Alternative Structures in Duplex DNA Formed within the Trinucleotide Repeats of the Myotonic Dystrophy and Fragile X Loci†

Christopher E. Pearson and Richard R. Sinden*

Center for Genome Research, Institute of Biosciences and Technology in the Texas Medical Center, Department of Biochemistry and Biophysics, Texas A&M University, 2121 West Holcombe Boulevard, Houston, Texas 77030-3303

Received January 16, 1996; Revised Manuscript Received February 19, 1996⊗

ABSTRACT: Most models proposed to explain the disease-associated expansion of $(CTG)_n$·$(CAG)_n$ and $(CGG)_n$·$(CCG)_n$ trinucleotide repeats include the formation of slipped strand DNA structures during replication; however, physical evidence for these alternative DNA secondary structures has not been reported. Using cloned fragments from the myotonic dystrophy (DM) and fragile X syndrome (FRAXA) loci containing normal, premutation, and full mutation lengths of repeats, we report the formation of novel alternative DNA secondary structures that map within the repeat tracts during reannealing of complementary strands, containing equal lengths of repeats, into linear duplex DNA molecules. Linear duplex DNA molecules containing these alternative DNA secondary structures are characterized by reduced electrophoretic mobilities in polyacrylamide gels. These alternative secondary structures are stable at physiological ionic strengths and to temperatures up to at least 55 °C. Following reduplexing, the CAG strand of the $(CTG)_n$·$(CAG)_n$ repeats is preferentially sensitive to mung bean nuclease, suggesting the presence of single-stranded DNA in the alternative DNA structure. For $(CTG)_{17}$, which is a repeat length found in normal individuals, less than 3% of the DNA molecules formed alternative DNA structures upon reduplexing. DNA molecules containing $(CTG)_{50}$ or $(CTG)_{255}$, which represent premutation and full mutation lengths of triplet repeats, respectively, formed a heterogeneous population of alternative DNA structures in >50% of DNA molecules. The complexity of the structures formed increased with the length of the triplet repeat. The relationship between repeat length and the propensity of formation and complexity of the novel structures correlates with the effect of repeat length on genetic instability in human diseases. These are the first results consistent with the existence of slipped strand DNA structures. The potential involvement of these structures, which we term S-DNA, in the genetic instability of triplet repeats is discussed.

Recently, the etiology of nine human genetic diseases, including myotonic dystrophy (DM)[1] (Fu *et al.*, 1992; Brook *et al.*, 1992; Mahadevan *et al.*, 1992) and fragile X syndrome (FRAXA) (Fu *et al.*, 1991; Kremer *et al.*, 1991), has been traced to genetic variation in the lengths of $(CTG)_n$·$(CAG)_n$ or $(CGG)_n$·$(CCG)_n$ triplet repeats in DNA [for review, see Sutherland and Richards (1995)]. Genetically, triplet repeat-associated diseases show the phenomenon of anticipation which is characterized by an earlier age at onset, an increased severity of the disease, and/or increased penetrance with successive generations. At the molecular level, genetic anticipation is accompanied by an increase in the number of trinucleotide repeats within the disease loci [for example, see Ashizawa *et al*. (1992) and Fu *et al.* (1991)]. The importance of the length of the repeat tract is evident for many of the diseases, including DM (Mahadevan *et al.*, 1992) and FRAXA (Fu *et al.*, 1991; Kremer *et al.*, 1991). In normal individuals, these loci contain a short length of triplet repeats (usually 5−37), which is polymorphic within the population. Increases in the lengths of the triplet repeats to 50−200 are associated with unstable premutation alleles. Explosive expansion to thousands of repeats can occur at the DM and FRAXA loci. The dependence of the probability of mutation upon the number of repeats has led to the appropriate term of "dynamic mutation" to describe this process (Richards & Sutherland, 1992).

In addition to the disease-linked fragile sites *FRAXA* (Fu *et al.*, 1991; Kremer *et al.*, 1991), *FRAXE* (Knight *et al.*, 1993), and *FRA11B* (Jacobsen syndrome) (Jones *et al.*, 1995), expansions of CGG trinucleotide repeats have also been associated with the chromosomal fragile site *FRAXF* (Parish *et al.*, 1994) and the autosomal fragile site *FRA16A* (Nancarrow *et al.*, 1994). Expansions of 800−1000 repeats have been detected at each of these fragile sites.

The mechanism of repeat expansion is completely unknown. Several hypotheses to explain this phenomenon involve aberrant replication or recombination at slipped strand intermediates within the triplet repeats (Fu *et al.*, 1991; Richards & Sutherland, 1992, 1994; Sinden & Wells, 1992; Wells & Sinden, 1993; Chong *et al.*, 1994; Eichler *et al.*, 1994; Jansen *et al.*, 1994; Kunst & Warren, 1994; Snow *et al.*, 1994; Kang *et al.*, 1995). For example, large expansions may occur when the length of the triplet repeat increases beyond the size of an Okazaki fragment, such that the 5′ and 3′ ends of a nascent fragment, which may be prone to slippage, are within the repeat tract (Richards & Sutherland,

† This work was supported by funds from the Institute of Biosciences and Technology and in part from NIH Grant GM52982.

⊗ Abstract published in *Advance ACS Abstracts,* April 1, 1996.

[1] Abbreviations: DM, myotonic dystrophy; PCR, polymerase chain reaction; *FMR1*, fragile X mental retardation gene 1; FRAXA, fragile X A.

0006-2960/96/0435-5041$12.00/0

1994). Although slipped strand DNA structures could theoretically form within long runs of a triplet repeat, no conclusive evidence has ever been presented for such structures. Using short oligonucleotides, it has recently been demonstrated that CTG, CAG, CGG, or CCG single strands can form intrastrand duplex hairpin structures (Chen *et al.*, 1995; Gacy *et al.*, 1995; Gao *et al.*, 1995; Mitas *et al.*, 1995; Smith *et al.*, 1995). Single strands containing CGG have also been shown to form quadruplex structures (Fry & Loeb, 1994). Kohwi *et al.* (1993) have described a Zn^{2+}- and Co^{2+}-dependent non-B-DNA conformation for CAG triplet repeats. Other alternative (non-B) DNA structures such as cruciforms, left-handed Z-DNA, and intramolecular triplex structures can form *in vivo* in the defined ordered sequence DNA elements: inverted repeats, alternating purine-pyrimidine tracts [(GC) or (AC)], and homopurine-homopyrimidine tracts containing mirror repeat symmetry [e.g. $(GA)_n$ and $(G)_n$], respectively [for review, see Sinden (1994)]. Spontaneous mutations in prokaryotic and eukaryotic (including human) cells are frequently associated with DNA sequence elements that can form alternative non-B-DNA structures [reviewed in Wells and Sinden (1993)]. Understanding the formation, structure, and stability of slipped strand structures in triplet repeats should contribute to our understanding of the molecular mechanisms responsible for their instability.

Using disease relevant lengths of triplet repeat tracts, we describe novel DNA secondary structures formed in otherwise duplex DNA between complementary strands of trinucleotide repeat-containing fragments from both DM and FRAXA genomic clones. This is the first report of alternative non-B-DNA structures, presumably slipped strand structures, formed within duplex $(CTG)_n$·$(CAG)_n$ and $(CGG)_n$·$(CCG)_n$ tracts. We term these structures S-DNA.

EXPERIMENTAL PROCEDURES

Plasmids/DNA. The $(CTG)_n$-containing DM clones containing the CTG-containing region [positions 357−433, as in Mahadevan *et al.* (1992)] where $n = 17$ (pRW3211), $n = 50$ (pRW3213), and $n = 255$ (pRW3222) have been described (Kang *et al.*, 1995; Chastain *et al.*, 1995) and are referred to here as $(CTG)_{17}$, $(CTG)_{50}$, and $(CTG)_{255}$, respectively. The repeat tract of $(CTG)_{17}$ and $(CTG)_{50}$ are pure, and that of the $(CTG)_{255}$ contains four ACT interruptions resulting in a tract of $(CTG)_{27}ACT(CTG)_{40}ACT(CTG)_{38}ACT(CTG)_{40}ACT(CTG)_{106\pm5}$.

The pFXA9.9.32 clone was created by using genomic DNA from a normal individual (FW). The CGG-containing region of the FRAXA loci, having a repeat pattern of 52 repeat units with two AGG interruptions $[(CGG)_9AGG(CGG)_9AGG(CGG)_{32}]$, was PCR amplified using the primers (571R and A) and conditions previously described (Chong *et al.*, 1994). The PCR product was then inserted into the PCRII vector (Invitrogen). The plasmid gives an identical *Mnl*I restriction digest (Eichler *et al.*, 1994) compared to the starting material, indicating that the plasmid clone has the same interspersion pattern.

Plasmid Preparation. Plasmid preparations were made from *Escherichia coli* cells that had grown overnight in Luria broth at 37 °C without amplification essentially as described (Kochel & Sinden, 1988). It is important that the cells not be maintained for too long a period at stationary phase as the amount of deletion products increases (unpublished results). Cells were harvested and lysed by treatment with lysozyme (Gibco) and a detergent solution of 1% Brij 58 (Sigma) and 0.4% deoxycholate (Sigma). It is important that cells are lysed by gentle means, i.e. not by boiling or alkaline lysis methods, as these methods may promote plasmid denaturation. Plasmids were treated with DNAse-free RNAse A and T1 (Sigma), phenol extracted, purified twice by isopycnic (cesium chloride/ethidium bromide) centrifugation, and stored in TE [10 mM Tris and 1 mM ethylenediaminetetraacetic acid (EDTA) (pH 7.6)] at −20 °C.

DNA Treatments and Reduplexing Reactions. Reduplexing reactions were performed as previously published (Pearson *et al.*, 1994, 1995). Essentially, DNA samples were ethanol precipitated, dried in a Speed Vac, then denatured by resuspension in a solution of 500 mM NaOH (pH 13) and 1.5 M NaCl, and incubated at room temperature for 5 min. Samples were neutralized by the addition of a 50-fold volume of 50 mM Tris-HCl (pH 8) and 5 mM EDTA, resulting in a solution having 0.01 M NaOH and 0.03 M NaCl (pH 8). These conditions promote full renaturation (Studier, 1969). Samples were immediately placed at 68 °C for 3 h, followed by purification by ethanol precipitation, resuspended in TE, and stored at −20 °C. This treatment does not affect the integrity of the phosphate backbone (Pearson *et al.*, 1994, 1995; Pouwels *et al.*, 1968; Rush & Warner, 1970; unpublished results). Precautionary measures were taken to avoid sample dehydration (Pearson *et al.*, 1994, 1995). All restriction enzymes were purchased from New England Biolabs, and reactions were performed as specified by the manufacturer. Procedures for end labeling [using T4 kinase (USB) to label the 5′ end and AMV reverse transcriptase (USB) to label the 3′ end] and electroelution were described (Pearson *et al.*, 1994, 1995).

Electrophoresis. Polyacrylamide gels (40:2 acrylamide:Bis) (4%) were cast in TBE [90 mM Tris, 90 mM borate, and 2.5 mM EDTA (pH 8.3)] and run at a constant voltage (150 V; 10−12 V/cm) at room temperature. Gels were then stained with ethidium bromide, photographed and/or dried on Whatman paper, and exposed to radiographic film (Kodak) in the presence of intensifying screens at −70 °C or a PhosphorImager screen (Molecular Dynamics). Quantitative analysis was performed using the PhosphorImager with ImageQuant software (Molecular Dynamics). The fraction of the *Hin*dIII/*Eco*RI fragments migrating as novel anomalous products represents the average of quantitative analysis from three to ten experiments. All relative migrations and base pair (bp) estimations were calculated with respect to the migration of the 123 bp ladder (Gibco) as previously described (Chastain *et al.*, 1995).

