



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



Research paper

The evolving world of protein-G-quadruplex recognition: A medicinal chemist's perspective

Claudia Sissi, Barbara Gatto, Manlio Palumbo*

Department of Pharmaceutical Sciences, University of Padova, Via Marzolo 5, 35131 Padova, Italy

ARTICLE INFO

Article history:

Received 26 February 2011

Accepted 20 April 2011

Available online 29 April 2011

Keywords:

G-quadruplex

Drug binding

Nucleic acids

Proteins

ABSTRACT

The physiological and pharmacological role of nucleic acids structures folded into the non canonical G-quadruplex conformation have recently emerged. Their activities are targeted at vital cellular processes including telomere maintenance, regulation of transcription and processing of the pre-messenger or telomeric RNA. In addition, severe conditions like cancer, fragile X syndrome, Bloom syndrome, Werner syndrome and Fanconi anemia are related to genomic defects that involve G-quadruplex forming sequences. In this connection G-quadruplex recognition and processing by nucleic acid directed proteins and enzymes represents a key event to activate or deactivate physiological or pathological pathways. In this review we examine protein-G-quadruplex recognition in physiologically significant conditions and discuss how to possibly exploit the interactions' selectivity for targeted therapeutic intervention.

© 2011 Elsevier Masson SAS. All rights reserved.

1. Introduction

G-Rich DNA or RNA sequences can fold into quadruplexes (G4s) consisting in non canonical structures stabilized by the stacking of G-quartets, in which four guanines are assembled in a planar arrangement by Hoogsteen hydrogen bonding [1,2].

G4s are characterized by the relative direction (parallel vs. antiparallel) of the strands connecting the guanines, by the syn vs. anti glycosyl conformation, by the nature and length of G–G connecting loops, by the intra- vs. inter-molecular nature of the structure and by the number of stacking tetrads. In addition, the conformational properties of G4 are largely influenced by the environmental conditions, in particular by the presence of monovalent ions such as K^+ or Na^+ . These features grant high degree of polymorphism to G4 arrangements and make them suitable to differential recognition. In addition, relatively mild changes in the experimental setting or addition of specific low molecular weight ligands may lead a G-rich sequence to fold/unfold, hence conferring G4s the characteristics of a molecular switch.

Due to these properties, G-quadruplex structures do not only represent novel nucleic acid arrangements worth of scientific investigation, they also emerge as biologically significant due to the presence of G-rich sequences in specific regions of the genome. In particular, guanines are over-represented in the terminal repeating sequences of chromosomes (telomeres) and in promoter regions of genes, especially proto-oncogenes, such as *c-myc*, *c-kit*, *bcl-2*, *VEGF*,

H-ras and *N-ras*, as well as in other human genes. In addition, G4s can be selectively formed at the RNA level further contributing to a modulation of the information flow leading to proteins [3,4]. These findings suggest a role of G4 in controlling biological events including chromosome protection and gene expression [5–8] and foresee several potential biophysical, diagnostic and therapeutic applications for G4. Recent reviews cover this matter thoroughly [9–11].

G4s effect on regulation of physiological (or pathological) processes can be considered in two ways. They can be exploited as targets for protein intervention, thus modulating their basal activity, or, alternatively, they can be potentially used as non-physiologic players to produce desired cellular effects. This latter concept is evidenced by the activity exhibited by some G4 folded sequences (aptamers) toward selected targets. Their strong affinity and specificity make them a sort of nucleic acid-based antibody. For recent reviews see references [7,12–14].

In this review, we will consider the most recent information available on the role played by G4s interactions with proteins, both to unravel naturally occurring recognition and regulation pathways with physiological and pathological relevance and to help identifying yet undisclosed non-physiologic nucleic acid structures (aptamer) able to interfere with biological processes.

In approaching this subject, we wish to remind the following important issues, which are relevant to our discussion, i.e.:

- a specific interaction occurring *in vitro* does not represent a safe proof for the existence of such an interaction *in vivo*;
- G4s are very plastic structures and can easily undergo conformational changes during recognition processes or coexist in

* Corresponding author. Tel./fax: +39 049 8275699.

E-mail address: manlio.palumbo@unipd.it (M. Palumbo).

different structures with diverse binding affinities to a given target;

- at present very few three-dimensional structures of G4-protein complexes are available, which renders molecular modeling and rational targeting a difficult task.

2. In vitro selected protein-G4 recognition

2.1. Protein-directed G4 aptamers

Aptamers are nucleic acids sequences able to appropriately fold to specifically recognize target molecules with specificities reminiscent those exhibited by antibodies. Aptamers include potential applications as anticoagulants, antivirals, antirheumatics, anti-proliferative agents and biosensors. For recent reviews see reference [15].

Although G4 represents just one of several possible options of aptamer architecture, a conspicuous number of G4-based aptamers are reported, the most celebrated one being the thrombin-binding 15mer GGTTGGTGTGGTTGG encompassing two G4 tetrads. Its complex with thrombin provides one of the very few X-ray structures of a G4 bound to a protein thus far available (Fig. 1) [16]. The aptamer is sandwiched between two thrombin molecules and binds (largely through salt bridges) the fibrinogen exosite of one and the heparin binding site in the C-terminal region of the other thrombin molecule. The TGT sequence forming one of the three loops is involved in a hydrophobic cluster near the fibrinogen exosite (Ile²⁴, His⁷¹, Ile⁷⁹, Tyr¹¹⁷), whereas a T belonging to one of

the TT loops associates with His⁹¹, Pro⁹² and Trp²³⁷ of the heparin binding site. Thus, both electrostatic and hydrophobic interactions explain stabilization and recognition of the G4-thrombin complex.

Current advances refer to an improved thrombin aptamer deriving from the incorporation of an unlocked nucleotide monomer (UNA), lacking the 2'-3' C–C bond in the D-ribose moiety, at each position of the 15mer. Substitution of D-ribose with UNA derivative may produce either destabilization or stabilization of the aptamer G4 structure depending upon its position along the DNA chain. This is likely related to a higher degree of local flexibility produced by the open ribose. Particularly effective was the oligomer modified at position 7, which was found to be more efficient than the original aptamer both in protein binding and in anticoagulant activity [17].

To target VEGF, an aptamer raised against the growth factor receptor-binding domain was identified as the 25mer TGTGGGGTG GACGGGCCGGGTAGA showing binding constants in the nanomolar range [18]. Its pharmacological interest is related to the fact that it shares the same target with Macugen (Pegaptanib), a non G4 RNA-based aptamer clinically approved against wet macular degeneration [19].

To further emphasize the potential clinical application of G4 aptamers, a cancer chemotherapeutic G-rich oligonucleotide, the 26mer AS1411 (GGTGGTGGTGGTTGTGGTGGTGGTGG), discovered over 10 years ago, is now undergoing phase II clinical trials for the treatment of AML and renal cell carcinoma [20]. Surprisingly, G4 forming potential seems to represent a general feature leading to cancer-selective antiproliferative activity as shown by the impressive data referring to 11 synthetic G-rich oligonucleotide libraries [21].

Potentially clinically useful G4s can target also non-human proteins. As an example, tetra-end linked oligonucleotides having a cluster of four TG₄T sequences, arranged in a parallel G4 were investigated for anti-HIV activity and showed potencies in the sub-micromolar range. The authors speculate that Arg190, the only conserved residue in the hypervariable V3 loop of GP120 represents a key site for aptamer binding, recognition being generally driven by electrostatic forces [22].

Finally, the field of G4 aptamer recognition can be also profitably exploited for developing biosensing systems for analytical application through the use of logic gate strategies. Very recently, synthetic receptors have been generated by functionalized G4 sequences, specifically an octanucleotide with a G₅ stretch, fit for targeting protein surfaces and for sensing specific intervening contacts. The assembled DNA quadruplexes are modified with three fluorophores exhibiting individual but overlapping emission spectra. This will produce distinctive emission patterns when interacting with different proteins. The observed differences can be used both to monitor protein concentration levels and to investigate analyte combinations producing conformational changes on the probe [23,24].

A list of G4 aptamers recognizing specific protein targets is reported in Table 1.

2.2. G4-directed antibodies and peptides

Using a sort of “mirror” approach, another line of research has led to major advances in the production and development of antibodies and peptide sequences able to bind G4s very efficiently. One of the first examples concerns a high-affinity (sub-nanomolar range) single-chain antibody fragment able to recognize the G4 formed by the *Styloynchia lemnae* telomeric repeat [38]. Peculiarly, this protein is effective in discriminating between parallel or anti-parallel quadruplex conformations formed by the same sequence.

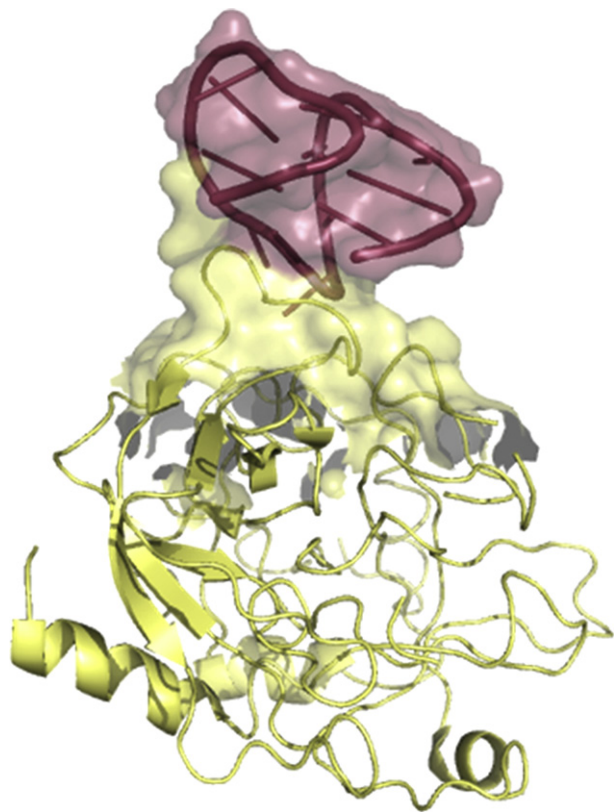


Fig. 1. Three-dimensional structure of the GGTTGGTGTGGTTGG aptamer bound to α -thrombin. PDF 1HUT. The G4 component is highlighted in red, the protein in yellow. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1
G4 structured DNA and RNA aptamers for specific protein recognition.

Aptamer sequence	Target protein	Reference
d(GGTGGTGTGGTTGG)	α -Thrombin	[16]
d(TTGGGGTT)	HIV gp120	[25]
GUGCGGGAUUGAGGGACGAUGGGGAAGU; (GGA) ₄	Prion protein	[26,27]
d(CGGTCGCTCCGTGTGGCTTGGGTGGGTGGCAGTGAC)	Human RNase H1	[28]
d(GGGC) ₄	StatT 3	[29]
AUCGGGAAGGGCUAGGGUGGGUUAU	NF- κ B	[30]
ACGGAUUCGUUAGGGUGGGAUCGGGAAGGGCUACGAACACCGU	NF- κ B receptor activator	[30]
d(GGGGTGGGAGGAGGGT)	HIV Integrase	[31]
d(ACAGGGGTGTGGG) ₂	Insulin	[32]
d(CAGCGTTAGGAAGGGCGTCGAAAGCAGGGTGGG)	HIV reverse transcriptase	[33]
d(GGTGGTGGTGGTGTGGTGGTGGTGG)	NF- κ B, nucleolin	[34]
d(AGCGGGCATATGGTGGTGGTGGTATGGTC)	Coronavirus helicase	[35]
d(TGTGGGGGTGGACGGCCGGGTAGA)	VEGF	[18]
d(AGCGTCAATACCACACGGGGTGGTGGGGGGCTGGTGTCTGGGGTGGGCTAATGGAGCTCGTGGTCAT)	Protein tyrosine phosphatase Shp-2	[36]
d(GGTTN) _n	Polyphosphate kinase 2	[37]

The finding that a single-chain antibody reacted specifically with the macronucleus of *S. lemnae* produced a clear-cut evidence that G4s do indeed represent biologically relevant structures, formed in the telomeres of living cells. Absence of antibody binding in the replication band, the region where telomere replication and elongation is occurring, indicated that reasonably G4s should be resolved during these stages.

