Insights into the characteristics and functions of G-quadruplexes

Dissertation zur Erlangung des akademischen Grades eines Doktors der Naturwissenschaften (Dr. rer. nat.)

vorgelegt von

Martin Benzler

an der

Universität Konstanz

Mathematisch-Naturwissenschaftliche Sektion

Fachbereich Chemie

Tag der mündlichen Prüfung: 22. März 2017
1. Referent: Professor Dr. Jörg Hartig
2. Referent: Professor Dr. Andreas Marx

Meinen Eltern

Table of contents

1. Introduction

1.1 Deoxyribonucleic acid

All living organisms as well as some viruses use the molecule deoxyribonucleic acid (DNA) for encoding their genome. The most prominent structure of DNA is the right-handed DNA double helix, also called B-DNA, discovered by James Watson and Francis Crick in 1953 [1]. DNA represents a polymer of nucleotides (Fig. 1.1).

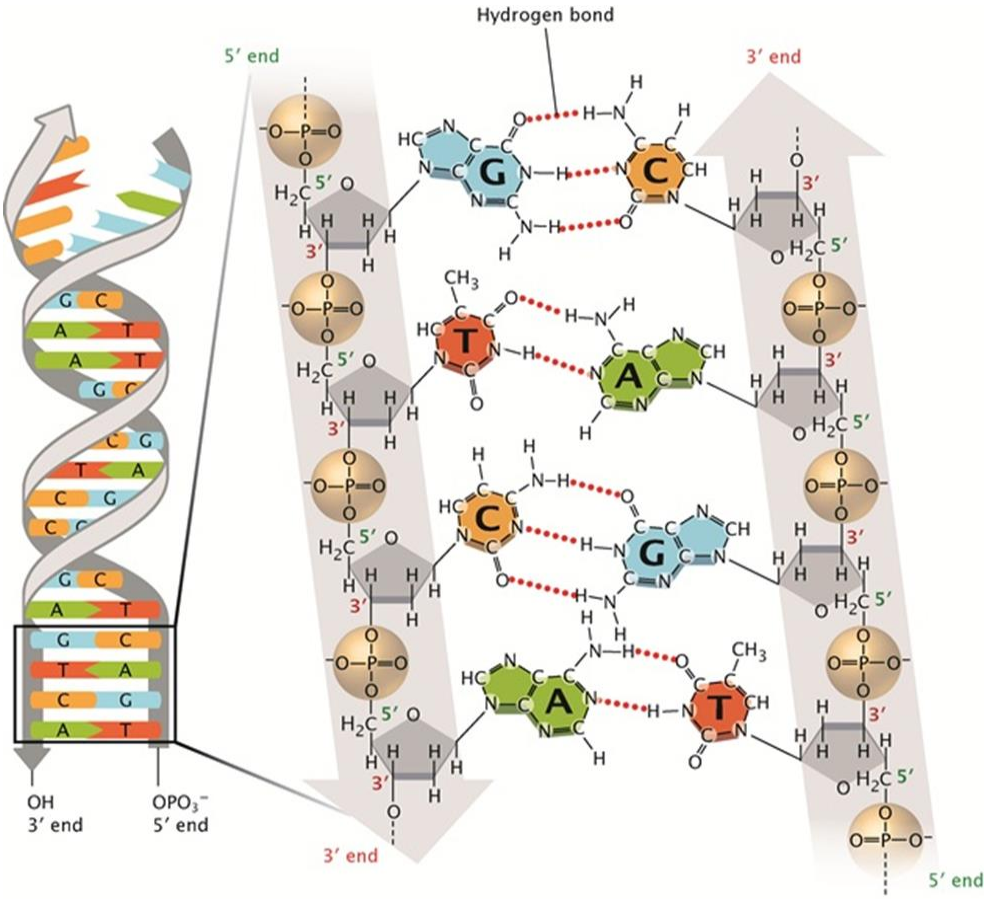


Fig. 1.1: Structure of DNA. Two anti-parallel DNA strands form a right-handed double helical structure in which guanine (G; blue) is connected via three hydrogen bonds to cytosine (C; orange) and adenine (A; green) is connected via two hydrogen bonds to thymine (T; red). The sugar-phosphate backbone, underlined in gray, gives each DNA strand a direction from the 5´ end with the free phosphate group to the 3´ end with the free hydroxyl group. (Figure from [2])

Each nucleotide consists of a nucleobase, a pentose and a phosphate group. Within the DNA the nucleotides are connected via phosphodiester bonds

between the 5´ carbon atom of the pentose (2´ deoxyribose) and the 3´ carbon atom of the following pentose. These repeats of alternating sugar and phosphate residues form the backbone of the DNA, thereby giving the DNA strand a direction from the 5´ end with a terminal free phosphate group to the 3´ end with the terminal free hydroxyl group of the sugar (Fig. 1.1). The nucleobases are connected via glycosidic bonds to the 1´ carbon of the pentose (Fig. 1.2 A). The four nucleobases found in DNA, adenine (A), guanine (G), thymine (T) and cytosine (C), are subdivided into purines (A and G) and pyrimidines (C and T) (Fig. 1.2 B).

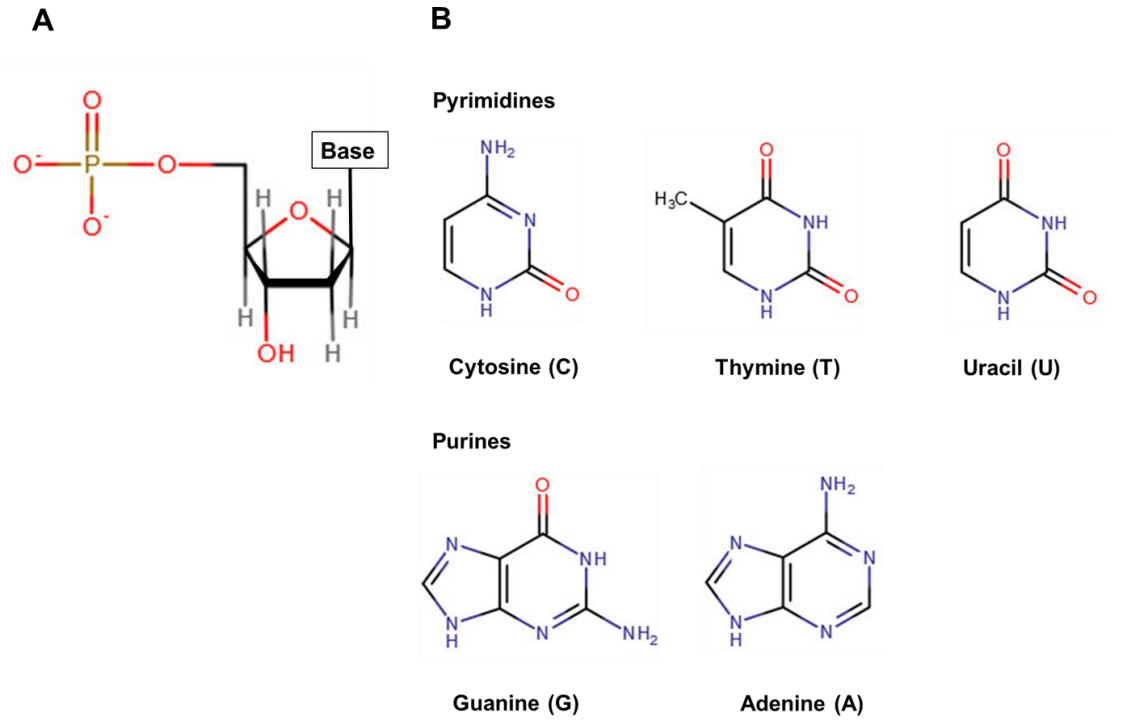


Fig. 1.2: Structure of a nucleotide and the nucleobases. A: A nucleotide consists of a nucleobase (Base), a pentose and a phosphate group. B: Nucleobases can be divided into the pyrimidines cytosine (C), thymine (T) and uracil (U) and the purines adenine (A) and guanine (G). While DNA contains the four nucleobases cytosine, guanine, adenine and thymine, the nucleobase thymine is replaced by uracil in RNA.

The DNA double helix is composed of two DNA strands in anti-parallel orientation in which the two strands are connected via hydrogen bonds between the nucleobases of each strand (Fig. 1.1). In this so-called complementary base

pairing adenine is connected to thymine via two hydrogen bonds and cytosine to guanine via three hydrogen bonds (Fig. 1.1). These hydrogen bonds are named after their discoverer as Watson-Crick base pairs. The stability of a DNA helix is mainly determined by the hydrogen bonds between the nucleobases and with a stronger impact by base-stacking interactions (π-π interactions) among the aromatic nucleobases [3].

While the most common structure of DNA in organisms is the B-DNA, also other double-stranded helical structures like the A-DNA [4] or Z-DNA [5,6] can be formed, although only the B-DNA and Z-DNA have been directly observed within organisms. Beside these double-stranded helical structures, DNA can also fold into other non-canonical structures, e.g. a triplex structure, in which a single DNA strand hybridizes with a DNA double strand [7-9], hairpin and cruciform structures, which include palindromic sequences [10-12], the i-motif, which comprises of two parallel duplexes with intercalated hemiprotonated cytosine^{+}-cytosine base pairs [13], or the Guanine-quadruplex, formed by sequences with at least four runs of two or more guanines [14] (cf. chapter 1.3). While these higher order structures were initially thought to be *in vitro* artefacts, bioinformatical analysis showed that such putative sequences with the potential to fold into different secondary and tertiary structures are highly conserved among the whole genome and therefore supporting the theory of their *in vivo* relevance [15-17]**.** Furthermore, there is evidence for these structures based on *in vitro* as well as in *vivo* studies that they are involved in regulatory processes, e.g. DNA transcription, stability of chromosomes, DNA recombination, cellular senescence and viral integration [18-24].

It is important to mention that DNA has to be considered as a dynamic structure which is not only dependent on the corresponding sequence but also on surrounding conditions, e. g. pH, ion concentrations, temperature, the presence of the complementary strand or interacting proteins. Therefore, alternative DNA structures may exist transiently instead of being in a steady-state form within the genome [25-29].

1.2 Ribonucleic acid

Ribonucleic acid (RNA), the molecule originated by the process of DNA transcription, is a multifunctional molecule involved in various processes in organisms, e.g. coding [30], regulation and expression of genes [31], but also catalyzing of biological reactions [32] or sensing and communicating responses to cellular signals [33].

Similar to DNA, RNA is assembled as a chain of nucleotides. However, there are three main differences between DNA and RNA: First, the ribose within the RNA backbone contains a hydroxyl group at the 2´ position. Second, while RNA, analog to DNA, contains the bases guanine, cytosine and adenine, thymine is replaced by uracil (U) (Fig. 1.2 B). Third, unlike DNA, RNA occurs mostly single-stranded [34]. Likewise to DNA, RNA can adopt a broad variety of different structures, including the Guanine-quadruplex motif [35,36]. However, the probability for RNA to form higher order structures might be increased by the absence of a complementary strand [37].

1.3 Guanine-quadruplexes

One of the most prominent non-canonical structures formed by nucleic acid sequences is the so-called Guanine-quadruplex (G-quadruplex). Already in 1910 Bang reported that guanylic acid forms polycrystalline gels at high concentration in aqueous solution [38]. However, the underlying structure was solved about 50 years later in 1962 by Gellert *et al.* via X-ray diffraction [14].

1.3.1 Structure of G-quadruplexes

The helical higher order structure of a G-quadruplex can be formed by G-rich DNA, RNA, PNA (peptide nucleic acid) or LNA (locked nucleic acid) sequences either within a single strand (intramolecular) or by multiple strands (intermolecular) [39]. Sequences with a minimum of four interspersed G-tracts (also named G-repeats or G-runs), which consist of two or more consecutive guanines, are able to form stable intramolecular G-quadruplex motifs. These conformations are composed of at least two stacking guanine tetrads (also

named guanine quartets) interacting via Hoogsteen hydrogen bonds and are stabilized by monovalent metal ions like potassium or sodium [40].

As shown in Fig. 1.3, four guanines form a (nearly) square planar tetrad in which the guanines interact via Hoogsteen hydrogen bonds between N2 and N7 and between N1 and O6 on adjacent guanines. Two or more of such G-tetrads can stack on top of each other and are stabilized by π-π interactions between the aromatic bases. This structure can be further stabilized by monovalent metal ions like potassium or sodium. The nucleotides between the G-tracts form loops which are extruded of the stacks and also influence the stability as well as the structure of the G-quadruplex depending on their number as well as their sequence [41-43].

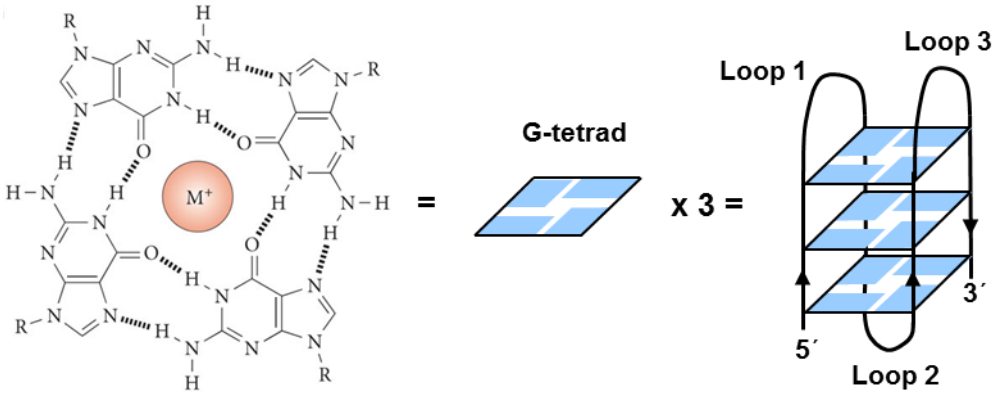


Fig. 1.3: Structure of a G-quadruplex. Four guanines form a square planar tetrad (G-tetrad) in which the guanines interact via Hoogsteen hydrogen bonds (Figure modified from [44]). Two or more of such G-tetrads stack on top of each other and are stabilized by π-π interactions between the aromatic bases. This structure can be further stabilized by monovalent cationic ions (M^+).

The sequence of an intramolecular G-quadruplex can be described with the following formula:

$$G_mX_nG_mX_oG_mX_pG_m$$

In this formula m describes the number of guanine residues in each G-tract and X_n, X_o and X_p describe the number and composition of nucleotides within the loops. The G-tracts have not to be identical and therefore the number of

guanine residues within a G-tract has not necessarily to be of equal length. In case one G-tract consists of more guanine residues than another, also guanine residues of the longer G-tract can be located within the loop region. As mentioned above a G-quadruplex can also be formed intermolecular by association of two (i) or four (ii) strands. Sequences of intermolecular bi- or tetramolecular G-quadruplexes can be described with the formula:

(i) $X_nG_mX_oG_mX_p$ (ii) $X_nG_mX_o$

The involved sequences forming the intermolecular G-quadruplex have not necessarily to be identical, however most of the described intermolecular G-quadruplexes are composed of identical sequences [39].

In general, a sequence, which is able to fold into an intramolecular G-quadruplex, is also able to fold into an intermolecular G-quadruplex, whereas higher strand concentrations promote the formation of an intermolecular G-quadruplex. While an intramolecular G-quadruplex consists of a sequence with at least four interspersed G-tracts, an intermolecular G-quadruplex needs a minimum of two G-tracts for a bimolecular G-quadruplex and at least one G-tract for a tetramolecular G-quadruplex [39].

1.3.2 Topologies of G-quadruplexes

G-quadruplexes display a broad variety of topologies, influenced by different parameters like the number of G-tracts, the sequence and size of the loops, the orientation of the strands, the stoichiometry and alignment and also the nature of the binding cations.

1.3.2.1 Strand orientation

With regard to the strand orientation from the 5´ to the 3´ end of the sequence, four different combinations of the strand orientations within a DNA G-quadruplex may arise (independently of an intra- or intermolecular G-quadruplex) (Fig. 1.4): (1) A G-quadruplex consisting of four strands with all strands having the same strand orientation is called a parallel G-quadruplex (Fig. 1.4 A). (2) A G-quadruplex with three strands in the same direction and

one strand in the opposite direction is named a mixed parallel/antiparallel (3+1) hybrid G-quadruplex (Fig. 1.4 B). (3) In terms a G-quadruplex with two neighboring strands having the same strand orientation and the remaining two strands having the opposite direction it is called an antiparallel G-quadruplex (Fig. 1.4 C). (4) This is also the case for a G-quadruplex consisting of two diagonal strands with the same orientation and the two remaining diagonal strands with the opposite orientation (Fig. 1.4 D).

Unlike DNA G-quadruplexes, RNA G-quadruplexes were found to be only monomorphic. This circumstance seems to be dependent on the 2´ OH group of the ribose which locks the glycosidic bond in the anti-conformation and therefore leading to a parallel-stranded G-quadruplex formation [45].

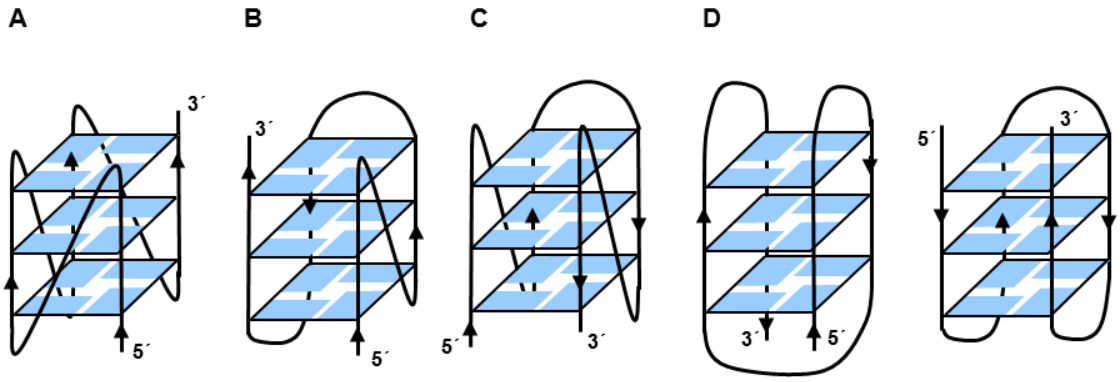


Fig. 1.4: Topologies of G-quadruplexes. A-D: Examples of different intramolecular G-quadruplex topologies. A: Parallel G-quadruplex (propeller type). B: Mixed parallel/antiparallel (3+1) hybrid G-quadruplex. C and D: Antiparallel G-quadruplexes (D: basket and chair type). (Figures modified from [46])

1.3.2.2 Glycosidic bonds

The nucleobase within a nucleotide is connected via a glycosidic bond to the 1´ carbon of the ribose and the nucleobase within a G-quadruplex can have either a syn- or an anti-conformation with regard to the pentose (Fig. 1.5 A). While in the right-handed DNA helix all nucleobases are in the anti-conformation, the different G-quadruplex topologies are based on different conformations of the glycosidic bonds. Whereas within a parallel G-quadruplex the glycosidic angles

are all in the anti-conformation [47-49] or occasionally all in the syn-conformation [50] (Fig. 1.5 B), the glycodsidic angles in a mixed parallel/antiparallel hybrid G-quadruplex are in syn-anti-anti-anti-conformation or in anti-syn-syn-syn-conformation [51] (Fig. 1.5 C). Within a G-quadruplex with two neighboring strands with the same direction and the two remaining strands having the opposite direction or within a G-quadruplex with each strand having antiparallel adjacent neighbors the glycosidic angles are in syn-syn-anti-anti-conformation or in syn-anti-syn-anti-conformation, respectively [52-56] (Fig. 1.5 D and E).

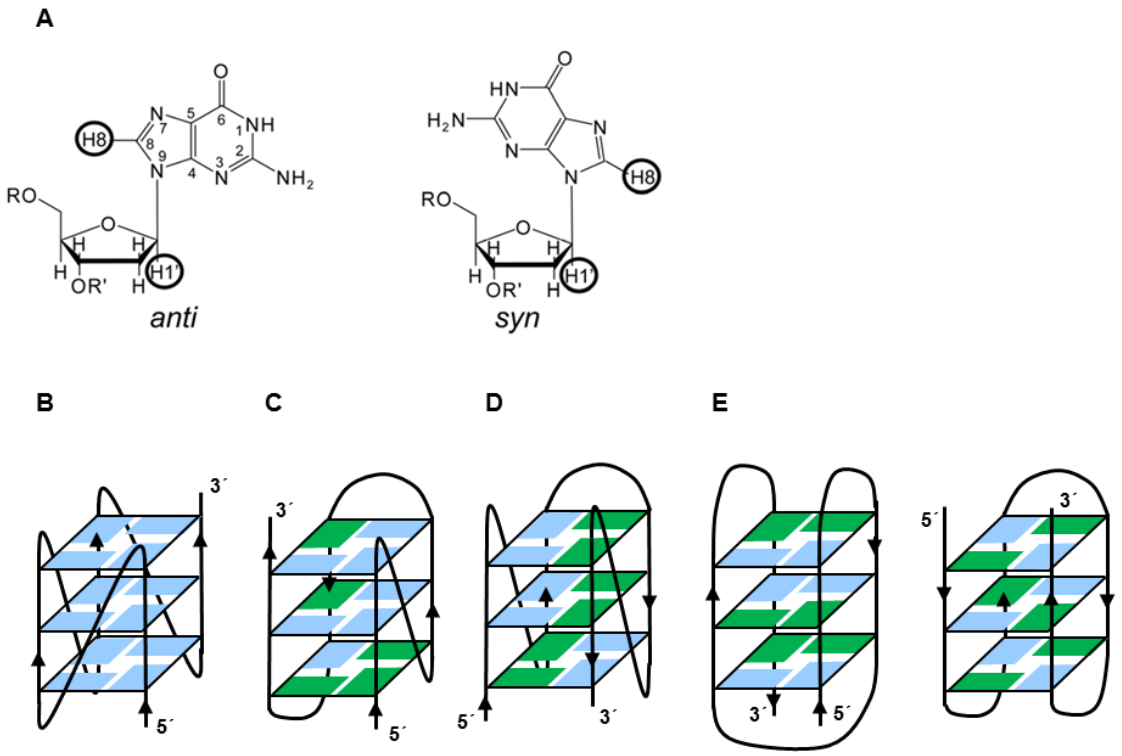


Fig. 1.5: Conformation of glycosidic bonds in different G-quadruplex topologies. A: Anti- or syn-conformation of guanine within a nucleotide (Figure from [57]). B-E: Examples of different intramolecular G-quadruplex topologies with their corresponding glycosidic angles. Glycosidic angles in syn-conformation are depicted in green and in anti-conformation in blue. B: Parallel G-quadruplex. C: Mixed parallel/antiparallel (3+1) hybrid G-quadruplex. D and E: Antiparallel G-quadruplexes (E: basket and chair type). (Figures modified from [46])

1.3.2.3 Loop sequences

The loop sequences of a G-quadruplex are usually short (between 1 to 7 nucleotides) and shorter loop sequences usually lead to a more stable G-quadruplex [41-43]. Furthermore, the length as well as the composition of the loop sequence may influence the G-quadruplex topology as indicated by *in vitro* experiments [42,58-60]. Three major types of loops can occur in G-quadruplexes formed by a single strand or two or three strands (Fig. 1.6): (1) a diagonal loop which connects two opposing antiparallel strands across the diagonal, (2) a lateral loop which connects two adjacent antiparallel strands and (3) a strand-reversal loop, also called propeller loop, which connects two adjacent parallel strands.

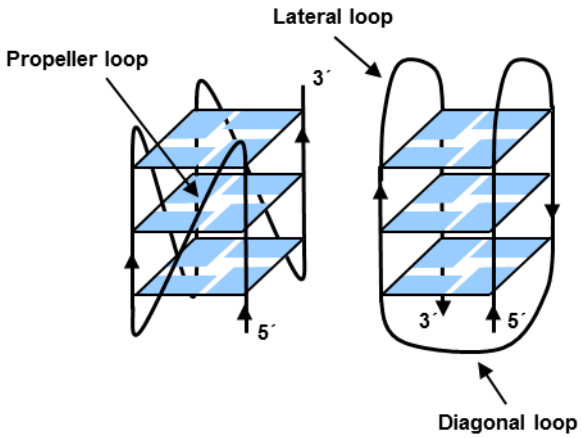


Fig. 1.6: Loop types occurring in G-quadruplexes. The three major types of loops occurring in G-quadruplexes are named propeller loop, diagonal loop and lateral loop. A propeller loop connects two adjacent parallel strands, a diagonal loop connects two opposing antiparallel strands across the diagonal and a lateral loop connects two adjacent antiparallel strands.

1.3.2.4 Grooves

The four grooves of a G-quadruplex, which are defined as the cavities restricted by the phosphodiester backbones, are dependent on the loops, but also on the topology of the G-quadruplex. Within a parallel G-quadruplex all strands have the same direction and therefore all nucleobases are in the anti-glycosidic-conformation leading to a symmetric arrangement with four grooves of identical width [61]. However, within an antiparallel G-quadruplex the nucleobases have

different glycosidic angles which affect the orientation of the backbone relative to the G-tetrad, thereby leading to grooves of different sizes [62].

1.3.3 Stability of G-quadruplexes

Comparisons between the stability of DNA and RNA G-quadruplexes revealed that RNA G-quadruplexes are thermodynamically more stable than their DNA counterparts [45,63]. It is assumed that this higher stability of RNA G-quadruplexes can be addressed to the 2´ OH group of the ribose which positively influences the number of water molecules which are bound to the G-quadruplex. Furthermore, it has been shown that the additional hydroxyl group of the ribose also leads to a more ordered water structure at the grooves of the RNA G-quadruplex [64].

Beside these differences of RNA and DNA G-quadruplexes in thermal stability, both are influenced by the ionic environment, the length and the composition of the loops, the number of G-tracts as well as by flanking nucleotides at the 5´ or 3´ end of a G-quadruplex forming sequence.

1.3.3.1 Influence of cations on G-quadruplexes

In addition to the Hoogsteen hydrogen bonds and π-π stacking interactions the stability of a G-quadruplex highly depends on the presence of monovalent or divalent cations. Within a G-tetrad the O6 atoms of the guanines are orientated towards the center, thereby creating a channel which is caused by their lone pairs of electrons. This strong negative electrostatic potential can be neutralized by cations of the correct size like the monovalent cations K^+ or Na^+. The exact location of a cation depends on the nature of the cation, but also on the structure of the G-quadruplex. In some structures the smaller Na^+ ion sits in the plane of the tetrad, while in other structures it is located between two G-tetrads [65]. The K^+ ion is always located between two G-tetrads forming the eight oxygen atoms in a symmetric tetragonal bipyramidal configuration [40]. In general, G-quadruplexes prefer K^+ over Na^+ based on the preferential hydration of Na^+ in comparison with K^+ [66]. In addition to K^+ and Na^+ also other cations are able to stabilize the G-quadruplex structure. The preference of DNA G-quadruplexes

for monovalent cations is K^+>>Na^+>Rb^+>Cs^+>>Li^+ and for divalent cations Sr^{2+}>>Ba^{2+}>Ca^{2+}>Mg^{2+}>Mn^{2+}>Co^{2+}>Zn^{2+} [67]. Interestingly, the G-quadruplex topology can be dependent on the cation as observed by NMR spectroscopy for the human telomeric sequence which shows different topologies in the presence of Na^+ or K^+, respectively (cf. chapter 1.3.4.1).

With regard to K^+ being the monovalent cation with the highest intracellular concentration (~150 mM), most studies on G-quadruplex topology were performed in the presence of K^+ [68]. It is worth to mention that under near physiological conditions G-quadruplexes can be thermodynamically very stable with melting temperatures above 90°C, indicating for their possible *in vivo* occurrence [67].

1.3.3.2 Influence of sequence, loop length and number of G-tracts on G-quadruplexes

Other important factors for the stability of G-quadruplexes are the length as well as the composition of their loop sequences. In general, the loop length is inversely related with DNA as well as RNA G-quadruplex stability [69-72]. Although the data are still sparse, also the composition of the loop sequences seems to influence G-quadruplex stability. Pandey *et al.* (2013) observed that cytosines within the loops of RNA G-quadruplexes with only two G-tetrads negatively influence their thermal stability [70]. Another publication from Mergny *et al.* (2008) indicates that in terms of single nucleotide loops within DNA G-quadruplexes pyrimidines lead to higher melting temperatures compared to adenines [73]. Interestingly, Halder *et al.* (2009) found that also the number of G-tracts seems to influence G-quadruplex stability. They observed higher stabilities for RNA G-quadruplexes formed within sequences with five or six G-tracts compared to a G-quadruplex formed by a sequence of only four G-tracts, although the corresponding G-quadruplexes were very likely composed of only four G-tracts [69].

1.3.3.3 Influence of flanking nucleotides on G-quadruplexes

While most of the research on G-quadruplexes only investigated the G-quadruplex forming sequence, also flanking nucleotides are able to influence the G-quadruplex stability. To date only a few publications addressing the influence of flanking nucleotides on G-quadruplexes are available. Yang and co-workers investigated the impact of short flanking sequences (up to 3 nt) at the 5´ and 3´ end on the G-quadruplex formed by the nuclease hypersensitive element III_1 (NHE III_1) originally located within the promoter region of *c-MYC*, thereby finding a positive influence of these nucleotides on G-quadruplex stability [74,75]. This increase of thermal stability seems to be accomplished by the formation of a capping structure by the flanking nucleotides as indicated by NMR spectroscopy experiments in potassium solution [74]. Experiments investigating the human telomeric sequence revealed an overall destabilizing effect for short flanking nucleotide sequences (1-3 nt) either at the 5´ or 3´ end [76,77]. However, molecular dynamic studies indicated for stacking interactions of the flanking nucleotide with the closing G-tetrad as well as the formation of additional hydrogen bonds. The authors argued that these local positive interactions may destabilize the overall structure by reducing the overall number of hydrogen bonds of the core G-tetrad [76]. Arora and co-workers were the first who published their experience on the influence of longer flanking sequences (4-12 nt) on the formation of the *c-KIT* DNA G-quadruplex in the presence of the complementary strand [78]. They found that with an increase of flanking bases at the 5´ and 3´ end of the G-quadruplex forming sequence this structural motif is driven towards normal duplex formation. Valuable information about the influence of flanking nucleotides on RNA G-quadruplex formation was gained by the group of Perreault who examined naturally occurring putative 5´ and 3´ UTR RNA G-quadruplex forming sequences in the context of their natural flanking sequences [79,80]. They observed that a high content of cytosines within the flanking sequences prevent G-quadruplex formation. They attributed this finding to the potential of cytosine tracts to form stable stem secondary structures with the guanines from the G-tracts. Furthermore, they developed a scoring system for the prediction of RNA G-quadruplex formation which included their findings of the effects of neighboring sequences [80]. While these are the only publications

dealing with the influence of flanking nucleotides on DNA and RNA G-quadruplex formation, some research has been done investigating the effect of dangling ends on duplex DNA, such as hairpins [81], and double-stranded RNA like small interfering RNA (siRNA) [82]. *In vitro* experiments showed a stabilizing effect of dangling ends on duplex formation for DNA as well as for RNA by the so-called dangling effect [83-85]. The strength of this stabilizing effect was dependent on the closing base pair as well as the following dangling end and correlated with the length of this overhang. It is suggested that the dangling nucleobase stacks together with the neighboring closing base pair and thereby acts as a hydrophobic cap at the end of the duplex. This shielding of the hydrogen bonds of the closing base pair leads to an increase of the thermodynamic stability of the DNA or RNA duplex structure [86]. Conversely, DNA duplex formation is more or equally stabilized by a 5´ dangling end, while RNA duplex formation is more or equally stabilized by 3´ dangling end [86].

1.3.4 Biological significance of G-quadruplexes

Computational analysis revealed more than 375.000 putative G-quadruplex forming sequences within the human genome by using the following algorithm [87,88]:

$$G_{3-5}\,N_{L1}\,G_{3-5}\,N_{L2}\,G_{3-5}\,N_{L3}\,G_{3-5}$$

G_{3-5} stands for a G-tract of 3 to 5 guanines and N_{L1-3} are the loop sequences with the limits $1 \geq N_{L1-3} \leq 7$ nt. This algorithm is called *quadparser* and is freely available [88]. Although G-quadruplexes with only two G-tetrads are possible, they were excluded from the search due to their relative instability. Furthermore, longer loop sequences (>7 nt) are possible, but were also excluded from the search for alleviating the search as well as due to their negative influence on G-quadruplex stability. It should be mentioned that computational analysis were performed not only within the human genome but also within other genomes, e.g. *S. cerevisiae* [89,90] or *E. coli* [91].

Interestingly, the putative G-quadruplex forming sequences are not equally distributed among the genome. Furthermore, the positions as well as the sequences are conserved among human populations and related yeast species.

They were found to be enriched at functional regions, such as the telomeric sequence [92], rDNA [93] and within promoter regions of various genes, especially of proto-oncogenes, e.g. *c-MYC* [94], *c-MYB* [95], *KRAS* [96], *BCL-2* [97], *VEGF* [98] and *c-KIT* [99]. They were also identified in RNAs, notably within the 5´ and 3´ untranslated regions (UTR) of mRNAs which are known to be crucial for the regulation of translation. Moreover, also the telomeric sequence was found to be transcribed into a putative G-quadruplex forming sequence, so-called telomeric repeat containing RNA (TERRA) [100]. In addition to the computational analysis there is strong evidence based on *in vitro* and *in vivo* experiments for a biological relevance of such putative G-quadruplex forming sequences. G-quadruplex forming sequences have been investigated by various techniques, such as NMR spectroscopy, X-ray crystallography, in-line probing and CD spectroscopy for structural analysis [79,80,101,102] or via reporter gene assays [103] and antibody detections [104] for direct or indirect *in vivo* evidence.

Depending on their different localizations several functions of DNA as well as RNA G-quadruplexes are under debate. Among the proposed functions in telomere organization[105] and the regulation of gene transcription [94] and translation [106], G-quadruplex formation seems also to be involved in many other processes like DNA replication [107], recombination [108], meiosis [49], pre-mRNA splicing [109], pre-mRNA polyadenylation [110] and mRNA targeting [111].

In contrast to RNA, DNA is mostly double-stranded and therefore formation of a G-quadruplex competes with the complementary strand for the formation of a duplex structure. The C-rich complementary sequence is able to form an alternative structure with intercalated hemiprotonated cytosine^{+}-cytosine base pairs (i-motif) [13]. Although G-quadruplexes are mostly of equal or even less stability compared to the competing duplex structure [112,113], it is assumed that G-quadruplexes can fold under certain circumstances. During the process of DNA transcription the two DNA strands become temporarily separated from each other and thereby G-quadruplex formation might take place. It is also suggested that G-quadruplex formation is actively induced with the participation of proteins. Furthermore, G-quadruplex formation might occur at locations at which the DNA is not double-stranded, e.g. the 3´ overhang of telomeres.

The following subchapters will focus on G-quadruplex forming sequences located in the telomeric sequence, within promoter regions as well as in mRNAs. Afterwards, the focus will lie on the state of the art of G-quadruplex-interacting proteins as well as G-quadruplex-interacting molecules. Finally, some examples for G-quadruplex detection via antibodies will be discussed. However, due to the huge amount of publications addressing G-quadruplex forming sequences this overview does not assert to claim on completeness.

1.3.4.1 G-quadruplexes in telomeres and telomeric RNA

Telomeres are specialized chromatin structures at the ends of all eukaryotic chromosomes. They were first described as essential structures by Hermann Müller in the 1930s [114]. In all vertebrates, telomeres are composed of double-stranded tandem repeats (2-10 kb) with the sequence 5´ TTAGGG 3´, ending in a short single-stranded G-rich 3´ overhang which represents a putative G-quadruplex forming sequence [101]. Other species have very similar sequences with G-runs and short intervening nucleotides, e.g. the protozoa *Oxytricha nova* (5´ TTTTGGGG 3´) [115]. Telomeres are associated with specialized proteins such as proteins from the shelterin complex or other chromatin or chromosome regulating factors [116]. Their main function is to protect the chromosome ends from being recognized as DNA breaks. Telomeres are also involved in the regulation of the life-span of the cell. During replication the telomeres become shorter (approximately 50-100 nucleotides per cell division) due to DNA end replication problems and nucleolytic processing [116]. Reaching of a critical point in telomer-shortening leads to cell-cycle arrest, senescence or apoptosis [116]. Due to their fundamental relevance in genomic stability and cell division telomeres play an important role in aging, but also in cancer development. The shortening of telomeres is counteracted by an enzyme called telomerase (also known as telomerase reverse transcriptase (Tert)). This enzyme is a ribonucleoprotein enzyme which is able to recognize and extend the 3´ OH end of the telomeres by using an associated RNA (also known as telomerase RNA component (Terc)) with the sequence 5´ AAUCCC 3´ as a template [117]. While stem and germ line cells express the telomerase to prevent cell death or loss of genetic functions, somatic cells lack sufficient amounts of telomerase and

therefore telomere shortening is an inevitable process in correlation with the number of cell divisions. Interestingly, 80-85% of all tumor cells express the telomerase and therefore their telomeres are not shorten in length after replication [118].

Due to the important role of the telomeres, the structure of the telomeric sequence was intensively studied and as expected its potential to fold into a G-quadruplex structure has been observed. However, the results for the exact G-quadruplex topology are controversial and are highly dependent on the experimental conditions. The first analysis of the telomeric sequence was performed by Wang *et al.* in 1993. They investigated the sequence $d[AG_3(T_2AG_3)_3]$ in a Na^+-containing solution by NMR spectroscopy. As a result they found an intramolecular anti-parallel G-quadruplex composed of three stacked tetrads connected via two lateral loops and one central diagonal loop [53]. In 2002 Parkinson *et al*. examined the same sequence, but in the presence of K^+ via X-ray crystallography. They identified a parallel G-quadruplex in which the G-repeats are connected via propeller loops [119]. In 2006 it was reported that the telomeric sequence forms a (3+1) hybrid-like G-quadruplex in the presence of K^+ using CD spectroscopy and NMR spectroscopy [120-122]. To analyze the telomeric structure under more cellular-like conditions, molecular crowding agents like polyethylene glycol (PEG) were used. For example, Xue *et al.* (2007) observed a parallel G-quadruplex in the presence of K^+ under simulated molecular crowding conditions using 40 % (w/v) PEG 200 [123]. In a recent publication Sharma *et al.* (2013) investigated the telomeric sequence by CD spectroscopy under different solution conditions in the presence or absence of PEG or acetonitrile. Thereby, a conformational change with increasing concentrations of PEG or acetonitrile from an unimolecular hybrid structure into a multimolecular parallel stranded structure occurred [124]. Other studies on the telomeric G-quadruplex structure were performed using electron paramagnetic resonance (EPR) techniques. As a result for analyzing the sequence $A(GGGTTA)_3GGG$ in K^+ solution, a mixture of the parallel propeller type and the antiparallel basket type G-quadruplex in a 1:1 ratio was observed [125]. With regard to the telomeric sequence consisting of tandem repeats of the sequence TTAGGG at the single-stranded 3´ overhang, also studies using multimers of

this sequence were carried out. These studies led to two different models: A “bead on a string” model, in which the G-quadruplexes are composed of four G-runs and connected by a TTA linker [126-128], and a “stacking” model, in which G-quadruplexes stack on one another, thereby creating a higher order structure [129,130]. Yu *et al.* (2006) obtained spectra of a mixture of parallel and antiparallel G-quadruplexes in the presence of K^+ for the sequences $(TTAGGG)_{4-12}$ by CD spectroscopy. Furthermore, they found a gradual increase in enthalpy and entropy for sequence repeats which are a multiple of four, indicating for the “beat on a string” arrangement [126]. To gain a more nucleus-like environment for structure analysis, Hartig and co-workers investigated a spin-labeled telomeric sequence $d[AGGG(TTAGGG)_3]$ in mature oocytes of *Xenopus laevis* by an elegant EPR technique. By doing so, they observed G-quadruplex formation in a 1:1 ratio of the parallel propeller type and the antiparallel basket type [131].

While a lot of research on elucidating the structure of the telomeric G-quadruplex forming sequence has been done, its possible function is still under debate. It is speculated that it might be involved in the capping function of the telomeres [132]. This theory is mainly supported by the identification and analysis of interactions between known telomere interacting proteins (e.g. of the shelterin complex) and G-quadruplexes formed by the telomeric sequence (cf. chapter 1.3.4.4) as well as by the influence of G-quadruplex-interacting molecules on the telomerase as well as on telomeric maintenance (cf. chapter 1.3.4.5). The formation of G-quadruplexes within the telomeric sequence became of particular interest by the finding that the formation of stable G-quadruplexes led to an inhibition of the telomerase [133]. Since the telomerase is overexpressed in most of all known cancer cells and this is thought to be a hallmark in cancer development, the inhibition of the telomerase might represent an interesting approach for anticancer therapy. Therefore, a lot of research has been done to find naturally occurring molecules or to develop drugs which specifically bind to and stabilize the telomeric G-quadruplexes [118] (cf. chapter 1.3.4.5).

