Published in final edited form as:

DNA Repair (Amst). 2014 July ; 19: 143–151. doi:10.1016/j.dnarep.2014.03.017.

Impact of Alternative DNA Structures on DNA Damage, DNA Repair, and Genetic Instability

Guliang Wang and Karen M. Vasquez*

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.

Abstract

Repetitive genomic sequences can adopt a number of alternative DNA structures that differ from the canonical B-form duplex (*i.e.* non-B DNA). These non-B DNA-forming sequences have been shown to have many important biological functions related to DNA metabolic processes; for example, they may have regulatory roles in DNA transcription and replication. In addition to these regulatory functions, non-B DNA can stimulate genetic instability in the presence or absence of DNA damage, *via* replication-dependent and/or replication-independent pathways. This review focuses on the interactions of non-B DNA conformations with DNA repair proteins and how these interactions impact genetic instability.

Keywords

DNA structure; DNA replication; DNA repair; Transcription; Genetic instability

More than half of the human genome comprises repetitive sequence elements such as transposon-derived repeats (*e.g.* Alu and LINE elements), pseudo-genes or duplications, and tandem simple repeats [1]. While many of these repetitive elements do not code for proteins, they play important roles in regulating chromatin structure and function. For example, many repetitive sequences have the capacity to adopt alternative DNA conformations that differ from the canonical B-DNA structure described by Watson and Crick more than 50 years ago, and are thus referred to as non-B DNA structures. Under appropriate physiological conditions, more than 10 types of non-B DNA conformations have been described [2–4]. Simple repeats can form slipped structures and/or looped regions when the two repetitive strands separate and misalign [5]. If a single-stranded looped-out region contains inverted repeats and can self-anneal to form intra-strand Watson and Crick base pairs, then a hairpin structure (or cruciform structure if both strands form hairpin structures at the same position) can form [6]. If a single-stranded region contains polypurines, with mirror-repeat symmetry,

© 2014 Elsevier B.V. All rights reserved.

*Address correspondence to: Karen M. Vasquez, PhD, 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. Tel.: +1 512 495 3040. Fax: +1 512 496 4946. karen.vasquez@austin.utexas.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

then it can pair with the purine-rich strand of the duplex *via* Hoogsteen hydrogen bonding to form a three-stranded helix, leaving the complementary strand unpaired [7, 8]. This particular type of non-B DNA is referred to as H-DNA or intramolecular triplex DNA. Purine bases in alternating purine/pyrimidine sequences, such as GT or GC repeats, can adopt a *syn* conformation while the pyrimidine nucleosides remain in an anti confirmation. Such a transition can bend the phosphate backbone into a zig-zag shape (referred to as Z-DNA) and alter the winding direction of each strand from right-handed to left-handed [9, 10]. In specific sequence contexts, four guanine bases can align *via* Hoogsteen hydrogen bonding to form a square planar structure called a guanine tetrad [11]. Further, regions containing four runs of three or more guanines have the potential to form stable G-quadruplexes where three or more guanine tetrads stack with each other. Other types of non-B DNA conformations include “sticky DNA”, an intramolecular structure adopted by two triplex-like structures and A-DNA, a DNA conformation that contains an increase in the number of base pairs per rotation, a deeper major groove, and a shallower minor groove than B-DNA (reviewed in ref. [2]). Some examples of non-B DNA structures are illustrated schematically in Figure 1.

Non-B DNA conformation *in vitro* and *in vivo*

Numerous techniques have been employed to characterize non-B DNA structures *in vitro* [12–15], and biological and genetic studies of non-B DNA structures, have revealed both physiological and pathological roles of non-B DNA *in vivo*. Accumulating evidence suggests that non-B DNA conformations play important biological roles in chromatin architecture, gene expression regulation, DNA replication, DNA damage and repair, and genetic instability, as summarized in the following reviews [4, 16–20]. As studies continue to elucidate the extent to which non-B DNA impacts biological functions in the cell, the following questions arise: “where” in the genome do non-B DNA-forming sequences reside?; “which” non-B conformations are formed at those sequences?; and “when” in the life-span of a cell do these sequences adopt non-B DNA structures and what are their functions within those sequence contexts?

As briefly described above, certain sequence motifs have the potential to adopt specific types of non-B DNA conformations. These sequence patterns allow for the identification of potential non-B DNA-forming sequences in the genome. With the advances in DNA genome sequencing, researchers have used this information to develop computer-based algorithms to develop search engines for potential non-B DNA structure-forming sequences [21, 22] (reviewed in [23]). Most search algorithms allow for user-defined settings to adjust the search parameters, and some can provide predicted free energy costs for such B- to non-B transitions. Thus, the user can evaluate the theoretical potential of a given conformation within sequences that may adopt more than one non-B DNA structure.

These computer-based search programs can be used to answer the questions of “where” and “which type”, but they do not provide information to address “when” does a non-B DNA conformation form *in vivo*. This question is confounded by many parameters, including but not limited to chromatin structure, epigenetic alterations, protein binding to the structure or structure-forming sequences, genomic context, DNA metabolism, and phase of the cell

cycle. Non-B DNA structures are high-energy states relative to Watson-Crick B-DNA, thus even an ideal non-B DNA-forming sequence is likely to remain in a B-DNA formation at most times under normal conditions. Further, genomic DNA is packaged into chromatin *in vivo*, and this interaction between DNA and histones also assists in maintaining DNA in a B-conformation. Thus, conditions that allow for unwinding of the DNA from the histones, separation of the DNA duplex (*e.g.* DNA replication, transcription, repair), and generation of negative supercoiling in the unwound DNA facilitate non-B DNA structure formation.

Non-B DNA contributes to genetic instability

Using algorithms as mentioned above to search genomic DNA for sequences with the capacity to adopt non-B DNA structures led to an important discovery; non-B DNA-forming sequences often co-localize with hotspots of DNA double-strand breaks (DSBs), deletions, rearrangements, and chromosomal translocations [24–26], implicating non-B DNA in genetic instability. For example, polypurine mirror-repeat H-DNA-forming sequences [27–34], mixed GT and GC Z-DNA-forming repeats [35, 36], and purine-rich tracts with the capacity to form intramolecular G-quadruplex structures [37, 38] were found within hundreds of bps adjacent to the major breakage hotspots within the P1 promoter of the *c-MYC* gene. H-DNA-forming sequences were also found in the major breakpoint region (Mbr) of the *BCL-2* gene, which is implicated in follicular lymphomas [39]. Changing the linear sequence in the Mbr region slightly (*i.e.* CCC to GGG), to prevent the formation of H-DNA, substantially reduced the frequency of translocation events in the *BCL-2* Mbr, suggesting a role for H-DNA in genetic instability [40]. Z-DNA-forming sequences were also found within hundreds of bps surrounding the translocation breakpoints in lymphoid tumors [41], and the DNA breakage hotspot cluster region from a subgroup of B-cell precursor acute lymphoblastic leukemia [42]. A long AT-rich inverted repeat on human chromosome 11q23 and 22q11 co-localizes with the breakpoints for t(11;22) translocations [43–45], not only in lymphoblasts and fibroblasts, but also in sperm from normal healthy males [46]. G-quadruplex-forming sequences have been found within 500 bp of mitotic and meiotic DNA breakpoints [47, 48], and are associated with mutation hotspots in human disease such as ataxias and Fragile X syndrome (FXS) [20].

Many other studies have provided solid evidence for non-B DNA in genetic instability both *in vitro* and *in vivo*, as summarized in the following reviews [16, 49–53] and within this issue. Of particular relevance, we found that when human sequences capable of forming H-DNA and Z-DNA were integrated into the genomes of transgenic mice, non-B DNA-induced genetic instability events in the genomes of ~7% of the F1 offspring (of ~100 mice analyzed), including large deletions, rearrangements and/or translocations, while no such events were detected in >60 mice carrying canonical B-DNA control transgenes [54]. Surprisingly, the percentage of non-B DNA-induced mutation on the transgenes in the mice was much higher than that observed from our studies using shuttle vector mutation-reporter plasmids in mammalian cells [54–56], and the frequency of *de novo* t(11;22) translocations mediated by AT-rich inverted repeats in human sperm [46]. A possible explanation for this result, aside from the differences in experimental systems used, is that the transgenes containing the non-B or B-DNA sequences were not functional or required by the cell and therefore were not under selective pressure. Because the system used allowed for detection

of a wide range of mutagenic events, our study may reflect an unexpectedly high level of non-B DNA instability in living cells. Thus, the mutagenic potential of endogenous non-B DNA-forming sequences may be under-estimated in many studies, since those cells suffering non-B DNA-induced mutagenic events in critical genes could be eliminated from the population prior to detection.