This seminal result was developed by exploring the possibility to generate a biologically significant structure selective recognition. Recently a single-chain antibody selected by phage display and competitive selection was reported to be effective for recognition of intramolecular G4s and clearly discriminating between two parallel arrangements found in a protooncogene promoter [39]. Furthermore, the same authors selected a single-chain G4 binding antibody by methods comparable to those reported above, but with negative selection against only duplex DNA. The identified protein showed a broad range of affinities toward selected genomic G4 sequences but no detectable binding to the duplex DNA. This antibody can hence represent a valuable tool to investigate the existence and function of G4 within the genome. In fact, a correlation was found between the effects of the antibody upon gene expression and the occurrence of putative quadruplex sequences at either the promoter or the terminus of genes. Interestingly, production of the antibody in human cells can significantly up- or down-regulate gene expression, which suggests a regulatory role of G4 in this process [40].

Antibodies are not the only proteins selected to recognize G4s. Indeed, chemically modified three zinc finger motifs were identified as G4 effective ligands by screening peptide libraries obtained by phage display [41]. The most effective polypeptide was found to inhibit DNA polymerase and telomerase action on a template containing the telomeric repeat at concentrations in the nanomolar range. Mutational and binding studies showed that any single zinc finger can be replaced with another known finger without significant loss of binding affinity, whereas two simultaneous replacements are deleterious to recognition, finger 2 being a crucial element for quadruplex complexation [42,43].

3. Physiological recognition and processing of G4

3.1. Telomere-related proteins

The possibility of physiological occurrence of G4s was originally located in the telomeres, nucleoprotein complexes at the ends of chromosomes, with the DNA component being characterized by repeating G-rich sequences (TTAGGG in humans). Indeed, many G4s stemming from telomeric fragments were characterized by

X-ray and spectroscopic techniques and shown indeed to be arranged in a variety of architectures [44].

Many factors are likely to play a role in the folding–unfolding equilibria of telomeric DNA. First, the occurrence of a complex loop organization (D-loop–T-loop), which masks telomere ends from the repair machinery, second, the presence of several functional proteins bound to the telomere, which produce an additional level of complexity. In fact the proteins could act as chaperones to favor G4 or be recruited by a stable quadruplex structure. On the contrary, they could bind preferentially to an unfolded form of G4 hence shifting the equilibrium away from folding. In all cases, G4 propensity of the DNA sequence can modulate protein assembly and DNA processing in the telomere.

3.2. Shelterin and CST

In vertebrates, telomeric DNA is capped by a group of six proteins which assumes the name of shelterin. This comprises TRF1 and TRF2 which directly bind the duplex structure and POT1 which interacts with the single-stranded overhang tail. Additionally, Tin2, TPP1 and Rap1 do not bind directly DNA but are required to complete this nucleoprotein complex creating a functional network connecting all components [45].

An alternative capping pathway was first described in budding yeast: it involves a heterotrimer complex named CST (Cdc13–Stn1–Ten1). Its requirement for telomere stability in a wide range of eukaryotes even in the presence of the shelterin proteins is now confirmed [46].

These proteins are responsible for a fine tuning of the conformational equilibria the telomeric portion undergoes and are directly involved not only in telomere preservation but also in regulation of telomerase activity [47].

It has been known since over 15 years that telomerase is unable to elongate G4 folded substrates, a principle that justifies a chemotherapeutic approach based on G4 stabilizer. However, other DNA binding proteins can promote G4 at telomeric level. In *Oxytricha nova*, the ends of macronuclear chromosomes are bound by telomere end-binding protein (TEBP). This protein is formed by two subunits, α and β . TEBP α efficiently binds single stranded 3' telomeric repeat allocating it in a deep cleft. TEBP β heterodimerizes in the presence of DNA leading to an α : β :single-stranded(ss) DNA ternary complex, in which the nucleic acid is bound between the N-terminal domain of α and a globular region of β [48]. Although the β subunit does not bind the terminal telomeric repeat by itself, it greatly accelerates G4 formation with oligonucleotide substrates corresponding to the *Oxytricha* (T₄G₄T₄G₄) and *Tetrahymena* (T₂G₄T₂G₄) telomeric repeats at physiological conditions [49]. Interestingly, a higher resolution electron

density map of the above mentioned ternary complex evidenced additional binding by a G4 structure located apart from the ssDNA binding site (Fig. 2). In this complex TEBP is in contact with the folded DNA through interactions which are extensively of electrostatic type. In addition, the protein–nucleic acid contact surface is relatively small. From the structural data it was not clear if the observed binding was physiologically relevant as it could have been promoted by the crystallization packaging [50]. Indeed, it is the C-terminal basic domain of the β subunit (absent in the crystallized complex) that is actually considered as the G4 binding element [49]. Interestingly, the G4 stabilizing effect was shown to be modulated *in vivo* by cell cycle dependent protein phosphorylation. However, the preference for stabilization of antiparallel arrangements was demonstrated to be the result of the co-presence of the α and β subunit [51]. The β subunit was subsequently shown to have a partial sequence homology to TPP1 in shelterin whereas TEBP α corresponds to vertebrate POT1 [52].

Nevertheless, POT1 interacts with G4s but it unfolds them to allow telomerase action. The protein forms a 1:1 stoichiometric complex with the folded DNA and “melts” it to free its 3'-terminus, thereby allowing telomere elongation [53]. The same conclusions were drawn using fission yeast or mouse telomeric DNA in the presence of POT1 DNA binding domain. The type of structure recognized by the protein is possibly antiparallel [54–56].

In yeast, a G4 binding mode similar to POT1 was shared by Cdc13 which, among its multiple roles to regulate telomere maintenance, promotes partial G4s denaturation [57]. In this organism, its counterpart is represented by Est1p which, similarly to TEBP β , stabilizes and promotes G4s in an Mg^{2+} dependent manner. This mechanism is important for *in vivo* function as shown by mutants deficient in G4 stimulation which produced gradual

telomere shortening and cellular senescence [58]. The experimental findings indicate that Est1p is not necessary for catalytic activity, but important for telomerase regulation [59,60].

As far as the effects of telomeric protein TRF2 (and RAP 1), its role possibly rests in protecting G-strand telomere overhang by recognizing the single-stranded/double-stranded DNA interface. Indeed, TRF2 and POT1 assemble with difficulty in the presence of a G4-organized overhang. However, involvement of DNA repair proteins such as WRN and MRE11 is not inhibited showing that the combined protein machinery at the telomere is not substantially impaired by G4 [61].

In line with the prevalence of protein effects over intramolecular G4 folding, a circular dichroism and plasmid uptake investigation indicates that the Arg-rich N-terminal TRF₂ domain stimulates the formation of parallel intramolecular G4s (in the presence of potassium ions but not sodium ions), as it was the case for its yeast homologue RAP1 [62]. Addition of the TRFH oligomerization domain produced single stranded oligonucleotides and intermolecular G4s but not parallel-stranded structures. In addition, the presence of TRF₂ overcame G-quadruplex-dependent inhibition of annealing and stimulated telomeric oligonucleotide uptake. These results point to different G4 handling abilities of the two protein domains [63].

Taken together, the results so far discussed indicate an important conserved *in vivo* functional role played by the formation/disruption of G4s in telomerase regulation processes, where Cdc13, TEBP α , POT1 and TRF₂ exhibit G4 unfolding properties, counterbalanced by the G4 stimulating effects exhibited by Est1p, TEBP β and TPP1.

3.3. Other telomere-related proteins

Besides telomeric relevant proteic complexes forming shelterin and CST, other proteins can be recruited at the telomere during particular phases of the cell cycle. Again, the G4s content represents a relevant factor to control their recruitment, a factor that can be appropriately regulated by the protein binding event.

An important process which involves protein localization at telomeric level is DNA damage sensing and repair. It is carried out in humans by a heterotrimeric complex, MRN, which is homologous to the yeast MRX complex, formed by Mre11, Rad50, and Xrs2 in a 2:2:1 stoichiometry [64].

Interestingly, MRE11 possesses much higher affinity for G4 sequences (21–54 nt) resembling or not telomeric arrays than for single- or double-stranded DNA. Protein–G4 complexation stimulated endonuclease activity at sites flanking the binding site. This activity might be required to produce free ends serving as substrates for subsequent action by helicases and telomerase [65].

A remarkable preference for G4 over single-stranded DNA was conserved also by the MRX heterotrimer. In this complex Rad50 suppresses endonuclease but not exonuclease activity in *Saccharomyces cerevisiae*. As a consequence, cleavage at selected sites occurs (single-stranded G residues and double-stranded DNA TG steps) which comprise also the G4 arrangement. This suggests a role of Rad50 in telomere homeostasis [66].

In analogy to MRE11, also the replication protein A (RPA) is involved both in DNA repair and telomere maintenance [67]. Human RPA binds and unfolds G4 forming 1:1 and 2:1 complexes with a linear telomeric sequence. The first binding step is proposed to destabilize the G4 structure. This unfolded DNA bound to one hRPA molecule at the extremity facilitates binding of a second RPA molecule to produce a 2:1 complex. Interestingly, RPA is known to inhibit or stimulate telomerase activity depending upon protein concentration which might correlate with the formation of the above two types of complex [68]. Spectroscopic and electrophoretic

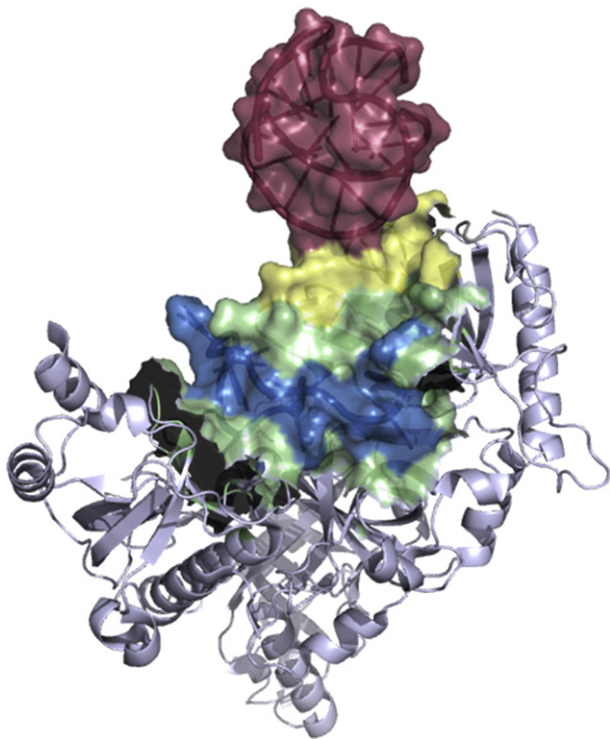


Fig. 2. Three-dimensional structure of the sequence GGGGTTTGGGG bound to *Oxytricha nova* TEBP heterodimer in a quaternary complex containing both single-stranded DNA and G4 folded DNA. PDB 1JB7. The G4 component is highlighted in red and the corresponding protein contacting surface in yellow, the single-stranded DNA fragment is in blue and the corresponding protein contacting surface in green. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

studies showed that RPA can recognize and unwind non-telomeric intramolecular G4 arrangements in a physiologically significant environment. However, the 1:1 complex can also bind the complementary pyrimidine rich strand, whereas it does not bind the annealed double-stranded structure [69].

Finally, ATRX is a widely expressed chromatin-associated protein, linked to syndromal mental retardation. The mechanism of action of this protein has been very recently examined and shown to be associated with telomeres, where it modulates histone replacement [70]. Interestingly, ATRX targets DNA regions having the potential to form G4 DNA and binds efficiently to sequences exhibiting a G4 (both parallel and antiparallel) structure, thereby suggesting a mechanism of selective protein nucleic acid recognition, possibly operating also in other non-telomeric regions [71].

3.4. Helicases

Helicases are DNA/RNA unwinding enzymes that use ATP as their energy source. Helicases generally work on duplex structures, but some of them are able to unwind a number of non canonical arrangements, including G4s where they are crucial to maintain the integrity of repetitive G-rich regions of the genome. Several helicase families including RecQ (hWRN, hBLM, Sgsl), FancJ, Pif1, Dog-1, RTEL and Dna2 are known to process telomeres, generally reducing their G4 levels [72]. Helicases directed toward the G-rich regions work both in telomerase-dependent cells as well as in those cells which use a non-conservative recombination-mediated telomere lengthening (ALT) pathway [73].

Deficiency of G4 active helicase causes severe impairment in DNA processing with aberrant genetic recombination and/or DNA replication [74]. As expected, inactivating mutations can produce dominant negative phenotypes where DNA replication, DNA damage response and protein trafficking are impaired [75]. Several critical pathological conditions are known to be associated with autosomal recessive mutations in genes coding for the G4-processing helicases [76–78].