The clarification of the role of the G-quadruplex forming sequence within the telomeres became even more complex, when another player showed up which

is suggested to be involved in telomere regulation. Recent findings revealed that the telomeric sequence, which for a long time was thought to be transcriptionally silent, is transcribed into the telomeric repeat containing RNA (TERRA) [100]. This putative G-quadruplex forming sequence is thought to play a crucial role in the regulation as well as protection of the chromosome ends [134,135]. Komiyama and co-workers investigated the structure and the location of this RNA in living cells by using a light-switching pyrene probe. They found that TERRA folds into a parallel G-quadruplex which is co-located at the telomeres, indicating for a role in telomere function [136]. This assumption is further supported by the detection of interactions between proteins from the shelterin complex and TERRA [137]. Furthermore, it has been shown that TERRA is able to hybridize with the telomeric sequence, thereby forming a so-called R-loop RNA:DNA hybrid structure [138]. This hybrid structure seems to be involved in the process of homologous recombination at the telomeres and therefore in telomere length homeostasis [139].

1.3.4.2 G-quadruplexes and the regulation of transcription

Among other regions, G-quadruplex forming sequences were found to be enriched within promoter regions of various genes, especially of proto-oncogenes. One or more G-quadruplex forming sequences were found within 1000 nt upstream of the transcription start site of 50% of all human genes [16]. This enrichment of G-quadruplex forming sequences in promoter regions leads to the suggestion of a possible function of G-quadruplexes in the regulation of transcription. Furthermore, formation of a G-quadruplex structure seems to be supported by the negative-supercoiling induced stress during transcription [140]. Intensive research has been done to investigate the possibility of these structures to fold. Experiments in cell culture systems, mostly using reporter gene constructs, which contain the corresponding sequence, were able to further indicate for the *in vivo* relevance of G-quadruplex formation [103].

Based on the obtained results, different mechanisms how G-quadruplex formation may influence gene transcription are under debate: The G-quadruplex forming sequences can be located either within the template or within the non-template strand of DNA. Although the formation of G-quadruplex structures in

putative duplex DNA is thermodynamically unfavored [141,142], during the process of replication the two strands become temporarily separated and therefore G-quadruplex formation may occur. Depending on the location of a G-quadruplex forming sequence either within the template strand or within the non-template strand, formation of a G-quadruplex structure might lead to inhibition or enhancement of DNA transcription. If positioned within the template strand, formation of a G-quadruplex might act as a steric hindrance for the transcription machinery, thereby blocking the transcriptional process. This scenario is proposed for a number of G-quadruplex forming sequences within promoter regions, e.g. G-quadruplex forming sequences within the promoter regions of *c-MYC* [94] and *VEGF* [143]. If localized within the non-template strand, formation of a G-quadruplex might be able to stabilize the single-stranded conformation, thereby supporting the transcriptional process. Another way how G-quadruplexes may influence gene transcription is to act as binding sites for either transcriptional enhancer, as indicated for the insulin-linked polymorphic region (ILPR) within the insulin gene promoter [144], or for repressors, as shown for nucleolin in terms of the *c-MYC* promoter [145]. It is worth mentioning that although the formation and resolving of G-quadruplexes seem to be used as a regulatory mechanism in gene transcription within cells, it might also bear the risk of genomic instability. Genome wide analysis of DNA breakpoints in cancer cell lines revealed an enrichment of such breakpoints at G-quadruplex forming loci [146]. It is assumed that this correlation between G-quadruplex forming sequences and genomic instability is associated with its proposed function as a steric hindrance for the transcription machinery. Since the G-quadruplex structure can be very stable under physiological conditions, this hindrance might lead to a stall or even a collapse of the replication fork which in turn may lead to DNA damage [147]. To that end, the formation as well as the resolving of G-quadruplex structures is supposed to be tightly regulated within cells (cf. chapter 1.3.4.4).

1.3.4.3 G-quadruplexes and the regulation of translation

The observation of putative G-quadruplex forming sequences within RNAs, especially within the 5´ UTR, but also within the open reading frame (ORF) and the 3´ UTR of mRNAs, led to the assumption for their influence on translation.

Most of the known G-quadruplex forming sequences are thought to have an inhibitory influence on translation. If a G-quadruplex is formed close to the cap structure of the mRNA (e.g. of *NRAS* [106]), assembly of the ribosome complex might be blocked. If the G-quadruplex is located in the center or close to the translation start site of the mRNA (e.g. of *BCL-2* [148]), G-quadruplex formation might act as a steric hindrance for the translation machinery. The G-quadruplex forming sequence within the 3´ UTR of *PIM1* is an example for a G-quadruplex which seems to reduce the level of translation [149]. Also some examples of G-quadruplex forming sequences are found within the open reading frame of mRNAs, e.g. of *APP1* [150] or *FMR1* [151]. However, to date it is unclear if these G-quadruplexes may have an inhibitory effect on the level of translation. Also some more systematically investigations on artificial G-quadruplex forming sequences within the mRNA were performed to elucidate the influence of G-quadruplex formation on translation *in vivo*. Hartig and co-workers analyzed the influence of artificial G-quadruplex forming sequences within the 5´ UTR of a reporter gene on translation in HEK 293 cells. The G-quadruplex forming sequences differed in the length of the loop sequences, in the number of G-tracts and in the position within the 5´ UTR of the mRNA. They observed predictable correlations between those parameters. While an increase of the loop length was associated with a decrease in G-quadruplex-mediated repression of translation, an increase in the number of G-tracts led to an increase in G-quadruplex-mediated repression of translation. The position-dependent experiment revealed the strongest inhibitory effect of the G-quadruplex on translation if it was located close to the translation start site [69]. On the contrary, Kumari and co-workers found in a similar cell culture experiment the strongest inhibitory effect for the G-quadruplex forming sequence of the *NRAS* mRNA on translation at a location close to the transcription start site [152]. This indicates that a location-dependent influence of

G-quadruplex forming sequences within a 5´ UTR of a mRNA seems to include yet unknown factors.

Beside these examples of inhibitory G-quadruplex forming sequences within mRNAs, there are also a few G-quadruplex forming sequences known to activate translation. The G-quadruplex sequences within the 5´ UTR of *FGF2* [153] and *VEGF* [154] were found to be located in internal ribosome entry sites (IRES) in which they seem to be important for initiation of translation.

1.3.4.4 G-quadruplex-interacting proteins

Since G-quadruplex forming sequences seem to play a crucial role in many cellular processes, it is supposed that formation as well as resolving of G-quadruplexes are tightly regulated within cells. Indeed, some proteins with the potential to interact with G-quadruplexes have been found. These specific recognitions of G-quadruplexes by proteins further indicate for the existence of G-quadruplex formation *in vivo*. Most of these proteins were shown to interact with DNA G-quadruplexes, however also a few proteins with the potential to interact with RNA G-quadruplexes were identified [155].

The shelterin complex, a complex composed of six proteins, is known to interact with the telomeric sequence [116]. One protein of this complex, the protection of telomeres protein 1 (POT1), has been shown by *in vitro* experiments to have G-quadruplex unwinding activity. This unwinding of G-quadruplexes by POT1 is assumed to be important for telomerase-dependent elongation of the 3´ single-stranded overhang. Therefore, it is suggested that resolving of the G-quadruplex structure by POT1 is important for proper telomerase-dependent elongation of the telomeres. Interestingly, this unwinding activity is probably topology-dependent, as POT1 shows stronger unwinding activity for antiparallel over parallel G-quadruplex structures [156]. Another function of POT1 seems to be facilitated when acting as a heterodimer together with the POT1-interacting protein1 (TPP1). POT1-TPP1 is able to bind to and protect G-quadruplexes within the single-stranded 3´ overhang from being recognized as a DNA damage site by replication protein A (RPA). RPA itself has also been shown to possess G-quadruplex unwinding activity. However in contrast to POT1, RPA

might be able to unfold both the antiparallel as well as the parallel G-quadruplex structure [157]. Another member of the shelterin complex is the telomere repeat binding factor 2 (TRF2). Balasubramanian and co-workers demonstrated that this protein can bind to G-quadruplex structures within the telomeric sequence, but also within the telomere repeat-containing RNA (TERRA), indicating for a possible role of TRF2 in telomere organization [158]. Two other proteins, which are thought to be involved in telomerase-dependent telomere elongation, are the heterogeneous ribonucleoprotein particles A1 (hnRNP A1) and A2 (hnRNP A2). Both, the unwinding protein 1 (UP1), which is a proteolytic product of hnRNP A1, and hnRNP A2*, which is a splice variant of hnRNP A2, have been shown to bind to and destabilize telomeric G-quadruplexes *in vitro* [159,160].

In addition to proteins interacting with telomeric G-quadruplexes, also some proteins interacting with G-quadruplexes within promoter regions were found. For example, the protein poly(ADP-ribose) polymerase-1 (PARP1) has been shown to interact with G-quadruplexes within promoter regions of various genes (e.g. *BCL-2*, *MYB*, *KIT*, *VEGF* and *KRAS*) [161]. Another example for a protein, which was identified to interact with G-quadruplexes within promoter regions, is the multifunctional protein nucleolin. The heterodimer nucleolin-hnRNP D was reported to bind to G-quadruplex structures *in vitro* and the overexpression of nucleolin in MCF10A cells caused an inhibition of *c-MYC* promoter-driven transcription [162,163].

While a growing number of proteins interacting with DNA G-quadruplexes has been identified, the number of proteins interacting with RNA G-quadruplexes is much less. The two members of the human DExH/D-box family of RNA helicases DHX36 and DHX9 were shown to have RNA G-quadruplex binding activities. While DHX9 was reported to unwind RNA G-quadruplex structures *in vitro* [164], DHX36 exhibited RNA G-quadruplex unwinding activities in HeLa cell lysates [165]. In a very recent publication some possible RNA G-quadruplex-interacting proteins have been identified via a pull-down assay. This pull-down experiment was performed using two naturally occurring G-quadruplex forming sequences originally located within the 5´ UTRs of *MMP16* and *ARPC2* mRNAs. Thereby, mainly proteins of the large as well as of the small ribosomal subunit and proteins belonging to the family of the heterogeneous nuclear

ribonucleoproteins (hnRNPs) were found [166]. Also some proteins interacting with TERRA have been described (cf. chapter 1.3.4.1). Most probably due to the structural similarities of DNA and RNA G-quadruplexes, some proteins were found to bind to both DNA and RNA G-quadruplexes. Among them are the already mentioned proteins DHX36 [165] and nucleolin [166], suggesting a role of such proteins in both DNA and RNA G-quadruplex regulation.

As mentioned at the end of chapter 1.3.4.2, the formation of a G-quadruplex structure bears the risk of genomic instability, if the cell loses the control of formation and resolving of this structure. Indeed, some helicases, which are associated with human diseases marked by genomic instability, showed G-quadruplex unwinding activities *in vitro*. These helicases include the RecQ helicases WRN [167], BLM [168], FANCJ [169] and Pif1 [170]. Altered or loss of function of these proteins is associated with an increased risk for cancer or premature aging processes. Interestingly, genomic instability was found to be increased at putative G-quadruplex forming loci, indicating for a role of G-quadruplexes in genome stability [147,171-173].

1.3.4.5 G-quadruplex-interacting small molecules

Due to their important role in the regulation of gene expression, G-quadruplexes became interesting as therapeutic targets. Especially the finding that the stabilization of the telomeric G-quadruplexes by a 2,6-diamidoantraquinone leads to an inhibition of the telomerase, which is expressed in most cancer cells, initialized the screening for G-quadruplex-interacting small molecules [133]. Since that time, a lot of research has been done to identify, but also to develop molecules with the potential to interact with G-quadruplexes and to elucidate the underlying mechanisms. This effort resulted in the discovery of naturally occurring G-quadruplex ligands like the alkaloid berberine [174] or the cyclic polyamine telomestatin [175], but also in the development of synthesized molecules. Although these molecules belong to a broad variety of different chemical classes like cationic porphyrins (e.g. TMPyP4 [176]), bisquinolinium derivatives (e.g. 360A [177]), naphthalene diimides [178], carbazole derivatives [179] and pentacyclic arcridines (e.g. RHPS4 [180]), they mostly share some structural

similarities, such as a flat aromatic surface and positively charged site chains [181].

In general, G-quadruplex ligands are suggested to act via the following mechanisms [182]: (1) via stacking interactions with the terminal G-tetrads: Most of the known G-quadruplex-interacting molecules exhibit planar aromatic ring systems to allow for π-π stacking interactions with the terminal G-tetrads. This mechanism includes electrostatic, hydrophobic and van der Waals interactions. (2) via recognition and binding to the grooves or loop sequences: The size and shape of the grooves of duplex DNA and G-quadruplexes are significantly different, therefore offering a suitable binding site for G-quadruplex-interacting molecules. Also the loop sequences between the G-tracts can show variance with regard to their composition as well as their length and therefore might serve as recognition sites for possible G-quadruplex ligands. G-quadruplex ligands with such features offer the possibility to distinguish not only between G-quadruplexes and B-DNA, but they might also be able to discriminate between different G-quadruplex topologies. Therefore, such recognition modes might be helpful for the development of highly specific therapeutics. (3) via intercalating between the G-tetrads: The mode of molecules that intercalate between base pairs of double-stranded DNA, like ethidium bromide or daunomycin, has been studied intensively. In general, also G-quadruplex-interacting molecules could act by this mechanism.

Since the finding of Sun and co-workers [133], a high number of studies investigating various G-quadruplex-interacting molecules has been performed and the understanding of the structural requirements of such molecules has been increased. Based on high throughput screenings, molecular modeling and biophysical methods, but also by virtual screenings of databases, a large number of promising G-quadruplex-interacting molecules has been developed [181]. Many of these molecules were tested for their antitumor activity in different tumor models, thereby showing their potential for cancer treatment [182]. However, so far only a tiny fraction of these molecules, e.g. RHPS4, has currently reached clinical trials, mostly in combination with other drugs like Topo I or PARP 1 inhibitors [181,183].

1.3.4.6 G-quadruplex visualization in vivo

Based on the huge amount of publications dealing with conditions for G-quadruplex formation or indirect *in vivo* investigations, the evidence of these structures to occur *in vivo* is quite high. This evidence is further strongly supported by the detection of G-quadruplex structures *in vivo* via antibodies or chemical labelling.

The first detection of G-quadruplex structures by an antibody *in vivo* was performed by using a high-affinity single-chain antibody for the visualization of G-quadruplexes formed by the telomeric sequence of *Stylonychia lemnae* [184]. Using the same antibody it could be observed that the telomere end-binding proteins α and β (TEBPα and TEBPβ) are involved in the control of telomeric G-quadruplex formation in *Stylonychia lemnae* [185]. However, very recently also G-quadruplex structures have been identified via antibodies in mammalian cells. Balasubramanian and co-workers were able to detect G-quadruplexes in genomic DNA by using a monoclonal single-chain antibody (BG4) against intramolecular G-quadruplex structures. The investigation of different human cell lines at different time points of the cell cycle revealed G-quadruplex formation during all phases, whereas the formation of G-quadruplexes seemed to be cell cycle-dependent with a maximum during the S phase. Furthermore, they observed an increase in G-quadruplex formation by the addition of a G-quadruplex stabilizing molecule [104]. In another publication the same group was able to also visualize RNA G-quadruplexes by antibody detection within the cytoplasm of human cell lines [186]. Moreover, also Lansdorp and co-workers developed a monoclonal antibody (1H6) against G-quadruplex structures and proofed its functionality in different mammalian cell lines [187]. Similar to the findings of Balasubramanian, they also detected an increase in G-quadruplex formation in the presence of a G-quadruplex stabilizing molecule. In addition to the visualization of G-quadruplex structures via antibody, also an elegant method using a binding-activated fluorescent ligand for G-quadruplex detection has been published by Laguerre and co-workers [188]. This ligand (NaphtoTASQ), which is based on a naphthalene core surrounded by four guanines, was used to visualize RNA G-quadruplex structures in MCF7, U2OS and B16F10 cells by

multiphoton microscopy. Unlike to the protocols for immunodetections, they were able to visualize G-quadruplex structures without fixation or permeabilization of the cells with this technique.

2. Aim of the studies

Within the last decades, G-quadruplex forming sequences have been intensively studied by various techniques. Thereby, deeper insights into the conditions how G-quadruplexes might fold as well as into their biological role were gained. However, a lot of questions still remain elusive. To that end, this thesis should shed more light on the structure, characteristics and functions of G-quadruplex forming sequences.

In the first part of this thesis the position-dependent influence of a naturally occurring G-quadruplex forming sequence within the 5´ UTR of an mRNA on translation should be investigated in HeLa 229 cells. Although only scarce information was available regarding a position-dependent influence of G-quadruplex forming sequences within 5´ UTRs on translation, contrary results were observed. Therefore, the G-quadruplex forming sequence, which is naturally located within the 5´ UTR of *MAPK2*, should be inserted at different positions within the 5´ UTR of a reporter construct. The measurement of the level of reporter expression should indicate for the influence of the G-quadruplex forming sequence on the level of translation. An additional purpose of applying this cell-based reporter gene system was to identify proteins which are interacting with the RNA *MAPK2* G-quadruplex within the 5´ UTR. Candidates of possible RNA G-quadruplex-interacting proteins should be found after their knock-down via siRNAs, followed by a change in the level of reporter gene expression. Furthermore, using the same reporter system in HeLa 229 cells, the three known DNA G-quadruplex-interacting molecules, namely the bisquinolinium derivatives 360A, Phen-DC_6 and Phen-DC_3, should be examined for their potential to also bind to and stabilize RNA G-quadruplexes.

The aim of the second part of this thesis was to identify proteins which are interacting with the putative G-quadruplex forming sequence of the telomeric repeat containing RNA (TERRA). This objective should be achieved by using a cell-based crosslinking approach in which the addition of 4-thiouridine (s4U) to HeLa S3 cells should result in the randomized incorporation of this uridine analogue into RNA and therefore also into TERRA. Crosslinking of s4U containing RNAs to interacting proteins followed by purification of TERRA via a

TERRA-specific hybridization linker DNA should enable the isolation of proteins bound to TERRA. The following analysis of these proteins via nano-LC-MS/MS would then reveal new TERRA-interacting proteins.

In the third part of this thesis the question how the number of G-tracts of a DNA G-quadruplex forming sequence influences the thermal stability of the resulting G-quadruplex was addressed. Previous findings revealed an increase in thermal stability of an RNA G-quadruplex formed within sequences containing five or six G-tracts, even though only four G-tracts are involved in the formation of an intramolecular G-quadruplex. This unexpected finding raised the question if this stabilizing effect of additional G-tracts on RNA G-quadruplexes is transferrable to DNA G-quadruplexes. Therefore, the thermal stability of DNA G-quadruplexes formed by sequences with more than four G-tracts should be analyzed by CD spectroscopy and CD melting experiments. Furthermore, the influence of the position of a G-quadruplex formed within sequences containing more than four G-tracts should be investigated via CD spectroscopy, CD melting experiments and DMS footprinting experiments. To that end, the generation of specific point mutations within the G-tracts should force the G-quadruplex formation at a specific position. Since the formation of a G-quadruplex within a sequence with more than four G-tracts might lead to flanking nucleotides either at the 5´ or the 3´ end of the G-quadruplex, the question was asked if such flanking nucleotides might influence the G-quadruplex thermal stability. Thus, nucleotides should be systematically attached either at the 5´ or the 3´ end of a G-quadruplex forming sequence consisting of four G-tracts; the corresponding structures and melting temperatures should be analyzed by CD spectroscopy and CD melting experiments.

In the last part of this thesis small molecules, which interact with the G-quadruplex formed by the human telomeric sequence, should be identified within two different substance libraries by a FRET-based screening approach (developed by Dr. Armin Benz). Small molecules found within this screen should be further verified for their potential to interact with the G-quadruplex formed by the human telomeric sequence, but also for their specificity towards this structure over duplex DNA.

3. Results and Discussion

3.1 Screening for proteins and small molecules interacting with G-quadruplexes within the 5´ UTR of mRNA

Among other functional regions G-quadruplex forming sequences were found to be enriched within the 5´ UTR of mRNAs, indicating for a regulatory function during translation [189]. According to bioinformatical analysis of the human transcriptome one or more G-quadruplex forming sequences are present within about 3000 5´ UTRs of mRNAs [106]. Based on cell culture experiments there is strong evidence that RNA G-quadruplexes are indeed formed *in vivo* and thereby seem to mostly negatively regulate the level of translation, e.g. the G-quadruplex forming sequences within the 5´ UTRs of *Zic-1* [190], *MT3-MPP* [191], *TRF2* [192] and *BCL2* [148]. The influence of these non-canonical structures on translation is likely to be dependent on different parameters such as the position of the G-quadruplex forming sequence within the 5´ UTR of the mRNA [152], the number of G-tracts as well as the length and composition of the loop sequences [69,70]. Compared to DNA G-quadruplex-interacting proteins, the number of known proteins interacting with RNA G-quadruplexes is very low. For example, the fragile X mental retardation 1 protein (FMRP) was shown to bind to different RNA G-quadruplexes [193]. It is suggested that this protein is able to stabilize the G-quadruplex structures within the 5´ UTRs of *MAP1B* and *PP2A*, thereby perturbing translation initiation [194,195]. However due to the very high thermodynamic stabilities of RNA G-quadruplexes *in vitro*, it is assumed that proteins are not only involved in the stabilization, but also in the unwinding of these structures. The two members of the human DEAH-box family of RNA helicases DHX36 and more recently identified DHX9 were found to bind to and unwind RNA G-quadruplex structures *in vitro* [164,165]. Furthermore, DHX36 was characterized as the major helicase for DNA as well as RNA G-quadruplex unwinding in HeLa cell lysates [196]. However beside the identification of these few RNA G-quadruplex-interacting proteins, the situation how RNA G-quadruplexes are regulated within cells remains still unclear.

3.1.1 Establishing of an appropriate system to investigate RNA G-quadruplex-dependent translation

To investigate the influence of proteins as well as small molecules on G-quadruplexes within the 5´ UTR of mRNA, an appropriate system had to be established to examine RNA G-quadruplex-dependent translation. To that end, the plasmid pcDNA5/FRT/TO-eGFP containing already an *eGFP* reporter gene was used (plasmid provided by Prof. Dr. Thomas U. Mayer, University of Konstanz) and the naturally occurring G-quadruplex forming sequence of the *MAPK2* 5` UTR was inserted into the 5´ UTR of the reporter eGFP (Fig. 3.1.1). As a control, a sequence, which is unable to form a G-quadruplex, was generated by replacing the underlined guanines of the *MAPK2* G-quadruplex forming sequence with thymines (Fig. 3.1.1 A). As mentioned above, the influence of a G-quadruplex within a 5´ UTR on translation seems to be position-dependent. Therefore, the *MAPK2* G-quadruplex forming sequence (GQ) or the control sequence (Ctrl) was inserted at different positions into the 5´ UTR of the *eGFP* reporter mRNA (Fig. 3.1.1 B).

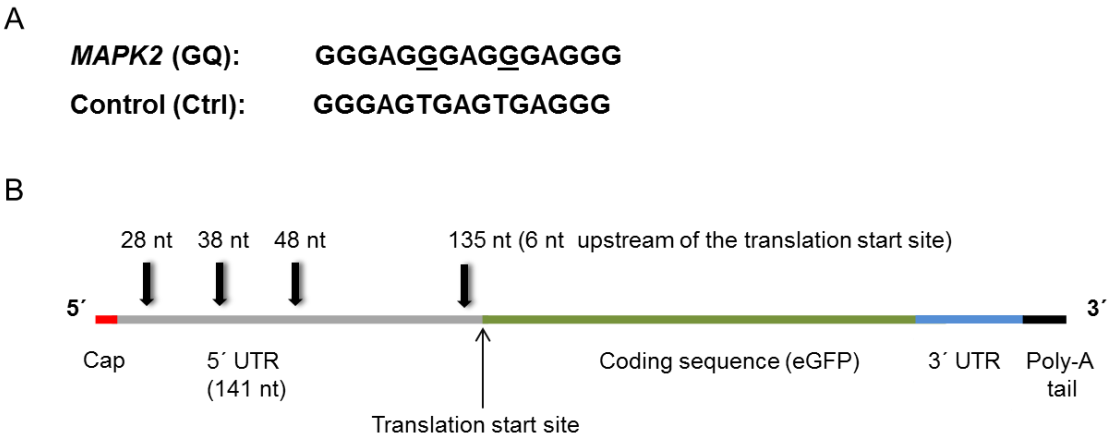


Fig. 3.1.1: Insertion of the *MAPK2* G-quadruplex forming sequence at different positions into the 5´ UTR of the *eGFP* mRNA. A: Sequences of the *MAPK2* RNA G-quadruplex forming sequence (GQ) and the corresponding control sequence (Ctrl). The underlined guanines of the *MAPK2* G-quadruplex forming sequence were replaced in the control sequence with thymines. These exchanges lead to a sequence which cannot fold into a G-quadruplex structure anymore. B: Schematic representation of the insertion of the *MAPK2* RNA G-quadruplex forming sequence or control sequence at positions 28 nt, 38 nt, 48 nt and 135 nt downstream of the 5´ UTR start of the *eGFP* mRNA.

Due to findings described by Kumari *et al*. (2008)[152], who screened for positions to insert a G-quadruplex forming sequence with the strongest inhibitory effect within the 5´ UTR of the mRNA encoding for NRAS, the *MAPK2* G-quadruplex forming sequence or the control sequence was inserted at the positions 28 nt, 38 nt and 48 nt downstream of the 5´ UTR start. In addition with regards to studies of Halder *et al.* (2009) [69], the *MAPK2* G-quadruplex forming sequence or the control sequence was inserted close to the translation start site (135 nt downstream of the 5´ UTR start) (Fig. 3.1.1 B).

To investigate the influence of the *MAPK2* RNA G-quadruplex at different positions within the 5´ UTR of the reporter mRNA on translation, HeLa 229 cells were transfected with the plasmid pcDNA5/FRT/TO-eGFP (wt) or pcDNA5/FRT/TO-eGFP containing the *MAPK2* G-quadruplex forming sequence or the corresponding control sequence within the 5´ UTR of the *eGFP* mRNA. To normalize the level of transfection, cells were always co-transfected with the plasmid pcDNA5/FRT/TO-mCherry. After 24 h upon reaching 80% of cell-confluence eGFP and mCherry expressions were measured by using the Tecan infinite M200 fluorescence-reader (Fig. 3.1.2).

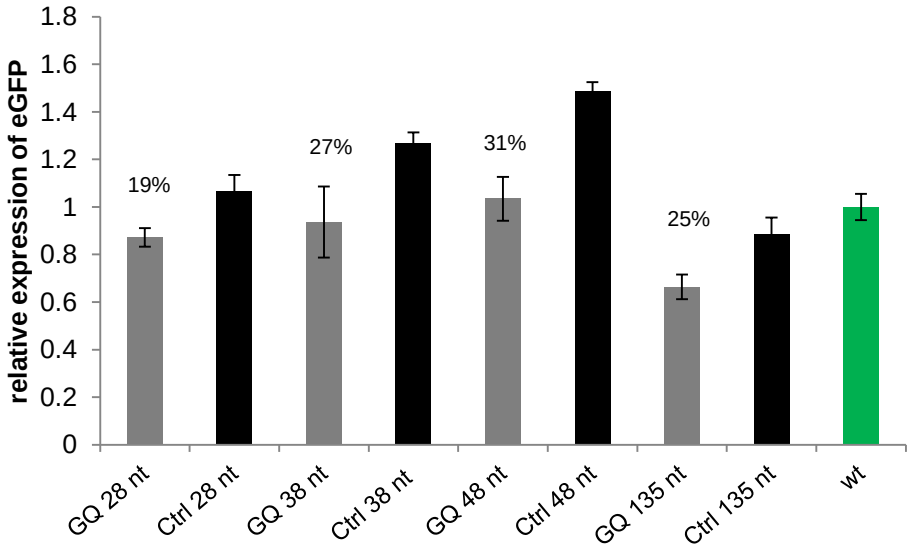


Fig. 3.1.2: Position-dependent influence of the *MAPK2* RNA G-quadruplex on eGFP expression. The influence of the *MAPK2* G-quadruplex forming sequence (GQ) inserted at different positions into the 5´ UTR of the reporter mRNA on expression was investigated by the detection of eGFP expression via a fluorescence reader. Level of expression was measured 24 h after co-transfection of HeLa 229 cells with 50 ng of pcDNA5/FRT/TO-eGFP (wt) or 50 ng of pcDNA5/FRT/TO-eGFP containing the *MAPK2* G-quadruplex forming sequence (GQ) or the

corresponding control sequence (Ctrl) within the 5´ UTR of the reporter mRNA and 50 ng of pcDNA5/FRT/TO-mCherry. Numbers on the top of the gray bars show the relative inhibition of reporter expression in comparison to the corresponding controls (black bars) in percentage. Numbers after the names of the different constructs mark the positions of the respective sequences with regard to the 5´ UTR start. The green bar shows the level of reporter expression of the wild type plasmid (wt) which was set to 1. The error bars represent the standard deviation of triplicates.

The observed data revealed a decrease in *eGFP* expression for all positions of the inserted *MAPK2* G-quadruplex forming sequences compared to the corresponding control sequences (Fig. 3.1.2). The *MAPK2* G-quadruplex forming sequence at the position 28 nt downstream of the 5´ UTR start caused a decrease in *eGFP* expression of 19% compared to the control sequence. At positions 38 nt and 48 nt, with regard to the 5´ UTR start, decreases of 27% and 31% were detected, respectively. At the position 6 nt upstream of the translation start site (135 nt downstream of the 5´ UTR start) the *eGFP* expression was decreased by 25%. Although it can only be assumed that the obtained effect of the G-quadruplex forming sequence on the level of reporter expression is based on a regulatory mechanism during translation and not during transcription, it is very likely because Halder *et al.* (2009) determined the amount of mRNA in a very similar experiment in a previous study by semi-quantitative PCR. They observed similar levels of reporter mRNA for a wild-type construct compared to a construct carrying a G-quadruplex forming sequence or a control sequence [69].

Discussion:

The obtained results indicate for the potential of the *MAPK2* G-quadruplex forming sequence to negatively regulate reporter expression if it is located within the 5´ UTR of the mRNA. As expected, the amount of reporter expression seems to be dependent on the position of the inserted G-quadruplex forming sequence, since the level of *eGFP* expression differs for the G-quadruplex forming sequences in comparison to their corresponding control sequences. Interestingly, also the corresponding control sequences are likely to influence the level of expression, as indicated by their differences in reporter expression compared to the level of wild type expression. This observation might be

explained by the destruction of other secondary structures within the 5´ UTR if a particular sequence is inserted. Therefore, also the insertion of a non G-quadruplex forming sequence might influence the level of reporter expression.

The results differ from the data observed by Kumari *et al.* (2008) who found a stronger inhibitory effect of a G-quadruplex forming sequence within the first 50 nt with regard to the 5´ UTR start and a weaker inhibitory effect close to the translation start site [152]. Although the strongest inhibitory effect of the *MAPK2* G-quadruplex forming sequence compared to the corresponding control sequence was observed at the position 48 nt downstream of the 5´ UTR start, also the weakest effect was detected within the same region, namely at position 28 nt downstream of the 5´ UTR start. Furthermore, the strongest inhibition of reporter expression compared to the wild type sequence was achieved by the insertion of the *MAPK2* G-quadruplex forming sequence at the position close to the translation start site (135 nt) which is more in line with the findings of Halder *et al.* (2009) [69]. Taken these findings together, it seems to be complicated to find reliable predictions on the position-dependent influence of the insertion of a G-quadruplex forming sequence into 5´ UTRs on expression. The insertion of a sequence itself might destroy an already existing secondary structure and therefore might influence the level of expression independently of its own structure. Therefore, the influence of a particular G-quadruplex forming sequence on expression might be investigated at a particular position in a particular 5´ UTR.

For further investigations of the influence of proteins and small molecules on *MAPK2* RNA G-quadruplex a sufficient G-quadruplex-dependent inhibition of translation is required. Therefore, the construct with the *MAPK2* RNA G-quadruplex forming sequence at position 48 nt downstream of the 5´ UTR start of the reporter mRNA, which revealed the strongest inhibition of expression, was used for the following experiments.

3.1.2 Investigation of possible RNA G-quadruplex-interacting proteins

G-quadruplex forming sequences seem to play an important role in the regulation of gene expression. Therefore, it is assumed that the formation as

well as the unwinding of G-quadruplexes is tightly regulated within cells. While many proteins interacting with DNA G-quadruplexes have been identified, only a few proteins interacting with RNA G-quadruplexes have been found yet. Among them are the two helicases DHX36 [196] and more recent DHX9 [197] which were shown to possess RNA G-quadruplex unwinding activity. Both helicases belong to the human DExH/D-box family of RNA helicases, indicating that members of this protein family might be involved in RNA G-quadruplex regulation. Therefore, further members of the human DExH/D-box family with RNA helicase activity, namely DHX29 [198], DDX3X [199], eIF4a1, eIF4a2 and eIF4a3 [200], were investigated for their ability to influence the *MAPK2* RNA G-quadruplex formation. In addition, DHX36 was also tested for its potential to unwind the *MAPK2* RNA G-quadruplex.

To study the ability of DHX36, DHX29, DDX3X, eIF4a1, eIF4a2 and eIF4a3 to unwind the *MAPK2* RNA G-quadruplex within the 5´ UTR of an mRNA, HeLa 229 cells were firstly transfected with 10 nM siRNAs against the respective proteins 24 h before transfection with the reporter constructs. As reporter constructs the plasmid pcDNA5/FRT/TO-eGFP containing the *MAPK2* G-quadruplex forming sequence at position 48 nt downstream of the 5´ UTR start or the corresponding control sequence and the plasmid pcDNA5/FRT/TO-mCherry were used. The level of reporter expression was measured 24 h after plasmid transfection of HeLa 229 cells via the Tecan infinite M200 fluorescence-reader (Fig. 3.1.4).

To proof the efficiency of the knock-down of possible RNA G-quadruplex-interacting proteins by siRNAs in HeLa 229 cells, the knock-down of DHX36 by siRNAs was exemplarily determined. Therefore, the expression of DHX36 after transfection of the cells with the respective siRNAs was examined via immunoblotting. For this purpose HeLa 229 cells were transfected with 10 nM siRNAs against DHX36 48 h before cell lysis. After separation of the cell lysate via SDS-gel electrophoresis, the siRNA-dependent knock-down of DHX36 was visualized by Western blot (Fig. 3.1.3).

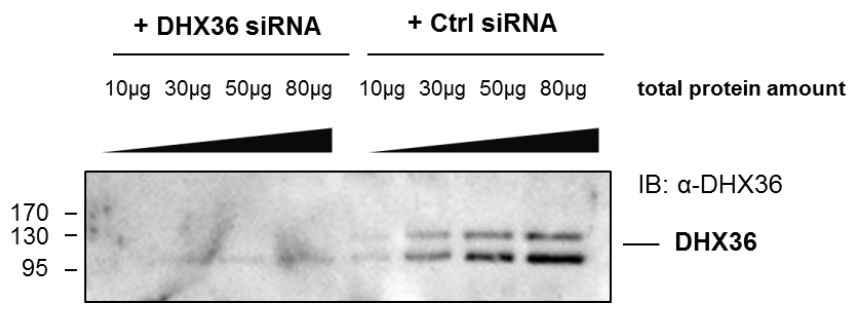


Fig. 3.1.3: Knock-down of DHX36 by siRNAs in HeLa 229 cells. HeLa 229 cells were transfected with siRNAs (each final conc. of 10 nM) against DHX36 48 h before lysis. As control HeLa 229 cells transfected with nonsilencing siRNA (Ctrl) were used. Different amounts of total protein (10-80 µg) were loaded onto a 10% SDS-gel. Immunoblotting was performed using a polyclonal antibody against DHX36.

The Western blot analysis of HeLa 229 cells transfected with siRNAs against DHX36 revealed a siRNA-dependent reduction of the DHX36 protein expression (115 kDa) for all tested protein concentrations (Fig. 3.1.3). In the control sample (cells transfected with nonsilencing siRNA) DHX36 could be detected in correlation to the loaded protein concentrations. These results argue for an efficient knock-down of DHX36 by the siRNAs.

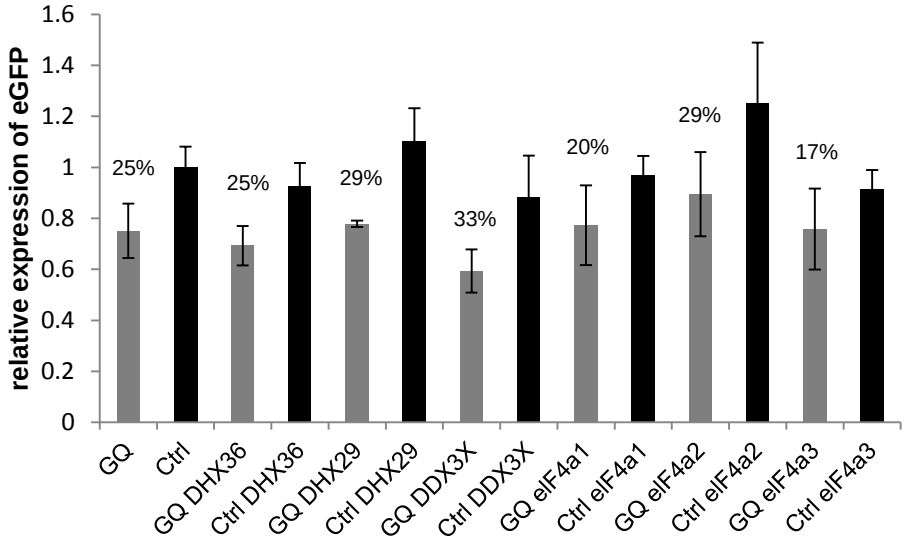


Fig. 3.1.4: Influence of the RNA helicases DHX36, DHX29, DDX3X, eIF4a1, eIF4a2 and eIF4a3 on the *MAPK2* RNA G-quadruplex within the 5´ UTR of the reporter mRNA. HeLa 229 cells were transfected with 10 nM of the respective siRNAs 24 h before co-transfection with the reporter constructs pcDNA5/FRT/TO-eGFP containing the *MAPK2* G-quadruplex forming

sequence (GQ) or the corresponding control sequence (Ctrl) within the 5´ UTR of the reporter mRNA and pcDNA5/FRT/TO-mCherry. Transfection with the reporter construct was performed by using 100 ng plasmid/well in a 96 well format. Level of reporter expression was measured 24 h after plasmid transfection of HeLa 229 cells via fluorescence-reader. Numbers on the top of the gray bars show the relative inhibition of expression in comparison to the corresponding control (black bars) in percentage. The level of reporter expression of Ctrl was set to 1. The error bars represent the standard deviation of triplicates.

For the *MAPK2* G-quadruplex forming sequence at position 48 nt downstream of the 5´ UTR (GQ), an inhibition of 25% of reporter expression was observed compared to the corresponding control (Ctrl) (Fig. 3.1.4). A further slight decrease in reporter expression could be only detected after the knock-down of DHX29, DDX3X or eIF4a2 by respective siRNAs. The knock-down of both DHX29 and eIF4a2 revealed a decrease of 29% in the reporter expression, which means a further decrease of only 4%. The knock-down of DDX3X resulted in a decrease of reporter expression of 33%. The knock-down of DHX36, whose ability for RNA G-quadruplex unwinding activity was already shown, did not show any effect. For the knock-down of eIF4a1 or eIF4a3 a slight increase in the reporter expression was detected.

Discussion:

In summary, the obtained results do not allow to make clear conclusions about the ability of the examined RNA helicases to influence the *MAPK2* RNA G-quadruplex within the 5´ UTR. No distinct reduction of the reporter expression could be detected after knocking-down the RNA helicases DHX36, DHX29, DDX3X, eIF4a1, eIF4a2 and eIF4a3 by siRNAs. Although slight decreases in the reporter expression were observed after knock-down of DDX3X, eIF4a2 and DHX29, none of these results can be evaluated as differences in reporter expression with a view to the high standard deviations. It is also unclear why the knock-down of DHX36, which has already been shown to possess RNA G-quadruplex unwinding activity, failed to further reduce G-quadruplex-dependent reporter expression. Although the Western blot against DHX36 displayed an almost complete knock-down of DHX36 in HeLa 229 cells transfected with siRNAs, DHX36 might not be the major source of G-quadruplex unwinding activity in HeLa 229 cells and therefore its effect was not detectable in the used

system. It is also possible that unwinding of RNA G-quadruplexes is facilitated by several different enzymes and therefore knock-down of a single enzyme is not sufficient to detect a change in G-quadruplex-dependent reporter expression. Furthermore, it is also conceivable that, although DHX36 has RNA G-quadruplex unwinding activity, its unwinding activity is specific for certain G-quadruplexes. In Creascy *et al.* (2008) the G-quadruplex unwinding activity by DHX36 was detected by performing *in vitro* experiments with a sequence which folds into an intermolecular G-quadruplex [196]. Therefore, the herein tested G-quadruplex forming sequence, which should fold into an intramolecular G-quadruplex structure, might not be a relevant target of DHX36.

