Published in final edited form as:

Methods. 2013 November ; 64(1): . doi:10.1016/j.ymeth.2013.08.004.

Methods to detect replication-dependent and replication-independent DNA structure-induced genetic instability

Guliang Wang¹, Sally Gaddis², and Karen M. Vasquez^{1,*}

¹Division 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.

²Department of Carcinogenesis, The University of Texas M.D. Anderson Cancer Center, Science Park-Research Division, Smithville, TX

Abstract

DNA can adopt a variety of alternative secondary (*i.e.*, non-B DNA) conformations that play important roles in cellular metabolism, including genetic instability, disease etiology, and evolution. While we still have much to learn, research in this field has expanded dramatically in the past decade. We have summarized in our previous *Methods* review (Wang *et al.*, *Methods*, 2009) some commonly used techniques to determine non-B DNA structural conformations and non-B DNA-induced genetic instability in prokaryotes and eukaryotes. Since that time, we and others have further characterized mechanisms involved in DNA structure-induced mutagenesis and have proposed both replication-dependent and replication-independent models. Thus, in this review, we highlight some current methodologies to identify DNA replication-related and replication-independent mutations occurring at non-B DNA regions to allow for a better understanding of the mechanisms underlying DNA structure-induced genetic instability. We also describe a new web-based search engine to identify potential intramolecular triplex (H-DNA) and left-handed Z-DNA-forming motifs in entire genomes or at selected sequences of interest.

Keywords

DNA structure; Genetic instability; Mutation; Replication; 2-D gel electrophoresis; Search engine

Under appropriate physiological conditions, more than 10 types of non-B DNA conformations, which differ in structure from the classic Watson and Crick B-DNA helix, can form in the genome [1-3]. Non-B DNA structures can form transiently during cellular metabolism, and depending on the type of DNA conformation and its location in the genome, these structures can impact DNA replication, gene transcription, recombination, DNA repair and genetic instability [4-8]. Factors such as negative supercoiling levels, pH, salt concentration, and the presence of DNA binding proteins are critical for the transition from B-DNA to non-B DNA. A critical factor required for the formation of non-B DNA structures is the sequence; the arrangement of the base-pairs dictates the potential for

© 2013 Elsevier Inc. 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.

interactions among different components in a DNA helix, and thus the ability to form non-B DNA structures [8]. Prediction and identification of sequences with the capacity to adopt non-B DNA structures in regions of interest is a useful first step in associating non-B DNA conformation with a phenotype. Here we introduce an easy-to-use, yet powerful search engine we recently developed to screen for those sequences that have the potential to form intramolecular triplexes (H-DNA) or left-handed Z-DNA structures. In addition, we describe techniques (*e.g.*, *in vitro* replication and mutagenesis assays, and 2-D agarose gel electrophoresis for replication fork stalling) to explore DNA replication-dependent and independent mechanisms of non-B DNA-induced mutations (see review [9]).

The purpose of this article (and our previous *Methods* review [10]) is to provide methodology options for DNA structure-related studies, with an emphasis on the application and limitation of each assay, rather than focusing on the details of each protocol. Combinations of these assays, and many other methods described elsewhere, have proven useful in advancing our understanding of DNA structure-induced genetic instability.

1. A web-based search engine to identify H-DNA-forming and Z-DNA-forming sequences

A search engine to identify sequences that are capable of forming non-B DNA structures is an important tool to further our understanding of their biological roles in cells. For example H-DNA forms at polypurine-polypyrimidine regions with mirror-repeat symmetry [11, 12] and Z-DNA forms in alternating purine/pyrimidine sequences [13]. These sequence requirements facilitate prediction and identification of potential non-B DNA-forming sequences in the genome. However, searching for potential non-B DNA-forming sequences can be tedious and potentially subject to error if performed manually.

We have recently developed a computer-based program that can search for potential H-DNA and Z-DNA-forming sequences in any given sequence in a text file or a provided Gene Name, Gene Symbol or Entrez Gene ID to retrieve the relevant nucleotide sequence. Sequences with the propensity to adopt H-DNA are **purine-** or **pyrimidine-rich** runs that have **mirror repeat** symmetry (PRMR) separated by a relatively short spacer of any composition of bases. The search algorithm begins by finding a purine run, part or all of which may become the 5' arm of a mirror repeat. The algorithm finds PRMR variations that are based on the initial purine run and groups these variations into a family. PRMR of the same family differ by the position of the spacer and/or overall length. PRMR that arise from adjacent purine runs are assigned to different families, yet may share an arm.