Among these, Bloom syndrome is associated with enhanced rate of leukemias, solid tumors and sarcomas [79] whereas Werner syndrome is characterized by a premature occurrence of age-related symptoms including diabetes, arteriosclerosis, atherosclerosis and osteoporosis. In addition, this condition produces an unusually high incidence of cancer [80]. Finally, Fanconi anemia J produces bone marrow failure and a predisposition to cancer [81]. Due to the well established correlations between genetic defects and onset of disease, helicases have been intensively investigated and the full picture of their biological functions dissected to a remarkable extent.

For what concern WRN and BLM, they participate in DNA replication, recombination and repair by processing the nucleic acid with directionality 3'–5', while the other human helicases normally unwind G4 5'–3' [82,83]. In particular WRN efficiently unwinds bimolecular G4s obtained from the nucleotide repeat $d(\text{CGG})_n$ related to the fragile X syndrome (see below) in the presence of ATP and Mg^{2+} . Interestingly, unwinding by WRN of a replication-blocking G4 domain in a DNA template enables DNA polymerase δ to traverse the hindering G-rich DNA tract. The tetraplex structures must be embedded into linear filaments to be resolved, since blunt-end sequences are not processed, nor are tetramolecular G4s. Hence, the helicase exhibits well defined requisites for DNA recognition, depending upon molecularity and sequence characteristic of the tetraplex substrate [84,85].

Although WRN belongs to the RecQ helicase family, its requirements for DNA substrate are not conserved for all members of this family. As an example, RECQ1 helicase cannot unwind G4 even in the presence of the protein helicase cofactor RPA, but it is effective in

resolving Holliday junctions thus highlighting substantially different complementary functions covered by different members of the RecQ family [86].

Subsequent studies on binding selectivity showed that a mandatory prerequisite for WRN to initiate unwinding is the presence of a single-stranded sequence at the 3' end. Similar requirements hold for BLM [76], which binds to G4s within G-loops. The principal site for binding forked DNA as well as G4 structures is located at the conserved RQC domain, consisting of a Zn^{2+} -binding region and a winged helix subdomain [87]. Unexpectedly, recent real-time fluorescence studies showed that BLM was poorly effective in unwinding intramolecular G4s adopted by different natural sequences, while it unwound double-stranded DNA far more efficiently. This challenges the information thus far available, although it was not obtained with intramolecularly folded G4s. In this case quadruplex formation is interpreted as an obstacle, hence a modulating element, for enzyme translocation along DNA [88].

Since oxidized dG represents one of the principal products of oxidative DNA damage, the molecular mechanism of G4 recognition was further examined by substituting dG with 8-oxo-dG in BLM quadruplex substrates [89]. Interestingly, the 8-oxo modified telomeric D-loops were preferentially unwound by WRN and BLM and were selectively recognized by POT1. These observations suggest involvement of these helicases in telomeric DNA repair processes with POT1 providing secondary structure disruption [90].

Studies on FANCI and its *C. elegans* ortholog DOG-1 suggest that also this enzyme is involved in the maintenance of unstable genomic G/C rich regions and protects them from being deleted during DNA processing. This is not limited to interstrand cross-link repair as formerly believed, but includes other pathways of DNA replication. In particular replication-blocking G4s, shown to be intrinsically mutagenic *in vivo*, are removed from genomes lacking DOG-1 [91–93].

Yeast Pif1 helicase effectively counteracts genetic instability of the G-rich human minisatellite CEB1 inserted in a yeast genome, while other helicases had no effect on the same construct. Studies on mutated CEB1 exhibiting different G-forming potential confirmed a relationship between enzyme unwinding activity and efficient G4 folding. In the human enzyme, the protein domain involved in preferential G4 resolvase action spans amino acids 206–620. The helicase substrate, besides the G-rich region, required a 5'-single stranded DNA flanking tail, which is not essential for enzyme binding to target [94,95].

The lagging strand DNA replication protein Dna2 is another helicase exhibiting nuclease activity. Both the yeast and the human Dna2 enzyme confirmed preferential G4 structure specific binding and efficient unwinding of this DNA structure. However, the 5'–3' nuclease activity is reduced by the folded DNA arrangement, but is restored in the presence of RPA [96].

Another non RecQ helicase isolated from human cells, DHX36, a G4 resolvase known also as RHAU belongs to the Asp-Glu-Ala-His box family. Its G4s selectivity was identified using affinity beads in the presence of ATP and Mg^{2+} [97]. The physiological relevance of this process is witnessed by the significant (over 50%) drop in helicase activity towards tetramolecular G4 following immunodepletion using a monoclonal antibody raised against DHX36 [98]. Enzyme down regulation decreased this biological process by almost one order of magnitude [99].

Interestingly, DHX36 resolvase activity is extended also to RNA substrates capable of folding into tetraplex. Binding affinity and selectivity for G4 was found to be located in the N-terminal (RHAU specific motif) conserved region. This was evidenced using the equivalent *Drosophila* protein, sharing with DHX the amino-terminal domain only. The high binding affinities exhibited towards G4 RNA (K_d in the pM range) suggests that this structure is

preferentially recognized by DHX and may represent a suitable *in vivo* target, which is unique among RNA helicases [100]. Very recently, the binding of DHX36 to a G-rich sequence of telomerase RNA (hTR), located in the 5'-terminal region, has been examined. Accordingly, recognition did not involve the helicase domain, but the N-terminal accessory domain of the enzyme, which binds specifically to parallel G4s resolved by the helicase domain. This process affects mature hTR formation and, as a result, the levels of telomere elongation by telomerase. The idea follows that G4 folding may protect nascent hTR from degradation, the subsequent resolvase action favoring RNA elongation to reach the mature form [101].

G4 motif modulation may occur also in viral systems. An example is represented by the SV40 large T-antigen helicase, a multifunctional protein required for viral replication and transformation. Besides double-stranded substrates, it unwinds also intramolecular G4s obtained from sequences occurring in the viral genome. This enzyme apparently binds different DNA conformations with similar efficiency, perhaps using different protein domains. Again a single stranded region 3' to the G4 is required for effective unfolding of the compact structure [102].

3.5. Heterogeneous nuclear ribonucleoproteins

The heterogeneous nuclear ribonucleoproteins (hnRNPs) are RNA binding proteins participating in multiple functions of nucleic acid metabolism (packaging of nascent transcripts, alternative splicing and translational regulation). They are characterized by distinctive features such as an RNA binding domain containing an Arg-Gly-Gly (RGG) box and an auxiliary domain rich in specific amino acids such as Gly, Asp/Glu or Pro. They are also able to cross the nuclear membrane [103].

Distinct hnRNP motifs involved in G4 destabilization of the fragile X repeat sequence were identified in the rat liver telomeric DNA binding protein 42 and in the mouse homologue CBF-A as well as in other members of the family like hnRPA1 and hnRNPA2. Incidentally, different protein domains are able to stabilize (rather than destabilize) the G4 structure formed by the telomeric repeat $d(\text{TTAGGG})_n$ [104,105].

hnRNP D consists of two nucleic acid binding domains able to recognize the human telomeric repeat $(\text{TTAGGG})_n$. NMR studies on the C-terminal binding domain in complex with the above repeat demonstrated that upon binding the protein unfolds a preformed quadruplex of telomeric DNA by binding to the linear oligonucleotide. This suggests a participation of hnRNP D in protecting a single-stranded telomerase template or eliminating the interference with the elongation of telomere DNA when folded into G4 at each translocation event [106].

Similar findings apply to the A/B hnRNPs, depletion of which substantially impaired telomerase activity. This was restored upon addition of hnRNP A1, a 320 amino acids protein with nucleic acid annealing activity [107]. Gel retardation, FRET spectroscopy and single molecule FRET microscopy confirmed that hnRNP A1 binds efficiently to single-stranded G-quadruplex telomeric sequences. In particular, it was shown that the telomere fragment binds to one or several hnRNP A1 molecules generating a rather tight complex structure. This fact suggests that binding of the ribonucleoprotein might reduce telomere DNA accessibility, actually, producing a telomere protecting effect [108].

The RNA recognition motifs of hnRNP A1, BD1 and BD2, are contained in its N-terminal 196 amino acids domain called unwinding protein 1 (UP1). In this protein, two highly conserved Arg-Arg-Met RNA binding structural motifs form a rigid combined entity fit for recognition of the nucleic acid [109]. Identification of UP1 as a G4 binder originates from studies aimed at selecting

proteins able to bind the mouse hypervariable minisatellite Pc-1, a sequence known to be G4 folded (likely as an intramolecular antiparallel quadruplex) at physiological conditions. Six Pc-1 binding proteins were isolated from NIH3T3 cellular extracts, and UP1 was identified among them. Subsequent electrophoretic mobility shift assays revealed tight binding (nanomolar K_d) of the Pc-1 oligonucleotides to hnRNP A1 and A3. Addition of nucleoproteins abolished DNA synthesis arrest, confirming their ability to unfold G4 architectures and contribute to preserving genome stability [110,111].

The role of UP1, at least in telomerase regulation, appears to be quite complex. Indeed, UP1 is able to enhance telomerase activity in a bell shaped fashion through unfolding of the G4 telomeric DNA and also through recruitment of telomerase to telomeric DNA. In fact, UP1 forms a ternary complex with telomerase and telomeric RNA. NMR chemical shift perturbation analysis suggests that telomerase RNA engages both BD1 and BD2 in the binary complex with UP1. Addition of telomeric DNA produces a ternary complex by displacing RNA from the BD1 site. Hence, the ribonucleoprotein can recruit (and unfold) telomeric DNA to telomerase by effectively bridging them. Incidentally, formation of a ternary complex may explain the bell shaped dependence of telomerase activity upon UP1 concentration. In fact, optimal activity depends on the relative ratios of ternary complex components and is not simply a monotonic function of the components' concentration [112].

3.6. G4-related control of transcription: proteins binding at gene promoters

Besides telomeres, other genomic locations exhibit high G4 probability as shown by computational analysis [113]. G-rich regions were found to be mainly located upstream of gene transcription start sites, hence in a context where they can be used to regulate gene expression. In particular almost 50% of all genes contain putative G4 sites, most abundantly at oncogene promoters but not at oncosuppressor genes. Hence, particular attention has been devoted to investigate the effects of possible G4 structures as modulators of the expression of several oncogenes, including c-myc, c-Kit, KRAS; pRb insensitivity; Bcl-2, VEGF and PDGF, representing a variety of differentiated conformational features. Considering that the double helical structure is under stress during transcription, it can easily compensate negative superhelicity with local G4 folding. Remember that the pyrimidine (C) rich strand can itself fold into a stable i-structure, further stabilizing an open DNA form [114,115].

The as yet more deeply investigated oncogene promoters are c-myc and KRAS. For thorough reviews on the state of the art concerning c-myc see [116,117]. Among the proteins shown to play an important role in G4 binding/processing on c-myc are the non metastatic protein, NM23-H2, nucleolin and the cellular nucleic acid binding protein, CNAB.

NM23-H2 is a member of a highly conserved family of protein kinases and is critically involved in cancer and development [118]. It binds c-myc at a nuclease responsive element (NHE III). Since it recognizes single stranded (purine- or pyrimidine-rich) nucleic acid sequences, it promotes unfolding of G4 (and C-rich structures) thus allowing transcriptional activation of the oncogene. Protein-mediated G4 unfolding was further confirmed by FRET experiments [119].

The same sequence but in its G4 form is recognized by nucleolin, one of the most abundant non ribosomal proteins of Eukaryotes that participates in ribosome biogenesis and cell proliferation. Its multifunctional role as well as its localization is dependent upon cellular stress conditions [120]. An investigation on deletion mutants showed that the C-terminal region of the protein is primarily

responsible for G4 binding, the minimal G4 recognition structure consisting of the two RNA binding domains and the RGG domain [121]. In particular, nucleolin binds and stabilizes NHE III of c-myc promoter thus it might act as an oncosuppressor by decreasing promoter activity in a G4 dependent way [122]. Remarkably, nucleolin binds NHE III in HeLa cells, which points to the *in vivo* significance of the observed interaction.

Finally, the cellular nucleic acid binding protein, CNAB, is a small highly conserved single stranded nucleic acid (DNA and RNA) binding protein involved in the control of cell proliferation and death. In its dimeric form it binds RNA as well as single-stranded DNA G-rich sequences. Its cellular targets include regulatory elements in gene promoters and in mRNA untranslated regions [123,124]. Upon binding to NHE III, CNAB enhances c-myc transcription. According to the data thus far discussed and the results obtained with nucleolin, it should be expected to melt the ordered G4. Instead, it appears to discriminate between parallel and antiparallel G4 folding, favouring the former. Hence, this represents an yet undisclosed gene regulation mechanism, whereby activation/repression is brought about by switching an ordered G4 topology into another [125].

