. Gleason-Guzman, T. Brooks, L. Hurley, *Biochemistry* 49 (2010) 4208–4219.
- [37] S. Cogoi, M. Paramasivam, A. Membrino, K.K. Yokoyama, L.E. Xodo, *J. Biol. Chem.* 285 (2010) 22003–22016.
- [38] K. McLuckie, Z. Waller, D. Sanders, D. Alves, R. Rodriguez, J. Dash, G. McKenzie, A. Venkitaraman, S. Balasubramanian, *J. Am. Chem. Soc.* 133 (2011) 2658–2663.
- [39] S. Rankin, A.P. Reszka, J. Huppert, M. Zloh, G.N. Parkinson, A.K. Todd, S. Ladame, S. Balasubramanian, S. Neidle, *J. Am. Chem. Soc.* 127 (2005) 10584–10589.
- [40] H. Fernando, A.P. Reszka, J. Huppert, S. Ladame, S. Rankin, A.R. Venkitaraman, S. Neidle, S. Balasubramanian, *Biochemistry* 45 (2006) 7854–7860.
- [41] T. Agarwal, S. Roy, T. Chakraborty, S. Maiti, *Bioorg. Med. Chem. Lett.* 20 (2010) 4346–4349.
- [42] T. Agarwal, S. Roy, T. Chakraborty, S. Maiti, *Biochemistry* 49 (2010) 8388–8397.
- [43] A. Arora, M. Dutkiewicz, V. Scaria, M. Hariharan, S. Maiti, J. Kurreck, *RNA* 14 (2008) 1290–1296.
- [44] M. Morris, S. Basu, *Biochemistry* 48 (2009) 5313–5319.
- [45] K. Halder, M. Wieland, J.S. Hartig, *Nucleic Acids Res.* 37 (2009) 6811–6817.
- [46] J.-D. Beaudoin, J.-P. Perreault, *Nucleic Acids Res.* 38 (2010) 7022–7036.
- [47] G. Balkwill, K. Derecka, T. Garner, C. Hodgman, A. Flint, M. Searle, *Biochemistry* 48 (2009) 11487–11495.
- [48] K. Derecka, G. Balkwill, T. Garner, C. Hodgman, A. Flint, M. Searle, *Biochemistry* 49 (2010) 7625–7633.
- [49] R. Shahid, A. Bugaut, S. Balasubramanian, *Biochemistry* 49 (2010) 8300–8306.
- [50] S. Bonnal, C. Schaeffer, L. Creancier, S. Clamens, H. Moine, A.-C. Prats, S. Vagner, *J. Biol. Chem.* 278 (2003) 39330–39336.
- [51] A. Decorsiere, A. Cayrel, S. Vagner, S. Millevoi, *Genes & Dev.* 25 (2011) 220–225.
- [52] A. Arora, B. Suess, *RNA, Biol* 8 (2011).
- [53] K. Halder, E. Laryg, M. Benzler, M. Teulader-Fichou, J. Hartig, *Chem. Biochem.* 12 (2011) 1663–1668.