Published in final edited form as:

Ann N Y Acad Sci. 2012 September ; 1267: 53–60. doi:10.1111/j.1749-6632.2012.06586.x.

G4 motifs in human genes

Nancy Maizels

Departments of Immunology and Biochemistry, University of Washington, Seattle, Washington

Abstract

The G4 motif, $G_3N_xG_3N_xG_3N_xG_3$, is enriched in some genomic regions and depleted in others. This motif confers the ability to form an unusual four-stranded DNA structure, G4 DNA. G4 DNA is associated with genomic instability, which may explain depletion of G4 motifs from some genes and genomic regions. Conversely, G4 motifs are enriched downstream of transcription start sites, where they correlate with pausing. The uneven distribution of G4 motifs in the genome strongly suggests that mechanisms of selection act not only on one-dimensional genomic sequence, but also on structures formed by genomic DNA. The biological roles of G4 structures illustrate that, to understand genome function, it is important to consider the dynamic structural potential implicit in the G4 motif.

Keywords

DNA; G-quadruplex; repeat; replication; transcription

Introduction

Some sequence motifs recur in the genome, and others are depleted. What does this tell us about the genome, its evolution, and its functions? And what can we learn about the functions of specific motifs by studying their distribution and abundance in the genome? This paper focuses on the G4 motif, $G_3N_xG_3N_xG_3N_xG_3$, which confers an intrinsic potential to form structures other than canonical B form duplex DNA. The structure of DNA may vary from the canonical double helix, sometimes dramatically so, in a sequence-dependent manner. These diverse noncanonical structures can have significant effects on genomic stability and gene function. The G4 motif appears throughout the genome. Three chromosomal domains with specific functions are greatly enriched for G4 motifs (“G4^{hi}”): the telomeres (repeat TTAGGG); ribosomal DNA; and the immunoglobulin switch (S) regions, which are targets for the region-specific recombination essential to class switch recombination (see Kenter *et al.*¹).

Although functionally important, G4 motifs can put the genome at risk because, as described later, replication through these structures can lead to genomic instability. Such instability has also been observed at other motifs with strong structure-forming potential. For example, trinucleotide repeats are a well-known example of deceptively simple sequences with strong structural potential. Many trinucleotide repeats can form stem-loop structures, which can be

© 2012 New York Academy of Sciences.

Address for correspondence: Nancy Maizels, Departments of Immunology and Biochemistry, University of Washington, 1959 N.E. Pacific Street, H474A HSB, Seattle, WA 98195-7650. maizels@u.washington.edu.

Conflicts of interest

The author declares no conflicts of interest.

sites of sequence expansion or contraction. Expansions at unstable trinucleotide repeats contribute to at least 30 different neurodegenerative diseases.²

G4 motifs and G4 structures

The structures formed by regions of DNA bearing the G4 motif are referred to as G4 (or G-quadruplex) DNA. G4 structures can form in DNA or in RNA. G4 structures readily form *in vitro*; there is mounting evidence for the importance of G4 structures in biology.

The essential unit of a G4 structure is the G-quartet (Fig. 1A), in which four guanines interact by pairwise hydrogen bonds between the N7 of one guanine and the extracyclic amine of another.^{3,4} Hydrogen bonds between guanines and stacking of the G-quartets upon one another confer very high thermodynamic stability upon G4 structures.

G4 structures are polymorphic

Nucleotides that do not participate in G-quartet formation (denoted as N_x in the consensus motif) form loops (Fig. 1B). Loop conformation is determined by local nucleotide sequence. Loops may extend over, under, or around the central stacked quartets to create a remarkable variety of structures.⁵

The icon typically used to represent G4 DNA does not give full credit to the potential structural diversity of even short sequences. Figure 1B shows a simple example of two of the numerous polymorphic structures that can be formed by the sequence G₃N₃G₃N₂G₄N₂G₅. The situation is undoubtedly more complex in the genome, where G4 motifs may contain more than four G-runs, and G-runs may be heterogeneous in length.

The potential for structural polymorphism makes it difficult to predict whether an individual G within a G4 motif will participate in G-quartet formation. Moreover, in order to form a G4 structure, the DNA duplex must undergo denaturation, as occurs in the course of transcription, replication, and recombination. These considerations suggest that G4 DNA structures are likely to differ from time to time within a cell and from cell to cell in a population. Structural polymorphism almost certainly explains the difficulty of identifying G-quartets by *in vivo* DNA footprinting, as this technique pools information from populations of molecules. Structural polymorphism also means that standard sequence alignments overlook evolutionarily conserved G4 motifs embedded in a less conserved sequence unless algorithms have been designed to search for the G4 motif explicitly.