Also, the nuclease hypersensitive element (NHE) of hKRAS was shown to adopt two distinct G4 conformations identified as a parallel and a mixed parallel/antiparallel form. NHE bound to paramagnetic beads to isolate G4 binding proteins from nuclear extracts allowed identification of proteins bound at this promoter like poly(ADP-ribose) polymerase-1 (PARP-1), Ku70 and hnRNP A1 [126]. Hence, hnRNPs not only contribute to essential interactions at the telomere level (see above), but they can also target other G-rich (DNA) sequences at oncogene promoters.

As it might be expected, KRAS G4s at the promoter are effectively unwound by hnRNP A1 and UP1, facilitating a quadruplex to duplex transition. This prompted the synthesis of decoy oligonucleotides directed to interfere with hKRAS promoter regulation by binding to promoter recognizing proteins, hence representing possible drugs directed towards KRAS-related disorders. Since these oligonucleotides should maintain a stable conformation in solution to efficiently recognize and sequester protein targets, they were modified with (*R*)-1-*O*-(4-(1-pyrenylethynyl) phenylmethyl) glycerol units (TINA) able to stack onto tetrads thus stabilizing G4. The modified decoys were tested against pancreatic adenocarcinoma cell lines, known to harbor mutations in the RAS gene. Interestingly, they showed remarkable antiproliferative activity with sub-micromolar IC₅₀ [127–129].

A GA rich element crucial for transcription was found also in the murine KRAS promoter. Pulldown and chromatin precipitation studies demonstrate that it recognizes the Myc-associated zinc finger protein (MAZ) and the polymerase PARP1. In particular, MAZ binds at the GA elements to both double-stranded and G4 structures whereas PARP-1 is specific for the latter only, binding with a protein:G4 stoichiometry 2:1 *in vitro*. These proteins were confirmed to be important for transcription since, when they are blocked by siRNA sequences, transcription is effectively down-regulated [130,131].

The c-myc oncogene represents a peculiar example. Indeed, spectroscopic and mutagenesis studies are consistent with an unusual G4 structure assumed by the GGA repeat. It derives from stacking of two tetrad/heptad sequences formed by two of the three (GGA)₄ repeats. Deletion of one of these repeating tetrads increased promoter activity, however, complete deletion of the GGA motifs impaired gene transcription; this clearly shows that the G-rich promoter sequence plays a role both as an activator and as a repressor. In addition, the GGA region is recognized by MAZ, which binds all GGA repeats and represses promoter activity. As with the murine KRAS MAZ is able to recognize both double-stranded DNA

and the peculiar G4s described above, without unfolding them [132].

Another biochemically and pharmacologically significant protein is the vascular endothelial growth factor VEGF. Compelling biochemical and biophysical evidence shows that G4 structures can form both *in vitro* and *in vivo* in a sequence located in the growth factor promoter region comprising four runs of three or more contiguous guanines separated by one or more bases. The secondary structure formed on the complementary C-rich strand is characterized by an intramolecular i-motif that involves six C–C(+) pairs [133–136]. Interestingly, the G-rich VEGF promoter region is recognized by the G4 binding protein nucleolin as it was the case for the c-myc promoter [121].

A G4 conformational shift to control gene expression is a mechanism not constrained to oncogenes but, as now extensively confirmed, it is common among other physiological pathways.

The ILPR (insulin linked polymorphic region) is a non-coding minisatellite upstream of the human insulin promoter associated with genetic susceptibility to insulin-dependent diabetes mellitus. It contains a G-rich two-repeat sequence, able to adopt an intramolecular G4 conformation. Interestingly, insulin tightly binds the ILPR sequences, the quadruplex parallel loops likely recognizing insulin's β chains. This raises the possibility of a direct interaction of a promoter with the correlated gene product [32].

Interestingly, besides insulin expression, polymorphism of ILPR can also affect the adjacent gene coding for insulin-like growth factor 2 (IGF-2). In analogy to insulin, IGF-2 binds efficiently to ILPR-related G4s confirming the possibility of an interplay between gene expression and promoter regulation [137].

Subsequent studies were carried out analyzing the interaction of insulin and IGF-2 with 3 selected ILPR variants by affinity MALDI and surface plasmon resonance techniques. The observed *K_d* values ranged between 10⁻⁸ and 10⁻¹³, showing a remarkable modulation by the nature of a G4-ordered structure and the possibility to powerfully detect efficient binding [138]. Protein digestion followed by MALDI detection of the G4s-IGF-2 complexes, showed a general Val-Cys-Gly-Asn-Arg-Gly-Phe consensus. It is not clear whether the results are in keeping with insulin β chains representing the major G4 binding site [139]. Clearly, another requirement for ILPR to display high-affinity binding for insulin is possibly related to hydrogen bonding (Watson–Crick) interactions between the loop components ACA and TGT. Furthermore, recognition occurs most effectively with an antiparallel G4 [140].

An interesting mechanism of gene regulation is proposed for myogenic factors (MRFs) which initiate and direct myogenesis. MRFs heterodimerize with E-proteins, producing an E-box motif which recognizes d(CANNTG) sequences in muscle tissue differentiating genes. However, homodimerization of MRF can switch binding preferences to G4 motifs. This raises the possibility that transcriptional activity might be modulated by the relative homo/heterodimer distribution. The presence of G4 would favour homodimers, hence acting as transcription silencing factors [141]. However, this is not the case, since G4s substantially enhance MRF(MyoD)/E-box driven expression of a reporter gene (luciferase). A possible mechanism compatible with all these findings envisages a G4 structure in the E-box promoter recruiting (transcriptionally ineffective) MyoD homodimers. In the presence of the E-protein partner, MyoD forms an heterodimer which releases the myogenic factor from the G4 and binds the neighboring E-box motif to activate transcription of the genes coding for muscle proteins [142].

Nucleophosmin (NMP1) is a phosphoprotein able to bind RNA and DNA. It has an RNase activity and preferentially cleaves pre-rRNA. In addition, it is involved in regulation of tumor suppressor genes p53 and p14arf, and mutations in the C-terminal domain are

observed in acute myeloid leukaemia [143]. NMP1 C-terminal domain binds DNA in a structure-dependent fashion, as it shows clear preference for G4 arrangements. Interestingly, a specific sequence found at the SO2 gene promoter, shown to be a selective target for the protein *in vivo*, folds as a G4 at physiological conditions. Interestingly, most of the mutations occurring in the leukemic phenotype map at the G4 binding domain and produce structural destabilization of the protein causing aberrant behavior [144].

The genome-wide search for G4 usually considers sequences potentially forming a folded structure with at least three stacking tetrads to grant adequate stability. However an investigation on the promoter of the ubiquitous enzyme human thymidine kinase showed a possibility for a G4 consisting of two tetrads only, similarly to the thrombin aptamer [16]. Indeed, biophysical and biochemical studies confirmed a stable quadruplex folding of this G-rich promoter fragment. In addition reporter gene experiments in cell systems supported a relationship between the folding state of the test promoter sequence and the level of gene transcription [145]. Thus, it is now evident that a more comprehensive search taking into account more flexible primary DNA sequence requirements is demanded in the near future.

3.7. G4-related control of translation: proteins binding at untranslated RNA sequences

RNA sequences are recognized depending on their primary or secondary structure and, often, the information for posttranscriptional regulation is localized in untranslated regions (UTR) of mRNAs. In particular, RNA regulatory sequences located in the 5'-UTR are likely to play a role in translational regulation, while RNA transport seems mostly regulated by sequences in the 3'-UTR of mRNAs, even if several exceptions are known [146,147]. Computational approaches of the transcripts of known protein coding genes indicate high incidence of potential G4 structures both at 5'- and 3'-UTR locations. Since G-rich RNA sequences are able to fold into the G4 structure similarly to DNA sequences [148], G4 occurrence at 5'-UTR can be envisaged in terms of translational repression [149,150].

From the above data G4 UTR binding proteins could modulate the translation process as others do during transcription and hence work as activators or repressors of RNA maturation.

A few examples are available in the present literature, but many more are expected to appear in the near future.

The first example of a G4 RNA able to suppress a translation process was originally reported for NRAS [151]. The 5'-UTR of this gene was found to be bound by Pat1 proteins. These are conserved across eukaryotes where they are involved in posttranslational mechanisms, translation repression, decapping activation and ribonucleoprotein degradation. For these proteins RNA binding has been recently demonstrated *in vitro*, with preference for G4 motifs, including the 5'-UTR of NRAS [152]. More interestingly, when tethered to reporter mRNA they were shown to repress translation, as it might be expected by a G4 stabilizing protein [153].

A second example deals with the SARS coronavirus unique domain SUD, which occurs in the highly pathogenic species only. A SUD fragment (SUD core), containing 264 of the 338 residues of the full-length domain, interacts efficiently with DNA and RNA sequences folded as G4s. Since it is not likely for the viral protein to enter the nucleus of the host cell, the most accessible targets are G4-adopting mRNA sequences coding for host proteins. This would allow control of the host cell response during infection [154].

Another case concerns the Epstein Barr Virus Nuclear Antigen 1 (EBNA-1), a virus-encoded protein essential for genome replication and maintenance during latency. EBNA-1 linking regions 1 and 2 are known to be involved in the cellular origin recognition complex.

Interestingly, in particular linking region 2 is arginine- and glycine-rich, hence resembling the RGG motifs that are known to operate in a number of RNA binding proteins. In fact linking region 2 (and 1) is found to bind G-rich RNA sequences predicted to adopt a G4 fold, suggesting that the proliferation and maintenance of the viral genome is mediated by G4 RNA recognition [155].

An additional well established example refers to a neurological pathology called Fragile X syndrome (FXS) recognized as the most common inherited form of mental retardation [156].

Genetically it is characterized by expansion of the number of a CGG repeat in the 5'-UTR of the FMR1 gene from below 50 in healthy individuals to over 200 in sick people. This abnormal expansion produces hypermethylation and silencing of the FMR1 gene which leads to loss of the fragile X mental retardation protein (FMRP). In the pathological conditions, the G-rich expanded repeat is occurring both at the genomic DNA (non-coding strand) and at an RNA level, both of which could be exploited to modulate FMR1 expression and translation. To assess the functional consequences of this hypothesis, a long CGG sequence containing 99 repeats was positioned upstream to a reporter gene: it efficiently reduced mRNA translation and depressed mRNA utilization in human cells transfected with plasmids bearing FMR1 promoter-dependent reporter gene. Parallel G-quadruplex disrupting proteins hnRNP A and CBF-A effectively increased reporter mRNA translation, whereas protein mutants unable to unwind G4 were inactive. The emerging mechanism foresees a translational block produced by the folded G4 structures of the long CGG repeat sequence, reversed by G4 unwinding proteins [157].

Protein FMRP itself (but not its autosomal paralogs FXR1P and FXR2P) binds efficiently to G4 RNA through its RGG box, without requiring additional binding motifs such as a stem or a stem/G4 junction as shown by several spectroscopic techniques [158]. This binding is still sustained by Arg residues of the RGG box, in which four Arg residues at positions 533, 538, 543 and 545 can be methylated post-translationally. Interestingly, the number of Arg required for RNA binding depends upon the nature of target RNA. For example only the first two residues are required for poly-ribosome association and G4 recognition, but all for mRNA association. Thus a role of methylation in regulating the bound RNA can be envisaged [159].

Several other proteins that bind to the G4 d(CGG)_n repeat were further identified. UP1 unfolds the trinucleotide repeat and reverses DNA synthesis arrest as it does on other G4 sequences [160]. Distinct is the behaviour of the telomeric DNA binding protein 42, rat homologue of the mouse CBF-A, previously identified as a single-stranded and quadruplex telomere DNA binding protein. In fact, while it stabilizes the human telomeric repeat, it destabilizes d(CGG) G4 repeats. Hence, the DNA sequences bind at different protein domains [161,162].

FRAXE is another form of mental retardation due to impairment of the FMR2 gene. The functions of the gene product are not yet fully understood, but the protein FMR2P specifically and effectively binds G4 RNA. In addition, the enhancer action of a G4 located in the mRNA of an alternatively spliced exon is reduced by FMR2P, suggesting that the protein is involved in the control of G4-mediated alternative splicing [163]. A thorough discussion on the role of G4 in RNA metabolism using FMRP and FMR2P as paradigmatic examples highlights the principal aspects of post-transcriptional gene expression, FMRP participating in translational regulation and FMR2P in alternative splicing processes [164].