Mung Bean Nuclease Treatment. Mung bean nuclease reactions were performed essentially as described (Umek & Kowalski, 1988) using 0.05 mg/mL DNA in 10 mM Tris-HCl (pH 7.5), 50 mM NaCl, and 10 u of mung bean nuclease (New England Biolabs). Reaction mixtures were incubated at 37 °C for 45 min, the reactions were stopped by addition of EDTA, and the DNA was purified by phenol/chloroform extraction and ethanol precipitation. Equal amounts of counts per minute were irreversibly denatured by treatment with glyoxal (McMaster & Carmichael, 1977) and separated on a 0.5% agarose gel (30.5 cm × 15.5 cm) in TBE by electrophoresis at 60 V for the indicated time. The gels were dried and exposed for radiography.

RESULTS

Detection of Novel Alternative Structures in Reduplexed Linear and Supercoiled DM DNAs Containing CTG Repeats. The cellular processes of transcription, replication, and recombination require that regions of DNA undergo a transformation from double- to single-stranded and back to the double-stranded state. We mimicked this process by alkali denaturation and renaturation (reduplexing) to ask if alternative DNA structures could form during this process. We used three genomic clones of the human DM locus that have identical human flanking sequences and 17, 50, or 255 CTG repeats (Figure 1A) (Kang *et al.*, 1995; Chastain *et al.*, 1995) corresponding to the normal, the premutation, and the full mutation number of repeat units, respectively (Fu *et al.*, 1992; Brook *et al.*, 1992; Mahadevan *et al.*, 1992).

Each plasmid was linearized by *Hin*dIII digestion, ^{32}P-end-labeled on either the 5′ end of the CAG strand or the 3′ end of the CTG strand (Figure 1B), and reduplexed as described in Experimental Procedures. Following reduplexing, the DNA was digested with *Eco*RI to generate a fragment containing the DM CTG insert (Figure 1B). Comparison of the polyacrylamide gel electrophoresis (PAGE) patterns revealed that the reduplexed plasmids gave rise to novel anomalously migrating products not present in the nontreated plasmids (Figure 1C). As we have previously described (Chastain *et al.*, 1995), each of the linear nontreated *Hin*dIII/*Eco*RI fragments migrated anomalously fast; the 164 bp $(CTG)_{17}$ fragment migrated as 156 bp, the 263 bp $(CTG)_{50}$ fragment migrated as 232 bp, and the 878 bp $(CTG)_{255}$ fragment migrated as 788 bp. In the case of reduplexed $(CTG)_{17}$ DNA, a single major novel product migrating as 210 bp comprised 2.7% of the $(CTG)_{17}$-containing *Hin*dIII/*Eco*RI fragment (Figure 1C, compare lane 1 with 6). For reduplexed $(CTG)_{50}$ DNA, a series of closely spaced novel bands migrating between 256 and 544 bp was observed (Figure 1C, compare lane 2 with 7 and lane 4 with 9). These novel products comprised 60.2% of the $(CTG)_{50}$-containing fragment, with a single region of greatest intensity migrating as 312 bp. Reduplexing the $(CTG)_{255}$ plasmid also yielded a series of closely spaced novel bands with three major bands centered at 1009, 1160, and 1310 bp (Figure 1C, compare lane 3 with 8 and lane 5 with 10). In total, 55.2% of the $(CTG)_{255}$-containing fragment migrated as novel anomalous products. A $(CTG)_{30}$-containing fragment formed anomalous products in ≈39% of the sample and yielded a pattern similar to $(CTG)_{50}$ (data not shown). The occurrence of the novel products was dependent upon reduplexing and independent of DNA concentration during renaturation. The concentration independence would argue against the formation of duplex−duplex complexes. Reduplexing plasmids lacking repeat tracts did not result in any novel anomalous bands (Pearson *et al.*, 1994, 1995; see below).

The novel anomalously migrating bands suggested that reduplexing had induced the formation of alternative, non-B-DNA structures in the repeat-containing *Hin*dIII/*Eco*RI fragments. The reduced mobility of these DNA fragments is consistent with the introduction of bends in the DNA resulting from short loops or three- or four-way junctions formed by CTG and CAG slipped out or hairpin structures (Leontis *et al.*, 1991; Welch *et al.*, 1993; Sinden, 1994; Lilley, 1995). The mobility was faster than that expected for multimeric aggregates. The heterogeneous population of products is indicative of a family of structural isomers with bends located at different (or multiple) positions within the molecules. For all three plasmids, the number of novel products, their relative mobilities, and relative intensities were indistinguishable when either the CAG or the CTG strands were radiolabeled (Figure 1C, compare lane 7 with 9 and lane 8 with 10). This indicates that the novel products were composed of both CTG- and CAG-containing strands. These novel products were not individual single strands since denatured, radiolabeled $(CTG)_n$- or $(CAG)_n$-containing strands migrated as a single band just above that for duplex DNA (data not shown). Unlike the formation of many alternative DNA structures, e.g. cruciforms, and even the non-B-DNA $(CTG)_n$ structure identified by Kohwi *et al*. (1993), the formation of the reduplex-induced alternative structures within the trinucleotide repeats did not require DNA supercoiling.

To further characterize the alternative structures, reduplexing was performed on supercoiled plasmids, where, following denaturation, the two DNA circular strands are catenated and cannot separate (Figure 1D). Following reduplexing, plasmids were digested with *Hin*dIII/*Eco*RI and the mobilities analyzed by PAGE. Reduplexing supercoiled plasmids yielded patterns of novel products similar to those obtained with the linearized plasmids (Figure 1E). Upon reduplexing the $(CTG)_{17}$-containing plasmid, the anomalous products could not reproducibly be visualized by staining with ethidium bromide (Figure 1E, compare lane 3 with 4). However, the major anomalous $(CTG)_{50}$-containing band and the three major anomalous $(CTG)_{255}$-containing bands were clearly visible (Figure 1E, compare lane 5 with 6 and lane 9 with 10). The reduplexed *Hin*dIII/*Eco*RI fragments from the parental plasmid (pUC19) did not display any novel bands (Figure 1E, compare lane 1 with 2 and lane 7 with 8). Although the pUC19 *Hin*dIII/*Eco*RI fragment is only 51 bp long and not visible on this gel, it is clear that the novel anomalous bands do not result from the degradation of the larger *Hin*dIII/*Eco*RI fragment. Since the alkali treatment used does not affect the integrity of the phosphate backbone (Pouwels *et al.*, 1968; Rush & Warner, 1970; Pearson *et al.*, 1994, 1995; C. E. Pearson and R. R. Sinden, unpublished results), the appearance of novel products from supercoiled plasmids indicates that they are the result of alternative interactions between the two complementary strands of the same double-stranded molecule, rather than the result of interactions between complementary strands from different duplex DNA molecules.

Alternative Structures Map within the Trinucleotide Repeats. To map the novel alternative DNA structures, each of the nontreated and reduplexed [^{32}P]*Hin*dIII/*Eco*RI fragments of Figure 1C was subjected to further restriction analysis using *Sau*3AI or *Msp*I, which cut 6 bp 3′ or 8 bp 5′ of the CTG repeats, respectively (Figure 2A). *Sau*3AI digestion removes the $(CTG)_n$ tract from the [^{32}P]*Hin*dIII label, while *Msp*I removes the nonrepetitive DNA distal to the [^{32}P]*Hin*dIII label, leaving the repeat tract. All *Sau*3AI digestions of either nontreated or reduplexed $(CTG)_{17}$-, $(CTG)_{50}$-, or $(CTG)_{255}$-containing plasmids gave rise to the expected 50 bp [^{32}P]*Hin*dIII/*Sau*3AI fragment (Figure 2B, lanes 2, 5, 8, 11, 14, and 17; see arrow). In contrast, all *Msp*I digestions yielded the expected [^{32}P]*Hin*dIII/*Msp*I duplex product, as well as novel products for the reduplexed reactions. Following *Msp*I digestion, the novel products

