Insights of potential G-quadruplex sequences in telomeres and proto-oncogenes

Rajendra N. Bhadane¹, Rupali R. Bhadane², Dhananjay B. Meshram³

SUMMARY

Guanine rich sequences have the ability to fold into stable 4 stranded structures called G-quadruplex under physiological concentrations of Na⁺ or K⁺. G-quadruplexes are found in telomeres, being stable structures under the control of telomerase binding proteins. They are also identified throughout the genome and are enriched in promoter regions of protein coding genes, upstream and downstream of the transcription initiation sites. A number of these promoter quadruplexes have been investigated for several proto-oncogenes. The formation of these quadruplexes can lead to chemical intervention of gene expression using a G-quadruplex binding ligand. We review location, configuration, and stabilization of these quadruplexes in some of the important promoters with regards to their potential as anticancer target.

KEY WORDS: G-Quadruplexes; Telomerase; Guanine; Telomere; Proto-Onecogenes

INTRODUCTION

The earliest research in the antineoplastic drug discovery was related to suppressing the synthesis and function of DNA. Today, a variety of other targets is under intensive investigation and they will provide oncologist with significant new approaches of therapy. Some of these approaches are inhibition of protease involved in metastasis, angiogenesis inhibitors, antisense technology, and G-quadruplexes.

G-quadruplexes are generally formed in DNA and RNA sequences containing repeated G-G-G-G called as G-tetrad. G-quadruplexes formed from planar stacking of hoogsteen bonded G-tetrads (1, 2) folded from a single G-rich sequence by intra- or inter-molecular association of 2 to 4 separate strands (3, 4). The core of G-quadruplexes are formed by the stacking of several G-tetrads and joined together by sugar phosphate backbone. The binding energy for this process arises from three main factors: hydrogen bonding between the guanines in a plane, π - π interaction between the guanines in adjacent planes and charge – charge interaction between partially negative oxygen of guanines and cations that typically sit in the octahedral position between the stacks (5-7). The monovalent cations such as K⁺ and Na⁺ at a physiological temperature and pH stabilize G-quadruplex by coordinating the carbonyl group of guanine at the center of G-tetrad core (5, 8).

It has been estimated that there are more than 376,000 potential quadruplex sequences found in number of important biological processes (9). Intramolecular G-quadruplexes formed by single-stranded DNA are currently under intensive research due to their potential formation in telomeres and promoter sequences (10, 11). The present review reports the G-quadruplexes formed in human telomeres and proto-oncogenes.

Telomere and telomerase

A simple, repeated DNA sequence of TTAGGG is found at the end of a linear chromosome called as the telomere. This DNA sequence caps the end of linear chromosomes (12). As DNA polymerase-α is not capable of completing the replication at the end of chromosomes so, telomerase acts by synthesizing and adding telomeric repeats to the end of chromosomes. Telomerase was discovered in 1985, through the ingenuity of Grieter and Blackburn (13, 14). In 1994, it was discovered that cancer cells require telomerase for survival. In somatic cell, telomeres are progressively shorten with 20-200 base pairs per cell division during each replication cycle. This erosion of telomeric DNA observed both in vivo and in primary cell culture, (15-17) continuing up to a critical length of 8-10 kilobase pairs (18, 19). After reaching this crucial point, which is also known as the first mortality stage (M1), most of the cells go into senescence, i.e. exit from cell cycle. A mutation in the tumor suppressor gene PS3 and/or the retinoblastoma protein (PRb) enables the cell to bypass M1 leading to the phase of extended life span (20, 21). At this stage, chromosomal instability continues until a second mortality stage (M2) or crisis has reached (22). In crisis average telomere lengths are 3-4Kb. The escape from the crisis is accomplished in two ways, either by reactivation of telomerase or by activation of telomerase restoration mechanism. This cellular immortalization is a potentially rate-limiting step in carcinogenesis that is important for the continuing evolution of most advanced cancers (Figure 1) (15, 20). This finding supports that telomere erosion is an important signal for the aging of the cell leading to the induction of cellular senescence and cell death (12).
Human telomerase

Human telomerase consists of tandem repeats of six nucleotides TTAGGG which are repeated for 5-25 kilobase pairs, 5'-3' towards the chromosome end (12). Mammalian telomerase is riboprotein consisting of the catalytic protein subunit, the protein component, and integral RNA component. The RNA component of human telomerase is about 445 nucleotides in length, which acts as a template for the synthesis of telomeric repeats (23). The protein domain (hTERT), homologous to reverse transcriptase (24-29), catalyses nucleotide polymerization (30, 31).

