816–823 Nucleic Acids Research, 1998, Vol. 26, No. 3

Structural analysis of slipped-strand DNA (S-DNA) formed in $(CTG)_n \cdot (CAG)_n$ repeats from the myotonic dystrophy locus

Christopher E. Pearson*, Yuh-Hwa Wang[1,+], Jack D. Griffith[1] and Richard R. Sinden

Center for Genome Research, Institute of Biosciences and Technology, Department of Biochemistry and Biophysics, Texas A&M University, Houston, TX 77030-3303, USA and [1]Lineberger Cancer Research Center, University of North Carolina, Chapel Hill, NC 27599-7295, USA

Received August 29, 1997; Revised and Accepted November 12, 1997

ABSTRACT

The mechanism of disease-associated trinucleotide repeat length variation may involve slippage of the triplet-containing strand at the replication fork, generating a slipped-strand DNA structure. We recently reported formation *in vitro* of slipped-strand DNA (S-DNA) structures when DNAs containing triplet repeat blocks of myotonic dystrophy or fragile X diseases were melted and allowed to reanneal to form duplexes. Here additional evidence is presented that is consistent with the existence of S-DNA structures. We demonstrate that S-DNA structures can form between two complementary strands containing equal numbers of repeats. In addition, we show that both the propensity for S-DNA formation and the structural complexity of S-DNAs formed increase with increasing repeat length. S-DNA structures were also analyzed by electron microscopy, confirming that the two strands are slipped out of register with respect to each other and confirming the structural polymorphism expected within long tracts of trinucleotide repeats. For $(CTG)_{50} \cdot (CAG)_{50}$ two distinct populations of slipped structures have been identified: those involving ≤10 repeats per slippage, which appear as bent/kinked DNA molecules, and those involving >10 repeats, which have multiple loops or hairpins indicative of complex alternative DNA secondary structures.

INTRODUCTION

Thirteen human genetic diseases, including myotonic dystrophy (DM) and fragile X, have been associated with mutations involving expansion of either $(CTG)_n \cdot (CAG)_n$ or $(CGG)_n \cdot (CCG)_n$ tracts (for a review see 1). The product of an expansion mutation is more likely to undergo subsequent mutation than the original substrate, hence this new class of genetic alterations are known as 'dynamic mutations' (2). Unlike the genome-wide repeat instability observed in certain cancers (3), which manifests as small changes in the length of many repeat tracts (±1–5 repeat units), the triplet disease-associated repeat instability is loci, repeat length and sequence specific. In some cases the genetic changes involve large expansions in repeat length (>2000 repeats). Chromosomal context and DNA secondary structure may contribute to triplet disease-associated instability.

The mechanism responsible for these dynamic mutations is unknown, but may involve sequence-specific non-Watson–Crick DNA structures, such as slipped-strand DNA (S-DNA). The triplet repeats associated with expansion and disease, $(CTG)_n \cdot (CAG)_n$, $(CGG)_n \cdot (CCG)_n$ and $(AAG)_n \cdot (CTT)_n$, can form alternative DNA structures (4–6). S-DNA could cause errors during DNA replication, repair, recombination or transcription (4,5). An understanding of the molecular mechanism of the expansion process is necessary for a complete understanding of the disease etiology.

We have recently reported the formation of S-DNA structures within $(CTG)_n \cdot (CAG)_n$ and $(CGG)_n \cdot (CCG)_n$ repeats following denaturation and renaturation (5) and these S-DNAs have properties consistent with those expected of slipped-strand structures. The anomalous migration of S-DNA in polyacrylamide gels suggests that renaturation of the repeats results in formation of slip-outs, while the non-repetitive DNA flanking the structural anomaly exists in a duplex linear B-DNA conformation. S-DNA formation does not require supercoil tension and S-DNA structures are remarkably stable at physiological salt concentrations. The CAG strand in S-DNA is more susceptible to mung bean nuclease attack, relative to its CTG complement, indicating greater single-stranded character of that strand. These initial data are consistent with a structure for S-DNA that involves out of register base pairing in the repeat tract (5).

In this report we have addressed the question of slipped structure formation between two complementary, equal length repeat-containing strands. We have also analyzed the S-DNA structures by electron microscopy (EM). Visualization of S-DNAs shows the presence of bends and loops that are localized to the triplet repeat tract and reveals a reduced contour length relative to linear duplex DNAs. The EM analysis confirms both the structural heterogeneity and the differential electrophoretic migration of various S-DNA isomers.