Replication-dependent mechanisms of non-B DNA-induced genetic instability

Many non-B DNA-forming sequences, such as triplet repeats capable of forming hairpin or looped structures, are more unstable in highly proliferative tissues [57, 58] or in rapidly dividing cells [59, 60], than in differentiated or slowly-replicating cells [61–63]. The frequencies and patterns of mutation caused by some of these non-B DNA-forming sequences are also dependent on the DNA replication status of the cell, and on the distance and orientation of the repetitive sequences relative to the replication origin [64–70]. These observations support at least two related models of replication-dependent/related genetic instability at non-B DNA-sequence containing regions: *i.e.* replication facilitates the formation of non-B DNA structures by separation of the duplex strands and/or by providing energy from negative supercoiling; and/or mutations occur during replication at non-B DNA regions, perhaps by structure-induced impediments to the replication machinery. Because the DNA replication machinery displaces nucleosomes and unwinds the genomic duplex DNA during replication, this generates long single-stranded regions on the lagging strand. These conditions can facilitate a B- to non-B transition. Moreover, many forms of non-B DNA can impact the processivity of DNA polymerases [71, 72] and cause replication fork stalling [66, 73–75], which further prolongs the existence of single strandedness on the lagging strand, thereby facilitating the formation of certain types of non-B DNA structures. On the other hand, components of the DNA replication machinery, *e.g.* DNA helicases, can unwind many types of non-B DNA structures in front of replication forks, and can assist in replication fork restart, in part by resolving non-B DNA structures [76]. However, it has been demonstrated that some energy stable non-B DNA conformations are resistant to helicase unwinding [73, 77]. Pre-existing non-B DNA structures can impact the fidelity of DNA polymerases, and result directly in the generation of mutations. For example, simple repeats such as hairpin-forming triplet repeats, can cause primer-template slippage during DNA replication, and can promote further slippage events [78], giving rise to repeat expansion or contraction, depending on strand orientation. If the stalling persists, DNA replication forks may collapse, resulting in DSBs [79–81]. 2-D electrophoresis of DNA replication intermediates, combined with Southern blotting, has demonstrated non-B DNA-induced replication fork stalling *in vivo* [82, 83], implicating DNA replication in non-B DNA-induced genetic instability.

Replication-independent mechanisms of non-B DNA-induced genetic instability

While the replication-dependent models of non-B DNA-induced genetic instability discussed above are well established mechanisms [84], non-B DNA structures can also

stimulate mutations in non-proliferative or slowly-proliferating tissues, *e.g.* brain [85–87] in Huntington disease and spinocerebellar ataxias, and muscle in patients with Kennedy's disease and Myotonic dystrophy [62, 88]. These findings suggest an alternative mechanism(s) of DNA structure-related instability that is independent of DNA replication. In support of such a model, we found that H-DNA, Z-DNA, and short cruciform-forming sequences were capable of stimulating the formation of DSBs and large-scale deletions on mutation-reporter plasmids in replication-deficient HeLa cell extracts [[56] and unpublished data]. Moreover, the spectra of mutations induced by ZDNA and H-DNA on plasmids in replication-deficient cells differed from those that occurred on the same plasmids in replication-proficient cells ([56] and our unpublished results).

Non-B DNA structure formation independent of replication

While replication can facilitate the formation of non-B DNA structures, genomic DNA remains in a dynamic state even in the absence of replication. Several biological processes result in negative supercoiling, opening of the chromatin structures and unwinding DNA, and thereby facilitate the formation of non-B DNA. Examples include DNA repair and DNA transcription, discussed below.

DNA damage and repair-induced non-B DNA conformation—DNA lesions can alter the conformation of the DNA helix, recruit or prevent the binding of DNA interacting proteins, and modify the distribution and modification of chromatin-related proteins, which can facilitate or impede the formation of non-B DNA. For example, an abasic site can destabilize duplex DNA in the canonical B-conformation, but has little effect when the abasic site is located in the loop of a hairpin structure [89]. One can speculate that an abasic site, or other bulky adduct that can destabilize B-DNA, when in the single-stranded region of an H-DNA or G-quadruplex structure may have a similar effect, although this has not yet been determined. A DNA interstrand crosslink within a cruciform/hairpin or loop structure may prevent a transition to B-DNA; while if a crosslink is formed within B-DNA, it may inhibit the formation of non-B conformations that require strand separation for their formation.

Repair of damage on genomic DNA will generate the conditions that favor non-B conformation, *e.g.* an open chromatin structure and the generation of negatively supercoiled single-stranded DNA. Moreover, the process of DNA repair *per se* can generate non-B DNA conformations at repetitive sequences. For example, it has been shown that the repair of a DSB near a short inverted repeat allowed the formation of a hairpin structure at the tip of the breakpoint. The end of the folded strand was then ligated to the complementary strand to form a large hairpin structure in *Saccharomyces cerevisiae* [90].

Transcription-induced non-B DNA formation—The DNA transcription machinery opens the chromatin structure and separates the coding and non-coding strands, and generates negative supercoiling downstream of the migrating complex, conditions that are conducive to the formation of non-B DNA structures. In fact, formation of a Z-DNA structure in the *c-MYC* promoter region (see above), as detected by Z-DNA-specific antibody binding in permeabilized mammalian cell nuclei, was dependent on transcription of

the *c-MYC* gene [36, 91, 92]. Formation of G-loops (G-quadruplex on one strand and an R-loop structure on the other strand) requires hybridization of an RNA strand to the complementary C-rich DNA strand, and is therefore dependent on transcription [93]. In collaboration with the Hanawalt group, we found that an H-DNA-forming sequence from the human *c-MYC* promoter (see Figure 2) resulted in stalling of T7 RNA polymerase within and downstream of the H-DNA-forming sequence [94]. We speculated that once T7 RNA polymerase passed the H-DNA-forming sequence, the generation of negative supercoiling allowed the formation of H-DNA at this sequence. The newly synthesized RNA could further bind to the template DNA strand to form a R-loop structure, or an RNA-RNA-DNA triplex structure on the otherwise single-stranded region within the H-DNA structure. This complex structure could cause architecture problems for RNA strand rotation and result in RNA polymerase stalling [94]. Thus, the process of transcription can facilitate the formation of non-B DNA structures, and also stabilize some types of non-B structures once they are formed.

Non-B DNA-induced mutation independent of replication

Roles of DNA repair protein in non-B DNA-induced genetic instability

DNA repair mechanisms have evolved to maintain genome integrity by removing DNA damage. To efficiently recognize damage in genomic DNA, many repair proteins sense DNA helical distortions (both chemical and physical) induced by the lesions [95–97]. Because the formation of non-B DNA can induce helical distortions, some of these structural alterations may share similarities with damaged DNA substrates, and thereby trigger repair activity in the absence of DNA damage per se. A variety of DNA repair-related proteins can recognize and bind specifically to non-B DNA structures, and cleave DNA at or near the non-B DNA conformation.

Nucleotide excision repair (NER) proteins are thought to recognize bulky DNA adducts *via* chemical and physical alterations in the Watson-Crick B-DNA helix [98]. A triplex structure formed by a triplex-forming oligonucleotide (TFO) binding to a polypurine/polypyrimidine duplex DNA *via* Hoogsteen hydrogen bonding [99], similar to the intramolecular triplex region of H-DNA, was shown to induce NER-dependent mutagenesis in mammalian cells [100]. The xeroderma pigmentosum group A (XPA), and group C (XPC) proteins were shown to stimulate repair synthesis at intermolecular triplex regions in HeLa cell extract [101], and the purified human recombinant XPA-RPA complex [102] and XPC-RAD23B complex [103] recognized and bound to intermolecular triplex structures *in vitro* [103]. We have also found that XPA deficiency reduced H-DNA-induced mutation frequencies in human XP12RO cells (Wang *et al*, unpublished results), suggesting that XPA plays a role in non-B DNA-induced genetic instability. In human cells, transcription-coupled NER proteins (XPA, CSB and ERCC1 and XPG) have been implicated in contraction of chromosomally integrated CAG repeats in a transcription-dependent fashion [104, 105].

Mismatch repair (MMR) proteins are responsible for the repair of base-pairing errors such as mismatches and small loops caused by replication-associated base miss-incorporation, replication slippage events, or certain types of DNA damage. The optimal substrates for MMR share similar features with some non-B DNA structures such as small loops formed at

simple repeats, hairpin structures formed at CNG triplet repeats where one mismatch exists in every three bp in the stem, and B to Z-DNA and B to H-DNA junctions. MMR has been shown to remove small cruciform structures (12 bp) in Chinese hamster ovary cells [106] and in yeast [107], but not large loops (*e.g.* 140 bp) [108–110]. Hairpin structures containing mismatches formed at CNG triplet repeats can recruit the MSH2-MSH3 complex [111]. MMR proteins have also been implicated in intermolecular triplex structure-induced mutagenesis [112–114]. MSH2-MSH3, together with XPA-RPA or XPC-RAD23B, has been shown to bind to intermolecular triplex structures with high affinity and specificity *in vitro* and *in vivo* [114]. H-DNA-forming GAA repeats induced chromosomal breakage in yeast, mediated by the MMR complexes MutSbeta and MutLalpha [115]. We have found that MSH2 deficiency in human cells reduced both H-DNA and Z-DNA-induced mutation frequencies by ~30–40% (Wang *et al.*, unpublished data), supporting a role for MMR proteins in non-B DNA induced genetic instability [115].