Role of telomerase in cancer progression

Telomerase activity is absent in normal cells, but is detected in more than 85% of cancer cells (32, 33). This observation implies a role of telomerase in cancer progression. Telomerase is necessary for sustained cell proliferation that characterizes the cancer (34-38). This is supported by the observation that early stage neuroblastomas have low telomerase activity (39). Other studies have suggested that telomerase activity is correlated with pathological stage (40, 41) or tumor aggressiveness (42). Thus, the lack of telomerase limits the growth of rapidly proliferating cells while the increase in telomerase permits indefinite proliferations. The end sequence of the telomere is guanines rich and is called as a G - tail. This G-tail is conserved feature of telomere and is essential for telomerase function (43-47). Due to the high concentration of guanines in G-tail, telomers are able to form stable G-quadruplex structures as shown in Figure 2 (48). The G-quadruplexes are a secondary structures of DNA and affects telomerase function (49). These alternative structures are an obstacle for both semi-conservative and telomerase mediated replication and must be resolved prior to these events (48). Molecules that promote and stabilize the formation of G-quadruplexes could potentially block the action of telomerase on telomeres (50, 51).

C-myc

The C-myc gene is present on human chromosome 8q24. It is expressed in normal cell by external signals such as growth factor and extracellular matrix contacts. An abnormal expression of c-myc in primary cells activates a protective pathway through induction of p19/p14ARF and a p53 dependent cell death pathway. Thus, normal cells, which over express c-myc are eliminated through apoptosis. Activation of c-myc gene, contribute to the development of human cancer. C-myc promotes cell proliferation and genomic instability by accelerating cells through G1 and S phase of the cell cycle. C-myc activates the telomerase hTERT which leads to the immortalization of cells (52). Over expression of c-myc proto-oncogene is associated with a broad spectrum of human cancers, including colon, breast, prostate, cervical and lung carcinomas, osteosarcoma, lymphomas and leukemias (53-61). C-myc over expression of 4 GGA repeats, (3 [GGA] 4), located 17 base pairs downstream of the transcription initiation site on bottom strand. The c-myc promoter can form an unusual secondary structure related to guanine tetrad (T) stacked onto guanine habitat (H) (77, 78). This unique quadruplex was the first G-quadruplex structure with pseudo-double chain reversal loop. DNA sequence with the two adjacent (GCA) 4 units can form a very stable higher ordered structure by intramolecular stacking to two T. H G-quadruplexes on the heptad plane resulting in T: H: H: T DNA structure (Figure 2) (79, 80). The relative orientation of DNA strands in G-quadruplexes were deduced by CD spectroscopy (81). The GGA repeat in gene promoter was the critical transcriptional activator binding site essential for c-myc promoter activity (82, 83). This specifies that a GGA sequence of the telomere is guanines rich and is called as a G - tail. This G-tail is conserved feature of telomere and is essential for telomerase function (43-47). Due to the high concentration of guanines in G-tail, telomers are able to form stable G-quadruplex structures as shown in Figure 2 (48). The G-quadruplexes are a secondary structures of DNA and affects telomerase function (49). These alternative structures are an obstacle for both semi-conservative and telomerase mediated replication and must be resolved prior to these events (48). Molecules that promote and stabilize the formation of G-quadruplexes could potentially block the action of telomerase on telomeres (50, 51).
repeat region have to be available for transcriptional activators to bind to the promoter. G-quadruplex formation, stabilized by small molecules, could also markedly inhibit c-myb expression, making the region unavailable for transcriptional factor binding (84, 85).

C-kit

The C-kit is a type-3 receptor tyrosine kinase, which is characterized by five extra cellular immunoglobin like domain and an intracellular split tyrosine kinase domain (86, 87). The C-kit proto-oncogene encodes a tyrosine kinase receptor for the growth promoting cytokine stem cell factor (SCF), which plays an important role in the control of differentiation (88, 89). The C-kit protein plays a role in oncogenic transformation of certain cell types and are found in several highly malignant human cancers such as gastrointestinal stromal tumors (GIST) (90, 91). The C-kit promoter has two non-overlapping quadruplex motifs (4, 92, 93). A 22 base g-rich sequence between -87 and -109 base pairs upstream from the transcription site of human c-kit gene, has been shown to fold into a quadruplex as in fig. 2 (c-kit-1) (4, 92, 94). A second g-rich sequence is situated -140 to -160 base pairs upstream from the transcription start site represented in fig. 3 (c-kit-2). The formation of G-quadruplexes at both sites was demonstrated by H-NMR, CD, and UV spectroscopy (90, 93).

KRAS

The KRAS gene is located in chromosome 12, and encodes for a 21kDa protein, p21RAS, which is anchored to the inner surface of the plasma membrane and act as molecular switch (95). The endogenous expression of an oncogenic KRAS allele in the mouse pancreas initiates the development of pancreatic ductal adenocarcinoma (95). Strategies aiming suppression of KRAS expression (96) or activation of RAS protein (97) have been proposed. Polypurine-polypyrimidine motifs are simple repeats distributed through the eukaryotic genome (98). These motifs are located in the KRAS promoter between -327 and -296 in human, -318 and -290 in mouse (99, 100). The anti KRAS activity of the endogenous TFO correlates with its ability to assume a G-quadruplex that recognized a nucleus protein binding to the NHPPPE duplex (96, 101). The purine strand of NHPPPE, located in the proximal promoter sequence of KRAS is able to form G-quadruplex (Figure 3) (69). The transfection experiments showed that the stabilization of the KRAS G-quadruplex with cationic porphyrin TMPyP4 results in a strong inhibition of KRAS transcription (69). The quadruplexes formed in the mouse and human promoter is characterized through electrophoretic mobility shift assay (EMSA), circular dichroism (CD), chemical probing with DMS/piperidine and polymerase stop assays (69).