The initial search casts a broad net and thus initial search results can include PRMR with low likelihood of forming H-DNA. Post-search filtering is included as part of the algorithm to eliminate these sequences. The algorithm we have developed searches a sequence for PRMR and allows user-defined parameters to specify the minimum length of the mirrored arms, the minimum and maximum allowable spacer length, and the number of mismatches allowed in the arms. The algorithm can be found in the **Supplementary Data**.

The algorithm performs best when either zero or one mismatch in mirror symmetry is allowed. If more than one mismatch is allowed the program will allow the first mismatch in the innermost L bases of the symmetry arm where L is the minimum arm length specified by the user. Additional mismatches are allowed in the outer boundary of the symmetry arms. Finally, the user is given the option to omit mirror repeats consisting of simple GA repeats or simple A repeats with a G at the extremities.

Sequences with the propensity to adopt Z-DNA are alternating purine-pyrimidine sequences such as GC and GT repeats. AT sequences often have a greater propensity to form hairpins or loop structures rather than Z-DNA [14]. The algorithm we have developed searches a sequence for alternating purine/pyrimidine tracts and scores the tracts, giving each GC dinucleotide a higher score than each GT dinucleotide. A user-defined minimum score is used to filter the search results. Because AT dinucleotide repeats are more likely to form hairpins than Z-DNA, the algorithm excludes the AT and TA dinucleotides. The algorithm can be found in the **Supplementary Data**.

The search results page displays the search parameters and a table of the identified potential H-DNA or Z-DNA-forming sequences and their positions in the gene (sequence provided); each sequence is also given a score that represents the possibility/stability of each non-B DNA conformation. The search engine can be accessed at: <http://www.utexas.edu/pharmacy/dnastructure/>.

Here we have described a brief example of the output of the search engine. **Figure 1** displays the result of a search for H-DNA and Z-DNA-forming sequences in the human *c-MYC* gene. Two H-DNA-forming sequences were identified, both located in the promoter region. The first one on the list was previously described as an H-DNA-forming sequence by Mirkin et al. (1987) [12] and Kinniburgh (1989) [15], and was found to be mutagenic in mammalian cells and in mice in our previous studies [16, 17]. The Z-DNA-forming sequences identified by the program, shown in the lower part of **Figure 1**, had also been shown to form Z-DNA in previously published studies [18], providing evidence for the utility and predictive power of the search engine.

Several other computer-based on-line algorithms are available to search for non-B DNA-forming sequences. For example, “MFold” (Michael Zuker, Rensselaer Polytechnic Institute, U.S.A. <http://mfold.rna.albany.edu/?q=mfold>) and “pknotsRG” (Universität Bielefeld, Germany, <http://bibiserv.techfak.uni-bielefeld.de/pknotsrg/>) were designed to predict secondary structures formed on RNA or ssDNA, such as stem-loop structures. The “einverted” (<http://emboss.bioinformatics.nl/cgi-bin/emboss/einverted>) or “palindrome” (<http://emboss.bioinformatics.nl/cgi-bin/emboss/palindrome>) were developed to search for inverted repeats, and “equicktandem” (<http://emboss.bioinformatics.nl/cgi-bin/emboss/equicktandem>) can identify tandem repeats. “QGRS Mapper” is a software program designed to search for Quadruplex forming G-Rich Sequences (QGRS) (<http://bioinformatics.ramapo.edu/QGRS/index.php>), based on published algorithms for recognition and mapping Quadruplex forming G-Rich Sequences in transcribed regions of genes [19].

It is important to appreciate that a computer-based search program generally uses simplified rules for the identification of candidate non-B DNA-forming sequences, but that some sequences that do not conform to those simplified “rules” can also adopt non-B conformations. For example, some sequences that do not contain pure alternating purine-pyrimidine sequences have been shown to adopt Z-DNA structures *in vitro* [20, 21]; moreover, chromatin structure and the presence of DNA binding proteins can change the ability of a given sequence to form non-B DNA [22-26]. Thus, the result of any computer prediction can only be used as a preliminary pre-screening step. Non-B DNA structure-forming potential can be confirmed by employing several *in vitro* techniques, as we have previously described [10].