Base excision repair (BER) proteins remove oxidative and alkylation-induced DNA damage. Although similarities between non-B DNA structures and optimal BER substrates are not obvious, BER proteins have been shown to be involved in non-B DNA-induced mutagenesis. Many non-B DNA structures represent targets for DNA damaging agents, and repair of these damaged sites at or near non-B DNA regions by long-patch BER involves re-synthesis of a long tract of DNA by polymerase beta (pol beta) [116]. Pol beta is a low processivity, low fidelity polymerase [117, 118], and its non-continuous synthesis can result in misalignment of the nascent strand and template, resulting in contraction and expansion of repetitive DNA. In a mouse model of Huntington's disease (HD), pol beta was enriched on expanded and unstable CAG repeats in the striatum, but not on the more stable CAG repeats in the cerebellum [119]. Deficiency of pol beta and a BER cofactor, HMGB1, was found to suppress CAG expansion [120]. Moreover, if the repetitive sequence was located in the 5'-flap structure generated during long-patch BER, the single-stranded sequence was able to adopt a non-B conformation such as a hairpin or a stem-loop. Once a stable structure was formed at the end (concealing the 5' end), the flap structure became resistant to endonuclease activity of the flap-removing enzyme, Fen-1 [121]. The repeats within the flap structure reannealed and religated, resulting in expansion during the next cycle of replication [120]. In contrast, deficiency in the BER initiation and cleavage proteins [*e.g.* DNA glycosylase 8-oxoguanine glycosylase (OGG1)], has been found to suppress triplet repeat expansion in mice [122, 123].

Non-B DNA impacts DNA damage patterns and repair efficiency

DNA damaging agents and repair proteins interact with DNA within chromatin *in vivo*, thus the type, location, and amount of damage, as well as the efficiency and fidelity of repair, can be affected by DNA structural features and chromatin organization. Non-B DNA conformations can dramatically alter both the amount and location of DNA damage, and the repair efficiencies, and thus impact genetic instability.

Non-B DNA structures affect the accessibility of DNA to damaging agents

Bases in canonical B-form DNA are protected inside the sugar-phosphate backbone. Many types on non-B DNA conformations alter this arrangement, making bases more or less

accessible to certain types of DNA damaging agents. Single-stranded DNA, which is a common feature in most non-B DNA conformations, is more accessible to damage than B-DNA since the bases are more exposed. For example, triplex-duplex junctions are hyperreactive to DNA intercalators [124, 125], while the triplex regions are more resistant to such agents because the third strand in the major groove can prevent the binding of major groove intercalators [126]. B-Z junctions, which contain base extrusion and single-stranded regions, are hyperreactive to DNA interactive agents [127, 128]. Guanine bases adopt *syn* positions in Z-DNA, and are flipped out the Z-shaped sugar-phosphate backbone. This transition renders guanines in Z-DNA more sensitive to dimethylsulfate and diethylsulfate modification [128, 129]. This was also the case when guanines were located in the loop region of a hairpin structure; rendering the exposed G bases hypersensitive to peroxyntirite (relative to B-DNA) [130]. The minor-groove is narrower in A-tracts containing T-A steps [131], and AT-rich sequences [132], which can enhance the negative electrostatic potential of the DNA. Such sequences can also reduce the accessibility of H4' and H5'2 by OH radicals, impacting the radiation-induced DNA breakage events.

Non-B DNA structures affect DNA repair efficiency and accuracy

Non-B DNA conformations alter the structural features of B-DNA helices and thus the accessibility of DNA damaging agents and DNA repair protein to the damaged sites. The binding affinity and repair efficiency of hOGG1 protein on 8-oxoG in a hairpin structure was significantly reduced compared to 8-oxoG in B-DNA [133]. In the striatal tissue of HD mice carrying expanded CAG repeats, abasic sites located at the 5' ends of the CAG repeats in a hairpin conformation were repaired less efficiently than abasic sites in B-DNA [134]. Purified MSH2-MSH3 complex can bind to CAG repeats in a hairpin structure, but the ATP hydrolysis activity of MSH2-MSH3 was reduced upon binding to a hairpin structure that contained A-A mis-paired bases in the stem. However, the effects of other types of mismatches within the triplet repeat were not tested [111]. As mentioned above, guanines in Z-DNA are more sensitive to alkylating modification [128, 129], and once formed, the damage in Z-DNA (*e.g.* N7-methylguanine and O6-methylguanine), were resistant to excision by their repair enzymes, DNA glycosylase and O6-methylguanine-DNA methyltransferase, respectively [135, 136]. The persisting damage in non-B DNA conformations could result in hotspots for genetic instability.

Non-B DNA conformations alter chromatin structure

DNA damage and repair occur in the context of chromatin, in which the DNA is wrapped around histone cores, with the arginines on the histones protruded into the minor groove of the DNA helix [137]. The electrostatic interactions of negatively charged DNA and positively charged arginines are important in maintaining the nucleosome structure. DNA conformations that interrupt this arrangement can destabilize or diminish the nucleosome structure. GAA repeats, when in B-form DNA, were packed into a canonical nucleosome structure, but were refractory to nucleosome assembly *in vitro* when in a triplex structure [138]. Similarly, GA repeats, when in B-form DNA, were efficiently packed into nucleosome structures, but when the sequence adopted a triplex structure, nucleosome assembly was not possible unless the triplex conformation was disassembled [139]. AT-rich regions have been shown to be excluded from nucleosomes [140], and while the core

nucleosome DNA is typically GC-rich, a CG(9) repeat capable of adopting a Z-DNA structure was excluded from nucleosomes in *S. cerevisiae* [141]. Further, CpG repeats in the promoter region of the chicken adult beta-globin gene, capable of forming Z-DNA, were not able to bind to the histone octamer *in vitro* [142]. Putative G-quadruplex-forming sequences have been shown to co-localize with nucleosome-depleted regions in both human cells and in *C. elegans* [143]. While short CGG(7) repeats can be packed into nucleosome structures, the longer CGG(74) repeats, capable of forming a hairpins or G-quadruplex structures, were resistant to nucleosome assembly [144]. Importantly, it is possible that non-B DNA conformations in the genome can affect the tertiary DNA structures of adjacent regions. In one study, intermolecular triplexes formed by a 16-mer oligonucleotide binding to the center of a 199-bp duplex interrupted histone-DNA contacts not only within the triplex region, but also the flanking sequences up to ~100 bp from the triplex structure [145]. Our group recently found that CG(14) repeats were refractory to nucleosome assembly, particularly when the cytosines within the CG repeats were methylated [a modification that facilitates/stabilizes Z-DNA structures (Wang *et al.* unpublished data)]. Thus, the impact of non-B DNA on chromatin structure may, in part, account for distal mutations (point mutations, small deletions, or insertions) seen hundreds of bps from the non-B DNA-forming sequences [146, 147].

In summary, a non-B DNA conformation in the genome can remodel the chromatin structure, alter the accessibility of DNA to damaging agents, impact the efficiency and accuracy of DNA repair pathways, and result in genetic instability within and surrounding (over hundreds of bps) the structure-forming sequences. The non-B structure itself, in the absence of DNA damage *per se*, can recruit and stimulate the activity of DNA repair proteins.

The complexity of non-B DNA structure formation *in vivo*

Alternative DNA structures were first identified and characterized many decades ago, using techniques such as circular dichroism (CD) to provide signature spectropolarimetry for a non-B DNA structure under specific conditions [148]; enzyme or chemical probing for many types of non-B DNA structures ([12] and references therein); structure-, rather than sequence-specific antibodies [40, 149–151]; and direct visualization of some non-B DNA structures by electron microscopy, such as cruciforms, H-DNA, and (large structures containing G4 DNA on the G-rich strand and a RNA/DNA hybrid on the other) [93, 152–155]. However, our understanding of the structural features and biological functions of non-B DNA *in vivo* is still in its infancy and future studies are warranted. Studies to characterize alternative DNA structures have, for the most part, been performed *in vitro* using simplified models [12–15, 156]. While these studies have been invaluable in advancing the non-B DNA field, it is critical to characterize these structures in a chromosomal context *in vivo*, an area of research in which direct evidence is still lacking. Many computer-based search programs have been designed to identify potential non-B DNA-forming sequences in genomes, however most algorithms are designed with simplified search parameters based on our knowledge of DNA secondary structure formation *in vitro*. Thus such programs may result in false positive or negatives for a particular structure within repetitive sequences. For example, Z-DNA is known to form at pure alternating purine-pyrimidine sequences;

however, it has been demonstrated that sequences containing interruptions of the purine-pyrimidine alternation have been shown to adopt Z-DNA structures *in vitro* [157, 158]. More confounding are the dynamic conditions that impact non-B DNA structure formation (e.g. negative supercoiling, binding proteins, chromatin structure, epigenetics, DNA transactions), are not typically included in the search criteria. In support of this, we have found that cytosine methylation within CpG repeats can facilitate and stabilize Z-form DNA, impact nucleosome assembly, and induce higher mutation frequencies in human cells and mouse chromosomes (Wang *et al*, unpublished data). Thus, structure predictions based on computer programs can only serve as a preliminary screening/identification step.