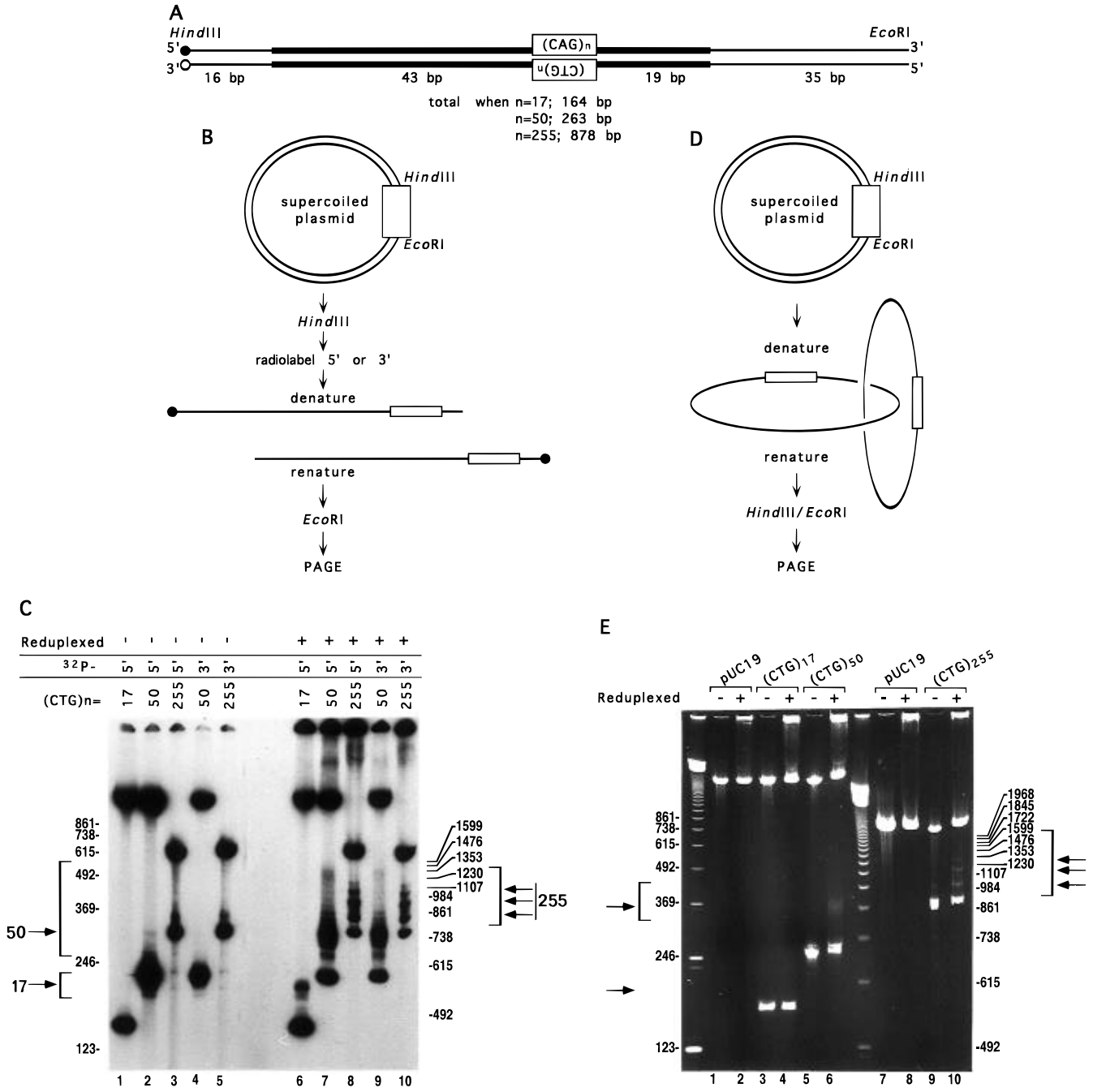

FIGURE 1: Reduplexing DM DNA fragments containing $(CTG)_n$ repeats results in slowly migrating novel DNAs. (A) Map of the *Hin*dIII/*Eco*RI fragments from DM genomic clones. Human nonrepetitive flanking sequences [positions 357−375 and 391−433, as in Mahadevan *et al.* (1992)] are in bold. The thin lines represent plasmid vector sequences. Positions of the 5′ and 3′ end labels are indicated by a closed and an open dot, respectively. (B) Schematic for the treatment of the repeat-containing linearized plasmids. Plasmids were digested with *Hin*dIII and ^{32}P-labeled on either the 5′ or the 3′ ends by T4 polynucleotide kinase or AMV reverse transcriptase. These DNAs were then reduplexed and digested with *Eco*RI, resulting in DNAs uniquely ^{32}P-labeled at the 5′ end of the CAG-containing strand or the 3′ end of the CTG-containing strand. (C) Non-reduplexed controls (lanes 1−5) and reduplexed (as in panel B) [^{32}P]*Hin*dIII/*Eco*RI restriction digestion products (lanes 6−10) were separated on a 4% polyacrylamide gel, dried, and exposed for autoradiography. The arrows and brackets denote the positions of the anomalous DNA fragments following reduplexing; the arrows denote major bands, while the brackets indicate the distribution of total radioactivity migrating at anomalous positions. (D) Schematic for the treatment of pUC19 and repeat-containing supercoiled plasmids. Native supercoiled plasmids were reduplexed and then digested with both *Hin*dIII and *Eco*RI. (E) Non-reduplexed controls (lanes 1, 3, 5, 8, and 9) and reduplexed (as in panel D) *Hin*dIII/*Eco*RI restriction-digested products (lanes 2, 4, 6, 8, and 10) were separated on a 4% polyacrylamide gel and stained with ethidium bromide. To increase resolution, the $(CTG)_{255}$ samples in the gels shown in both panels C and E were loaded earlier than all other samples. In both panels C and E, the 123 bp ladder loaded with the $(CTG)_{17}$ and $(CTG)_{50}$ samples is shown to the left of the gels while that loaded with the $(CTG)_{255}$ samples is shown to the right of the gels.

migrated faster while the basic pattern of the anomalous bands was preserved (Figure 2B, compare lane 4 with 6, lane 10 with 12, and lane 16 with 18; see brackets). In the *Msp*I digestions of the reduplexed $(CTG)_{17}$- and $(CTG)_{50}$-containing DNA, the retarded migration of the novel products relative to the linear [^{32}P]*Hin*dIII/*Msp*I product was proportional to their retarded migration prior to *Msp*I digestion (Figure 2B, compare lane 4 with 6 and lane 10 with 12; see brackets). While the increase in mobility was small for the $(CTG)_{255}$ products in the gel shown in Figure 2B (compare

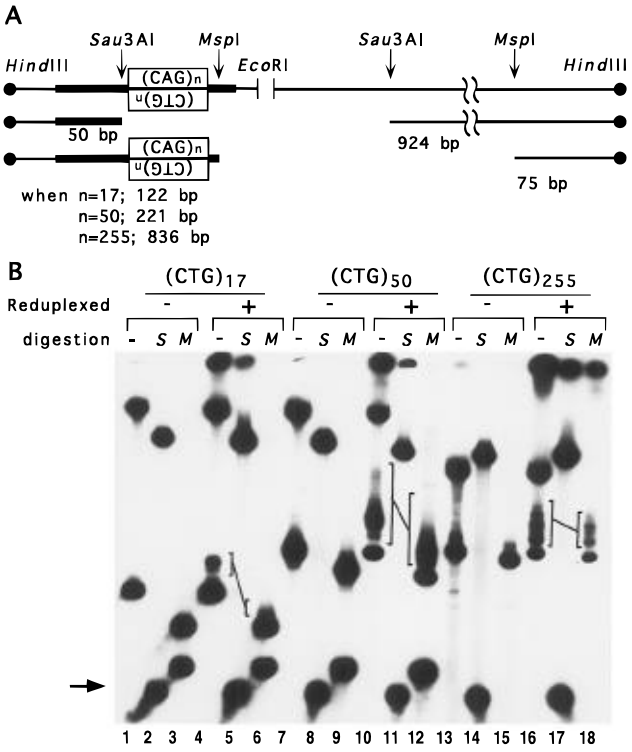

FIGURE 2: Source of the anomalous migration maps within the $(CTG)_n$ triplet repeat. (A) Restriction sites for *Hin*dIII, *Eco*RI, *Sau*3AI, and *Msp*I and the sizes of the *Hin*dIII/*Eco*RI, *Hin*dIII/*Sau*3AI, and *Hin*dIII/*Msp*I restriction-digested products are indicated. Human sequences are in bold, and the ^{32}P radiolabel is indicated by a dot. (B) *Sau*3AI and *Msp*I restriction digests of nontreated and reduplexed [^{32}P]*Hin*dIII/*Eco*RI reactions (see Figure 1C) were analyzed on a 4% polyacrylamide gel, dried, and exposed for autoradiography. The 5′-labeled [^{32}P]*Hin*dIII/*Eco*RI products are shown. The position of the [^{32}P]*Hin*dIII/*Sau*3AI product from each $(CTG)_n$-containing [^{32}P]*Hin*dIII/*Eco*RI fragment is indicated by an arrow. The reduplexed novel $(CTG)_n$-containing [^{32}P]*Hin*dIII/*Eco*RI products as well as their respective *Msp*I-digested products are indicated by connected brackets for each reduplexed $(CTG)_n$ repeat. To increase resolution, each of the $(CTG)_{255}$ samples and their *Msp*I digests (lanes 13 and 16 and 15 and 18, respectively) were loaded earlier than all other samples. The *Sau*3AI digests of the $(CTG)_{255}$ samples (lanes 14 and 17) were loaded with the other samples. pBR325 DNA (500 ng) was included in each digestion as an internal positive control for complete digestion, and completion was monitored by ethidium staining of gels (data not shown).

lane 16 with 18; see brackets), longer electrophoresis more clearly demonstrated a decrease in the length of the $(CTG)_{255}$ [^{32}P]*Hin*dIII/*Msp*I products (data not shown). In addition, when restriction digests for all DNAs were analyzed on DNA sequencing gels, the patterns of bands for nontreated and the reduplexed DNAs were identical. This indicated that complete digestion had occurred at both the *Sau*3AI and *Msp*I sites on both the linear duplex and DNAs containing the novel structures. Since the novel bands were abolished by *Sau*3AI digestion and the mobility of the novel bands was proportionally increased by *Msp*I digestion, these results delimit the structural anomaly of the reduplexed DM fragment to within 6 bp 3′ and 8 bp 5′ of the trinucleotide repeat tract. Moreover, since neither *Sau*3AI nor *Msp*I is able to restrict single-stranded DNA, the digestion of both the nontreated linear and reduplexed novel products indicates that the DNA flanking the repeat tracts is in a linear duplex conformation.

We also observed digestion of both the nontreated and reduplexed repeat tracts with *Fnu*4HI, *Bso*FI, and *Bbv*I, each of which recognize sites within the repeat tracts (5′-GCNGC-3′, 5′-GCNGC-3′, and 5′-GCAGC-3′, respectively) (data not shown). Digestion by these enzymes yielded [^{32}P]*Hin*dIII fragments cut at their recognition sites within the repeat tracts that were most proximal to the [^{32}P]*Hin*dIII end. These results indicate either that this region of the repeat tracts was in a B-like conformation or that the anomalous structure is efficiently recognized and restricted by these enzymes. Since *Bbv*I does not cut single-stranded DNA, the results may be consistent with linear duplex DNA within the immediate end of the repeat tract.

Alternative Structures in CTG Repeats Exhibit Remarkable Stability. To investigate the stability of the alternative DNA structures, we examined their electrophoretic migration following purification from a polyacrylamide gel. The region of greatest intensity of the [^{32}P]*Hin*dIII/*Eco*RI $(CTG)_{50}$ novel band (see Figure 1C, lane 7 or 9) was electroeluted from a polyacrylamide gel (Pearson *et al.*, 1994, 1995), as was the reduplexed product that comigrated with the linear nontreated [^{32}P]*Hin*dIII/*Eco*RI $(CTG)_{50}$ DNA fragment. Each of the three major novel $(CTG)_{255}$-containing fragments and the reduplexed product that comigrated with the linear nontreated [^{32}P]*Hin*dIII/*Eco*RI $(CTG)_{255}$ DNA fragment (see Figure 1C, lane 8 or 10) were also isolated. The mobilities of the isolated novel products were identical to those prior to purification (data not shown) and to those of purified samples incubated at 37 °C (Figure 3A, compare lanes 3 and 4 with lane 2; Figure 3B, compare lanes 3−6 with lane 2). This indicates that the novel DNA structures were stable through elution, multiple buffer changes, ethanol precipitation, as well as phenol extraction with strong vortexing [which is known to accelerate the reassociation of DNA (Kohne *et al.*, 1977)].

To further investigate the stability of the $(CTG)_{50}$ and $(CTG)_{255}$ novel products, nontreated (Figure 3A, lane 1; Figure 3B, lane 1) and reduplexed [^{32}P]*Hin*dIII/*Eco*RI samples (Figure 3A, lane 2; Figure 3B, lane 2), as well as the isolated products (Figure 3A, lanes 3 and 4; Figure 3B, lanes 3−6), were incubated for 60 min at 37, 55, or 85 °C. After incubation at 37 and 55 °C, the electrophoretic pattern of each sample was identical to that prior to temperature treatment (Figure 3A, compare lanes 1−4 with 5−8; Figure 3B, compare lanes 1−6 with 7−12). These results demonstrate that both the $(CTG)_{50}$ and $(CTG)_{255}$ linear duplexes and novel structures were stable up to 55 °C. However, following incubation at 85 °C, novel products similar to those present in the reduplexed reactions appeared in the nonreduplexed [^{32}P]*Hin*dIII/*Eco*RI samples (Figure 3A, compare lane 1 with 9; Figure 3B, compare lane 1 with 13). Moreover, for the individual purified products, treatment at 85 °C resulted in the appearance of multiple novel bands (including products that did not enter the gel) as well as a band comigrating with the linear nontreated duplex fragment (Figure 3A, compare lanes 3 and 4 with 11 and 12; Figure 3B, compare lanes 3−6 with 15−18). The isolated novel structures, in $(CTG)_{50}$ and $(CTG)_{255}$, were stable following overnight incubation at 4, 25, 37, or 55 °C, as well as after months of storage at −20 °C. Only after long autoradiographic exposure was some (≈2%) interconversion detected (C. E. Pearson and R. R. Sinden, unpublished results), which is similar to that observed for similarly sized pseudocruciform heteroduplexes (Pearson *et al.*, 1994, 1995). These results indicate that at 37 °C and at a physiological salt concentration there appears to be minimal interconversion

