Secondary structure formation and DNA instability at fragile site FRA16B

Allison A. Burrow, Allison Marullo, Lindsay R. Holder and Yuh-Hwa Wang*

Department of Biochemistry, Wake Forest University School of Medicine, Medical Center Boulevard, Winston-Salem, NC 27157-1016, USA

Received October 13, 2009; Revised December 22, 2009; Accepted December 27, 2009

ABSTRACT

Human chromosomal fragile sites are specific loci that are especially susceptible to DNA breakage following conditions of partial replication stress. They often are found in genes involved in tumorigenesis and map to over half of all known cancer-specific recurrent translocation breakpoints. While their molecular basis remains elusive, most fragile DNAs contain AT-rich flexibility islands predicted to form stable secondary structures. To understand the mechanism of fragile site instability, we examined the contribution of secondary structure formation to breakage at FRA16B. Here, we show that FRA16B forms an alternative DNA structure in vitro. During replication in human cells, FRA16B exhibited reduced replication efficiency and expansions and deletions, depending on replication orientation and distance from the origin. Furthermore, the examination of a FRA16B replication fork template demonstrated that the majority of the constructs contained DNA polymerase paused within the FRA16B sequence, and among the molecules, which completed DNA synthesis, 81% of them underwent fork reversal. These results strongly suggest that the secondary-structure-forming ability of FRA16B contributes to its fragility by stalling DNA replication, and this mechanism may be shared among other fragile DNAs.

INTRODUCTION

Fragile sites are specific chromosomal regions located throughout the human genome that are especially susceptible to DNA breakage. These regions are defined cytogenetically as gaps or breaks on metaphase chromosomes following conditions of partial replication stress. Fragile sites are divided into two major classes based on their frequency in the population and are subdivided according to their mode of induction in cultured cells. Rare fragile sites are found in <5% of the population and are inherited in a Mendelian manner (1,2). The majority of rare fragile sites can be induced under folate-deficient conditions and contain a microsatellite (CGG)n repeat (3), whereas the rare, non-folate-sensitive sites are comprised of an AT-rich minisatellite element (2). In contrast, common fragile sites have been observed in all individuals and are believed to represent a normal component of chromosome structure (4). Most common fragile sites are observed after exposure to low doses of aphidicolin, an inhibitor of DNA polymerases α, δ and ε (5,6). To date, over 80 common fragile sites are listed in the Human Genome Database (GDB). Most have not yet been investigated at the molecular level, but it is known that regions of fragility can extend over megabases of DNA with gaps or breaks occurring throughout (7). Although a consensus sequence has not yet been identified among common fragile sites, the DNAs examined thus far contain frequent, AT-rich flexibility islands capable of forming secondary structures that are much more stable compared to other regions of the genome (8), similar to what has been reported for most rare sites.

Fragile sites are normally stable in cultured cells. However, these sites are hotspots for sister chromatid exchanges, deletions and rearrangements after induction with replication inhibitors (9,10). Moreover, many fragile sites are frequently associated with sites of chromosomal breakage in tumors (11,12). While the exact mechanism of fragile site expression remains elusive, replication timing experiments have shown that all fragile sites studied to date, including FRAXA (13), FRA3B (14), FRA7H (15), FRA10B (16), FRA16B (16), FRA1H (17) and FRA2G (17) exhibit delayed replication. The delay is further exacerbated by the addition of replication inhibitors, with some fragile site alleles remaining unreplicated in late G2 phase (14,15). Although it is not entirely clear how delayed replication at fragile sites results in chromosome breakage, evidence suggests that DNA sequences with the potential to form stable secondary structures can present significant difficulties during replication,
which may lead to unreplicated regions of the genome that are visible as gaps and breaks during metaphase (18). The (CGG)_n repeat within rare, folate-sensitive sites has been shown to form hairpin (19) and quadruplex structures (20) that present a significant block to replication both in vitro and in vivo (21, 22), whereas a polymorphic AT-rich sequence with the ability of forming a cruciform within common fragile site FRA16D blocked replication in yeast, resulting in increased chromosome breakage (23).

Several studies have demonstrated a critical role for the Ataxia-Telangiectasia and Rad3-Related (ATR)-dependent DNA damage checkpoint pathway in the maintenance of fragile sites. Although their direct roles remain unclear, proteins including the S-phase and G_2/M checkpoint kinase ATR (18), as well as its downstream targets BRCA1 (24) and CHK1 (25), are required for fragile site stability, as their deficiencies result in significantly increased fragile site breakage. ATR is a major component of the checkpoint pathway where it functions by sensing and responding to DNA damage, including stalled and collapsed replication forks (26, 27). Based on this evidence, it is hypothesized that ATR maintains fragile site stability by sensing and binding to single-stranded DNA resulting from stalled replication forks at sites of secondary structure formation (18), and that a deficiency or defect in ATR can prevent repair, leading to increased fragile site expression. Supporting this hypothesis, cells from patients with Seckel syndrome, who have low levels of ATR protein, show increased instability at fragile sites compared to normal cells following replication stress (28). This difference in instability suggests that the low level of ATR present in Seckel syndrome patients is not sufficient to respond to replication stress, consistent with the hypothesis that the breaks observed on metaphase chromosomes are unreplicated fragile sites that escape the ATR checkpoint (18).