More than 300,000 G4 motifs can be identified in the human genome using a search algorithm that scores four or more adjacent runs of three or more guanines, in which loop size is constrained at 7 nt (G₃N₁₋₇G₃N₁₋₇G₃N₁₋₇G₃).⁶ The relatively short length searched by this algorithm (on the order of 30 nt) was chosen somewhat arbitrarily, and the distance over which G4 structures form *in vivo* may be longer.⁷ Not surprisingly, the number of G4 motifs that can be identified in the genome increases with increasing loop length.

G4 structures form in single-stranded DNA

G4 structures form only when DNA is single-stranded. Transcription of regions bearing G4 motifs results in formation of characteristic structures, G-loops (Fig. 2), which contain a stable cotranscriptional RNA/DNA hybrid on the template strand and G4 DNA interspersed with single-stranded regions on the nontemplate strand.⁸⁻¹⁰ G-loop and G4 DNA formation occur in only one transcriptional orientation, with a C-rich template strand. Transient denaturation that accompanies replication similarly enables formation of G4 DNA (Fig. 2).

G4 helicases unwind G4 DNA

G4 DNA that forms during replication is unwound by G4 helicases associated with the replication apparatus. These include the human helicases Bloom syndrome (BLM),¹¹ Werner syndrome (WRN),¹² and Fanconi anemia (FANCF).^{13,14} If G4 DNA is not resolved, it can lead to genomic instability.

The importance of replicative G4 DNA helicases has been dramatically demonstrated by experiments with model organisms. Nematodes lacking the FANCF homolog *dog-1* exhibit genomic instability mapping to regions bearing G4 motifs.¹⁵ *Saccharomyces cerevisiae* lacking the distinct G4 helicase Pif-1 exhibit instability at a G4^{hi} reporter sequence, which is exacerbated by small molecule ligands selective for G4 structures.^{16,17}

Structure-function analysis has shown that the RecQ family helicases, such as BLM and WRN, bind to G4 DNA using a highly conserved domain found only in enzymes of this family.¹⁸ This small RQC (“RecQ conserved”) domain can bind G4 DNA independently of other domains of the enzyme, and with affinity in the nanomolar range. Comparable high affinity characterizes other protein/G4 DNA interactions. There has as yet been no structural analysis of protein bound to G4 DNA or to define the features of the G4 structure that are targets for G4 DNA-binding proteins.

G4 DNA causes instability in the human genome

Deficiencies in G4 helicases cause human diseases associated with genomic instability and cancer predisposition: BLM, WRN, and FANCF.^{19–23} Deficiency in a G4 helicase may have particular impact on a G4^{hi} chromosomal domain. BLM is characterized by immunodeficiency due to faulty class-switch recombination. Cells from individuals with WRN exhibit telomere shortening, which contributes to the premature aging characteristic of this disease.²⁴

The potential for instability conferred by the G4 motif almost certainly accounts for the fact that some of the most polymorphic variable number tandem repeats in the human genome bear a G4 signature motif, including D4S43, GGGGAGGGGAAGA; the insulin-linked hyper-variable repeat, ACAGGGGTGTGGG; MS1, AG-GGTGGAG; MS32, CAGAATGGAGCAGGTGGC-CAGGGGTGACT; and CEB1, GGGGGGAGGG-AGGGTGGCCTGCGGAGGTCCCTGGGCTGA.

G4 motifs are selected in the human genome

The risk of genomic instability at G4 motifs suggests that selection might purge these motifs from functionally significant sequences, such as genes and their regulatory regions. However, this is not the case. We have shown that G4 motifs exhibit a very characteristic and highly uneven distribution among genes and gene regulatory regions.^{25–28} This distribution argues for evolutionary selection of G4 motifs and for their likely importance in genomic function.

To score G4 motifs in human genes, we analyzed regions of sequence 100 nt in length in sliding windows of 100 nt, scoring each region that contained a G4 motif as a “hit,” then quantifying the percentage of hits in the total numbers of windows searched.²⁵ We evaluated the distribution of gene ontology terms for each human gene across the spectrum of G4 DNA potentials. Most genes proved to be relatively depleted for G4 motifs (“G4^{lo}”), but some were considerably enriched (“G4^{hi}”). These differences in abundance of G4 motifs do not reflect coding capacity, but apply over entire genes, including both exons and introns.