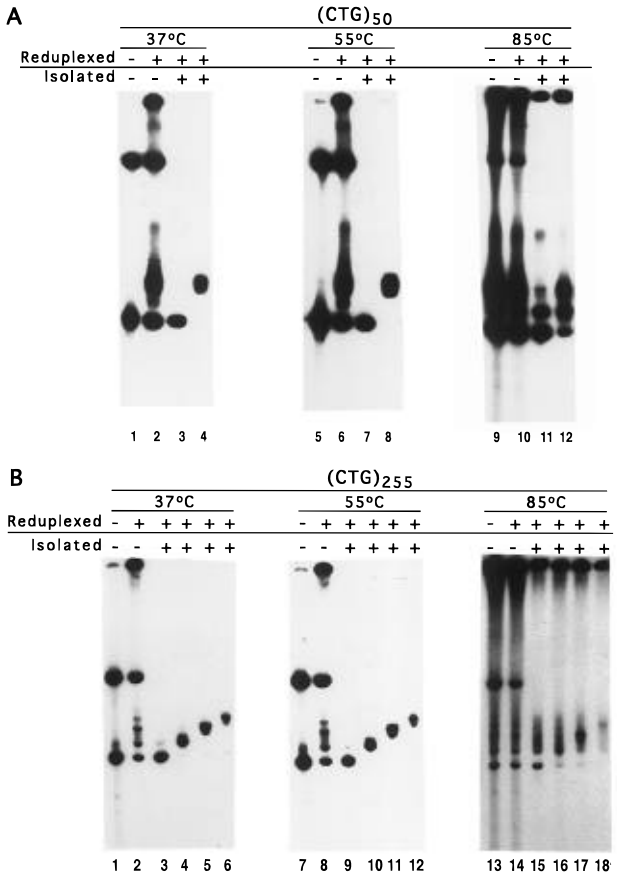

FIGURE 3: Temperature stability of novel products. (A) ^{32}P-labeled $(CTG)_{50}$-containing *Hin*dIII/*Eco*RI fragments (lanes 1, 5, and 9), reduplexed *Hin*dIII/*Eco*RI products (lanes 2, 6, and 10), the gel-purified $(CTG)_{50}$-containing *Hin*dIII/*Eco*RI band that comigrated with the nontreated linear fragment (lanes 3, 7, and 11), and the gel-purified $(CTG)_{50}$-containing *Hin*dIII/*Eco*RI novel band (lanes 4, 8, and 12) in TEN [50 mM NaCl, 1 mM EDTA, 10 mM Tris (pH 7.6)] were incubated for 60 min at 37, 55, or 85 °C, allowed to cool at room temperature for 30 min, and then separated on 4% polyacrylamide gels. (B) Shown are ^{32}P-labeled $(CTG)_{255}$-containing *Hin*dIII/*Eco*RI fragments (lanes 1, 7, and 13), reduplexed *Hin*dIII/*Eco*RI products (lanes 2, 8, and 14), and the gel-purified $(CTG)_{255}$-containing *Hin*dIII/*Eco*RI band that comigrated with the nontreated linear fragment (lanes 3, 9, and 15). The last three lanes of each set of samples (4, 10, and 16; 5, 11, and 17; and 6, 12, and 18) correspond to each of the three major gel-purified $(CTG)_{255}$-containing *Hin*dIII/*Eco*RI novel bands produced after reduplexing. Samples were treated at 37, 55, and 85 °C, as indicated. Panels A and B show analysis of 5′-labeled [^{32}P]*Hin*dIII/*Eco*RI products. Results were identical using 3′-labeled [^{32}P]*Hin*dIII/*Eco*RI products (data not shown).

between the various structural isomers formed by each reduplexing reaction. However, treatment at 85 °C for 60 min was sufficient to induce the formation of the alternative structures.

Alternative Structures Are Sensitive to Mung Bean Nuclease. Slipped strand DNA structures would be expected to contain single-stranded regions of DNA. Individual CTG or CAG strands may fold into independent secondary structures (Chen *et al.*, 1995; Gacy *et al.*, 1995; Gao *et al.*, 1995; Mitas *et al.*, 1995; Smith *et al.*, 1995), where single-stranded loops would exist. Due to both the length of the triplet repeat tracts and the broad spectrum of different structural isomers that could form (see Discussion), the fine details of the structures cannot be readily determined. In order to test for the presence of single-stranded regions in the complete population of alternative DNA structures, the mung bean nuclease sensitivity of the repeat tracts in linearized full length duplex or reduplexed plasmids was investigated. At neutral pH, mung bean nuclease specifically recognizes single-stranded DNAs, unwound regions, cruciforms, and hairpin loops (Umek & Kowalski, 1988). $(CTG)_{50}$ or $(CTG)_{255}$ plasmids were linearized with *Hin*dIII and labeled at either the 5′ or 3′ ends, and an aliquot of each was reduplexed (Figure 4A). Each sample was then digested with *Pst*I (removing 12 bp and the proximal ^{32}P end label), resulting in DNAs uniquely labeled at the 5′ end of the CTG-containing strand or the 3′ end of the CAG-containing strand (Figure 4A). Samples were then digested with mung bean nuclease, irreversibly denatured with glyoxal (McMaster & Carmichael, 1977), and analyzed on an agarose gel (Figure 4B). Digestion of the radiolabeled linear plasmids with *Bam*HI, which cuts 29 nucleotides 5′ of the $(CTG)_n$ tract (Figure 4A−C, lanes 9 and 19), provided a convenient marker for digestion within the repeat tracts.

A low-background level mung bean nuclease digestion occurred on both 5′- and 3′-radiolabeled strands throughout the entire linear duplex $(CTG)_{50}$ and $(CTG)_{255}$ plasmids with no preferential sites of cutting (Figure 4B, compare lane 1 with 3, lane 2 with 4, lane 11 with 13, and lane 12 with 14). In contrast, mung bean nuclease digestion of the 3′-radiolabeled (CAG) strands of reduplexed $(CTG)_{50}$ and $(CTG)_{255}$ plasmids resulted in digestion products migrating just above the [^{32}P]*Hin*dIII/*Bam*HI marker (Figure 4B, compare lane 7 with 8 and lane 17 with 18). This is more clearly seen in Figure 4C, where the mung bean nuclease-digested products were better resolved by longer electrophoresis (lanes 7−10 and 17−20). In contrast to the (CAG)-containing strands, under these conditions, specific digestion of the (CTG)-containing strands was not observed. These results demonstrate that (1) the preferential digestion was dependent upon an alternative DNA structure, (2) the sensitivity was within the repeat tracts, and (3) the sensitivity was preferential for the CAG strand of the repeat tract.

Detection of Novel Alternative Structures in Reduplexed Linear and Supercoiled FRAXA *DNAs Containing CGG Repeats*. Reduplexing analysis was also performed on pFXA9.9.32 containing a tract of 52 repeats from the *FMR-1* gene $[(CGG)_9AGG(CGG)_9AGG(CGG)_{32}]$ (Figure 5A). In a fashion similar to that shown in Figure 1B, the pFXA9.9.32 plasmid was linearized by *Hin*dIII digestion, ^{32}P-end-labeled on either the 5′ end of the CCG strand or the 3′ end of the CGG strand, and reduplexed by alkali denaturation and renaturation. This was followed by *Xba*I digestion to liberate the FRAXA CGG insert. The linear nontreated 600 bp CGG-containing *Hin*dIII/*Xba*I fragment migrated as 583 bp. Reduplexed plasmid gave rise to a set of novel anomalously slow migrating bands not present in the nontreated plasmid (Figure 5B, compare lanes 3 and 4 with 1 and 2). The novel products, which comprised 15% of the $(CGG)_{52}$-containing *Hin*dIII/*Xba*I fragment, migrated as a sharp distinct band at 897 bp and a fainter band at 792 bp (Figure 5B, lanes 3 and 4; see upper arrows). The fraction and pattern of novel products were indistinguishable when either the CCG- or the CGG-containing strand was labeled (Figure 5B, compare lane 3 with 4), indicating that the novel products were composed of both complementary strands. The presence of the novel products suggested that an alternative, non-B-DNA structure had been induced by reduplexing.

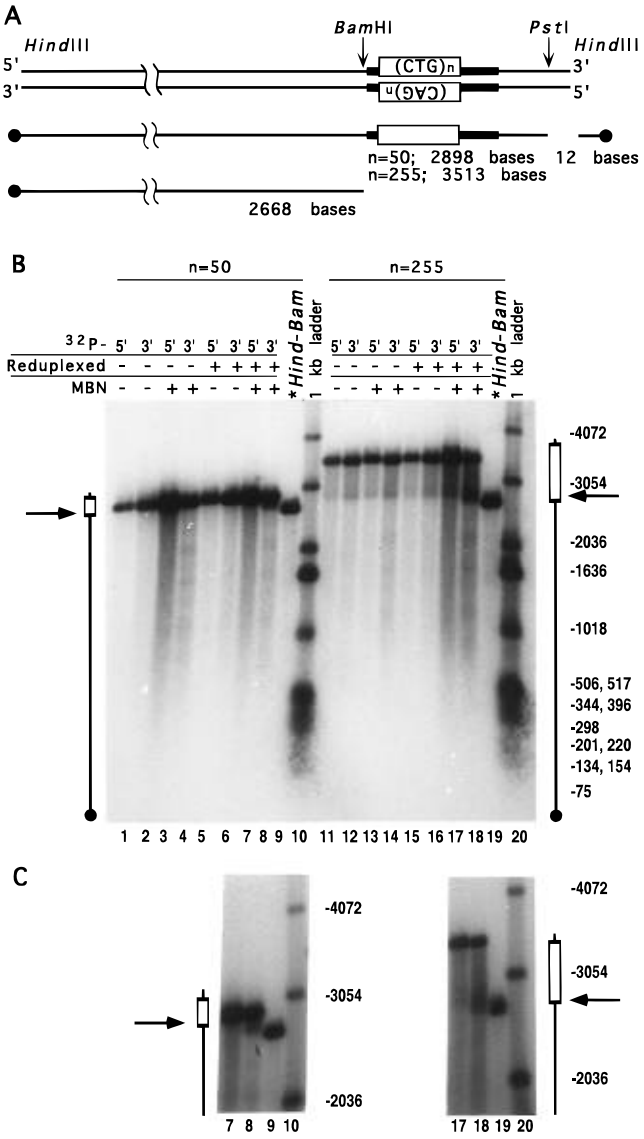

FIGURE 4: Sensitivity of $(CTG)_n$ triplet repeat tracts to mung bean nuclease. (A) Restriction map of the $(CTG)_n$-containing plasmids. Restriction sites and sizes of the various restriction fragments are indicated. Human sequences are in bold, and the ^{32}P radiolabel is indicated by a dot. For labeling with ^{32}P, $(CTG)_{50}$ and $(CTG)_{255}$ plasmids were linearized with *Hin*dIII and radiolabeled on either the 5′ or the 3′ end with T4 polynucleotide kinase or AMV reverse transcriptase, respectively. To avoid loss of the radiolabel on the 5′ single-stranded overhangs by mung bean nuclease digestion, the four base overhangs were filled with cold nucleotides using AMV reverse transcriptase. Some of the radiolabeled plasmids were reduplexed, and these and the nontreated radiolabeled linear plasmids were then digested with *Pst*I, resulting in DNAs uniquely ^{32}P-labeled at the 5′ end of the CTG-containing strand or the 3′ end of the CAG-containing strand. The 12 bp [^{32}P]*Hin*dIII/*Pst*I fragment migrated off the agarose gel. (B) Lanes 1−4 and 11−14 show duplex linear *Hin*dIII/*Pst*I fragments from the $(CTG)_{50}$- and the $(CTG)_{255}$-containing plasmids, respectively, labeled at the 5′ and the 3′ end as indicated. Samples in lanes 3 and 4 and 13 and 14 were treated with mung bean nuclease as described in Experimental Procedures. Lanes 5−8 and 15−18 show reduplexed *Hin*dIII/*Pst*I fragments from the $(CTG)_{50}$- and the $(CTG)_{255}$-containing plasmids, respectively, labeled at the 5′ and the 3′ end as indicated. Samples in lanes 7 and 8 and 17 and 18 were treated with mung bean nuclease. *Bam*HI digests of the [^{32}P]*Hin*dIII/*Pst*I-linearized plasmids and radiolabeled 1 kb ladder (Gibco) served as markers (lanes 9 and 19 and 10 and 20, respectively). Following mung bean nuclease digestion, equal amounts of radioactivity were irreversibly denatured by treatment with glyoxal (McMaster & Carmichael, 1977), loaded onto a 0.5% agarose gel, and electrophoresed at 60 V for 22 h. Gels were dried and exposed for radiography. Only the lower section of the gel is shown. (C) To increase resolution, aliquots of samples shown in lanes 7−10 and 17−20 were loaded on a 0.5% agarose gel and electrophoresed for 72 h. Only the lower section of the gel is shown. The faint band migrating faster than the band of full length *Hin*dIII/*Pst*I $(CTG)_{255}$ fragment and slower than the *Hin*dIII/*Bam*HI marker that is evident in all $(CTG)_{255}$ lanes (Figure 4B, lanes 11−18) is due to a small amount (<1%) of plasmids having only 5, 10, or 30 repeat units, natural deletion products in the $(CTG)_{255}$ plasmid preparation.