*To whom correspondence should be addressed. Tel: +1 713 677 7676; Fax: +1 713 677 7689; Email: cpearson@ibt.tamu.edu

[+]Present address: Department of Biochemistry Robert Wood Johnson Medical School, University of Medicine and Dentistry of New Jersey, Piscataway, NJ 08855, USA

MATERIALS AND METHODS

Plasmid/DNA treatments and denaturation/renaturation reactions

The genomic clones from the human DM locus (positions 357–433, as in 7) containing the $(CTG)_n \cdot (CAG)_n$ repeat, where $n = 30$, $n = 50$ and $n = 255$, have been described (4,5,8). The $n = 30$ and $n = 50$ repeat tracts are free of sequence interruptions, while the $n = 255$ tract contains four ACT interruptions resulting in a tract of $(CTG)_{27}ACT(CTG)_{40}ACT(CTG)_{38}ACT(CTG)_{40}ACT(CTG)_{106\pm5}$. Plasmids were prepared using detergent lysis as previously described (5). Plasmids were treated with DNase-free RNases A and T1 (Sigma), phenol extracted, purified twice by isopycnic centrifugation and stored in TE (10 mM Tris, 1 mM EDTA, pH 7.6) at –20°C. All precautions were taken to avoid deletions during plasmid propagation in *Escherichia coli* (5,9).

End-labeling of the 5′- or 3′-ends was performed using T4 polynucleotide kinase (USB) or AMV reverse transcriptase (USB) respectively as described (4,5,10,11). All denaturation/renaturation reactions were performed as described and precautionary measures were taken to avoid sample dehydration (4,5,10,11). All restriction enzymes were purchased from New England Biolabs and reactions were performed as specified by the manufacturer. The slipped-strand structures were excised from polyacrylamide gels and the DNA electroeluted as described (5,10,11).

Electrophoresis

Polyacrylamide gels (4%, 13 × 9 × 0.15 cm, 40:2 acrylamide:bis-acrylamide) were cast in TBE (90 mM Tris, 90 mM borate, 2.5 mM EDTA, pH 8.3) and subjected to a constant voltage (150 V, 10–12 V/cm) at room temperature, unless otherwise stated. Gels were then stained with ethidium bromide, photographed and/or dried on Whatman paper and exposed to radiographic film (Kodak) in the presence of an intensifying screen at –70°C or a PhosphorImager screen (Molecular Dynamics). The amount of S-DNA formed was measured by densitometric analyses as a percentage of the total population of repeat-containing molecules. All quantitation was done on 3–10 experiments by PhosphorImager analysis using ImageQuant software (Molecular Dynamics). All relative migrations and base pair estimations were calculated with respect to the migration of a 123 bp ladder (Gibco BRL) as previously described (8).

Electron microscopy

DNA samples were prepared and EM was performed as described (12). In brief, the samples were mixed with a buffer containing 0.15 M NaCl, 1 mM $MgCl_2$ and 2 mM spermidine, adsorbed to glow-charged thin carbon films, washed with a water/graded ethanol series and rotary shadow cast with tungsten. Samples were examined using a Phillips CM12 electron microscope. Micrographs are shown in reverse contrast. A Nikon multiformat film scanner attached to a Macintosh computer and Adobe Photoshop was used to form montages of the images.