To further understand the mechanism of breakage at human chromosomal fragile sites, we have investigated the cause of instability at fragile site FRA16B (29). Spontaneous FRA16B expression has been observed among individuals and can be induced by chemicals that bind AT-rich DNA, such as distamycin A (30) or berenil (31). Studies of FRA16B, located at 16q22.1, have shown that it spans the same genomic region as the common fragile site FRA16C and is also apparent following treatment with aphidicolin (8). After induction, the heterozygote frequency of FRA16B is about 5% in populations of European descent, representing the most frequently expressed rare fragile site (32). Positional cloning has revealed that FRA16B-expressing chromosomes may contain up to 2000 copies of a 33-bp AT-rich minisatellite repeat (ATAATTATATATTATACTAATATATATC/ATA), whereas normal chromosomes consist of only 7–12 copies of the repetitive element (33). Competitive nucleosome reconstitution assays demonstrate that, in the presence of distamycin A, FRA16B DNA displays nucleosome exclusion (29, 34) that increases in proportion to the number of repeats (29, 34). These results, along with similar studies on various other fragile sites (35–37), suggest a common feature for the chromatin structure surrounding fragile DNAs, which may play an important role in their expression. Therefore, FRA16B serves as a model for the examination of fragile sites, since it exhibits characteristics of both common and rare fragile sites, and has been mapped to cancer-specific rearrangements (12).

CGG repeats underlying the basis of fragility at rare, folate-sensitive fragile sites have been shown to form a secondary structure in vitro (38), but there is no physical evidence of alternative DNA structures for AT-rich fragile DNA, which comprises the majority of fragile sites. Therefore, it is important to determine whether these sequences are capable of forming secondary structures. It is also critical to investigate the contribution of secondary structure formation on chromosome instability. Recently, Ragland et al. (39) demonstrated the ability of common fragile DNAs to induce chromosome instability at ectopic sites in HCT116 cells, although the relationship between common fragile site instability and secondary structure was not investigated. In addition, the role of secondary structure formation in chromosome instability has been determined in yeast (23) but has not yet been investigated in human cells.

In this study, the ability of FRA16B to form a secondary structure in vitro was demonstrated by a reduction of electrophoretic mobility in polyacrylamide gels and visualization of short, branched structures by electron microscopy (EM). It was also determined that both replication orientation and distance from FRA16B to the origin affect FRA16B replication efficiency and instability in human cells. Furthermore, examination of FRA16B replication fork templates by EM revealed a tendency for FRA16B DNA to promote spontaneous fork reversal and an even greater occurrence of polymerase pausing at specific sites within the FRA16B region, which was confirmed by DNA sequencing gels. Overall, these results strongly suggest that the secondary structure-forming ability of FRA16B contributes to its fragility by stalling DNA replication, and this mechanism may be shared among other fragile DNAs.

**MATERIALS AND METHODS**

**Plasmids**

FRA16B-containing plasmids were created using genomic DNA from an individual expressing the FRA16B chromosome. A 921-bp fragment of FRA16B consisting of 14 perfect copies of the 33-bp AT-rich minisatellite repeat and some imperfect repeats with AT-rich flanking sequences was cloned in (34) to generate pFRA16B18.

To construct SV40 replication templates, the FRA16B fragment described above and the SV40 origin of replication (ori) (40) were cloned into pGEM3zf(+) (Promega). The FRA16B sequence was inserted at one of four different distances relative to the ori, and in one of two possible replication orientations. This generated eight constructs with the FRA16B sequence located 30, 300, 400 or 700 bp from the origin of replication in two different replication orientations (Figure 2A). A plasmid containing the SV40 ori only with no FRA16B sequence (pGEM-SV40ori) served as a control.
To construct FRA16B replication fork templates, complimentary oligonucleotides (5'-AGCTTGAGCTGCCCAGGGCTGAGGACA3' and 5'-CCTACGCTTGAGGCCATGCA-3') containing a site for the nicking endonuclease Nb.BbvCI (New England Biolabs) were annealed and cloned into the EcoRI and HindIII sites of pFRA16B18 to create pFRA16B37 (Supplementary Figure S1). The FRA16B DNA-containing sequence of pFRA16B37, and the Nb.BbvCI recognition site are also indicated in Supplementary Figure S1. The construction of these plasmids will allow us to examine DNA polymerase stalling, and replication fork regression of a 487-bp fragment of FRA16B DNA (see below).