2. Methods for exploring DNA replication-dependent and replication-independent mechanisms of DNA structure-induced genetic instability

The mechanisms of non-B structure-induced genetic instability are complex and not yet fully elucidated, but it is clear that more than one mechanism is involved in processing various non-B DNA structures [9]. Depending on the type of DNA structure, the nearby cis elements, and the status of replication, transcription, or DNA repair within close proximity to the non-B DNA-forming sequences, different mechanisms may play different roles in processing non-B DNA.

2-1. Non-B DNA-mediated impediments to DNA replication

The formation of structures such as cruciforms, H-DNA, G4 DNA, or slipped DNA, requires the melting of the B-DNA duplex into single-strands. During replication the DNA duplex is denatured and single-stranded DNA (ssDNA) is exposed, allowing for the formation of non-B DNA structures, particularly during lagging strand synthesis [27]. Furthermore, DNA is unwrapped from histone cores during replication, which results in negative supercoiling of the duplex DNA, as required for the formation of many types of non-B DNA structures. DNA helicases associated with the replication machinery can resolve some types of non-B DNA structures formed in front of the progressing DNA polymerase, but not in all cases [7, 28]. If the non-B DNA structures are left unresolved, then they can cause impediments to DNA polymerases, resulting in replication fork collapse and DNA double-strand breaks (DSBs) [29]. Stalled replication forks can also result in extended exposure of ssDNA on either the template strand or the nascent strand, which could stimulate the formation of additional non-B DNA structures on these single-stranded regions, resulting in expansion or contraction events [30, 31]. Thus, many approaches to explore the effects of non-B DNA structures on DNA replication have been developed. We briefly summarize three of these approaches below.

1. *In vitro* replication assay—A simple approach to detect non-B DNA-based impediments to replication is to determine the efficiency of nascent DNA synthesis by measuring radiolabeled nucleotide incorporation into B-DNA versus that of non-B DNA [32]. Briefly, in cell-free extracts that contain SV40 large T antigen, a shuttle vector containing an SV40 replication origin can be replicated in the presence of [³²P]dNTP, dNTPs and an ATP-regenerating system. The test plasmid should contain a non-B DNA sequence of interest or a control B-DNA sequence and also the same amount of a second shuttle plasmid of a different size, but the same replication origin to serve as an internal control. Plasmids should be incubated in the extracts for an appropriate amount of time to allow for replication, purified from the cell extracts and linearized by restriction digestion, separated on an agarose gel, and then analyzed by use of a PhosphorImager. More details of the *in vitro* replication assay in cell extracts will be described below. The radioactive incorporation resulting from replication of the test plasmid containing a non-B DNA sequence (after being normalized by replication efficiency of the internal control plasmid) can be compared with that of the same plasmid, but one which contains a control B-DNA sequence. A caveat of this approach is that if replication is only paused temporarily and the replication machinery can resolve the non-B DNA structure before the second round of replication occurs on the control plasmid, then the differences in replication between the non-B and B-DNA plasmids may be indistinguishable [32].

2. DNA combing—Fluorescence microscopy-based DNA combing (fiber) technique allows visualization of DNA replication elongation in cell extracts or in cultured cells, and allows for direct measurements of the rate of movement of active replication forks. This technique stretches the otherwise randomly coiled DNA uniformly on a smooth surface,

which is suitable for nucleic acid hybridization, allowing for subsequent measurement of labeled segments [33]. Briefly, nascent DNA is pulse labeled by incorporation of two different halogenated nucleotides (*e.g.* CldU, IdU and/or BrdU) before the cells are lysed, and then individual fibers of DNA are stretched onto a microscope slide. The newly synthesized DNA with the halogenated nucleotides incorporated is stained with specific antibodies and the position of the replication origins and the rate of fork progression can be visualized by fluorescence microscopy [34, 35]. This approach works well for visualization and semi-quantification of replication fork movement impaired by long tracts (Mb in length) of non-B DNA-forming sequences [36]. A stretched DNA fiber of 1 μm in length is approximately equivalent to 2 kb. However, it is not sensitive enough to visualize subtle differences between short DNA segments (*e.g.* <200 kb).