3.1.3 Investigation of possible RNA G-quadruplex-interacting small molecules

The discovery of G-quadruplex forming sequences within the 5´ UTRs of several mRNAs of clinical interest highlights G-quadruplex structures as suitable therapeutic targets. In general, small molecules interacting with RNA G-quadruplexes might act either by stabilizing or destabilizing the G-quadruplex structure or by interference with a G-quadruplex-interacting protein. While most of the research on the development of G-quadruplex-interacting small molecules has been done for DNA G-quadruplexes, mainly in connection with the telomeric sequence, the knowledge about RNA G-quadruplex-interacting small molecules is much less advanced.

There are only a few publications addressing RNA G-quadruplex-interacting small molecules. For example, Balasubramanian and co-workers found in an *in vitro* system a quinolino-dicarboxamide derivative which stabilizes the G-quadruplex within the naturally occurring 5´ UTR of the NRAS mRNA [201]. The three bisquinolinium compounds Phen-DC_3, Phen-DC_6 and 360A (Fig. 3.1.5) were successfully tested *in vitro* by Gomez *et al.* (2010) for their ability to bind to the G-quadruplex formed within the 5´ UTR of the TRF2 mRNA [202].

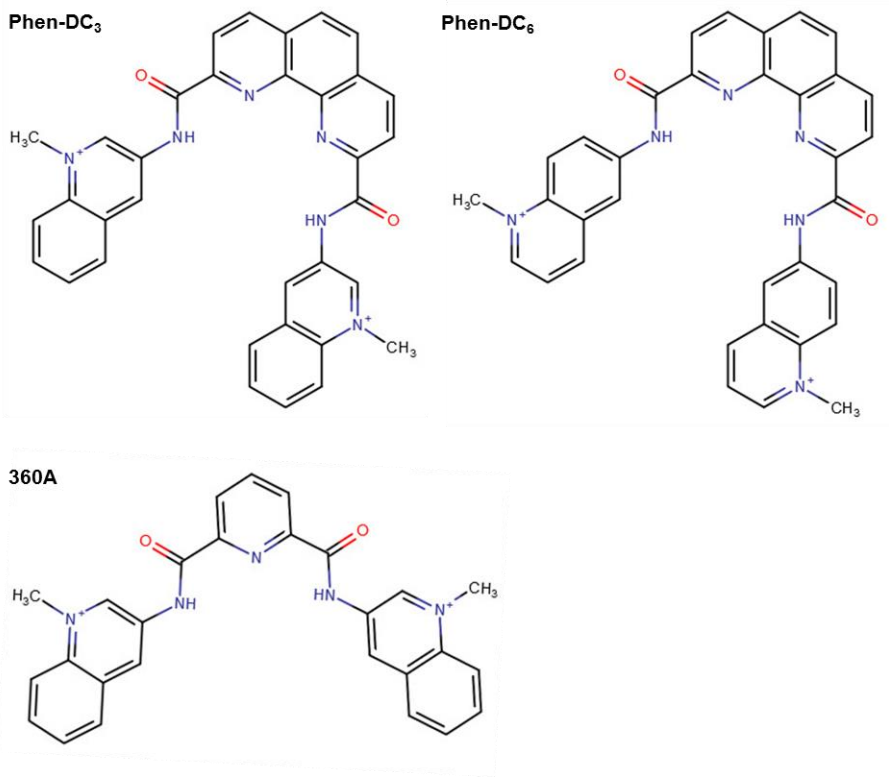


Fig. 3.1.5: Chemical structures of the bisquinolinium dicarboxamide derivatives Phen-DC_3, Phen-DC_6 and 360A.

Later on, our group could show by fluorescence intercalating displacement (FID) and CD spectroscopy experiments that the same molecules can bind to and stabilize artificial RNA G-quadruplexes [203]. In addition, we demonstrated in a cell culture system (HEK 293 cells) that the molecules Phen-DC_3, Phen-DC_6 and 360A are able to reduce reporter expression by stabilization of an artificial G-quadruplex within the 5´ UTR of a reporter mRNA [203].

To further investigate the G-quadruplex stabilizing potential of these three molecules, HeLa 229 cells were co-transfected with the plasmid pcDNA5/FTR/TO-eGFP, carrying the *MAPK2* G-quadruplex forming sequence (GQ) or the corresponding control sequence (Ctrl) at position 48 nt downstream of the 5´ UTR start site, and the plasmid pcDNA5/FTR/TO-mCherry. Five hours post-transfection the respective small molecule was added to the cells to a final concentration of 10 µM. Reporter expression was measured 24 h after transfection upon reaching 80% of cell-confluence via the Tecan infinite M200 fluorescence-reader (Fig. 3.1.6).

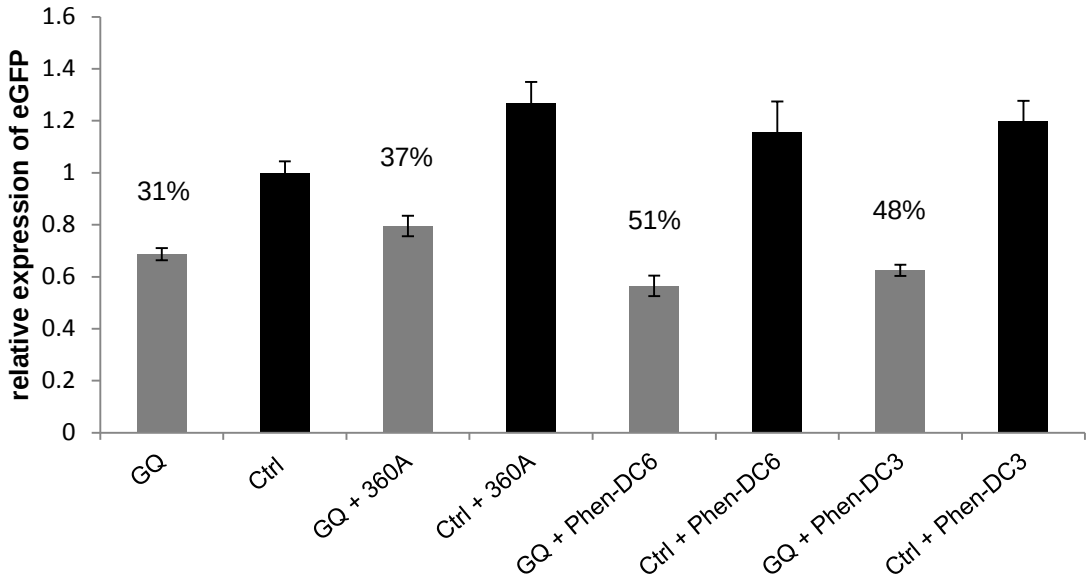


Fig. 3.1.6: Influence of the small molecules Phen-DC$_3$, Phen-DC$_6$ and 360A on *MAPK2* RNA G-quadruplex within the 5` UTR of the reporter mRNA. HeLa 229 cells were transfected with 200 ng/well of the reporter construct pcDNA5/FRT/TO-eGFP containing the *MAPK2* G-quadruplex forming sequence (GQ) or the corresponding control sequence (Ctrl) within the 5´ UTR and with the plasmid pcDNA5/FRT/TO-mCherry. Five hours post-transfection the respective small molecule was added to the cells to a final concentration of 10 µM. Level of reporter expression was measured 24 h after plasmid transfection of HeLa 229 cells via fluorescence-reader. Numbers on the top of the gray bars show the relative inhibition of expression in comparison to the corresponding control (black bars) in percentage. The level of reporter expression of Ctrl was set to 1. The error bars represent the standard deviation of triplicates.

Compared to the corresponding control sequence (Ctrl), the *MAPK2* G-quadruplex forming sequence (GQ) led to an inhibition of reporter expression of about 31% (Fig. 3.1.6). The addition of 10 µM of 360A only slightly further increased the inhibition by about 6% in comparison to its corresponding control (Ctrl + 360A), whereas the addition of 10 µM of the molecule Phen-DC$_6$ or Phen-DC$_3$ resulted in a further decrease of about 20% and 17% compared to the controls (Ctrl + Phen-DC$_6$ and Ctrl + Phen-DC$_3$). Although it cannot be excluded that the effect of these molecules on the reporter expression might be due to inhibition of transcription, it is unlikely because we determined the amount of transcribed mRNA in our previous study by semi-quantitative PCR and did not observe a small molecule-dependent change in the level of mRNA transcription [203].

Discussion:

The obtained results emphasize the RNA G-quadruplex stabilizing potential of the small molecules Phen-DC_3, Phen-DC_6 and to a smaller extend of 360A. These results are in line with our previous observations where we tested these molecules for their RNA G-quadruplex stabilizing potential in a different setting with artificial RNA G-quadruplex forming sequences located 11 nt upstream of the translation start site of an mRNA [203]. Similar to the herein described results we found a stronger reduction of RNA G-quadruplex-mediated reporter expression caused by Phen-DC_3 and Phen-DC_6 compared to 360A, if used in a cellular system with 10 µM concentrations in the context of a G-quadruplex forming sequence with a single nucleotide within the loop. Also Gomez *et al.* (2010) detected a stronger inhibitory effect by Phen-DC_3 and Phen-DC_6 compared to 360A on the level of reporter expression in an RNA G-quadruplex-controlled *in vitro* transcription/translation experiment [202].

The three molecules share some structural features which are thought to be important for their G-quadruplex stabilizing properties. They comprise of a crescent shape which is due to internal hydrogen bonds between the NH amid groups and the nitrogen (360A) or the nitrogens (Phen-DC_3 and Phen-DC_6) of the core unit. This shape is believed to be perfectly suited to overlap with a G-tetrad. In addition to the beneficial shape, all three molecules consist of large aromatic ring systems which are expected to contribute to excellent stacking interactions with the external G-tetrads. The observation of a stronger effect of Phen-DC_3 and Phen-DC_6 compared to 360A might be explained by their expanded aromaticity of their phenanthrolinedicarboxamide core compared to the pyridodicarboxamide core of 360A. Another feature of all three molecules is the presence of two positively charged side-arms which are thought to recognize the negatively charged phosphate backbone [204]. Recently, the structure of Phen-DC_3 bound to an intramolecular parallel DNA G-quadruplex, which was formed by a G-quadruplex forming sequence of the *c-MYC* promoter region, was solved via NMR-spectroscopy [205]. Thereby, strong π-π stacking interactions of Phen-DC_3 with the guanines of the top G-tetrad were observed. Even though this G-quadruplex was based on a DNA sequence, the mode of

binding of Phen-DC_3 to an RNA G-quadruplex might be similar or at least might include π-π stacking interactions with the external G-tetrads.

Interestingly, all three molecules seem to positively influence the level of the reporter expression of the corresponding controls. This observation was neither detected in our previous studies nor described by Gomez *et al*. (2010). It remains unclear whether these unexpected results might be a consequence of the quite high standard deviations or if the molecules are able to somehow positively influence expression in a so far not described manner, e.g. by interactions with other RNA structures within the mRNA.

3.2 Screening for TERRA-interacting proteins

The discovery that the human telomeric sequence is transcribed into the so-called telomeric repeat containing RNA (TERRA) by Azzalin and co-workers raised the question of the *in vivo* relevance of TERRA [100]. TERRA consists of 5` UUAGGG 3` tandem repeats with heterogeneous length and therefore belongs to the putative G-quadruplex forming RNA sequences [100]. The possibility of this sequence to fold into a G-quadruplex structure has been shown by NMR and X-ray crystallography [206-208]. Based on the finding that TERRA is co-localized at the telomeres, a regulatory and protective role of TERRA for the chromosomal ends is assumed [136]. This assumption has been further strengthened by the identification of several known telomere-associated proteins found to bind to a TERRA oligonucleotide [137]. Furthermore, the two members of the shelterin complex telomeric repeat factor 1 and 2 (TRF1 and TRF2) have been shown to interact with TERRA *in vivo* [137]. However apart from this *in vivo* study, the situation of proteins binding to TERRA within the cell remains still unclear.

To identify proteins interacting with TERRA *in vivo*, a cell-based crosslinking approach was performed (Fig. 3.2.1). The idea for the use of this method evoked in relation to a publication by Hafner *et al.* (2010) who developed the method in order to identify RNA binding proteins and microRNA target sites and referred to as PAR-CLIP (Photoactivatable-Ribonucleoside-Enhanced Crosslinking and Immunoprecipitation) [209]. This method is based on the random incorporation of the uridine analog 4-thiouridine (s4U) into cellular RNA. 4-thiouridine, in which the oxygen at position 4 of the pyrimidine ring is replaced by a sulfur atom, is photoactivatable with UV light above 320 nm wavelength [210]. The UV light exposure leads to specific crosslinks between proteins, which are in close proximity to s4U containing RNA, and the RNA itself [211]. Compared to conventional UV 254 nm crosslinking, the photoactivatable nucleotides should improve RNA recovery 100- to 1000-fold, using the same amount of radiation energy [209].

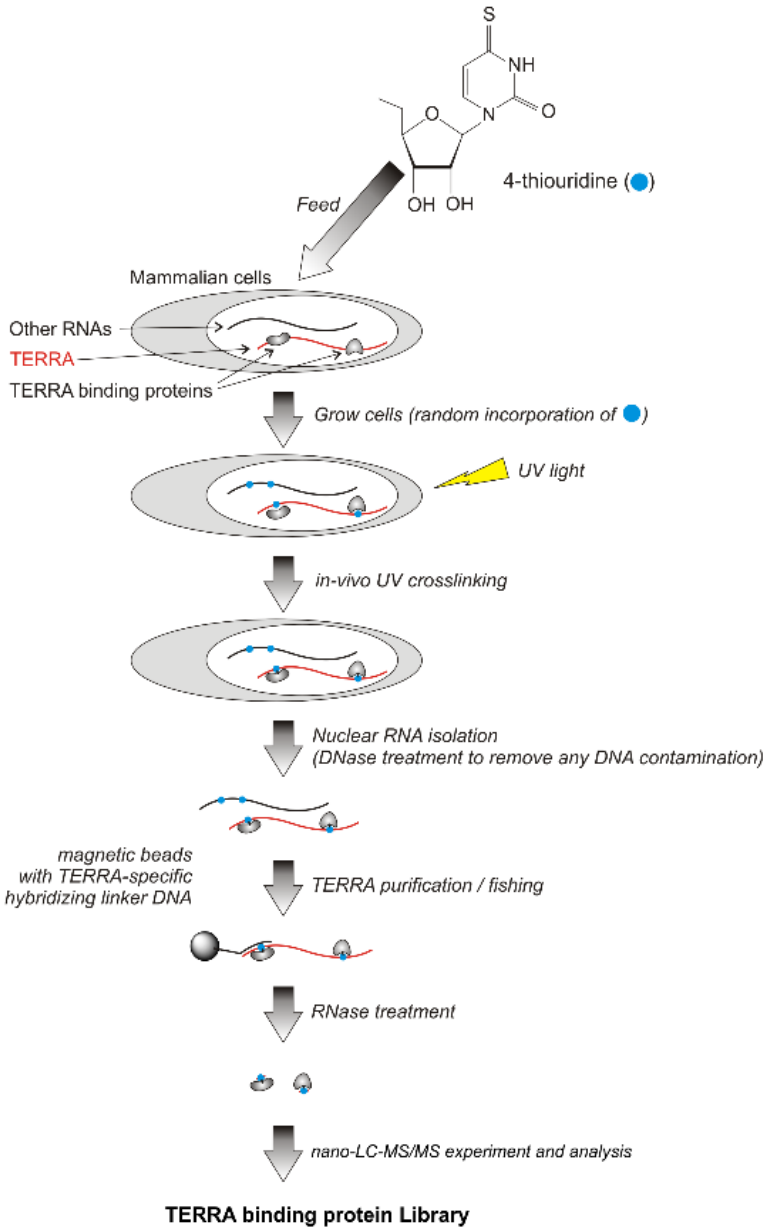


Fig. 3.2.1: Schematic representation of the cell-based crosslinking approach to purify *in vivo* TERRA-interacting proteins. HeLa S3 cells were fed with the photoactivatable uridine analog 4-thiouridine (s4U) (final conc. = 100 µM) for 48 h, followed by *in vivo* crosslinking with UV light of 365 nm (0.15 J/cm^2 for 30 min). Nuclear RNA was isolated and TERRA together with crosslinked proteins were fished using magnetic beads coupled to a TERRA-specific hybridization linker DNA (NH_2-(C_{12})-$(CCCTAA)_8$). After the release of TERRA-crosslinked proteins by RNase treatment, the proteins were separated via SDS-gel electrophoresis. To identify the crosslinked proteins, excised protein bands were analyzed using nano-LC MS/MS. Scheme by Dr. Kangkan Halder.

HeLa S3 cells were grown for 48 h in medium containing s4U. After random incorporation of s4U into RNA, the cells were exposed to UV light (365 nm).

Following UV crosslinking of s4U containing RNAs with interacting proteins, the nuclear extract was isolated and treated with DNase to remove DNA contaminations. To purify TERRA (TERRA-fishing), this extract was incubated with magnetic beads coupled to a TERRA-specific hybridizing linker DNA (NH_2-(C_{12})-$(CCCTAA)_8$). As a control, a randomized sequence of the TERRA-specific hybridizing linker DNA was used. After washing the magnetic beads, proteins crosslinked to TERRA were released by RNase treatment. To identify the crosslinked proteins, the released proteins were separated via SDS gel electrophoresis, protein bands were excised from the SDS gel and the samples were analyzed by nano-LC-MS/MS (performed by the proteomics facility of the University of Konstanz).

Fig. 3.2.2: SDS-PAGE analysis of possible TERRA-interacting proteins. The protein eluates after TERRA-fishing were separated on a 12% SDS gel followed by silver staining. First and fourth lane: Marker (M); second lane: protein eluate after fishing with the TERRA-specific hybridization linker DNA (TERRA); third lane: protein eluate after fishing with the randomized hybridization linker DNA (Ctrl).

After the separation of the proteins via SDS gel electrophoresis, a silver staining of the gel was performed to visualize the protein bands (Fig. 3.2.2). At first glance differences in the pattern of the samples between the TERRA-bound proteins and the control were hardly noticeable. Nevertheless, the protein bands were excised from the gel and investigated via nano-LC MS/MS. Unfortunately, the mass spectrometry analysis of the excised protein bands revealed no usable results. According to the proteomics facility only keratin was found within the samples.

Discussion:

The analysis of the excised protein bands via nano-LC MS/MS by the proteomics facility of the University of Konstanz identified solely keratin in both eluates. This result was not expected with regard to the pattern of the silver stained SDS gel.

Although no useable results could be revealed after nano-LC MS/MS, the herein described method can be a suitable and promising approach for the detection of TERRA-interacting proteins *in vivo*. Therefore, it might be worth to repeat the experiment. Due to the time-consuming procedure to generate sufficient amounts of cell nuclei lysate for TERRA-fishing, it was not feasible to repeat the experiment in terms of the current study.

The situation for TERRA-interacting proteins *in vivo* is still quite unclear. However based on RNA pull-down experiments followed by mass spectrometric analysis, a large number of possible TERRA-interacting proteins have been detected. As expected, most of these proteins are known telomere-associated proteins, e.g. proteins of the shelterin complex, the origin recognition complex (ORC) or proteins which belong to the heterogeneous nuclear ribonucleoprotein (HNRNP) family [137,212]. More recently, Scheibe *et al.* (2013) performed an elegant RNA pull-down experiment combined with stable isotope labeling by amino acids in cell culture (SILAC), thereby allowing also protein quantification [213]. In addition to the proteins found in previous studies, they were able to identify 15 novel TERRA-interacting proteins. Based on endoribonuclease-prepared short-interference RNA (esiRNA) knock-down experiments they found that SRRT/ARS2, which is known to be involved in miRNA processing, and

MORF4L2, a component of the NuA4 histone acetyltransferase complex, are involved in the regulation of detectable total TERRA as well as telomere-bound TERRA amounts. Although the herein mentioned RNA pull-down techniques seem well-suited for the identification of TERRA-interacting proteins, they all require nuclei lysis prior to the “fishing” of TERRA-interacting proteins. This lysis step might influence possible TERRA-protein interactions and therefore differ to the situation *in vivo*. It is therefore conceivable to lose possible protein-RNA interactions. Furthermore, these RNA pull-down techniques might be limited by the requirement of kinetically stable interactions between the respective RNA and the bound proteins and therefore may also lead to a loss of less stable protein-RNA interactions. To that end, the s4U-based screening approach might be helpful to overcome this issue because the crosslinking of proteins to RNA occurs prior to the nuclei lysis and stable RNA-protein interactions are not required.

3.3 Investigation of G-quadruplex formation *in vitro*

The formation of an intramolecular G-quadruplex requires a minimum of four consecutive G-tracts. Interestingly, naturally occurring G-quadruplex forming sequences are often composed of more than four G-tracts (e.g. *c-MYC* [214], *BCL-2* [97], *VEGF* [215]), thereby theoretically enable the formation of a G-quadruplex motif at different positions [39]. This circumstance raises the question of a biological benefit of G-quadruplex forming sequences containing more than four G-tracts.

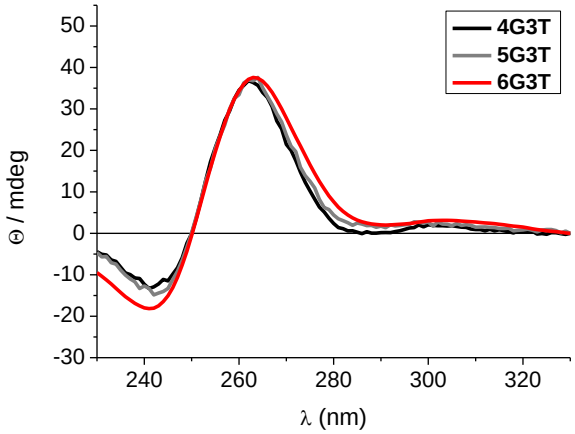
3.3.1 Investigation of structure and thermal stability of DNA G-quadruplexes formed by sequences with more than four G-tracts

Halder *et al.* (2009) investigated the structure and thermal stability of artificial RNA G-quadruplexes formed by sequences with four, five and six consecutive G-tracts *in vitro* by CD spectroscopy and CD melting experiments [69]. They found that an increasing number of G-tracts lead to an increase in G-quadruplex thermal stability, even though only four G-tracts should be involved in the formation of an intramolecular G-quadruplex. This unexpected finding raised the question if this stabilizing effect of additional G-tracts on RNA G-quadruplexes is transferrable to DNA G-quadruplexes. Therefore, the structure and thermal stability of DNA G-quadruplexes formed by sequences with four (4G3T), five (5G3T) and six (6G3T) G-tracts were examined *in vitro* via CD spectroscopy and CD melting experiments (Table 3.3.1 and Fig. 3.3.1).

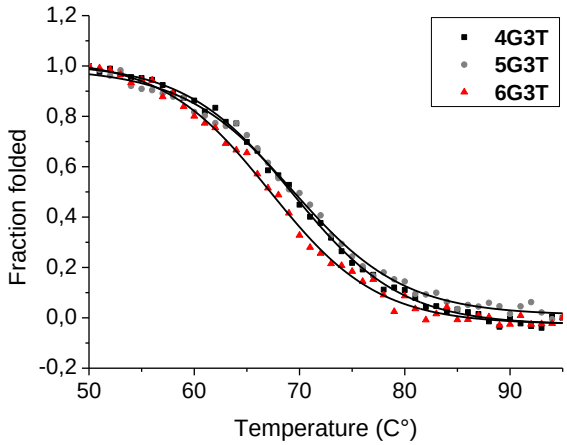
Table 3.3.1: Names and sequences of oligonucleotides used to investigate the structure and thermal stability of DNA G-quadruplexes formed by sequences with more than four G-tracts.

Name	Sequence 5´→ 3´
4G3T	GGGTGGGTGGGTGGG
5G3T	GGGTGGGTGGGTGGGTGGG
6G3T	GGGTGGGTGGGTGGGTGGGTGGG

A



B



C

Name	Tm (°C)
4G3T	69.4
5G3T	69.4
6G3T	67.2

Fig. 3.3.1: Influence of the number of G-tracts on DNA G-quadruplex structure and thermal stability. A: CD spectra of 5 µM 4G3T, 5G3T and 6G3T in the presence of 1 mM KCl. B: Normalized CD melting curves of 5 µM 4G3T, 5G3T and 6G3T at 262 nm in the presence of 1 mM KCl. C: Corresponding melting temperature midpoints (Tm) at 262 nm in the presence of 1 mM KCl.

The CD spectra of the G-quadruplex forming sequences 4G3T, 5G3T and 6G3T showed a positive peak at around 262 nm and a negative peak at around 242 nm in the presence of 1 mM KCl (Fig. 3.3.1 A). Such a CD spectrum is typical for a parallel G-quadruplex structure. The corresponding melting curves revealed similar melting temperatures (Tm) for G-quadruplexes formed by 4G3T and 5G3T (Tm = 69.4 °C) and a decrease of about 2°C for the G-quadruplex formed by 6G3T (Tm = 67.2°C) (Fig. 3.3.1 B and C). These results indicate that all three tested sequences fold into a parallel G-quadruplex and that the G-

quadruplex folded by the sequence 6G3T is slightly less stable than the G-quadruplexes folded by the sequences 4G3T and 5G3T.

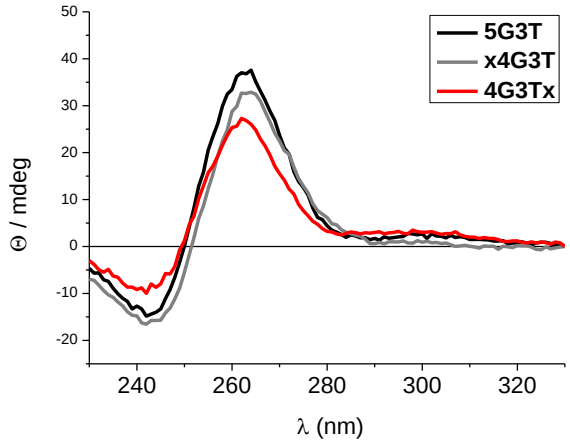
While the sequence 4G3T can only fold into one possible G-quadruplex, the sequences 5G3T and 6G3T have different possibilities for G-quadruplex formations. The formation of a G-quadruplex within the sequences 5G3T and 6G3T can lead to flanking sites either at the 5´ end or at the 3´ end or, as possible for 6G3T, at both ends of the G-quadruplex. In addition, it is also possible that one or, as possible for 6G3T, two of the G-tracts in the middle of the sequence become part of a loop within the G-quadruplex.

To analyze the influence of the additional G-tracts as flanking sites on G-quadruplex structure and thermal stability, further CD spectroscopy and CD melting experiments with 5G3T and 6G3T were performed (Fig. 3.3.2 and Fig. 3.3.3). The possibilities for G-quadruplex formation within the sequences 5G3T and 6G3T were reduced by selected point mutations. The x in the name of the sequences represents a G-tract in which the guanosine in the middle is substituted by a thymidine (Table 3.3.2 and Table 3.3.3). This alteration leads to a modified “G-tract” which is unable to participate in G-quadruplex formation.

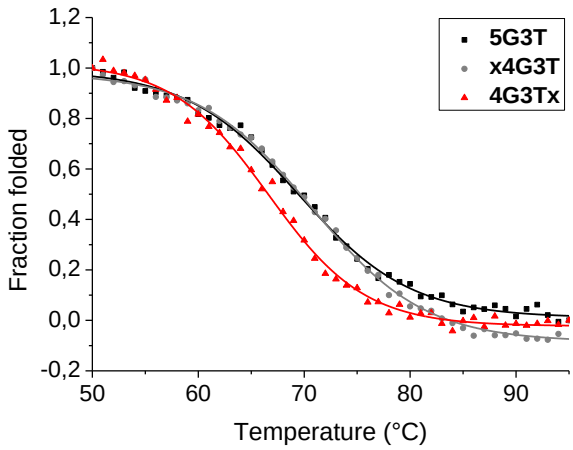
Table 3.3.2: Names and sequences of oligonucleotides used to investigate the influence of the G-quadruplex position on the structure and thermal stability of the DNA G-quadruplex formed by 5G3T.

Name	Sequence 5´→ 3´
5G3T	GGGTGGGTGGGTGGGTGGG
x4G3T	GTGTGGGTGGGTGGGTGGG
4G3Tx	GGGTGGGTGGGTGGGTGTG

A



B



C

Name	Tm (°C)
5G3T	69.4
x4G3T	70.7
4G3Tx	66.6

Fig. 3.3.2: Influence of the G-quadruplex position on the structure and thermal stability of the DNA G-quadruplex formed by 5G3T. A: CD spectra of 5 µM 5G3T, x4G3T and 4G3Tx in the presence of 1 mM KCl. B: Normalized CD melting curves of 5 µM 5G3T, x4G3T and 4G3Tx at 262 nm in the presence of 1 mM KCl. C: Corresponding melting temperature midpoints (Tm) at 262 nm in the presence of 1 mM KCl.

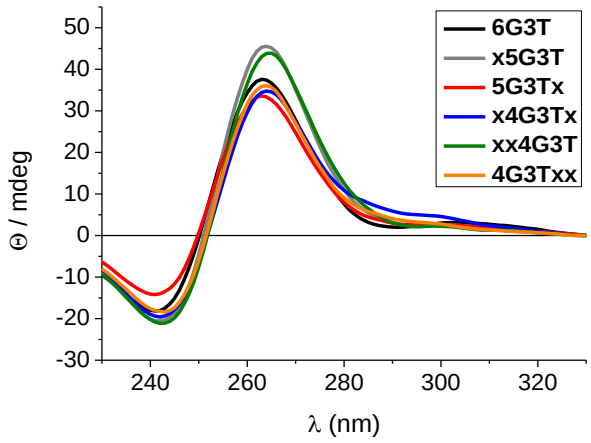
The CD spectra of the G-quadruplex forming sequences 5G3T, x4G3T and 4G3Tx showed typical spectra for the formation of a parallel G-quadruplex with a maximum at about 262 nm and a minimum at about 242 nm (Fig. 3.3.2 A). The G-quadruplex formed by the sequence x4G3T, which represents 5G3T, but with the G-quadruplex forced to form at the 3´ end, revealed a Tm value slightly higher than the G-quadruplex formed by 5G3T. In contrast, a decrease in thermal stability of about 3°C was observed for the G-quadruplex folded by the sequence 4G3Tx, which represents 5G3T with the G-quadruplex forced to form

at the 5´ end (Fig. 3.3.2 B and C). Based on these results the conclusion can be made that the formation of the G-quadruplex at the 3´ end with an overhang at the 5´ end is thermodynamically more stable compared to G-quadruplex formation at the 5´ end with an overhang at the 3´ end.

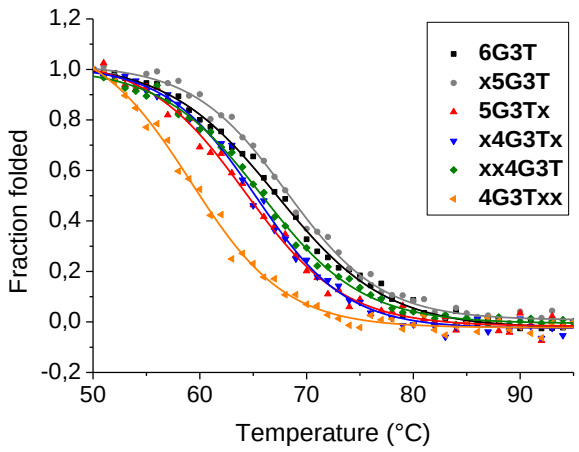
Table 3.3.3: Names and sequences of oligonucleotides used to investigate the influence of the DNA G-quadruplex position on structure and thermal stability of 6G3T.

Name	Sequence 5´→ 3´
6G3T	GGGTGGGTGGGTGGGTGGGTGGG
x5G3T	GTGTGGGTGGGTGGGTGGGTGGG
5G3Tx	GGGTGGGTGGGTGGGTGGGTGTG
x4G3Tx	GTGTGGGTGGGTGGGTGGGTGTG
xx4G3T	GTGTGTGTGGGTGGGTGGGTGGG
4G3Txx	GGGTGGGTGGGTGGGTGTGTGTG

A



B



C

Name	Tm (°C)
6G3T	67.2
x5G3T	68.1
5G3Tx	64.3
x4G3Tx	65.3
xx4G3T	65.8
4G3Txx	59.1

Fig. 3.3.3: Influence of the G-quadruplex position on the structure and thermal stability of the DNA G-quadruplex formed by 6G3T. A: CD spectra of 5 µM 6G3T, x5G3T, 5G3Tx, x4G3Tx, xx4G3T and 4G3Txx in the presence of 1 mM KCl. B: Normalized CD melting curves

of 5 µM 6G3T, x5G3T, 5G3Tx, x4G3Tx, xx4G3T and 4G3Txx at 262 nm in the presence of 1 mM KCl. C: Corresponding melting temperature midpoints (Tm) at 262 nm in the presence of 1 mM KCl.

Similar to the obtained CD spectra of the 5G3T variants, all six CD spectra of the G-quadruplex forming sequences of the 6G3T variants displayed typical spectra for a parallel G-quadruplex formation with a positive peak at about 262 nm and a negative peak at about 242 nm (Fig. 3.3.3 A). The G-quadruplex formed by the sequence x5G3T, in which the G-quadruplex is forced to form either in the middle or at the 3´ end, revealed a slightly higher Tm value of about 1°C than the G-quadruplex formed by 6G3T. For the G-quadruplex formed by the sequence 5G3Tx a decrease in thermal stability of about 3°C was detected. The G-quadruplex folded by the sequence x4G3Tx, in which the G-quadruplex can only fold in the middle of the sequence, exhibited a Tm value 2°C lower compared to the G-quadruplex folded by 6G3T. For the G-quadruplexes formed by the sequences xx4G3T and 4G3Txx Tm values were reduced by about 2°C and 8°C, respectively (Fig. 3.3.3 B and C).

Discussion:

The obtained Tm values for the DNA G-quadruplexes formed by 5G3T (Tm = 69.4°C) and 6G3T (Tm = 67.2°C) did not reveal an increase in thermal stability in comparison to the Tm value of the G-quadruplex formed by 4G3T (Tm = 69.4°C). While the G-quadruplexes formed by 4G3T and 5G3T showed similar results, the thermal stability of the G-quadruplex folded by 6G3T was reduced by about 2°C. These observations are in contrast to the results observed by Halder *et al*. (2009) in similar experiments with RNA G-quadruplexes. They found an increase in thermal stability with increasing number of G-tracts [69]. It can only be speculated that the flanking G-tracts have a stronger stabilizing impact on RNA G-quadruplexes compared to their DNA counterparts.

The performed experiments, in which the G-quadruplex formation is forced at specific positions within the sequence, gave further insights into the effect of flanking nucleotides on G-quadruplex thermal stability. The G-quadruplex formed by the sequence x4G3T showed a minimal increase in thermal stability

of about 1°C compared to the G-quadruplex folded by 5G3T. In case of x4G3T the G-quadruplex is located at the 3´ end of the sequence with a short nucleotide sequence which is not involved in G-quadruplex formation at the 5´ end. In contrast to the G-quadruplex formed by x4G3T, the G-quadruplex formed by the sequence 4G3Tx, in which the formation of the G-quadruplex is forced to fold at the 5´ end with a short overhang at the 3´ end, exhibited a decrease in thermal stability of about 3°C. These results indicate that flanking nucleotides at the 5´ end are able to stabilize the G-quadruplex structure, while flanking nucleotides at the 3´ end seem to have a negative effect on G-quadruplex thermal stability. This hypothesis is further supported by the results obtained for x5G3T and 5G3Tx. The G-quadruplex in the sequence x5G3T, which can be formed either in the middle or at the 3´ end of the sequence, showed, similar to the G-quadruplex formed by x4G3T, a slight increase in thermal stability of about 1°C compared to the G-quadruplex folded by 6G3T. The G-quadruplex folded by 5G3Tx, which is located either in the middle or at the 5´ end of sequence, exhibited a decrease of about 3°C compared to the G-quadruplex formed by 6G3T and about 4°C to the G-quadruplex formed by x5G3T.

The obtained Tm values for the G-quadruplexes formed by xx4G3T and 4G3Txx were 65.8°C and 59.1°C, respectively. Although both G-quadruplexes are less stable than the G-quadruplex formed by 6G3T, the G-quadruplex of xx4G3T, which has to fold at the 3` end of the sequence, is about 7°C above the received Tm value of the G-quadruplex folded by 4G3Txx. These data indicate that, apart from a slight stabilizing effect of flanking nucleotides at the 5´ end of a DNA G-quadruplex, the increasing number of flanking nucleotides at both the 5´ end and at the 3´ end is likely to negatively affect G-quadruplex thermal stability.

The G-quadruplex folded by the sequence x4G3Tx, in which the G-quadruplex can only fold in the middle of the sequence, revealed a Tm value of 65.3°C. It can be speculated that the slightly positive effect of flanking nucleotides at the 5´ end on G-quadruplex thermal stability is counteracted by the stronger negative effect of flanking nucleotides at the 3´ end, thereby leading to a less

stable G-quadruplex compared to the G-quadruplexes formed by 6G3T and x5G3T. It can also not be excluded that the differences in the sequences, originated by the G to T substitutions, may as well influence the G-quadruplex thermal stability. This suspicion is further strengthened by comparison of x5G3T, x4G3Tx and xx4G3T. In case of x5G3T the G-quadruplex can fold either in the middle or at the 3´ end of the sequence, whereas the G-quadruplex in x4G3Tx is forced to fold in the middle. In case of xx4G3T the G-quadruplex formation can only occur at the 3´ end of the sequence. The finding that x4G3Tx as well as xx4G3T, which should represent the two possibilities of G-quadruplex formation in x5G3T, revealed less stable G-quadruplexes compared to x5G3T raises the question if also the type of flanking nucleotide influences G-quadruplex thermal stability (cf. chapter 3.3.3). It is worth to mention that the G-quadruplex folded by 6G3T, 5G3T, x5G3T or 5G3Tx might theoretically contain G-tracts within its loops. However, this seems to be unlikely due to the DMS footprint result obtained for the investigation of a preferred position of G-quadruplex formation within T10-6G3T-T10 (cf. chapter 3.3.2).

3.3.2 Investigation of a preferred position of G-quadruplex formation within the DNA sequence 6G3T

The appearance of naturally occurring G-quadruplex forming sequences with more than four G-tracts raises the question of a preferred position of G-quadruplex formation in such sequences. The sequence 6G3T represents a G-quadruplex forming sequence in which a G-quadruplex can theoretically fold at different positions, either starting from the 5´ end, in the middle or from the 3´ end of the sequence (Fig. 3.3.4).

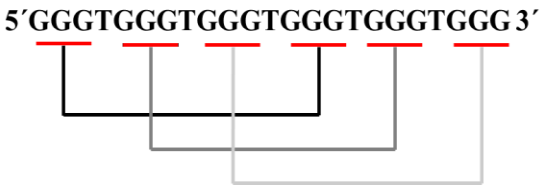


Fig. 3.3.4: Possible positions of G-quadruplex formation within the sequence 6G3T. The sequence 6G3T comprises three possible positions for G-quadruplex formation, starting either at the 5´ end, in the middle or at the 3´ end of the sequence. The three possible positions for G-quadruplex formation are indicated by the brackets.

In addition to these three possible positions of G-quadruplex formation, it is also possible that G-tracts might be included in loop sequences, thereby enabling further G-quadruplex formations. To investigate the structure and preferred position of a G-quadruplex within the sequence 6G3T, CD spectroscopy as well as DMS footprinting experiments of 6G3T were performed (Fig. 3.3.5 and Fig. 3.3.6). For a better separation of the cleavage products on a denaturating PAGE, a linker of ten thymidines (T10) was attached at the 5´ end as well as at the 3´ end of the 6G3T sequence (Table 3.3.4).

Table 3.3.4: Name and sequence of the oligonucleotide used to investigate the preferred position for G-quadruplex formation within T10-6G3T-T10.

Name	Sequence 5´→ 3´
T10-6G3T-T10	TTTTTTTTTTGGGTGGGTGGGTGGGTGGGTGGGTTTTTTTTTT

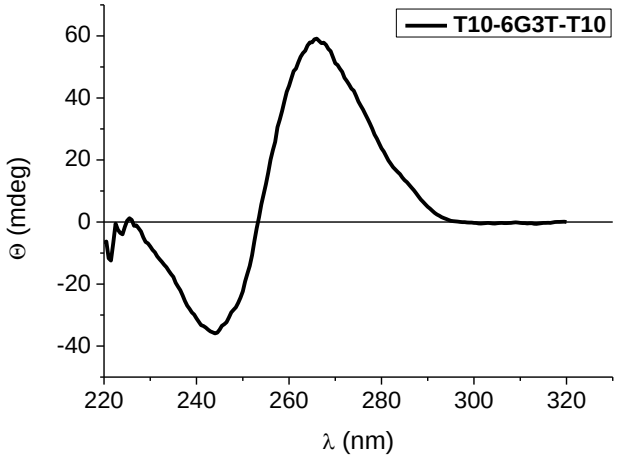


Fig. 3.3.5: CD spectrum of the G-quadruplex forming sequence T10-6G3T-T10. CD spectrum of 5 µM T10-6G3T-T10 in the presence of 10 mM KCl.