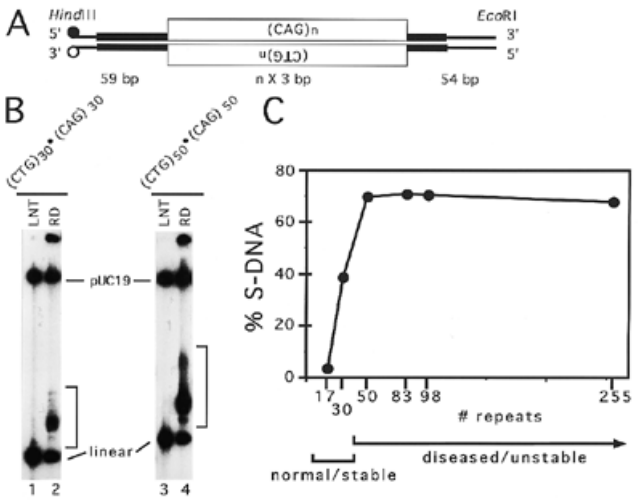

Figure 1. Denaturing and renaturing DM DNA fragments containing $(CTG)_n \cdot (CAG)_n$ repeats results in slowly migrating DNAs with respect to DNAs that have never been denatured/renatured. (**A**) Map of the DM $(CTG)_n \cdot (CAG)_n$ *Hin*dIII–*Eco*RI fragment. Human non-repetitive flanking sequences (positions 357–375 and 391–433, as in 7) are in bold. The thin lines represent plasmid vector sequences. (**B**) $(CTG)_n \cdot (CAG)_n$-containing fragments with $n = 30$ and $n = 50$ are shown. Linear non-treated (LNT) controls (lanes 1 and 3) and reanealed (RD) (lanes 2 and 4) ^{32}P-labeled *Hin*dIII–*Eco*RI restriction digestion products were separated on a 4% polyacrylamide gel, dried and exposed for autoradiography. The brackets indicate the distribution of total radioactivity migrating anomalously following reannealing. The linear duplex DNAs are also indicated. The pattern of slow migrating products was indistinguishable when either the CTG or the CAG strands were radiolabeled, indicating that the new products were composed of both CTG- and CAG-containing strands. (**C**) Propensity of S-DNA formation as a function of repeat length. The percentages were measured as described in Materials and Methods.

RESULTS

Formation of S-DNA in $(CTG)_n \cdot (CAG)_n$ repeats

The denaturation and renaturation of DNAs containing $(CTG)_n \cdot (CAG)_n$ or $(CGG)_n \cdot (CCG)_n$ repeats results in S-DNA structures that migrate slowly on polyacrylamide gels (5). Plasmids containing between 17 and 255 (CTG)·(CAG) repeats were linearized by *Hin*dIII digestion, ^{32}P-end-labeled and denatured and renatured as described (5). Following denaturation and renaturation the DNA was digested with *Eco*RI to liberate the DM $(CTG)_n \cdot (CAG)_n$ inserts (Fig. 1A). The polyacrylamide gel electrophoresis patterns revealed a set of anomalously migrating products in the denatured/renatured plasmids not present in the non-treated plasmids (Fig. 1B). DNA with mismatches, multiple unpaired bases or three- or four-way junctions migrates anomalously slowly in polyacrylamide gels (10,11,13–16) due to introduction of bends into the DNA molecule. The slow mobility of DNA following reannealing is consistent with the presence of slipped repeats and three-way junctions.

Reannealing DNA containing $(CTG)_{30} \cdot (CAG)_{30}$ or $(CTG)_{50} \cdot (CAG)_{50}$ resulted in a series of closely spaced slowly migrating products (Fig. 1B, compare lanes 1 with 2 and 3 with 4). There were two distinct populations of slowly migrating products for $(CTG)_{50} \cdot (CAG)_{50}$. These populations had apparent size ratios, R, where $R = bp_{apparent}/bp_{actual}$, of 1.25 and 2.0 (see Fig. 3 or 4 for positions of relevant size markers). As observed

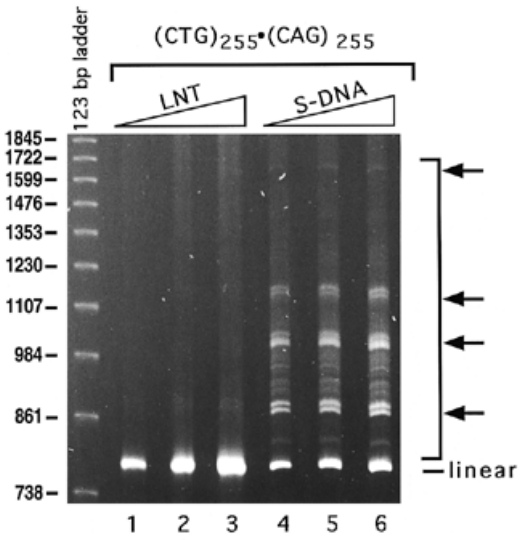

Figure 2. Heterogeneity in $(CTG)_{255}\cdot(CAG)_{255}$ S-DNAs. Linear non-treated (LNT) and reannealed (RD) samples of the *Hin*dIII-linearized $(CTG)_{255}\cdot(CAG)_{255}$ plasmid were digested with *Eco*RI and separated on 4% polyacrylamide gels. Increasing amounts of DNA were loaded: 1 µg, lanes 1 and 4; 2 µg, lanes 2 and 5; 3 µg, lanes 3 and 6. For clarity only the lower portion of the gel is shown. The bracket indicates the total distribution of anomalously migrating DNAs following reannealing. Arrows indicate the major species of S-DNAs. The 123 bp marker is shown.