3. Two-dimensional gel electrophoresis of DNA replication intermediates—

This technique can detect transient pausing of the DNA replication machinery at non-B DNA regions. The basic principle of this analysis is to use gel electrophoresis to separate DNA replication intermediates in a mixture of different molecules of different lengths and shapes (see below), by size in the first dimension, and by shape and size in the second dimension. This technique allows determination of the location of a replication-stalling site within the fragment under analysis. An advantage of this technique over methods to determine replication rates is that transient pausing of the replication machinery can be detected even when there is no obvious delay in the overall replication rate. In our studies of *in vitro* replication using HeLa cell extracts supplied with SV40 large T antigen (see above), the incorporation of ^{32}P -dCTP into a plasmid containing a Z-DNA-forming sequence and control B-DNA-forming plasmid were very similar [32], while the same Z-DNA-forming sequence represented a strong pausing site when probed by 2-D gel analysis (Wang and Vasquez, unpublished data). These results suggest that this Z-DNA-forming sequence can significantly impede replication forks, perhaps in a transient fashion that is not detectable by measuring the overall replication rate of a plasmid *in vivo* over multiple rounds of replication. Below is a brief introduction of the 2-D gel electrophoresis assay to determine DNA structure-induced replication stalling on a plasmid in mammalian cells.

For this example, we used a shuttle vector based on the pSP189 plasmid that contains an SV40 origin of replication and an SV40 T antigen-expressing gene that can initiate bi-directional replication of the plasmid in human cells [32]. Non-B DNA structure-forming sequences were cloned into the vector ~1-3 kb away from replication origin. If the fragment to be analyzed does not include the DNA replication origin, then it should result in a typical Y-shaped intermediate (**Figure 2A**). If a non-B DNA structure stalls the replication fork for an extended period of time, then the replication fork progresses into the fragment from the other direction, resulting in a double Y-shape (**Figure 2B**). If possible, the non-B DNA-forming sequence should be located within the first 1/3 of the fragment, avoiding the center, so that the stalling “bulge” will be detected on the right arm of the arc, and not the tip. If the fragment to be analyzed contains the replication origin, then the replication intermediates will contain two replication forks moving outward, resulting in bubble-shaped replication intermediates (**Figure 2C**) until one fork progresses outside of the fragment, which would lead to a Y-shaped structure. If a non-B DNA structure stalls a replication fork for an extended time frame, then this bubble to single Y transition will progress to a double Y-shape when the replication fork from the other direction progresses into the fragment (**Figure 2D**). Therefore, the choice of restriction sites determines the shape of the arc and the location of the stalling “bulge” in 2-D gel analysis.

Transfected plasmids should be purified from mammalian cells using a method that allows for recovery of short linear molecules, such as Hirt's method [37]. DNA should be digested

with *DpnI* to remove the plasmid DNA that was not replicated in the mammalian cells, to reduce the signal of the 1x linear DNA.

These nonlinear (bubble or Y-shape) DNA replication intermediates are then separated by two-dimensional gel electrophoresis. The first dimension is a conventional agarose gel electrophoresed at low voltage with no ethidium bromide added in the gel or buffer to separate the DNA predominately by size. The gel slices containing the appropriate DNA fragment (1x - 2x the size of restriction fragment to be analyzed) are placed in a perpendicular fashion on the second dimension, high concentration agarose gel, which contains ethidium bromide in the gel and buffer. By performing the gel electrophoresis at a higher voltage in the presence of ethidium bromide, the DNA molecules are separated not only by size, but also by their shape. The bulkier branched Y-shaped or bubble-shaped replication intermediates migrate slower than linear DNA of the same size in the second dimension gel. After electrophoresis, the linear restriction fragment and the replication intermediates can be visualized by standard Southern blotting using probes specific to the region of interest. A more detailed protocol can be found in references [38, 39].

Single Y-, double Y- and bubble-shaped intermediates can co-exist as a mixture, and can transition from one shape to another (for example, from a single Y to a double Y when another fork progresses into the region of interest, or form a bubble to a Y shape when one of the replication forks progresses beyond the region being probed), as shown in **Figures 2B and D**, which can confound interpretation of the results.