To further characterize the structural anomaly, reduplexing was performed on a supercoiled plasmid as described in the legend to Figure 1D, followed by *Hin*dIII/*Xba*I digestion and analysis by PAGE (Figure 5C). Reduplexing the supercoiled plasmid yielded a major and a minor band (Figure 5C, compare lane 1 with 2; see arrows to the right) similar to that obtained with the linearized plasmids. Thus, reduplexing of closed-circular supercoiled CGG-containing plasmids, like supercoiled CTG-containing plasmids, resulted in the appearance of novel products that are the result of alternative interactions between the two equal length complementary strands of the same double-stranded molecule.

Using a mapping strategy similar to that performed on the CTG-containing DM fragments (described in the legend to Figure 2), the alternative DNA structures of the FRAXA fragments were mapped to the $(CGG)_n$ tracts. Both the nontreated and reduplexed [^{32}P]*Hin*dIII/*Xba*I fragments were subjected to further restriction analysis using *Xho*I or *Nar*I, which cut 7 bp 3′ and 26 bp 5′, respectively, to the CGG repeat tract (Figure 6A). *Xho*I digestion of either nontreated or reduplexed DNAs produced the expected 191 bp [^{32}P]-*Hin*dIII/*Xho*I fragment (Figure 6B, lanes 2 and 5; see arrow to the lower left). *Nar*I digestion of both nontreated and reduplexed DNAs, although incomplete (as is common for this enzyme), yielded the expected full length duplex [^{32}P]-*Hin*dIII/*Nar*I fragment, as well as novel bands for the reduplexed reaction (Figure 6B, lanes 3 and 6). Following *Nar*I digestion, the novel bands migrated faster, but still much slower than duplex [^{32}P]*Hin*dIII/*Nar*I fragments (Figure 6B, compare lane 6 with 4; see brackets). These results delimit the structural anomaly to within 7 bp 3′ and 26 bp 5′ of the CGG trinucleotide repeat region.

Intermolecular Complexes Are Formed by CTG or CGG Triplet Repeats. In addition to the slowly migrating novel products, the reduplexing process of either CTG- or CGG-containing fragments also yielded some products that were unable to enter the gel (Figures 1C,E and 5B,C). These products were not formed upon isolation of either the fully duplex or novel products (see Figure 3). However, treatment of the isolated products at 85 °C resulted in these products (see Figure 3A, lanes 11 and 12; see Figure 3B, lanes 15−18). This indicates that the DNA retained in the wells is composed of the triplet repeat-containing DNA fragments. It is possible that this DNA represents a reduplex-induced multimeric complex or an aggregated form (Studier, 1969).

DISCUSSION

The process where single-stranded DNAs reanneal to become double-stranded occurs in a variety of biological processes such as transcription, recombination, repair, and replication. Mimicking this process by melting and reannealing DNA, we have detected novel DNA structures

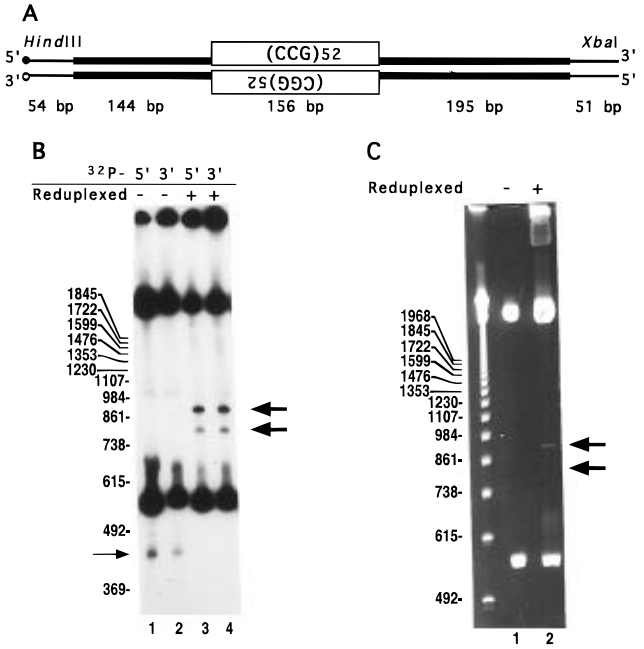

FIGURE 5: Reduplex analysis of FRAXA DNA fragments. (A) Map of the *Hin*dIII/*Xba*I restriction fragment from pFXA9.9.32 containing the FRAXA genomic sequence. Human nonrepetitive flanking sequences [positions 1−195 and 286−429, as in Eichler *et al.* (1994)] are in bold, while the thin lines are vector sequences. Positions of the 5′ and 3′ end labels are indicated by a closed and an open dot, respectively. Lengths of the triplet region and the human and vector sequences are indicated. (B) The reduplexing strategy of the *Hin*dIII-linearized plasmid was as described in the legend to Figure 1B. The secondary restriction digestion was with *Xba*I. Lanes 1 and 2 show linear duplex [^{32}P]*Hin*dIII/*Xba*I restriction-digested products, and lanes 3 and 4 show reduplexed samples. In both panels B and C, the arrows on the right side of the gel denote the position of the two major novel bands observed following reduplexing, and the migration position of the 123 bp ladder is shown. In plasmid preparations of pFXA9.9.32, there is a small (<1%) proportion of natural distinct deletion products, which differ in the number of repeat units (see arrow to the left). (C) The reduplexing strategy of the native supercoiled pFXA9.9.32 was as described in the legend to Figure 1D, where, following treatment, supercoiled samples were digested with both *Hin*dIII and *Xba*I. Restriction digests were separated on a 4% polyacrylamide gel and stained with ethidium bromide.

formed within the $(CTG)_n$ and $(CGG)_n$ tracts of the DM and FRAXA loci, respectively. This is the first report of an alternative DNA secondary structure formed in disease relevant lengths of complementary trinucleotide repeats in double-stranded DNA in the absence of supercoiling. The alternative DNA structures, which we term S-DNA, have the following characteristics. (1) The novel products are composed of complementary strands. (2) They can be formed from equal lengths of triplet repeats in the two strands. (3) The structural anomaly occurs within the repeat tract. (4) The structural anomaly is subtended by linear duplex DNA. (5) Structure formation does not require superhelical tension. (6) The structures are remarkably stable under physiological conditions. (7) In the novel structure, the CAG strand has a greater single-stranded character than its CTG complement, as indicated by the preferential mung bean nuclease reactivity of the CAG-containing strand. The results are consistent with the formation of slipped strand DNA structures (Figure 7A). These characteristics are consistent with the possibility that in human chromosomes these alternative structures may form within the triplet repeats at a particular locus.

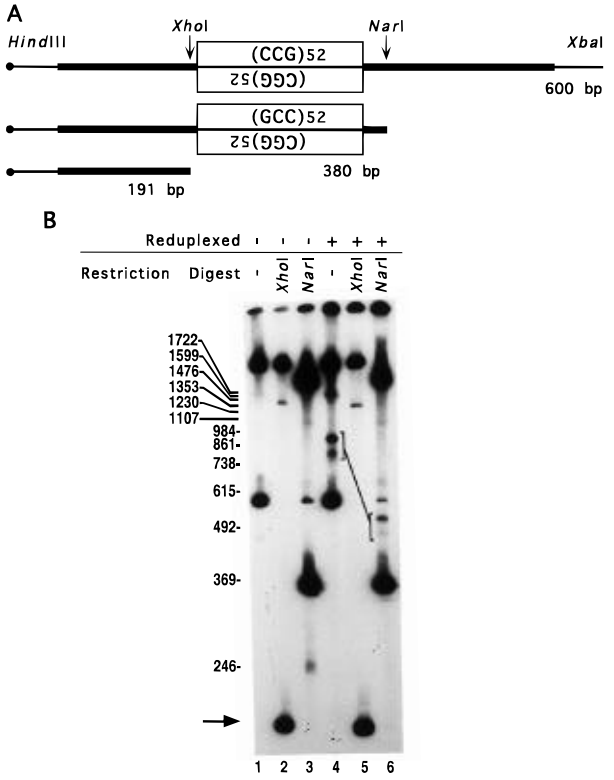

FIGURE 6: Source of the anomalous migration maps within the $(CGG)_{52}$ triplet repeat. (A) Restriction sites for *Hin*dIII, *Xba*I, *Xho*I, and *Nar*I and the sizes of the *Hin*dIII/*Xba*I, *Hin*dIII/*Xho*I, and *Hin*dIII/*Nar*I restriction-digested products are indicated. Human sequences are in bold, and the ^{32}P radiolabel is indicated by a dot. (B) Lanes 1−3 show *Xho*I and *Nar*I restriction digests of nontreated linear duplex [^{32}P]*Hin*dIII/*Xba*I fragments, and lanes 4−6 show reduplexed samples. The position of the [^{32}P]*Hin*dIII/*Xho*I product from the $(CGG)_{52}$-containing [^{32}P]*Hin*dIII/*Xba*I fragment is indicated by an arrow to the left. The reduplexed novel $(CGG)_{52}$-containing [^{32}P]*Hin*dIII/*Xba*I products and their *Nar*I-digested products are indicated by connected brackets. *Nar*I digestion is incomplete for both the nontreated and the reduplexed samples. The migration positions of the 123 bp ladder are shown.

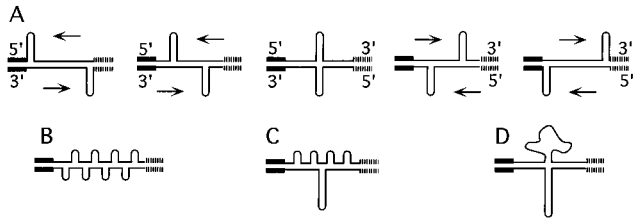

FIGURE 7: Models for S-DNA structure. (A) Due to the repetitive nature of the triplet sequences, a variety of slipped strand DNA (S-DNA) structures could form within a triplet repeat tract. A number of different structures are possible in which the position and length of the slipped out regions vary. In the structures shown, the lengths of the slipped out strands in both complementary strands are identical, although this may not be a requirement for alternative triplet repeat structures. Thin lines represent trinucleotide repeats, while thick solid or interrupted lines represent flanking, unique sequence DNA. (B) Possible S-DNA with multiple slipped out regions in both complementary strands. (C) Structure with multiple slipped out regions in one strand and a single hairpin arm in the opposite strand. (D) Possible alternative triplet repeat structures in which a hairpin has formed in only one strand. The opposite strand may be unpaired (or in a collapsed but non-base-paired structure).