Finally, it is worth recalling how it is not always possible to predict the binding preference by monitoring the presence of conserved protein structural domains. As an example, in the preceding reports, RGG box appeared to represent a general selective G4 binding motif. This is however not a rule since the RGG of the

Herpes Simplex Virus-1 ICP27 protein binds flexible non G4 RNA structures [165,166].

3.8. Consequences of G4-protein recognition in drug design and development

A large body of information is available on G4-interacting small molecules, both in structural, functional and pharmacological studies [10,167,168].

The basic chemical requirements for effective G4 recognition by small molecules have been dissected to a large extent and modeling/structural studies give useful hints on specific ligand–nucleic acid interactions [169].

G4 binders have clearly a large potential as therapeutic agents [170]. First of all telomere-directed compounds behave as anti-proliferative agents interfering with both telomerase-active organisms and systems exhibiting alternative mechanisms of telomere maintenance, exploiting inhibition of telomerase function in the first case and telomere dysfunction in the second.

As a result, some of the known ligands are being considered for cancer treatment [171]. Interestingly, also “generic” G-rich oligonucleotides behave as antiproliferative agents which, among possible mechanisms, might imply competition effects for recognition of and interference with G4-mediated biological processes. Given the ubiquitous, although not random, occurrence of G4 motifs in the genome and in RNA transcripts, a G4 binding agent is expected to cause a number of side effects by impairment of several physiological processes not related to disease. Hence, unwanted toxicities are to be expected. In addition, G4 folding shows a remarkable polymorphism in DNA, which on one hand might improve differential targeting among topological forms, but on the other might render the nucleic acid target ill-defined, hence elusive [172].

The relatively poor specificity of a ligand directed to G4 might be turned into a strong one by considering protein–G4 complexes as novel drug targets, similarly to well-known protein–nucleic acid complexes exploited as drug targets, such as the DNA–topoisomerase cleavage complexes.

The selectivity could be generated by drug features recognizing the G4-protein interface, in principle affecting the stability of the binary complex through stabilizing or destabilizing effects. We could envisage ternary complex formation, drug competition with the G4 binding site and/or allosteric effects to strengthen or impair quadruplex-related biological response. This includes also interference with mechanisms of protein homo- and hetero- dimer formation which are known to participate in several critical regulatory processes. Conversely destabilization of an unwanted G4 fold may be carried out by devising a ligand tethered to stabilize a single stranded binding protein–unfolded nucleic acid complex.

While structural information on the latter type of complex is readily available in the literature [173], very little three-dimensional information is available on proteins bound to G4 in a complex [16,50]. Therefore it is at present not easy to pinpoint precise structural features involved in G4 binding specificity. From the nucleic acid side the quadruplex scaffold can be considered as a central core, like a flower's corolla, with the petals being represented by the loops (each characterized by the direction, number and nature of connecting bases), able to mediate selective recognition of a target protein in a way reminiscent of antigen binding to antibody. In this connection, many G4-based aptamers have been identified and are being investigated as drugs and as diagnostic agents.

As far as newly emerged targets, the G-rich complementary strand in double-stranded chromatin is a C-rich sequence, which is known to fold into an i-motif, a four-stranded structure consisting of parallel-stranded duplexes zipped together in an antiparallel

orientation by intercalated, hemiprotonated cytosine⁺–cytosine base pairs. It is found in gene promoters such as c-myc, bcl-2 and ILPR [174–176]. The i-motif, besides representing per se a potential drug target, might be also recognized by nucleic acid binding proteins, such as NM23-H2, again a G4-protein complex representing a unique feature for highly specific ternary interactions [117].

In addition, it has been recently shown that eukaryotic telomeres are transcribed into telomeric repeat containing RNA (TERRA). Telomeric tract-derived 5'-UUAGGG-3' repeats span about 200 bases and are able to fold into full RNA G4 structures and RNA-DNA inter-molecular hybrids [177].

TERRA appears to participate in telomeric DNA replication, heterochromatin formation and telomerase regulation. A number of TERRA-binding proteins including TRF1, TRF2 and hTERT have been identified, which suggests the possible existence of G4 TERRA recognizing proteins and hence the possible selective modulation of the TERRA functions by disrupting or stabilizing G4-protein complexes [178,179].

In medical applications binary protein-G4 complexes might be specifically targeted by drug molecules combining efficacy and selectivity to treat a series of severe diseases, starting with cancer (telomere and oncogene promoter targets), mental conditions such as fragile X, diabetes and illnesses associated with autosomal recessive mutations in genes coding for G4-processing helicases.

Some known G4 binders have already been shown to interfere with particular protein-G4 systems. Many were essentially employed as biochemical tools, but could be possibly turned into drugs as they behave as effective modulators of biological responses. On turn, specific G4 topologies could be stabilized in solution by chemical modifications of naturally occurring oligonucleotides, to act as decoys for pharmacological applications [127].

In conclusion, a medicinal chemist will be amazed by the variety of opportunities for drug discovery offered by an exciting, fast growing field of research such as the world of G4.

Acknowledgments

This work was supported by AIRC (Associazione Italiana per la Ricerca sul Cancro) grant #5286. The authors acknowledge Dott. Marson G. for graphical assistance.

References

- [1] A. Majumdar, D.J. Patel, Identifying hydrogen bond alignments in multi-stranded DNA architectures by NMR, *Acc. Chem. Res.* 35 (2002) 1–11.
- [2] S. Neidle, S. Balasubramanian, *Quadruplex Nucleic Acids*. RSC, Cambridge, 2006.
- [3] M. Wieland, J.S. Hartig, RNA quadruplex-based modulation of gene expression, *Chem. Biol.* 14 (2007) 757–763.
- [4] M.A. Mullen, K.J. Olson, P. Dallaire, F. Major, S.M. Assmann, P.C. Bevilacqua, RNA G-Quadruplexes in the model plant species *Arabidopsis thaliana*: prevalence and possible functional roles, *Nucleic Acids Res.* 38 (2010) 8149–8163.
- [5] H. Arthanari, P.H. Bolton, Functional and dysfunctional roles of quadruplex DNA in cells, *Chem. Biol.* 8 (2001) 221–230.
- [6] L. Oganessian, T.M. Bryan, Physiological relevance of telomeric G-quadruplex formation: a potential drug target, *Bioessays* 29 (2007) 155–165.
- [7] J.E. Johnson, J.S. Smith, M.L. Kozak, F.B. Johnson, In vivo veritas: using yeast to probe the biological functions of G-quadruplexes, *Biochimie* 90 (2008) 1250–1263.
- [8] J.S. Smith, Q. Chen, L.A. Yatsunyk, J.M. Nicoludis, M.S. Garcia, R. Kranaster, S. Balasubramanian, D. Monchaud, M.P. Teulade-Fichou, L. Abramowitz, D.C. Schultz, F.B. Johnson, Rudimentary G-quadruplex-based telomere capping in *Saccharomyces cerevisiae*, *Nat. Struct. Mol. Biol.* 18 (2011) 478–485.
- [9] T.M. Bryan, P. Baumann, G-quadruplexes: from guanine gels to chemotherapeutics, *Methods Mol. Biol.* 608 (2010) 1–16.
- [10] S. Neidle, Human telomeric G-quadruplex: the current status of telomeric G-quadruplexes as therapeutic targets in human cancer, *FEBS J.* 277 (2010) 1118–1125.