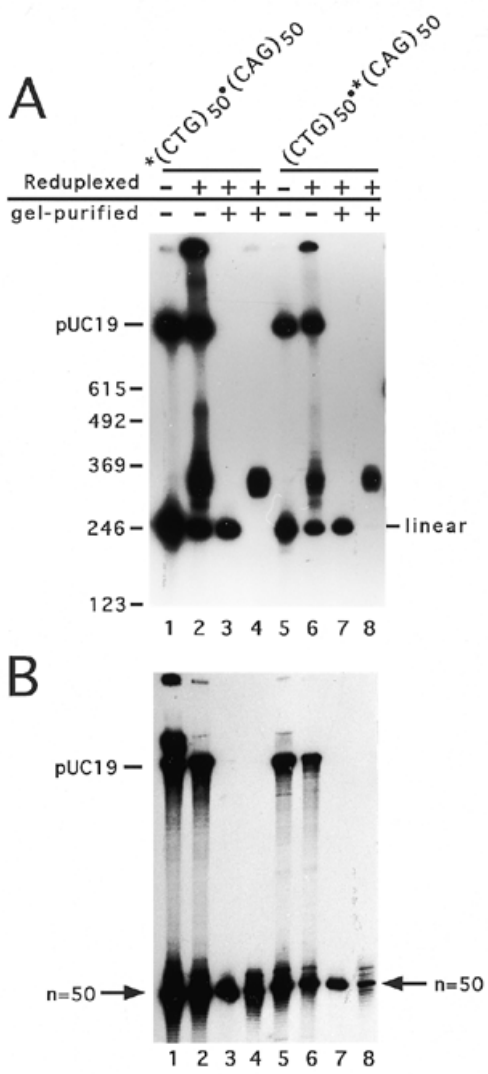

Figure 3. Gel purification and length distribution of DM $(CTG)_{50}\cdot(CAG)_{50}$ DNAs. DM $(CTG)_{50}\cdot(CAG)_{50}$ plasmid was digested with *Hin*dIII, radiolabeled (*) on either the 5′-end of the CAG strand (lanes 1–4) or the 3′-end of the CTG strand (lanes 5–8). Following denaturation/renaturation samples were digested with *Eco*RI. (**A**) All DNAs were separated on 4% native polyacrylamide gels. Shown are: linear non-treated DNA (LNT, lanes 1 and 5); reannealed samples (RD, lanes 2 and 6); gel-purified post-reannealing bands that co-migrated with the linear non-treated DNA (lanes 3 and 7); gel-purified $R = 1.25$ major S-DNA (lanes 4 and 8). Position of the linear duplex DNA and of the 123 bp marker are shown. (**B**) Samples from (A) were heat denatured in formamide and loaded immediately on a denaturing 5% urea/formamide polyacrylamide gel. Sample lane numbers in (B) correspond to those in (A). The arrow indicates the position of the $(CTG)_{50}$ or $(CAG)_{50}$ repeat fragments.

previously, each of the native linear forms migrated anomalously rapidly (8). For $(CTG)_{50}\cdot(CAG)_{50}$ 67–70% of the repeat-containing fragment formed S-DNA. In the case of $(CTG)_{30}\cdot(CAG)_{30}$, 39% of the repeat-containing fragment formed S-DNA. The propensity to form S-DNA increased from 2% for 17 repeats to 70% of the total repeat-containing DNA for 50 or more repeats (Fig. 1C). The electrophoretic patterns of S-DNA were similar for both lengths of repeats and, in fact, reannealing $(CTG)_n\cdot(CAG)_n$-containing DNA fragments from the DM gene with 30–83 repeats resulted in basically similar, but not identical, patterns of slipped products (data not shown).