By the very nature of the repeating units, which could form multiple modes of out-of-register mispairings, one would expect a high degree of structural variability in slipped strand DNA structures within a run of triplet repeats. This is in

contrast to defined conformations such as cruciforms, Z-DNA, or intramolecular triplex structures in which the alternative helix structure is clearly defined by the primary DNA sequence (Sinden, 1994). In long triplet repeat tracts, both the length of the slipped out regions as well as their positions can vary as illustrated in Figure 7A. Varying extents of slippage may occur in either the 3′ or 5′ direction (Figure 7A), and the number of slipped out regions per repeat-containing strand can also vary (Figure 7B,C). The complex pattern of multiple products indicated a high degree of conformational polymorphism following the CTG- or CGG-reduplexing reactions (Figures 1 and 5). In fact, analysis of the reduplexing reactions on a higher resolution gel resulted in greater apparent heterogeneity of the anomalous bands (C. E. Pearson and R. R. Sinden, unpublished results). Electron microscopic analysis of these DNAs confirms the heterogeneous nature of the structural variants (C. E. Pearson, Y.-H. Wang, J. D. Griffith, and R. R. Sinden, unpublished results). However, the various structural isomers did not result in an evenly distributed smear of products, as analyzed by PAGE. Rather, certain structural variants were favorable, resulting in major alternative products observed as major bands. Polymorphism of structural variants of a single sequence has been observed for single-stranded telomere DNA repeating units [reviewed in Williamson (1993)] and for long tracts of polypurine-polypyrimidine sequences in supercoiled plasmids (Shimizu *et al.*, 1990).

Short CTG, CAG, CGG, and CCG single-stranded oligonucleotides can independently assume noncomplementary duplex hairpin structures (Chen *et al.*, 1995; Gacy *et al.*, 1995; Gao *et al.*, 1995; Mitas *et al.*, 1995; Smith *et al.*, 1995). Therefore, within an otherwise duplex DNA region, the slipped out repeat tracts of each strand may form completely independent (intrastrand) hairpin structures (Figure 7A−D), while the mixed sequence DNA flanking the repeat tracts is linear duplex. However, other possibilities such as random coil, compact/collapsed (Mitchell *et al.*, 1995), or intrastrand quadruplex (Fry & Loeb, 1994) conformations cannot be ruled out. Alternatively, the slipped out regions of each strand may interact with each other, resulting in interstrand knotted structures (Studier, 1969; Broker *et al.*, 1977; Coggins & O'Prey, 1989; Coggins *et al.*, 1992).

The CAG strand of the alternative DNA structure was preferentially sensitive to mung bean nuclease digestion compared to the CTG strand. Several interpretations can explain this difference. (1) The sensitivity of the CAG strand to mung bean nuclease digestion may reflect greater single-strand character, for example, in a less stable hairpin or loop (Figure 7D). This is consistent with the demonstration that, although both the CTG and the CAG strands could form hairpins (Chen *et al.*, 1995; Gacy *et al.*, 1995; Gao *et al.*, 1995; Mitas *et al.*, 1995; Smith *et al.*, 1995), the CTG strand forms a more stable hairpin (Gacy *et al.*, 1995; Smith *et al.*, 1995). (2) The structures formed by the CAG- and CTG-containing strands may be different intrastrand structures (Figure 7C,D). (3) In a given alternative structure, the number of regions with single-stranded character on the CAG strand may be greater than that on the CTG strand (Figure 7C). The presence of more single-stranded regions on the CAG strand would result in more mung bean nuclease sensitive sites. (4) The alternative triplet repeat structures on each strand may preferentially form in the 3′ direction (as opposed to the 5′ direction) (see the right side of Figure 7A). This would result in a polar positioning on each individual strand of the single-stranded sensitive sites. On the substrates used for mung bean nuclease attack, this would result in sensitive sites of the CAG strand toward the unique 3′-*Hin*dIII ^{32}P label while sensitive sites of the CTG strand would be away from the unique 5′-*Hin*dIII ^{32}P label. If this were the case, mung bean digestion products of the CTG strand would result in a minimal loss of approximately 35−50 nucleotides, and these products would not be resolved from the full length plasmid.

It is known that these CTG-containing DM and CGG-containing FRAXA plasmids are unstable during propagation in *E. coli*, in that the number of repeats can vary among the products of a plasmid preparation (Fu *et al.*, 1991; Kremer *et al.*, 1991; Kang *et al.*, 1995). The analysis of [^{32}P]*Hin*dIII/*Eco*RI fragments containing $(CTG)_{50}$ and $(CTG)_{255}$ on a sequencing gel revealed varying numbers of triplet repeats (±5 repeats) centered at $(CTG)_{50}$ and $(CTG)_{255}$, where >85% of the DNA was of the correct length (C. E. Pearson and R. R. Sinden, unpublished results). Plasmid preparations of $(CTG)_{255}$ also contained a small amount (<1%) of plasmids having only 5, 10, or 30 repeat units. The hybridization of two strands containing different lengths of triplet repeats will produce a heteroduplex containing a bulge for short repeat length differences, or possibly a hairpin-like structure for longer repeat lengths (Chen *et al.*, 1995; Gacy *et al.*, 1995; Gao *et al.*, 1995; Mitas *et al.*, 1995; Smith *et al.*, 1995) resulting in the formation of a three-way junction. DNAs containing bulges and three-way junctions are known to migrate anomalously slowly in polyacrylamide gels (Leontis *et al.*, 1991; Welch *et al.*, 1993; Pearson *et al.*, 1994; Lilley, 1995). Several lines of evidence demonstrate that the reduplex-induced novel products from linear DNA are not entirely the result of the formation of heteroduplex molecules. First, since the novel DNA structures were formed in supercoiled plasmids, where the catenated complementary strands contain equal lengths of triplet repeats, these alternative DNA structures *could not* be the result of the formation of heteroduplex molecules (Figure 1E; Figure 5C) (no anomalous products were present until reduplexing, demonstrating that detectable heteroduplex structures were not present in the original plasmid preparation). Second, following reduplexing of a preparation of $(CTG)_{50}$ in which the lengths of triplet repeats in both strands were homogeneous, a pattern of novel products similar to those shown in Figures 1C and 3A was obtained, again indicating that the novel structures formed in the repeats do not require heterogeneity of repeat tract length between the hybridizing strands [these experiments will be described elsewhere (C. E. Pearson and R. R. Sinden, unpublished results)]. Thus, while some of the anomalous products formed following reduplexing linear DNAs are the result of heteroduplex formation [which results in reduced mobility (C. E. Pearson and R. R. Sinden, unpublished results)], many of the novel products from linear reduplexing and all of the products from supercoiled DNA are composed of two equal length complementary strands.

The results presented for mapping S-DNA to the triplet repeat region are also consistent with the formation of slipped strand DNA structures. For either the CTG- or the CGG-containing fragments, when the majority of one nonrepetitive flanking sequence was removed (by *Msp*I or *Nar*I digestion, respectively), the basic pattern of the novel products in each

case was preserved. Clearly, the triplet repeat region does not reorganize into linear duplex DNA upon removal of one nonrepetitive flanking region. This demonstrates that both nonrepetitive flanking duplex regions are unnecessary for the stability of the novel products. Apparently, interactions between complementary strands within the triplet repeat regions (i.e. possibly slipped mispairings) are sufficient to stabilize the novel structures generated by reduplexing. These structures may also be stabilized in part by intrastrand (hairpin) interactions within each strand (Chen *et al.*, 1995; Gacy *et al.*, 1995; Gao *et al.*, 1995; Mitas *et al.*, 1995; Smith *et al.*, 1995), as suggested for DNA structural intermediates involved in deletion between direct repeats flanking an inverted repeat (Trinh & Sinden, 1991).

Alternative Triplet Repeat Structures and Disease Association. Genetic analyses of DNA from individuals with myotonic dystrophy or fragile X indicate that there is a close relationship between the number of repeats and both the age of onset and severity of the symptoms, and/or increased penetrance (Fu *et al.*, 1991, 1992; Kremer *et al.*, 1991; Brook *et al.*, 1992; Mahadevan *et al.*, 1992). We observe that a larger percentage of the premutation and fully expanded repeat lengths form novel structures than the repeat lengths found in normal individuals, indicating a relationship between the number of repeats and the propensity to form alternative structures. The relationship between repeat tract size and the propensity of formation of alternative structures correlates well with the increased genetic instability of larger repeat tracts (Fu *et al.*, 1992; Brook *et al.*, 1992; Mahadevan *et al.*, 1992). To our knowledge, this is the first report of such a striking association of the *in vitro* formation of alternative DNA structures with a human genetic disease. We also observe an increase in the number of major alternative DNA structures for the reduplexed DM fragments as the number of $(CTG)_n$ repeats increased from 17 to 50 to 255. As noted above, the heterogeneous population is in agreement with the possibility for multiple modes of out-of-register base pairing. However, there are major alternative products within the population. These major products, each observed as numerous individual bands observed on the gels, likely represent a population of repeat length-dependent structures. The number of different structural isomers formed (in addition to the propensity for formation) may be dependent on the sequence homogeneity of the repeat tract. The $(CTG)_{17}$ and the $(CTG)_{50}$ sequences are pure repeats, while the $(CTG)_{255}$ has four ACT interruptions, $[(CTG)_{27}ACT(CTG)_{40}ACT(CTG)_{38}ACT(CTG)_{40}ACT(CTG)_{106\pm5}]$. It is possible that the four ACT interruptions within the $(CTG)_{255}$ tract are responsible for the formation of the three major anomalous products, compared to the one major heterogeneous band present in the $(CTG)_{50}$ tract. Sequence interruptions may favor or exclude the formation of certain structural isomers. The effect of sequence interruptions within the triplet repeats on the alternative DNA structures formed remains to be determined.

The novel DM CTG-containing structures are similar to the novel FRAXA CGG-containing structures to the extent that each is contained within the repeat tracts and they can be composed of complementary strands of equal lengths. The pattern and percentage of novel reduplex-induced bands for the $(CGG)_{52}$ tract from a FRAXA locus was different from that observed for the reduplexed DM $(CTG)_{50}$ tract. Two major products, presumably representing two major alternative structures, were formed within the $(CGG)_{52}$ tract, while a single major heterogeneous band was observed for $(CTG)_{50}$. The lower percentage of novel products and fewer structural isomers may be related to the AGG interruptions within the FRAXA sequence $[(CGG)_9AGG(CGG)_9AGG(CGG)_{32}]$. AGG interruptions within the CGG tracts of the *FMR-1* locus confer increased genetic stability to the repeat tract, such that interrupted tracts are less likely to undergo expansions (Eichler *et al.*, 1994; Kunst & Warren, 1994; Snow *et al.*, 1994). Eichler *et al.* (1994) established 34−38 pure repeats as the threshold above which increased instability occurs. The 32 pure repeats in the pFXA9.9.32 clone is just below this threshold. Similarly, CAT interruptions in the CAG repeat of the *SCA1* locus also seem to provide increased genetic stability to those repeats (Chung *et al.*, 1993). Interruptions, and their distribution within the trinucleotide repeats, may strongly influence the propensity of alternative structure formation and the particular structural isomers formed.