2-2. DNA replication-independent genetic instability caused by non-B DNA-forming sequences

Genetic instability induced by non-B DNA can occur in a variety of tissues regardless of proliferative status. However, some non-B DNA-forming sequences, such as hairpin-forming triplet repeats, are more unstable in highly proliferative tissues [40, 41] or in rapidly dividing cells [42, 43]. Further, the location of the repeat relative to the replication origin can have dramatic effects on the mutagenesis caused by non-B DNA [31, 44-49], implicating semi-conservative DNA replication in non-B DNA-induced genetic instability. However, non-B DNA-forming sequences can also induce mutations in tissues with low proliferative status or non-proliferating tissues or cells, implying the existence of chromosomal replication-independent mechanisms of DNA structure-related instability [50-53].

The association of non-B DNA-induced genetic instability with DNA replication has been inferred from the comparison of data from tissues or cells with different rates of proliferation. In an effort to examine the role of replication more directly, plasmids lacking compatible mammalian replication origins have been transfected into mammalian cells. While this approach can eliminate genomic DNA replication, it is possible that plasmid replication may occur even in the absence of a dedicated replication origin [54]. Because only a small fraction of the transfected plasmid DNA will actually enter the nuclei, measurements of replication-independent processes are diluted by unprocessed plasmid. For example, we have used a cell-free extract system to reestablish cellular functions in a test tube. For example, we used an SV40-based replication system in HeLa cell-free extracts to determine replication-dependent or replication-independent DNA structure-induced mutagenesis [10, 16, 32]. The shuttle vector contains an SV40 replication origin and a pBR327 origin, which support replication in SV40 transformed mammalian cells (expressing SV40 T antigen) and bacteria [10]. We have used both the *supF* gene and the *lacZ* gene as mutation reporters in the shuttle vectors. The *supF* gene encodes a suppressor tRNA that suppresses amber mutations in the - *galactosidase* gene in MBM7070 reporter bacterial cells and generates blue colonies; the *lacZ* gene expresses the amino-terminal

fragment of the *lacZ* gene product (- galactosidase) and also generates blue colonies, in compatible host bacteria such as DH5⁺, on plates containing 5-bromo-4-chloro-3-indolyl- -D-galactoside (X-Gal) and isopropyl -D-thiogalactoside (IPTG). For more details, see [10, 16, 32]. Briefly, depending on the activity of the extracts, 10-50 ng of plasmid DNA is incubated with 180 µg HeLa cell-free extract supplied in the reaction buffer (30 mM HEPES, pH 7.5, 7 mM MgCl₂, 0.5 mM dithiothreitol, 4 mM ATP, 100 mM each of dNTP, 50 mM each of NTP, 40 mM phosphocreatine, 0.625 units of creatine phosphokinase) at 37°C for 8 hours. Since HeLa cells do not express T antigen, the extracts will not support replication of the SV40 origin-containing plasmids. To permit the same plasmid to be tested under replication-proficient conditions, 1 µg of SV40 large T antigen (CHIMERX, Madison, WI) is supplied. Therefore, this system can provide comparable conditions to determine non-B DNA-induced genetic instability on plasmids in cell-free extracts in the presence or absence of DNA replication. After incubation, the plasmid DNA can be purified from cell extracts after RNase A and proteinase K digestion and phenol and chloroform extraction. The DNA can be subjected to reporter mutagenesis assays to characterize mutations, LM-PCR to detect DSBs near non-B DNA structures, or T4 polynucleotide kinase and [³²P] ATP to radiolabel the breakpoints of single-strand breaks (SSBs) or DSBs generated near non-B DNA [16, 32]. To verify the replication status of the plasmids in the cell-free extracts when SV40 T antigen is supplied, [³²P]dCTP can be added into the reaction to determine the efficiency of plasmid replication via nucleotide incorporation measurements. As an alternative approach, purified plasmid DNA can be digested with the *DpnI* restriction enzyme, which cleaves methylated GATC sites, a signature methylation modification specific to dam⁺ bacterial strains. Plasmids that are replicated in mammalian cell-free extracts should be resistant to *DpnI* digestion. Using this system, we have found that H-DNA-forming sequences and Z-DNA-forming sequences induced SSBs, DSBs, and large-scale deletions and rearrangements of plasmid DNA in both replication-deficient and replication-proficient HeLa cell extracts ([32] and our unpublished results), indicating the presence of both replication-independent and replication-dependent mechanisms of non-B DNA induced mutagenesis in HeLa cell extracts.