Reduplexing assay

Reduplexing reactions were performed as previously described in (38,41,42). Briefly, the FRA16B fragment from pFRA16B18 was obtained by restriction enzyme digest using EcoRI and HindIII (New England Biolabs), and gel-purified. The FRA16B fragment was then dephosphorylated at the 5' end with calf intestinal alkaline phosphatase (New England Biolabs) and end-labeled with [α-32P] ATP (PerkinElmer) using T4 kinase (New England Biolabs). End-labeled DNA (1 ng) was added to solutions of 500 μl containing 1, 0.58, 0.3 or 0.1 M NaCl in TE buffer and incubated at 95°C for 5 min. The samples were cooled to room temperature, and the DNA was ethanol-precipitated in the presence of glycogen (Roche). The DNA pellets were then air-dried and resuspended in TE buffer. As a control, pGEM3zf(+) was digested with EcoRI and Eco109I (New England Biolabs) to generate an 892-bp fragment, and was subjected to the same conditions as FRA16B. DNA samples were electrophoresed in a 4% polyacrylamide gel cast in TBE at 50 V for 18 h at room temperature. The gel was dried and visualized by phosphorimaging (GE Healthcare).

EM

Reduplexed DNAs or replication fork template reaction mixtures were directly mounted onto glow-charged carbon-coated copper EM grids followed by washing in a water/ethanol gradient and rotary shadowcasting with tungsten, as described previously (43). DNAs were visualized on a Philips transmission electron microscope 400. Measurements of the DNA lengths were determined using Image J software (NIH).

Replication efficiency

The human embryonic kidney HEK293T cells were grown in Dulbecco’s modified Eagle’s medium (Gibco) containing 10% fetal bovine serum (HyClone) and 1% penicillin streptomycin (Gibco) on 100-mm-diameter plates. The cells were 50% confluent when co-transfected with 2.5 μg of pGEM-SV40ori and 2.5 μg of each FRA16B/SV40 plasmid using the CaPO4 method. The cells were allowed to grow for 16 h following transfection before replacing the media. Low-molecular-weight DNA was extracted by the Hirt’s lysis method (44) 48 h after transfection. The SV40 viral DNAs were further purified by incubation with proteinase K, phenol/chloroform extraction and alcohol precipitation. To determine replication efficiency of the constructs (Figure 2B), SV40-replicated DNAs were digested with HindIII and NdeI (New England Biolabs) to linearize the plasmids, and with DpnI (New England Biolabs) to remove unreplicated parental templates. For clone 46, EcoRI was used in place of HindIII. Southern blot analysis was then used to identify replicated DNAs using an [α-32P] dCTP-labeled probe hybridizing to nucleotide numbers 1725–2132 of pGEM3zf(+), which is present in all nine constructs. Replication efficiency was determined by the ratio of replicated FRA16B DNA to the control (pGEM-SV40ori) using ImageQuant version 5.2 to measure the intensity of each band. Student’s t-test was performed to determine statistically significant differences between clones.

Mutation assay

To investigate instability associated with FRA16B, a modified version of the stability of trinucleotide repeat by individual product (STRIP) assay (45) was used to examine individual products following replication in human cells. Essentially, products of replication from transfected HEK293T cells were digested with DpnI to eliminate any unreplicated parental templates and transformed into SURE-2 cells (Stratagene) (Figure 2B). Approximately 60 single colonies from each clone were picked at random, and the DNAs were isolated and digested with appropriate restriction enzymes to release the FRA16B insert. Samples were then run on 1.5% agarose gels and scored for insertion or deletion events, characterized by slower or faster migrating bands, respectively, compared to unreplicated DNA. As an additional control to measure the background instability in Escherichia coli, DNAs that were not transfected into HEK293T cells were directly transformed into SURE-2 cells, and scored for instability events. Fisher’s exact test was used to determine the statistical significance of instability resulting from replication in HEK293T cells compared to the background instability in E. coli, as well as differences between FRA16B-containing constructs compared to the control.