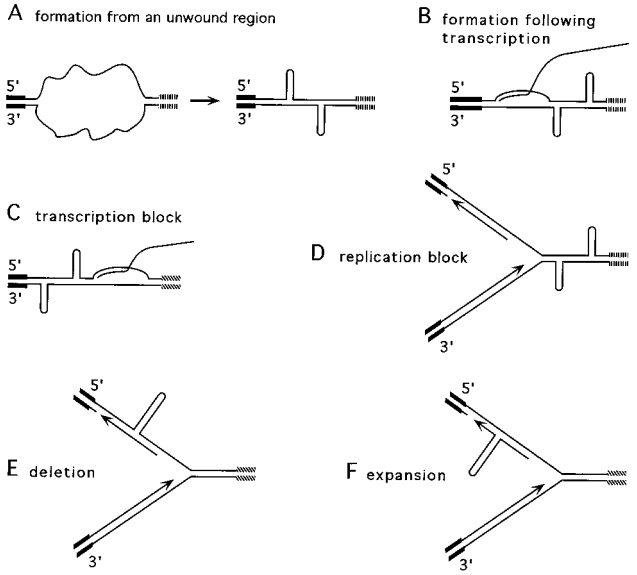

FIGURE 8: Models for the formation and biological effects of S-DNA. (A) Unwinding of a triplet repeat tract and reannealing may result in the formation of alternative triplet repeat slipped strand (S-DNA) structures. (B) S-DNA might form during the reannealing of triplet repeats following transcription. (C) S-DNA may act as a block to transcription, resulting in an alteration of gene expression from triplet repeat-associated genes. (D) S-DNA may act as a block/pause to replication fork progression. (E) Hairpin structure formation in the template strand during replication of the lagging strand would result in deletion of triplet repeats. (F) Hairpin structure formation in the nascent strand during replication of the lagging strand would result in expansion of triplet repeats.

Potential Pathways to Formation of Alternative Triplet Repeat Structures. There may be a number of ways in which the novel reduplex-induced structures can form in human cells. Conditions conducive to formation may occur when triplet repeat expansion occurs. The formation of alternative triplet repeat structures may require large regions of unwound DNA (Figure 8A). Very long regions (0.1−60 kb) of single-stranded DNA have been detected in a variety of eukaryotic organisms, including human cells (Henson, 1978; Bjursell *et al.*, 1979, and references therein). In many instances, these unwound regions were associated with DNA replication (Micheli *et al.*, 1982). Long unwound regions not associated with replication have also been reported (Bjursell *et al.*, 1979, and references therein; Henson, 1978; Wortzman & Baker,

1981). Single-stranded regions have also been associated with recombination, early development (Micheli *et al.*, 1982), and meiosis (Klein & Byers, 1978). In one instance, where these single-stranded regions were localized, the sequences were composed of tracts of short tandemly repeating units (Wortzman & Baker, 1981). In the cell, reannealing following the unwinding of the premutation or full length triplet repeats may give rise to the novel DNA secondary structures described here (Figure 8A).

These alternative triplet repeat structures might also form during transcription, replication, or recombination. It is possible that following passage of a transcription complex, which transiently produces single-stranded DNA, alternative triplet repeat structures could form during reannealing (Figure 8B). Alternative triplet repeat structures might also form at a replication fork either behind the replication complex between the nascent and template strands and/or ahead of the replication machinery between the two template strands (Figure 8D−F). The occurrence of single-stranded regions during genetic recombination could also lead to the formation of alternative triplet repeat structures.

Alternative DNA secondary structure formation within triplet repeats may be greatly facilitated by the presence of a free end within the repeat tract. Free ends are known to enhance the duplex-to-hairpin (Scheffler *et al.*, 1970; Xodo *et al.*, 1989, and references therein) and duplex-to-random coil (Pouwels *et al.*, 1968; Rush & Warner, 1970; Scheffler *et al.*, 1970) transitions, as well as the reverse reactions. Free ends within the repeat tract could occur during DNA replication, specifically during RNA priming by polα-primase on the unwound template. Multiple rounds of RNA priming are required for discontinuous synthesis on the lagging strand of replication forks (Burhans *et al.*, 1990). As suggested (Eichler *et al.*, 1994; Kunst & Warren, 1994; Snow *et al.*, 1994; Richards & Sutherland, 1994), it may not be pure coincidence that the size of the premutation repeat tracts (50−60 repeat units) is close to that of an Okazaki fragment (150−180 bp) (Burhans *et al.*, 1990). The longer the repeat tract, the more likely that the initiation of Okazaki fragment synthesis occurs within the repeat tract and the higher the probability of formation of an alternative DNA structure.

The CTG reduplex-induced alternative structures are remarkably stable under physiological conditions. As mentioned above, each reduplexing reaction results in a broad spectrum of structural isomers and there does not appear to be a rapid interconversion between them. Branch migration within a given duplex would presumably be required for the interconversion between the various alternative structures or for their removal. Recent evidence indicates that branch migration, long thought to occur spontaneously at a rapid rate, occurs very slowly (Panyutin & Hsieh, 1994). Moreover, proteins are typically involved in this process (Shiba *et al.*, 1991). Additionally, mismatched bases can severely inhibit spontaneous branch migration (Panyutin & Hsieh, 1993). Unpaired nucleotides can increase the stability of three-way DNA junctions (Leontis *et al.*, 1991; Welch *et al.*, 1993), suggesting that mismatched base pairings, expected to occur in S-DNA, may inhibit interconversion between the various alternative structures or stabilize them from branch migration back into the linear duplex form. If the alternative triplet repeat structures described here are in fact slipped strand DNA structures, the apparently slow interconversion between them suggests that a cellular protein may be required to drive branch migration, thus facilitating the removal of slipped strand DNA structures in tandemly repeated tracts.

Possible Involvement of Alternative Triplet Repeat Structures in the Genetic Instability of Triplet Repeats in Human Diseases. A unique and puzzling feature of human triplet repeat diseases is the propensity of the CTG and CGG repeats to expand rather than contract (Sinden & Wells, 1992; Sutherland & Richards, 1995). Although deletions of triplet repeats in humans have been observed, they are rare (O'Hoy *et al*., 1993; Graaff *et al*., 1995). In contrast, deletions are the preferential event in bacterial cells (Jaworski *et al*., 1995; Kang *et al*., 1995). Primer−template misalignment during DNA replication is believed to be a major source of genome instability, especially at repetitive DNA elements [reviewed by Sinden and Wells (1992) and Wells and Sinden (1993)]. The alternative triplet repeat structures described here do not require unrestrained superhelical tension for formation and can form in linear (relaxed) DNA fragments. Consequently, these structures could form at a replication fork either behind the replication complex between the nascent and template strands and/or ahead of the replication machinery between the two template strands (Figure 8D−F). Slippage of the DNA polymerase and the nascent strand forward on the template strand would result in the formation of structures behind the replication fork, with an excess of repeats on the template strand (Figure 8E), giving rise to deletion products. Slippage of the DNA polymerase and the nascent (primer) strand backward on the template strand would result in the formation of structures behind the replication fork, with an excess of repeats on the nascent strand (Figure 8F). Such a situation would give rise to expansion products. Alternative triplet repeat structures ahead of the replication complex may act as barriers to the progression of the replication fork and lead to delayed replication or reiterative synthesis (Figure 8D), which could explain the occurrence of large expansions in a single genetic step (Sinden & Wells, 1992).

If repeat instability is due to slippage during DNA replication, then both deletion and expansion will involve replicated duplex DNAs containing a stretch of mismatched repeats (a heteroduplex region) (Figure 8E,F). It is not known whether heteroduplex or homoduplex trinucleotide repeat secondary structures are recognized by the human mismatch repair system. It is possible that the alternative triplet repeat structures can be stably maintained through the completion of the S, G2, and M phases until the next round of replication. The efficiency and specificity of mismatch or single-stranded loop correction varies with the mismatched or looped template (Brown & Jiricny, 1988; Umar *et al.*, 1994), as well as the sequence context surrounding the mismatch (Umar *et al.*, 1994; Jones *et al.*, 1987). An extreme case of the variable repairability of mismatched DNAs is fully base-paired palindromes. Heteroduplex DNAs containing palindromes, which form three-way junctions, inhibit mismatch repair in yeast (Nag & Petes, 1991). In fact, palindromic sequences in heteroduplex DNA are stable in both yeast (Nag & Petes, 1991; Nag *et al.*, 1989) and mammalian cells (Bollag *et al.*, 1992). Alternative trinucleotide repeat structures, like palindromic structures, may be poorly recognized by the cellular mismatch repair systems. If, on the other hand, alternative trinucleotide repeat structures are efficiently recognized and corrected, the repeat tracts

may be constantly undergoing the processes of structure formation and mismatch repair, which might increase the possibility for replication errors during repair. Although repeat expansion may involve slippage during DNA replication, there are other proposed mechanisms such as recombination gap repair (Jansen *et al.*, 1994) and unequal crossover between sister chromatids (Smith, 1976; Nussbaum *et al.*, 1986). These models may also involve alternative DNA structures within the repeats, such as those described here.

While the S-DNA alternative DNA structures described here, presumably slipped strand DNA, may be involved in repeat instability, other possible biological functions of the alternative structures may exist. DNA slippage at tandem repeats has been hypothesized to participate in many different biological processes, including genetic evolution, mutagenesis via small insertions and/or deletions, as well as human genetic diseases involving deletions and/or expansions. These structures may be recognized by cellular proteins in a structure-specific fashion (Richards *et al.*, 1993; Pearson *et al.*, 1994, 1995; Chen *et al.*, 1995; Yano-Yanagisawa *et al.*, 1995) and influence transcription or replication. It is possible that long tracts of triplet repeats containing alternative structures alter chromosomal structure (Wang *et al.*, 1994; Otten & Tapscott, 1995) in such a way that there is altered methylation (Hansen *et al.*, 1992; Hornstra *et al.*, 1993; Smith *et al.*, 1994; Chen *et al.*, 1995), delayed replication (Hansen *et al.*, 1993) (Figure 8D), decreased transcription (Fu *et al.*, 1993) (Figure 8C), or chromosomal fragility.

The identification of these novel alternative DNA structures now allows an analysis of their possible involvement in DNA replication, mismatch repair, and recombination which may lead to the expansion of triplet repeats in human genetic disease. The establishment of an assay to detect these novel alternative DNA structures is an important advance that should prove useful in determination of the mechanism(s) of their formation and the possible role of these unusual structures in the instability of repeated DNAs.

ACKNOWLEDGMENT

We are indebted to David L. Nelson and Evan E. Eichler of Baylor College of Medicine for providing the FRAXA clone pFXA9.9.32. We thank David Nelson, Evan E. Eichler, Jan Klysik, Paul D. Chastain, M. Angie Parsons, Jennifer L. Jennings, and especially Vladimir Potaman for stimulating discussions and critical comments on the manuscript.

REFERENCES

Ashizawa, T., Dubel, J. J., Dunne, P. W., Dunne, C. J., Fu, Y.-H., Pizzuti, A., Caskey, C. T., Boerwinkle, E., Perryman, M. B., Epstein, H. F., & Hejtmancik, J. F. (1992) *Neurology 42*, 1877−1883.

Bjursell, G., Gussander, E., & Lindahl, T. (1979) *Nature 280*, 420−423.

Bollag, R. J., Elwood, D. R., Tobin, E. D., Godwin, A. R., & Liskay, R. M. (1992) *Mol. Cell. Biol. 12*, 1546−1552.

Broker, T. R., Soll, L., & Chow, L. T. (1977) *J. Mol. Biol. 113*, 579−589.

Brook, J. D., McCurrach, M. E., Harley, H. G., Buckler, A. J., Church, D., Aburatani, H., Hunter, K., Stanton, V. P., Thirion, J. P., Hudson, T., Sohn, R., Zemelman, B., Snell, R. G., Rundle, S. A., Crow, S., Davies, J., Shelbourne, P., Buxton, J., Jones, C., Juvonen, V., Johnson, K., Harper, P. S., Shaw, D. J., & Housman, D. E. (1992) *Cell 68*, 799−808.