Technologies developed in recent years have provided very powerful tools to further advance our knowledge of this field. For example, we have recently applied next-generation sequencing to determine non-B DNA-induced genetic instability directly without screening or selection of a reporter gene, and have characterized Z-DNA-induced mutations in human cells. Data collected from these experiments could be very informative when evaluated together with those from other traditional methods.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Dr. Rick A. Finch for useful comments and discussion. Support was provided by an NIH/NCI grant to K.M.V. (CA93729).

REFERENCES

1. Choi J, Majima T. Chem. Soc. Rev. 2011; 40:5893–5909. [PubMed: 21901191]
2. Wells RD. J. Biol. Chem. 1988; 263:1095–1098. [PubMed: 3275663]
3. Wang G, Vasquez KM. Mutat. Res. 2006; 598:103–119. [PubMed: 16516932]
4. Zhao J, Bacolla A, Wang G, Vasquez KM. Cell. Mol. Life Sci. 2010; 67:43–62. [PubMed: 19727556]

5. Kim YS, Kang HS. *Nucleic Acids Res.* 1989; 17:9279–9289. [PubMed: 2555790]
6. Lin S, Kowalski D. *J. Mol. Biol.* 1994; 235:496–507. [PubMed: 8289278]
7. Vasquez KM, Wang G. *Mutat. Res.* 2012
8. Sinden RR, Wells RD. *Curr. Opin. Biotechnol.* 1992; 3:612–622. [PubMed: 1369117]
9. Wang G, Vasquez KM. *Mol. Carcinog.* 2009; 48:286–298. [PubMed: 19123200]
10. Wang G, Zhao J, Vasquez KM. *Methods.* 2009; 48:54–62. [PubMed: 19245837]
11. Htun H, Dahlberg JE. *Science.* 1988; 241:1791–1796. [PubMed: 3175620]
12. Mirkin SM, Lyamichev VI, Drushlyak KN, Dobrynin VN, Filippov SA, Frank-Kamenetskii MD. *Nature.* 1987; 330:495–497. [PubMed: 2825028]
13. Rich A, Nordheim A, Wang AH. *Annu. Rev. Biochem.* 1984; 53:791–846. [PubMed: 6383204]
14. Wang AH, Hakoshima T, van der Marel G, van Boom JH, Rich A. *Cell.* 1984; 37:321–331. [PubMed: 6722876]
15. Kinniburgh AJ. *Nucleic Acids Res.* 1989; 17:7771–7778. [PubMed: 2678005]
16. Wang G, Vasquez KM. *Proc. Natl. Acad. Sci. U. S. A.* 2004; 101:13448–13453. [PubMed: 15342911]
17. Wang G, Carbajal S, Vijg J, DiGiovanni J, Vasquez KM, Natl J. *Cancer Inst.* 2008; 100:1815–1817.
18. Wittig B, Wolf S, Dorbic T, Vahrson W, Rich A. *EMBO J.* 1992; 11:4653–4663. [PubMed: 1330542]
19. Kikin O, D'Antonio L, Bagga PS. *Nucleic Acids Res.* 2006; 34:W676–682. [PubMed: 16845096]
20. Feigon J, Wang AH, van der Marel GA, van Boom JH, Rich A. *Science.* 1985; 230:82–84. [PubMed: 4035359]
21. Eichman BF, Schroth GP, Basham BE, Ho PS. *Nucleic Acids Res.* 1999; 27:543–550. [PubMed: 9862978]
22. Rothenburg S, Koch-Nolte F, Haag F. *Immunol. Rev.* 2001; 184:286–298. [PubMed: 12086319]
23. Tippin DB, Ramakrishnan B, Sundaralingam M. *J. Mol. Biol.* 1997; 270:247–258. [PubMed: 9236126]
24. Zacharias W, Jaworski A, Wells RD. *Bacteriol J.* 1990; 172:3278–3283.
25. Koeris M, Funke L, Shrestha J, Rich A, Maas S. *Nucleic Acids Res.* 2005; 33:5362–5370. [PubMed: 16177183]
26. Herbert A, Schade M, Lowenhaupt K, Alfken J, Schwartz T, Shlyakhtenko LS, Lyubchenko YL, Rich A. *Nucleic Acids Res.* 1998; 26:3486–3493. [PubMed: 9671809]
27. Pearson CE, Sinden RR. *Curr. Opin. Struct. Biol.* 1998; 8:321–330. [PubMed: 9666328]
28. Peleg M, Kopel V, Borowiec JA, Manor H. *Nucleic Acids Res.* 1995; 23:1292–1299. [PubMed: 7753619]
29. Betous R, Rey L, Wang G, Pillaire MJ, Puget N, Selves J, Biard DS, Shinya K, Vasquez KM, Cazaux C, Hoffmann JS. *Mol. Carcinog.* 2009; 48:369–378. [PubMed: 19117014]
30. Krasilnikova MM, Mirkin SM. *Mol. Cell. Biol.* 2004; 24:2286–2295. [PubMed: 14993268]
31. Samadashwily GM, Raca G, Mirkin SM. *Nat. Genet.* 1997; 17:298–304. [PubMed: 9354793]
32. Wang G, Christensen LA, Vasquez KM. *Proc. Natl. Acad. Sci. U. S. A.* 2006; 103:2677–2682. [PubMed: 16473937]
33. Bensimon A, Simon A, Chiffaudel A, Croquette V, Heslot F, Bensimon D. *Science.* 1994; 265:2096–2098. [PubMed: 7522347]
34. Schwab RA, Blackford AN, Niedzwiedz W. *The EMBO journal.* 2010; 29:806–818. [PubMed: 20057355]
35. Pillaire MJ, Betous R, Conti C, Czaplicki J, Pasero P, Bensimon A, Cazaux C, Hoffmann JS. *Cell Cycle.* 2007; 6:471–477. [PubMed: 17329970]
36. Schurra C, Bensimon A. *Methods Mol. Biol.* 2009; 464:71–90. [PubMed: 18951180]
37. Hirt B. *J. Mol. Biol.* 1967; 26:365–369. [PubMed: 4291934]
38. Krasilnikova MM, Mirkin SM. *Methods Mol. Biol.* 2004; 277:19–28. [PubMed: 15201446]
39. Kohwi, Y. *Trinucleotide repeat protocols.* Springer; New York: 2013.