Increases in repeat length increased the structural complexity of the S-DNA isomers formed (5; compare Fig. 2 with 1B). To further analyze the structural complexity of S-DNA, the *Hin*dIII–*Eco*RI fragments (for map see Fig. 1A) of denatured/renatured DNAs containing $(CTG)_{255}\cdot(CAG)_{255}$ were analyzed on a 4% polyacrylamide gel run at a lower voltage (60 V) for a longer length of time (Fig. 2). Many more distinct S-DNA isomers were apparent than in our previous results (5). Each of the three major S-DNA bands observed previously were resolved into at least two specific products (Fig. 2, see arrows). The major doublet bands may represent similar but polar isomers (i.e. 3′ slipped strands versus 5′ slipped strands). This higher resolution gel also revealed many minor specific isomers migrating as distinct bands (Fig. 2, see bracket).

Analysis of length heterogeneity in denatured/renatured DNAs

Preparations of triplet repeat-containing plasmids contain low percentages of plasmids containing short deletions and expansions within the triplet repeat (5,9,17). When bacteria are grown under conditions to minimize length heterogeneity [bacterial cultures shaken at low r.p.m. (<150) and cells harvested prior to the end of log phase growth] these length differences are a few repeats below and above that of the predominant length (see below). Reannealing plasmids with this length heterogeneity will result in both heteroduplexes as well as slipped homoduplexes. To determine the exact lengths of the repeats in each of the strands comprising the S-DNA structures individual bands were purified (Fig. 3A) and analyzed on a denaturing polyacrylamide gel (Fig. 3B). In the

$(CTG)_{50}\cdot(CAG)_{50}$-containing plasmid preparation the repeat tract contained some length heterogeneity, where >90% of the strands had a length of 50 repeats and <10% had lengths of 45–49 or 51–55 repeats (Fig. 3B, lanes 1 and 5). Both the 5′- (CAG) and 3′-radiolabeled (CTG) strands of the gel-purified R = 1.25 major S-DNA contained similar heterogeneity (Fig. 3B, compare lane 4 with 1 and 8 with 5). In contrast, the post-denaturation/renaturation fragment that co-migrated with the linear non-treated DNA containied *only* 50 repeats in each strand (Fig. 3B, lanes 3 and 7).

S-DNA forms in DNA with equal length repeats in the complementary strands

As discussed above, due to length heterogeneity of the repeats within the plasmid preparation (Fig. 3B), following denaturation and renaturation, heteroduplexes will be produced by annealing two strands having different numbers of repeats. However, not all slowly migrating DNAs could be heteroduplexes, since the percentage of S-DNA is much greater than the amount of DNA with $(CTG)_n\cdot(CAG)_n$ where $n \neq 50$. It is expected that two complementary strands with the same number of repeats may anneal in an out of register fashion and that these slipped homoduplexes would also migrate anomalously slowly. To demonstrate this the $(CTG)_{50}\cdot(CAG)_{50}$-containing *Hin*dIII–*Eco*RI fragment that co-migrated with the linear non-treated fragment following reannealing, which contains *only* 50 repeats, was isolated (Fig. 4, lane 4) and subjected to a second denaturation/renaturation treatment (Fig. 4, lane 8). This second reannealing resulted in a pattern of products which was very similar to that generated by the initial reannealing (Fig. 4, compare lane 8 with 2). Actually, more of the R = 2.0 S-DNA was formed relative to the R = 1.25 S-DNA. These results demonstrate that S-DNA structures can form between complementary strands containing equal lengths of repeats. The formation of new bands following reannealing of trinucleotide repeat-containing supercoiled plasmids, whose strands are catenated, also demonstrated that slipped homoduplexes can form from two strands containing the same number of triplet repeats (5).

The gel-purified R = 1.25 major S-DNA sample, which contains some heterogeneity in the number of CAG/CTG repeats (see Fig. 3B, lanes 4 and 8), was also subjected to a second reannealing treatment (see Fig. 4, lanes 5 and 9). The electrophoretic pattern of products formed was very similar to those generated by the initial reannealing protocol (Fig. 4, compare lane 9 with 2) and to the pattern formed following a 60 min treatment at 85°C (see 5). Apparently, the full set of original $(CTG)_{50}\cdot(CAG)_{50}$ S-DNA products (Fig. 4, lane 2) is a mixture of individual slipped DNA molecules in which the complementary strands are predominantly homogeneous, but also contains a minor fraction of heterogeneous repeat lengths. As a control a purified native ^{32}P-labeled *Hin*dIII–*Eco*RI $(CTG)_{50}\cdot(CAG)_{50}$ fragment was denatured and renatured. This treatment resulted in a full distribution of products similar to those obtained from denaturation/renaturation of the full-length *Hin*dIII-linearized plasmid (Fig. 4, compare lanes 3 and 7 with 1 and 2).