- [11] S. Balasubramanian, L.H. Hurley, S. Neidle, Targeting G-quadruplexes in gene promoters: a novel anticancer strategy? *Nat. Rev. Drug Discov.* 10 (2011) 261–275.
- [12] T.M. Bryan, M.B. Jarstfer, Interrogation of G-quadruplex–protein interactions, *Methods* 43 (2007) 332–339.
- [13] H.J. Lipps, D. Rhodes, G-quadruplex structures: in vivo evidence and function, *Trends Cell Biol.* 19 (2009) 414–422.
- [14] M. Fry, Tetraplex DNA and its interacting proteins, *Front. Biosci.* 12 (2007) 4336–4351.
- [15] B. Gatto, M. Palumbo, C. Sissi, Nucleic acid aptamers based on the G-quadruplex structure: therapeutic and diagnostic potential, *Curr. Med. Chem.* 16 (2009) 1248–1265.
- [16] K. Padmanabhan, K.P. Padmanabhan, J.D. Ferrara, J.E. Sadler, A. Tulinsky, The structure of alpha-thrombin inhibited by a 15-mer single-stranded DNA aptamer, *J. Biol. Chem.* 268 (1993) 17651–17654.
- [17] A. Pasternak, F.J. Hernandez, L.M. Rasmussen, B. Vester, J. Wengel, Improved thrombin binding aptamer by incorporation of a single unlocked nucleic acid monomer, *Nucleic Acids Res.* 39 (2011) 1155–1164.
- [18] Y. Nonaka, K. Sode, K. Ikebukuro, Screening and improvement of an anti-VEGF DNA aptamer, *Molecules* 15 (2010) 215–225.
- [19] D.H. Bunka, O. Platonova, P.G. Stockley, Development of aptamer therapeutics, *Curr. Opin. Pharmacol.* 10 (2010) 557–562.
- [20] P.J. Bates, D.A. Laber, D.M. Miller, S.D. Thomas, J.O. Trent, Discovery and development of the G-rich oligonucleotide AS1411 as a novel treatment for cancer, *Exp. Mol. Pathol.* 86 (2009) 151–164.
- [21] E.W. Choi, L.V. Nayak, P.J. Bates, Cancer-selective antiproliferative activity is a general property of some G-rich oligodeoxynucleotides, *Nucleic Acids Res.* 38 (2010) 1623–1635.
- [22] G. Oliviero, J. Amato, N. Borbone, S. D'Errico, A. Galeone, L. Mayol, S. Haider, O. Olubiyi, B. Hoorelbeke, J. Balzarini, G. Piccialli, Tetra-end-linked oligonucleotides forming DNA G-quadruplexes: a new class of aptamers showing anti-HIV activity, *Chem. Commun. (Camb.)* 46 (2010) 8971–8973.
- [23] D. Margulies, A.D. Hamilton, Digital analysis of protein properties by an ensemble of DNA quadruplexes, *J. Am. Chem. Soc.* 131 (2009) 9142–9143.
- [24] D. Margulies, A.D. Hamilton, Protein recognition by an ensemble of fluorescent DNA G-quadruplexes, *Angew. Chem. Int. Ed. Engl.* 48 (2009) 1771–1774.
- [25] J.R. Wyatt, T.A. Vickers, J.L. Roberson, R.W. Buckheit Jr., T. Klimkait, E. DeBaets, P.W. Davis, B. Rayner, J.L. Imbach, D.J. Ecker, Combinatorially selected guanosine-quartet structure is a potent inhibitor of human immunodeficiency virus envelope-mediated cell fusion, *Proc. Natl. Acad. Sci. U.S.A.* 91 (1994) 1356–1360.
- [26] S. Weiss, D. Proske, M. Neumann, M.H. Groschup, H.A. Kretzschmar, M. Famulok, E.L. Winnacker, RNA aptamers specifically interact with the prion protein PrP, *J. Virol.* 71 (1997) 8790–8797.
- [27] A. Matsugami, T. Mashima, F. Nishikawa, K. Murakami, S. Nishikawa, K. Noda, T. Yokoyama, M. Katahira, Structural analysis of r(GGA)₄ found in RNA aptamer for bovine prion protein, *Nucleic Acids Symp. Ser. (Oxf.)* (2008) 179–180.
- [28] F. Pileur, M.L. Andreola, E. Dausse, J. Michel, S. Moreau, H. Yamada, S.A. Gaidamakov, R.J. Crouch, J.J. Toulme, C. Cazenave, Selective inhibitory DNA aptamers of the human RNase H1, *Nucleic Acids Res.* 31 (2003) 5776–5788.
- [29] N. Jing, Y. Li, X. Xu, W. Sha, P. Li, L. Feng, D.J. Tweardy, Targeting Stat3 with G-quartet oligodeoxynucleotides in human cancer cells, *DNA Cell Biol.* 22 (2003) 685–696.
- [30] T. Mori, A. Oguro, T. Ohtsu, Y. Nakamura, RNA aptamers selected against the receptor activator of NF-kappaB acquire general affinity to proteins of the tumor necrosis factor receptor family, *Nucleic Acids Res.* 32 (2004) 6120–6128.
- [31] A.T. Phan, V. Kuryavyi, J.B. Ma, A. Faure, M.L. Andreola, D.J. Patel, An interlocked dimeric parallel-stranded DNA quadruplex: a potent inhibitor of HIV-1 integrase, *Proc. Natl. Acad. Sci. U.S.A.* 102 (2005) 634–639.
- [32] A.C. Connor, K.A. Frederick, E.J. Morgan, L.B. McGown, Insulin capture by an insulin-linked polymorphic region G-quadruplex DNA oligonucleotide, *J. Am. Chem. Soc.* 128 (2006) 4986–4991.
- [33] D. Michalowski, R. Chitima-Matsiga, D.M. Held, D.H. Burke, Novel bimodular DNA aptamers with guanosine quadruplexes inhibit phylogenetically diverse HIV-1 reverse transcriptases, *Nucleic Acids Res.* 36 (2008) 7124–7135.
- [34] A.C. Girvan, Y. Teng, L.K. Casson, S.D. Thomas, S. Juliger, M.W. Ball, J.B. Klein, W.M. Pierce Jr., S.S. Barve, P.J. Bates, AGRO100 inhibits activation of nuclear factor-kappaB (NF-kappaB) by forming a complex with NF-kappaB essential modulator (NEMO) and nucleolin, *Mol. Cancer Ther.* 5 (2006) 1790–1799.
- [35] K.T. Shum, J.A. Tanner, Differential inhibitory activities and stabilisation of DNA aptamers against the SARS coronavirus helicase, *Chembiochem* 9 (2008) 3037–3045.
- [36] J. Hu, J. Wu, C. Li, L. Zhu, W.Y. Zhang, G. Kong, Z. Lu, C.J. Yang, A G-quadruplex aptamer inhibits the phosphatase activity of oncogenic protein Shp2 in vitro, *Chembiochem* 12 (2011) 424–430.
- [37] K.T. Shum, E.L. Lui, S.C. Wong, P. Yeung, L. Sam, Y. Wang, R.M. Watt, J.A. Tanner, Aptamer-mediated inhibition of *Mycobacterium tuberculosis* polyphosphate kinase 2, *Biochemistry* 50 (2011) 3261–3271.
- [38] C. Schaffitzel, I. Berger, J. Postberg, J. Hanes, H.J. Lipps, A. Pluckthun, In vitro generated antibodies specific for telomeric guanine-quadruplex DNA react with *Stylomychia lemnae* macronuclei, *Proc. Natl. Acad. Sci. U.S.A.* 98 (2001) 8572–8577.
- [39] H. Fernando, R. Rodriguez, S. Balasubramanian, Selective recognition of a DNA G-quadruplex by an engineered antibody, *Biochemistry* 47 (2008) 9365–9371.
- [40] H. Fernando, S. Sewitz, J. Darot, S. Tavare, J.L. Huppert, S. Balasubramanian, Genome-wide analysis of a G-quadruplex-specific single-chain antibody that regulates gene expression, *Nucleic Acids Res.* 37 (2009) 6716–6722.
- [41] M. Isalan, S.D. Patel, S. Balasubramanian, Y. Choo, Selection of zinc fingers that bind single-stranded telomeric DNA in the G-quadruplex conformation, *Biochemistry* 40 (2001) 830–836.
- [42] S.D. Patel, M. Isalan, G. Gavory, S. Ladame, Y. Choo, S. Balasubramanian, Inhibition of human telomerase activity by an engineered zinc finger protein that binds G-quadruplexes, *Biochemistry* 43 (2004) 13452–13458.
- [43] S. Ladame, J.A. Schouten, J. Roldan, J.E. Redman, S. Neidle, S. Balasubramanian, Exploring the recognition of quadruplex DNA by an engineered Cys2–His2 zinc finger protein, *Biochemistry* 45 (2006) 1393–1399.
- [44] J. Dai, M. Carver, D. Yang, Polymorphism of human telomeric quadruplex structures, *Biochimie* 90 (2008) 1172–1183.
- [45] W. Palm, T. de Lange, How shelterin protects mammalian telomeres, *Annu. Rev. Genet.* 42 (2008) 301–334.
- [46] M.J. Giraud-Panis, M.T. Teixeira, V. Geli, E. Gilson, CST meets shelterin to keep telomeres in check, *Mol. Cell.* 39 (2010) 665–676.
- [47] A. Smogorzewska, T. de Lange, Regulation of telomerase by telomeric proteins, *Annu. Rev. Biochem.* 73 (2004) 177–208.
- [48] M.P. Horvath, V.L. Schweiker, J.M. Bevilacqua, S.C. Schultz, Crystal structure of the *Oxytricha nova* telomere end binding protein complexed with single strand DNA, *Cell* 95 (1998) 963–974.
- [49] G. Fang, T.R. Cech, The beta subunit of *Oxytricha* telomere-binding protein promotes G-quartet formation by telomeric DNA, *Cell* 74 (1993) 875–885.
- [50] M.P. Horvath, S.C. Schultz, DNA G-quartets in a 1.86 Å resolution structure of an *Oxytricha nova* telomeric protein–DNA complex, *J. Mol. Biol.* 310 (2001) 367–377.
- [51] K. Paeschke, T. Simonsson, J. Postberg, D. Rhodes, H.J. Lipps, Telomere end-binding proteins control the formation of G-quadruplex DNA structures in vivo, *Nat. Struct. Mol. Biol.* 12 (2005) 847–854.
- [52] F. Wang, E.R. Podell, A.J. Zaugg, Y. Yang, P. Baciú, T.R. Cech, M. Lei, The POT1–TPP1 telomere complex is a telomerase processivity factor, *Nature* 445 (2007) 506–510.
- [53] A.J. Zaugg, E.R. Podell, T.R. Cech, Human POT1 disrupts telomeric G-quadruplexes allowing telomerase extension in vitro, *Proc. Natl. Acad. Sci. U.S.A.* 102 (2005) 10864–10869.
- [54] A. Furukawa, H. Torigoe, Tetraplex structure of fission yeast telomeric DNA and its unfolding by the interaction with telomeric DNA binding protein Pot1, *Nucleic Acids Symp. Ser. (Oxf.)* (2005) 63–64.
- [55] K. Kaneda, H. Torigoe, Interaction between tetraplex structure of mouse telomeric DNA and telomeric DNA binding domains of mouse telomere binding protein Pot1, *Nucleic Acids Symp. Ser. (Oxf.)* (2009) 245–246.
- [56] H. Torigoe, Fission yeast telomeric DNA binding protein Pot1 has the ability to unfold tetraplex structure of telomeric DNA, *Nucleosides Nucleotides* 26 (2007) 1255–1260.
- [57] Y.C. Lin, J.W. Shih, C.L. Hsu, J.J. Lin, Binding and partial denaturing of G-quartet DNA by Cdc13p of *Saccharomyces cerevisiae*, *J. Biol. Chem.* 276 (2001) 47671–47674.
- [58] M.L. Zhang, X.J. Tong, X.H. Fu, B.O. Zhou, J. Wang, X.H. Liao, Q.J. Li, N. Shen, J. Ding, J.Q. Zhou, Yeast telomerase subunit Est1p has guanine quadruplex-promoting activity that is required for telomere elongation, *Nat. Struct. Mol. Biol.* 17 (2010) 202–209.
- [59] M. Cohn, E.H. Blackburn, Telomerase in yeast, *Science* 269 (1995) 396–400.
- [60] J. Lingner, T.R. Cech, T.R. Hughes, V. Lundblad, Three ever shorter telomere (EST) genes are dispensable for in vitro yeast telomerase activity, *Proc. Natl. Acad. Sci. U.S.A.* 94 (1997) 11190–11195.
- [61] G.H. Yanez, S.J. Khan, A.M. Locovei, I.M. Pedroso, T.M. Fletcher, DNA structure-dependent recruitment of telomeric proteins to single-stranded/double-stranded DNA junctions, *Biochem. Biophys. Res. Commun.* 328 (2005) 49–56.
- [62] R. Giraldo, M. Suzuki, L. Chapman, D. Rhodes, Promotion of parallel DNA quadruplexes by a yeast telomere binding protein: a circular dichroism study, *Proc. Natl. Acad. Sci. U.S.A.* 91 (1994) 7658–7662.
- [63] I.M. Pedroso, W. Hayward, T.M. Fletcher, The effect of the TRF2 N-terminal and TRFH regions on telomeric G-quadruplex structures, *Nucleic Acids Res.* 37 (2009) 1541–1554.
- [64] A. Rupnik, N.F. Lowndes, M. Grenon, MRN and the race to the break, *Chromosoma* 119 (2010) 115–135.
- [65] G. Ghosal, K. Muniyappa, *Saccharomyces cerevisiae* Mre11 is a high-affinity G4 DNA-binding protein and a G-rich DNA-specific endonuclease: implications for replication of telomeric DNA, *Nucleic Acids Res.* 33 (2005) 4692–4703.
- [66] G. Ghosal, K. Muniyappa, The characterization of *Saccharomyces cerevisiae* Mre11/Rad50/Xrs2 complex reveals that Rad50 negatively regulates Mre11 endonucleolytic but not the exonucleolytic activity, *J. Mol. Biol.* 372 (2007) 864–882.
- [67] M. Ueno, Roles of DNA repair proteins in telomere maintenance, *Biosci. Biotechnol. Biochem.* 74 (2010) 1–6.
- [68] T.R. Salas, I. Petrusseva, O. Lavrik, A. Bourdoncle, J.L. Mergny, A. Favre, C. Saintome, Human replication protein A unfolds telomeric G-quadruplexes, *Nucleic Acids Res.* 34 (2006) 4857–4865.