G4 motifs correlate with gene function

Strikingly, we found that depletion or enrichment of G4 motifs correlates with gene function.²⁵ G4^{lo} genes function in nucleic acid binding, nucleosome assembly, ubiquitin-dependent proteolysis, cell adhesion, and regulation of cell division. In contrast, G4^{hi} genes function as transcription factors, growth factors, and cytokines, or in development, cell signaling, and muscle contraction. Tumor suppressor genes and oncogenes are at opposite ends of the spectrum: tumor suppressor genes are G4^{lo}, and oncogenes are G4^{hi}. The G4^{lo} status of tumor suppressor genes could be explained by selection for maximal genomic stability. But there is no obvious ready explanation for the G4^{hi} status of oncogenes.

Abundance of G4 motifs is independent of genomic environment

Our genomes consist of large segments of fairly homogeneous guanine + cytosine content, defined as isochores²⁹ (also see Ref. 30). Although the concept of isochores was put forward long before the sequence of the genome was known, it has proven to be robust.³¹ We asked if abundance or depletion of G4 motifs might reflect gene position within larger isochores by computing the relative density of G4 motifs in a gene and the 20 kb of upstream and downstream flanking sequence.²⁵ Those calculations showed that, on average, tumor suppressor genes are reduced and oncogenes are enriched in G4 motifs, relative to their flanking sequences. Thus, at least for these classes of genes, distribution of G4 motifs within a gene correlates with gene function, but not with local genomic environment.

G4 structures can be targets of regulation

We have recently begun to appreciate that G4 motifs function in diverse and unanticipated contexts. Several examples in which G4 structures have been unambiguously correlated with gene regulation illustrate this functionality. A G4 DNA target controls antigen switching in *Neisseria gonorrhoeae*.³² G4 DNA structures have been implicated in epigenetic regulation of human genes by ATRX, a chromatin remodelling factor deficient in an X-linked human genetic disease characterized by alpha thalassemia accompanied by mental retardation.³⁶ G4 RNA structures formed within transcripts of Fragile X genes that carry expansions of the trinucleotide repeat sequence CGG can titrate RNA processing factors to contribute to Fragile X disease pathology.³⁴ In addition, G4 structures have been shown to communicate changes in the cellular environment. G4 RNA structures formed within a 100 nt region of the p53 gene transcript determine 3'-processing of this mRNA upon induction of DNA damage.⁷ This raises the possibility that a subset of G4 motifs may respond to and communicate stress genome-wide.

G4 motifs in regulatory regions of human genes

Figure 3 diagrams an idealized gene, with a regulatory region upstream of the transcription start site (TSS), and downstream exons and introns. We and others have shown that G4 motifs are enriched in the regions flanking the TSS.^{26,35,36}

Strand-biased distribution of G4 motifs

The regions upstream and downstream of the TSS have very distinct potential to form nonduplex structures in the course of transcription. The upstream region may experience positive supercoiling that could be conducive to local formation of nonduplex structures. In contrast, the downstream region undergoes transient denaturation to enable passage of RNA polymerase. This prompted us to ask if there was evident strand bias of enrichment near the TSS. We found that upstream of the TSS, G4 motifs are comparably enriched on the template and nontemplate DNA strands.²⁶ Downstream of the TSS, G4 motifs are enriched on the nontemplate DNA strand.

To avoid overstating the case for G4 DNA in biology, it was important to ask if other mechanisms of regulation might explain G-richness flanking the TSS. We did this by masking sequence motifs for well-defined regulatory processes, such as CpG dinucleotides that are sites of methylation, and G-rich recognition motifs for common transcription factors, such as SP1 or KLF (Krüppel-like factor), and then recalculating distribution of G4 motifs.²⁶ Masking eliminated most of the G4^{hi} character of the region upstream of the TSS, suggesting that the G4^{hi} character of the region upstream of the TSS might reflect these well-defined processes.

The notion that regulatory factors might recognize G4 structures to control transcription had garnered considerable interest, particularly because many oncogene promoters contain at least one G4 motif upstream of the TSS. However, the modest residual G4 character remaining after masking suggested that duplex DNA was the predominant binding target. Nonetheless, there may be at least one exception to this rule. SP1 was recently shown to bind to G4 DNA with high affinity,³⁷ raising the possibility that SP1 might regulate gene expression by binding not only duplex DNA but also G4 DNA structures.