For each of the denaturation/renaturation reactions using the gel-purified ^{32}P-labeled *Hin*dIII–*Eco*RI $(CTG)_{50}\cdot(CAG)_{50}$ fragments a band just above the position of the duplex DNA was observed (R = 1.05) (Fig. 4, lanes 7–9, see arrowhead). This band was not present following reannealing linearized plasmid (Fig. 4, lane 2). In the short *Hin*dIII–*Eco*RI restriction fragment triplet repeats constitute the majority of the DNA molecule. During

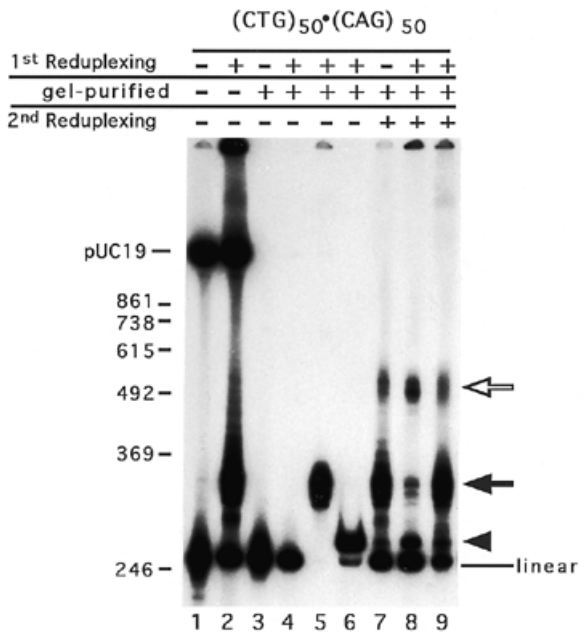

Figure 4. Effect of secondary reannealing on S-DNA. The DM $(CTG)_{50}\cdot(CAG)_{50}$ plasmid was digested with *Hin*dIII, radiolabeled, denatured and renatured and then digested with *Eco*RI. All DNAs were separated on a 4% polyacrylamide gel. Shown are: linear non-treated DNA (lane 1); denatured/renatured DNA (lane 2); gel-purified linear non-treated DNA (lane 3); gel-purified post-reannealing DNAs that co-migrated with the linear non-treated DNA (lane 4); gel-purified R = 1.25 major population of S-DNA (lane 5); single-stranded DNA following heat denaturation and rapid cooling of gel-purified linear DNA (lane 6). The gel-purified DNA samples of lanes 3–5 were subjected to denaturation and renaturation and loaded in lanes 7–9 respectively. Filled and hollow arrows indicate the R = 1.25 and R = 2.0 S-DNA populations respectively. The arrowhead indicates the single-stranded *Hin*dIII–*Eco*RI DNA. Although only the duplex DNAs radiolabeled on the 5′-end of the CAG strand are shown, the pattern of products was identical for DNAs radiolabeled on the 3′-end of the CTG strand.

renaturation an intra-strand hairpin structure may form within the triplet repeats and reduce the possibility of complementary strand hybridization. The single-stranded nature of the R = 1.05 product was confirmed by denaturation and rapid chilling (which prevents renaturation) of the ^{32}P-labeled *Hin*dIII–*Eco*RI $(CTG)_{50}\cdot(CAG)_{50}$-containing fragment which resulted in this R = 1.05 band (Fig. 4, compare lane 6 with 7–9, see arrowhead).

EM analysis of gel-purified $(CTG)_{50}\cdot(CAG)_{50}$ S-DNA

The gel-purified linear duplex and both the two slowly migrating S-DNA populations were analyzed by EM. First, the linear non-treated *Hin*dIII–*Eco*RI restriction fragment containing $(CTG)_{50}\cdot(CAG)_{50}$ (Fig. 5A) appeared as linear molecules with smooth contours, an appearance identical to random sequence B-DNA (14; Fig. 5B).