The CD spectrum of the G-quadruplex forming sequence T10-6G3T-T10 was recorded in the presence of a higher KCl concentration (10 mM), since it was assumed that the long thymidine linkers could result in a less stable G-quadruplex structure in comparison to the former used sequence 6G3T (cf. chapter 3.3.1). The measurement revealed, similar to the spectrum of the G-quadruplex formed by 6G3T, a typical spectrum for a parallel G-quadruplex

structure, as indicated by a positive peak at about 262 nm and a negative peak at about 242 nm wavelength (Fig. 3.3.5).

To determine the position of G-quadruplex formation within the sequence T10-6G3T-T10, T10-6G3T-T10 was analyzed by DMS footprinting (Fig. 3.3.6).

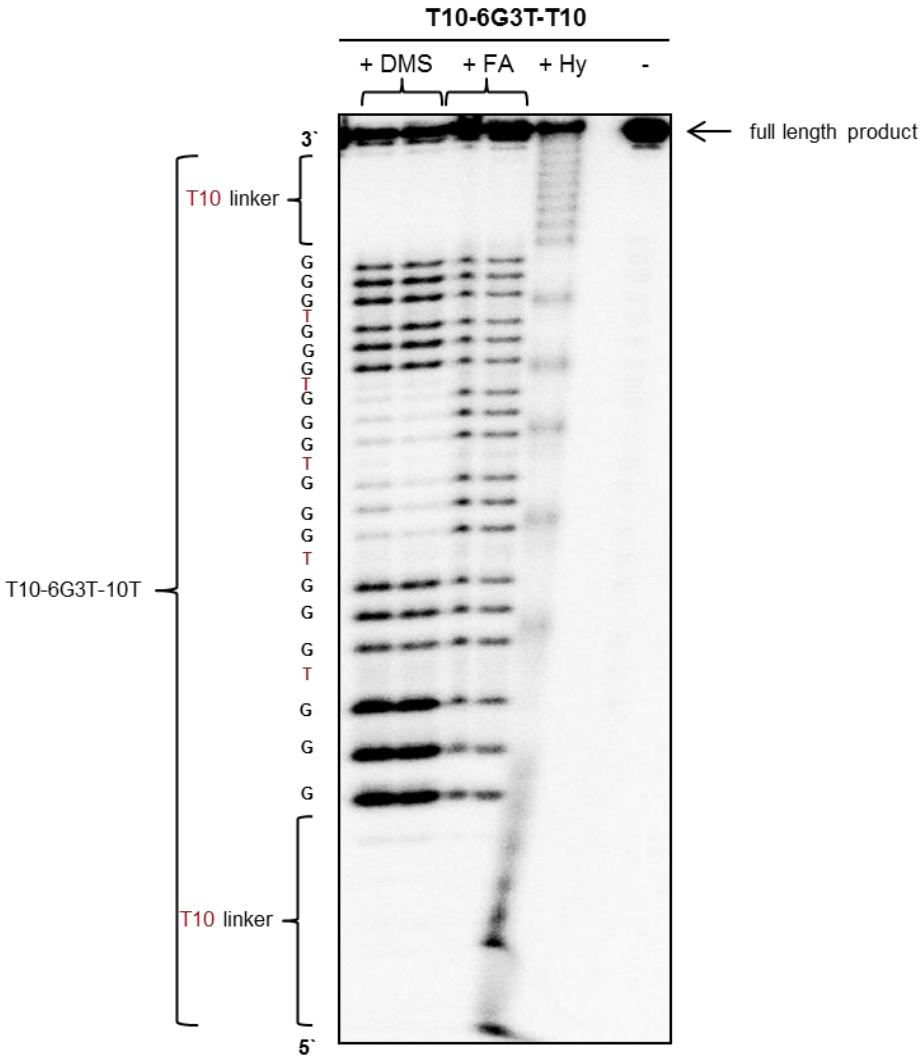


Fig. 3.3.6: Investigation of a preferred position of the G-quadruplex formed within the DNA sequence T10-6G3T-T10. DMS footprint of the T10-6G3T-T10 oligonucleotide analyzed on a 12% denaturating PAGE. Folding of G-quadruplex was performed in the presence of 20 mM KCl. Cleavage of purines and pyrimidines was performed via chemical sequencing using formic acid (FA) and hydrazine (Hy), according to Maxam and Gilbert [216]. The first two lanes show the cleavage products after DMS treatment (+ DMS). The next two lanes exhibit the cleavage products after cleavage of purines (+ FA), followed by a lane showing the cleavage products after cleavage of pyrimidines (+ Hy). The last lane displays the full length product (-). The corresponding sequence of T10-6G3T-T10 is depicted on the left site.

The results obtained by the DMS footprint suggest that the two G-tracts in the middle of the sequence T10-6G3T-T10 seem to be protected from DMS-mediated cleavage, while the two outer G-tracts at the 5´ end as well as at the 3´ end show cleavage products in the presence of DMS (Fig. 3.3.6 +DMS). This result leads to the conclusion that within the G-quadruplex forming sequence T10-6G3T-T10 all three possible G-quadruplex formations exist under the tested conditions. The G-tract, which is close to the 5´ end, shows the strongest cleavage product. Therefore, it can be assumed that the G-quadruplex motif starting from the 5´ end of the sequence is less frequently folded compared to the G-quadruplex motif in the middle or at the 3´ end of the sequence. Since the two G-tracts in the middle of the sequence seem to be protected from DMS-mediated cleavage, it can be claimed that they are involved in all G-quadruplex formations and therefore are not part of loop sequences.

Discussion:

The sequence T10-6G3T-T10, which is an artificial, symmetrical G-quadruplex forming sequence with long sequence overhangs at the 5´ end and the 3´ end, revealed a CD spectrum of a parallel G-quadruplex structure. The DMS footprinting experiment displayed a mixture of G-quadruplex formations at all three possible positions. Our previous findings of a more positive effect of flanking nucleotides at the 5´ end of the G-quadruplex might be seen as strengthened by the observation that the first G-tract, with regard to the 5´ end, seems to be less often included into G-quadruplex formation than the five other G-tracts, thereby preventing a longer 3´ overhang. However, the G-quadruplex does not exclusively fold from the 3´ end which would lead to the longest possible 5´ overhang and the shortest 3´ overhang. Due to the long T10 linker sequences the sequence T10-6G3T-T10 can hardly be compared with the former used sequence 6G3T (cf. chapter 3.3.1), as the long thymidine overhangs might influence G-quadruplex formation. Notwithstanding, the sequence might be better suited for representing *in vivo* cases in which a G-quadruplex forming sequence is usually flanked by long sequences at the 5´ and the 3´ end. Although DNA is usually double-stranded, the two strands become temporarily single-stranded during DNA transcription and G-quadruplex

formation might occur. It can be only speculated about possible reasons for G-quadruplex forming sequences with more than four G-tracts within the genome. G-quadruplex forming sequences with more than four G-tracts might be a consequence of repeat expansion. This phenomenon, which can appear during DNA replication, DNA repair and recombination, has been described as a major source of genomic instability and contributes to several hereditary neurological, neurodegenerative or developmental diseases [217,218]. For example, trinucleotide expansion, which is the predominant form of repeat expansions, is associated with some neuromuscular diseases [219]. In addition, also other repeat expansions like tetranucleotide [220] or pentanucleotide expansions [221] have been shown to correlate with hereditary diseases. However unlike to disease-associated repeat expansions, the repeat expansion of a G-quadruplex forming sequence might be beneficial. It is conceivable that the resulting increase of possibilities for G-quadruplex formation within a G-quadruplex forming sequence provides an advantage for the G-quadruplex-dependent influence on gene expression. Another advantage of G-quadruplex forming sequences with more than four G-tracts might be that they are less susceptible for mutations. In case that one of the G-tracts is destructed by mutations, the G-quadruplex might still be able to fold and therefore perform its regulatory function.

3.3.3 Influence of flanking nucleotides on DNA G-quadruplex structure and thermal stability

In previous studies the formation and stability of DNA G-quadruplexes have been mostly investigated without the context of flanking nucleotides. Studies on the influence of naturally occurring flanking nucleotides on the stability of G-quadruplexes formed by G-quadruplex forming sequences in the promoter region of the *c-KIT* oncogene [78], the NHE III_1 of the *c-MYC* promoter [74,75] and the telomeric sequence [76,77] revealed controversial findings. The data obtained for DNA G-quadruplexes formed by sequences with more than four G-tracts suggest that the length, the composition and also the site of flanking nucleotides might affect G-quadruplex stability (cf. chapter 3.3.1). According to the very few publications addressing the influence of flanking nucleotides on G-quadruplex stability our knowledge how flanking nucleotides might influence G-quadruplex

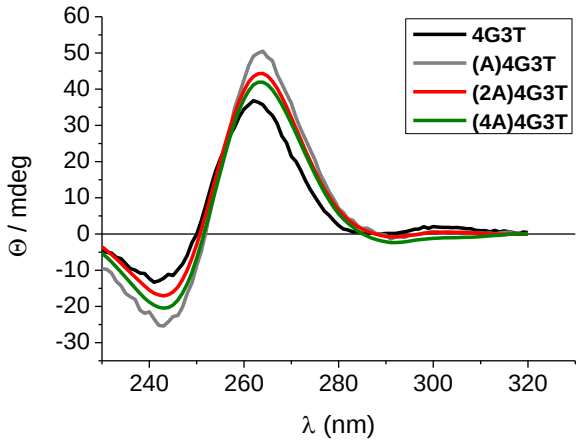
formation is quite limited. A better understanding how flanking sites affect G-quadruplex thermal stability is necessary for a more accurate prediction of the existence of putative naturally occurring G-quadruplexes. To obtain a more complete picture of the influence of flanking nucleotides at the 5´ or 3´ end of G-quadruplexes, the structure and thermal stability of the artificial G-quadruplex formed by the sequence 4G3T with increasing number of adenosines or thymidines as flanking nucleotides were analyzed.

To investigate the influence of flanking nucleotides at the 5´ end on DNA G-quadruplex structure and thermal stability, CD spectroscopy as well as CD melting experiments were performed with 4G3T sequences containing increasing numbers of either adenosines or thymidines at the 5´ end (Table 3.3.5, Fig. 3.3.7 and Fig. 3.3.8).

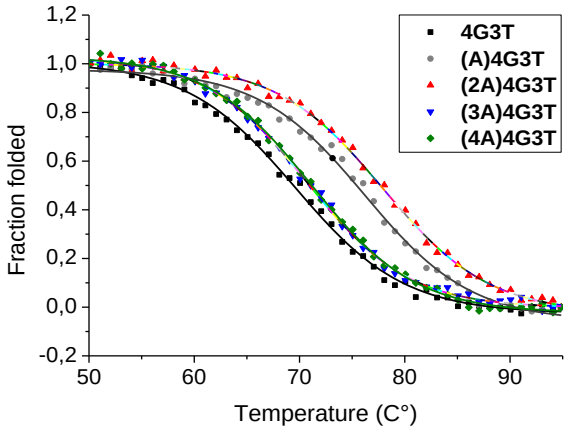
Table 3.3.5: Names and sequences of oligonucleotides used to investigate the influence of flanking nucleotides at the 5´ end on the structure and thermal stability of the DNA G-quadruplex formed by 4G3T.

Name	Sequence 5´→ 3´
4G3T	GGGTGGGTGGGTGGG
(A)4G3T	AGGGTGGGTGGGTGGG
(2A)4G3T	AAGGGTGGGTGGGTGGG
(3A)4G3T	AAAGGGTGGGTGGGTGGG
(4A)4G3T	AAAAGGGTGGGTGGGTGGG
(T)4G3T	TGGGTGGGTGGGTGGG
(2T)4G3T	TTGGGTGGGTGGGTGGG
(3T)4G3T	TTTGGGTGGGTGGGTGGG
(4T)4G3T	TTTTGGGTGGGTGGGTGGG

A



B

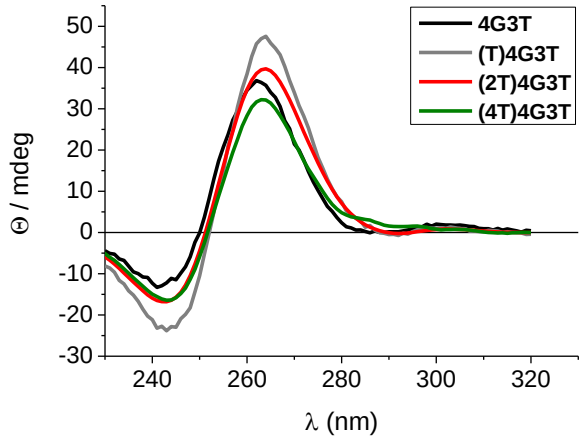


C

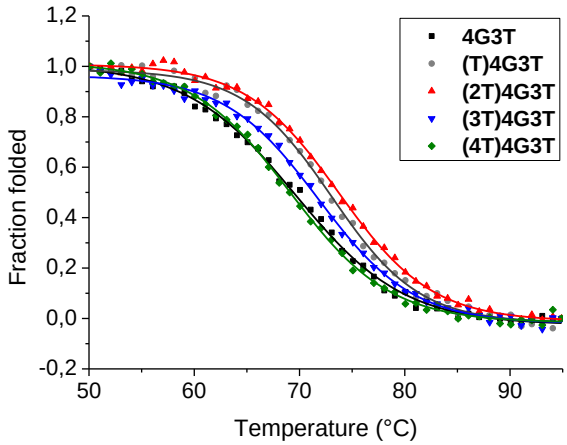
Name	Tm (°C)
4G3T	69.6
(A)4G3T	76.2
(2A)4G3T	77.9
(3A)4G3T	70.7
(4A)4G3T	71.1

Fig. 3.3.7: Influence of flanking adenosines at the 5´ end on the structure and thermal stability of the DNA G-quadruplex formed by the sequence 4G3T. A: CD spectra of 5 µM 4G3T, (A)4G3T, (2A)4G3T and (4A)4G3T in the presence of 1 mM KCl. B: Normalized CD melting curves of 5 µM 4G3T, (A)4G3T, (2A)4G3T, (3A)4G3T and (4A)4G3T at 262 nm in the presence of 1 mM KCl. C: Corresponding melting temperature midpoints (Tm) at 262 nm in the presence of 1 mM KCl.

A



B



C

Name	Tm (°C)
4G3T	69.6
(T)4G3T	73.0
(2T)4G3T	73.7
(3T)4G3T	71.8
(4T)4G3T	69.1

Fig. 3.3.8: Influence of flanking thymidines at the 5´ end on the structure and thermal stability of the DNA G-quadruplex formed by the sequence 4G3T. A: CD spectra of 5 µM 4G3T, (T)4G3T, (2T)4G3T and (4T)4G3T in the presence of 1 mM KCl. B: Normalized CD melting curves of 5 µM 4G3T, (T)4G3T, (2T)4G3T, (3T)4G3T and (4T)4G3T at 262 nm in the presence of 1 mM KCl. C: Corresponding melting temperature midpoints (Tm) at 262 nm in the presence of 1 mM KCl.

The CD spectra of all 4G3T sequences with flanking adenosines or thymidines at the 5´ end showed a negative peak at about 242 nm and a positive peak at 262 nm, indicating that all sequences fold into the parallel G-quadruplex conformation (Fig. 3.3.7 A and Fig. 3.3.8 A). The corresponding melting curves revealed an increase of the Tm value by the addition of a single adenosine ((A)4G3T) or two adenosines ((2A)4G3T) of about 7°C and 8°C, respectively (Fig. 3.3.7 B and C). The addition of a single thymidine ((T)4G3T) or two thymidines ((2T)4G3T) resulted in an increase of the Tm value of about 3°C and

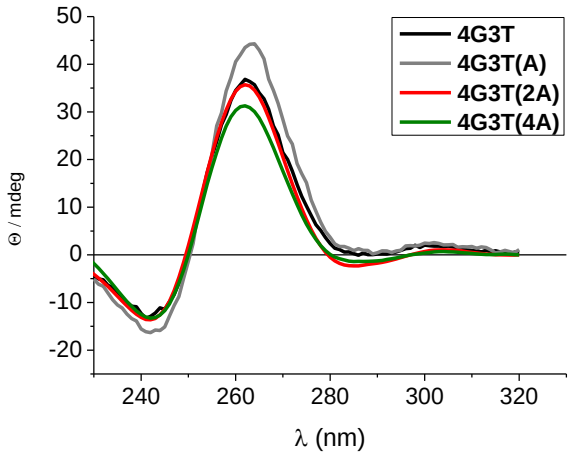
4°C, compared to 4G3T (Fig. 3.3.8 B and C). These results suggest a stabilizing effect of up to two adenosines or thymidines at the 5´ end on this DNA G-quadruplex. The addition of three or four adenosines ((3A)4G3T and (4A)4G3T) at the 5´ end of the 4G3T sequence lead to a decrease of the Tm value of the corresponding G-quadruplex of about 7°C compared to (2A)4G3T. However, the thermal stability of these G-quadruplexes is still slightly above the value observed for the G-quadruplex formed by 4G3T (Fig. 3.3.7 B and C). Similar to our findings for the addition of three or four adenosines at the 5´ end of the G-quadruplex forming sequence, also the addition of three or four thymidines ((3T)4G3T and (4T)4G3T) caused a decrease of the Tm values of the formed G-quadruplexes, compared to the G-quadruplex folded by (2T)4G3T (Fig. 3.3.8 B and C). While for the G-quadruplex folded by the sequence (3T)4G3T a decrease in the Tm value compared to the G-quadruplex formed by (2T)4G3T of about 2°C was observed, the CD melting curve for (4T)4G3T revealed a decrease of about 5°C. These results indicate for a stronger stabilizing effect of adenosines at the 5´ end on the G-quadruplex structure compared to thymidines. Furthermore, the stabilizing effect of adenosines or thymidines at the 5´ end on the G-quadruplex structure seems to be counteracted by a destabilizing effect which starts with the third nucleotide of the flanking sequence and is further increased with additional nucleotides.

To analyze the influence of flanking nucleotides at the 3´ end on DNA G-quadruplex structure and thermal stability, CD spectroscopy as well as CD melting experiments were performed with 4G3T sequences containing increasing numbers of either adenosines or thymidines at the 3´ end (Table 3.3.6, Fig. 3.3.9 and Fig. 3.3.10).

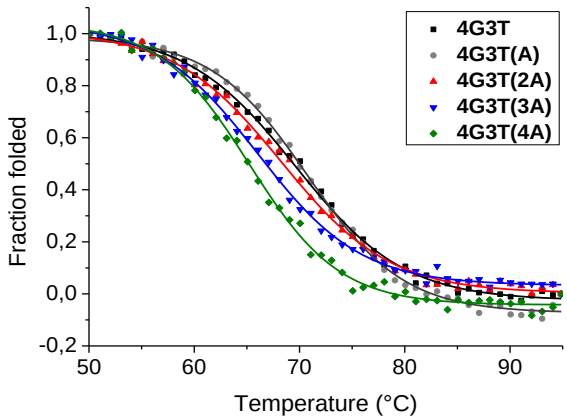
Table 3.3.6: Names and sequences of oligonucleotides used to investigate the influence of flanking nucleotides at the 3´ end on the structure and thermal stability of the DNA G-quadruplex formed by 4G3T.

Name	Sequence 5´→ 3´
4G3T	GGGTGGGTGGGTGGG
4G3T(A)	GGGTGGGTGGGTGGGA
4G3T(2A)	GGGTGGGTGGGTGGGAA
4G3T(3A)	GGGTGGGTGGGTGGGAAA
4G3T(4A)	GGGTGGGTGGGTGGGAAAA
4G3T(T)	GGGTGGGTGGGTGGGT
4G3T(2T)	GGGTGGGTGGGTGGGTT
4G3T(3T)	GGGTGGGTGGGTGGGTTT
4G3T(4T)	GGGTGGGTGGGTGGGTTTT

A



B

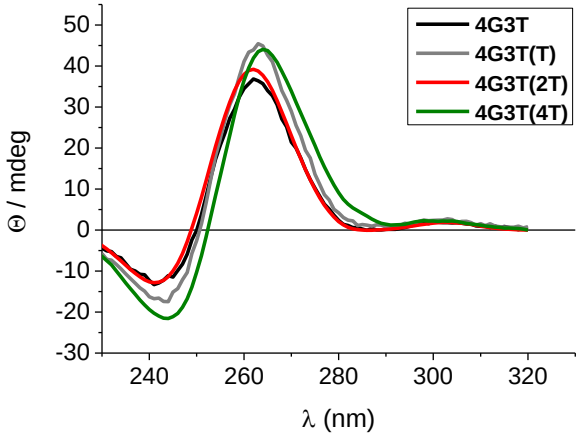


C

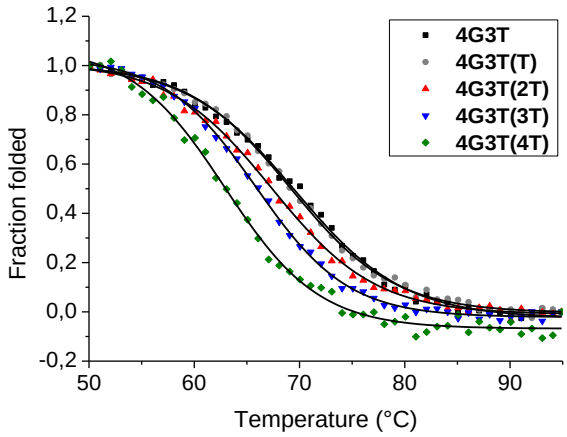
Name	Tm (°C)
4G3T	69.6
4G3T(A)	70.7
4G3T(2A)	68.3
4G3T(3A)	66.1
4G3T(4A)	65.2

Fig. 3.3.9: Influence of flanking adenosines at the 3´ end on the structure and thermal stability of the DNA G-quadruplex formed by the sequence 4G3T. A: CD spectra of 5 µM 4G3T, 4G3T(A), 4G3T(2A) and 4G3T(4A) in the presence of 1 mM KCl. B: Normalized CD melting curves of 5 µM 4G3T, 4G3T(A), 4G3T(2A), 4G3T(3A) and 4G3T(4A) at 262 nm in the presence of 1 mM KCl. C: Corresponding melting temperature midpoints (Tm) at 262 nm in the presence of 1 mM KCl.

A



B



C

Name	Tm (°C)
4G3T	69.6
4G3T(T)	69.2
4G3T(2T)	67.4
4G3T(3T)	65.9
4G3T(4T)	62.9

Fig. 3.3.10: Influence of flanking thymidines at the 3´ end on the structure and thermal stability of the DNA G-quadruplex formed by the sequence 4G3T. A: CD spectra of 5 µM 4G3T, 4G3T(T), 4G3T(2T) and 4G3T(4T) in the presence of 1 mM KCl. B: Normalized CD melting curves of 5 µM 4G3T, 4G3T(T), 4G3T(2T), 4G3T(3T) and 4G3T(4T) at 262 nm in the presence of 1 mM KCl. C: Corresponding melting temperature midpoints (Tm) at 262 nm in the presence of 1 mM KCl.

Similar to the observations for nucleotides at the 5´ end, the CD spectra of all 4G3T sequences with adenosines and thymidines at the 3´ end showed a minimum at about 242 nm and a maximum at 262 nm, indicating for the formation of a parallel G-quadruplex structure (Fig. 3.3.9 A and Fig. 3.3.10 A). The corresponding melting curves at 262 nm revealed a slight increase of the Tm value of the G-quadruplex formed by 4G3T only for the addition of a single adenosine at the 3´ end (4G3T(A)). For all other G-quadruplexes formed by sequences with flanking nucleotides at the 3´ end a decrease in thermal stability

was observed. The addition of a single adenosine at the 3´ end of the 4G3T sequence resulted in a minimal increase of the Tm value of the respective G-quadruplex of about 1°C. Adding further adenosines (4G3T(2A), 4G3T(3A) and 4G3T(4A)) caused a progressing decrease of the Tm values of the G-quadruplexes of 1°C, 3°C and 4°C compared to the G-quadruplex formed by 4G3T (Fig. 3.3.9 B and C). The findings with regard to the addition of thymidines at the 3´ end of the sequence were similar to those observed for adenosines. However for the addition of thymidines at the 3´ end of the sequence, even the addition of a single thymidine (4G3T(T)) already led to a minimal decrease of the observed G-quadruplex Tm value of about 0.5°C (Fig. 3.3.10 B and C). Similar to the observed results for the addition of adenosines at the 3´ end of the sequence, the addition of further thymidines at the 3´ end further destabilizes the G-quadruplex structure. While the G-quadruplex folded by 4G3T(2T) showed a decrease in the observed Tm value of about 2°C, the G-quadruplexes formed by the sequences 4G3T(3T) and 4G3T(4T) revealed a decline of 4°C and 7°C in their Tm values, respectively (Fig. 3.3.10 B and C).

Discussion

The results obtained for the addition of a single adenosine or thymidine and two adenosines or thymidines at the 5´ end of the G-quadruplex forming sequence 4G3T indicate for a stabilizing effect of flanking nucleotides at the 5´ end of a G-quadruplex structure on its thermal stability. The addition of a single adenosine to the 5´ end of 4G3T (4G3T(A)) caused an increase of the detected G-quadruplex Tm value of about 7°C compared to the 4G3T G-quadruplex. By adding a second adenosine the Tm value of the folded G-quadruplex was further increased of about 1°C. Similar results, but with lower impact were obtained for the addition of a single or two thymidines at the 5´ end of 4G3T. The addition of a single thymidine led to an increase of the detected G-quadruplex Tm value of about 3°C and the addition of two thymidines further increased G-quadruplex thermal stability of about 1°C. Interestingly, the addition of a third and fourth adenosine ((3A)4G3T and (4A)4G3T) resulted in a drastic decrease of G-quadruplex thermal stability, as shown by the decrease of the Tm value of about 7°C compared to the G-quadruplex folded by (2A)4G3T.

Although the Tm values of the G-quadruplexes formed by (3A)4G3T and (4A)4G3T are still slightly above the detected value for the G-quadruplex folded by 4G3T, the addition of a third and fourth adenosine seems to have a strong negative effect on G-quadruplex thermal stability. The results obtained for the addition of three and four thymidines at the 5´ end of 4G3T also revealed a decrease of G-quadruplex thermal stability compared to (2T)4G3T. However, the destabilizing effect of a third and fourth thymidine is less dramatic, compared to the effect observed for adenosines. The data for adenosines and thymidines at the 5´ end of 4G3T lead to the conclusion that only the first two nucleotides are involved in the stabilizing effect on the G-quadruplex structure, while the addition of further nucleotides is likely to destabilize the G-quadruplex structure. The observed stabilizing effect on the G-quadruplex structure was stronger in case of the addition of a single adenosine or two adenosines compared to a single thymidine or two thymidines.

The data obtained for flanking sites at the 3´ end of 4G3T showed a very weak stabilizing effect on G-quadruplex thermal stability. Only in case of adding a single adenosine at the 3´ end of 4G3T a weak stabilizing effect of about 1°C on G-quadruplex thermal stability, compared to the 4G3T G-quadruplex, was observed. In contrast, the addition of a single thymidine led to a decrease of G-quadruplex thermal stability of about 0.5°C. The addition of further adenosines as well as thymidines at the 3´ end of 4G3T resulted in a continuous decrease of G-quadruplex thermal stability, indicating that longer flanking sites at the 3´ end, similar to the observations for longer flanking sites at the 5´ end, have a destabilizing effect on G-quadruplex thermal stability.

The fact that flanking nucleotides might have a stabilizing effect on DNA and RNA structures has already been described for DNA and RNA double-stranded structures, such as hairpins [81] and siRNA [82]. *In vitro* experiments showed a stabilizing role of flanking ends on duplex formation for DNA as well as RNA by the so-called dangling effect [83-85]. The strength of this stabilizing effect was dependent on the closing base pair as well as the following dangling end and correlated with the length of this overhang. It is suggested that the dangling nucleobase stacks together with the neighboring closing base pair and thereby

acts as a hydrophobic cap at the end of the duplex. This shielding of the hydrogen bonds of the closing base pair leads to an increase of the thermodynamic stability of the DNA or RNA duplex structure [86]. In addition to the investigations of the influence of flanking sequences on the stability of double-stranded DNA and RNA, some work has been done to address the influence of flanking sequences on DNA G-quadruplexes. Studies of flanking nucleotides of the G-quadruplex forming sequence in the promotor region of the *c-KIT* oncogene revealed a destabilizing effect on G-quadruplex stability [78]. However, it is important to mention that the experiments were performed using the G-quadruplex forming sequence in the context of the adjacent 4, 6, 8 and 12 nucleotides at both flanking sites together, therefore they were unable to distinguish between effects based on 5´ or 3´ flanking sites. These results are in line with the findings obtained in this study, since a destabilizing effect of four flanking nucleotides either at the 5´ end or at the 3´ end on G-quadruplex thermal stability was detected. Experiments investigating modified G-quadruplex forming sequences of the NHE III_1 of the *c-MYC* promoter with one or three nucleotides at the 5´ end and one or two nucleotides at the 3´ end revealed a stabilizing effect of flanking sites on the G-quadruplex structure [75]. However, likewise to the experiments on the *c-KIT* G-quadruplex forming sequence, a distinction between the influences of the 5´ and the 3´ flanking sites on the G-quadruplex thermal stability was not determined. Two publications, one addressing the influence of 5´ flanking sites [76] and another one dealing with the influence of 3´ flanking sites on the G-quadruplex of the telomeric sequence [77], both revealed a decrease of G-quadruplex stability. Ambrus and co-workers examined the structure of a slightly modified G-quadruplex forming sequence of the NHE III_1 of the *c-MYC* promoter with the first three adjacent 5´ and 3´ nucleotides by means of NMR spectroscopy [74]. They observed that the flanking nucleotides at the 5´ and 3´ end of the G-quadruplex were interacting via stacking interactions with the bottom and the top G-tetrads, respectively. Valuable information about the influence of the composition of flanking sequences on RNA G-quadruplex formation was gained by the group of Perreault [79,80]. By investigating 5´ and 3´ UTR RNA G-quadruplex forming sequences in the context of their natural flanking sequences, they observed an inhibitory effect of cytosine tracts within the

flanking sequences on G-quadruplex formation. They attributed this finding to the formation of stable stem structures of those cytosine tracts with guanosines from the G-tracts. Although their observation was made in the context of RNA G-quadruplex forming sequences, this influence of cytosine tracts within flanking sequences might also count for DNA G-quadruplex forming sequences. Besides these controversial findings of the influence of flanking sites on G-quadruplex stability, the data obtained in this study lead to the assumption that the first two adenosines as well as the first two thymidines at the 5´ end are able to stabilize the 4G3T G-quadruplex. This stabilizing effect might be caused by stacking interactions with the bottom G-tetrad, thereby acting as a hydrophobic shield. Since adenosine has a larger aromatic ring system than thymidine, it might allow for stronger stacking interactions with the bottom G-tetrad. This would also explain the weaker impact of thymidine on G-quadruplex stability compared to adenosine. Indeed, it has already been shown that the stacking free energy for purine-purine stacks is greater than for purine-pyrimidine stacks [222,223]. The stabilizing effect of the first two nucleotides on G-quadruplex stability seems to be counteracted by a negative impact of longer 5´ flanking sites as observed for (3A)4G3T, (3T)4G3T, (4A)4G3T and (4T)4G3T. Since the stability of the G-quadruplex is likely to become drastically reduced by the addition of a third or fourth nucleotide (especially in case of adenosine), it is thinkable that these nucleotides might disturb the capping function of the first and/or second nucleotide, maybe by interactions with these nucleotides. Interestingly, a stabilizing effect of nucleotides at the 3´ end of the G-quadruplex structure was only observed for a single adenosine and this stabilizing effect was much weaker compared to its 5´ end counterpart. It can be only speculated that the flanking nucleotides at the 3´ end of the G-quadruplex formed by 4G3T are less enabled for stacking interactions with the top G-tetrad. This might be due to an unfavourable steric arrangement of the flanking nucleotides in regard to the top G-tetrad.

The results obtained in this study provide a systematic investigation of the influence of flanking nucleotides on DNA G-quadruplex structure and thermal stability, thereby emphasizing the importance of these nucleotides on the G-quadruplex stability. Most of the *in vitro* studies on G-quadruplex structure and

stability have been performed in the absence of flanking sites. However, in a biological context G-quadruplex forming sequences are usually surrounded by 5´ and 3´ sequences. Therefore, addressing this question can significantly improve our understanding how flanking sites can influence G-quadruplex structures and their stability. A better knowledge of the conditions that influence G-quadruplex structure and stability might contribute to a more accurate prediction of G-quadruplex formation in biological contexts. Furthermore, it would also be beneficial in terms of biological engineering where G-quadruplexes are used as biological tools.

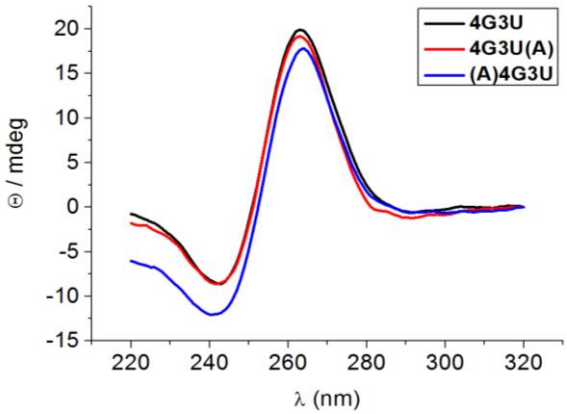
3.3.4 Influence of flanking nucleotides on structure and thermal stability of the RNA G-quadruplex formed by 4G3U

The findings that flanking nucleotides at the 5´ end and 3´ end of DNA G-quadruplexes influence their thermal stability led to the question if these results are transferable to RNA G-quadruplexes. To that end, a single adenosine was added either at the 5´ end or the 3´ end of the RNA G-quadruplex forming sequence 4G3U (Table 3.3.7) and these sequences were investigated via CD spectroscopy as well as CD melting experiments (Fig. 3.3.11).

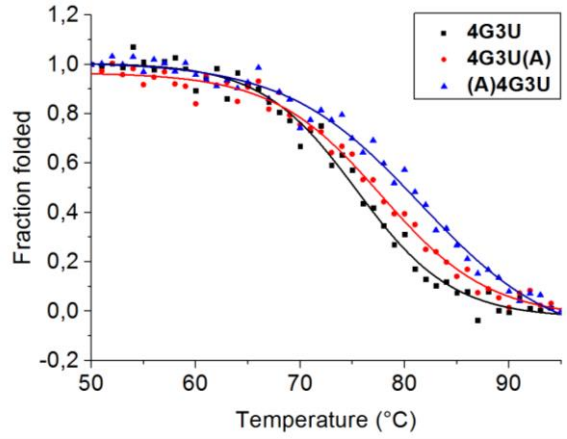
Table 3.3.7: Names and sequences of oligonucleotides used to investigate the influence of a flanking nucleotide either at the 5´ end or 3´ end on the structure and thermal stability of the RNA G-quadruplex formed by 4G3U.

Name	Sequence 5´→ 3´
4G3U	GGGUGGGUGGGUGGG
(A)4G3U	AGGGUGGGUGGGUGGG
4G3U(A)	GGGUGGGUGGGUGGGA

A



B



C

Name	Tm (°C)
4G3U	75.4
4G3U(A)	77.7
(A)4G3U	81.5

Fig. 3.3.11: Influence of a flanking adenosine either at the 5´ end or the 3´ end on the structure and thermal stability of the RNA G-quadruplex formed by 4G3U. A: CD spectra of 3 µM 4G3U, (A)4G3U and 4G3U(A) in the presence of 1 mM KCl. B: Normalized CD melting curves of 3 µM 4G3U, 4G3U(A) and (A)4G3U at 262 nm in the presence of 1 mM KCl.

The CD spectra of the investigated RNA sequences 4G3U, (A)4G3U and 4G3U(A) revealed a positive peak at around 262 nm and a negative peak at around 240 nm which are indicators of the formation of a parallel G-quadruplex structure. The corresponding melting curves at 262 nm wavelength resulted in a melting temperature midpoint of 75.4°C for the G-quadruplex formed by 4G3U, 81.5°C for the (A)4G3U G-quadruplex and 77.7°C for the 4G3U(A) G-quadruplex.

Discussion:

The RNA G-quadruplex forming sequence 4G3U, which represents the RNA equivalent of the former used DNA G-quadruplex forming sequence 4G3T, folded into a parallel G-quadruplex structure in the presence of 1 mM KCl. This result was expected, since RNA G-quadruplexes are thought to be monomorphic and are always found to adopt this G-quadruplex topology (cf. chapter 1.3.2.4). The CD melting experiment revealed that the RNA G-quadruplex has a higher thermal stability compared to the DNA G-quadruplex (Tm of 4G3U: 75.4°C; Tm of 4G3T: 69.6°C). Furthermore, the CD melting experiments resulted in a stabilizing effect of both, the addition of an adenosine at the 5´ end as well as at the 3´ end. Although there were less data collected on the influence of flanking nucleotides at the 5´ end or 3´ end on RNA G-quadruplex structure and stability, compared to the data obtained for DNA G-quadruplexes, the results are quite similar. As observed in case of the influence of a single adenosine at the 5´ end or the 3´ end on 4G3T G-quadruplex thermal stability, the addition of a single adenosine to the 5´ end of the sequence 4G3U led to a stronger increase in G-quadruplex thermal stability than the addition of an adenosine to the 3´ end. Even the observed changes in the Tm values by the addition of a single adenosine at the 5´ end or the 3´ end of 4G3U seem to be comparable to the DNA G-quadruplex folded by 4G3T. While adding a single adenosine at the 5´ end resulted in an increase of the G-quadruplex Tm values of about 6-7°C, the addition of a single adenosine to the 3´ end only increased the Tm values of about 1-2°C for the 4G3U as well as 4G3T G-quadruplexes. According to the similarities in the obtained CD spectra and the performed CD melting experiments it can be speculated that adenosines at the 5´ end and 3´ end stabilize the G-quadruplexes formed by 4G3T and 4G3U in a similar manner.

3.4 Screening for small molecules interacting with the human telomeric sequence

The telomeres are specialized DNA-protein structures that cap the end of all eukaryotic chromosomes. In humans they consist of tandem repeats of the sequence $(TTAGGG)_n$ with a length of 3-15 kb, ending in a single-stranded 3´ overhang of about 200 nt [224,225]. Telomeres play a crucial role in the protection of the chromosome ends and genetic integrity. This function seems to be facilitated by the formation of higher order structures [226-228]. Many proteins are known to interact with the telomeric sequence, e.g. proteins of the shelterin complex, and are thought to enable and control the particular higher order structures [116]. Due to the so-called end replication problem, the telomeres shorten after each round of cell division until they reach a critical length (so-called Hayflick limit), leading to senescence or apoptosis [229-231]. Telomere shortening can be counteracted by an enzyme called telomerase. This enzyme is composed of the telomerase reverse transcriptase (TERT) and the human telomerase RNA (hTR) which is used by TERT as a template for the elongation of the telomeric sequence [232,233]. Telomerase activity is only detectable in highly proliferating cells, like germ cells and stem cells, but not in somatic tissue [234]. However, in about 85-90% of all cancer cells and primary tumors the telomerase is overexpressed, thereby protecting the cells against telomere shortening and finally senescence and apoptosis [234,235]. Interestingly, the 3´ single-stranded telomeric overhang, which is the binding site of the human telomerase RNA (hTR) for telomere elongation, represents a putative G-quadruplex forming sequence. It is assumed that by stabilization of the G-quadruplexes within this sequence the binding of hTR and thereby telomerase-dependent elongation of the telomeres is blocked [236]. Furthermore, stabilization of the G-quadruplex structure can hinder telomere replication, thereby leading to DNA damage responses at the telomeres [237,238]. According to the pivotal role of the telomerase in cancer development, inhibition of telomerase activity seems to be an appropriate target for cancer therapy.