The strand asymmetry of G4 motifs downstream of the TSS suggests that, in a fraction of genes, transcription of regions bearing G4 motifs may promote formation of nucleic acid structures that could serve as targets for regulation. Three kinds of unusual structures are associated with G-loop formation (Fig. 2): G4 DNA structures in the nontemplate strand of a G-loop, which may be recognized by regulatory factors; G4 structures in the pre-mRNA or mRNA, which may bind specific factors to determine mRNA processing or functions; and cotranscriptional RNA/DNA hybrids, which place special demands on RNA processing,^{38,39} and may thus be subject to unusual regulation.

Conserved elements carrying G4 motifs in the first intron

The first intron represents a potentially privileged position within the genomic sequence, because it is very close to the TSS but does not appear in the mature mRNA or protein coding sequence. Proximity to the TSS would enable an element to load factors critical for regulation of gene expression, and enable their access to the core transcription complex. We found that nearly half of human genes carry one or more G4 motifs at the very 5' end of the first intron, on the nontemplate DNA strand (Fig. 3, middle).²⁶ This enrichment displays clear strand bias, and is evident predominately on the nontemplate strand.

It is significant that G-rich intron 1 elements are conserved from humans to frogs in their potential to form G4 structures, even though the sequences themselves are not conserved. In many cases, these elements are quite long, comprising eight or more G-runs. Both these observations suggest that factors recognize structures formed in these regions. How and why remains to be discovered. This emphasizes, again, the importance of algorithms that search explicitly for G4 structures.

G4 motifs correlate with promoter-proximal pausing

RNA Pol II pauses just downstream of the promoter in many human genes. We showed that promoter-proximal pausing correlates with G-richness downstream of the promoter, and with G-rich intron 1 elements.²⁸ This raises the possibility that structures formed either in the DNA (by a pioneering round of transcription) or mRNA transcript may regulate pausing. This could occur either directly or via protein-mediated regulatory interactions. Pausing—and especially the release from pausing—may enable a rapid response to environmental or developmental signals, and thus provide another connection between genomic structures and the cellular response to stress.

Coding regions and UTRs are not enriched for G4 motifs

The enrichment of G4 motifs around the TSS and at the 5' end of the first intron contrasts with the depletion and apparent selection against these motifs within coding regions. We found that 85% of human exons contain no G4 motifs.²⁷ A graph of distribution of G4 motifs within 1 kb downstream of ATG initiator codons and within 1 kb upstream of termination codons shows that relatively few genes carry these motifs, and where they do occur they predominate on the template rather than nontemplate DNA strand (Fig. 3, bottom). There are also few genes with G4 motifs in the 5' - and 3' -UTRs (note that there appears to be some enrichment for G4 motifs at the very 5' -end of the 5' -UTR and at the 5' -end of coding regions, but this can probably be ascribed to the contribution from the region just downstream of the TSS). The paucity of G4 motifs within genes provides further evidence for counterselection of these motifs within specific genomic regions.

For the future: what explains the distribution of G4 motifs in the genomic landscape?

The distinctive patterns of enrichment and depletion of G4 motifs support the view that these motifs have been selected during evolution and play key roles in genome biology. The genomic instability associated with G4 motifs can explain their depletion. Regulation by G4 structures may explain their enrichment, as suggested by the very intriguing examples that have come to light over the past few years. As we learn more about genomic regulation, more examples of function of these structures in regulation will undoubtedly be discovered. For example, systems analyses have the potential to correlate G4 motifs with gene expression under a variety of conditions, and thereby establish if and how G4 structures could exert global regulation in response to changing environmental conditions or cell stimuli. In addition, as genomewide analyses begin to systematically tally the spectrum of factors enriched at G4 motifs, we will learn whether G4 structures form rarely or frequently, and whether they are occasional or common regulatory targets.

The selection of G4 motifs in the genome highlights the importance of designing genomic search algorithms that look beyond primary DNA sequence to consider the dynamic structural potential of DNA.

