Mutation Research 743–744 (2013) 118–131

Contents lists available at SciVerse ScienceDirect

Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis

journal homepage: www.elsevier.com/locate/molmut
Community address: www.elsevier.com/locate/mutres

Review

The yin and yang of repair mechanisms in DNA structure-induced genetic instability

Karen M. Vasquez*, Guliang Wang

Division of Pharmacology and Toxicology, College of Pharmacy, The University of Texas at Austin, Dell Pediatric Research Institute, 1400 Barbara Jordan Blvd. R1800, Austin, TX 78723, United States

ARTICLE INFO

Article history:
Received 28 August 2012
Received in revised form
21 November 2012
Accepted 24 November 2012
Available online 3 December 2012

Keywords:
DNA structure
Triplex DNA
DNA repair
Z-DNA
Genetic instability

ABSTRACT

DNA can adopt a variety of secondary structures that deviate from the canonical Watson–Crick B-DNA form. More than 10 types of non-canonical or non-B DNA secondary structures have been characterized, and the sequences that have the capacity to adopt such structures are very abundant in the human genome. Non-B DNA structures have been implicated in many important biological processes and can serve as sources of genetic instability, implicating them in disease and evolution. Non-B DNA conformations interact with a wide variety of proteins involved in replication, transcription, DNA repair, and chromatin architectural regulation. In this review, we will focus on the interactions of DNA repair proteins with non-B DNA and their roles in genetic instability, as the proteins and DNA involved in such interactions may represent plausible targets for selective therapeutic intervention.

Contents

Repetitive sequences that do not code for proteins account for more than half of the total DNA in the human genome, which includes transposon-derived repeats, (*e.g.*, Alu and LINE elements), pseudo-genes or duplications, and tandem simple repeats [1]. In contrast to its former nickname, "junk DNA", repetitive sequences are now known to play important roles in regulating

* Corresponding author. Tel.: +1 512 495 3040; fax: +1 512 496 4946.
E-mail address: karen.vasquez@austin.utexas.edu (K.M. Vasquez).

0027-5107/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.mrfmmm.2012.11.005

the genome, including putative involvement in shaping chromatin structure, regulating gene expression, and stimulating genomic rearrangement. Of particular interest, some simple repeats, which represent approximately 3% of the total human genome, have been found to form various types of non-canonical structures that differ from the Watson–Crick B-form (*i.e.*, non-B DNA structures). Extensively studied microsatellite repeats, such as triplet repeats can contain tens, hundreds, or even thousands of trinucleotide repeat units [2,3]. Expanded triplet repeats (such as CTG and CGG) have been implicated in more than 20 genetic diseases, including Fragile X syndrome, Huntington's disease, Friedreich's ataxia, and myotonic dystrophy [4,5]. The role of microsatellites in disease is not limited to triplet repeats, as evidenced by Myotonic dystrophy type 2 (DM2), which is caused by the extreme expansion of CCTG tetranucleotide repeats from <30 repeats in normal individuals to up to 11,000 in some patients [6], and spinocerebellar ataxia type 10 patients, in which the number of pentanucleotide ATTCT repeats can be expanded from approximately 14 in normal individuals to 4500 [7].

In cells, the majority of genomic DNA exits in the B-conformation at any given time (Fig. 1A), but the conformational state of DNA is dynamic and is generally negatively supercoiled corresponding to a high-energy state. The DNA is subjected to manipulation *via* replication, transcription, and DNA repair proteins. These processes unwind the DNA from the histones, generating more negative supercoiling stress and can open the double helix, giving rise to non-B structures at sequences (*e.g.*, repetitive sequences) conducive to their formation. To date, more than 10 different types of non-B DNA conformations have been identified and characterized [8]. Examples include, slipped structures formed at direct repeats, hairpin or cruciform structures formed at inverted repeats, intra-molecular triplex DNA structures (H-DNA) formed at polypurine–polypyrimidine elements with mirror repeat symmetry, four-stranded G-quartet structures formed at sequences comprising 4 guanine tracks of at least 3 continuous guanosine residues separated by 1–7 nucleotides, and left-handed Z-DNA structures formed at alternating pyrimidine–purine sequences (see Fig. 1; for review, see [9]). Notably, under certain conditions, imperfect repeat sequences can also form non-B conformations. For example, short CG repeats (2–4 repeats) that are separated by 3 bp interruptions can adopt Z-DNA structures and form Z-Z junctions on plasmids *in vitro*, as assessed by two-dimensional gel electrophoresis and chemical and enzymatic probing [10]. Hairpin structure formation typically requires an inverted repeat symmetry, but can also form at CNG triplet repeats, which contain a mismatch after every two C-G base-pairs [11,12]. Thus, sequences with the potential to form non-B DNA structures (or non-B DNA-forming sequences) are very abundant in the human genome.