The inhibition of the human telomerase by the G-quadruplex stabilizing compound 2,6-diamidoantraquinone was first described by Sun and co-workers

in 1997 [133]. Subsequently to the finding of Sun and co-workers, plenty of other G-quadruplex binding molecules with the potential to inhibit the telomerase have been developed. Most of them contain a polycyclic heteroaromatic structure to enable π-π interactions with the external G-tetrads and positively charged side chains for interactions with the negatively charged DNA backbone [181,239,240]. In addition to synthesized molecules some naturally occurring small molecules with G-quadruplex binding properties were found, e.g. the cyclic polyamine telomestatin [241] or the alkaloid berberine [174,175]. It is worth to mention that stabilization of the telomeric G-quadruplexes via G-quadruplex stabilizing compounds might not only inhibit the telomerase activity, but seems also to negatively affect tumor cell growth in additional ways, e.g. by telomere uncapping [156,242-244] or by inhibition of transcription or translation of G-quadruplex containing oncogenes [244,245]. Therefore, some G-quadruplex stabilizing molecules may also be useful for treating some cancers which use the alternative lengthening of telomeres (ALT) pathway [246,247]. Although plenty of promising telomeric G-quadruplex binders have been developed, so far only a tiny fraction of these molecules has currently reached clinical trials, mostly in combination with other drugs, e.g. Topo I or PARP 1 inhibitors [181,183]. Therefore, there is still need for G-quadruplex binders with higher binding affinity and specificity.

3.4.1 Automated screening of small molecule libraries for telomeric G-quadruplex binders

To identify new candidates or lead structures of putative telomeric G-quadruplex binders, a FRET-based screening approach, developed by Dr. Armin Benz, was performed (in detail [248]). This screening was shown to be a reliable approach to identify molecules interacting with the telomeric G-quadruplex. Therefore, the screening was carried out using the same settings as described by Benz *et al.* (2011) [248]. Out of two small molecule libraries (ChemBioNet library (CBN) and the library of Prof. Dr. E. Daltrozzo (D), University of Konstanz) about 3800 structurally diverse compounds were automatically screened for their potential to stabilize G-quadruplex formation. For this screening approach an oligonucleotide composed of the sequence $GGG(TTAGGG)_3$ (HT) and labelled

with fluorescein (FAM) at the 5’ end and tetramethylrhodamine (TAMRA) at the 3’ end was used (Fig. 3.4.1 A). It has been shown that as long as the oligonucleotide is unstructured, emission of FAM (measured at 535 nm) can be measured after excitation at 485 nm wavelength. However in case of G-quadruplex formation, either induced by the presence of a monovalent cation like potassium or induced by a small molecule, FAM and TAMRA come into close vicinity, thereby enabling a fluorescence resonance energy transfer (FRET) between FAM and TAMRA (Fig. 3.4.1 B) [248,249]. Therefore, G-quadruplex formation can be measured by the decrease of FAM emission and the increase of TAMRA emission (measured at 590 nm) after excitation of FAM at 485 nm wavelength.

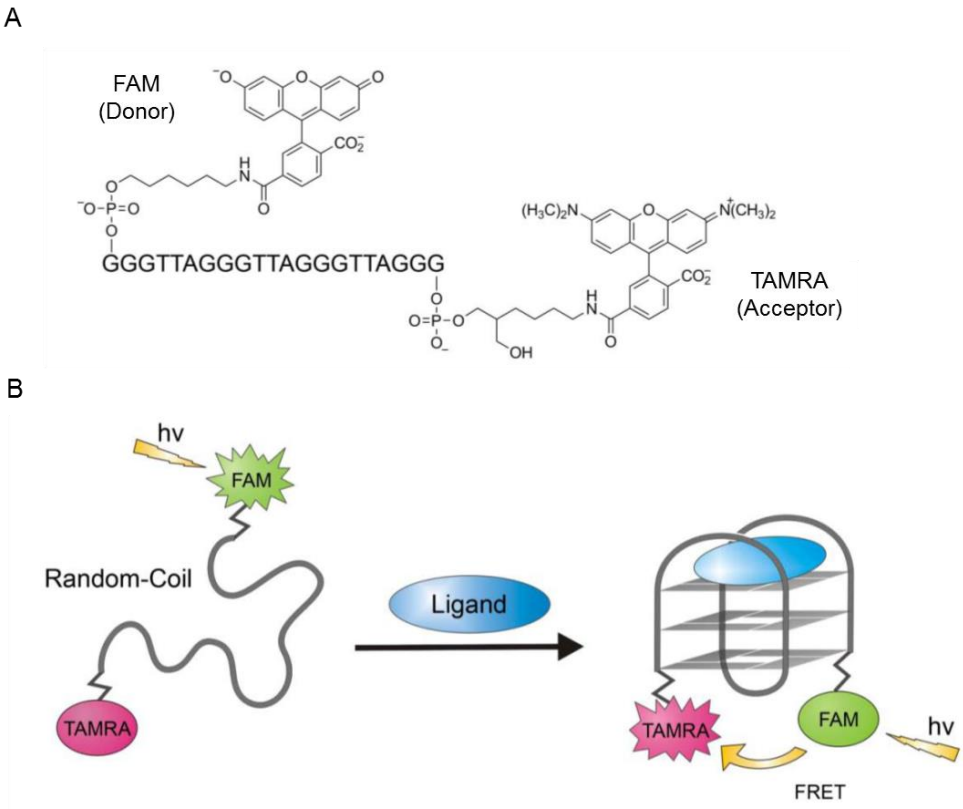


Fig. 3.4.1: FRET-based screening approach for the identification of putative telomeric G-quadruplex binders. A: Sequence of the HT-DNA oligonucleotide with the chemical structures of the 5´ coupled fluorescein (FAM) and 3´ coupled tetramethylrhodamine (TAMRA). B: Scheme of the underlying mechanism of the screening approach. In the absence of an appropriate ligand, the FAM/TAMRA labeled HT-DNA (FHT-DNA) is unstructured and therefore the two fluorophores FAM and TAMRA are situated away from each other. In this case the excitation of FAM with light of 485 nm wavelength is detectable by its emission at 535 nm wavelength. However in terms of a ligand-induced G-quadruplex formation, the two fluorophores come into close vicinity, thereby enabling fluorescence resonance energy transfer (FRET) between FAM and TAMRA. Therefore, the formation of a G-quadruplex can be measured by the decrease in

FAM emission (measured at 535 nm) and the increase in TAMRA emission (measured at 590 nm) after excitation at 485 nm wavelength. (Figure modified from A. Benz [250])

The screening was performed in a 384 well format using a robot system of the Tecan Freedom EVO® series and fluorescence was measured with a Tecan Infinite F500 plate reader. 0.2 µM of the FAM/TAMRA labelled HT (FHT) was incubated with ~ 95 µM compound in 10 mM Tris-HCl buffer (pH 7.5) for 20 min at RT, followed by detection of FAM and TAMRA emission after excitation at 485 nm wavelength in the fluorescence reader. As controls, FHT was incubated in the presence (positive control) or absence (negative control) of 100 mM KCl. The ratio of FAM/FRET for the negative control was set to 1 and the ratio of FAM/FRET for the positive control was set to 0. As a criterion for compound-induced G-quadruplex formation the FAM/FRET ratio was defined to be smaller than 0.4.

Via this automated screening 31 putative G-quadruplex binders were identified and further characterized by concentration dependency experiments. To that end, 0.2 µM of the FHT oligonucleotide was incubated together with a serial dilution of the identified compounds ranging from 100 µM to 1.6 µM. Like in the automated screen described above, the compound-induced G-quadruplex formation within FHT was analyzed via FRET measurement. The results of the 10 most promising compounds are depicted in Fig. 3.4.3; within these compounds seven hits (CBN1-7) were out of the ChemBioNet library (CBN) and three hits (D1-D3) out of the Daltrozzo library (D). Their corresponding chemical structures are shown in Fig. 3.4.2.

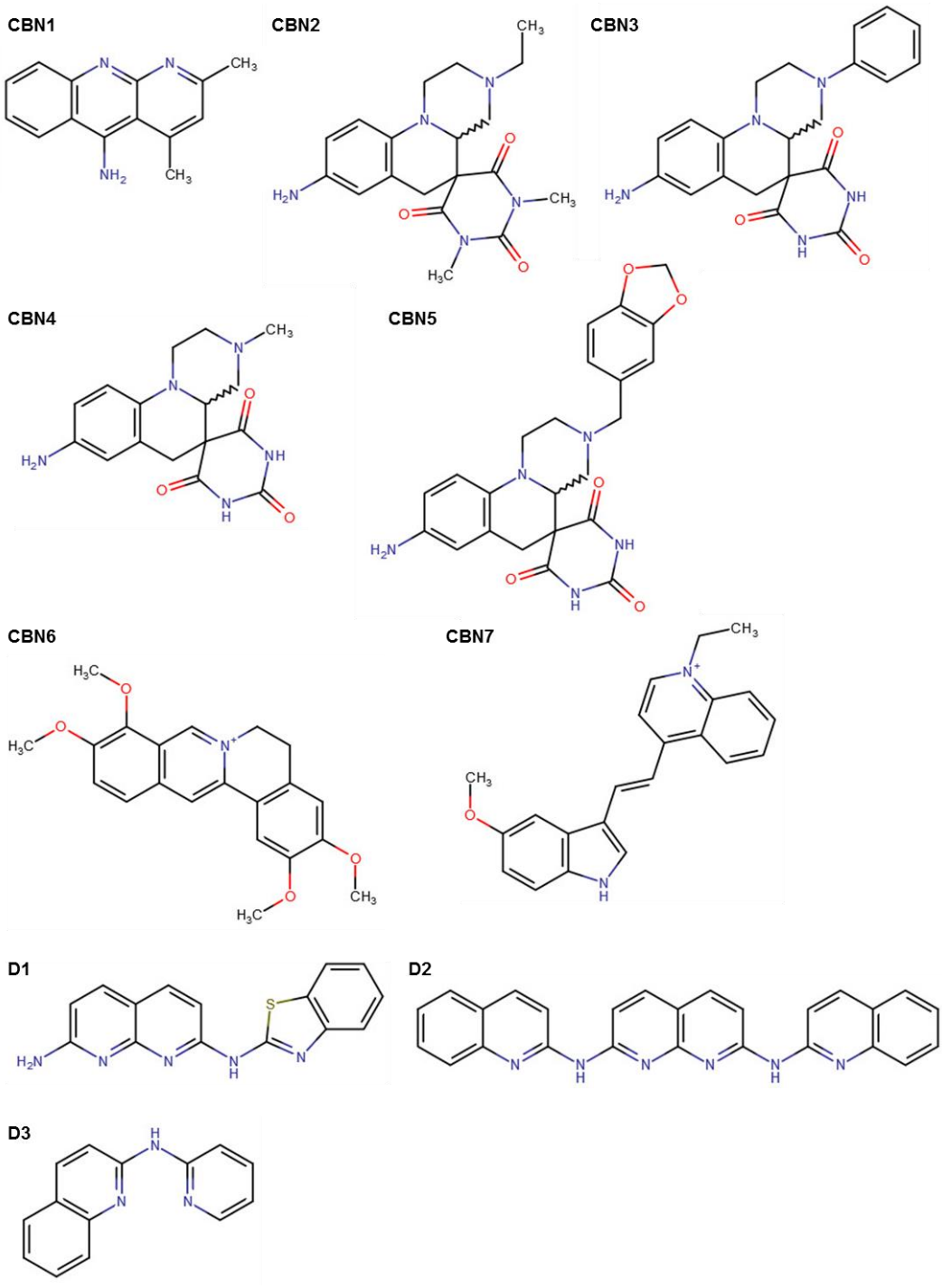


Fig. 3.4.2: Chemical structures of the putative telomeric G-quadruplex binders CBN1-7 and D1-D3 identified in an automated screening of 3800 compounds of the ChemBioNet (CBN) and Daltrozzo (D) libraries.

All 10 identified compounds seem to show a concentration-dependent ability to induce G-quadruplex formation within the FHT oligonucleotide, as indicated by the changes in the FAM/FRET ratio (Fig. 3.4.3). These data suggest that all of

the depicted molecules are able to maximally induce G-quadruplex formation at 100 µM and even show an effect at 1.6 µM. The obtained results imply that the molecules found within the ChemBioNet library (CBN1-7) are more effective compared to the molecules found within the Daltrozzo library (D1-3).

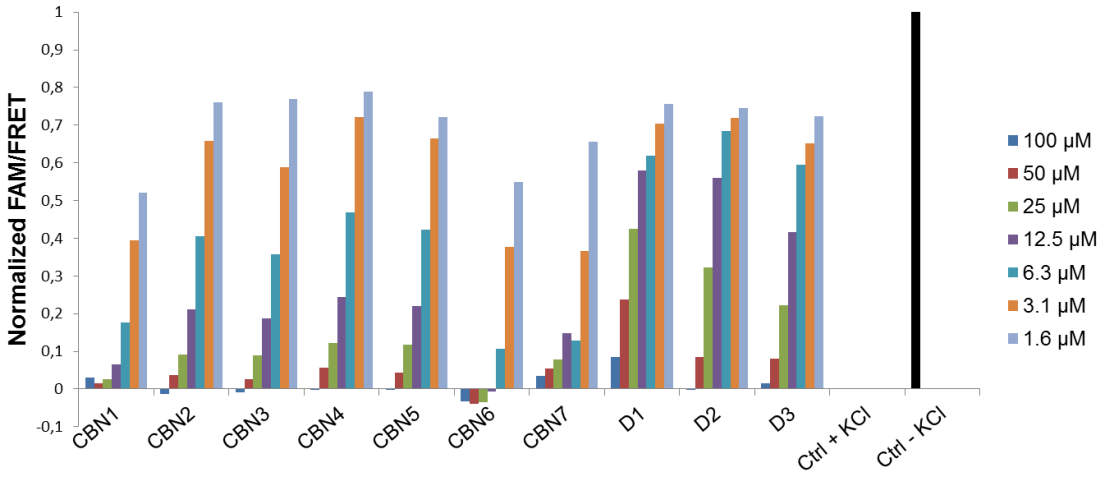


Fig. 3.4.3: Concentration-dependent ability of the 10 identified compounds to induce G-quadruplex formation within the FHT-DNA. Out of a screen using about 3800 compounds of the ChemBioNet (CBN) and Daltrozzo (D) libraries, the 10 most promising compounds (CBN1-7 and D1-3) were further characterized with the focus on their ability to induce G-quadruplex formation within the FHT-DNA in the absence of KCl. Different compound concentrations (from 1.6 µM to 100 µM) and a FHT-DNA concentration of 0.2 µM were tested in 10 mM Tris-HCl (pH 7.5). As controls, the FHT-DNA was measured in the absence (Ctrl – KCl) or in the presence (Ctrl + KCl) of 100 mM KCl.

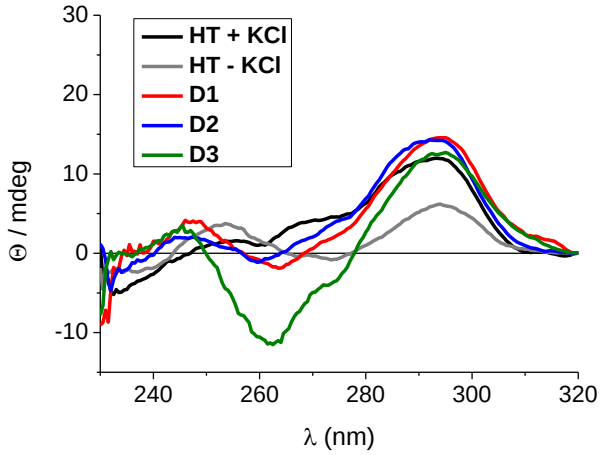
Approximately similar results were obtained for the compounds CBN2, CBN3, CBN4 and CBN5. This outcome is very likely due to their shared chemical substructure (Fig. 3.4.2 and Fig. 3.4.3). CBN6 is the berberine derivative palmatine which already has been described as a G-quadruplex binder [251]. This finding supports the principal functionality of the used screening approach for the identification of telomeric G-quadruplex-interacting molecules. Interestingly, all detected hits contain a quinoline-like and/or a naphthyridine-like substructure which are characteristic features of already known G-quadruplex binders [239,252,253]. Although CBN1, CBN6 and CBN7 showed the best results (Fig.

3.4.3), only the compounds D1, D2 and D3 were further characterized due to their availability (provided by Prof. Dr. E. Daltrozzo, University of Konstanz).

3.4.2 Characterization of the putative G-quadruplex ligands D1-D3 via CD spectroscopy, CD melting and competitive melting experiments

To verify the potential of the identified compounds D1-D3 to induce G-quadruplex formation within the human telomeric sequence HT-DNA and to investigate the induced G-quadruplex topology, CD spectroscopy experiments were performed. Therefore, 3 µM of the oligonucleotide $GGG(TTAGGG)_3$ (HT-DNA) were incubated with 100 µM of the respective compound in 10 mM Tris-HCl buffer (pH 7.5) either in the absence (Fig. 3.4.4 A) or presence of 10 mM KCl (Fig. 3.4.4 B).

A



B

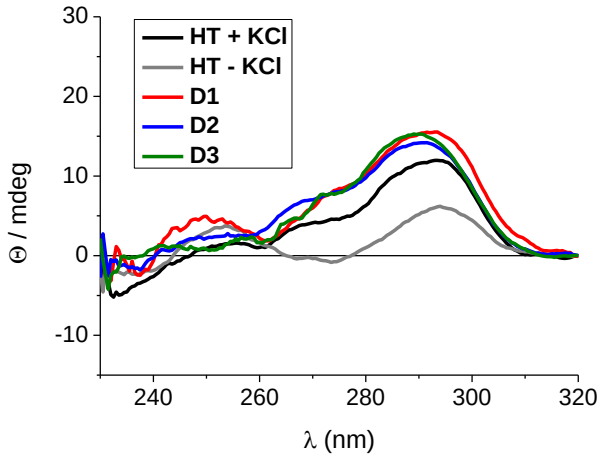


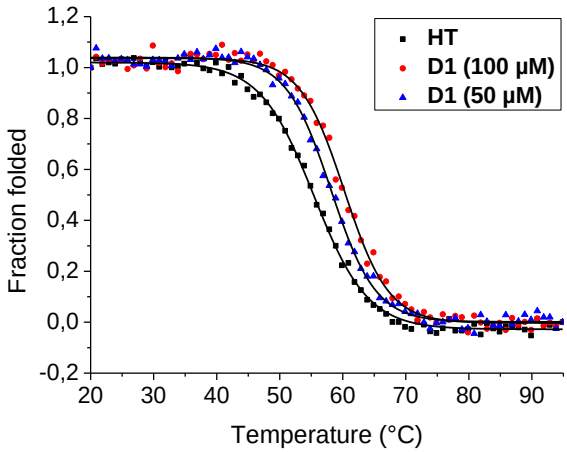
Fig. 3.4.4: Formation and topology of the HT-DNA G-quadruplex induced by the compounds D1-D3. A and B: CD spectra of the HT-DNA (3 µM) together with 100 µM of the respective compound in 10 mM Tris-HCl (pH 7.5) and in the absence (A) or in the presence (B) of 10 mM KCl.

While in the absence of monovalent cations (HT – KCl) the telomeric sequence HT-DNA did not fold into a G-quadruplex structure, a CD spectrum of the HT-DNA in the presence of potassium (HT + KCl) showed a positive peak at around 290 nm and a shoulder at around 265 nm (Fig. 3.4.4 A and B). This spectrum is typical for a mixture of parallel and antiparallel G-quadruplex structures. In the absence of KCl and in the presence of 100 µM of the two compounds D1 and D2, both compounds induced a mixture of parallel and antiparallel G-quadruplex

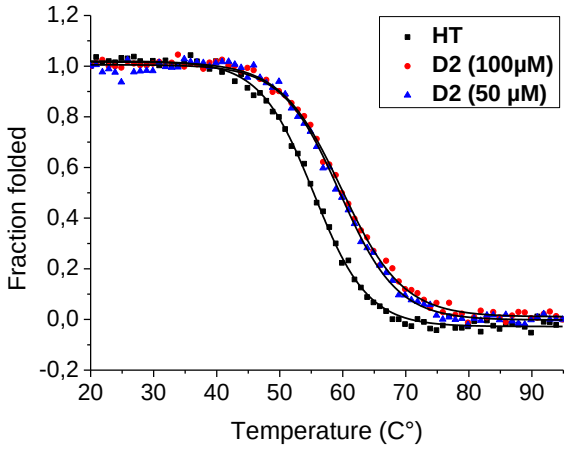
structures within the HT-DNA, indicated by the positive peak at around 290 nm and a shoulder at around 265 nm (Fig. 3.4.4 A). Interestingly, the CD spectrum of the HT-DNA in the absence of KCl, but in the presence of 100 µM of the compound D3 displayed a spectrum with a maximum at around 290 nm and a minimum at around 265 nm which is typical for an antiparallel G-quadruplex structure (Fig. 3.4.4 A). All CD spectra of the HT-DNA in the presence of 10 mM KCl and 100 µM compound concentration showed the formation of a mixture of parallel and antiparallel G-quadruplex structures, as indicated by the positive peak at around 290 nm and a shoulder at around 265 nm (Fig. 3.4.4 B).

To examine whether the three compounds D1, D2 and D3 are not only able to induce G-quadruplex formation within the HT-DNA, but also to stabilize the G-quadruplex structure, CD melting experiments were performed. For this purpose 5 µM HT-DNA were incubated with 100 µM or 50 µM of the respective compound in 10 mM Tris-HCl (pH 7.5) and in the presence of 10 mM KCl. The samples were heated from 15°C to 95°C and the changes of the CD signal were measured at 290 nm wavelength (Fig. 3.4.5 A-C).

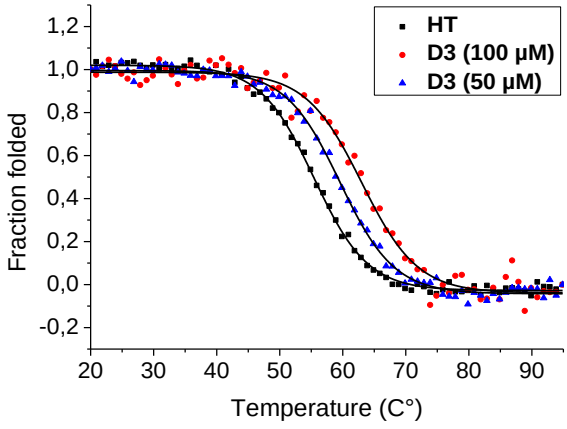
A



B



C



D

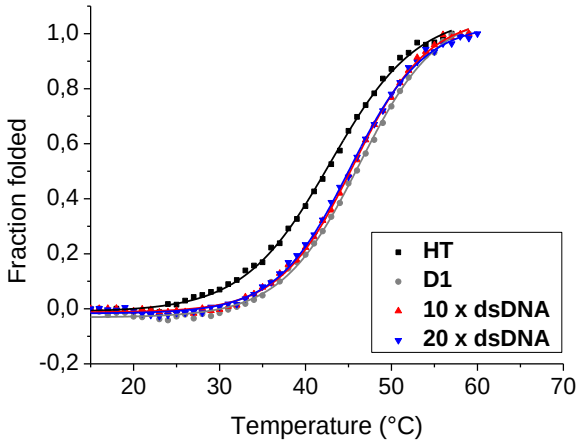
Name	Tm (°C)
HT-DNA	55.4
+ D1 (100 µM)	60.2
+ D1 (50 µM)	58.1
+ D2 (100 µM)	59.5
+ D2 (50 µM)	59.8
+ D3 (100 µM)	63.0
+ D3 (50 µM)	59.6

Fig. 3.4.5: Stabilization of the G-quadruplex structure within the HT-DNA by the three compounds D1, D2 and D3. A-C: CD melting curves of the HT-DNA (5 µM) at 290 nm in the presence of 10 mM KCl and in the absence (black curve) or presence of 100 µM (red curve) and 50 µM (blue curve) of the compound D1 (A), D2 (B) or D3 (C) in 10 mM Tris-HCl (pH 7.5). D: Corresponding melting temperature midpoints (Tm) at 290 nm in the presence of 10 mM KCl.

For all three compounds a moderate stabilization of the G-quadruplex within the HT-DNA was observed (Fig. 3.4.5 A-C). The HT-DNA G-quadruplex showed a Tm value of 55.4°C which was further increased by the addition of 50 µM and 100 µM of D1 to 58.1 °C and 60.2°C, respectively. The addition of 50 µM and 100 µM of D2 led to an increase of the Tm value to 59.5 °C and 59.8°C and the addition of 50 µM and 100 µM of D3 resulted in an increase of the Tm value to 59.6 °C and 63.0°C, respectively.

Since one challenging feature of G-quadruplex binders is their specificity towards G-quadruplex structures over double-stranded DNA, competitive melting experiments in the absence or presence of the double-stranded DNA 5´-CAATCGGATCGAATTCGATCCGATTG-3´ (dsDNA) were performed. Therefore, 0.5 µM of the double-labelled HT-DNA (FHT-DNA) was incubated with 10 µM of the different compounds D1-D3 either in the absence or in the presence of 5 µM or 10 µM of dsDNA (Fig. 3.4.6 A-C).

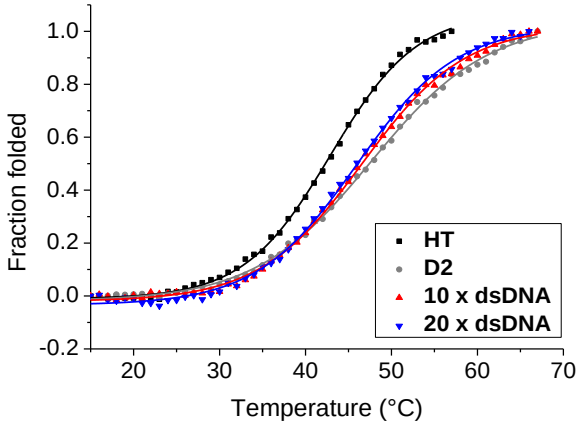
A



D

Name	Tm (°C)
HT-DNA	42.9
+ D1	46.3
+ 10 x dsDNA	45.6
+ 20 x dsDNA	45.2

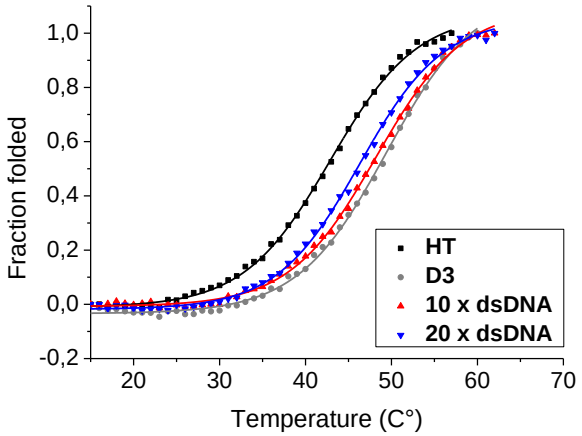
B



E

Name	Tm (°C)
HT-DNA	42.9
+ D2	47.6
+ 10 x dsDNA	46.6
+ 20 x dsDNA	45.9

C



F

Name	Tm (°C)
HT-DNA	42.9
+ D3	49.4
+ 10 x dsDNA	48.2
+ 20 x dsDNA	46.3

Fig. 3.4.6: Specificity of the identified compounds D1, D2 and D3 towards the G-quadruplex structure within the FHT-DNA over dsDNA. A-C: Competitive meltings of the FAM and TAMRA labelled HT-DNA oligonucleotide (HT) in the presence of the compounds D1 (A), D2 (B) or D3 (C) and in the presence of the double-stranded DNA (dsDNA). The HT-DNA (0.5 µM) was incubated with 10 µM of the compounds D1, D2 or D3 in the presence of 10 mM KCl and without or together with 5 µM (10 x dsDNA) or 10 µM (20 x dsDNA) of dsDNA in 10 mM

Tris-HCl (pH 7.5). D-F: Corresponding melting temperature midpoints (Tm) in the presence of 10 mM KCl.

For the G-quadruplex formed by the FAM and TAMRA labelled HT-DNA (HT) a Tm value of 42.9°C was detected in the presence of 10 mM KCl (Fig. 3.4.6 A-F). The addition of 10 µM of the compound D1 increased the Tm value of the HT-DNA G-quadruplex to 46.3°C (Fig. 3.4.6 A and D). After adding 5 µM or 10 µM of the double-stranded DNA (dsDNA), which represents 10 fold and 20 fold higher concentrations compared to the HT-DNA, the observed Tm values of the G-quadruplexes were 45.6°C and 45.2°C, respectively. The addition of 10 µM D2 led to an increased G-quadruplex Tm value of 47.6°C which decreased in the presence of 5 µM and 10 µM dsDNA to 46.6°C and 45.9°C, respectively (Fig. 3.4.6 B and E). The compound D3, which already showed the strongest stabilizing effect in the CD melting experiments, increased the Tm value of the HT-DNA G-quadruplex to 49.4°C, and in the presence of 5 µM or 10 µM of dsDNA Tm values of 48.2°C and 46.3°C were measured (Fig. 3.4.6 C and F).

Discussion:

By using a FRET-based screening approach for small molecules which interact with the telomeric sequence, several promising compounds have been identified in the substance libraries ChemBioNet and Daltrozzo. Although the most promising compounds were found within the ChemBioNet library, due to their availability the three most promising compounds D1, D2 and D3 detected within the Daltrozzo library were further characterized. Using 100 µM compound concentration, the three compounds D1, D2 and D3 exhibited their potential to induce G-quadruplex formation within the HT-DNA (3 µM) in the absence of potassium. Interestingly, while D1 and D2 induced the formation of a mixture of the parallel and antiparallel G-quadruplex topology, the CD spectrum of D3 indicated for the formation of the antiparallel G-quadruplex topology. The CD spectra in the presence of 10 mM KCl showed a mixture of the parallel and antiparallel G-quadruplex conformation for all three compounds. The potential of these molecules to stabilize the HT-DNA G-quadruplex structure was investigated via CD melting experiments using 50 µM and 100 µM of the

corresponding molecule and 3 µM HT-DNA. Thereby, moderate G-quadruplex stabilization properties of all three molecules were observed. The obtained data indicate that D3 has the strongest stabilizing effect on the HT-DNA G-quadruplex. However, it also showed the strongest decrease in the detected Tm value after the molecule concentration was reduced from 100 µM to 50 µM. Therefore, it can be concluded that the stabilizing effect of D3 is dependent on high molecule concentrations. Interestingly, in case of the compound D2, which showed the weakest stabilizing effect on the HT-DNA G-quadruplex at 100 µM molecule concentration, no change in the Tm value for the HT-DNA G-quadruplex was observed after reducing of molecule concentration from 100 µM to 50 µM. This result may indicate that, although the stabilizing effect of D2 seems to be weaker compared to D3 with the used concentrations, D2 might act at lower concentrations. In comparison to compounds identified and characterized by Benz *et al.* (2011), who found under the same conditions a molecule-dependent stabilizing effect of up to 17°C with 50 µM molecule concentration, the stabilizing effect of all three molecules has to be estimated as relatively weak [248]. However, the performed competitive melting studies revealed a high specificity of all three molecules on the human telomeric G-quadruplex structure over duplex DNA, especially in case of the two molecules D1 and D2.

It can be only speculated about the mechanism how the molecules D1, D2 and D3 interact with and stabilize the HT-DNA G-quadruplex structure. All three molecules contain extended aromatic systems with quinoline-like and/or naphthyridine-like substructures. These features have been observed for many other G-quadruplex-interacting molecules as well, since it may allow for stacking interactions with the topmost and/or lowermost G-tetrad of the G-quadruplex structure [205,239,252]. Therefore, it can be assumed that this mechanism is also involved in the G-quadruplex stabilizing properties of the compounds D1-D3 identified in this study. Although the molecules showed only moderate stabilization of the HT-DNA G-quadruplex, their high specificity towards the G-quadruplex structure over double-stranded DNA might advise them as suitable lead structures for further improvements via rational design.

4. Summary

Nucleic acid sequences can adopt a broad variety of different secondary structures. Among them is the so-called guanine-quadruplex (G-quadruplex). This non-canonical structure can be formed by G-rich nucleic acid sequences with a minimum of four interspersed G-tracts. Bioinformatical analysis revealed an enrichment of G-quadruplex forming sequences at functional regions, like the telomeric sequence, the promoter region of many genes, especially proto-oncogenes, but also the 5´ and 3´ untranslated regions (UTRs) of mRNAs. A lot of research has been done to investigate the structure as well as the functions of G-quadruplex forming sequences by *in vitro* and *in vivo* experiments. These studies were enabled to elucidate the polymorphic structure of G-quadruplexes and to indicate for their regulatory role in different cellular processes, such as transcription, translation, replication or telomere maintenance. However, a number of questions regarding the characteristics and functions of G-quadruplexes remain still unanswered. To that end, this thesis tries to provide more insights into different aspects of the structural/physical properties and the functions of G-quadruplexes.

In the first part of this thesis the influence of an RNA G-quadruplex forming sequence located at different positions within the 5´ UTR of an mRNA on the level of translation was analyzed. In a cell-based reporter system, a position-dependent influence of the naturally occurring *MAPK2* RNA G-quadruplex forming sequence inserted into the 5´ UTR of an mRNA on the level of reporter gene expression was shown. By using this system, possible RNA G-quadruplex-interacting proteins and already known DNA G-quadruplex stabilizing small molecules, namely the three bisquinolinium derivatives 360A, Phen-DC_3 and Phen-DC_6, were examined for their potential to influence the stability of RNA G-quadruplexes. Whereas knock-down of the putative RNA G-quadruplex-interacting proteins via siRNAs did not significantly affect the level of reporter gene expression, the potency of 360A, Phen-DC_3 and Phen-DC_6 to bind to and to stabilize the examined RNA G-quadruplex structure within the 5´ UTR could be demonstrated by a reduction of reporter gene expression.

In the second part of this thesis proteins, which are interacting with the putative G-quadruplex forming sequence of the telomeric repeat containing RNA (TERRA), were tried to identify by using an *in vivo* crosslinking approach. TERRA, the transcript of the human telomeric sequence, is suggested to be involved in the regulation of telomere maintenance. Unfortunately, no TERRA-interacting proteins could thus far be detected due to experimental problems. Nonetheless, the method used in this thesis might turn out to be a suitable tool for *in vivo* screening of TERRA-interacting proteins.

The third part of this thesis focused on the question how the number of G-tracts, but also the number and composition of flanking nucleotides of a DNA G-quadruplex influence its thermal stability. Previous studies revealed an increase in thermal stability of RNA G-quadruplexes consisting of sequences with more than four G-tracts (five or six G-tracts), even though only four G-tracts are involved in the formation of an intramolecular G-quadruplex. The analysis of DNA G-quadruplex forming sequences with four, five or six G-tracts via CD spectroscopy and CD melting experiments led to the assumption that the stabilizing effect of additional G-tracts observed for RNA G-quadruplexes is not transferrable one-to-one to DNA G-quadruplexes: Equal thermal stabilities could be observed for DNA G-quadruplexes formed by sequences comprising of four or five G-tracts. However, a decrease in thermal stability was detected for a DNA G-quadruplex formed by a sequence containing six G-tracts. To determine the impact of nucleotides, which are not involved in G-quadruplex formation, on DNA G-quadruplex thermal stability, DNA G-quadruplex forming sequences were analyzed which contain more than four G-tracts with deletions within the G-tracts either at the 5` or at the 3´ end, preventing these G-tracts from participating in G-quadruplex formation. Additionally, thymidines or adenosines were systematically attached at the 5´ or 3´ end of a DNA G-quadruplex forming sequence consisting of four G-tracts. The obtained results emphasized the importance of flanking nucleotides on DNA G-quadruplex thermal stability. They indicated that the first two flanking nucleotides at the 5´ end have a stabilizing effect on the thermal stability of the G-quadruplex structure, while at the 3´ end only the addition of a single adenosine results in a slight stabilization of the G-quadruplex. The addition of further nucleotides at the 5´ as well as the 3´ end

led to the destabilization of the G-quadruplex structure. In general, the stabilizing effect of flanking nucleotides was stronger for nucleotides located at the 5´ end than for nucleotides located at the 3´ end. Furthermore, the stabilizing effect of adenosines was stronger than compared to that of thymidines. This observed stabilizing effect is proposed to be based on stacking interactions of the flanking nucleotides with the external tetrads of the G-quadruplex. Less detailed experiments on an RNA G-quadruplex forming sequence indicated for similar influences of flanking nucleotides on RNA G-quadruplex thermal stability as observed for its DNA counterpart.

In the last part of this thesis new candidates of small molecules, which are capable to bind to and to stabilize the DNA G-quadruplex formed by the human telomeric sequence, were identified within two different substance libraries by using a FRET-based screening approach. Detailed analysis of these small molecules via CD spectroscopy, CD melting and competitive melting experiments further verified their potential as G-quadruplex stabilizing ligands and revealed their specificity for the G-quadruplex structure of the human telomeric sequence over duplex DNA. The G-quadruplex-interacting small molecules identified in this study might be useful as lead structures for the development of highly selective G-quadruplex ligands.

5. Zusammenfassung

Nukleinsäuresequenzen können ein breites Spektrum an verschiedenen Sekundärstrukturen annehmen; zu ihnen zählt unter anderem der sogenannte Guanin-Quadruplex (G-Quadruplex). Diese nicht-kanonische Struktur kann von Guanin-reichen Nukleinsäuresequenzen ausgebildet werden, welche eine Abfolge von mindestens vier *G-Tracts* aufweisen. Bioinformatische Analysen ergaben, dass potentielle G-Quadruplex bildende Sequenzen vermehrt in funktionellen Regionen wie z. B. in der telomerischen Sequenz, in der Promotorregion vieler Gene, besonders von Proto-Onkogenen, aber auch in den 5´ und 3` nicht-translatierten Regionen (UTR) von mRNAs vorkommen. Zur Aufklärung der Struktur sowie der Funktionen von G-Quadruplex bildenden Sequenzen wurden zahlreiche *in vitro* und *in vivo* Studien durchgeführt. Diese Studien zeigten, dass G-Quadruplexe eine polymorphe Struktur aufweisen, und lieferten Hinweise auf die Beteiligung von G-Quadruplexen an der Regulation verschiedenster zellulärer Prozesse, wie beispielsweise der Transkription, der Translation, der Replikation oder der Integrität von Telomeren. Dennoch bleiben viele Fragen bezüglich der strukturellen Eigenschaften und der Funktionen von G-Quadruplexen unbeantwortet. Diese Dissertation versucht daher, weitere detailliertere Einblicke in die Struktur und Funktionen von G-Quadruplexen zu gewähren.

Im ersten Teil dieser Arbeit wurde der Einfluss einer RNA G-Quadruplex bildenden Sequenz, welche jeweils an unterschiedlichen Positionen innerhalb der 5` UTR einer Reporter-mRNA inseriert wurde, auf das Expressionslevel des Reportergens analysiert. In diesem zellbasierten Reportersystem konnte ein positionsabhängiger Einfluss der natürlich vorkommenden RNA G-Quadruplex bildenden Sequenz von *MAPK2* auf die Expression des Reportergens gezeigt werden. Nachfolgend wurde mit Hilfe dieses Systems untersucht, ob sowohl potentielle RNA G-Quadruplex bindende Proteine als auch die bereits bekannten DNA G-Quadruplex bindenden niedermolekularen Substanzen 360A, Phen-DC_3 und Phen-DC_6 vermögen, die Stabilität des *MAPK2* RNA G-Quadruplexes zu beeinflussen. Während der *Knock-down* von möglichen RNA G-Quadruplex bindenden Proteinen via siRNAs keinen eindeutigen Einfluss auf

das Expressionslevel des Reportergens zeigte, konnte anhand einer Reduktion der Reportergenexpression nachgewiesen werden, dass die drei niedermolekularen Substanzen 360A, Phen-DC_3 und Phen-DC_6 die untersuchte RNA G-Quadruplex Struktur innerhalb der 5´ UTR binden und stabilisieren können.

Im zweiten Teil dieser Arbeit wurde mittels einer in *vivo crosslinking* Methode versucht, Proteine zu identifizieren, welche mit der möglichen G-Quadruplex bildenden Sequenz der *telomeric repeat containing* RNA (TERRA) interagieren. Es wird vermutet, dass TERRA, das Transkript der humanen telomerischen Sequenz, eine wesentliche Rolle bei der Organisation der Telomere einnimmt. Unglücklicherweise konnten jedoch auf Grund von experimentellen Problemen keine mit TERRA interagierenden Proteine identifiziert werden. Nichtsdestotrotz könnte die hier verwendete Methode sehr gut dazu geeignet sein, um Proteine, welche *in vivo* mit TERRA interagieren, zu identifizieren.