References

1. Kenter AL, Feldman S, Wuerffel R, Achour I, Wang L, Kumar S. Three dimensional architecture of the Igh locus facilitates class switch recombination. *Ann NY Acad Sci.* 2012; 1267:86–94. This volume. [PubMed: 22954221]
2. Mirkin SM. Expandable DNA repeats and human disease. *Nature.* 2007; 447:932–940. [PubMed: 17581576]
3. Gellert M, Lipsett MN, Davies DR. Helix formation by guanylic acid. *Proc Natl Acad Sci USA.* 1962; 48:2014–2018.
4. Sen D, Gilbert W. Formation of parallel four-stranded complexes by guanine rich motifs in DNA and its implications for meiosis. *Nature.* 1988; 334:364–366. [PubMed: 3393228]
5. Phan AT, Kuryavyy V, Patel DJ. DNA architecture: from G to Z. *Curr Opin Struct Biol.* 2006; 16:288–298. [PubMed: 16714104]
6. Huppert JL, Balasubramanian S. Prevalence of quadruplexes in the human genome. *Nucleic Acids Res.* 2005; 33:2908–2916. [PubMed: 15914667]
7. Decorsiere A, Cayrel A, Vagner S, Millevoi S. Essential role for the interaction between hnRNP H/F and a G quadruplex in maintaining p53 pre-mRNA 3' -end processing and function during DNA damage. *Genes Dev.* 2011; 25:220–225. [PubMed: 21289067]
8. Duquette ML, Handa P, Vincent JA, et al. Intra-cellular transcription of G-rich DNAs induces formation of G-loops, novel structures containing G4 DNA. *Genes Dev.* 2004; 18:1618–1629.

9. Duquette ML, Huber MD, Maizels N. G-rich proto-oncogenes are targeted for genomic instability in B-cell lymphomas. *Cancer Res.* 2007; 67:2586–2594. [PubMed: 17363577]
10. Duquette ML, Pham P, Goodman MF, Maizels N. AID binds to transcription-induced structures in c-MYC that map to regions associated with translocation and hypermutation. *Oncogene.* 2005; 24:5791–5798. [PubMed: 15940261]
11. Sun H, Karow JK, Hickson ID, Maizels N. The Bloom's syndrome helicase unwinds G4 DNA. *J Biol Chem.* 1998; 273:27587–27592. [PubMed: 9765292]
12. Fry M, Loeb LA. Human Werner syndrome DNA helicase unwinds tetrahelical structures of the fragile X syndrome repeat sequence d(CGG)_n. *J Biol Chem.* 1999; 274:12797–12802. [PubMed: 10212265]
13. London TB, et al. FANCI is a structure-specific DNA helicase associated with the maintenance of genomic G/C tracts. *J Biol Chem.* 2008; 283:36132–36139. [PubMed: 18978354]
14. Wu Y, Shinya K, Brosh RM Jr. FANCI helicase defective in Fanconi anemia and breast cancer unwinds G-quadruplex DNA to defend genomic stability. *Mol Cell Biol.* 2008; 28:4116–4128. [PubMed: 18426915]
15. Kruisselbrink E, et al. Mutagenic capacity of endogenous G4 DNA underlies genome instability in FANCI-defective *C. elegans*. *Curr Biol.* 2008; 18:900–905. [PubMed: 18538569]
16. Ribeyre C, et al. The yeast Pif1 helicase prevents genomic instability caused by G-quadruplex-forming CEB1 sequences *in vivo*. *PLoS Genet.* 2009; 5:e1000475. [PubMed: 19424434]
17. Piazza A, et al. Genetic instability triggered by G-quadruplex interacting Phen-DC compounds in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* 2010; 38:4337–4348. [PubMed: 20223771]
18. Huber MD, Duquette ML, Shiels JC, Maizels N. A conserved G4 DNA binding domain in RecQ family helicases. *J Mol Biol.* 2006; 358:1071–1080. [PubMed: 16530788]
19. Sissi C, Gatto B, Palumbo M. The evolving world of protein-G-quadruplex recognition: a medicinal chemist's perspective. *Biochimie.* 2010; 93:1219–1230. [PubMed: 21549174]
20. Monnat RJ Jr. Human RECQ helicases: roles in DNA metabolism, mutagenesis and cancer biology. *Semin Cancer Biol.* 2010; 20:329–339. [PubMed: 20934517]
21. Rossi ML, Ghosh AK, Bohr VA. Roles of Werner syndrome protein in protection of genome integrity. *DNA Repair (Amst).* 2010; 9:331–344. [PubMed: 20075015]
22. Deans AJ, West SC. DNA interstrand crosslink repair and cancer. *Nat Rev Cancer.* 2011; 11:467–480. [PubMed: 21701511]
23. Deakynne JS, Mazin AV. Fanconi anemia: at the crossroads of DNA repair. *Biochemistry (Mosc).* 2011; 76:36–48. [PubMed: 21568838]
24. Crabbe L, Jauch A, Naeger CM, et al. Telomere dysfunction as a cause of genomic instability in Werner syndrome. *Proc Natl Acad Sci USA.* 2007; 104:2205–2210. [PubMed: 17284601]
25. Eddy J, Maizels N. Gene function correlates with potential for G4 DNA formation in the human genome. *Nucleic Acids Res.* 2006; 34:3887–3896. [PubMed: 16914419]
26. Eddy J, Maizels N. Conserved elements with potential to form polymorphic G-quadruplex structures in the first intron of human genes. *Nucleic Acids Res.* 2008; 36:1321–1333. [PubMed: 18187510]
27. Eddy J, Maizels N. Selection for the G4 DNA motif at the 5' end of human genes. *Mol Carcinog.* 2009; 48:319–325. [PubMed: 19306310]
28. Eddy J, et al. G4 motifs correlate with promoter-proximal transcriptional pausing in human genes. *Nucleic Acids Res.* 2011; 39:4975–4983. [PubMed: 21371997]
29. Bernardi G, et al. The mosaic genome of warm-blooded vertebrates. *Science.* 1985; 228:953–958. [PubMed: 4001930]
30. Bernardi G. The genome: an isochore ensemble and its evolution. *Ann NY Acad Sci.* 2012; 1267:31–34. This volume. [PubMed: 22954213]
31. Costantini M, Clay O, Auletta F, Bernardi G. An isochore map of human chromosomes. *Genome Res.* 2006; 16:536–541. [PubMed: 16597586]
32. Cahoon LA, Seifert HS. An alternative DNA structure is necessary for pilin antigenic variation in *Neisseria gonorrhoeae*. *Science.* 2009; 325:764–767. [PubMed: 19661435]