Importantly, it appears that the distribution of non-B DNA-forming sequences in genomes is not random. There are many regions in the genome that comprise clusters of different types of overlapping non-B DNA-forming sequences. For example, rare fragile sites in the human genome contain sequences that have the capacity to adopt a number of non-B DNA secondary structures such as hairpins, cruciforms, and quadruplexes [159–161]. Another example includes the 21st intron of the human *PKD1* gene, which contains a 2.5 kb polypyrimidine tract with 23 mirror repeats with stem lengths of at least 10 nucleotides [162]. These repeats can adopt H-DNA structures *in vitro* [163]. This region is highly mutagenic in both germ line and somatic cells from autosomal dominant polycystic kidney disease patients [164]. However, the extent to which of the polypurine/polypyrimidine mirror repeat tracts form H-DNA at any given time and their direct contribution to the high mutation frequencies is difficult to determine. Another example is the human *c-MYC* promoter region (Figure 2), where multiple overlapping non-B DNA-forming sequences have been identified within a small region (~400 bp) surrounding the promoter P0 [23]. These include sequences with the capacity to adopt H-DNA [39], G-quadruplexes [37] and Z-DNA conformations [92], and each of these non-B DNA elements individually has been shown to be mutagenic in simplified models. However, the ability to demonstrate the extent to which one or another structure within this sequence impacts the formation of another and/or is solely responsible for the induced genetic instability *in vivo* remains a challenge. GAA repeats can adopt H-DNA structures, but they also have the propensity to form slippage loops; GC repeats can readily form Z-DNA, but also can form hairpins or small slippage loops; many large repetitive regions, such as Alu repeats, contain smaller units that can form various non-B DNA structures. In addition, the formation of a non-B DNA structure on a molecule will relax the negative supercoiling, which is required for most non-B DNA conformations. Thus, the extent to which one structure-forming sequence impacts another within a given region must be considered, as these may not represent topologically independent regions. Thus in a genomic area that contains multiple non-B DNA-forming sequences, one tract may form a specific non-B DNA conformation at a given time, and may inhibit or promote another form of non-B DNA at the same or nearby tracts under given conditions. Moreover, DNA is in a dynamic state such that transitions between non-B and B-DNA may be on-going, where a transition from non-B to B-DNA in one tract may give rise to non-B formation in other regions. Until a direct and conclusive method is developed to detect a specific DNA structure within a region of multiple non-B DNA forming-sequences, careful consideration is warranted when studying the biological effects of such a complex and dynamic region of the genome.

Acknowledgments

We thank Dr. Rick A. Finch for critical review of the manuscript. This work was supported by an NIH/NCI grant to K.M.V. (CA093729).

Abbreviations

DSB	DNA double-strand break
NER	Nucleotide excision repair
XPA	xeroderma pigmentosum group A
XPC	xeroderma pigmentosum group C
MMR	Mismatch repair
BER	base excision repair
Mbr	Major breakpoint region
OGG	DNA glycosylase 8-oxoguanine glycosylase
FXS	Fragile X syndrome

Literature Cited

1. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, et al. Initial sequencing and analysis of the human genome. *Nature*. 2001; 409:860–921. [PubMed: 11237011]
2. Choi J, Majima T. Conformational changes of non-B DNA. *Chem. Soc. Rev.* 2011; 40:5893–5909. [PubMed: 21901191]
3. Wells RD. Unusual DNA structures. *J. Biol. Chem.* 1988; 263:1095–1098. [PubMed: 3275663]
4. Wang G, Vasquez KM. Non-B DNA structure-induced genetic instability. *Mutat. Res.* 2006; 598:103–119. [PubMed: 16516932]
5. Djian P. Evolution of simple repeats in DNA and their relation to human disease. *Cell*. 1998; 94:155–160. [PubMed: 9695944]
6. Sinden RR, Potaman VN, Oussatcheva EA, Pearson CE, Lyubchenko YL, Shlyakhtenko LS. Triplet repeat DNA structures and human genetic disease: dynamic mutations from dynamic DNA. *J Biosci.* 2002; 27:53–65. [PubMed: 11927777]
7. Mirkin SM, Frank-Kamenetskii MD. H-DNA and related structures. *Annu. Rev. Biophys. Biomol. Struct.* 1994; 23:541–576. [PubMed: 7919793]
8. Htun H, Dahlberg JE. Topology and formation of triple-stranded H-DNA. *Science*. 1989; 243:1571–1576. [PubMed: 2648571]
9. Malfoy B, Rousseau N, Vogt N, Viegas-Pequignot E, Dutrillaux B, Leng M. Nucleotide sequence of an heterochromatic segment recognized by the antibodies to Z-DNA in fixed metaphase chromosomes. *Nucleic Acids Res.* 1986; 14:3197–3214. [PubMed: 3010230]
10. Johnston BH. Generation and detection of Z-DNA. *Methods Enzymol.* 1992; 211:127–158. [PubMed: 1406305]
11. Sen D, Gilbert W. A sodium-potassium switch in the formation of four-stranded G4-DNA. *Nature*. 1990; 344:410–414. [PubMed: 2320109]
12. Wang G, Zhao J, Vasquez KM. Methods to determine DNA structural alterations and genetic instability. *Methods*. 2009; 48:54–62. [PubMed: 19245837]
13. Palecek E. Probing DNA structure with osmium tetroxide complexes in vitro. *Methods Enzymol.* 1992; 212:139–155. [PubMed: 1518446]

14. Kohwi-Shigematsu T, Kohwi Y. Detection of non-B-DNA structures at specific sites in supercoiled plasmid DNA and chromatin with haloacetaldehyde and diethyl pyrocarbonate. *Methods Enzymol.* 1992; 212:155–180. [PubMed: 1518447]
15. Romier C, Dominguez R, Lahm A, Dahl O, Suck D. Recognition of single-stranded DNA by nuclease P1: high resolution crystal structures of complexes with substrate analogs. *Proteins.* 1998; 32:414–424. [PubMed: 9726413]
16. Wang G, Vasquez KM. Models for chromosomal replication-independent non-B DNA structure-induced genetic instability. *Mol. Carcinog.* 2009; 48:286–298. [PubMed: 19123200]
17. Sinden RR, Wells RD. DNA structure, mutations, and human genetic disease. *Curr. Opin. Biotechnol.* 1992; 3:612–622. [PubMed: 1369117]
18. McMurray CT. DNA secondary structure: a common and causative factor for expansion in human disease. *Proc. Natl. Acad. Sci. U. S. A.* 1999; 96:1823–1825. [PubMed: 10051552]
19. Balasubramanian S, Hurley LH, Neidle S. Targeting G-quadruplexes in gene promoters: a novel anticancer strategy? *Nat. Rev. Drug Discov.* 2011; 10:261–275. [PubMed: 21455236]
20. Wu Y, Brosh RM Jr. G-quadruplex nucleic acids and human disease. *The FEBS journal.* 2010; 277:3470–3488. [PubMed: 20670277]
21. Kikin O, D'Antonio L, Bagga PS. QGRS Mapper: a web-based server for predicting G-quadruplexes in nucleotide sequences. *Nucleic Acids Res.* 2006; 34:W676–W682. [PubMed: 16845096]
22. Cer RZ, Bruce KH, Donohue DE, Temiz NA, Mudunuri US, Yi M, Volfovsky N, Bacolla A, Luke BT, Collins JR, Stephens RM. Searching for non-B DNA-forming motifs using nBMST (non-B DNA motif search tool). *Curr Protoc Hum Genet.* 2012; Chapter 18(Unit 18 17):11–22.
23. Wang G, Gaddis S, Vasquez KM. Methods to detect replication-dependent and replication-independent DNA structure-induced genetic instability. *Methods.* 2013; 64:67–72. [PubMed: 23954565]
24. Raghavan SC, Lieber MR. Chromosomal translocations and non-B DNA structures in the human genome. *Cell Cycle.* 2004; 3:762–768. [PubMed: 15254430]
25. Rogozin IB, Pavlov YI. Theoretical analysis of mutation hotspots and their DNA sequence context specificity. *Mutat. Res.* 2003; 544:65–85. [PubMed: 12888108]
26. Chen X, Shen Y, Zhang F, Chiang C, Pillalamarri V, Blumenthal I, Talkowski M, Wu BL, Gusella JF. Molecular analysis of a deletion hotspot in the NRXN1 region reveals the involvement of short inverted repeats in deletion CNVs. *Am. J. Hum. Genet.* 2013; 92:375–386. [PubMed: 23472757]
27. Wiener F, Ohno S, Babonits M, Sumegi J, Wirschubsky Z, Klein G, Mushinski JF, Potter M. Hemizygous interstitial deletion of chromosome 15 (band D) in three translocation-negative murine plasmacytomas. *Proc. Natl. Acad. Sci. U. S. A.* 1984; 81:1159–1163. [PubMed: 6422467]
28. Akasaka T, Akasaka H, Ueda C, Yonetani N, Maesako Y, Shimizu A, Yamabe H, Fukuhara S, Uchiyama T, Ohno H. Molecular and clinical features of non-Burkitt's, diffuse large-cell lymphoma of B-cell type associated with the c-MYC/immunoglobulin heavy-chain fusion gene. *J. Clin. Oncol.* 2000; 18:510–518. [PubMed: 10653866]
29. Kovalchuk AL, Muller JR, Janz S. Deletional remodeling of c-myc-deregulating chromosomal translocations. *Oncogene.* 1997; 15:2369–2377. [PubMed: 9393881]
30. Joos S, Haluska FG, Falk MH, Henglein B, Hameister H, Croce CM, Bornkamm GW. Mapping chromosomal breakpoints of Burkitt's t(8;14) translocations far upstream of c-myc. *Cancer Res.* 1992; 52:6547–6552. [PubMed: 1330296]
31. Haluska FG, Tsujimoto Y, Croce CM. The t(8;14) breakpoint of the EW 36 undifferentiated lymphoma cell line lies 5' of MYC in a region prone to involvement in endemic Burkitt's lymphomas. *Nucleic Acids Res.* 1988; 16:2077–2085. [PubMed: 3128772]
32. Saglio G, Grazia Borrello M, Guerrasio A, Sozzi G, Serra A, di Celle PF, Foa R, Ferrarini M, Roncella S, Borgna Pignatti C, et al. Preferential clustering of chromosomal breakpoints in Burkitt's lymphomas and L3 type acute lymphoblastic leukemias with a t(8;14) translocation. *Genes Chromosomes Cancer.* 1993; 8:1–7. [PubMed: 7691153]
33. Care A, Cianetti L, Giampaolo A, Sposi NM, Zappavigna V, Mavilio F, Alimena G, Amadori S, Mandelli F, Peschle C. Translocation of c-myc into the immunoglobulin heavy-chain locus in