Neuroblastoma RAS viral oncogene homolog (NRAS)

The neuroblastoma RAS viral oncogene homolog (NRAS) encoded protein p21 mediate both signal transduction across the cell membrane and the intracellular signaling pathway responsible for cell proliferation and differentiation (102). A mutation in the coding region of NRAS is responsible for increased cell proliferation (103). The G-quadruplex forming sequence in the 5’ UTR of the human NRAS proto-oncogene mRNA was found (104). It is 254 nucleotides long NRAS 5’ UTR containing NRAS RNA G-quadruplex (NRQ) motif, located 14 nucleotides downstream of the 5’ cap and 222 nucleotide upstream of the translocation start site (105). The configuration of RNA quadruplex NRQ folds into stable quadruplex was done by CD spectroscopy (104). Therefore, inhibition of the expression of oncogenic NRAS can be one of the potential therapeutic strategies.

Vascular endothelial growth factor (VEGF)

The proximal promoter region of the human vascular endothelial growth factor (VEGF) gene contains a polypurine /polypyrimidine tract that serves as multiple binding sites for Sp1 and Egr 1 transcription factor. This tract contains a guanines rich sequence consisting of 4 runs of three or more contiguous guanines separated by one or more bases (106-108) corresponding to be general motifs for the formation of intermolecular G-quadruplex. Starving tumor cells can become angiogenic by gaining the ability to direct the formation of new blood vessels for their survival (109, 110). The angiogenic switch in cancer cell is often initiated by increased expression of vascular endothelial growth factor (VEGF). The VEGF is pluripotent cytokinine and angiogenic growth factor consisting two identical subunits that bind to VEGF receptor on the surface of endothelial cells (111). The interaction of VEGF with receptor stimulates the proliferation, migration, survival, and permeability of endothelial cells, resulting in the formation of new blood vessels (112). The human VEGF promoter region contains binding sites for several transcriptional factors such as HIF-1, AP-1, AP-2, Egr-1, Sp1 and many others (113, 114). It is revealed that the proximal 36 base pair region is essential for inducible VEGF promoter activity in several human cancer...
cells (113, 114). The polyurine tract of the VEGF promoter consists of five runs of at least three contiguous guanines separated by one or more bases capable of forming quadruplex structures. The formation of the G-quadruplexes from these G-rich sequences in the human VEGF promoter were confirmed by NMR, CD spectroscopy, X-ray crystallography (106).

**Bcl-2**

Bcl-2 is a potent oncoprotein that plays an essential role in survival and functions as an inhibitor of cell apoptosis (115). It was first discovered in human follicular lymphoma (116). Bcl-2 is over expressed in a wide range of human tumors, including B-cell and T-cell lymphomas, breast, prostate, cervical, colorectal and non-small lung carcinomas (117). The p1 promoter located 1386-1423 base pairs upstream of the translocation site is the major transcriptional promoter for bcl-2 (118). This is GC rich promoter located 58 to 19 base pairs upstream of the p1 promoter with 39 base pair sequence (118). This GC rich element (bcl2pu39) contains six runs of guanines each forming a series of these distinct intramolecular G-quadruplexes in K+-containing solution (Fig. 3) (119). The structure of these major G-quadruplexes was confirmed by NMR and biochemical studies (119, 120). These major G-quadruplexes in the bcl-2 promoter represents an important target for designing of new anticancer drug that specifically binds these structures and modulates bcl-2 gene expression (120).

**The retinoblastoma (Rb) gene**

The retinoblastoma (Rb) gene encodes nuclear phosphoproteins, which act as a tumor suppressor by affecting the cell cycle. The control of this pathway is disturbed in all human cancers (121, 122). Xu et al. demonstrated that the regions at the 5’ termini of Rb gene sequence are extremely rich in G and C residues (123, 124). The result of their study shows that G-strand at the 5’ terminus of Rb gene is capable of forming G-quadruplex in vitro and encumbers the progress of DNA polymerase. This indicates a possible role for the quadruplex structures in Rb gene (Figure 3) (125).

**Conclusion**

The biological significance of G-quadruplexes has been recognized by numerous research efforts. It appears to be the next great hope in a long line of promising targets. With the advancement of x-ray crystallography, CD, and NMR spectroscopy studies, the structure and topology of G-quadruplexes have become clearer than before. As there are many G-rich regions along chromosomes and some of these are associated with the promoter of oncogenes. This provides opportunities to target telomeric and oncogene promoter sequences of G-quadruplexes using quadruplex binding ligands.

**Acknowledgements**

We are thankful to Board of College and University Development, University of Pune, for grant BCUD/OSD/184. We also thank Mr. Vishwasrao Bhadane for proofreading and editing the final version of the text.

**Conflict of Interest**

We declare no conflicts of interest.

**REFERENCES:**


44 Chan SW, Blackburn EH. Telomerase and ATM/Tel1p protect telomeres from nonhomologous end joining. Mol Cell. 2003;11(5):1379-87.


sis.