Analysis of FRA16B synthesis by the Klenow fragment of DNA polymerase I

A DNA template (pFRA16B37) was synthesized which mimics a replication fork that has progressed through clone 17 FRA16B sequence. The plasmid was digested with Nb.BbvCl to generate a nick directly in front of the FRA16B fragment that does not contain any cytosine bases for a length of 487 bp (Supplementary Figure S1). The nicked DNA was then incubated with 5 U of the Klenow fragment (exo-) of DNA polymerase I (New England Biolabs) in a reaction mixture containing 75 mM Tris, pH 7.5, 5 mM MgCl2 and 5 mM DTT with 2.5 mM each dATP, dTTP and dGTP nucleotides. The absence of cytosine caused DNA synthesis to stop at the end of the repetitive tract when the first guanine was
encountered on the template strand. This generated a single-strand tail by strand displacement. This resulted in a duplex circle with a single-strand tail where the tail represents the lagging strand, one arm of the circle is the leading strand and the other arm serves as the template (Figure 3A). To examine DNA synthesis on the lagging strand, a 50-fold molar excess of the complementary oligonucleotide (5′-ATATAATA-TATTATTATATCTA ATA-3′) was annealed to the lagging strand at the 3′ end for 30 min at 37 °C in a buffer containing 7.5 mM Tris, pH 7.5, 5 mM MgCl₂ and 5 mM DTT with 2.5 mM each dATP, dTTP and dCTP nucleotides, and further incubation with 5 U of Klenow fragment (exo-) for 30 min at 37 °C. The reaction mixture was purified and examined by EM for DNA synthesis. To serve as a control, a plasmid containing non-fragile site DNA (pGLGAP) was kindly provided by Dr. Jack Griffith (UNC-CH) and constructed using the same conditions as the FRA16B replication fork template (46).

To detect the pause sites on the lagging strand, the same oligonucleotide was radiolabeled and annealed to the lagging strand in the same reaction conditions as described above. DNA synthesis was carried out using 0.5, 5 and 15 U of Klenow fragment (exo-). The reaction mixtures were purified, resuspended in formamide/dye solution, electrophoresed through 12% polyacrylamide-7 M urea gels, and examined by phosphorimager analysis.

To locate the position of the replication fork for detecting fork regression, the reaction mixtures were linearized with AhdI or XmnI (New England Biolabs), to produce asymmetrical arms for measuring the distance from the fork junction to the DNA ends, and the samples were examined by EM.

RESULTS

Secondary structure formation of FRA16B DNA

Since there has been no physical evidence of alternative DNA structures formed for AT-rich fragile DNAs, we investigated secondary structure formation by subjecting a 921-bp fragment of FRA16B to reduplexing in the presence of various concentrations of NaCl to allow re-annealing of the single strands following denaturation. Separation by polyacrylamide gel electrophoresis (PAGE) showed that the reduplicated FRA16B fragment gave rise to two slower-migrating products over a range of NaCl concentrations (Figure 1A). These products were not present in the untreated FRA16B sample or the reduplicated pGEM3zf(+) control, suggesting the formation of a secondary structure during reduplexing of FRA16B DNA. Furthermore, when FRA16B was denatured and only one strand of the duplex was labeled with ³²P, the labeled strand corresponded uniquely to one of the bands with reduced electrophoretic mobility, demonstrating that each reduplicated band is produced from one of the two strands of FRA16B (Supplementary Figure S2). The examination of these molecules by EM confirmed that FRA16B folded into branched duplexes following denaturation and re-annealing (Figure 1B). These structures were not observed in the untreated FRA16B or reduplicated pGEM3zf(+) samples. The lengths of 100 DNA molecules from each sample were measured, and the comparison revealed that the reduplicated FRA16B molecules were shorter than the untreated FRA16B DNAs, indicating the presence of slippaged-out regions participating in the formation of secondary structure. These data are the first to demonstrate that FRA16B is indeed able to form an alternative DNA structure in vitro.

Reduced replication efficiency and increased instability of FRA16B in human cells

To analyze FRA16B replication in human cells, an SV40 system was used to evaluate replication efficiency and instability of FRA16B-containing constructs. Since several studies have shown that the ability of a sequence to inhibit replication depends on cis-acting factors, including replication orientation and distance relative to the origin (22,23,45,47–49), the 921-bp FRA16B insert was placed at varying distances from the SV40 ori and in one of two replication orientations to generate eight constructs containing FRA16B (Figure 2A). A control plasmid that does not contain FRA16B DNA was also constructed. An equal amount of the control plasmid and each FRA16B-containing plasmid was co-transfected into HEK293T cells (Figure 2B). Following SV40 replication, the DNAs were linearized and digested with DpnI to remove unreplicated parental templates. Replication efficiency was determined on Southern blots as the ratio of completely replicated FRA16B DNA to the control (Figure 2C). Clone 17 exhibited a statistically significant reduction in replication efficiency compared to clone 15 (P = 0.045), indicating that orientation affects replication efficiency. The efficiency of replication for clone 17 was also reduced when compared to clone 34 (P = 0.073), clone 10 (P = 0.003), clone 12 (P = 0.005) and clone 46 (P = 0.083). Our results demonstrate that both the replication orientation and distance from the origin affect FRA16B replication.