Brown, T. C., & Jiricny, J. (1988) *Cell 54*, 705−711.

Burhans, W. C., Vassilev, L. T., Caddle, M. S., Heintz, N. H., & DePamphilis, M. L. (1990) *Cell 62*, 955−965.

Chastain, P. D., Eichler, E. E., Kang, S., Nelson, D. L., Levene, S. D., & Sinden, R. R. (1995) *Biochemistry 34*, 16125−16131.

Chen, X., Mariappan, S. V. S., Catasti, P., Ratliff, R., Moyzis, R. K., Laayoun, A., Smith, S. S., Bradbury, E. M., & Gupta, G. (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 5199−5203.

Chong, S. S., Eichler, E. E., Nelson, D. L., & Hughes, M. R. (1994) *Am. J. Med. Genet. 51*, 522−526.

Chung, M., Ranum, L. P. W., Duvick, L. A., Servadio, A., Zoghbi, H. Y., & Orr, H. T. (1993) *Nat. Genet. 5*, 254−258.

Coggins, L. W., & O'Prey, M. (1989) *Nucleic Acids Res. 17*, 7417−7426.

Coggins, L. W., O'Prey, M., & Akhter, S. (1992) *J. Mol. Biol. 121*, 279−285.

Eichler, E. E., Holden, J. J. A., Popovich, B. W., Reiss, A. L., Snow, K., Thibodeau, S. N., Richards, C. S., Ward, P. A., & Nelson, D. L. (1994) *Nat. Genet. 8*, 88−94.

Fry, M., & Loeb, L. A. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 4950−4954.

Fu, F.-H., Friedman, D. L., Richards, S., Pearlman, J. A., Gibbs, R. A., Pizzuti, A., Ashizawa, T., Perryman, M. B., Scarlato, G., Fenwick, R. G., Jr., & Caskey, C. T. (1993) *Science 260*, 235−238.

Fu, Y.-H., Kuhl, D. P. A., Pizzuti, A., Pieretti, M., Sutcliffe, J. S., Richards, S., Verkerk, A. J. M. H., Holden, J. J. A., Fenwick, R. G., Jr., Warren, S. T., Oostra, B. A., Nelson, D. L., & Caskey, C. T. (1991) *Cell 67*, 1047−1058.

Fu, Y.-H., Pizzuti, A., Fenwick, R. G., Jr., King, J., Rajnarayan, S., Dunne, P. W., Dubel, J., Nasser, G. A., Ashizawa, T., DeJong, P., Wieringa, B., Korneluk, R., Perryman, M. B., Epstein, H. F., & Caskey, C. T. (1992) *Science 255*, 1256−1258.

Gacy, A. M., Goellner, G., Juranic, N., Macura, S., & McMurry, C. T. (1995) *Cell 81*, 533−540.

Gao, X., Huang, X., Smith, G. K., Zheng, M., & Liu, H. (1995) *J. Am. Chem. Soc. 117*, 8883−8884.

Graff, E., Rouillard, P., Willems, P. J., Smits, A. P. T., Rousseau, F., & Oostra, B. A. (1995) *Hum. Mol. Genet. 4*, 45−49.

Hansen, R. S., Gartler, S. M., Scott, C. R., Chen, S.-H., & Laird, C. D. (1992) *Hum. Mol. Genet. 1*, 571−578.

Hansen, R. S., Canfield, T. K., Lamb, M. M., Gartler, S. M., & Laird, C. D. (1993) *Cell 73*, 1403−1409.

Henson, P. (1978) *J. Mol. Biol. 119*, 487−506.

Hornstra, I. K., Nelson, D. L., Warren, S. T., & Yang, T. P. (1993) *Hum. Mol. Genet. 2*, 1659−1665.

Jansen, G., Willems, P., Coerwinkel, M., Nillesen, W., Smeets, H., Vits, L., Howeler, C., Brunner, H., & Wieringa, B. (1994) *Am. J. Hum. Genet. 54*, 575−585.

Jaworski, A., Rosche, W. A., Gellibolian, R., Kang, S., Shimizu, M., Bowater, R. P., Sinden, R. R., & Wells, R. D. (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 11019−11023.

Jones, C., Penny, L., Mattina, T., Yu, S., Baker, E., Voullaire, L., Langdon, W. Y., Sutherland, G. R., Richards, R. I., & Tunnacliffe, A. (1995) *Nature 376*, 145−149.

Jones, M., Wagner, R., & Radman, M. (1987) *Genetics 115*, 605−610.

Kang, S., Jaworski, A., Ohshima, K., & Wells, R. D. (1995) *Nat. Genet. 10*, 213−218.

Klein, H. L., & Byers, B. (1978) *J. Bacteriol. 134*, 629−635.

Knight, S. J. L., Flannery, A. V., Hirst, M. C., Campbell, L., Christodoulou, Z., Phelps, S. R., Pointon, J., Middleton-Price, H. R., Oostra, B. A., & Davies, K. E. (1993) *Cell 74*, 127−134.

Kochel, T. J., & Sinden, R. R. (1988) *BioTechniques 6*, 532−543.

Kohne, D. E., Levison, S. A., & Byers, M. J. (1977) *Biochemistry 16*, 5329−5341.

Kohwi, Y., Wang, H., & Kohwi-Shigematsu, T. (1993) *Nucleic Acids Res. 21*, 5651−5655.

Kremer, E. J., Pritchard, M., Lynch, M., Yu, S., Holman, K., Baker, E., Warren, S. T., Schlessinger, D., Sutherland, G. R., & Richards, R. I. (1991) *Science 252*, 1711−1714.

Kunst, C. B., & Warren, S. T. (1994) *Cell 77*, 853−861.

Leontis, N. B., Kwok, W., & Newman, J. S. (1991) *Nucleic Acids Res. 19*, 759−766.

Lilley, D. M. J. (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 7140−7142.
Mahadevan, M., Tsilfidis, C., Sabourin, L., Shutler, G., Amemiya, C., Jansen, G., Neville, C., Narang, M., Barcelo, J., O'Hoy, K., Leblond, S., Earle-McDonald, J., de Jons, P. J., Wieringa, B., & Korneluk, R. G. (1992) *Science 255*, 1253−1255.
McMaster, G. K., & Carmichael, G. G. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 4835−4838.
Micheli, G., Baldari, C. T., Carri, M. T., DiCello, G., & Buongiorno-Nardelli, M. (1982) *Exp. Cell Res. 137*, 127−140.
Mitas, M., Yu, A., Kamp, T. J., Chambers, E. J., & Haworth, I. S. (1995) *Nucleic Acids Res. 23*, 1050−1059.
Mitchell, J. E., Newbury, S. F., & McClellan, J. A. (1995) *Nucleic Acids Res. 23*, 1876−1881.
Nag, D. K., & Petes, T. D. (1991) *Genetics 129*, 669−673.
Nag, D. K., White, M. A., & Petes, T. D. (1989) *Nature 340*, 318−320.
Nancarrow, J. K., Kremer, E., Holman, K., Eyre, H., Doggett, N. A., LePasiler, D., Callen, D. F., Sutherland, G. R., & Richards, R. I. (1994) *Science 264*, 1938−1941.
Nussbaum, R. L., Airhart, S. D., & Ledbetter, D. H. (1986) *Am. J. Hum. Genet. 23*, 715−721.
O'Hoy, K. L., Tsilfidis, C., Mahadevan, M. S., Neville, C. E., Barcelo, J., Hunter, A. G. W., & Korneluk, A. G. (1993) *Science 259*, 809−812.
Otten, A. D., & Tapscott, S. J. (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 5465−5469.
Panyutin, I. G., & Hsieh, P. (1993) *J. Mol. Biol. 230*, 413−424.
Panyutin, I. G., & Hsieh, P. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 2021−2025.
Parish, J. E., Oostra, B. A., Berkerk, A. J. M. H., Richards, C. S., Reynolds, J., Spikes, A. S., Shaffer, L. G., & Nelson, D. L. (1994) *Nat. Genet. 8*, 229−235.
Pearson, C. E., Ruiz, M. T., Price, G. B., & Zannis-Hadjopoulos, M. (1994) *Biochemistry 33*, 14185−14196.
Pearson, C. E., Zannis-Hadjopoulos, M., Price, G. B., & Zorbas, H. (1995) *EMBO J. 14*, 1571−1580.
Pouwels, P. H., Knijnenburg, C. M., van Rotterdam, J., & Cohen, J. A. (1968) *J. Mol. Biol. 32*, 169−182.
Richards, R. I., & Sutherland, G. R. (1992) *Nat. Genet. 1*, 7−9.
Richards, R. I., & Sutherland, G. R. (1994) *Nat. Genet. 6*, 114−116.
Richards, R. I., Holman, K., Yu, S., & Sutherland, G. R. (1993) *Hum. Mol. Genet. 2*, 1429−1435.
Rush, M. G., & Warner, R. C. (1970) *J. Biol. Chem. 245*, 2704−2708.
Scheffler, I. E., Elson, E. L., & Baldwin, R. L. (1970) *J. Mol. Biol. 48*, 145−147.
Shiba, T., Iwasaki, H., Nakata, A., & Shinagawa, H. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 8445−8449.
Shimizu, M., Hanvey, J. C., & Wells, R. D. (1990) *Biochemistry 29*, 4704−4713.
Sinden, R. R. (1994) *DNA Structure and Function*, Academic Press, San Diego.
Sinden, R. R., & Wells, R. D. (1992) *Curr. Opin. Biotechnol. 3*, 612−622.
Smith, G. K., Jie, J., Fox, G. E., & Gao, X. (1995) *Nucleic Acids Res. 23*, 4303−4311.
Smith, G. P. (1976) *Science 191*, 528−535.
Smith, S. S., Laayoun, A., Lingeman, R. G., Baker, D. J., & Riley, J. (1994) *J. Mol. Biol. 243*, 143−151.
Snow, K., Tester, D. J., Kruckeberg, K. E., Schaid, D. J., & Thibodeau, S. N. (1994) *Hum. Mol. Genet. 9*, 1543−1551.
Studier, F. W. (1969) *J. Mol. Biol. 41*, 199−209.
Sutherland, G. R., & Richards, R. I. (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 3636−3641.
Trinh, T. Q., & Sinden, R. R. (1991) *Nature 352*, 544−547.
Umar, A., Boyer, J. C., & Kunkel, T. A. (1994) *Science 266*, 814−816.
Umek, R. M., & Kowalski, D. (1988) *Cell 52*, 559−567.
Wang, Y.-H., Amirhaeri, S., Kang, S., Wells, R. D., & Griffith, J. D. (1994) *Science 265*, 669−671.
Welch, J. B., Duckett, D. R., & Lilley, D. M. (1993) *Nucleic Acids Res. 21*, 4548−4555.
Wells, R. D., & Sinden, R. R. (1993) in *Genome Analysis Volume 7: Genome Rearrangement and Stability* (Davies, K., & Warren, S., Eds.) pp 107−138, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Williamson, J. R. (1993) *Curr. Biol. 3*, 357−362.
Wortzman, M. S., & Baker, R. F. (1981) *Science 211*, 588−590.
Xodo, L. E., Manzini, G., Quadrifoglio, F., Yathindra, N., Van der Marel, G. A., & Van Boom, J. H. (1989) *J. Mol. Biol. 205*, 777−781.
Yano-Yanagisawa, H., Li, Y., Wang, H., & Kohwi, Y. (1995) *Nucleic Acids Res. 23*, 2654−2660.

BI9601013