- [69] J.H. Fan, E. Bochkareva, A. Bochkarev, D.M. Gray, Circular dichroism spectra and electrophoretic mobility shift assays show that human replication protein A binds and melts intramolecular G-quadruplex structures, *Biochemistry* 48 (2009) 1099–1111.
- [70] P.W. Lewis, S.J. Elsaesser, K.M. Noh, S.C. Stadler, C.D. Allis, Daxx is an H3.3-specific histone chaperone and cooperates with ATRX in replication-independent chromatin assembly at telomeres, *Proc. Natl. Acad. Sci. U.S.A.* 107 (2010) 14075–14080.
- [71] M.J. Law, K.M. Lower, H.P. Voon, J.R. Hughes, D. Garrick, V. Viprakasit, M. Mitson, M. De Gobbi, M. Marra, A. Morris, A. Abbott, S.P. Wilder, S. Taylor, G.M. Santos, J. Cross, H. Ayyub, S. Jones, J. Ragoussis, D. Rhodes, I. Dunham, D.R. Higgs, R.J. Gibbons, ATR-X syndrome protein targets tandem repeats and influences allele-specific expression in a size-dependent manner, *Cell* 143 (2010) 367–378.
- [72] K. Paeschke, K.R. McDonald, V.A. Zakian, Telomeres: structures in need of unwinding, *FEBS Lett.* 584 (2010) 3760–3772.
- [73] S. Bhattacharyya, A. Sandy, J. Groden, Unwinding protein complexes in ALTernative telomere maintenance, *J. Cell. Biochem.* 109 (2010) 7–15.
- [74] K. Hanada, I.D. Hickson, Molecular genetics of RecQ helicase disorders, *Cell. Mol. Life Sci.* 64 (2007) 2306–2322.
- [75] Y. Wu, R.M. Brosh Jr., Helicase-inactivating mutations as a basis for dominant negative phenotypes, *Cell Cycle* 9 (2010) 4080–4090.
- [76] P. Mohaghegh, J.K. Karow, R.M. Brosh Jr., V.A. Bohr, I.D. Hickson, The Bloom's and Werner's syndrome proteins are DNA structure-specific helicases, *Nucleic Acids Res.* 29 (2001) 2843–2849.
- [77] Y. Wu, R.M. Brosh Jr., G-quadruplex nucleic acids and human disease, *FEBS J.* 277 (2010) 3470–3488.
- [78] C. Molenaar, K. Wiesmeijer, N.P. Verwoerd, S. Khazen, R. Eils, H.J. Tanke, R.W. Dirks, Visualizing telomere dynamics in living mammalian cells using PNA probes, *EMBO J.* 22 (2003) 6631–6641.
- [79] H. Kaneko, N. Kondo, Clinical features of Bloom syndrome and function of the causative gene, BLM helicase, *Expert Rev. Mol. Diagn.* 4 (2004) 393–401.
- [80] M. Fry, The Werner syndrome helicase–nuclease – one protein, many mysteries, *Sci. Aging Knowledge Environ.* 2002 (2002) re2.
- [81] J.P. de Winter, H. Joenje, The genetic and molecular basis of Fanconi anemia, *Mutat. Res.* 668 (2009) 11–19.
- [82] M.L. Rossi, A.K. Ghosh, V.A. Bohr, Roles of Werner syndrome protein in protection of genome integrity, *DNA Repair (Amst.)* 9 (2010) 331–344.
- [83] L. Wu, Role of the BLM helicase in replication fork management, *DNA Repair (Amst.)* 6 (2007) 936–944.
- [84] M. Fry, L.A. Loeb, Human Werner syndrome DNA helicase unwinds tetrahelical structures of the fragile X syndrome repeat sequence d(CGG)_n, *J. Biol. Chem.* 274 (1999) 12797–12802.
- [85] A.S. Kamath-Loeb, L.A. Loeb, E. Johansson, P.M. Burgers, M. Fry, Interactions between the Werner syndrome helicase and DNA polymerase delta specifically facilitate copying of tetraplex and hairpin structures of the d(CGG)_n trinucleotide repeat sequence, *J. Biol. Chem.* 276 (2001) 16439–16446.
- [86] V. Popuri, C.Z. Bachrati, L. Muzzolini, G. Mosedale, S. Costantini, E. Giacomini, I.D. Hickson, A. Vindigni, The Human RecQ helicases, BLM and RECQ1, display distinct DNA substrate specificities, *J. Biol. Chem.* 283 (2008) 17766–17776.
- [87] M.D. Huber, M.L. Duquette, J.C. Shiels, N. Maizels, A conserved G4 DNA binding domain in RecQ family helicases, *J. Mol. Biol.* 358 (2006) 1071–1080.
- [88] J.Q. Liu, C.Y. Chen, Y. Xue, Y.H. Hao, Z. Tan, G-quadruplex hinders translocation of BLM helicase on DNA: a real-time fluorescence spectroscopic unwinding study and comparison with duplex substrates, *J. Am. Chem. Soc.* 132 (2010) 10521–10527.
- [89] M.T. Russo, G. De Luca, P. Degan, M. Bignami, Different DNA repair strategies to combat the threat from 8-oxoguanine, *Mutat. Res.* 614 (2007) 69–76.
- [90] A. Ghosh, M.L. Rossi, J. Aulds, D. Croteau, V.A. Bohr, Telomeric D-loops containing 8-oxo-2'-deoxyguanosine are preferred substrates for Werner and Bloom syndrome helicases and are bound by POT1, *J. Biol. Chem.* 284 (2009) 31074–31084.
- [91] E. Kruisselbrink, V. Guryev, K. Brouwer, D.B. Pontier, E. Cuppen, M. Tijsterman, Mutagenic capacity of endogenous G4 DNA underlies genome instability in FANCD1-defective *C. elegans*, *Curr. Biol.* 18 (2008) 900–905.
- [92] T.B. London, L.J. Barber, G. Mosedale, G.P. Kelly, S. Balasubramanian, I.D. Hickson, S.J. Boulton, K. Hiom, FANCD1 is a structure-specific DNA helicase associated with the maintenance of genomic G/C tracts, *J. Biol. Chem.* 283 (2008) 36132–36139.
- [93] Y. Wu, K. Shin-ya, R.M. Brosh Jr., FANCD1 helicase defective in Fanconi anemia and breast cancer unwinds G-quadruplex DNA to defend genomic stability, *Mol. Cell. Biol.* 28 (2008) 4116–4128.
- [94] C. Ribeyre, J. Lopes, J.B. Boule, A. Piazza, A. Guedin, V.A. Zakian, J.L. Mergny, A. Nicolas, The yeast Pif1 helicase prevents genomic instability caused by G-quadruplex-forming CEB1 sequences in vivo, *PLoS Genet.* 5 (2009) e1000475.
- [95] C.M. Sanders, Human Pif1 helicase is a G-quadruplex DNA-binding protein with G-quadruplex DNA-unwinding activity, *Biochem. J.* 430 (2010) 119–128.
- [96] T. Masuda-Sasa, P. Polaczek, X.P. Peng, L. Chen, J.L. Campbell, Processing of G4 DNA by Dna2 helicase/nuclease and replication protein A (RPA) provides insights into the mechanism of Dna2/RPA substrate recognition, *J. Biol. Chem.* 283 (2008) 24359–24373.
- [97] C. Harrington, Y. Lan, S.A. Akman, The identification and characterization of a G4-DNA resolvase activity, *J. Biol. Chem.* 272 (1997) 24631–24636.
- [98] J.P. Vaughn, S.D. Creacy, E.D. Routh, C. Joyner-Butt, G.S. Jenkins, S. Pauli, Y. Nagamine, S.A. Akman, The DEXH protein product of the DHX36 gene is the major source of tetramolecular quadruplex G4-DNA resolving activity in HeLa cell lysates, *J. Biol. Chem.* 280 (2005) 38117–38120.
- [99] S.D. Creacy, E.D. Routh, F. Iwamoto, Y. Nagamine, S.A. Akman, J.P. Vaughn, G4 resolvase 1 binds both DNA and RNA tetramolecular quadruplex with high affinity and is the major source of tetramolecular quadruplex G4-DNA and G4-RNA resolving activity in HeLa cell lysates, *J. Biol. Chem.* 283 (2008) 34626–34634.
- [100] S. Lattmann, B. Giri, J.P. Vaughn, S.A. Akman, Y. Nagamine, Role of the amino terminal RHAU-specific motif in the recognition and resolution of guanine quadruplex-RNA by the DEAH-box RNA helicase RHAU, *Nucleic Acids Res.* 38 (2010) 6219–6233.
- [101] A.N. Sexton, K. Collins, The 5' guanosine tracts of human telomerase RNA are recognized by the G-quadruplex binding domain of the RNA helicase DHX36 and function to increase RNA accumulation, *Mol. Cell. Biol.* 31 (2011) 736–743.
- [102] N. Baran, L. Pucshansky, Y. Marco, S. Benjamin, H. Manor, The SV40 large T-antigen helicase can unwind four stranded DNA structures linked by G-quartets, *Nucleic Acids Res.* 25 (1997) 297–303.
- [103] S.P. Han, Y.H. Tang, R. Smith, Functional diversity of the hnRNPs: past, present and perspectives, *Biochem. J.* 430 (2010) 379–392.
- [104] S. Khateb, P. Weisman-Shomer, I. Hershco, L.A. Loeb, M. Fry, Destabilization of tetraplex structures of the fragile X repeat sequence (CGG)_n is mediated by homolog-conserved domains in three members of the hnRNP family, *Nucleic Acids Res.* 32 (2004) 4145–4154.
- [105] P. Weisman-Shomer, E. Cohen, M. Fry, Distinct domains in the CarG-box binding factor A destabilize tetraplex forms of the fragile X expanded sequence d(CGG)_n, *Nucleic Acids Res.* 30 (2002) 3672–3681.
- [106] Y. Enokizono, Y. Konishi, K. Nagata, K. Ouhashi, S. Uesugi, F. Ishikawa, M. Katahira, Structure of hnRNP D complexed with single-stranded telomere DNA and unfolding of the quadruplex by heterogeneous nuclear ribonucleoprotein D, *J. Biol. Chem.* 280 (2005) 18862–18870.
- [107] Q.S. Zhang, L. Manche, R.M. Xu, A.R. Krainer, hnRNP A1 associates with telomere ends and stimulates telomerase activity, *Rna* 12 (2006) 1116–1128.
- [108] A.C. Kruger, M.K. Raarup, M.M. Nielsen, M. Kristensen, F. Besenbacher, J. Kjems, V. Birkedal, Interaction of hnRNP A1 with telomere DNA G-quadruplex structures studied at the single molecule level, *Eur. Biophys. J.* 39 (2010) 1343–1350.
- [109] R.M. Xu, L. Jokhan, X. Cheng, A. Mayeda, A.R. Krainer, Crystal structure of human UP1, the domain of hnRNP A1 that contains two RNA-recognition motifs, *Structure* 5 (1997) 559–570.
- [110] H. Fukuda, M. Katahira, N. Tsuchiya, Y. Enokizono, T. Sugimura, M. Nagao, H. Nakagama, Unfolding of quadruplex structure in the G-rich strand of the minisatellite repeat by the binding protein UP1, *Proc. Natl. Acad. Sci. U.S.A.* 99 (2002) 12685–12690.
- [111] H. Nakagama, K. Higuchi, E. Tanaka, N. Tsuchiya, K. Nakashima, M. Katahira, H. Fukuda, Molecular mechanisms for maintenance of G-rich short tandem repeats capable of adopting G4 DNA structures, *Mutat. Res.* 598 (2006) 120–131.
- [112] T. Nagata, Y. Takada, A. Ono, K. Nagata, Y. Konishi, T. Nukina, M. Ono, A. Matsugami, A. Furukawa, N. Fujimoto, H. Fukuda, H. Nakagama, M. Katahira, Elucidation of the mode of interaction in the UP1–telomerase RNA–telomeric DNA ternary complex which serves to recruit telomerase to telomeric DNA and to enhance the telomerase activity, *Nucleic Acids Res.* 36 (2008) 6816–6824.
- [113] J.L. Huppert, S. Balasubramanian, G-quadruplexes in promoters throughout the human genome, *Nucleic Acids Res.* 35 (2007) 406–413.
- [114] J.L. Huppert, Structure, location and interactions of G-quadruplexes, *FEBS J.* 277 (2010) 3452–3458.
- [115] T.A. Brooks, S. Kendrick, L. Hurley, Making sense of G-quadruplex and i-motif functions in oncogene promoters, *FEBS J.* 277 (2010) 3459–3469.
- [116] T.A. Brooks, L.H. Hurley, Targeting MYC expression through G-quadruplexes, *Genes Cancer* 1 (2010) 641–649.
- [117] V. Gonzalez, L.H. Hurley, The c-MYC NHE III(1): function and regulation, *Annu. Rev. Pharmacol. Toxicol.* 50 (2010) 111–129.
- [118] E.H. Postel, Multiple biochemical activities of NM23/NDP kinase in gene regulation, *J. Bioenerg. Biomembr.* 35 (2003) 31–40.
- [119] R.K. Thakur, P. Kumar, K. Halder, A. Verma, A. Kar, J.L. Parent, R. Basundra, A. Kumar, S. Chowdhury, Metastases suppressor NM23-H2 interaction with G-quadruplex DNA within c-MYC promoter nuclease hypersensitive element induces c-MYC expression, *Nucleic Acids Res.* 37 (2009) 172–183.
- [120] F. Mongelard, P. Bouvet, Nucleolin: a multifunctional protein, *Trends Cell. Biol.* 17 (2007) 80–86.
- [121] V. Gonzalez, L.H. Hurley, The C-terminus of nucleolin promotes the formation of the c-MYC G-quadruplex and inhibits c-MYC promoter activity, *Biochemistry* 49 (2010) 9706–9714.
- [122] V. Gonzalez, K. Guo, L. Hurley, D. Sun, Identification and characterization of nucleolin as a c-myc G-quadruplex-binding protein, *J. Biol. Chem.* 284 (2009) 23622–23635.
- [123] P. Armas, S. Nasif, N.B. Calcaterra, Cellular nucleic acid binding protein binds G-rich single-stranded nucleic acids and may function as a nucleic acid chaperone, *J. Cell. Biochem.* 103 (2008) 1013–1036.
- [124] N.B. Calcaterra, P. Armas, A.M. Weiner, M. Borgognone, CNBP: a multifunctional nucleic acid chaperone involved in cell death and proliferation control, *IUBMB Life* 62 (2010) 707–714.
- [125] M. Borgognone, P. Armas, N.B. Calcaterra, Cellular nucleic-acid-binding protein, a transcriptional enhancer of c-Myc, promotes the formation of parallel G-quadruplexes, *Biochem. J.* 428 (2010) 491–498.