40. Martorell L, Monckton DG, Gamez J, Johnson KJ, Gich I, Lopez de Munain A, Baiget M. Hum. Mol. Genet. 1998; 7:307–312. [PubMed: 9425239]
41. Wong LJ, Ashizawa T, Monckton DG, Caskey CT, Richards CS. Am. J. Hum. Genet. 1995; 56:114–122. [PubMed: 7825566]
42. Martorell L, Martinez JM, Carey N, Johnson K, Baiget M. J. Med. Genet. 1995; 32:593–596. [PubMed: 7473648]
43. Wohrle D, Kennerknecht I, Wolf M, Enders H, Schwemmle S, Steinbach P. Hum. Mol. Genet. 1995; 4:1147–1153. [PubMed: 8528201]
44. Kang S, Jaworski A, Ohshima K, Wells RD. Nat. Genet. 1995; 10:213–218. [PubMed: 7663518]
45. Miret JJ, Pessoa-Brandao L, Lahue RS. Proc. Natl. Acad. Sci. U. S. A. 1998; 95:12438–12443. [PubMed: 9770504]
46. Trinh TQ, Sinden RR. Nature. 1991; 352:544–547. [PubMed: 1865910]
47. Hashem VI, Sinden RR. Mutat. Res. 2005; 570:215–226. [PubMed: 15708580]
48. Iyer RR, Wells RD. J. Biol. Chem. 1999; 274:3865–3877. [PubMed: 9920942]
49. Pelletier R, Krasilnikova MM, Samadashwily GM, Lahue R, Mirkin SM. Mol. Cell. Biol. 2003; 23:1349–1357. [PubMed: 12556494]
50. Chong SS, McCall AE, Cota J, Subramony SH, Orr HT, Hughes MR, Zoghbi HY. Nat. Genet. 1995; 10:344–350. [PubMed: 7670474]
51. Telenius H, Kremer B, Goldberg YP, Theilmann J, Andrew SE, Zeisler J, Adam S, Greenberg C, Ives EJ, Clarke LA, et al. Nat. Genet. 1994; 6:409–414. [PubMed: 8054984]
52. Hashida H, Goto J, Kurisaki H, Mizusawa H, Kanazawa I. Ann. Neurol. 1997; 41:505–511. [PubMed: 9124808]
53. Gomes-Pereira M, Fortune MT, Monckton DG. Hum. Mol. Genet. 2001; 10:845–854. [PubMed: 11285250]
54. Vashee S, Cvetcic C, Lu W, Simancek P, Kelly TJ, Walter JC. Genes Dev. 2003; 17:1894–1908. [PubMed: 12897055]