- human acute B-cell leukemia. A molecular analysis. *EMBO J.* 1986; 5:905–911. [PubMed: 3013623]
34. Wilda M, Busch K, Klose I, Keller T, Woessmann W, Kreuder J, Harbott J, Borkhardt A. Level of MYC overexpression in pediatric Burkitt's lymphoma is strongly dependent on genomic breakpoint location within the MYC locus. *Genes Chromosomes Cancer.* 2004; 41:178–182. [PubMed: 15287031]
 35. Rimokh R, Rouault JP, Wahbi K, Gadoux M, Lafage M, Archimbaud E, Charrin C, Gentilhomme O, Germain D, Samarut J, et al. A chromosome 12 coding region is juxtaposed to the MYC protooncogene locus in a t(8;12)(q24;q22) translocation in a case of B-cell chronic lymphocytic leukemia. *Genes Chromosomes Cancer.* 1991; 3:24–36. [PubMed: 2069907]
 36. Wolf S, Wittig B, Rich A. Identification of transcriptionally induced Z-DNA segments in the human c-myc gene. *Biochim. Biophys. Acta.* 1995; 1264:294–302. [PubMed: 8547317]
 37. Siddiqui-Jain A, Grand CL, Bearss DJ, Hurley LH. Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. *Proc. Natl. Acad. Sci. U. S. A.* 2002; 99:11593–11598. [PubMed: 12195017]
 38. Grand CL, Han H, Munoz RM, Weitman S, Von Hoff DD, Hurley LH, Bearss DJ. The cationic porphyrin TMPyP4 down-regulates c-MYC and human telomerase reverse transcriptase expression and inhibits tumor growth in vivo. *Mol. Cancer Ther.* 2002; 1:565–573. [PubMed: 12479216]
 39. Kinniburgh AJ. A cis-acting transcription element of the c-myc gene can assume an H-DNA conformation. *Nucleic Acids Res.* 1989; 17:7771–7778. [PubMed: 2678005]
 40. Raghavan SC, Chastain P, Lee JS, Hegde BG, Houston S, Langen R, Hsieh CL, Haworth IS, Lieber MR. Evidence for a triplex DNA conformation at the bcl-2 major breakpoint region of the t(14;18) translocation. *J. Biol. Chem.* 2005; 280:22749–22760. [PubMed: 15840562]
 41. Boehm T, Mingle-Gaw L, Kees UR, Spurr N, Lavenir I, Forster A, Rabbitts TH. Alternating purine-pyrimidine tracts may promote chromosomal translocations seen in a variety of human lymphoid tumours. *EMBO J.* 1989; 8:2621–2631. [PubMed: 2531086]
 42. Sinclair PB, Parker H, An Q, Rand V, Ensor H, Harrison CJ, Strefford JC. Analysis of a breakpoint cluster reveals insight into the mechanism of intrachromosomal amplification in a lymphoid malignancy. *Hum. Mol. Genet.* 2011; 20:2591–2602. [PubMed: 21487021]
 43. Kurahashi H, Emanuel BS. Long AT-rich palindromes and the constitutional t(11;22) breakpoint. *Hum. Mol. Genet.* 2001; 10:2605–2617. [PubMed: 11726547]
 44. Edelmann L, Spiteri E, Koren K, Pulijaal V, Bialer MG, Shanske A, Goldberg R, Morrow BE. AT-rich palindromes mediate the constitutional t(11;22) translocation. *Am. J. Hum. Genet.* 2001; 68:1–13. [PubMed: 11095996]
 45. Kurahashi H, Shaikh TH, Hu P, Roe BA, Emanuel BS, Budarf ML. Regions of genomic instability on 22q11 and 11q23 as the etiology for the recurrent constitutional t(11;22). *Hum. Mol. Genet.* 2000; 9:1665–1670. [PubMed: 10861293]
 46. Kurahashi H, Inagaki H, Ohye T, Kogo H, Kato T, Emanuel BS. Palindrome-mediated chromosomal translocations in humans. *DNA Repair (Amst).* 2006; 5:1136–1145. [PubMed: 16829213]
 47. Juranek SA, Paeschke K. Cell cycle regulation of G-quadruplex DNA structures at telomeres. *Curr. Pharm. Des.* 2012; 18:1867–1872. [PubMed: 22376109]
 48. Capra JA, Paeschke K, Singh M, Zakian VA. G-quadruplex DNA sequences are evolutionarily conserved and associated with distinct genomic features in *Saccharomyces cerevisiae*. *PLoS Comput. Biol.* 2010; 6:e1000861. [PubMed: 20676380]
 49. Vasquez KM, Wang G. The yin and yang of repair mechanisms in DNA structure-induced genetic instability. *Mutat. Res.* 2013; 743–744:118–131.
 50. Zhao J, Bacolla A, Wang G, Vasquez KM. Non-B DNA structure-induced genetic instability and evolution. *Cell. Mol. Life Sci.* 2010; 67:43–62. [PubMed: 19727556]
 51. Raghavan SC, Lieber MR. DNA structures at chromosomal translocation sites. *Bioessays.* 2006; 28:480–494. [PubMed: 16615081]
 52. Pearson CE, Sinden RR. Trinucleotide repeat DNA structures: dynamic mutations from dynamic DNA. *Curr. Opin. Struct. Biol.* 1998; 8:321–330. [PubMed: 9666328]