Der dritte Teil dieser Arbeit widmete sich der Frage, in wieweit sowohl die Anzahl der *G-Tracts*, als auch die Anzahl und Zusammensetzung flankierender Nukleotide eines DNA G-Quadruplexes seine thermische Stabilität beeinflussen. Vorangegangene Studien ergaben, dass ein RNA G-Quadruplex, welcher in seiner Sequenz mehr als vier *G-Tracts* (fünf bzw. sechs *G-Tracts*) enthielt, eine höhere thermische Stabilität aufwies, obgleich angenommen wird, dass lediglich vier der *G-Tracts* an der Ausbildung eines intramolekularen G-Quadruplexes beteiligt sind. Die in dieser Arbeit angestellte Untersuchung von DNA G-Quadruplex Sequenzen mit vier, fünf oder sechs *G-Tracts* mittels CD Spektroskopie und CD Schmelzpunktanalyse zeigte, dass der stabilisierende Effekt zusätzlicher *G-Tracts,* welcher für RNA G-Quadruplexe beobachtet wurde, nicht eins zu eins auf DNA G-Quadruplexe übertragbar ist: DNA G-Quadruplexe, die aus G-Quadruplex bildenden Sequenzen mit vier oder fünf *G-Tracts* gebildet wurden, unterschieden sich in ihrer thermischen Stabilität nicht, jedoch wies die DNA G-Quadruplex bildende Sequenz mit sechs *G-Tracts* eine niedrigere thermische Stabilität des gebildeten G-Quadruplexes auf. Um den Einfluss von Nukleotiden, welche nicht an der Ausbildung des G-Quadruplexes beteiligt sind, auf die thermische Stabilität eines DNA G-Quadruplexes zu

untersuchen, wurde durch gezielte Punktmutationen die Bildung eines DNA G-Quadruplexes in DNA G-Quadruplex bildenden Sequenzen mit mehr als vier *G-Tracts* entweder am 5´ Ende, am 3´ Ende oder in der Mitte der Sequenz erzwungen. Zusätzlich wurden systematisch Adenosin(e) oder Thymidin(e) an das 5´ oder 3´ Ende einer DNA G-Quadruplex Sequenz bestehend aus vier *G-Tracts* angehängt. Die erlangten Ergebnisse unterstreichen die Bedeutung von flankierenden Nukleotiden für die thermische Stabilität eines DNA G-Quadruplexes. Sie zeigten, dass die ersten beiden flankierenden Nukleotide am 5´ Ende zu einer thermischen Stabilisierung des G-Quadruplexes beizutragen scheinen, hingegen am 3´ Ende lediglich das Hinzufügen eines einzelnen Adenosins eine schwache Stabilisierung des G-Quadruplexes zur Folge hat. Das Hinzufügen weiterer Nukleotide sowohl am 5´ Ende, als auch am 3´ Ende führte jedoch zu einer Destabilisierung der G-Quadruplex Struktur. Generell fiel der stabilisierende Effekt für flankierende Nukleotide am 5´ Ende stärker aus als für flankierende Nukleotide am 3´ Ende. Des Weiteren war zu beobachten, dass die thermische Stabilisierung der G-Quadruplex Struktur durch Adenosine stärker als die durch Thymidine war. Die stabilisierenden Effekte der flankierenden Nukleotide sind wahrscheinlich π-π Interaktionen mit den äußeren G-Tetraden des G-Quadruplexes geschuldet. Weniger detaillierte Untersuchungen, die den Einfluss von flankierenden Nukleotiden auf die thermische Stabilität eines RNA G-Quadruplexes analysierten, ließen einen vergleichbaren Einfluss auf die thermische Stabilität des RNA G-Quadruplexes wie bei dem entsprechenden DNA G-Quadruplex vermuten.

Im letzten Teil dieser Arbeit wurden mit Hilfe einer auf FRET-basierenden Methode in zwei verschiedenen Substanzbibliotheken niedermolekulare Substanzen identifiziert, welche mit dem von der humanen telomerischen DNA Sequenz gebildeten G-Quadruplex interagieren und diesen zu stabilisieren vermögen. Die anschließende detaillierte Analyse dieser niedermolekularen Substanzen via CD Spektroskopie, CD Schmelzpunktanalyse und kompetitiver Schmelzpunktanalyse verifizierte ihr Potential als G-Quadruplex stabilisierende Liganden und zeigte ihre Spezifität für die durch die humane telomerische Sequenz gebildete G-Quadruplex Struktur gegenüber doppelsträngiger DNA. Die im Rahmen dieser Arbeit identifizierten DNA G-Quadruplex Liganden

könnten daher als mögliche Leitstrukturen der Entwicklung von hoch spezifischen G-Quadruplex Liganden dienen.

6. Material and Methods

6.1 Material

6.1.1 Chemicals and reagents

All chemicals and reagents were of p.a. or molecular biology grade. Water was drawn from a combined reverse osmosis/ultrapure water system.

Table 6.1: Chemicals and reagents used in this study

Chemicals/Reagents	Supplier
Acrylamide	Roth
Agarose	Roth
Ammonium persulfate (APS)	Merck
Boric acid (H_3BO_3)	Roth
Dichlorodimethylsilane	Sigma-Aldrich
Dimethyl sulfate (DMS)	Sigma-Aldrich
Dimethyl sulfoxide (DMSO)	Fluka
Disodium phosphate (Na_2HPO_4)	Sigma-Aldrich
Dynabeads® MyOne™ Carboxylic Acid	Invitrogen
Ethanol	Riedel de Haën
Ethylene-diamine-tetraacetic acid (EDTA)	Roth
Ethylene-glycol-tetraacetic acid (EGTA)	Roth
Formic acid	Sigma-Aldrich
Glycine	Roth
Hydrazine	Sigma-Aldrich
Hydrochloric acid (HCl)	Merck
HEPES	Sigma-Aldrich
Lithium chloride (LiCl)	Sigma-Aldrich
2-Mercaptoethanol	Roth
2-(N-morpholino)ethanesulfonic acid (MES)	Sigma-Aldrich
Magnesium chloride ($MgCl_2$)	Merck
Magnesium sulfate ($MgSO_4$)	Sigma-Aldrich

Monopotassium phosphate (KH_2PO_4)	Sigma-Aldrich
N-(3-Dimethylaminopropyl)-N′-ethylcarbodiimide (EDC)	Sigma-Aldrich
Phenylmethylsulfonyl fluoride (PMSF)	Sigma-Aldrich
Piperidine	Sigma-Aldrich
Potassium chloride (KCl)	Sigma-Aldrich
Sodium chloride (NaCl)	VWR
Sodium dodecyl sulfate (SDS)	Sigma-Aldrich
Sodium fluoride (NaF)	Sigma-Aldrich
Sodium vanadate (Na_3VO_4)	Sigma-Aldrich
Sodium acetate	Merck
Tetramethylethylenediamine (TEMED)	Roth
Tris(hydroxymethyl)aminomethane (Tris)	Sigma-Aldrich
Tryptone	MP Biomedicals
Tween 20	Roth
Urea	Roth
Yeast extract	MP Biomedicals

6.1.2 Buffers, solutions and media

Table 6.2: Buffers and solutions used in this study

Buffer and solutions	Components
Agarose loading buffer	Glycerol (30%) Bromphenolblue Xylencyanol
Blocking buffer	Non-fat milk powder (5%) in PBST buffer
Denaturating PAGE loading dye	Formamide (80%) EDTA (2 mM)
DMS stop solution	2-Mercaptoethanol (1 M) Sodium acetate (0.5 M) carrier DNA (50 µg/µl)

DNA sequencing stop solution	Sodium acetate (0.5 M) carrier DNA (50 µg/µl) pH 6.0
EDC buffer	EDC (1.25 M) in MES buffer
Laemmli loading dye (2x)	Tris-HCl (100 mM; pH 6.8) SDS (4%) Glycerol (20%) DTT (0.2 M) Bromphenolblue
MES buffer	MES (100 mM) pH 4.8
Nuclei lysis buffer (NLB)	HEPES (40 mM; pH 7.6) LiCl (200 mM) EDTA (80 mM) EGTA (2 mM) PMSF (1 mM) Protease inhibitor cocktail (1x) RNase inhibitor (1x)
Nuclei preparation buffer (NPB)	HEPES (20 mM; pH 7.9) $MgCl_2$ (5 mM) KCl (20 mM) PMSF (0.2 mM) DTT (1 mM) Protease inhibitor cocktail (1x)
Nuclei storage buffer (NSB)	HEPES (20 mM; pH 7.5) $MgCl_2$ (1.5 mM) KCl (20 mM) Glycerol (50%) PMSF (0.2 mM) EGTA (5 mM) DTT (1 mM) Protease inhibitor cocktail (1x)
PBS buffer	NaCl (137 mM) KCl (2.7 mM) Na_2HPO_4 (4.3 mM) KH_2PO_4 (1.47 mM) pH 7.4

PBST buffer	Tween 20 (20 g/L) in PBS
SDS running buffer (10x)	Tris base (30.3 g/L) Glycine (144 g/L) SDS (1g/L) pH 8.3
TBE buffer	Tris Base (89 mM) Boric acid (89 mM) EDTA pH 8.0 (2 mM)
TE buffer	Tris-HCl (10 mM; pH 8) EDTA (1 mM)
Transfer buffer	Tris-HCl (25 mM; pH 8.3) Glycine (192 mM) Methanol (20%)
Tris buffer	Tris-HCl (10 mM) pH 7.5
TT buffer	Tris-HCl (250 mM; pH 8) Tween 20 (0.1%)

Table 6.3: Media used in this study

Medium	Components
LB-Agar	Tryptone (10 g/L) Yeast extract (5 g/L) NaCl (10 g/L) Agar (20 g/L) pH 9.0
LB medium	Tryptone (10 g/L) Yeast extract (5 g/L) NaCl (10 g/L) pH 7.0
SOC medium	Tryptone (2% (w/v)) Yeast extract (0.5% (w/v)) Glucose (20 mM) NaCl (10 mM) KCl (2.5 mM) $MgCl_2$ (10 mM) $MgSO_4$ (10 mM)

6.1.3 Bacterial strains

E. coli XL10-Gold:

endA1 glnV44 recA1 thi-1 gyrA96 relA1 lac Hte Δ(mcrA)183 Δ(mcrCB-hsdSMR-mrr)173 tet^R F'[proAB $lacI^qZ\Delta M15$ Tn10(Tet^R Amy Cm_R)] (Invitrogen)

6.1.4 Mammalian cell lines

HeLa 229 cells:

human cervix adenocarcinoma cells [254]

HeLa S3 cells:

human cervix adenocarcinoma cells [254,255]

6.1.5 Cell culture material

Table 6.4: Cell culture material used in this study

Reagents	Supplier
DMEM	Gibco
Fetal calf serum (FCS)	Gibco
HiPerFect	Qiagen
Phosphate buffered saline (PBS)	Sigma-Aldrich
Penicillin/Streptomycin	Biochrom
Trypsine	Gibco
TurboFect	Thermo scientific

6.1.6 Oligonucleotides and primers

All oligonucleotides and primers were synthesized and HPLC purified by Metabion GmbH.

Table 6.5: Oligonucleotides used for the investigation of G-quadruplex formation *in vitro*

Name	Sequence 5´→ 3´
4G3T	GGGTGGGTGGGTGGG
5G3T	GGGTGGGTGGGTGGGTGGG
6G3T	GGGTGGGTGGGTGGGTGGGTGGG
4G3T(A)	GGGTGGGTGGGTGGGA
4G3T(2A)	GGGTGGGTGGGTGGGAA
4G3T(3A)	GGGTGGGTGGGTGGGAAA
4G3T(4A)	GGGTGGGTGGGTGGGAAAA
4G3T(T)	GGGTGGGTGGGTGGGT
4G3T(2T)	GGGTGGGTGGGTGGGTT
4G3T(3T)	GGGTGGGTGGGTGGGTTT
4G3T(4T)	GGGTGGGTGGGTGGGTTTT
(A)4G3T	AGGGTGGGTGGGTGGG
(2A)4G3T	AAGGGTGGGTGGGTGGG
(3A)4G3T	AAAGGGTGGGTGGGTGGG
(4A)4G3T	AAAAGGGTGGGTGGGTGGG
(T)4G3T	TGGGTGGGTGGGTGGG
(2T)4G3T	TTGGGTGGGTGGGTGGG
(3T)4G3T	TTTGGGTGGGTGGGTGGG
(4T)4G3T	TTTTGGGTGGGTGGGTGGG
T10-6G3T-T10	TTTTTTTTTTGGGTGGGTGGGTGGGTGGGTGGGTTTTTTTTTT
4G3U	GGGUGGGUGGGUGGG
(A)4G3U	AGGGUGGGUGGGUGGG
4G3U(A)	GGGUGGGUGGGUGGGA

Table 6.6: Oligonucleotides used in the screen for small molecules interacting with the human telomeric sequence

Name	Sequence 5´→ 3´
dsDNA	CAATCGGATCGAATTCGATCCGATTG
FHT-DNA	FAM-GGGTTAGGGTTAGGGTTAGGG-TAMRA
HT-DNA	GGGTTAGGGTTAGGGTTAGGG

Table 6.7: Primers for the insertion of the *MAPK2* G-quadruplex forming sequence or control sequence into pcDNA5/FRT/TO-eGFP

Name	Sequence 5´→ 3´
MAPK2 GQ 135 nt fw	AGGGAGGGAGGGCTTAAGATGGTGAGCAAGG
MAPK2 Ctrl 135 nt fw	AGTGAGTGAGTGCTTAAGATGGTGAGCAAGG
MAPK2 135 nt rev	CCCTTTAAACGCTAGAGTCCG
MAPK2 GQ 28 nt fw	AGGGAGGGAGGGAGCTCGTTTAGTGAACC
MAPK2 Ctrl 28 nt fw	AGTGAGTGAGTGAGCTCGTTTAGTGAACC
MAPK2 28 nt rev	CCCGTCGACGATCTCTATCAC
MAPK2 GQ 38 nt fw	AGGGAGGGAGGGAGTGAACCGTCAGATCG
MAPK2 Ctrl 38 nt fw	AGTGAGTGAGTGAGTGAACCGTCAGATCG
MAPK2 38 nt rev	CCCAAACGAGCTCGTCGAC
MAPK2 GQ 48 nt fw	AGGGAGGGAGGGCAGATCGCCTGGAGAC
MAPK2 Ctrl 48 nt fw	AGTGAGTGAGTGCAGATCGCCTGGAGAC
MAPK2 48 nt rev	CCCACGGTTCACTAAACGAG
Sequencing primer	TATGGGACTTTCCTACTTGG

6.1.7 siRNAs and hybridization linker DNA

All siRNAs were purchased from Qiagen and the hybridization linker DNAs were ordered from IBA GmbH.

Table 6.8: siRNAs used for the investigation of possible RNA G-quadruplex-interacting proteins

Name	Target sequence 5´→ 3´
Hs_DHX36_4	CTGGCTTATCTATCACCTAAA
Hs_DHX36_6	TCCAGTGTTAGTCATATCGTA
Hs_DHX36_7	CCCGTTGAGCCACATATTGGA
Hs_DHX36_8	TTCGATCTGACTTGAAAGTAA
Hs_DDX3X_5	CACAGGTGTGATACAACTTAA
Hs_DDX3X_8	AACGAGAGAGTTGGCAGTACA
Hs_DHX29_6	CATGACCAATACACTCTTATA
Hs_DHX29_10	CTGCTCGAAGTTGGACTGGAA
Hs_EIF4A1_12	CTAATCCATTTCCCTAACCTA
Hs_EIF4A1_13	AAGAAGGAACCGTGAACATTT
Hs_EIF4A2_2	CGCGGGATTGATGTGCAACAA
Hs_EIF4A2_7	CACCCGGGAGAGTGTTTGATA
Hs_EIF4A3_2	AAAGAGCAGATTTACGATGTA
Hs_EIF4A3_5	CCGCATCTTGGTGAAACGTGA
Ctrl_AllStars_1	proprietary – not disclosed

Table 6.9: Hybridization linker DNA used for the purification of TERRA

Name	Sequence 5´→ 3´
TERRA-specific	H_2N-(C12)-$(CCCTAA)_8$
randomized	H_2N-(C12)-CTCACATCACCATCACACACCATCCTCACAACACCTACTCACCATACC

6.1.8 Nucleotides and radiochemicals

Table 6.10: Nucleotides and radiochemicals used in this study

Reagents	Supplier
dNTPs	Fermentas
[γ^{32}-P]-ATP	Hartmann Analytic
Salmon sperm DNA sodium salt (Carrier DNA)	Roth

6.1.9 G-quadruplex-interacting molecules

Table 6.11: G-quadruplex-interacting molecules used in this study

G-quadruplex-interacting molecule	Supplier
360A	Corinne Guetta, Institute Curie-UMR176
Phen-DC_3	Corinne Guetta, Institute Curie-UMR176
Phen-DC_6	Corinne Guetta, Institute Curie-UMR176

6.1.10 Substance libraries

Table 6.12: Substance libraries used in this study

Library	Source
ChemBioNet (CBN) (P1001 – P1010)	MPI Berlin
Daltrozzo (D)	Prof. Dr. E. Daltrozzo, University of Konstanz

6.1.11 Antibodies

Table 6.13: Antibodies used in this study

Antibody	Species & Type	Supplier	Dillution
anti-DHX36	rabbit/polyclonal	Thermo scientific	1:5000
HRP-coupled anti-rabbit	goat	Thermo scientific	1:2000

6.1.12 Enzymes and kits

Table 6.14: Enzymes used in this study

Enzymes	Supplier
DpnI	NEB
HindIII	NEB
Phusion HotStart II DNA-polymerase	Thermo Scientific
Protease inhibitor cocktail (cOMPLETE Mini)	Roche
Quick T4 DNA ligase	NEB
RNase inhibitor cocktail (RiboLock)	Thermo Scientific
S1 Nuclease	Promega
T4 Polynucleotidekinase (PNK)	NEB

Table 6.15: Kits used in this study

Kits	Supplier
Quick Ligation™ Kit	NEB
Zyppy™ Plasmid Miniprep Kit	Zymo Research
DNA Clean & Concentrator ™ Kit	Zymo Research
Zymoclean™ Gel DNA Recovery Kit	Zymo Research

6.1.13 Plasmids

The two plasmids pcDNA5/FRT/TO-eGFP and pcDNA5/FRT/TO-mCherry were kindly provided by Prof. Dr. Thomas U. Mayer, University of Konstanz (cf. chapter 9).

Table 6.16: Plasmids used and constructed in this study

Name	Insert	Vector	Primers
pcDNA5/FRT/TO-eGFP	*eGFP*	pcDNA5/FRT/TO	-
pcDNA5/FRT/TO-mCherry	*mCherry*	pcDNA5/FRT/TO	-

pcDNA5-MAPK2-135nt-eGFP	*MAPK2*-GQ*	pcDNA5/FRT/TO-eGFP	MAPK2 GQ 135nt fw MAPK2 135nt rev
pcDNA5-Ctrl-135nt-eGFP	Ctrl**	pcDNA5/FRT/TO-eGFP	MAPK2 Ctrl 135nt fw MAPK2 135nt rev
pcDNA5-MAPK2-28nt-eGFP	*MAPK2*-GQ*	pcDNA5/FRT/TO-eGFP	MAPK2 GQ 28nt fw MAPK2 28nt rev
pcDNA5-Ctrl-28nt-eGFP	Ctrl**	pcDNA5/FRT/TO-eGFP	MAPK2 Ctrl 28nt fw MAPK2 28nt rev
pcDNA5-MAPK2-38nt-eGFP	*MAPK2*-GQ*	pcDNA5/FRT/TO-eGFP	MAPK2 GQ 38nt fw MAPK2 38nt rev
pcDNA5-Ctrl-38nt-eGFP	Ctrl**	pcDNA5/FRT/TO-eGFP	MAPK2 Ctrl 38nt fw MAPK2 38nt rev
pcDNA5-MAPK2-48nt-eGFP	*MAPK2*-GQ*	pcDNA5/FRT/TO-eGFP	MAPK2 GQ 48nt fw MAPK2 48nt rev
pcDNA5-Ctrl-48nt-eGFP	Ctrl**	pcDNA5/FRT/TO-eGFP	MAPK2 Ctrl 48nt fw MAPK2 48nt rev

**MAPK2*-GQ: GGGAGGGAGGGAGGG
**Ctrl: GGGAGTGAGTGAGGG

6.1.14 DNA and protein markers

Table 6.17: DNA and protein markers used in this study

Name	Supplier
GeneRuler^{TM} 1kb DNA ladder	Thermo scientific
PageRuler^{TM} Protein ladder (10-180 KDa)	Thermo scientific

6.1.15 Equipment

Table 6.18: Equipment used in this study

Equipment	Supplier; Model designation
Circular dichroism spectrometer	Jasco; J-815
UV-Crosslinker	Vilber Lourmat; Bio-Link BLX
Electroporator	Eppendorf; Elektroporator 2510

Fluorescence plate reader	Tecan; infinite M200 and F500
Gel documentation device	Biometra
Gel drier	Bio-Rad
Incubator	Heraeus
Imaging system	Fujifilm; LAS-3000
NanoQuant plate	Tecan
PAGE gadget	Biorad; SequiGel GT
PCR cycler	Biometra; Thermocycler Gradient
Real-time PCR system	Bio-Rad; Chromo4™
Robot system	Tecan; Freedom EVO® series
Steril hood	HERA safe
UV light table	Biometra
X-ray screen cassette	Fujifilm; BAS Cassette 2 2048
X-ray screen reader	Bio-Rad
X-ray screens	Fuji

6.1.16 Software

Table 6.19: Software used in this study

Name	Producer
Clone Manager	Sci-Ed
i-Control 1.4.9	Tecan
IDT Oligo Analyzer	IDT (http://eu.idtdna.com/calc/analyzer)
MolPort Structure Search	MolPort (https://www.molport.com/shop/index)
MS Office 2010 Excel	Microsoft
MS Office 2010 PowerPoint	Microsoft
MS Office 2010 Word	Microsoft
Origin 8.1	OriginLab

6.2 Methods

6.2.1 Cloning procedure

6.2.1.1 Whole plasmid PCR

For the insertion of the *MAPK2* G-quadruplex forming sequence or the corresponding control sequence at different positions within the 5´ UTR, whole plasmid PCRs were carried out with the plasmid pcDNA5/FRT/TO-eGFP.

All PCRs were performed with the Phusion Hot Start II Polymerase using a BioMetra thermal cycler. The composition of the reaction mixture and the PCR program are depicted in Table 6.20 and Table 6.21. The respective primers are listed in Table 6.7.

Table 6.20: PCR reaction mixture

Reagent	Volume (µl)	Final concentration
HF buffer (5x)	30	1x
dNTPs (2 mM)	15	200 µM
Primer forward (10 µM)	9	0.6 µM
Primer reverse (10 µM)	9	0.6 µM
Template (20 ng/µl)	1.5	0.2 ng/µl
DMSO (100%)	4.5	3%
Phusion Hot Start II Polymerase (2 U/µl)	1.5	0.02 U/µl
Water	79.5	

Table 6.21: PCR program

PCR cycle	Temperature (°C)	Time (s)
Lid preheating	98	-
1. Initial denaturation	98	120
2. Denaturation*	98	50
3. Annealing*	58	60
4. Elongation*	72	120
5. Final elongation	72	300
cooling	4	-

*Step 2. to step 4. were repeated for 30 cycles.

After concentrating the PCR amplified DNA via ethanol precipitation (cf. chapter 6.2.1.2), the template DNA was digested by the restriction enzyme DpnI for 90 min at 37°C, followed by inactivation of DpnI for 20 min at 65°C. The digestion products were than separated via agarose gel electrophoresis (cf. chapter 6.2.1.3) and extracted from the gel (cf. chapter 6.2.1.4). Following ligation (cf. chapter 6.2.1.5), the plasmid DNA was transformed into *E. coli* XL 10-Gold cells (cf. chapter 6.2.1.6).

6.2.1.2 Ethanol precipitation

To precipitate DNA, three volumes of ethanol (100%) and one tenth volume of sodium acetate (3 M, pH 5.4) were added to the DNA solution, followed by incubation for at least 20 min at -80°C. Afterwards, the solution was centrifuged for 30 min at 14000 x g. The supernatant was discarded and the remaining DNA pellet was dissolved in the required volume of water or piperidine.

6.2.1.3 Agarose gel electrophoresis

Agarose was boiled in 0.5x TBE buffer until agarose was completely dissolved. Afterwards, the agarose solution was poured into a gel tray and cooled down to RT. Next, the agarose gel was submerged in 0.5x TBE buffer in a gel electrophoresis chamber. Following loading of the respective samples (mixed with agarose loading buffer), the gel run at 5-10 V/cm. Staining of the gel was

performed in 0.01% (w/v) ethidiumbromide solution, followed by analysis at 260 nm wavelength in a gel documentation system (Biometra).

6.2.1.4 Purification of DNA from agarose gels

DNA was excised from the agarose gel under UV light and purified using the Zymoclean™ Gel DNA Recovery Kit (Zymo Research).

6.2.1.5 Ligation

Ligation reactions were carried out using the Quick Ligation™ Kit (NEB). The obtained ligation products were purified via the DNA Clean & Concentrator™ Kit (Zymo Research).

6.2.1.6 Transformation of DNA into electrocompetent E.coli cells

To transform the respective plasmid into *E. coli* cells, 80 µl of electrocompetent *E.coli* cells were mixed with ~50 ng of plasmid DNA and transferred into a pre-chilled electroporation cuvette. Following electroporation at 1800 V for 5 ms, the cells were incubated in 1 ml SOC medium at 37°C for 1 h and plated on LB plates containing the appropriate antibiotic.

6.2.1.7 Preparation of plasmid DNA

Preparation of plasmid DNA from *E. coli* over-night cultures was performed using the Zyppy™ Plasmid Miniprep Kit (Zymo Research).

6.2.1.8 DNA sequencing

DNA sequencing was performed by GATC Biotech AG.

6.2.2 Measurement of DNA and RNA concentration

To determine DNA or RNA concentration, the specific absorption of DNA or RNA was measured at 260 nm using a Tecan Infinite M200 NanoQuant Plate. As reference, water or the respective buffer was used. After determination of the extinction coefficient and the molecular weight via the online program “IDT Oligo Analyzer”, the DNA or RNA concentration was calculated according to the Lambert-Beer’s law.

6.2.3 Maintenance and cultivation of bacterial cultures

Bacteria were cultivated either in LB media or on LB plates containing the respective antibiotic at 37°C. Glycerol stocks of bacteria were prepared in a Cryovial by mixing sterile glycerol (50%) and overnight cultures in a 1:1 ratio and stored at -80°C.

6.2.4 Maintenance and cultivation of mammalian cell lines

HeLa 229 and HeLa S3 cells were maintained in a 5% CO_2-humidified incubator at 37°C. For cultivation of the cells, 75 cm^2 cell culture bottles with Dulbecco's Modified Eagle Medium (DMEM) were used. Medium was supplemented with 10% fetal calf serum (FCS), 100 U/ml penicillin and 100 µg/ml streptomycin. For freezing mammalian cells in liquid nitrogen, cells grown to ~80% cell confluency were trypsinized and harvested by centrifugation. After resuspension of the cell pellet (from about 10^7 cells) in 1 ml FCS containing 10% DMSO (v/v), the cell solution was transferred into a Cryovial and kept in a freezing container at -80°C overnight. Afterwards, the Cryovial was finally stored in liquid nitrogen.

6.2.5 Transient transfection

6.2.5.1 Transfection of HeLa 229 cells with plasmids

One day before transfection with plasmids, HeLa 229 cells were trypsinized, washed once in 1x PBS, diluted in fresh medium, transferred to a 96 well plate (20.000 cells/well in 100 µl medium) and incubated at 37°C for 24 h. On the next day, 50-200 ng of the respective plasmid containing *eGFP* as reporter gene and 50-200 ng of the respective plasmid containing *mCherry* as reporter gene were mixed together with 0.3 µl TurboFect (Thermo scientific) per well in 200 µl serum-free medium and incubated for 30 min at RT. The medium was removed from the cells and the transfection mixture was pipetted drop-wise on the cells. After incubation at 37°C for 5 h, transfection medium was replaced by 200 µl medium containing 10% FCS and 1% Pen/Strep for further 19 h. Next, cells (grown to about 70-80% confluency) were washed once in 1x PBS, followed by trypsinization (25 µl/well) for 5 min at 37°C. Afterwards, cells were transferred to an opaque 96-well plate and the level of eGFP and mCherry

expression was measured with a Tecan infinite M200 fluorescence plate reader (eGFP: excitation wavelength = 488 nm and emission wavelength = 535 nm; mCherry: excitation wavelength = 587 nm and emission wavelength = 620 nm).

6.2.5.2 Transfection of HeLa 229 cells with siRNAs

One day before transfection with siRNAs, cells were trypsinized, washed once in 1x PBS, diluted in fresh medium, transferred to a 96 well plate (10.000 cells/well in 100 µl medium) and incubated at 37°C. After 24 h, 0.2 µl of the respective siRNAs (stock 10 µM) were mixed together with 1 µl HiPerFect (Qiagen), filled up with serum-free medium to 200 µl and incubated for 30 min at RT. The transfection complexes were added dropwise to the cells and incubated for 6 h. Afterwards, the medium was replaced against medium containing 10% FCS. 18 h later, transfection of cells with plasmids was performed (cf. chapter 6.2.5.1).

For the knock-down of DHX36 cells were transfected with a combination of four different siRNAs, while for knocking-down DHX29, DDX3X, eIF4a1, eIF4a2 and eIF4a3 cells were transfected with a combination of two different siRNAs (cf. chapter 6.1.7). Knock-down efficiency of combined siRNAs against DHX36 was validated by immunoblotting (cf. chapter 6.2.9).

6.2.6 Lysis of mammalian cells

Lysis of HeLa 229 cells for immunoblotting experiments was performed by using the Passive Lysis 5x Buffer (Promega) in accordance to the manufacture´s protocol. Lysis of HeLa S3 cells for screening of TERRA-interacting molecules is described in chapter 6.2.12.3.

6.2.7 Protein quantification via Bradford assay

For quantification of protein concentration, a Bradford assay was performed using the Bio-Rad Protein Assay (Bio-Rad) in accordance to the manufacture´s protocol. The absorption was measured at 595 nm in the Tecan infinite M200 fluorescence-reader.

6.2.8 SDS-polyacrylamid gel electrophoresis (SDS-PAGE)

SDS-polyacrylamide gel electrophoresis (SDS-PAGE) was performed according to the protocol of Laemmli [256]. Protein samples were mixed with Laemmli loading dye and heated at 98°C for 5 min. After loading of the samples, the electrophoresis was carried out at a constant current of 120 V in a Bio-Rad MINI protean gel apparatus. Gels were analyzed either by immunoblotting or silver staining.

Table 6.22: Reagent composition for 12% SDS-PAGE

Reagent	Volume (µl)	Final concentration
	Stacking gel	
Water	4295	
Rotiphorese 40 (37.5:1)	500	4%
Tris-HCl (0.5 M, pH 6.8)	125	
10% SDS	50	
10% APS	25	
TEMED	5	
	Separating gel	
Water	4345	
Rotiphorese 40 (37.5:1)	3000	12%
Tris-HCl (1.5 M, pH 8.8)	2500	
10% SDS	100	
10% APS	50	
TEMED	5	

6.2.9 Immunoblotting

After SDS-PAGE (cf. chapter 6.2.8) proteins were transferred onto a PVDF membrane (0.45 µm; Millipore). To that end, the wet transfer apparatus Bio-Rad Mini Trans-Blot Cell was used (Bio-Rad). The transfer procedure was performed according to the manufacture´s protocol. Afterwards, non-specific binding of the antibody was prevented by incubation of the membrane in TBST containing 5%

non-fat milk overnight at 4°C. Following washing of the membrane three times for 10 min in TBST, the membrane was incubated with the primary antibody for 1 h at RT. Next, the membrane was again washed for three times for 10 min in TBST and subsequently the secondary antibody was applied for 1 h at RT. Afterwards, the membrane was again washed for three times for 10 min in TBST and the blot was developed using the Pierce™ ECL Western Blotting Substrate (Thermo scientific) in the imaging system LAS-3000 (Fujifilm).

6.2.10 CD spectroscopy

6.2.10.1 Structural analysis via CD spectroscopy

Oligonucleotides were purchased from Metabion in HPLC grade. The following reaction mixture was used for CD spectroscopy measurements:

Table 6.23: Reaction mixture for CD spectroscopy measurements

Reagent	Volume (µl)	Final concentration
Tris-HCl (100 mM, pH 7.5)	60	10 mM
Oligonucleotide (100 µM)	30 (18)	5 µM (3 µM)
KCl (1 M)	0 - 6	0 - 10 mM
Ligand (10 mM)	0 - 6	0 - 100 µM
H_2O	filled up to 600	

For structure formation the reaction mixture was heated up to 95°C for 5 min, followed by slowly cooling to RT over 16 h. Dependent on the experiment different concentrations of KCl and G-quadruplex-interacting molecule (ligand) were added to the reaction mixture. CD spectra were recorded from 220 nm to 320 nm in pentaplets in capped-cuvettes with 1 cm path length with a Jasco 815 spectrometer using a resolution of 0.5 nm, a band width of 1 nm, a speed of 500 nm/min and a response time of 0.5 s at 25°C. The CD spectrum of each sample was accumulated five times and averaged. The spectrum of the buffer was subtracted from the spectra of the samples and the value at 320 nm wavelength was set to zero.

6.2.10.2 Thermal stability analysis via CD spectroscopy

For melting experiments, the above described reaction batch (cf. chapter 6.2.10.1) was heated from 20°C to 95°C with a heating rate of 0.4°C/min. The CD signal at 265 nm (for parallel G-quadruplexes) or at 290 nm (for antiparallel G-quadruplexes) was recorded in 1°C steps. Results were analyzed using Origin 8.1 software and the Tm values were calculated from the corresponding melting curves at the half-maximum decrease of the CD signal by curve fitting via the formula: y = A2 + (A1-A2)/(1 + exp((x-x0)/dx)).

6.2.11 Fluorescence spectroscopy

6.2.11.1 Screening of substance libraries for G-quadruplex-interacting compounds

The screening for G-quadruplex-interacting compounds within substance libraries was performed using a FRET-based assay, developed by Dr. Armin Benz [248]. Therefore, black 384 well plates with a flat bottom (Greiner) were filled with the master mix depicted in Table 6.24.

Table 6.24: Composition of the master mix used for screening substance libraries for G-quadruplex-interacting compounds

Reagent	Volume (µl)	Final concentration
Tris-HCl (100 mM, pH 7.5)	1.5	10 mM
FHT oligonucleotide (10 µM)	0.3	0.2 µM
H_2O	13.2	

The compounds were transferred full-automatically via a pin tool device by a pipetting robot of the Freedom Evo series (Tecan) from the substance library plates to the acceptor plates containing the master mix. Thereby, 73 nl (± 2.7) of 10 mM of compound stock solution dissolved in DMSO were added twice to the master mix in the acceptor plates, leading to a final volume of 146 nl and a final compound concentration of ~ 95 µM. After each compound transfer, the pin tool was washed in DMSO, then in a supersonic bath containing Milli Q H_2O with 10% DMSO, followed by a methanol bath. Afterwards, the pin tool was dried by airflow. After adding the compounds to the master mix, the solution was briefly

mixed and incubated for 20 min at RT. The following fluorescence read out was performed by a Tecan Infinite F500 plate reader. To that end, the emission of FAM at 535 nm and TAMRA at 590 nm was measured after excitation of FAM at 485 nm wavelength. As controls, the master mix containing the FHT oligonucleotide in the presence (positive control) or in the absence (negative control) of 100 mM KCl was used. To determine the level of background emission, the master mix without FHT oligonucleotide was used. After subtraction of the background emission, the ratio of FAM/FRET for the negative control was set to 1 and the ratio of the FAM/FRET for the positive control was set to 0.

6.2.11.2 Concentration-dependent studies of G-quadruplex formation via FRET

0.2 µM FHT oligonucleotide was incubated together with a serial dilution of the respective compound ranging from 1.6 µM to 100 µM in 10 mM Tris-HCl (pH 7.5) for 20 min at RT. The experiment was carried out in black 384 well plates with a flat bottom (Greiner) using a total reaction volume of 20 µl. As controls, the mentioned reaction mix containing no compound in the presence (positive control) or in the absence (negative control) of 100 mM KCl was used. To determine the level of background emission, the master mix without FHT oligonucleotide was used. The FRET signal was measured by the Tecan Infinite M200 plate reader which detected the emission of FAM at 535 nm and the emission of TAMRA at 590 nm after excitation at 485 nm wavelength. After subtraction of the background emission, the ratio of FAM/FRET for the negative control was set to 1 and the ratio of the FAM/FRET for the positive control was set to 0.

6.2.11.3 Competitive melting

Competitive melting experiments were performed in a Chromo4^{TM} system (Bio-Rad). To that end, 0.5 µM of the FHT oligonucleotide was mixed together with 10 µM of the respective compound and either in the presence or in the absence of 5 µM or 10 µM of double-stranded DNA (dsDNA). To allow for G-quadruplex formation, the reaction mixture was heated to 95°C and incubated for 2 min, followed by cooling down to 15°C with a cooling rate of 1°C/s and equilibrating for 10 min. Afterwards, the reaction mixture was heated to 90°C with a heating

rate of 1°C/min and the emission of FAM was measured every heating step. The data analysis was performed with Origin 8.1 and the Tm values were calculated by fitting the corresponding melting curves via the formula: y = A2 + (A1-A2)/(1 + exp((x-x0)/dx)).

6.2.12 Screening for TERRA-interacting proteins

6.2.12.1 Cultivation of HeLa S3 cells

3 x 10^6 HeLa S3 cells were seeded in 10 cm^2 dishes using DMEM (with 10% fetal calf serum (FCS), 100 U/ml penicillin and 100 µg/ml streptomycin) and grew at 37°C with 5% CO_2. After 24 h, medium was replaced by medium containing 10% FCS, 100 U/ml penicillin, 100 µg/ml streptomycin and 100 µM 4-thiouridine (s4U). After additional 14 h and reaching about 80% confluence (~8 x 10^6 cells/dish), the cells were ready for use.

6.2.12.2 In vivo crosslinking

HeLa S3 cells (cf. chapter 6.2.12.1) were quickly washed twice with cold 1x PBS. Next, cells were placed in the crosslinker (Bio-Link 365 from VILBER LOURMAT) and exposed to UV light of 365 nm at 0.15 J/cm^2 for 30 min. Afterwards, the cells were washed again twice with cold 1x PBS and then pooled in a 50 ml falcon by using a cell-scraper. Cells were centrifuged at 3000 x g for 10 min at 4°C and were ready for the preparation and lysis of cell nuclei.

6.2.12.3 Preparation and lysis of nuclei

HeLa S3 cell pellets (cf. chapter 6.2.12.2) were resuspended in three volumes of nuclei preparation buffer (NPB) and incubated on ice for 30 min. Afterwards, the cells were sheared by using a dounce tissue grinder (Sigma-Aldrich) until the outer cell membrane burst. Next, nuclei were pelletized by centrifugation for 10 min at 8000 x g at 4°C. Then, nuclei pellets were resuspended in two volume of nuclei storage buffer (NSB) and stored at -80°C. Pooled nuclei of 2 x 10^9 HeLa S3 cells were thawed on ice, centrifuged at 8000 x g, resuspended in two volumes of nuclei lysis buffer (NLB) and incubated for 45 min at RT. After lysis of the nuclei was completed, the extract was centrifuged at 13000 x g for 15 min at 4°C and the supernatant was collected and ready for TERRA purification.

6.2.12.4 TERRA purification

Prior to the TERRA purification, the magnetic beads Dynabeads® MyOne™ Carboxylic Acid (Invitrogen) were activated and coupled to the TERRA-specific hybridization linker DNA or to the randomized control linker DNA (cf. chapter 6.1.7). Therefore, 100 nmol of the respective hybridization linker DNA were coupled to 20 mg of the magnetic beads by using EDC in accordance to the manufacture´s protocol. To hybridize TERRA crosslinked to proteins with the linker DNA coupled to the magnetic beads, the supernatant from the lysed nuclear pellets of HeLa S3 cells was incubated with the beads at 45°C for 3 h, followed by further incubation at 25°C for 20 min. Afterwards, the magnetic beads were separated and washed three times with 1x PBS using a magnetic device (Dynal MPC™, Magnetic Particle Concentrator) in accordance to the manufacture´s protocol. Next, Nuclease S1 was added in accordance to the manufacture´s protocol and the solution was incubated for 30 min at 37°C, followed by incubation for 10 min at 25°C. The solution was shortly spun down by centrifugation at low speed to remove the magnetic beads. The supernatant with the protein eluate was collected and loaded on a 12% SDS gel (cf. chapter 6.2.8), followed by visualization of the protein bands via silver staining. Protein bands were excised from the gel for mass spectrometric analysis (cf. chapter 6.2.12.5).

6.2.12.5 Mass spectrometry

The excised bands from the SDS gel were analyzed via Nano-LC-MS/MS by the proteomics facility of the University of Konstanz.

6.2.13 DMS footprinting

6.2.13.1 Purification of oligonucleotides via preparative PAGE

Prior to footprinting experiments, purification of the used oligonucleotides was achieved by denaturating polyacrylamide-urea gel electrophoresis (PAGE) (Table 6.25).