The two predominant S-DNA populations formed following reannealing $(CTG)_{50}\cdot(CAG)_{50}$-containing DNA [the major R = 1.25 S-DNA (Fig. 4, filled arrow) and the minor R = 2.0 S-DNA population (Fig. 4, hollow arrow)] were gel purified and analyzed by EM. Sixty five of 100 gel-purified R = 1.25 S-DNA molecules contained single or multiple kinks (Fig. 5C and D). A good indication of the presence of bends or kinks is provided by an analysis of the ratio of the straight line distance between the ends of

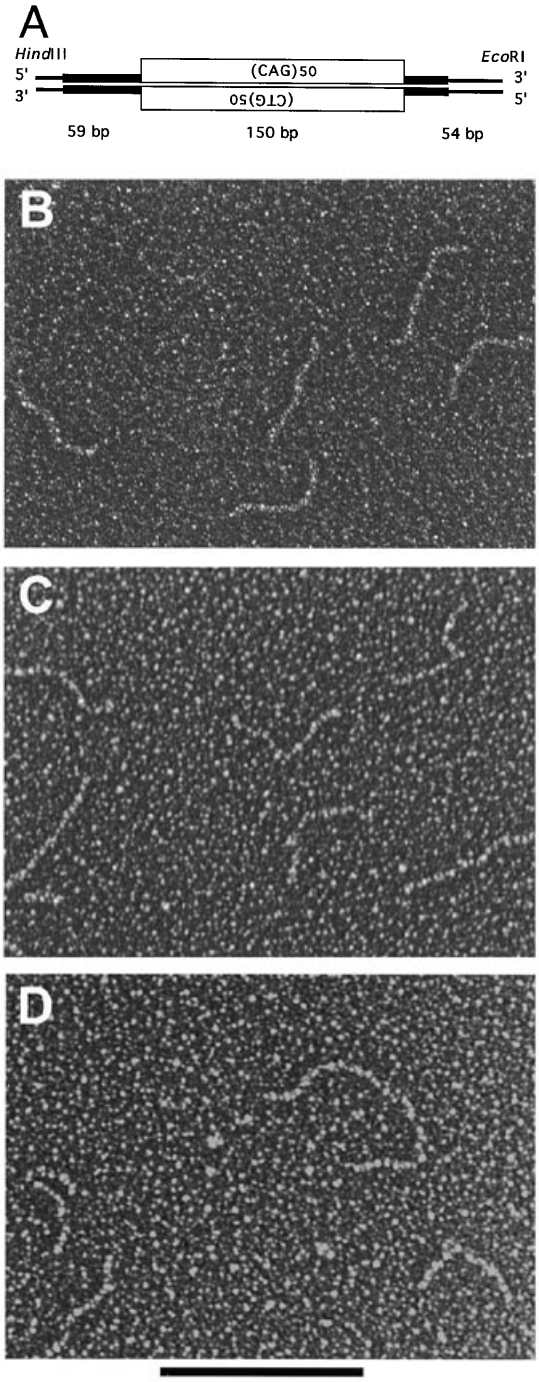

Figure 5. EM analysis of the DM $(CTG)_{50}\cdot(CAG)_{50}$ gel-purified major S-DNA population. (**A**) Map of the DM $(CTG)_{50}\cdot(CAG)_{50}$ *Hin*dIII–*Eco*RI fragment. Human non-repetitive flanking sequences (positions 357–375 and 391–433, as in 7) are in bold. The thin lines represent plasmid vector sequences. The map is drawn to scale. All samples analyzed were gel-purified *Hin*dIII–*Eco*RI fragments. (**B**) Micrograph of linear non-treated DNAs. (**C**, **D**) Micrographs of the $R = 1.25$ major S-DNA population, which indicate that the size and number of the kinks varied between various molecules. See text for detailed descriptions. Micrographs are shown in reverse contrast. Bar 170 nm.