Next, a mutation assay was performed to determine if instability events were occurring within FRA16B during replication in HEK293T cells. Individual replication products were examined using a modified version of the STRIP assay (45), originally designed to determine the role of mammalian replication on the stability of triplet repeats. Replication products from transfected HEK293T cells were digested with DpnI and transformed into E. coli (Figure 2B). Only DpnI-resistant material yielded colonies, from which DNAs were isolated and digested with restriction enzymes to release the FRA16B insert. FRA16B inserts were resolved on agarose gels and scored for insertions or deletions (Figure 2D). Due to the resolution of agarose gels, mutations with only large size changes (≥50 bp) were detected. The mutation frequencies for each construct is shown in Figure 2E. Interestingly, clone 17, which displayed the greatest reduction in replication efficiency, also demonstrated a statistically significant increase in instability events compared to the background instability in E. coli (P = 0.022) and the control construct without FRA16B DNA in HEK293T.
cells ($P = 0.016$). Replication of clone 17 in HEK293T cells resulted in the highest mutation rate ($9\%$) among all of the FRA16B constructs, with four deletions (ranging from 99 to 378 bp) and one expansion (33 bp) (Figure 2F), which were confirmed by DNA sequencing. The 378-bp deletion in mutant 17-4 and the 165-bp deletion in mutant 17-18 begin, respectively, at nt 58 and nt 207 of the lagging strand template of the FRA16B sequence, located at the first Okazaki initiation zone (OIZ). The 99-bp deletion in mutant 17-8 and the 224-bp deletion in mutant 17-34 begin, respectively, at nt 364 and nt 406 of the lagging strand template of the FRA16B sequence, located at the second OIZ. The 33-bp insertion in mutant 17-48 begins at nt 325, and is one perfect copy of the 33-bp FRA16B repeat. Mapping of the mutation sites to the most energetically favorable predicted DNA secondary structures of the lagging strand template of the FRA16B sequence, as determined by Mfold (50) suggests that the deleted regions may correspond to sites of extensive secondary structure formation, including multiple hairpins (Supplementary Figure S3).

