Methods 64 (2013) 67–72

Contents lists available at ScienceDirect

Methods

journal homepage: www.elsevier.com/locate/ymeth

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

Guliang Wang [a], Sally Gaddis [b], Karen M. Vasquez [a,*]

[a] Division of Pharmacology and Toxicology, College of Pharmacy, The University of Texas at Austin, Dell Pediatric Research Institute, 1400 Barbara Jordan Blvd. R1800, Austin, TX 78723, United States
[b] Department of Molecular Carcinogenesis, The University of Texas, MD Anderson Cancer Center, Smithville, TX, United States

ARTICLE INFO

Article history:
Available online 15 August 2013

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

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.

1. Introduction

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 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.

2. 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.

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

1046-2023/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ymeth.2013.08.004

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. Fig. 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. [12] and Kinniburgh [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 Fig. 1, had also been shown to form Z-DNA in a previously published study [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].

3. 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.

3.1. 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.

3.1.1. 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