Table 6.25: Reagents and composition of preparative PAGE

Reagent	Volume (ml)	Final concentration
Acrylamide 25% (19:1)	48	12%
Urea (9 M)	42	
TBE (10x) containing urea (9 M)	10	1x
APS (10%)	0.8	
TEMED	0.04	

Oligonucleotides were mixed in a 1:1 ratio with denaturating PAGE loading dye, followed by heating at 95°C for 3 min. After a pre-run of the gel for at least 45 min at 800 V in 1x TBE buffer, the samples were loaded onto the gel and the gel was running at 800 V for 4.5 h. Subsequently, the gel was wrapped into saran wrap. The bands for the full-length oligonucleotides were visualized on a silica gel plate with fluorescence indicator using UV light at 254 nm wavelength. The bands were encircled with a marker, cut out via scalpel and transferred into 1.5 ml eppendorf tubes. The gel slices were crushed into small pieces using a pipette tip and the equal weight of water was added to the gel pieces. After incubation on a thermomixer overnight at 4°C, the oligonucleotides were filtered via glass wool and desalted via ethanol precipitation (cf. chapter 6.2.1.2).

6.2.13.2 Radioactive labeling of oligonucleotides

Oligonucleotides were 5´ end labeled with γ-^{32}P-ATP using the reaction mix depicted in Table 6.26.

Table 6.26: Reaction mixture used for radioactive labeling of oligonucleotides

Reagent	Volume (µl)	Final concentration
T4-PNK buffer (10x)	5	1x
Oligonucleotide (50 µM)	4	4 µM
γ-^{32}P-ATP (10 mCi/ml)	2	0.4 mCi/ml
T4-PNK (10 U/µl)	2	0.4 U/µl
H_2O	37	

The reaction mixture was incubated for 30 min at 37°C. To inactivate T4-PNK, the mixture was incubated at 95°C for further 10 min. Subsequently to the labeling, the 5´ end γ-^{32}P-ATP labeled oligonucleotides were purified via gel filtration by Sephadex G-25 columns to remove remaining salts, T4-PNK and free γ-^{32}P-ATP.

6.2.13.3 Folding of 5´ end γ-^{32}P-ATP labeled oligonucleotides

To enable correct folding of 5´ end γ-^{32}P-ATP labeled oligonucleotides, the oligonucleotides were heated in the reaction mixture (depicted in Table 6.27) at 95°C for 5 min and cooled down to RT within 3 h.

Table 6.27: Reaction mixture used for folding of 5´ end γ-^{32}P-ATP labeled oligonucleotides

Reagent	Volume (µl)	Final concentration
Oligonucleotide (2 µM)	4	0.4 µM
TBE (2x) (+/- 40 mM KCl)	10	1 x TBE; +/- 20 mM KCl
H_2O	6	

6.2.13.4 DMS treatment

After incubation of the γ-^{32}P-ATP labeled oligonucleotides in TBE buffer containing +/- 20 mM KCl (cf. chapter 6.2.13.3), 2 µl of DMS solution (10% DMS, 50% ethanol) were added to the reaction mixture for 30 s at RT. To stop the DMS-induced DNA methylation, 5.5 µl of DMS-stop solution was added. Subsequently, an ethanol precipitation was performed (cf. chapter 6.2.1.2).

6.2.13.5 Chemical sequencing

After incubation of the γ-^{32}P-ATP labeled oligonucleotides in TBE buffer (cf. chapter 6.2.13.3), the oligonucleotides were incubated with 20 µl of formic acid (for purine residue specific cleavage) or with 40 µl of hydrazine (for pyrimidine residue specific cleavage) for 20 min at RT. The reaction was stopped by the addition of 60 µl of sequencing stop solution and an ethanol precipitation was performed (cf. chapter 6.2.1.2).

6.2.13.6 Piperidine treatment

The oligonucleotides purified via ethanol precipitation were incubated with 100 µl of 1 M piperidine solution for 30 min at 95°C. Afterwards, the oligonucleotides were dried in a SpeedVac, washed in 100 µl water, again dried in a SpeedVac and resuspended in 20 µl 1x PAGE loading dye.

6.2.13.7 Separation of cleavage products via denaturating PAGE

For separation of the cleavage products from DMS treatment and chemical sequencing a 12% denaturating PAGE was performed (cf. chapter 6.2.13.1). After a pre-run of at least 30 min at 1800 V, samples were heated at 95°C for 2 min and 5 µl of each sample were loaded on the gel. The gel was running for 3 h in 1x TBE buffer at 1800 V. Afterwards, the gel was dried in a vacuum gel dryer, placed into an X-ray film cassette together with an X-ray film for 14 h and was screened on a phosphor imager.

7. Abbreviations

°C	degree celsius
µg	microgram
µl	microliter
µM	micromolar
360A	2,6-pyridine-dicarboxamide bisquinolinium
A	adenine
APS	ammonium persulfate
ATP	adenosine triphosphate
bp	basepair
c	concentration
C	cytosine
CD	circular dichroism
cm	centimeter
DMEM	dulbecco's modified eagle's medium
DMS	dimethylsulfate
DNA	deoxyribonucleic acid
dsDNA	double-stranded DNA
EDTA	ethylene-diamine-tetraacetic acid
EGTA	ethylene-glycol-tetraacetic acid
FCS	fetal calf serum
eGFP	enhanced green fluorescent protein
FAM	fluorescein amidite
FRET	fluorescence resonance energy transfer
g	gram
G	guanine
GMP	guanosine monophosphate
h	hour
HEPES	N-2-hydroxyethylpiperazine-N'-2-ethanesulfonic acid
kb	kilobase

L	liter
LC	liquid chromatography
LNA	locked nucleic acid
M	molar
min	minute
ml	milliliter
mm	millimeter
mRNA	messenger RNA
MS	mass spectrometry
nm	nanometer
NMR	nuclear magnetic resonance
nt	nucleotide
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffered saline
PCR	polymerase chain reaction
Phen DC_3	6,6'-disubstituted-2,2'-bipyridine
Phen DC_6	2,9-disubstituted-1,10-phenanthroline
PMSF	phenylmethanesulfonyl fluoride
PNA	peptide nucleic acid
RNA	ribonucleic acid
rDNA	ribosomal DNA
rpm	rounds per minute
RT	room temperature
s4U	4-thiouridine
sec	second
SDS	sodium dodecyl sulfate
siRNA	small interfering RNA
T	thymine
T4-PNK	T4 Polynucleotide kinase
TAMRA	tetramethylrhodamine
TBE	Tris/Borate/EDTA

TEMED	N,N,N’,N’-tetramethyleneethylenediamine
TERRA	telomeric repeat containing RNA
Tm	temperature midpoint
Tris	Tris(hydroxymethyl)aminomethane
tRNA	transfer RNA
U	uracil
U	unit
UTR	untranslated region
UV	ultraviolet
V	volt

8. References

1 Watson, J. D. & Crick, F. H. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* **171**, 737-738 (1953).

2 Pray, L. Discovery of DNA structure and function: Watson and Crick. *Nature Education* **1** (2008).

3 Yakovchuk, P., Protozanova, E. & Frank-Kamenetskii, M. D. Base-stacking and base-pairing contributions into thermal stability of the DNA double helix. *Nucleic acids research* **34**, 564-574, doi:34/2/564 [pii] 10.1093/nar/gkj454 (2006).

4 Wahl, M. C. & Sundaralingam, M. Crystal structures of A-DNA duplexes. *Biopolymers* **44**, 45-63, doi:10.1002/(SICI)1097-0282(1997)44:1<45::AID-BIP4>3.0.CO;2-# [pii] 10.1002/(SICI)1097-0282(1997)44:1<45::AID-BIP4>3.0.CO;2-# (1997).

5 Wang, A. H. *et al.* Molecular structure of a left-handed double helical DNA fragment at atomic resolution. *Nature* **282**, 680-686 (1979).

6 Rich, A. & Zhang, S. Timeline: Z-DNA: the long road to biological function. *Nature reviews. Genetics* **4**, 566-572, doi:10.1038/nrg1115 (2003).

7 Lyamichev, V. I., Mirkin, S. M. & Frank-Kamenetskii, M. D. A pH-dependent structural transition in the homopurine-homopyrimidine tract in superhelical DNA. *J Biomol Struct Dyn* **3**, 327-338, doi:10.1080/07391102.1985.10508420 (1985).

8 Lyamichev, V. I., Mirkin, S. M. & Frank-Kamenetskii, M. D. Structures of homopurine-homopyrimidine tract in superhelical DNA. *J Biomol Struct Dyn* **3**, 667-669, doi:10.1080/07391102.1986.10508454 (1986).

9 Mirkin, S. M. *et al.* DNA H form requires a homopurine-homopyrimidine mirror repeat. *Nature* **330**, 495-497, doi:10.1038/330495a0 (1987).

10 Panayotatos, N. & Wells, R. D. Cruciform structures in supercoiled DNA. *Nature* **289**, 466-470 (1981).

11 Oussatcheva, E. A. *et al.* Influence of global DNA topology on cruciform formation in supercoiled DNA. *Journal of molecular biology* **338**, 735-743, doi:10.1016/j.jmb.2004.02.075 S0022283604003134 [pii] (2004).

12 Lilley, D. M. The inverted repeat as a recognizable structural feature in supercoiled DNA molecules. *Proceedings of the National Academy of Sciences of the United States of America* **77**, 6468-6472 (1980).

13 Gehring, K., Leroy, J. L. & Gueron, M. A tetrameric DNA structure with protonated cytosine.cytosine base pairs. *Nature* **363**, 561-565, doi:10.1038/363561a0 (1993).

14 Gellert, M., Lipsett, M. N. & Davies, D. R. Helix formation by guanylic acid. *Proceedings of the National Academy of Sciences of the United States of America* **48**, 2013-2018 (1962).

15 Du, X. *et al.* The genome-wide distribution of non-B DNA motifs is shaped by operon structure and suggests the transcriptional importance of non-B DNA structures in Escherichia coli. *Nucleic acids research* **41**, 5965-5977, doi:gkt308 [pii] 10.1093/nar/gkt308.

16 Huppert, J. L. & Balasubramanian, S. G-quadruplexes in promoters throughout the human genome. *Nucleic acids research* **35**, 406-413, doi:gkl1057 [pii] 10.1093/nar/gkl1057 (2007).

17 Strawbridge, E. M., Benson, G., Gelfand, Y. & Benham, C. J. The distribution of inverted repeat sequences in the Saccharomyces cerevisiae genome. *Curr Genet* **56**, 321-340, doi:10.1007/s00294-010-0302-6.

18 Murat, P. & Balasubramanian, S. Existence and consequences of G-quadruplex structures in DNA. *Current opinion in genetics & development* **25**, 22-29, doi:S0959-437X(13)00149-4 [pii] 10.1016/j.gde.2013.10.012.

19 Zhao, J., Bacolla, A., Wang, G. & Vasquez, K. M. Non-B DNA structure-induced genetic instability and evolution. *Cell Mol Life Sci* **67**, 43-62, doi:10.1007/s00018-009-0131-2.

20 Boyer, A. S., Grgurevic, S., Cazaux, C. & Hoffmann, J. S. The human specialized DNA polymerases and non-B DNA: vital relationships to preserve genome integrity. *Journal of molecular biology* **425**, 4767-4781, doi:S0022-2836(13)00606-2 [pii] 10.1016/j.jmb.2013.09.022.

21 Bansal, M., Kumar, A. & Yella, V. R. Role of DNA sequence based structural features of promoters in transcription initiation and gene expression. *Current opinion in structural biology* **25**, 77-85, doi:S0959-440X(14)00008-6 [pii] 10.1016/j.sbi.2014.01.007.

22 Cherepanov, P. *et al.* Mode of interaction of G-quartets with the integrase of human immunodeficiency virus type 1. *Molecular pharmacology* **52**, 771-780 (1997).

23 Incles, C. M. *et al.* A G-quadruplex telomere targeting agent produces p16-associated senescence and chromosomal fusions in human prostate cancer cells. *Mol Cancer Ther* **3**, 1201-1206, doi:3/10/1201 [pii] (2004).

24 McAllister, R. G. *et al.* Lentivector integration sites in ependymal cells from a model of metachromatic leukodystrophy: non-B DNA as a new factor influencing integration. *Mol Ther Nucleic Acids* **3**, e187, doi:mtna201439 [pii] 10.1038/mtna.2014.39.

25 Jayaram, B. & Beyeridge, D. L. Modeling DNA in aqueous solutions: theoretical and computer simulation studies on the ion atmosphere of DNA. *Annu Rev Biophys Biomol Struct* **25**, 367-394, doi:10.1146/annurev.bb.25.060196.002055 (1996).

26 Nikolova, E. N. *et al.* Transient Hoogsteen base pairs in canonical duplex DNA. *Nature* **470**, 498-502, doi:10.1038/nature09775 (2011).

27 Hamelberg, D., Williams, L. D. & Wilson, W. D. Influence of the dynamic positions of cations on the structure of the DNA minor groove: sequence-dependent effects. *Journal of the American Chemical Society* **123**, 7745-7755 (2001).

28 van Erp, T. S., Cuesta-Lopez, S. & Peyrard, M. Bubbles and denaturation in DNA. *Eur Phys J E Soft Matter* **20**, 421-434, doi:10.1140/epje/i2006-10032-2 (2006).

29 Harteis, S. & Schneider, S. Making the bend: DNA tertiary structure and protein-DNA interactions. *Int J Mol Sci* **15**, 12335-12363, doi:10.3390/ijms150712335 (2014).

30 Gros, F. *et al.* Molecular and biological characterization of messenger RNA. *Cold Spring Harb Symp Quant Biol* **26**, 111-132 (1961).

31 Pelechano, V. & Steinmetz, L. M. Gene regulation by antisense transcription. *Nature reviews. Genetics* **14**, 880-893, doi:10.1038/nrg3594 (2013).

32 Serganov, A. & Nudler, E. A decade of riboswitches. *Cell* **152**, 17-24, doi:10.1016/j.cell.2012.12.024 (2013).

33 Valadi, H. *et al.* Exosome-mediated transfer of mRNAs and microRNAs is a novel mechanism of genetic exchange between cells. *Nature cell biology* **9**, 654-659, doi:10.1038/ncb1596 (2007).

34 Rana, T. M. Illuminating the silence: understanding the structure and function of small RNAs. *Nature reviews. Molecular cell biology* **8**, 23-36, doi:10.1038/nrm2085 (2007).

35 Halder, S. & Bhattacharyya, D. RNA structure and dynamics: a base pairing perspective. *Progress in biophysics and molecular biology* **113**, 264-283, doi:10.1016/j.pbiomolbio.2013.07.003 (2013).

36 Huppert, J. L. Hunting G-quadruplexes. *Biochimie* **90**, 1140-1148, doi:10.1016/j.biochi.2008.01.014 (2008).

37 Millevoi, S., Moine, H. & Vagner, S. G-quadruplexes in RNA biology. *Wiley interdisciplinary reviews. RNA* **3**, 495-507, doi:10.1002/wrna.1113 (2012).

38 Bang, I. Untersuchungen über die Guanylsäure. *Biochemische Zeitschrift* **26** (1910).

39 Burge, S., Parkinson, G. N., Hazel, P., Todd, A. K. & Neidle, S. Quadruplex DNA: sequence, topology and structure. *Nucleic acids research* **34**, 5402-5415, doi:gkl655 [pii] 10.1093/nar/gkl655 (2006).

40 van Mourik, T. & Dingley, A. J. Characterization of the monovalent ion position and hydrogen-bond network in guanine quartets by DFT calculations of NMR parameters. *Chemistry* **11**, 6064-6079, doi:10.1002/chem.200500198 (2005).

41 Guedin, A., Gros, J., Alberti, P. & Mergny, J. L. How long is too long? Effects of loop size on G-quadruplex stability. *Nucleic acids research* **38**, 7858-7868, doi:gkq639 [pii] 10.1093/nar/gkq639.

42 Hazel, P., Huppert, J., Balasubramanian, S. & Neidle, S. Loop-length-dependent folding of G-quadruplexes. *Journal of the American Chemical Society* **126**, 16405-16415, doi:10.1021/ja045154j (2004).

43 Bugaut, A. & Balasubramanian, S. 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, doi:10.1021/bi701873c (2008).

44 Bochman, M. L., Paeschke, K. & Zakian, V. A. DNA secondary structures: stability and function of G-quadruplex structures. *Nature reviews. Genetics* **13**, 770-780, doi:10.1038/nrg3296 (2012).

45 Zhang, D. H. *et al.* Monomorphic RNA G-quadruplex and polymorphic DNA G-quadruplex structures responding to cellular environmental factors. *Biochemistry* **49**, 4554-4563, doi:10.1021/bi1002822 (2010).

46 Holder, I. T., Drescher, M. & Hartig, J. S. Structural characterization of quadruplex DNA with in-cell EPR approaches. *Bioorganic & medicinal chemistry* **21**, 6156-6161, doi:10.1016/j.bmc.2013.04.014 (2013).

47 Wang, Y. & Patel, D. J. Guanine residues in d(T2AG3) and d(T2G4) form parallel-stranded potassium cation stabilized G-quadruplexes with anti glycosidic torsion angles in solution. *Biochemistry* **31**, 8112-8119 (1992).

48 Laughlan, G. *et al.* The high-resolution crystal structure of a parallel-stranded guanine tetraplex. *Science* **265**, 520-524 (1994).

49 Sen, D. & Gilbert, W. Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. *Nature* **334**, 364-366, doi:10.1038/334364a0 (1988).

50 Pan, B., Shi, K. & Sundaralingam, M. Base-tetrad swapping results in dimerization of RNA quadruplexes: implications for formation of the i-motif RNA octaplex. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 3130-3134, doi:0507730103 [pii] 10.1073/pnas.0507730103 (2006).

51 Wang, Y. & Patel, D. J. Solution structure of the Tetrahymena telomeric repeat d(T2G4)4 G-tetraplex. *Structure* **2**, 1141-1156 (1994).

52 Smith, F. W. & Feigon, J. Quadruplex structure of Oxytricha telomeric DNA oligonucleotides. *Nature* **356**, 164-168, doi:10.1038/356164a0 (1992).

53 Wang, Y. & Patel, D. J. Solution structure of the human telomeric repeat d[AG3(T2AG3)3] G-tetraplex. *Structure* **1**, 263-282 (1993).

54 Sundquist, W. I. & Klug, A. Telomeric DNA dimerizes by formation of guanine tetrads between hairpin loops. *Nature* **342**, 825-829, doi:10.1038/342825a0 (1989).

55 Williamson, J. R., Raghuraman, M. K. & Cech, T. R. Monovalent cation-induced structure of telomeric DNA: the G-quartet model. *Cell* **59**, 871-880, doi:0092-8674(89)90610-7 [pii] (1989).

56 Kettani, A. *et al.* Solution structure of a Na cation stabilized DNA quadruplex containing G.G.G.G and G.C.G.C tetrads formed by G-G-G-C repeats observed in adeno-associated viral DNA. *Journal of molecular biology* **282**, 619-636, doi:S0022-2836(98)92030-7 [pii] 10.1006/jmbi.1998.2030 (1998).

57 Karsisiotis, A. I. & Webba da Silva, M. Structural probes in quadruplex nucleic acid structure determination by NMR. *Molecules* **17**, 13073-13086, doi:10.3390/molecules171113073 (2012).

58 Pandey, S., Agarwala, P. & Maiti, S. Effect of loops and G-quartets on the stability of RNA G-quadruplexes. *J Phys Chem B* **117**, 6896-6905, doi:10.1021/jp401739m.

59 Risitano, A. & Fox, K. R. Influence of loop size on the stability of intramolecular DNA quadruplexes. *Nucleic acids research* **32**, 2598-2606, doi:10.1093/nar/gkh598 32/8/2598 [pii] (2004).

60 Risitano, A. & Fox, K. R. Stability of intramolecular DNA quadruplexes: comparison with DNA duplexes. *Biochemistry* **42**, 6507-6513, doi:10.1021/bi026997v (2003).

61 Cosconati, S. *et al.* Tandem application of virtual screening and NMR experiments in the discovery of brand new DNA quadruplex groove binders. *Journal of the American Chemical Society* **131**, 16336-16337, doi:10.1021/ja9063662 (2009).

62 Huppert, J. L. Four-stranded nucleic acids: structure, function and targeting of G-quadruplexes. *Chem Soc Rev* **37**, 1375-1384, doi:10.1039/b702491f (2008).

63 Joachimi, A., Benz, A. & Hartig, J. S. A comparison of DNA and RNA quadruplex structures and stabilities. *Bioorganic & medicinal chemistry* **17**, 6811-6815, doi:10.1016/j.bmc.2009.08.043 (2009).

64 Egli, M., Portmann, S. & Usman, N. RNA hydration: a detailed look. *Biochemistry* **35**, 8489-8494, doi:10.1021/bi9607214 (1996).

65 Guschlbauer, W., Chantot, J. F. & Thiele, D. Four-stranded nucleic acid structures 25 years later: from guanosine gels to telomer DNA. *J Biomol Struct Dyn* **8**, 491-511, doi:10.1080/07391102.1990.10507825 (1990).

66 Hud, N. V., Smith, F. W., Anet, F. A. & Feigon, J. The selectivity for K+ versus Na+ in DNA quadruplexes is dominated by relative free energies of hydration: a thermodynamic analysis by 1H NMR. *Biochemistry* **35**, 15383-15390, doi:10.1021/bi9620565 bi9620565 [pii] (1996).

67 Halder, K. & Hartig, J. S. RNA quadruplexes. *Met Ions Life Sci* **9**, 125-139 (2011).

68 Yang, D. & Okamoto, K. Structural insights into G-quadruplexes: towards new anticancer drugs. *Future Med Chem* **2**, 619-646.

69 Halder, K., Wieland, M. & Hartig, J. S. Predictable suppression of gene expression by 5'-UTR-based RNA quadruplexes. *Nucleic acids research* **37**, 6811-6817, doi:gkp696 [pii] 10.1093/nar/gkp696 (2009).

70 Pandey, S., Agarwala, P. & Maiti, S. Effect of loops and G-quartets on the stability of RNA G-quadruplexes. *J Phys Chem B* **117**, 6896-6905, doi:10.1021/jp401739m (2013).

71 Guedin, A., Gros, J., Alberti, P. & Mergny, J. L. How long is too long? Effects of loop size on G-quadruplex stability. *Nucleic acids research* **38**, 7858-7868, doi:10.1093/nar/gkq639 (2010).

72 Zhang, A. Y., Bugaut, A. & Balasubramanian, S. A sequence-independent analysis of the loop length dependence of intramolecular RNA G-quadruplex stability and topology. *Biochemistry* **50**, 7251-7258, doi:10.1021/bi200805j (2011).

73 Guedin, A., De Cian, A., Gros, J., Lacroix, L. & Mergny, J. L. Sequence effects in single-base loops for quadruplexes. *Biochimie* **90**, 686-696, doi:10.1016/j.biochi.2008.01.009 (2008).

74 Ambrus, A., Chen, D., Dai, J., Jones, R. A. & Yang, D. Solution structure of the biologically relevant G-quadruplex element in the human c-MYC promoter. Implications for G-quadruplex stabilization. *Biochemistry* **44**, 2048-2058, doi:10.1021/bi048242p (2005).

75 Hatzakis, E., Okamoto, K. & Yang, D. Thermodynamic stability and folding kinetics of the major G-quadruplex and its loop isomers formed in the nuclease hypersensitive element in the human c-Myc promoter: effect of loops and flanking segments on the stability of parallel-stranded intramolecular G-quadruplexes. *Biochemistry* **49**, 9152-9160, doi:10.1021/bi100946g.

76 Hayden, K. L. & Graves, D. E. Addition of bases to the 5'-end of human telomeric DNA: influences on thermal stability and energetics of unfolding. *Molecules* **19**, 2286-2298, doi:molecules19022286 [pii] 10.3390/molecules19022286.

77 Petraccone, L. *et al.* Structure and stability of higher-order human telomeric quadruplexes. *Journal of the American Chemical Society* **133**, 20951-20961, doi:10.1021/ja209192a.

78 Arora, A., Nair, D. R. & Maiti, S. Effect of flanking bases on quadruplex stability and Watson-Crick duplex competition. *FEBS J* **276**, 3628-3640, doi:EJB7082 [pii] 10.1111/j.1742-4658.2009.07082.x (2009).

79 Beaudoin, J. D. & Perreault, J. P. 5'-UTR G-quadruplex structures acting as translational repressors. *Nucleic acids research* **38**, 7022-7036, doi:10.1093/nar/gkq557 (2010).

80 Beaudoin, J. D., Jodoin, R. & Perreault, J. P. New scoring system to identify RNA G-quadruplex folding. *Nucleic acids research* **42**, 1209-1223, doi:10.1093/nar/gkt904 (2014).

81 Riccelli, P. V., Mandell, K. E. & Benight, A. S. Melting studies of dangling-ended DNA hairpins: effects of end length, loop sequence and biotinylation of loop bases. *Nucleic acids research* **30**, 4088-4093 (2002).

82 O'Toole, A. S., Miller, S. & Serra, M. J. Stability of 3' double nucleotide overhangs that model the 3' ends of siRNA. *Rna* **11**, 512-516, doi:10.1261/rna.7254905 (2005).

83 Bommarito, S., Peyret, N. & SantaLucia, J., Jr. Thermodynamic parameters for DNA sequences with dangling ends. *Nucleic acids research* **28**, 1929-1934, doi:gkd309 [pii] (2000).

84 Ohmichi, T., Nakano, S., Miyoshi, D. & Sugimoto, N. Long RNA dangling end has large energetic contribution to duplex stability. *Journal of the American Chemical Society* **124**, 10367-10372, doi:ja0255406 [pii] (2002).

85 Clanton-Arrowood, K., McGurk, J. & Schroeder, S. J. 3' terminal nucleotides determine thermodynamic stabilities of mismatches at the ends of RNA helices. *Biochemistry* **47**, 13418-13427, doi:10.1021/bi801594k 10.1021/bi801594k [pii] (2008).

86 Isaksson, J. & Chattopadhyaya, J. A uniform mechanism correlating dangling-end stabilization and stacking geometry. *Biochemistry* **44**, 5390-5401, doi:10.1021/bi047414f (2005).

87 Todd, A. K., Johnston, M. & Neidle, S. Highly prevalent putative quadruplex sequence motifs in human DNA. *Nucleic acids research* **33**, 2901-2907, doi:33/9/2901 [pii] 10.1093/nar/gki553 (2005).

88 Huppert, J. L. & Balasubramanian, S. Prevalence of quadruplexes in the human genome. *Nucleic acids research* **33**, 2908-2916, doi:33/9/2908 [pii] 10.1093/nar/gki609 (2005).

89 Capra, J. A., Paeschke, K., Singh, M. & Zakian, V. A. G-quadruplex DNA sequences are evolutionarily conserved and associated with distinct genomic features in Saccharomyces cerevisiae. *PLoS Comput Biol* **6**, e1000861, doi:10.1371/journal.pcbi.1000861.

90 Hershman, S. G. *et al.* Genomic distribution and functional analyses of potential G-quadruplex-forming sequences in Saccharomyces cerevisiae. *Nucleic acids research* **36**, 144-156, doi:gkm986 [pii] 10.1093/nar/gkm986 (2008).

91 Rawal, P. *et al.* Genome-wide prediction of G4 DNA as regulatory motifs: role in Escherichia coli global regulation. *Genome research* **16**, 644-655, doi:16/5/644 [pii] 10.1101/gr.4508806 (2006).

92 Blackburn, E. H. Structure and function of telomeres. *Nature* **350**, 569-573, doi:10.1038/350569a0 (1991).

93 Drygin, D. *et al.* Anticancer activity of CX-3543: a direct inhibitor of rRNA biogenesis. *Cancer research* **69**, 7653-7661, doi:0008-5472.CAN-09-1304 [pii] 10.1158/0008-5472.CAN-09-1304 (2009).

94 Siddiqui-Jain, A., Grand, C. L., Bearss, D. J. & Hurley, L. H. Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 11593-11598, doi:10.1073/pnas.182256799 182256799 [pii] (2002).

95 Palumbo, S. L. *et al.* A novel G-quadruplex-forming GGA repeat region in the c-myb promoter is a critical regulator of promoter activity. *Nucleic acids research* **36**, 1755-1769, doi:gkm1069 [pii] 10.1093/nar/gkm1069 (2008).

96 Cogoi, S. & Xodo, L. E. G-quadruplex formation within the promoter of the KRAS proto-oncogene and its effect on transcription. *Nucleic acids research* **34**, 2536-2549, doi:34/9/2536 [pii] 10.1093/nar/gkl286 (2006).

97 Dexheimer, T. S., Sun, D. & Hurley, L. H. Deconvoluting the structural and drug-recognition complexity of the G-quadruplex-forming region upstream of the bcl-2 P1 promoter. *Journal of the American Chemical Society* **128**, 5404-5415, doi:10.1021/ja0563861 (2006).

98 Sun, D., Guo, K., Rusche, J. J. & Hurley, L. H. Facilitation of a structural transition in the polypurine/polypyrimidine tract within the proximal promoter region of the human VEGF gene by the presence of potassium and G-quadruplex-interactive agents. *Nucleic acids research* **33**, 6070-6080, doi:33/18/6070 [pii] 10.1093/nar/gki917 (2005).

99 Hsu, S. T. *et al.* A G-rich sequence within the c-kit oncogene promoter forms a parallel G-quadruplex having asymmetric G-tetrad dynamics. *Journal of the American Chemical Society* **131**, 13399-13409, doi:10.1021/ja904007p (2009).

100 Azzalin, C. M., Reichenbach, P., Khoriauli, L., Giulotto, E. & Lingner, J. Telomeric repeat containing RNA and RNA surveillance factors at mammalian chromosome ends. *Science* **318**, 798-801, doi:10.1126/science.1147182 (2007).

101 Neidle, S. & Parkinson, G. N. The structure of telomeric DNA. *Current opinion in structural biology* **13**, 275-283 (2003).

102 Vorlickova, M. *et al.* Circular dichroism and guanine quadruplexes. *Methods* **57**, 64-75, doi:10.1016/j.ymeth.2012.03.011 (2012).

103 Halder, K., Benzler, M. & Hartig, J. S. Reporter assays for studying quadruplex nucleic acids. *Methods* **57**, 115-121, doi:10.1016/j.ymeth.2012.02.005 (2012).

104 Biffi, G., Tannahill, D., McCafferty, J. & Balasubramanian, S. Quantitative visualization of DNA G-quadruplex structures in human cells. *Nature chemistry* **5**, 182-186, doi:10.1038/nchem.1548 (2013).

105 Xu, Y., Sato, H., Shinohara, K., Komiyama, M. & Sugiyama, H. T-loop formation by human telomeric G-quadruplex. *Nucleic acids symposium series*, 243-244, doi:10.1093/nass/nrm122 (2007).

106 Kumari, S., Bugaut, A., Huppert, J. L. & Balasubramanian, S. An RNA G-quadruplex in the 5' UTR of the NRAS proto-oncogene modulates translation. *Nat Chem Biol* **3**, 218-221, doi:nchembio864 [pii] 10.1038/nchembio864 (2007).

107 Besnard, E. *et al.* Unraveling cell type-specific and reprogrammable human replication origin signatures associated with G-quadruplex consensus motifs. *Nature structural & molecular biology* **19**, 837-844, doi:10.1038/nsmb.2339 (2012).

108 Cahoon, L. A. & Seifert, H. S. An alternative DNA structure is necessary for pilin antigenic variation in Neisseria gonorrhoeae. *Science* **325**, 764-767, doi:10.1126/science.1175653 (2009).

109 Gomez, D. *et al.* Telomerase downregulation induced by the G-quadruplex ligand 12459 in A549 cells is mediated by hTERT RNA alternative splicing. *Nucleic acids research* **32**, 371-379, doi:10.1093/nar/gkh181 (2004).

110 Decorsiere, A., Cayrel, A., Vagner, S. & Millevoi, S. Essential role for the interaction between hnRNP H/F and a G quadruplex in maintaining p53 pre-mRNA 3'-end

processing and function during DNA damage. *Genes & development* **25**, 220-225, doi:10.1101/gad.607011 (2011).

111 Subramanian, M. *et al.* G-quadruplex RNA structure as a signal for neurite mRNA targeting. *EMBO reports* **12**, 697-704, doi:10.1038/embor.2011.76 (2011).

112 Konig, S. L., Huppert, J. L., Sigel, R. K. & Evans, A. C. Distance-dependent duplex DNA destabilization proximal to G-quadruplex/i-motif sequences. *Nucleic acids research* **41**, 7453-7461, doi:10.1093/nar/gkt476 (2013).

113 Kumar, N. & Maiti, S. The effect of osmolytes and small molecule on Quadruplex-WC duplex equilibrium: a fluorescence resonance energy transfer study. *Nucleic acids research* **33**, 6723-6732, doi:10.1093/nar/gki961 (2005).

114 Müller, H. The remarking of chromosomes. *Collecting Net* **13** (1938).

115 Wang, Y. & Patel, D. J. Solution structure of the Oxytricha telomeric repeat d[G4(T4G4)3] G-tetraplex. *Journal of molecular biology* **251**, 76-94, doi:10.1006/jmbi.1995.0417 (1995).

116 de Lange, T. Shelterin: the protein complex that shapes and safeguards human telomeres. *Genes & development* **19**, 2100-2110, doi:19/18/2100 [pii] 10.1101/gad.1346005 (2005).

117 Blasco, M. A. Telomeres and human disease: ageing, cancer and beyond. *Nature reviews. Genetics* **6**, 611-622, doi:10.1038/nrg1656 (2005).

118 Ruden, M. & Puri, N. Novel anticancer therapeutics targeting telomerase. *Cancer Treat Rev* **39**, 444-456, doi:10.1016/j.ctrv.2012.06.007 (2013).

119 Parkinson, G. N., Lee, M. P. & Neidle, S. Crystal structure of parallel quadruplexes from human telomeric DNA. *Nature* **417**, 876-880, doi:10.1038/nature755 nature755 [pii] (2002).

120 Ambrus, A. *et al.* Human telomeric sequence forms a hybrid-type intramolecular G-quadruplex structure with mixed parallel/antiparallel strands in potassium solution. *Nucleic acids research* **34**, 2723-2735, doi:34/9/2723 [pii] 10.1093/nar/gkl348 (2006).

121 Xu, Y., Noguchi, Y. & Sugiyama, H. The new models of the human telomere d[AGGG(TTAGGG)3] in K+ solution. *Bioorganic & medicinal chemistry* **14**, 5584-5591, doi:S0968-0896(06)00300-2 [pii] 10.1016/j.bmc.2006.04.033 (2006).

122 Luu, K. N., Phan, A. T., Kuryavyi, V., Lacroix, L. & Patel, D. J. Structure of the human telomere in K+ solution: an intramolecular (3 + 1) G-quadruplex scaffold. *Journal of the American Chemical Society* **128**, 9963-9970, doi:10.1021/ja062791w (2006).

123 Xue, Y. *et al.* Human telomeric DNA forms parallel-stranded intramolecular G-quadruplex in K+ solution under molecular crowding condition. *Journal of the American Chemical Society* **129**, 11185-11191, doi:10.1021/ja0730462 (2007).

124 Sharma, V. R. & Sheardy, R. D. The Human Telomere Sequence, (TTAGGG)4, in the Absence and Presence of Cosolutes: A Spectroscopic Investigation. *Molecules* **19**, 595-608, doi:molecules19010595 [pii] 10.3390/molecules19010595.

125 Singh, V., Azarkh, M., Exner, T. E., Hartig, J. S. & Drescher, M. Human telomeric quadruplex conformations studied by pulsed EPR. *Angewandte Chemie* **48**, 9728-9730, doi:10.1002/anie.200902146 (2009).

126 Yu, H. Q., Miyoshi, D. & Sugimoto, N. Characterization of structure and stability of long telomeric DNA G-quadruplexes. *Journal of the American Chemical Society* **128**, 15461-15468, doi:10.1021/ja064536h (2006).

127 Vorlickova, M., Chladkova, J., Kejnovska, I., Fialova, M. & Kypr, J. Guanine tetraplex topology of human telomere DNA is governed by the number of (TTAGGG) repeats. *Nucleic acids research* **33**, 5851-5860, doi:33/18/5851 [pii] 10.1093/nar/gki898 (2005).

128 Bauer, L., Tluckova, K., Tohova, P. & Viglasky, V. G-quadruplex motifs arranged in tandem occurring in telomeric repeats and the insulin-linked polymorphic region. *Biochemistry* **50**, 7484-7492, doi:10.1021/bi2003235.

129 Petraccone, L., Trent, J. O. & Chaires, J. B. The tail of the telomere. *Journal of the American Chemical Society* **130**, 16530-16532, doi:10.1021/ja8075567 10.1021/ja8075567 [pii] (2008).

130 Haider, S., Parkinson, G. N. & Neidle, S. Molecular dynamics and principal components analysis of human telomeric quadruplex multimers. *Biophysical journal* **95**, 296-311, doi:S0006-3495(08)70304-7 [pii] 10.1529/biophysj.107.120501 (2008).

131 Azarkh, M. *et al.* Intracellular conformations of human telomeric quadruplexes studied by electron paramagnetic resonance spectroscopy. *Chemphyschem* **13**, 1444-1447, doi:10.1002/cphc.201100980 (2012).

132 Maizels, N. Dynamic roles for G4 DNA in the biology of eukaryotic cells. *Nature structural & molecular biology* **13**, 1055-1059, doi:10.1038/nsmb1171 (2006).

133 Sun, D. *et al.* Inhibition of human telomerase by a G-quadruplex-interactive compound. *Journal of medicinal chemistry* **40**, 2113-2116, doi:10.1021/jm970199z jm970199z [pii] (1997).

134 Azzalin, C. M. & Lingner, J. Telomere functions grounding on TERRA firma. *Trends Cell Biol* **25**, 29-36, doi:10.1016/j.tcb.2014.08.007 (2015).

135 Arora, R., Brun, C. M. & Azzalin, C. M. TERRA: Long Noncoding RNA at Eukaryotic Telomeres. *Prog Mol Subcell Biol* **51**, 65-94, doi:10.1007/978-3-642-16502-3_4 (2011).

136 Xu, Y., Suzuki, Y., Ito, K. & Komiyama, M. Telomeric repeat-containing RNA structure in living cells. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 14579-14584, doi:10.1073/pnas.1001177107 (2010).

137 Deng, Z., Norseen, J., Wiedmer, A., Riethman, H. & Lieberman, P. M. TERRA RNA binding to TRF2 facilitates heterochromatin formation and ORC recruitment at telomeres. *Molecular cell* **35**, 403-413, doi:10.1016/j.molcel.2009.06.025 (2009).

138 Aguilera, A. & Garcia-Muse, T. R loops: from transcription byproducts to threats to genome stability. *Molecular cell* **46**, 115-124, doi:10.1016/j.molcel.2012.04.009 (2012).

139 Arora, R. *et al.* RNaseH1 regulates TERRA-telomeric DNA hybrids and telomere maintenance in ALT tumour cells. *Nature communications* **5**, 5220, doi:10.1038/ncomms6220 (2014).

140 Sun, D. & Hurley, L. H. The importance of negative superhelicity in inducing the formation of G-quadruplex and i-motif structures in the c-Myc promoter: implications for drug targeting and control of gene expression. *Journal of medicinal chemistry* **52**, 2863-2874, doi:10.1021/jm900055s (2009).

141 Li, W., Miyoshi, D., Nakano, S. & Sugimoto, N. Structural competition involving G-quadruplex DNA and its complement. *Biochemistry* **42**, 11736-11744, doi:10.1021/bi034168j (2003).

142 Li, W., Wu, P., Ohmichi, T. & Sugimoto, N. Characterization and thermodynamic properties of quadruplex/duplex competition. *FEBS Lett* **526**, 77-81 (2002).

143 Wu, Y. *et al.* Stabilization of VEGF G-quadruplex and inhibition of angiogenesis by quindoline derivatives. *Biochim Biophys Acta* **1840**, 2970-2977, doi:10.1016/j.bbagen.2014.06.002 (2014).

144 Connor, A. C., Frederick, K. A., Morgan, E. J. & McGown, L. B. Insulin capture by an insulin-linked polymorphic region G-quadruplex DNA oligonucleotide. *Journal of the American Chemical Society* **128**, 4986-4991, doi:10.1021/ja056097c (2006).

145 Gonzalez, V., Guo, K., Hurley, L. & Sun, D. Identification and characterization of nucleolin as a c-myc G-quadruplex-binding protein. *The Journal of biological chemistry* **284**, 23622-23635, doi:10.1074/jbc.M109.018028 (2009).

146 De, S. & Michor, F. DNA secondary structures and epigenetic determinants of cancer genome evolution. *Nature structural & molecular biology* **18**, 950-955, doi:10.1038/nsmb.2089 (2011).

147 Lambert, S. & Carr, A. M. Impediments to replication fork movement: stabilisation, reactivation and genome instability. *Chromosoma* **122**, 33-45, doi:10.1007/s00412-013-0398-9 (2013).

148 Shahid, R., Bugaut, A. & Balasubramanian, S. The BCL-2 5' untranslated region contains an RNA G-quadruplex-forming motif that modulates protein expression. *Biochemistry* **49**, 8300-8306, doi:10.1021/bi100957h.

149 Arora, A. & Suess, B. An RNA G-quadruplex in the 3' UTR of the proto-oncogene PIM1 represses translation. *RNA Biol* **8**, 802-805, doi:10.4161/rna.8.5.16038 (2011).