DNA polymerase stalling at FRA16B during DNA synthesis

Evidence has shown that the formation of secondary structures by CGG repeats can block replication within...
Figure 2. Analysis of FRA16B replication efficiency and instability in human cells using an SV40 replication system. (A) Diagram of replication constructs used for replication efficiency and instability assays. The bidirectional SV40 ori (open circles) was cloned into pGEM3zf(+) to generate pGEM-SV40ori, used as a control. The 921-bp FRA16B fragment (shaded box) was inserted 30, 300, 400 or 700 bp from the SV40 ori, and in one of two replication orientations (represented by arrows). (B) Schematic of strategy used to determine the replication efficiency of FRA16B constructs and instability events following replication in HEK293T cells. To determine replication efficiency, equal amounts of pGEM-SV40ori and each FRA16B-containing construct were co-transfected into HEK293T cells. After replication for 48 h, SV40 DNAs were extracted, purified and digested with DpnI to remove any unreplicated templates. Replicated molecules were then subjected to Southern blot analysis. To examine FRA16B-associated instability, each FRA16B-containing construct was transfected and replicated in HEK293T cells. SV40 DNA was extracted, purified and digested with DpnI to remove any unreplicated templates. These DNAs were transformed into E. coli, isolated, and digested with restriction enzymes to release the FRA16B insert. Replication products were analyzed on 1.5% agarose gels to score insertions and deletions. As a control, replication constructs were directly transformed into E. coli, without replication in human cells, to account for background instability in E. coli. (C) Replication efficiency of FRA16B constructs in HEK293T cells. The replication efficiency of each FRA16B-containing construct was determined by the ratio of replicated FRA16B DNA to the control (pGEM-SV40ori). Results are shown as the mean ± SD from six individual experiments. (D) Identification of mutation events of replication in HEK293T cells using gel electrophoresis. A representative agarose gel shows (continued)
rare, folate-sensitive fragile sites (21,22), and that replica-
tion forks have a high tendency to regress while moving
through difficult-to-replicate DNA sequences (30,51). To
further investigate these possible mechanisms of instabil-
ity, we constructed replication fork templates
pFRA16B37, a plasmid containing FRA16B that
mimicked clone 17, and pGLGAP, a control containing
non-fragile site DNA (Figure 3A). Following DNA syn-
thesis by Klenow fragment, circular pFRA16B37 and
pGLGAP molecules were examined by EM (Figure 3C,
bottom panels). We found that 73% of pGLGAP molecules
had double-stranded tails, indicating that they underwent
completed DNA synthesis, whereas 24% did not (Figure
3B). In contrast, only 10% of pFRA16B37 molecules
completed DNA synthesis and had double-stranded
tails. Moreover, a majority of pFRA16B37 molecules
(73%) had no detectable DNA synthesis (i.e. no
double-stranded tails) and contained polymerase ‘stuck’
to the DNA, even after phenol/chloroform extraction
(Figure 3C). To locate the FRA16B sequence, the
samples were further linearized with restriction enzymes
to generate asymmetric arms (Figure 3C, bottom right
panel). We observed that most (68%) of the
polymerase-bound molecules had polymerases located
within the FRA16B sequence (black boxes). Clone 17 demonstrates a statistically significant increase in instability events in HEK293T cells relative to background instability (*P = 0.022), and to the control (**P = 0.016). (F) Analysis of mutations generated by replication of clone 17 with SV40 ori in HEK293T cells. For the five mutants, the type of mutation, size of the mutation (in bp) and distance (in bp) from SV40 ori are reported in the table.
Figure 3. FRA16B DNA synthesis by Klenow fragment of *E. coli* DNA polymerase I. (A) Construction of replication fork templates to examine fork regression. Plasmids containing FRA16B clone 17 fragment (pFRA16B37, Supplementary Figure S1) or a G-less cassette (pGLGAP) were constructed, nicked with Nb.BbvCl at a site immediately upstream of the inserts (bold), and incubated with Klenow fragment to generate a single-strand tail by strand displacement. The single-strand tail was converted into a double-strand tail by annealing a complimentary oligonucleotide at the 3’ end of the displaced strand and further incubation with Klenow fragment. To map the location of the double-strand tail, plasmids were linearized so measurements of the two asymmetrical tails could be obtained to determine the amount of fork regression. (B) Classification of circular DNA molecules after Klenow fragment reaction by EM. After the second Klenow reaction to create a double-strand tail, reaction mixtures were directly mounted onto carbon-coated copper EM grids and rotary shadowcasted with tungsten. At least 100 molecules from each sample were examined to determine the percentage of molecules with tails, protein-bound, or without a tail. (C) Visualization of circular and linear pGLGAP and pFRA16B37 DNAs after synthesis reaction with Klenow fragment. Representative images were montaged and are shown in reverse contrast. (D) Location of bound polymerases within FRA16B replication fork templates. The location of the polymerase as a percentage of the total length is plotted as a histogram for all FRA16B molecules that displayed bound polymerase and did not contain a double-stranded tail. The FRA16B fragment is located at 33–50% of the total length from the nearest end. (E) Analysis of DNA polymerase pause sites during lagging strand DNA synthesis. Radiolabeled oligomers were annealed to the displaced lagging strand of pGLGAP (lanes 1–3) and pFRA16B37 (lanes 5–7) replication fork templates, and synthesis was carried out with 0.5 U (lanes 1 and 5), 5 U (lanes 2 and 6) or 15 U (lanes 3 and 7) of Klenow fragment. Lane 4 (C) contained the radiolabeled 25-nt oligonucleotide only. The first 12 nt of the newly synthesized complimentary strand are indicated, and the pause sites are boxed. (F) Detection of FRA16B replication fork regression. Circular replication fork molecules were linearized with AhdI to produce asymmetrical arms. The location of the double-strand tail was defined as the percentage of total DNA length from the nearest end.
polymerase-trapped molecules by EM, and suggest that FRA16B DNA inhibits the progression of DNA polymerase, possibly due to its ability to form secondary structure.

Next, we analyzed fork regression with the FRA16B molecules containing double-stranded tails. These molecules were a small fraction (10%) of the total population but could represent a fully regressed, non-regressed, or partially regressed replication fork, depending on the location of the double-stranded tails. The location of the double-strand tail was indicated as the percentage of total DNA length from the nearest end. For the FRA16B template, if no regression occurred, the tail would be located at 46% of the total length, whereas under complete regression, templates would have the tail located at 34% of the total length (Figure 3F). FRA16B forks were considered to have regressed when the double-stranded tails were located ≤40% of the total length, whereas under complete regression, templates would have the tail located at 34% of the total length (Figure 3F). FRA16B forks were considered to have regressed when the double-stranded tails were located ≤40% of the total length, whereas under complete regression, templates would have the tail located at 34% of the total length (Figure 3F). FRA16B forks were considered to have regressed when the double-stranded tails were located ≤40% of the total length, whereas under complete regression, templates would have the tail located at 34% of the total length (Figure 3F). FRA16B forks were considered to have regressed when the double-stranded tails were located ≤40% of the total length, whereas under complete regression, templates would have the tail located at 34% of the total length (Figure 3F). FRA16B forks were considered to have regressed when the double-stranded tails were located ≤40% of the total length, whereas under complete regression, templates would have the tail located at 34% of the total length (Figure 3F). FRA16B forks were considered to have regressed when the double-stranded tails were located ≤40% of the total length, whereas under complete regression, templates would have the tail located at 34% of the total length (Figure 3F). FRA16B forks were considered to have regressed when the double-stranded tails were located ≤40% of the total length, whereas under complete regression, templates would have the tail located at 34% of the total length (Figure 3F).