33. Law MJ, et al. ATR-X syndrome protein targets tandem repeats and influences allele-specific expression in a size-dependent manner. *Cell*. 2010; 143:367–378. [PubMed: 21029860]
34. Santoro MR, Bray SM, Warren ST. Molecular mechanisms of fragile X syndrome: a twenty-year perspective. *Annu Rev Pathol*. 2011; 7:219–245. [PubMed: 22017584]
35. Huppert JL, Balasubramanian S. G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Res*. 2007; 35:406–413. [PubMed: 17169996]
36. Zhao Y, Du Z, Li N. Extensive selection for the enrichment of G4 DNA motifs in transcriptional regulatory regions of warm blooded animals. *FEBS Lett*. 2007; 581:1951–1956. [PubMed: 17462634]
37. Raiber EA, Kranaster R, Lam E, et al. A non-canonical DNA structure is a binding motif for the transcription factor SP1 *in vitro*. *Nucleic Acids Res*. 2011; 40:1499–1508. [PubMed: 22021377]
38. Aguilera A. mRNA processing and genomic instability. *Nat Struct Mol Biol*. 2005; 12:737–738. [PubMed: 16142225]
39. Gan W, et al. R-loop-mediated genomic instability is caused by impairment of replication fork progression. *Genes Dev*. 2011; 25:2041–2056. [PubMed: 21979917]

Figure 1. G-quartets and G4 DNA. (A) G-quartet formed by interactions among four guanine bases in DNA or RNA. (B) Intramolecular G4 DNA. The diagram illustrates two of the possible structures formed by the sequence $G_3N_3G_3N_2G_4N_2G_5$.

Figure 2.

Formation of G4 DNA during transcription or replication. Right, transcription of two adjacent G4 motifs results in formation of a G-loop, containing a stable RNA/DNA hybrid (gray) and G4 DNA on the nontemplate strand. Left, G4 DNA forms on the template for lagging-strand replication.

Figure 3.

Distribution of G4 motifs in human genes. Percentage of genes with one or more G4 motifs per 100 bp interval was calculated for the nontemplate and template strands of the indicated regions flanking the transcription start site (TSS, top); at the 5' ends of the first or second introns (middle); and at the 5' -ends of the 5' -UTR and coding region, or the 3' -ends of the coding region and 3' -UTR. Exons, shaded; 3' -UTR, hatched. Nontemplate and template strands are represented by solid and dashed lines, respectively. UTR, untranslated region.