150 Westmark, C. J. & Malter, J. S. FMRP mediates mGluR5-dependent translation of amyloid precursor protein. *PLoS Biol* **5**, e52, doi:10.1371/journal.pbio.0050052 (2007).

151 Schaeffer, C. *et al.* The fragile X mental retardation protein binds specifically to its mRNA via a purine quartet motif. *EMBO J* **20**, 4803-4813, doi:10.1093/emboj/20.17.4803 (2001).

152 Kumari, S., Bugaut, A. & Balasubramanian, S. 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, doi:10.1021/bi8010797 (2008).

153 Bonnal, S. *et al.* A single internal ribosome entry site containing a G quartet RNA structure drives fibroblast growth factor 2 gene expression at four alternative translation initiation codons. *The Journal of biological chemistry* **278**, 39330-39336, doi:10.1074/jbc.M305580200 (2003).

154 Morris, M. J., Negishi, Y., Pazsint, C., Schonhoft, J. D. & Basu, S. An RNA G-quadruplex is essential for cap-independent translation initiation in human VEGF IRES. *Journal of the American Chemical Society* **132**, 17831-17839, doi:10.1021/ja106287x (2010).

155 Brazda, V., Haronikova, L., Liao, J. C. & Fojta, M. DNA and RNA quadruplex-binding proteins. *Int J Mol Sci* **15**, 17493-17517, doi:10.3390/ijms151017493 (2014).

156 Zaug, A. J., Podell, E. R. & Cech, T. R. Human POT1 disrupts telomeric G-quadruplexes allowing telomerase extension in vitro. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 10864-10869, doi:10.1073/pnas.0504744102 (2005).

157 Ray, S., Bandaria, J. N., Qureshi, M. H., Yildiz, A. & Balci, H. G-quadruplex formation in telomeres enhances POT1/TPP1 protection against RPA binding. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 2990-2995, doi:10.1073/pnas.1321436111 (2014).

158 Biffi, G., Tannahill, D. & Balasubramanian, S. An intramolecular G-quadruplex structure is required for binding of telomeric repeat-containing RNA to the telomeric protein TRF2. *Journal of the American Chemical Society* **134**, 11974-11976, doi:10.1021/ja305734x (2012).

159 Hudson, J. S., Ding, L., Le, V., Lewis, E. & Graves, D. Recognition and binding of human telomeric G-quadruplex DNA by unfolding protein 1. *Biochemistry* **53**, 3347-3356, doi:10.1021/bi500351u (2014).

160 Wang, F. *et al.* Telomere- and telomerase-interacting protein that unfolds telomere G-quadruplex and promotes telomere extension in mammalian cells. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 20413-20418, doi:10.1073/pnas.1200232109 (2012).

161 Soldatenkov, V. A., Vetcher, A. A., Duka, T. & Ladame, S. First evidence of a functional interaction between DNA quadruplexes and poly(ADP-ribose) polymerase-1. *ACS Chem Biol* **3**, 214-219, doi:10.1021/cb700234f (2008).

162 Cogoi, S., Paramasivam, M., Membrino, A., Yokoyama, K. K. & Xodo, L. E. The KRAS promoter responds to Myc-associated zinc finger and poly(ADP-ribose) polymerase 1 proteins, which recognize a critical quadruplex-forming GA-element. *The Journal of biological chemistry* **285**, 22003-22016, doi:10.1074/jbc.M110.101923 (2010).

163 Dempsey, L. A., Sun, H., Hanakahi, L. A. & Maizels, N. G4 DNA binding by LR1 and its subunits, nucleolin and hnRNP D, A role for G-G pairing in immunoglobulin switch recombination. *The Journal of biological chemistry* **274**, 1066-1071 (1999).

164 Chakraborty, P. & Grosse, F. Human DHX9 helicase preferentially unwinds RNA-containing displacement loops (R-loops) and G-quadruplexes. *DNA repair* **10**, 654-665, doi:S1568-7864(11)00110-8 [pii] 10.1016/j.dnarep.2011.04.013.

165 Vaughn, J. P. *et al.* The DEXH protein product of the DHX36 gene is the major source of tetramolecular quadruplex G4-DNA resolving activity in HeLa cell lysates. *The Journal of biological chemistry* **280**, 38117-38120, doi:C500348200 [pii] 10.1074/jbc.C500348200 (2005).

166 von Hacht, A. *et al.* Identification and characterization of RNA guanine-quadruplex binding proteins. *Nucleic acids research* **42**, 6630-6644, doi:10.1093/nar/gku290 (2014).

167 Mohaghegh, P., Karow, J. K., Brosh, R. M., Jr., Bohr, V. A. & Hickson, I. D. The Bloom's and Werner's syndrome proteins are DNA structure-specific helicases. *Nucleic acids research* **29**, 2843-2849 (2001).

168 Huber, M. D., Lee, D. C. & Maizels, N. G4 DNA unwinding by BLM and Sgs1p: substrate specificity and substrate-specific inhibition. *Nucleic acids research* **30**, 3954-3961 (2002).

169 London, T. B. *et al.* FANCJ is a structure-specific DNA helicase associated with the maintenance of genomic G/C tracts. *The Journal of biological chemistry* **283**, 36132-36139, doi:10.1074/jbc.M808152200 (2008).

170 Chung, W. H. To peep into Pif1 helicase: multifaceted all the way from genome stability to repair-associated DNA synthesis. *Journal of microbiology* **52**, 89-98, doi:10.1007/s12275-014-3524-3 (2014).

171 Johnson, J. E., Cao, K., Ryvkin, P., Wang, L. S. & Johnson, F. B. Altered gene expression in the Werner and Bloom syndromes is associated with sequences having G-quadruplex forming potential. *Nucleic acids research* **38**, 1114-1122, doi:10.1093/nar/gkp1103 (2010).

172 Paeschke, K., Capra, J. A. & Zakian, V. A. DNA replication through G-quadruplex motifs is promoted by the Saccharomyces cerevisiae Pif1 DNA helicase. *Cell* **145**, 678-691, doi:10.1016/j.cell.2011.04.015 (2011).

173 Sarkies, P. *et al.* FANCJ coordinates two pathways that maintain epigenetic stability at G-quadruplex DNA. *Nucleic acids research* **40**, 1485-1498, doi:10.1093/nar/gkr868 (2012).

174 Ren, J. & Chaires, J. B. Sequence and structural selectivity of nucleic acid binding ligands. *Biochemistry* **38**, 16067-16075, doi:bi992070s [pii] (1999).

175 Naasani, I., Seimiya, H., Yamori, T. & Tsuruo, T. FJ5002: a potent telomerase inhibitor identified by exploiting the disease-oriented screening program with COMPARE analysis. *Cancer research* **59**, 4004-4011 (1999).

176 Izbicka, E. *et al.* Effects of cationic porphyrins as G-quadruplex interactive agents in human tumor cells. *Cancer research* **59**, 639-644 (1999).

177 Granotier, C. *et al.* Preferential binding of a G-quadruplex ligand to human chromosome ends. *Nucleic acids research* **33**, 4182-4190, doi:10.1093/nar/gki722 (2005).

178 Cuenca, F. *et al.* Tri- and tetra-substituted naphthalene diimides as potent G-quadruplex ligands. *Bioorganic & medicinal chemistry letters* **18**, 1668-1673, doi:10.1016/j.bmcl.2008.01.050 (2008).

179 Chang, C. C. *et al.* A fluorescent carbazole derivative: high sensitivity for quadruplex DNA. *Analytical chemistry* **75**, 6177-6183, doi:10.1021/ac034789i (2003).

180 Gowan, S. M., Heald, R., Stevens, M. F. & Kelland, L. R. Potent inhibition of telomerase by small-molecule pentacyclic acridines capable of interacting with G-quadruplexes. *Molecular pharmacology* **60**, 981-988 (2001).

181 Bidzinska, J., Cimino-Reale, G., Zaffaroni, N. & Folini, M. G-quadruplex structures in the human genome as novel therapeutic targets. *Molecules* **18**, 12368-12395, doi:molecules181012368 [pii] 10.3390/molecules181012368.

182 Shalaby, T. *et al.* G-quadruplexes as potential therapeutic targets for embryonal tumors. *Molecules* **18**, 12500-12537, doi:10.3390/molecules181012500 (2013).

183 Ali, A. & Bhattacharya, S. DNA binders in clinical trials and chemotherapy. *Bioorganic & medicinal chemistry* **22**, 4506-4521, doi:10.1016/j.bmc.2014.05.030 (2014).

184 Schaffitzel, C. *et al.* In vitro generated antibodies specific for telomeric guanine-quadruplex DNA react with Stylonychia lemnae macronuclei. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 8572-8577, doi:10.1073/pnas.141229498 (2001).

185 Paeschke, K. *et al.* Telomerase recruitment by the telomere end binding protein-beta facilitates G-quadruplex DNA unfolding in ciliates. *Nature structural & molecular biology* **15**, 598-604, doi:10.1038/nsmb.1422 (2008).

186 Biffi, G., Di Antonio, M., Tannahill, D. & Balasubramanian, S. Visualization and selective chemical targeting of RNA G-quadruplex structures in the cytoplasm of human cells. *Nature chemistry* **6**, 75-80, doi:10.1038/nchem.1805 (2014).

187 Henderson, A. *et al.* Detection of G-quadruplex DNA in mammalian cells. *Nucleic acids research* **42**, 860-869, doi:10.1093/nar/gkt957 (2014).

188 Laguerre, A. *et al.* Visualization of RNA-Quadruplexes in Live Cells. *Journal of the American Chemical Society* **137**, 8521-8525, doi:10.1021/jacs.5b03413 (2015).

189 Huppert, J. L., Bugaut, A., Kumari, S. & Balasubramanian, S. G-quadruplexes: the beginning and end of UTRs. *Nucleic acids research* **36**, 6260-6268, doi:gkn511 [pii] 10.1093/nar/gkn511 (2008).

190 Arora, A. *et al.* Inhibition of translation in living eukaryotic cells by an RNA G-quadruplex motif. *Rna* **14**, 1290-1296, doi:rna.1001708 [pii] 10.1261/rna.1001708 (2008).

191 Morris, M. J. & Basu, S. An unusually stable G-quadruplex within the 5'-UTR of the MT3 matrix metalloproteinase mRNA represses translation in eukaryotic cells. *Biochemistry* **48**, 5313-5319, doi:10.1021/bi900498z (2009).

192 Gomez, D. *et al.* A G-quadruplex structure within the 5'-UTR of TRF2 mRNA represses translation in human cells. *Nucleic acids research* **38**, 7187-7198, doi:gkq563 [pii] 10.1093/nar/gkq563.

193 Brown, V. *et al.* Microarray identification of FMRP-associated brain mRNAs and altered mRNA translational profiles in fragile X syndrome. *Cell* **107**, 477-487, doi:S0092-8674(01)00568-2 [pii] (2001).

194 Lu, R. *et al.* The fragile X protein controls microtubule-associated protein 1B translation and microtubule stability in brain neuron development. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 15201-15206, doi:0404995101 [pii] 10.1073/pnas.0404995101 (2004).

195 Castets, M. *et al.* FMRP interferes with the Rac1 pathway and controls actin cytoskeleton dynamics in murine fibroblasts. *Hum Mol Genet* **14**, 835-844, doi:ddi077 [pii] 10.1093/hmg/ddi077 (2005).

196 Creacy, S. D. *et al.* G4 resolvase 1 binds both DNA and RNA tetramolecular quadruplex with high affinity and is the major source of tetramolecular quadruplex G4-DNA and G4-RNA resolving activity in HeLa cell lysates. *The Journal of biological chemistry* **283**, 34626-34634, doi:M806277200 [pii] 10.1074/jbc.M806277200 (2008).

197 Chakraborty, P. & Grosse, F. Human DHX9 helicase preferentially unwinds RNA-containing displacement loops (R-loops) and G-quadruplexes. *DNA repair* **10**, 654-665, doi:10.1016/j.dnarep.2011.04.013 (2011).

198 Pisareva, V. P., Pisarev, A. V., Komar, A. A., Hellen, C. U. & Pestova, T. V. Translation initiation on mammalian mRNAs with structured 5'UTRs requires DExH-box protein DHX29. *Cell* **135**, 1237-1250, doi:S0092-8674(08)01374-3 [pii] 10.1016/j.cell.2008.10.037 (2008).

199 Soto-Rifo, R. *et al.* DEAD-box protein DDX3 associates with eIF4F to promote translation of selected mRNAs. *EMBO J* **31**, 3745-3756, doi:emboj2012220 [pii] 10.1038/emboj.2012.220.

200 Parsyan, A. *et al.* mRNA helicases: the tacticians of translational control. *Nature reviews. Molecular cell biology* **12**, 235-245, doi:nrm3083 [pii] 10.1038/nrm3083.

201 Bugaut, A., Rodriguez, R., Kumari, S., Hsu, S. T. & Balasubramanian, S. Small molecule-mediated inhibition of translation by targeting a native RNA G-quadruplex. *Organic & biomolecular chemistry* **8**, 2771-2776, doi:10.1039/c002418j (2010).

202 Gomez, D. *et al.* A G-quadruplex structure within the 5'-UTR of TRF2 mRNA represses translation in human cells. *Nucleic acids research* **38**, 7187-7198, doi:10.1093/nar/gkq563 (2010).

203 Halder, K., Largy, E., Benzler, M., Teulade-Fichou, M. P. & Hartig, J. S. Efficient suppression of gene expression by targeting 5'-UTR-based RNA quadruplexes with bisquinolinium compounds. *Chembiochem : a European journal of chemical biology* **12**, 1663-1668, doi:10.1002/cbic.201100228 (2011).

204 De Cian, A., Delemos, E., Mergny, J. L., Teulade-Fichou, M. P. & Monchaud, D. Highly efficient G-quadruplex recognition by bisquinolinium compounds. *Journal of the American Chemical Society* **129**, 1856-1857, doi:10.1021/ja067352b (2007).

205 Chung, W. J., Heddi, B., Hamon, F., Teulade-Fichou, M. P. & Phan, A. T. Solution structure of a G-quadruplex bound to the bisquinolinium compound Phen-DC(3). *Angewandte Chemie* **53**, 999-1002, doi:10.1002/anie.201308063 (2014).

206 Xu, Y., Kaminaga, K. & Komiyama, M. G-quadruplex formation by human telomeric repeats-containing RNA in Na+ solution. *Journal of the American Chemical Society* **130**, 11179-11184, doi:10.1021/ja8031532 (2008).

207 Schoeftner, S. & Blasco, M. A. Developmentally regulated transcription of mammalian telomeres by DNA-dependent RNA polymerase II. *Nature cell biology* **10**, 228-236, doi:10.1038/ncb1685 (2008).

208 Collie, G. W., Haider, S. M., Neidle, S. & Parkinson, G. N. A crystallographic and modelling study of a human telomeric RNA (TERRA) quadruplex. *Nucleic acids research* **38**, 5569-5580, doi:10.1093/nar/gkq259 (2010).

209 Hafner, M. *et al.* Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* **141**, 129-141, doi:10.1016/j.cell.2010.03.009 (2010).

210 Favre, A. *et al.* 4-Thiouridine photosensitized RNA-protein crosslinking in mammalian cells. *Biochemical and biophysical research communications* **141**, 847-854 (1986).

211 Meisenheimer, K. M. & Koch, T. H. Photocross-linking of nucleic acids to associated proteins. *Critical reviews in biochemistry and molecular biology* **32**, 101-140, doi:10.3109/10409239709108550 (1997).

212 Lopez de Silanes, I., Stagno d'Alcontres, M. & Blasco, M. A. TERRA transcripts are bound by a complex array of RNA-binding proteins. *Nature communications* **1**, 33, doi:10.1038/ncomms1032 (2010).

213 Scheibe, M. *et al.* Quantitative interaction screen of telomeric repeat-containing RNA reveals novel TERRA regulators. *Genome research* **23**, 2149-2157, doi:10.1101/gr.151878.112 (2013).

214 Mathad, R. I., Hatzakis, E., Dai, J. & Yang, D. c-MYC promoter G-quadruplex formed at the 5'-end of NHE III1 element: insights into biological relevance and parallel-stranded G-quadruplex stability. *Nucleic Acids Res* **39**, 9023-9033, doi:gkr612 [pii] 10.1093/nar/gkr612.

215 Sun, D. *et al.* The proximal promoter region of the human vascular endothelial growth factor gene has a G-quadruplex structure that can be targeted by G-quadruplex-interactive agents. *Mol Cancer Ther* **7**, 880-889, doi:7/4/880 [pii] 10.1158/1535-7163.MCT-07-2119 (2008).

216 Maxam, A. M. & Gilbert, W. A new method for sequencing DNA. *Proceedings of the National Academy of Sciences of the United States of America* **74**, 560-564 (1977).

217 Kim, J. C. & Mirkin, S. M. The balancing act of DNA repeat expansions. *Current opinion in genetics & development* **23**, 280-288, doi:10.1016/j.gde.2013.04.009 (2013).

218 Mirkin, S. M. Expandable DNA repeats and human disease. *Nature* **447**, 932-940, doi:10.1038/nature05977 (2007).

219 Margolis, R. L. & Ross, C. A. Expansion explosion: new clues to the pathogenesis of repeat expansion neurodegenerative diseases. *Trends in molecular medicine* **7**, 479-482 (2001).

220 Liquori, C. L. *et al.* Myotonic dystrophy type 2 caused by a CCTG expansion in intron 1 of ZNF9. *Science* **293**, 864-867, doi:10.1126/science.1062125 (2001).

221 Matsuura, T. *et al.* Large expansion of the ATTCT pentanucleotide repeat in spinocerebellar ataxia type 10. *Nature genetics* **26**, 191-194, doi:10.1038/79911 (2000).

222 Friedman, R. A. & Honig, B. A free energy analysis of nucleic acid base stacking in aqueous solution. *Biophysical journal* **69**, 1528-1535, doi:10.1016/S0006-3495(95)80023-8 (1995).

223 Krueger, A., Protozanova, E. & Frank-Kamenetskii, M. D. Sequence-dependent base pair opening in DNA double helix. *Biophysical journal* **90**, 3091-3099, doi:10.1529/biophysj.105.078774 (2006).

224 Moyzis, R. K. *et al.* A highly conserved repetitive DNA sequence, (TTAGGG)n, present at the telomeres of human chromosomes. *Proceedings of the National Academy of Sciences of the United States of America* **85**, 6622-6626 (1988).

225 Wright, W. E., Tesmer, V. M., Huffman, K. E., Levene, S. D. & Shay, J. W. Normal human chromosomes have long G-rich telomeric overhangs at one end. *Genes & development* **11**, 2801-2809 (1997).

226 Griffith, J. D. *et al.* Mammalian telomeres end in a large duplex loop. *Cell* **97**, 503-514, doi:S0092-8674(00)80760-6 [pii] (1999).

227 Williamson, J. R. G-quartet structures in telomeric DNA. *Annu Rev Biophys Biomol Struct* **23**, 703-730, doi:10.1146/annurev.bb.23.060194.003415 (1994).

228 de Lange, T. T-loops and the origin of telomeres. *Nature reviews. Molecular cell biology* **5**, 323-329, doi:10.1038/nrm1359 nrm1359 [pii] (2004).

229 Hayflick, L. Antecedents of cell aging research. *Exp Gerontol* **24**, 355-365 (1989).

230 Hayflick, L. & Moorhead, P. S. The serial cultivation of human diploid cell strains. *Exp Cell Res* **25**, 585-621 (1961).

231 Shay, J. W. & Wright, W. E. Senescence and immortalization: role of telomeres and telomerase. *Carcinogenesis* **26**, 867-874, doi:bgh296 [pii] 10.1093/carcin/bgh296 (2005).

232 Greider, C. W. & Blackburn, E. H. A telomeric sequence in the RNA of Tetrahymena telomerase required for telomere repeat synthesis. *Nature* **337**, 331-337, doi:10.1038/337331a0 (1989).

233 Greider, C. W. & Blackburn, E. H. Identification of a specific telomere terminal transferase activity in Tetrahymena extracts. *Cell* **43**, 405-413, doi:0092-8674(85)90170-9 [pii] (1985).

234 Kim, N. W. *et al.* Specific association of human telomerase activity with immortal cells and cancer. *Science* **266**, 2011-2015 (1994).

235 Shay, J. W. & Bacchetti, S. A survey of telomerase activity in human cancer. *Eur J Cancer* **33**, 787-791, doi:S0959-8049(97)00062-2 [pii] 10.1016/S0959-8049(97)00062-2 (1997).

236 Zahler, A. M., Williamson, J. R., Cech, T. R. & Prescott, D. M. Inhibition of telomerase by G-quartet DNA structures. *Nature* **350**, 718-720, doi:10.1038/350718a0 (1991).

237 Salvati, E. *et al.* Telomere damage induced by the G-quadruplex ligand RHPS4 has an antitumor effect. *J Clin Invest* **117**, 3236-3247, doi:10.1172/JCI32461 (2007).

238 Rizzo, A. *et al.* Stabilization of quadruplex DNA perturbs telomere replication leading to the activation of an ATR-dependent ATM signaling pathway. *Nucleic acids research* **37**, 5353-5364, doi:gkp582 [pii] 10.1093/nar/gkp582 (2009).

239 Monchaud, D. & Teulade-Fichou, M. P. A hitchhiker's guide to G-quadruplex ligands. *Organic & biomolecular chemistry* **6**, 627-636, doi:10.1039/b714772b (2008).

240 Neidle, S. Human telomeric G-quadruplex: the current status of telomeric G-quadruplexes as therapeutic targets in human cancer. *FEBS J* **277**, 1118-1125, doi:EJB7463 [pii] 10.1111/j.1742-4658.2009.07463.x.

241 Kim, M. Y., Vankayalapati, H., Shin-Ya, K., Wierzba, K. & Hurley, L. H. Telomestatin, a potent telomerase inhibitor that interacts quite specifically with the human telomeric intramolecular g-quadruplex. *Journal of the American Chemical Society* **124**, 2098-2099, doi:ja017308q [pii] (2002).

242 Gomez, D. *et al.* The G-quadruplex ligand telomestatin inhibits POT1 binding to telomeric sequences in vitro and induces GFP-POT1 dissociation from telomeres in human cells. *Cancer research* **66**, 6908-6912, doi:66/14/6908 [pii] 10.1158/0008-5472.CAN-06-1581 (2006).

243 Gomez, D. *et al.* Telomestatin-induced telomere uncapping is modulated by POT1 through G-overhang extension in HT1080 human tumor cells. *The Journal of biological chemistry* **281**, 38721-38729, doi:M605828200 [pii] 10.1074/jbc.M605828200 (2006).

244 Hampel, S. M. *et al.* Mechanism of the antiproliferative activity of some naphthalene diimide G-quadruplex ligands. *Molecular pharmacology* **83**, 470-480, doi:mol.112.081075 [pii] 10.1124/mol.112.081075.

245 Balasubramanian, S., Hurley, L. H. & Neidle, S. Targeting G-quadruplexes in gene promoters: a novel anticancer strategy? *Nat Rev Drug Discov* **10**, 261-275, doi:nrd3428 [pii] 10.1038/nrd3428.

246 Cesare, A. J. & Reddel, R. R. Alternative lengthening of telomeres: models, mechanisms and implications. *Nature reviews. Genetics* **11**, 319-330, doi:nrg2763 [pii] 10.1038/nrg2763.

247 Temime-Smaali, N. *et al.* The G-quadruplex ligand telomestatin impairs binding of topoisomerase IIIalpha to G-quadruplex-forming oligonucleotides and uncaps telomeres in ALT cells. *PLoS One* **4**, e6919, doi:10.1371/journal.pone.0006919 (2009).

248 Benz, A., Singh, V., Mayer, T. U. & Hartig, J. S. Identification of novel quadruplex ligands from small molecule libraries by FRET-based high-throughput screening. *Chembiochem : a European journal of chemical biology* **12**, 1422-1426, doi:10.1002/cbic.201100094.

249 Ueyama, H., Takagi, M. & Takenaka, S. A novel potassium sensing in aqueous media with a synthetic oligonucleotide derivative. Fluorescence resonance energy transfer associated with Guanine quartet-potassium ion complex formation. *Journal of the American Chemical Society* **124**, 14286-14287, doi:ja026892f [pii] (2002).

250 Benz, A. *Artifizielle G-Quadruplexe und Identifizierung von neuen G-Quadruplexliganden*. (2011).

251 Zhou, Q. *et al.* Screening potential antitumor agents from natural plant extracts by G-quadruplex recognition and NMR methods. *Angewandte Chemie* **47**, 5590-5592, doi:10.1002/anie.200800913 (2008).

252 Dhamodharan, V., Harikrishna, S., Jagadeeswaran, C., Halder, K. & Pradeepkumar, P. I. Selective G-quadruplex DNA stabilizing agents based on bisquinolinium and bispyridinium derivatives of 1,8-naphthyridine. *The Journal of organic chemistry* **77**, 229-242, doi:10.1021/jo201816g (2012).

253 Koirala, D. *et al.* A single-molecule platform for investigation of interactions between G-quadruplexes and small-molecule ligands. *Nature chemistry* **3**, 782-787, doi:10.1038/nchem.1126 (2011).

254 Yee, C., Krishnan-Hewlett, I., Baker, C. C., Schlegel, R. & Howley, P. M. Presence and expression of human papillomavirus sequences in human cervical carcinoma cell lines. *The American journal of pathology* **119**, 361-366 (1985).

255 Chen, T. R. Re-evaluation of HeLa, HeLa S3, and HEp-2 karyotypes. *Cytogenetics and cell genetics* **48**, 19-24 (1988).

256 Laemmli, U. K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680-685 (1970).

9. Appendix

9.1 Sequences of the plasmids pcDNA5/FRT/TO-eGFP and pcDNA5/FRT/TO-mCherry

9.1.1 pcDNA5/FRT/TO-eGFP

Within the following plasmid sequence the TATA-box is highlighted in green, the 5´ UTR of the mRNA is shaded in gray and the translation start is shown in red.

GACGGATCGGGAGATCTCCCGATCCCCTATGGTGCACTCTCAGTACAATCTGCTCTGATG
CCGCATAGTTAAGCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTAGTGCG
CGAGCAAAATTTAAGCTACAACAAGGCAAGGCTTGACCGACAATTGCATGAAGAATCTGCT
TAGGGTTAGGCGTTTTGCGCTGCTTCGCGATGTACGGGCCAGATATACGCGTTGACATTG
ATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGA
GTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCG
CCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGAC
GTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATAT
GCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCA
GTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTA
CCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGG
GATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACG
GGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGT
ACGGTGGGAGGTCTATATAAGCAGAGCTCTCCCTATCAGTGATAGAGATCTCCCTATCAGT
GATAGAGATCGTCGACGAGCTCGTTTAGTGAACCGTCAGATCGCCTGGAGACGCCATCCA
CGCTGTTTTGACCTCCATAGAAGACACCGGGACCGATCCAGCCTCCGGACTCTAGCGTTT
AAACttaagATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCG
AGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGAT
GCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCC
TGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGAC
CACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGC
ACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGC
GACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATC
CTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAG
CAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTG
CAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCC
CGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCG
ATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGC
TGTACAAGGGCCGGCCCAGGCCTCTCGAGggcgcgccgggccCGTTTAAACCCGCTGATCAGC
CTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTG
ACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATT
GTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAG
GATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGCTTCTGAGGC
GGAAAGAACCAGCTGGGGCTCTAGGGGGTATCCCCACGCGCCCTGTAGCGGCGCATTAA
GCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCG
CCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAG
CTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAA
AAAACTTGATTAGGGTGATGGTTCACGTACCTAGAAGTTCCTATTCCGAAGTTCCTATTCTC

TAGAAAGTATAGGAACTTCCTTGGCCAAAAAGCCTGAACTCACCGCGACGTCTGTCGAGAA
GTTTCTGATCGAAAAGTTCGACAGCGTCTCCGACCTGATGCAGCTCTCGGAGGGCGAAGA
ATCTCGTGCTTTCAGCTTCGATGTAGGAGGGCGTGGATATGTCCTGCGGGTAAATAGCTG
CGCCGATGGTTTCTACAAAGATCGTTATGTTTATCGGCACTTTGCATCGGCCGCGCTCCCG
ATTCCGGAAGTGCTTGACATTGGGGAATTCAGCGAGAGCCTGACCTATTGCATCTCCCGC
CGTGCACAGGGTGTCACGTTGCAAGACCTGCCTGAAACCGAACTGCCCGCTGTTCTGCAG
CCGGTCGCGGAGGCCATGGATGCGATCGCTGCGGCCGATCTTAGCCAGACGAGCGGGTT
CGGCCCATTCGGACCGCAAGGAATCGGTCAATACACTACATGGCGTGATTTCATATGCGC
GATTGCTGATCCCCATGTGTATCACTGGCAAACTGTGATGGACGACACCGTCAGTGCGTC
CGTCGCGCAGGCTCTCGATGAGCTGATGCTTTGGGCCGAGGACTGCCCCGAAGTCCGGC
ACCTCGTGCACGCGGATTTCGGCTCCAACAATGTCCTGACGGACAATGGCCGCATAACAG
CGGTCATTGACTGGAGCGAGGCGATGTTCGGGGATTCCCAATACGAGGTCGCCAACATCT
TCTTCTGGAGGCCGTGGTTGGCTTGTATGGAGCAGCAGACGCGCTACTTCGAGCGGAGG
CATCCGGAGCTTGCAGGATCGCCGCGGCTCCGGGCGTATATGCTCCGCATTGGTCTTGAC
CAACTCTATCAGAGCTTGGTTGACGGCAATTTCGATGATGCAGCTTGGGCGCAGGGTCGA
TGCGACGCAATCGTCCGATCCGGAGCCGGGACTGTCGGGCGTACACAAATCGCCCGCAG
AAGCGCGGCCGTCTGGACCGATGGCTGTGTAGAAGTACTCGCCGATAGTGGAAACCGAC
GCCCCAGCACTCGTCCGAGGGCAAAGGAATAGCACGTACTACGAGATTTCGATTCCACCG
CCGCCTTCTATGAAAGGTTGGGCTTCGGAATCGTTTTCCGGGACGCCGGCTGGATGATCC
TCCAGCGCGGGGATCTCATGCTGGAGTTCTTCGCCCACCCCAACTTGTTTATTGCAGCTTA
TAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCACTGCA
TTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTATCATGTCTGTATACCGTCGACCT
CTAGCTAGAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTC
ACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGA
GTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGT
CGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGG
CGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCG
GTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGA
AAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTG
GCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGA
GGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCG
TGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGG
GAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCG
CTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCG
GTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCA
CTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGT
GGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGT
TACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGG
TGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTT
TGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGT
CATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATC
AATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCAC
CTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATA
ACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCA
CGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGA
AGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAG
TAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGT
GTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTT
ACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCA
GAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACT
GTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAG
AATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGC
CACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTC

AAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCT
TCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCG
CAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATAT
TATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAA
AATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTC

9.1.2 pcDNA5/FRT/TO-mCherry

GACGGATCGGGAGATCTCCCGATCCCCTATGGTGCACTCTCAGTACAATCTGCTCTGATG
CCGCATAGTTAAGCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTAGTGCG
CGAGCAAAATTTAAGCTACAACAAGGCAAGGCTTGACCGACAATTGCATGAAGAATCTGCT
TAGGGTTAGGCGTTTTGCGCTGCTTCGCGATGTACGGGCCAGATATACGCGTTGACATTG
ATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGA
GTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCG
CCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGAC
GTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATAT
GCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCA
GTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTA
CCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGG
GATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACG
GGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGT
ACGGTGGGAGGTCTATATAAGCAGAGCTCTCCCTATCAGTGATAGAGATCTCCCTATCAGT
GATAGAGATCGTCGACGAGCTCGTTTAGTGAACCGTCAGATCGCCTGGAGACGCCATCCA
CGCTGTTTTGACCTCCATAGAAGACACCGGGACCGATCCAGCCTCCGGACTCTAGCGTTT
AAACttaagATGGTGAGCAAGGGCGAGGAGGATAACATGGCCATCATCAAGGAGTTCATGCG
CTTCAAGGTGCACATGGAGGGCTCCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGG
GCGAGGGCCGCCCCTACGAGGGCACCCAGACCGCCAAGCTGAAGGTGACCAAGGGTGG
CCCCCTGCCCTTCGCCTGGGACATCCTGTCCCCTCAGTTCATGTACGGCTCCAAGGCCTA
CGTGAAGCACCCCGCCGACATCCCCGACTACTTGAAGCTGTCCTTCCCCGAGGGCTTCAA
GTGGGAGCGCGTGATGAACTTCGAGGACGGCGGCGTGGTGACCGTGACCCAGGACTCCT
CCCTGCAGGACGGCGAGTTCATCTACAAGGTGAAGCTGCGCGGCACCAACTTCCCCTCCG
ACGGCCCCGTAATGCAGAAGAAGACCATGGGCTGGGAGGCCTCCTCCGAGCGGATGTAC
CCCGAGGACGGCGCCCTGAAGGGCGAGATCAAGCAGAGGCTGAAGCTGAAGGACGGCG
GCCACTACGACGCTGAGGTCAAGACCACCTACAAGGCCAAGAAGCCCGTGCAGCTGCCC
GGCGCCTACAACGTCAACATCAAGTTGGACATCACCTCCCACAACGAGGACTACACCATC
GTGGAACAGTACGAACGCGCCGAGGGCCGCCACTCCACCGGCGGCATGGACGAGCTGTA
CAAGGGCCGGCCAGAATTCAAGGCCTCTCGAGCCTCTAGTggcgcgccgggccCGTTTAAACC
CGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCG
TGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAAT
TGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAG
CAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGG
CTTCTGAGGCGGAAAGAACCAGCTGGGGCTCTAGGGGGTATCCCCACGCGCCCTGTAGC
GGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAG
CGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTT
CCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACC
TCGACCCCAAAAAACTTGATTAGGGTGATGGTTCACGTACCTAGAAGTTCCTATTCCGAAG
TTCCTATTCTCTAGAAAGTATAGGAACTTCCTTGGCCAAAAAGCCTGAACTCACCGCGACG
TCTGTCGAGAAGTTTCTGATCGAAAAGTTCGACAGCGTCTCCGACCTGATGCAGCTCTCGG
AGGGCGAAGAATCTCGTGCTTTCAGCTTCGATGTAGGAGGGCGTGGATATGTCCTGCGGG
TAAATAGCTGCGCCGATGGTTTCTACAAAGATCGTTATGTTTATCGGCACTTTGCATCGGC
CGCGCTCCCGATTCCGGAAGTGCTTGACATTGGGGAATTCAGCGAGAGCCTGACCTATTG
CATCTCCCGCCGTGCACAGGGTGTCACGTTGCAAGACCTGCCTGAAACCGAACTGCCCGC

TGTTCTGCAGCCGGTCGCGGAGGCCATGGATGCGATCGCTGCGGCCGATCTTAGCCAGA
CGAGCGGGTTCGGCCCATTCGGACCGCAAGGAATCGGTCAATACACTACATGGCGTGATT
TCATATGCGCGATTGCTGATCCCCATGTGTATCACTGGCAAACTGTGATGGACGACACCGT
CAGTGCGTCCGTCGCGCAGGCTCTCGATGAGCTGATGCTTTGGGCCGAGGACTGCCCCG
AAGTCCGGCACCTCGTGCACGCGGATTTCGGCTCCAACAATGTCCTGACGGACAATGGCC
GCATAACAGCGGTCATTGACTGGAGCGAGGCGATGTTCGGGGATTCCCAATACGAGGTCG
CCAACATCTTCTTCTGGAGGCCGTGGTTGGCTTGTATGGAGCAGCAGACGCGCTACTTCG
AGCGGAGGCATCCGGAGCTTGCAGGATCGCCGCGGCTCCGGGCGTATATGCTCCGCATT
GGTCTTGACCAACTCTATCAGAGCTTGGTTGACGGCAATTTCGATGATGCAGCTTGGGCGC
AGGGTCGATGCGACGCAATCGTCCGATCCGGAGCCGGGACTGTCGGGCGTACACAAATC
GCCCGCAGAAGCGCGGCCGTCTGGACCGATGGCTGTGTAGAAGTACTCGCCGATAGTGG
AAACCGACGCCCCAGCACTCGTCCGAGGGCAAAGGAATAGCACGTACTACGAGATTTCGA
TTCCACCGCCGCCTTCTATGAAAGGTTGGGCTTCGGAATCGTTTTCCGGGACGCCGGCTG
GATGATCCTCCAGCGCGGGGATCTCATGCTGGAGTTCTTCGCCCACCCCAACTTGTTTATT
GCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTTTT
TCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTATCATGTCTGTATACCG
TCGACCTCTAGCTAGAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTA
TCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGC
CTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGA
AACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGT
ATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGG
CGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAAC
GCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGC
GTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCA
AGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGC
TCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCC
CTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGT
CGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTT
ATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCA
GCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAG
TGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGC
CAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAG
CGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGAT
CCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTT
TGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTA
AATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAG
GCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTA
GATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGA
CCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCG
CAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCT
AGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCG
TGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCG
AGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTT
GTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTC
TTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTC
TGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACC
GCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAAC
TCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTG
ATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAAT
GCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTC
AATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTT
AGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTC

10. Publications

Giacani L, Brandt SL, Puray-Chavez M, Reid TB, Godornes C, Molini BJ, Benzler M, Hartig JS, Lukehart SA, Centurion-Lara A **(2012)**. Comparative investigation of the genomic regions involved in antigenic variation of the TprK antigen among treponemal species, subspecies, and strains. *J Bacteriol.* 194(16):4208-25.

Halder K, Benzler M, Hartig JS **(2012)**. Reporter assays for studying quadruplex nucleic acids. *Methods.* 57(1):115-21.

Halder K, Largy E, Benzler M, Teulade-Fichou MP, Hartig JS **(2011)**. Efficient suppression of gene expression by targeting 5'-UTR-based RNA quadruplexes with bisquinolinium compounds. *Chembiochem.* 12(11):1663-8.

11. Danksagung

Mein ganz besonderer Dank gilt meinem Doktorvater Professor Dr. Jörg Hartig. Er ermöglichte mir nicht nur, meine Dissertation in seiner Arbeitsgruppe zu erstellen, sondern er ließ mir auch fortwährend jedwede Unterstützung bei der Bearbeitung meiner Forschungsprojekte zuteilwerden. Mit großem Interesse, Verständnis und Geduld verfolgte er das Werden dieser Arbeit und half jederzeit mit seiner exzellenten wissenschaftlichen Expertise.

Herrn Professor Dr. Andreas Marx danke ich ganz herzlich für die Übernahme des Zweitgutachtens.

Ebenfalls gilt mein Dank Herrn Professor Dr. Martin Scheffner für die Übernahme des Prüfungsvorsitzes meiner mündlichen Promotionsprüfung.

Der Konstanz Research School Chemical Biology (KoRS-CB) danke ich für die Vergabe eines einjährigen Promotionsstipendiums und für die Möglichkeit, dass ich an interessanten und lehrreichen Kursen und Seminaren teilnehmen durfte.

Weiterer Dank gebührt Frau Professor Dr. Elke Deuerling und Herrn Professor Dr. Thomas U. Mayer für viele hilfreiche Diskussionen und Ratschläge im Rahmen meines Thesis Committees der Graduiertenschule KoRS-CB. Darüber hinaus bin ich Herrn Professor Dr. Thomas U. Mayer für die Bereitstellung von Plasmiden und für die Hilfe bei der Durchführung des High-Throughput-Screenings sehr dankbar. Ebenfalls danke ich Herrn Professor Dr. Ewald Daltrozzo für die Bereitstellung seiner Substanzbibliothek und der Proteomics Facility der Universität Konstanz für die massenspektrometrische Analyse.

Nicht vergessen werden darf ein herzliches Dankeschön für die ausgezeichnete Arbeitsatmosphäre innerhalb der gesamten AG Hartig: Astrid Joachimi, Dr. Charlotte Rehm, Dr. Isabelle Holder, Dr. Stefanie Wagner, Dr. Rashi Halder, Dr. Benedikt Klauser, Dr. Athanasios Saragliadis, Dr. Kangkan Halder, Dr. Vijay Singh, Dr. Armin Benz und Dr. Markus Wieland trugen sowohl mit wissenschaftlichen, als auch mit nicht-wissenschaftlichen Gesprächen und Aktivitäten maßgeblich dazu bei, dass ich eine tolle Promotionszeit in Konstanz verbringen durfte. Insbesondere danke ich Dr. Rashi Halder, Astrid Joachimi

und Dr. Kangkan Halder für ihre Hilfestellungen und Ratschläge bei Experimenten zu Beginn meiner Promotion.

Meiner Freundin Dr. Elisabeth Stürner danke ich für ihre Anteilnahme und Unterstützung.

Mein ganz aufrichtiger Dank gebührt meinen Eltern Centa und Dr. Bernhard Benzler. Sie haben mich immer bedingungslos und mit großer Anteilnahme unterstützt und gefördert. Ihnen ist diese Arbeit gewidmet.