Out of the 32 molecules analyzed, 26 (81%) underwent fork reversal, whereas six (19%) did not display fork reversal. As a control, pGLGAP templates were examined (n=81), and our results were consistent with previous studies, demonstrating that ~10–20% of random, nonrepetitive DNAs undergo some degree of spontaneous fork regression (46,52). These results show that fragile site-containing replication forks have a greater tendency to undergo fork reversal than non-fragile site sequences.

We also found that even in non-regressed molecules which underwent DNA synthesis, the average length of FRA16B tails was only 21% of the full length, whereas the average length of the control pGLGAP template tails was 85% of the total length. These results suggest that the shorter double-stranded tails generated by DNA synthesis of FRA16B templates could result from bypass of a slipped-out secondary structure formed on the lagging strand template, leading to the generation of deletion mutations, as we observed following FRA16B replication in human cells. Together, these data suggest that polymerase pausing, fork reversal, or polymerase skipping at FRA16B could contribute to its instability.

**DISCUSSION**

To gain a better understanding of the mechanism of instability at chromosomal fragile sites, we examined the ability of FRA16B to form a secondary structure in vitro and evaluated the effects of cis-acting factors on replication of FRA16B DNA. Previously, it has been shown that the CGG repeat, which underlies the basis of fragility at rare, folate-sensitive fragile sites, is able to form a stable secondary structure (19,20) that presents significant difficulties during replication (21,22). However, the ability of an AT-rich fragile DNA to form a secondary structure in vitro has not been demonstrated until now. Our data show that FRA16B DNA forms an alternative structure...
following denaturation and re-annealing. To our knowledge, this is the first experimental evidence of an AT-rich fragile DNA, which represents the majority of fragile sites, forming a secondary structure in vitro. This observation supports the concept that formation of highly stable secondary structures could be a general mechanism that contributes to fragile site breakage during DNA replication (18).

Since prior reports have indicated potential problems with replication due to DNA secondary structures, FRA16B replication in human cells was also examined. We determined that, similar to previous results, cis-acting factors including replication orientation and distance from the origin do play a role in both replication efficiency and instability of FRA16B-containing constructs in human cells. Numerous studies have suggested that replication orientation, which determines leading and lagging strand synthesis, has a significant effect on replication efficiency as well as both the type and frequency of mutations (45,47–49). In general, an increase in the number of mutations is observed when the more structure-prone strand serves as the lagging strand template. Thus, when the strand with the most potential to form a stable secondary structure serves as the lagging strand template, the DNA is more likely to fold into an alternative structure within the regions that maintain single-strandedness. Consistent with previous studies, we found that orientation greatly affects replication efficiency in that clone 17 possessed a markedly lower efficiency than clone 15, yet they differed only in orientation. Clone 17 also displayed a higher frequency of mutations than clone 15 during replication in HEK293T cells. Although both strands of FRA16B are able to form secondary structure, as shown in Figure 1A, we found that there are clearly different structures being formed by each strand, based upon their different electrophoretic mobilities (Supplementary Figure S2). The more rapid mobility of the template strand in Supplementary Figure S2 supports the idea that the lagging strand template in the unstable orientation forms a more stable secondary structure than the template strand in the stable orientation. These data prompt us to propose that one structure may be more deleterious to DNA replication than the other, thus leading to an orientation effect similar to what has been observed when only one strand forms a stable alternative structure. Although significant differences were not observed among the other sets of FRA16B constructs (clones 31 and 34, clones 10 and 12 and clones 46 and 39), this could be due to the fact that in these constructs, FRA16B DNA is at least 300 bp away from the origin, and orientation may only be a factor when the fragile site is in close proximity to the origin. This is likely due to the location of secondary structures within the Okazaki initiation zone (OIZ), a region of ~300 nt that remains single stranded. Placing the beginning of the 921 bp FRA16B DNA fragment 30 bp from the SV40 ori, as was the case for clones 15 and 17, would have FRA16B occupy the first OIZ, promoting the formation of an alternative DNA structure. The other sets of FRA16B constructs have the start of the FRA16B sequence located at the second or the third OIZs. The placement of the beginning of the FRA16B sequence in different OIZs might affect the ability of FRA16B to form stable secondary structure, due to the binding cooperativity of RPA protein (53). The binding of RPA in the first OIZ could propagate to the second and third OIZs, and might prevent the formation of alternative structure by FRA16B, when FRA16B sequence begins at the second and third OIZs. In contrast, clone 17 locates in the first OIZ, and its ability to form alternative structure might trump the initial binding of RPA protein. Consequently, both the orientation and location of FRA16B within clone 17 played a significant role in reducing DNA replication efficiency and increasing the number of mutations. Furthermore, the large base pair deletions that occurred within FRA16B during replication in human cells recapitulate what has been observed at fragile sites in numerous tumors in vivo. The most energetically favorable predicted secondary structure of the lagging strand template of clone 17 (50) suggested that the deleted regions may occur at sites of extensive secondary structure formation, including multiple hairpins (Supplementary Figure S3), further supporting a potential role for secondary structure at sites of DNA breakage within fragile sites. Possible explanations for the increase in mutation frequency and decrease in replication efficiency observed for the FRA16B constructs in the SV40 replication system are: (i) the formation of a secondary structure close to the origin hinders loading of the replication machinery, leading to decreased replication efficiency, and/or (ii) secondary structure formation within single-stranded DNA regions during replication causes insertions via replication restart, or deletions following bypass of the structure or cleavage of a recessed fork—all of which are consequences of replication fork stalling. Since clone 17, which is located 30 bp from the SV40 ori, exhibited a statistically significant decrease in replication efficiency, it is conceivable that inhibition of origin firing is contributing to what is observed. In contrast, clones 31, 34, 10, 12, 46 and 39, which all place FRA16B at least 300 bp from the origin of replication, do not demonstrate a significant reduction in replication efficiency, and this suggests replication may initiate unperturbed. However, the increased frequency of instability events for clone 17 following replication in HEK293T cells supports replication fork stalling, which is a consequence of DNA elongation. Analysis of all five mutants derived from clone 17 demonstrates that sites of deletions and expansions occurred as much as ~400 bp from the origin of replication (Figure 2F), suggesting that DNA replication was initiated but paused at a site located at a greater distance from the origin. A similar study, which examined the ability of the CCTG repeat to cause mutations, reports comparable results, demonstrating that when the repeat is located ~25 bp from the SV40 ori, an increase in expansions and deletions is observed, also suggesting replication fork stalling during DNA synthesis (47).