53. Mirkin SM. Expandable DNA repeats and human disease. *Nature*. 2007; 447:932–940. [PubMed: 17581576]
54. Wang G, Carbajal S, Vijg J, DiGiovanni J, Vasquez KM. DNA structure-induced genomic instability in vivo. *J. Natl. Cancer Inst.* 2008; 100:1815–1817. [PubMed: 19066276]
55. Wang G, Vasquez KM. Naturally occurring H-DNA-forming sequences are mutagenic in mammalian cells. *Proc. Natl. Acad. Sci. U. S. A.* 2004; 101:13448–13453. [PubMed: 15342911]
56. Wang G, Christensen LA, Vasquez KM. Z-DNA-forming sequences generate large-scale deletions in mammalian cells. *Proc. Natl. Acad. Sci. U. S. A.* 2006; 103:2677–2682. [PubMed: 16473937]
57. Martorell L, Monckton DG, Gamez J, Johnson KJ, Gich I, Lopez de Munain A, Baiget M. Progression of somatic CTG repeat length heterogeneity in the blood cells of myotonic dystrophy patients. *Hum. Mol. Genet.* 1998; 7:307–312. [PubMed: 9425239]
58. Wong LJ, Ashizawa T, Monckton DG, Caskey CT, Richards CS. Somatic heterogeneity of the CTG repeat in myotonic dystrophy is age and size dependent. *Am. J. Hum. Genet.* 1995; 56:114–122. [PubMed: 7825566]
59. Martorell L, Martinez JM, Carey N, Johnson K, Baiget M. Comparison of CTG repeat length expansion and clinical progression of myotonic dystrophy over a five year period. *J. Med. Genet.* 1995; 32:593–596. [PubMed: 7473648]
60. Wohrle D, Kennerknecht I, Wolf M, Enders H, Schwemmle S, Steinbach P. Heterogeneity of DM kinase repeat expansion in different fetal tissues and further expansion during cell proliferation in vitro: evidence for a casual involvement of methyl-directed DNA mismatch repair in triplet repeat stability. *Hum. Mol. Genet.* 1995; 4:1147–1153. [PubMed: 8528201]
61. Anvret M, Ahlberg G, Grandell U, Hedberg B, Johnson K, Edstrom L. Larger expansions of the CTG repeat in muscle compared to lymphocytes from patients with myotonic dystrophy. *Hum. Mol. Genet.* 1993; 2:1397–1400. [PubMed: 8242063]
62. Thornton CA, Johnson K, Moxley RT 3rd. Myotonic dystrophy patients have larger CTG expansions in skeletal muscle than in leukocytes. *Ann. Neurol.* 1994; 35:104–107. [PubMed: 8285579]
63. Zatz M, Passos-Bueno MR, Cerqueira A, Marie SK, Vainzof M, Pavanetto RC. Analysis of the CTG repeat in skeletal muscle of young and adult myotonic dystrophy patients: when does the expansion occur? *Hum. Mol. Genet.* 1995; 4:401–406. [PubMed: 7795594]
64. Kang S, Jaworski A, Ohshima K, Wells RD. Expansion and deletion of CTG repeats from human disease genes are determined by the direction of replication in *E. coli*. *Nat. Genet.* 1995; 10:213–218. [PubMed: 7663518]
65. Miret JJ, Pessoa-Brandao L, Lahue RS. Orientation-dependent and sequence-specific expansions of CTG/CAG trinucleotide repeats in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. U. S. A.* 1998; 95:12438–12443. [PubMed: 9770504]
66. Samadashwily GM, Raca G, Mirkin SM. Trinucleotide repeats affect DNA replication in vivo. *Nat. Genet.* 1997; 17:298–304. [PubMed: 9354793]
67. Trinh TQ, Sinden RR. Preferential DNA secondary structure mutagenesis in the lagging strand of replication in *E. coli*. *Nature*. 1991; 352:544–547. [PubMed: 1865910]
68. Hashem VI, Sinden RR. Duplications between direct repeats stabilized by DNA secondary structure occur preferentially in the leading strand during DNA replication. *Mutat. Res.* 2005; 570:215–226. [PubMed: 15708580]
69. Iyer RR, Wells RD. Expansion and deletion of triplet repeat sequences in *Escherichia coli* occur on the leading strand of DNA replication. *J. Biol. Chem.* 1999; 274:3865–3877. [PubMed: 9920942]
70. Pelletier R, Krasilnikova MM, Samadashwily GM, Lahue R, Mirkin SM. Replication and expansion of trinucleotide repeats in yeast. *Mol. Cell. Biol.* 2003; 23:1349–1357. [PubMed: 12556494]
71. Hoyne PR, Maher LJ 3rd. Functional studies of potential intrastrand triplex elements in the *Escherichia coli* genome. *J. Mol. Biol.* 2002; 318:373–386. [PubMed: 12051844]
72. Hile SE, Eckert KA. Positive correlation between DNA polymerase alpha-primase pausing and mutagenesis within polypyrimidine/polypurine microsatellite sequences. *J. Mol. Biol.* 2004; 335:745–759. [PubMed: 14687571]

73. Rao BS. Pausing of simian virus 40 DNA replication fork movement in vivo by (dG-dA)_n.(dT-dC)_n tracts. *Gene*. 1994; 140:233–237. [PubMed: 8144031]
74. Krasilnikova MM, Mirkin SM. Replication stalling at Friedreich's ataxia (GAA)_n repeats in vivo. *Mol. Cell. Biol.* 2004; 24:2286–2295. [PubMed: 14993268]
75. Voineagu I, Narayanan V, Lobachev KS, Mirkin SM. Replication stalling at unstable inverted repeats: interplay between DNA hairpins and fork stabilizing proteins. *Proc. Natl. Acad. Sci. U. S. A.* 2008; 105:9936–9941. [PubMed: 18632578]
76. Chu WK, Hickson ID. RecQ helicases: multifunctional genome caretakers. *Nat. Rev. Cancer*. 2009; 9:644–654. [PubMed: 19657341]
77. Peleg M, Kopel V, Borowiec JA, Manor H. Formation of DNA triple helices inhibits DNA unwinding by the SV40 large T-antigen helicase. *Nucleic Acids Res.* 1995; 23:1292–1299. [PubMed: 7753619]
78. Hartenstine MJ, Goodman MF, Petruska J. Base stacking and even/odd behavior of hairpin loops in DNA triplet repeat slippage and expansion with DNA polymerase. *J. Biol. Chem.* 2000; 275:18382–18390. [PubMed: 10849445]
79. Wang Q, Liu JQ, Chen Z, Zheng KW, Chen CY, Hao YH, Tan Z. G-quadruplex formation at the 3' end of telomere DNA inhibits its extension by telomerase, polymerase and unwinding by helicase. *Nucleic Acids Res.* 2011; 39:6229–6237. [PubMed: 21441540]
80. Han H, Hurley LH, Salazar M. A DNA polymerase stop assay for G-quadruplex-interactive compounds. *Nucleic Acids Res.* 1999; 27:537–542. [PubMed: 9862977]
81. Paeschke K, Capra JA, Zakian VA. DNA replication through G-quadruplex motifs is promoted by the *Saccharomyces cerevisiae* Pif1 DNA helicase. *Cell*. 2011; 145:678–691. [PubMed: 21620135]
82. Anand RP, Shah KA, Niu H, Sung P, Mirkin SM, Freudenreich CH. Overcoming natural replication barriers: differential helicase requirements. *Nucleic Acids Res.* 2012; 40:1091–1105. [PubMed: 21984413]
83. Voineagu I, Surka CF, Shishkin AA, Krasilnikova MM, Mirkin SM. Replisome stalling and stabilization at CGG repeats, which are responsible for chromosomal fragility. *Nat. Struct. Mol. Biol.* 2009; 16:226–228. [PubMed: 19136957]
84. Levinson G, Gutman GA. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Mol. Biol. Evol.* 1987; 4:203–221. [PubMed: 3328815]
85. Chong SS, McCall AE, Cota J, Subramony SH, Orr HT, Hughes MR, Zoghbi HY. Gametic and somatic tissue-specific heterogeneity of the expanded SCA1 CAG repeat in spinocerebellar ataxia type 1. *Nat. Genet.* 1995; 10:344–350. [PubMed: 7670474]
86. Telenius H, Kremer B, Goldberg YP, Theilmann J, Andrew SE, Zeisler J, Adam S, Greenberg C, Ives EJ, Clarke LA, et al. Somatic and gonadal mosaicism of the Huntington disease gene CAG repeat in brain and sperm. *Nat. Genet.* 1994; 6:409–414. [PubMed: 8054984]
87. Hashida H, Goto J, Kurisaki H, Mizusawa H, Kanazawa I. Brain regional differences in the expansion of a CAG repeat in the spinocerebellar ataxias: dentatorubral-pallidolysian atrophy, Machado-Joseph disease, and spinocerebellar ataxia type 1. *Ann. Neurol.* 1997; 41:505–511. [PubMed: 9124808]
88. Ansved T, Lundin A, Anvret M. Larger CAG expansions in skeletal muscle compared with lymphocytes in Kennedy disease but not in Huntington disease. *Neurology*. 1998; 51:1442–1444. [PubMed: 9818876]
89. Volker J, Plum GE, Klump HH, Breslauer KJ. DNA repair and DNA triplet repeat expansion: the impact of abasic lesions on triplet repeat DNA energetics. *J. Am. Chem. Soc.* 2009; 131:9354–9360. [PubMed: 19566100]
90. Butler DK, Yasuda LE, Yao MC. Induction of large DNA palindrome formation in yeast: implications for gene amplification and genome stability in eukaryotes. *Cell*. 1996; 87:1115–1122. [PubMed: 8978615]
91. Wittig B, Dorbic T, Rich A. Transcription is associated with Z-DNA formation in metabolically active permeabilized mammalian cell nuclei. *Proc. Natl. Acad. Sci. U. S. A.* 1991; 88:2259–2263. [PubMed: 2006166]