DNA Structure Search

THE UNIVERSITY OF TEXAS AT AUSTIN

About

Search

Results

Gene information

Species	Gene Name	Entrez Gene ID	Alternate Symbols	Description	Chromosome	Transcription Direction
Hs	MYC	4609	MRTL, MYCC, bHLHe39, c-Myc	v-myc avian myelocytomatosis viral oncogene homolog	8	forward

Transcript information¹

Transcript (link is to NCBI GenBank)	Transcript Length	Location on Chromosome
NM_002467.4	5366	128748315-128753680

Mirror Repeat Search Parameters and Results

Minimum Arm Length	Minimum Spacer Length	Maximum Spacer Length	Allowable Number of Mismatches	Minimum Arm Length for Mismatches to be Allowed	Minimum %G in Arms	Omit Simple Repeats	Putative Promoter Region Length	Number of Mirror Repeat Families Found
7	3	10	1	15	0	A, GA	2000	2

Mirror repeats

Mirror Repeat ²	Family ³	Length	% G	Strand	Start ⁴ in Chromosome	Start ⁵ in Gene	Region ⁵
GGGAGGGGCGCTTATGGGAGGG	2	23	86.7	reverse	128748214	-100	putative-promoter
AGGGAGAGGGTTTGGAGGGA	1	21	66.7	forward	128748049	-265	putative-promoter

Alternating Purine/Pyrimidine Tract Search Parameters and Results

Minimum Score	Putative Promoter Region Length	Number of Alternating Purine/Pyrimidine Tracts Found
75	2000	6

Alternating Purine/Pyrimidine Tracts

Alternating Purine/Pyrimidine Tract ²	Length	Score	Start ⁴ in Chromosome	Start ⁵ in Gene	Region ⁵
GCGCGCA	7	78.0	128747936	-378	putative-promoter
GCGCGCGT	8	78.0	128748087	-227	putative-promoter
GCGCGCAC	8	78.0	128750410	2096	intron_1
TGCGCGCA	8	78.0	128750970	2656	exon_2
ACGCGCGC	8	75.0	128747814	-500	putative-promoter
TGCGCGC	7	75.0	128750418	2104	intron_1

Figure 1. Results of a search for H-DNA- and Z-DNA-forming sequences in the human *c-Myc* gene

Results reveal two H-DNA-forming sequences in the promoter region of the human *c-MYC* gene. Mirror-repeat polypurine sequences are shown in black and the linker region between the two mirror-repeat sequences is shown in orange. Multiple Z-DNA-forming sequences were also identified by the search engine.

A

Figure 2. Two-dimensional gel electrophoresis of DNA replication intermediates for detecting non-B DNA-mediated impediments to replication

A. Schematic diagram of the shuttle plasmid containing the SV40 replication origin and the non-B DNA structure-forming sequence. The restriction sites and the probes for analysis are labeled on the schematic. The replication origin is not included in the restriction fragment to be probed, such that a stalled replication fork by the non-B DNA structure (red octagon) results in the single-Y shaped arc (black bulge). **B.** If a replication fork is stalled for an extended period of time, then the replication fork from the other direction will also be located within the restriction fragment, resulting in a double Y arc. **C.** The replication origin is included in the restriction fragment to be probed such that the replication intermediate is

initiated with a bubble shape. The schematic diagram depicts the transient stalling of a replication fork by the non-B DNA structure (red octagon) within the bubble arc (black bulge). The broken line in grey shows the position of the relative Y-shape arc. **D.** If a replication fork is stalled for an extended period of time, or if the replication origin is not centrally located within the fragment, then one fork will progress beyond the restriction fragment and the replication intermediates become a single Y shape. If the replication fork from the opposite direction progresses into the restriction fragment, then this results in a double Y arc. Adapted from Krasilnikova *et al* [38].