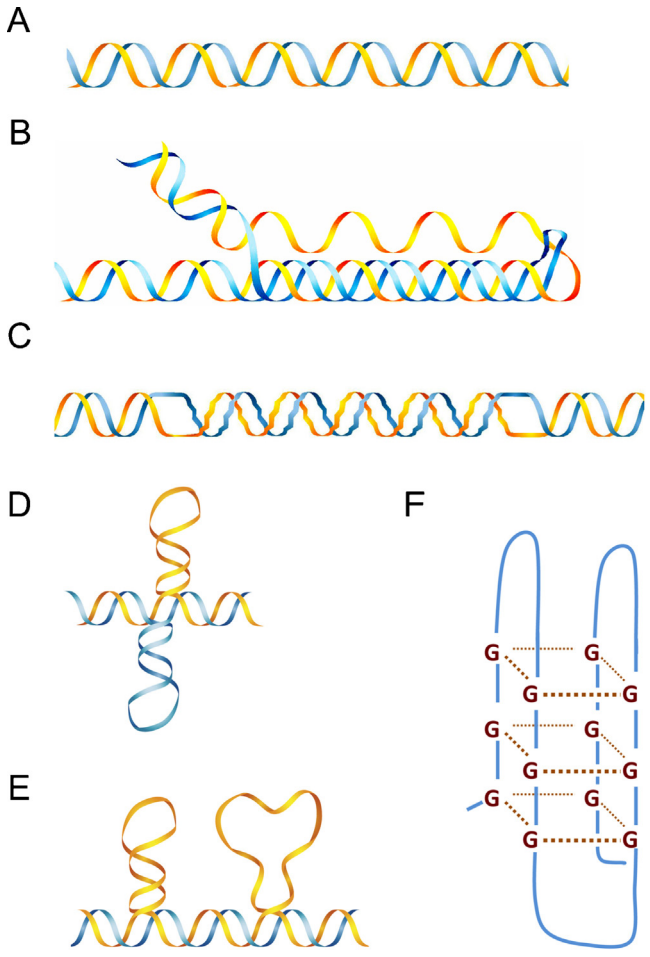

Fig. 1. Schematic diagram of DNA conformations. (A) Canonical B-DNA; (B) intramolecular triplex, H-DNA; (C) left-handed Z-DNA; (D) cruciform; (E) hairpin (left) and slippage loop (right); (F) G-quadruplex.

1. Non-B DNA structure and genetic instability

Neurological disorders related to triplet repeat expansion have been studied extensively (for reviews see [13,14]); however, many types of non-B DNA conformations in addition to loop structures due to slippage and misalignment (*e.g.*, slipped DNA) of repetitive units or hairpins/cruciform formed at triplet repeat sequences have been implicated in genetic instability associated with a variety of human disorders [15–18], as we briefly outline below.

1.1. Intra-molecular triplex DNA (H-DNA) and genetic instability

H-DNA can form at polypurine–polypyrimidine sequences with mirror repeat symmetry, where half of the symmetry, when single-stranded, binds to the purine-rich strand of the duplex containing the other half of the symmetry *via* Hoogsteen hydrogen bonding through the major groove, forming a three-stranded DNA structure (Fig. 1B) [19]. Genes containing long polypurine–polypyrimidine sequences show higher frequencies of alternative splicing and chromosomal translocations [20], and H-DNA-forming sequences frequently co-localize with chromosomal breakpoint hotspots and genomic rearrangements in disease-related human genes. For example, several polypurine–polypyrimidine tracks can be found near (within hundreds of bps) the major breakage hotspots in the *c-MYC* gene found in leukemias and lymphomas [21–28], and in the *BCL-2* gene major breakpoint region (Mbr) in follicular lymphomas. *In vitro* experiments demonstrated that these sequences were capable of forming non-B DNA structures, including H-DNA [29]. Furthermore, disruption of the H-DNA conformation by disruption of the mirror symmetry reduced the frequency of translocation events in the *BCL-2* Mbr [30]. The 21st intron of the human *PKD1* gene, which contains a 2.5 kb polypyrimidine tract with 23 mirror repeats [31] that can adopt H-DNA structures *in vitro* [32], has been implicated in the high mutation rate in this region in both germ line and somatic cells from autosomal dominant polycystic kidney disease patients (ADPKD) [33]. This 2.5 kb polypyrimidine sequence in a plasmid induced DNA double-strand breaks (DSBs) and resulted in large-scale deletions in *Escherichia coli* [34]. In our previous work, we discovered that both an endogenous H-DNA-forming sequence from the human *c-MYC* promoter and model H-DNA-forming sequences, when cloned into a mutation reporter shuttle plasmid, induced DSBs near the H-DNA locus and induced