92. Wittig B, Wolfl S, Dorbic T, Vahrson W, Rich A. Transcription of human c-myc in permeabilized nuclei is associated with formation of Z-DNA in three discrete regions of the gene. *EMBO J.* 1992; 11:4653–4663. [PubMed: 1330542]
93. Duquette ML, Handa P, Vincent JA, Taylor AF, Maizels N. Intracellular transcription of G-rich DNAs induces formation of G-loops, novel structures containing G4 DNA. *Genes Dev.* 2004; 18:1618–1629. [PubMed: 15231739]
94. Belotserkovskii BP, di Silva E, Tornaletti S, Wang G, Vasquez KM, Hanawalt PC. A triplex-forming sequence from the human c-Myc promoter interferes with DNA transcription. *J. Biol. Chem.* 2007; 282:32433–32441. [PubMed: 17785457]
95. Mitchell DL, Nairn RS. The biology of the (6-4) photoproduct. *Photochem. Photobiol.* 1989; 49:805–819. [PubMed: 2672059]
96. Gunz D, Hess MT, Naegeli H. Recognition of DNA adducts by human nucleotide excision repair. Evidence for a thermodynamic probing mechanism. *The Journal of biological chemistry.* 1996; 271:25089–25098. [PubMed: 8810263]
97. Buschta-Hedayat N, Buterin T, Hess MT, Missura M, Naegeli H. Recognition of nonhybridizing base pairs during nucleotide excision repair of DNA. *Proc. Natl. Acad. Sci. U. S. A.* 1999; 96:6090–6095. [PubMed: 10339546]
98. Wood RD. DNA damage recognition during nucleotide excision repair in mammalian cells. *Biochimie.* 1999; 81:39–44. [PubMed: 10214908]
99. Cooney M, Czernuszewicz G, Postel EH, Flint SJ, Hogan ME. Site-specific oligonucleotide binding represses transcription of the human c-myc gene in vitro. *Science.* 1988; 241:456–459. [PubMed: 3293213]
100. Wang G, Seidman MM, Glazer PM. Mutagenesis in mammalian cells induced by triple helix formation and transcription-coupled repair. *Science.* 1996; 271:802–805. [PubMed: 8628995]
101. Wang G, Chen Z, Zhang S, Wilson GL, Jing K. Detection and determination of oligonucleotide triplex formation-mediated transcription-coupled DNA repair in HeLa nuclear extracts. *Nucleic Acids Res.* 2001; 29:1801–1807. [PubMed: 11292853]
102. Vasquez KM, Christensen J, Li L, Finch RA, Glazer PM. Human XPA and RPA DNA repair proteins participate in specific recognition of triplex-induced helical distortions. *Proc. Natl. Acad. Sci. U. S. A.* 2002; 99:5848–5853. [PubMed: 11972036]
103. Thoma BS, Wakasugi M, Christensen J, Reddy MC, Vasquez KM. Human XPC-hHR23B interacts with XPA-RPA in the recognition of triplex-directed psoralen DNA interstrand crosslinks. *Nucleic Acids Res.* 2005; 33:2993–3001. [PubMed: 15914671]
104. Lin Y, Dion V, Wilson JH. Transcription promotes contraction of CAG repeat tracts in human cells. *Nat. Struct. Mol. Biol.* 2006; 13:179–180. [PubMed: 16388310]
105. Lin Y, Wilson JH. Transcription-induced CAG repeat contraction in human cells is mediated in part by transcription-coupled nucleotide excision repair. *Mol. Cell. Biol.* 2007; 27:6209–6217. [PubMed: 17591697]
106. Taghian DG, Hough H, Nickoloff JA. Biased short tract repair of palindromic loop mismatches in mammalian cells. *Genetics.* 1998; 148:1257–1268. [PubMed: 9539440]
107. Nag DK, Petes TD. Seven-base-pair inverted repeats in DNA form stable hairpins in vivo in *Saccharomyces cerevisiae*. *Genetics.* 1991; 129:669–673. [PubMed: 1752412]
108. Tran H, Degtyareva N, Gordenin D, Resnick MA. Altered replication and inverted repeats induce mismatch repair-independent recombination between highly diverged DNAs in yeast. *Mol. Cell. Biol.* 1997; 17:1027–1036. [PubMed: 9001255]
109. Nag DK, Kurtz A. A 140-bp-long palindromic sequence induces double-strand breaks during meiosis in the yeast *Saccharomyces cerevisiae*. *Genetics.* 1997; 146:835–847. [PubMed: 9215890]
110. Bill CA, Taghian DG, Duran WA, Nickoloff JA. Repair bias of large loop mismatches during recombination in mammalian cells depends on loop length and structure. *Mutat. Res.* 2001; 485:255–265. [PubMed: 11267836]
111. Owen BA, Yang Z, Lai M, Gajek M, Badger JD, Hayes JJ, Edelmann W, Kucherlapati R, Wilson TM, McMurray CT. (CAG)(n)-hairpin DNA binds to Msh2-Msh3 and changes properties of mismatch recognition. *Nat. Struct. Mol. Biol.* 2005; 12:663–670. [PubMed: 16025128]

112. Wu Q, Christensen LA, Legerski RJ, Vasquez KM. Mismatch repair participates in error-free processing of DNA interstrand crosslinks in human cells. *EMBO Rep.* 2005; 6:551–557. [PubMed: 15891767]
113. Wu Q, Vasquez KM. Human MLH1 protein participates in genomic damage checkpoint signaling in response to DNA interstrand crosslinks, while MSH2 functions in DNA repair. *PLoS Genet.* 2008; 4:e1000189. [PubMed: 18787700]
114. Zhao J, Jain A, Iyer RR, Modrich PL, Vasquez KM. Mismatch repair and nucleotide excision repair proteins cooperate in the recognition of DNA interstrand crosslinks. *Nucleic Acids Res.* 2009; 37:4420–4429. [PubMed: 19468048]
115. Kim HM, Narayanan V, Mieczkowski PA, Petes TD, Krasilnikova MM, Mirkin SM, Lobachev KS. Chromosome fragility at GAA tracts in yeast depends on repeat orientation and requires mismatch repair. *EMBO J.* 2008; 27:2896–2906. [PubMed: 18833189]
116. Dogliotti E, Fortini P, Pascucci B, Parlanti E. The mechanism of switching among multiple BER pathways. *Prog. Nucleic Acid Res. Mol. Biol.* 2001; 68:3–27. [PubMed: 11554307]
117. Sweasy JB. Fidelity mechanisms of DNA polymerase beta. *Prog. Nucleic Acid Res. Mol. Biol.* 2003; 73:137–169. [PubMed: 12882517]
118. Beard WA, Shock DD, Wilson SH. Influence of DNA structure on DNA polymerase beta active site function: extension of mutagenic DNA intermediates. *J. Biol. Chem.* 2004; 279:31921–31929. [PubMed: 15145936]
119. Goula AV, Berquist BR, Wilson DM 3rd, Wheeler VC, Trottier Y, Merienne K. Stoichiometry of base excision repair proteins correlates with increased somatic CAG instability in striatum over cerebellum in Huntington's disease transgenic mice. *PLoS Genet.* 2009; 5:e1000749. [PubMed: 19997493]
120. Liu Y, Prasad R, Beard WA, Hou EW, Horton JK, McMurray CT, Wilson SH. Coordination between polymerase beta and FEN1 can modulate CAG repeat expansion. *The Journal of biological chemistry.* 2009; 284:28352–28366. [PubMed: 19674974]
121. Ruggiero BL, Topal MD. Triplet repeat expansion generated by DNA slippage is suppressed by human flap endonuclease 1. *J. Biol. Chem.* 2004; 279:23088–23097. [PubMed: 15037629]
122. Kovtun IV, Liu Y, Bjoras M, Klungland A, Wilson SH, McMurray CT. OGG1 initiates age-dependent CAG trinucleotide expansion in somatic cells. *Nature.* 2007; 447:447–452. [PubMed: 17450122]
123. Kovtun IV, Johnson KO, McMurray CT. Cockayne syndrome B protein antagonizes OGG1 in modulating CAG repeat length in vivo. *Aging.* 2011; 3:509–514. [PubMed: 21566259]
124. Sun JS, Lavery R, Chomilier J, Zakrzewska K, Montenay-Garestier T, Helene C. Theoretical study of ethidium intercalation in triple-stranded DNA and at triplex-duplex junctions. *J. Biomol. Struct. Dyn.* 1991; 9:425–436. [PubMed: 1815636]
125. Collier DA, Mergny JL, Thuong NT, Helene C. Site-specific intercalation at the triplex-duplex junction induces a conformational change which is detectable by hypersensitivity to diethylpyrocarbonate. *Nucleic Acids Res.* 1991; 19:4219–4224. [PubMed: 1870975]
126. Sun JS, Francois JC, Montenay-Garestier T, Saison-Behmoaras T, Roig V, Thuong NT, Helene C. Sequence-specific intercalating agents: intercalation at specific sequences on duplex DNA via major groove recognition by oligonucleotide-intercalator conjugates. *Proc. Natl. Acad. Sci. U. S. A.* 1989; 86:9198–9202. [PubMed: 2594761]
127. Rodolfo C, Lanza A, Tornaletti S, Fronza G, Pedrini AM. The ultimate carcinogen of 4-nitroquinoline 1-oxide does not react with Z-DNA and hyperreacts with B-Z junctions. *Nucleic Acids Res.* 1994; 22:314–320. [PubMed: 8127667]
128. Johnston BH, Rich A. Chemical probes of DNA conformation: detection of Z-DNA at nucleotide resolution. *Cell.* 1985; 42:713–724. [PubMed: 4053183]
129. Zimmerman SB. The three-dimensional structure of DNA. *Annu. Rev. Biochem.* 1982; 51:395–427. [PubMed: 7051961]
130. Jarem DA, Wilson NR, Delaney S. Structure-dependent DNA damage and repair in a trinucleotide repeat sequence. *Biochemistry (Mosc.).* 2009; 48:6655–6663.
131. Haran TE, Mohanty U. The unique structure of A-tracts and intrinsic DNA bending. *Q. Rev. Biophys.* 2009; 42:41–81. [PubMed: 19508739]