- [126] S. Cogoi, M. Paramasivam, B. Spolaore, L.E. Xodo, Structural polymorphism within a regulatory element of the human KRAS promoter: formation of G4-DNA recognized by nuclear proteins, *Nucleic Acids Res.* 36 (2008) 3765–3780.
- [127] S. Cogoi, M. Paramasivam, V. Filichev, I. Geci, E.B. Pedersen, L.E. Xodo, Identification of a new G-quadruplex motif in the KRAS promoter and design of pyrene-modified G4-decoys with antiproliferative activity in pancreatic cancer cells, *J. Med. Chem.* 52 (2009) 564–568.
- [128] M. Paramasivam, A. Membrino, S. Cogoi, H. Fukuda, H. Nakagama, L.E. Xodo, Protein hnRNP A1 and its derivative Up1 unfold quadruplex DNA in the human KRAS promoter: implications for transcription, *Nucleic Acids Res.* 37 (2009) 2841–2853.
- [129] L. Xodo, M. Paramasivam, A. Membrino, S. Cogoi, Protein hnRNP1 binds to a critical G-rich element of KRAS and unwinds G-quadruplex structures: implications in transcription, *Nucleic Acids Symp. Ser. (Oxf.)* (2008) 159–160.
- [130] S. Cogoi, M. Paramasivam, A. Membrino, K.K. Yokoyama, L.E. Xodo, The KRAS promoter responds to Myc-associated zinc finger and poly(ADP-ribose) polymerase 1 p.oteins, which recognize a critical quadruplex-forming GA-element, *J. Biol. Chem.* 285 (2010) 22003–22016.
- [131] V.A. Soldatenkov, A.A. Vetcher, T. Duka, S. Ladame, First evidence of a functional interaction between DNA quadruplexes and poly(ADP-ribose) polymerase-1, *ACS Chem. Biol.* 3 (2008) 214–219.
- [132] S.L. Palumbo, R.M. Memmott, D.J. Uribe, Y. Krotova-Khan, L.H. Hurley, S.W. Ebbinghaus, A novel G-quadruplex-forming GGA repeat region in the c-myc promoter is a critical regulator of promoter activity, *Nucleic Acids Res.* 36 (2008) 1755–1769.
- [133] K. Guo, V. Gokhale, L.H. Hurley, D. Sun, Intramolecularly folded G-quadruplex and i-motif structures in the proximal promoter of the vascular endothelial growth factor gene, *Nucleic Acids Res.* 36 (2008) 4598–4608.
- [134] D. Sun, K. Guo, J.J. Rusche, L.H. Hurley, 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 Res.* 33 (2005) 6070–6080.
- [135] D. Sun, K. Guo, Y.J. Shin, Evidence of the formation of G-quadruplex structures in the promoter region of the human vascular endothelial growth factor gene, *Nucleic Acids Res.* 39 (2011) 1256–1265.
- [136] D. Sun, W.J. Liu, K. Guo, J.J. Rusche, S. Ebbinghaus, V. Gokhale, L.H. Hurley, 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 (2008) 880–889.
- [137] Y. Wang, H. Zhang, L.A. Ligon, L.B. McGown, Association of insulin-like growth factor 2 with the insulin-linked polymorphic region in cultured fetal thymus cells, *Biochemistry* 48 (2009) 8189–8194.
- [138] J. Xiao, J.A. Carter, K.A. Frederick, L.B. McGown, A genome-inspired DNA ligand for the affinity capture of insulin and insulin-like growth factor-2, *J. Sep. Sci.* 32 (2009) 1654–1664.
- [139] J. Xiao, L.B. McGown, Mass spectrometric determination of ILPR G-quadruplex binding sites in insulin and IGF-2, *J. Am. Soc. Mass Spectrom.* 20 (2009) 1974–1982.
- [140] J.D. Schonhoft, A. Das, F. Achamyeh, S. Samdani, A. Sewell, H. Mao, S. Basu, ILPR repeats adopt diverse G-quadruplex conformations that determine insulin binding, *Biopolymers* 93 (2010) 21–31.
- [141] A. Yafe, J. Shklover, P. Weisman-Shomer, E. Bengal, M. Fry, Differential binding of quadruplex structures of muscle-specific genes regulatory sequences by MyoD, MRF4 and myogenin, *Nucleic Acids Res.* 36 (2008) 3916–3925.
- [142] J. Shklover, P. Weisman-Shomer, A. Yafe, M. Fry, Quadruplex structures of muscle gene promoter sequences enhance in vivo MyoD-dependent gene expression, *Nucleic Acids Res.* 38 (2010) 2369–2377.
- [143] M. Nakagawa, Y. Kameoka, R. Suzuki, Nucleophosmin in acute myelogenous leukemia, *N. Engl. J. Med.* 352 (2005) 1819–1820; author reply.
- [144] L. Federici, A. Arcovito, G.L. Scaglione, F. Scaloni, C. Lo Sterzo, A. Di Matteo, B. Falini, B. Giardina, M. Brunori, Nucleophosmin C-terminal leukemia-associated domain interacts with G-rich quadruplex forming DNA, *J. Biol. Chem.* 285 (2010) 37138–37149.
- [145] R. Basundra, A. Kumar, S. Amrane, A. Verma, A.T. Phan, S. Chowdhury, A novel G-quadruplex motif modulates promoter activity of human thymidine kinase 1, *FEBS J.* 277 (2010) 4254–4264.
- [146] H. Tiedge, RNA reigns in neurons, *Neuron* 48 (2005) 13–16.
- [147] G.S. Wilkie, K.S. Dickson, N.K. Gray, Regulation of mRNA translation by 5'- and 3'-UTR-binding factors, *Trends Biochem. Sci.* 28 (2003) 182–188.
- [148] J.D. Beaudoin, J.P. Perreault, 5'-UTR G-quadruplex structures acting as translational repressors, *Nucleic Acids Res.* 38 (2010) 7022–7036.
- [149] J.L. Huppert, A. Bugaut, S. Kumari, S. Balasubramanian, G-quadruplexes: the beginning and end of UTRs, *Nucleic Acids Res.* 36 (2008) 6260–6268.
- [150] D.J. Patel, A.T. Phan, V. Kuryavyi, Human telomere, oncogenic promoter and 5'-UTR G-quadruplexes: diverse higher order DNA and RNA targets for cancer therapeutics, *Nucleic Acids Res.* 35 (2007) 7429–7455.
- [151] S. Kumari, A. Bugaut, J.L. Huppert, S. Balasubramanian, An RNA G-quadruplex in the 5' UTR of the NRAS proto-oncogene modulates translation, *Nat. Chem. Biol.* 3 (2007) 218–221.
- [152] A. Marnef, N. Standart, Pat1 p.oteins: a life in translation, translation repression and mRNA decay, *Biochem. Soc. Trans.* 38 (2010) 1602–1607.
- [153] A. Marnef, M. Maldonado, A. Bugaut, S. Balasubramanian, M. Kress, D. Weil, N. Standart, Distinct functions of maternal and somatic Pat1 p.oteins paralogs, *Rna* 16 (2010) 2094–2107.
- [154] J. Tan, C. Vonrhein, O.S. Smart, G. Bricogne, M. Bollati, Y. Kusov, G. Hansen, J.R. Mesters, C.L. Schmidt, R. Hilgenfeld, The SARS-unique domain (SUD) of SARS coronavirus contains two macrodomains that bind G-quadruplexes, *PLoS Pathog.* 5 (2009) e1000428.
- [155] J. Norseen, F.B. Johnson, P.M. Lieberman, Role for G-quadruplex RNA binding by Epstein-Barr virus nuclear antigen 1 in DNA replication and metaphase chromosome attachment, *J. Virol.* 83 (2009) 10336–10346.
- [156] J.R. Brouwer, R. Willemsen, B.A. Oostra, The FMR1 gene and fragile X-associated tremor/ataxia syndrome, *Am. J. Med. Genet. B Neuropsychiatr. Genet.* 150B (2009) 782–798.
- [157] S. Khateb, P. Weisman-Shomer, I. Hershco-Shani, A.L. Ludwig, M. Fry, The tetraplex (CGG)_n destabilizing proteins hnRNP A2 and CBF-A enhance the in vivo translation of fragile X premutation mRNA, *Nucleic Acids Res.* 35 (2007) 5775–5788.
- [158] M. Bole, L. Menon, M.R. Mihailescu, Fragile X mental retardation protein recognition of G quadruplex structure per se is sufficient for high affinity binding to RNA, *Mol. Biosyst.* 4 (2008) 1212–1219.
- [159] E. Blackwell, X. Zhang, S. Ceman, Arginines of the RGG box regulate FMRP association with polyribosomes and mRNA, *Hum. Mol. Genet.* 19 (2010) 1314–1323.
- [160] H. Fukuda, M. Katahira, E. Tanaka, Y. Enokizono, N. Tsuchiya, K. Higuchi, M. Nagao, H. Nakagama, Unfolding of higher DNA structures formed by the d(CGG) triplet repeat by UP1 p.oteins, *Genes Cells* 10 (2005) 953–962.
- [161] G. Sarig, P. Weisman-Shomer, M. Fry, Telomeric and tetraplex DNA binding properties of qTBP42: a homologue of the CarG box binding protein CBF-A, *Biochem. Biophys. Res. Commun.* 237 (1997) 617–623.
- [162] P. Weisman-Shomer, Y. Naot, M. Fry, Tetrahelical forms of the fragile X syndrome expanded sequence d(CGG)_n are destabilized by two heterogeneous nuclear ribonucleoprotein-related telomeric DNA-binding proteins, *J. Biol. Chem.* 275 (2000) 2231–2238.
- [163] M. Bensaid, M. Melko, E.G. Bechara, L. Davidovic, A. Berretta, M.V. Catania, J. Gecez, E. Lalli, B. Bardoni, FRA-XE-associated mental retardation protein (FMR2) is an RNA-binding protein with high affinity for G-quartet RNA forming structure, *Nucleic Acids Res.* 37 (2009) 1269–1279.
- [164] M. Melko, B. Bardoni, The role of G-quadruplex in RNA metabolism: involvement of FMRP and FMR2P, *Biochimie* 92 (2010) 919–926.
- [165] K.A. Corbin-Lickfett, I.H. Chen, M.J. Cocco, R.M. Sandri-Goldin, The HSV-1 ICP27 RGG box specifically binds flexible, GC-rich sequences but not G-quartet structures, *Nucleic Acids Res.* 37 (2009) 7290–7301.
- [166] J.C. Darnell, C.E. Fraser, O. Mostovetsky, R.B. Darnell, Discrimination of common and unique RNA-binding activities among Fragile X mental retardation protein paralogs, *Hum. Mol. Genet.* 18 (2009) 3164–3177.
- [167] S. Neidle, The structures of quadruplex nucleic acids and their drug complexes, *Curr. Opin. Struct. Biol.* 19 (2009) 239–250.
- [168] M. Egli, P.S. Pallan, The many twists and turns of DNA: template, telomere, tool, and target, *Curr. Opin. Struct. Biol.* 20 (2010) 262–275.
- [169] S. Neidle, G.N. Parkinson, Quadruplex DNA crystal structures and drug design, *Biochimie* 90 (2008) 1184–1196.
- [170] S. Balasubramanian, S. Neidle, G-quadruplex nucleic acids as therapeutic targets, *Curr. Opin. Chem. Biol.* 13 (2009) 345–353.
- [171] J.L. Huppert, Four-stranded DNA: cancer, gene regulation and drug development, *Philos. Transact. A Math. Phys. Eng. Sci.* 365 (2007) 2969–2984.
- [172] D.H. Zhang, T. Fujimoto, S. Saxena, H.Q. Yu, D. Miyoshi, N. Sugimoto, Monomeric RNA G-quadruplex and polymorphic DNA G-quadruplex structures responding to cellular environmental factors, *Biochemistry* 49 (2010) 4554–4563.
- [173] S.S. Patel, K.M. Picha, Structure and function of hexameric helicases, *Annu. Rev. Biochem.* 69 (2000) 651–697.
- [174] J. Dai, E. Hatzakis, L.H. Hurley, D. Yang, I-motif structures formed in the human c-MYC promoter are highly dynamic—insights into sequence redundancy and I-motif stability, *PLoS One* 5 (2010) e11647.
- [175] S. Kendrick, Y. Akiyama, S.M. Hecht, L.H. Hurley, The i-motif in the bcl-2 P1 p.otein forms an unexpectedly stable structure with a unique 8:5:7 loop folding pattern, *J. Am. Chem. Soc.* 131 (2009) 17667–17676.
- [176] S. Dhakal, J.D. Schonhoft, D. Koirala, Z. Yu, S. Basu, H. Mao, Coexistence of an ILPR i-motif and a partially folded structure with comparable mechanical stability revealed at the single-molecule level, *J. Am. Chem. Soc.* 132 (2010) 8991–8997.
- [177] S. Feuerhahn, N. Iglesias, A. Panza, A. Porro, J. Lingner, TERRA biogenesis, turnover and implications for function, *FEBS Lett.* 584 (2010) 3812–3818.
- [178] Z. Deng, J. Norseen, A. Wiedmer, H. Riethman, P.M. Lieberman, TERRA RNA binding to TRF2 facilitates heterochromatin formation and ORC recruitment at telomeres, *Mol. Cell.* 35 (2009) 403–413.
- [179] S. Redon, P. Reichenbach, J. Lingner, The non-coding RNA TERRA is a natural ligand and direct inhibitor of human telomerase, *Nucleic Acids Res.* 38 (2010) 5797–5806.