To dissect the mechanism for the generation of such mutations within FRA16B during replication, we constructed a replication fork template mimicking clone 17 with FRA16B DNA ~30 bp from the initial replication fork and the same replication orientation as clone 17.
Visualization of FRA16B replication fork molecules by EM showed that the majority of DNAs contained polymerases ‘stuck’ at the FRA16B region, and analysis of the lagging strand synthesis confirmed that the polymerase had stalled at specific sites within the FRA16B tract. As these polymerase-trapped molecules were not observed with the non-fragile DNA control template (Figure 3B), we suggest that the formation of secondary structure by FRA16B DNA may be responsible for polymerase pausing. Further, examination of replication fork templates containing either the telomeric or CTG repeats, which have been shown to form hairpin structures, did not reveal polymerase pausing within the repetitive DNAs by the same EM analysis (54), although Oshima and Wells have used DNA sequencing gels to demonstrate polymerase pausing by CTG repeats (55). These results suggest the secondary structure formed within FRA16B is likely more extensive than a single hairpin and highly stable. A highly stable replication fork with exposed single-stranded regions can trigger the ATR-dependent DNA damage checkpoint pathway, which in turn inhibits further firing of replication origins, blocks entry into mitosis and promotes DNA repair, recombination or apoptosis (56,57). However, a deficiency of proteins in the ATR-dependent cell cycle checkpoint pathway will dramatically increase fragile site breakage, as several studies have reported (18,24,25,58).

It is important to note that polymerase stalling at sites of secondary structure is probably only likely when the FRA16B sequence is located very close to the origin of replication, and our results are limited due to the fact that only one construct was analyzed by DNA sequencing gels, and contained FRA16B located 30 bp from the replication origin.

This study is the first to provide evidence of spontaneous replication fork reversal within a fragile site sequence during DNA synthesis. Based on this information, we propose a model whereby FRA16B instability primarily arises from polymerase stalling caused by the formation of secondary structure at the fragile site region (Figure 4B), coupled with the failure of the ATR checkpoint pathway. Other mechanisms contributing to fragile site instability could be due to replication fork regression (Figure 4C), the formation of stable DNA secondary structures on the lagging strand template leading to polymerase skipping (Figure 4D) or replication restart with secondary structure formed on the newly synthesized strand (Figure 4E). These mechanisms may be shared among other fragile DNA sequences. Overall, our results provide insight into the mechanism of fragile site instability by demonstrating the ability of a fragile DNA to form a stable secondary structure that affects replication efficiency and instability, and causes polymerase pausing during DNA synthesis.
SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Dr. Jack Griffith (UNC-CH) for providing pGLGAP plasmid.

FUNDING

The National Cancer Institute (CA85826 and CA113863 to Y.-H. W), Funding for open access charges: National Institutes of Health.

Conflict of interest statement. None declared.

REFERENCES

17. Pearson,C.E. and Sinden,R.R. (1996) Alternative structures in duplex DNA formed within the trinucleotide repeats of the