132. Stella S, Cascio D, Johnson RC. The shape of the DNA minor groove directs binding by the DNA-bending protein Fis. *Genes Dev.* 2010; 24:814–826. [PubMed: 20395367]
133. Jarem DA, Wilson NR, Schermerhorn KM, Delaney S. Incidence and persistence of 8-oxo-7,8-dihydroguanine within a hairpin intermediate exacerbates a toxic oxidation cycle associated with trinucleotide repeat expansion. *DNA Repair.* 2011; 10:887–896. [PubMed: 21727036]
134. Goula AV, Pearson CE, Della Maria J, Trottier Y, Tomkinson AE, Wilson DM 3rd, Merienne K. The nucleotide sequence, DNA damage location, and protein stoichiometry influence the base excision repair outcome at CAG/CTG repeats. *Biochemistry (Mosc.).* 2012; 51:3919–3932.
135. Lagravere C, Malfoy B, Leng M, Laval J. Ring-opened alkylated guanine is not repaired in Z-DNA. *Nature.* 1984; 310:798–800. [PubMed: 6382026]
136. Boiteux S, Costa de Oliveira R, Laval J. The Escherichia coli O6-methylguanine-DNA methyltransferase does not repair promutagenic O6-methylguanine residues when present in Z-DNA. *J. Biol. Chem.* 1985; 260:8711–8715. [PubMed: 3894348]
137. Luger K, Mader AW, Richmond RK, Sargent DF, Richmond TJ. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature.* 1997; 389:251–260. [PubMed: 9305837]
138. Ruan H, Wang YH. Friedreich's ataxia GAA.TTC duplex and GAA.GAA.TTC triplex structures exclude nucleosome assembly. *J. Mol. Biol.* 2008; 383:292–300. [PubMed: 18775729]
139. Espinas ML, Jimenez-Garcia E, Martinez-Balbas A, Azorin F. Formation of triple-stranded DNA at d(GA.TC)_n sequences prevents nucleosome assembly and is hindered by nucleosomes. *J. Biol. Chem.* 1996; 271:31807–31812. [PubMed: 8943221]
140. Takagi H, Inai Y, Watanabe S, Tatemoto S, Yajima M, Akasaka K, Yamamoto T, Sakamoto N. Nucleosome exclusion from the interspecies-conserved central AT-rich region of the Ars insulator. *J. Biochem. (Tokyo).* 2012; 151:75–87. [PubMed: 21930654]
141. Wong B, Chen S, Kwon JA, Rich A. Characterization of Z-DNA as a nucleosome-boundary element in yeast *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. U. S. A.* 2007; 104:2229–2234. [PubMed: 17284586]
142. Davey C, Pennings S, Allan J. CpG methylation remodels chromatin structure in vitro. *J. Mol. Biol.* 1997; 267:276–288. [PubMed: 9096225]
143. Wong HM, Huppert JL. Stable G-quadruplexes are found outside nucleosome-bound regions. *Mol. Biosyst.* 2009; 5:1713–1719. [PubMed: 19585004]
144. Godde JS, Kass SU, Hirst MC, Wolffe AP. Nucleosome assembly on methylated CGG triplet repeats in the fragile X mental retardation gene 1 promoter. *The Journal of biological chemistry.* 1996; 271:24325–24328. [PubMed: 8798682]
145. Westin L, Blomquist P, Milligan JF, Wrangé O. Triple helix DNA alters nucleosomal histone-DNA interactions and acts as a nucleosome barrier. *Nucleic Acids Res.* 1995; 23:2184–2191. [PubMed: 7610046]
146. Bacolla A, Wang G, Jain A, Chuzhanova NA, Cer RZ, Collins JR, Cooper DN, Bohr VA, Vasquez KM. Non-B DNA-forming sequences and WRN deficiency independently increase the frequency of base substitution in human cells. *The Journal of biological chemistry.* 2011; 286:10017–10026. [PubMed: 21285356]
147. Shishkin AA, Voineagu I, Matera R, Cherng N, Chernet BT, Krasilnikova MM, Narayanan V, Lobachev KS, Mirkin SM. Large-scale expansions of Friedreich's ataxia GAA repeats in yeast. *Mol. Cell.* 2009; 35:82–92. [PubMed: 19595718]
148. Baase WA, Johnson WC Jr. Circular dichroism and DNA secondary structure. *Nucleic Acids Res.* 1979; 6:797–814. [PubMed: 424316]
149. Agazie YM, Lee JS, Burkholder GD. Characterization of a new monoclonal antibody to triplex DNA and immunofluorescent staining of mammalian chromosomes. *J. Biol. Chem.* 1994; 269:7019–7023. [PubMed: 7509814]
150. Lee JS, Burkholder GD, Latimer LJ, Haug BL, Braun RP. A monoclonal antibody to triplex DNA binds to eucaryotic chromosomes. *Nucleic Acids Res.* 1987; 15:1047–1061. [PubMed: 2434928]
151. Lee JS, Latimer LJ, Haug BL, Pulleyblank DE, Skinner DM, Burkholder GD. Triplex DNA in plasmids and chromosomes. *Gene.* 1989; 82:191–199. [PubMed: 2583520]
152. Mikheikin AL, Lushnikov AY, Lyubchenko YL. Effect of DNA supercoiling on the geometry of holliday junctions. *Biochemistry (Mosc.).* 2006; 45:12998–13006.

153. Shlyakhtenko LS, Potaman VN, Sinden RR, Lyubchenko YL. Structure and dynamics of supercoil-stabilized DNA cruciforms. *J. Mol. Biol.* 1998; 280:61–72. [PubMed: 9653031]
154. Kurahashi H, Inagaki H, Yamada K, Ohye T, Taniguchi M, Emanuel BS, Toda T. Cruciform DNA structure underlies the etiology for palindrome-mediated human chromosomal translocations. *J. Biol. Chem.* 2004; 279:35377–35383. [PubMed: 15208332]
155. Vetcher AA, Napierala M, Iyer RR, Chastain PD, Griffith JD, Wells RD. Sticky DNA, a long GAA.GAA.TTC triplex that is formed intramolecularly, in the sequence of intron 1 of the frataxin gene. *J. Biol. Chem.* 2002; 277:39217–39227. [PubMed: 12161437]
156. Raghavan SC, Tsai A, Hsieh CL, Lieber MR. Analysis of non-B DNA structure at chromosomal sites in the mammalian genome. *Methods Enzymol.* 2006; 409:301–316. [PubMed: 16793408]
157. Feigon J, Wang AH, van der Marel GA, van Boom JH, Rich A. Z-DNA forms without an alternating purine-pyrimidine sequence in solution. *Science.* 1985; 230:82–84. [PubMed: 4035359]
158. Eichman BF, Schroth GP, Basham BE, Ho PS. The intrinsic structure and stability of out-of-alternation base pairs in Z-DNA. *Nucleic Acids Res.* 1999; 27:543–550. [PubMed: 9862978]
159. Freudenreich CH. Chromosome fragility: molecular mechanisms and cellular consequences. *Front. Biosci.* 2007; 12:4911–4924. [PubMed: 17569619]
160. Zlotorynski E, Rahat A, Skaug J, Ben-Porat N, Ozeri E, Hershberg R, Levi A, Scherer SW, Margalit H, Kerem B. Molecular basis for expression of common and rare fragile sites. *Mol. Cell. Biol.* 2003; 23:7143–7151. [PubMed: 14517285]
161. Sutherland GR. Rare fragile sites. *Cytogenet. Genome Res.* 2003; 100:77–84. [PubMed: 14526166]
162. Van Raay TJ, Burn TC, Connors TD, Petry LR, Germino GG, Klinger KW, Landes GM. A 2.5 kb polypyrimidine tract in the PKD1 gene contains at least 23 H-DNA-forming sequences. *Microb. Comp. Genomics.* 1996; 1:317–327. [PubMed: 9689215]
163. Blaszak RT, Potaman V, Sinden RR, Bissler JJ. DNA structural transitions within the PKD1 gene. *Nucleic Acids Res.* 1999; 27:2610–2617. [PubMed: 10373576]
164. Watnick TJ, Piontek KB, Cordal TM, Weber H, Gandolph MA, Qian F, Lens XM, Neumann HP, Germino GG. An unusual pattern of mutation in the duplicated portion of PKD1 is revealed by use of a novel strategy for mutation detection. *Hum. Mol. Genet.* 1997; 6:1473–1481. [PubMed: 9285784]

Figure 1. Non-B DNA structures. (A) Cruciform DNA formed at inverted repeats, (B) left-handed Z-DNA formed at alternating purine-pyrimidine sequences, (C) intermolecular triplex H-DNA formed at mirror repeat symmetric polypurine/polypyrimidine regions, (D) G-quadruplex (tetraplex) DNA formed at sequences containing at least four runs of three or more guanines, and (E) Slipped DNA formed by misalignment of simple repeats. From reference [50] with permission.

Figure 2.

Schematic of the human *c-MYC* gene promoter and clusters of sequences with the potential to adopt non-B DNA structures. The alternative promoters P0, P1, P2, and Exon 1 are shown. Within a region of 400 bp surrounding P0, there are clusters of non-B DNA-forming sequences, including four Z-DNA-forming sequences [92] (red); two H-DNA-forming sequences [39] (black); and one G-quadruplex-forming sequence [37] (blue). The G-quadruplex and H-DNA-forming sequences overlap each other.