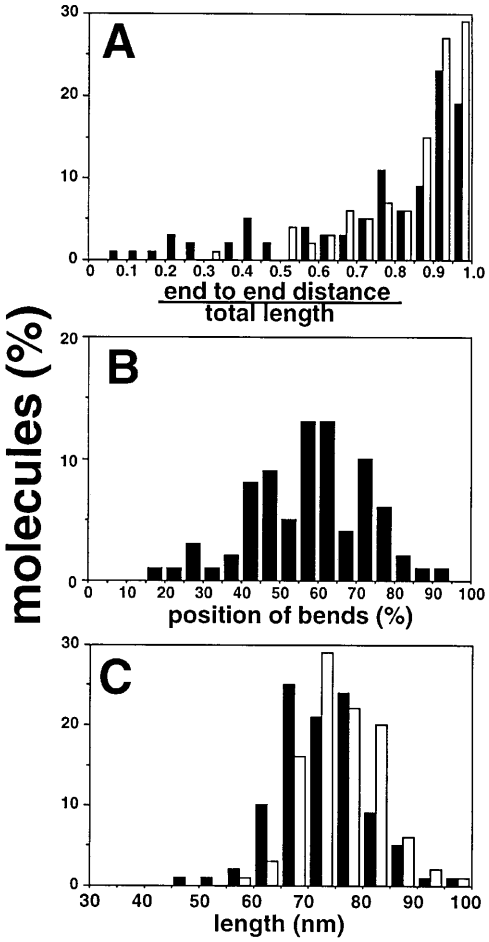

Figure 6. EM analysis of the 263 bp DM $(CTG)_{50}\cdot(CAG)_{50}$ linear and S-DNAs. (**A**) Analysis of the ratio of the straight line distance between the ends of a $(CTG)_{50}\cdot(CAG)_{50}$ *Hin*dIII–*Eco*RI fragment (ε–ε) relative to its total contour length L (as in 14). (ε–ε)/L measurements of the linear non-treated DNAs, hollow bars; (ε–ε)/L measurements of the $R = 1.25$ major S-DNA population, filled bars. (**B**) Analysis of position of bends in the 263 bp $R = 1.25$ major S-DNA population. (**C**) Analysis of contour length measurements of the linear non-treated $(CTG)_{50}\cdot(CAG)_{50}$ *Hin*dIII–*Eco*RI fragments, hollow bars; analysis of contour length measurements of the $R = 1.25$ S-DNA population, filled bars. Using Student's *t*-test the differences in the length measurements between the linear non-treated and the S-DNAs were statistically significant, with $P < 0.05$.

a DNA fragment (ε–ε) divided by the total contour length L (14). (ε–ε)/L measurements of the linear non-treated $(CTG)_{50}\cdot(CAG)_{50}$ *Hin*dIII–*Eco*RI fragments (Fig. 6A, open bars) were typical of random sequence DNA (14). However, analysis of the $R = 1.25$ S-DNAs (Fig. 6A, solid bars) showed a significant degree of compaction, as is evident in the broadening of the distribution and a shift in the distribution of (ε–ε)/L values to lower numbers. In addition, all kinks mapped within the trinucleotide repeat tract, although they did not have a preferred location within the repeat tract (Fig. 6B), in agreement with our previous mapping of S-DNA (5). Although the specific ends were not determined, analysis of these measurements where the length from the end to the bend was measured from the short arm shows that there is no bias of nucleation point, indicating that the kinks likely formed randomly throughout the repeat tract (data not shown).

If the strands of the repeat tracts in the S-DNA structures are slipped with respect to each other by out of register mispairings then the length of the Watson–Crick duplex should be shortened by the number of repeats slipped out on either strand. Contour

lengths of both the linear non-treated $(CTG)_{50}\cdot(CAG)_{50}$-containing DNA (Fig. 5B) and the $R = 1.25$ major S-DNA population (Fig. 5C and D) are shown in Figure 6C. The linear non-treated DNA had an average length of 757 ± 67.9 Å (with a range of 583–954 Å), typical of random sequence DNA (Y.-H.Wang and J.D.Griffith, unpublished results). Other studies of native linear triplet repeat DNAs have not revealed any detectable differences from random sequence B-DNA (J.D.Griffith, unpublished results). However, the gel-purified $(CTG)_{50}\cdot(CAG)_{50}$ $R = 1.25$ major S-DNA had an average length of 729 ± 82.9 Å (with a range of 493–966 Å). If the repeat tract in the $R = 1.25$ S-DNA population is composed of slipped out repeats on both strands separated by non-slipped out repeats which are Watson–Crick base paired in an out of register fashion, the length of the slipped out region can be calculated: The average length of the linear form is 757 Å and the average length of S-DNA is 729 Å. This represents a 3.7% decrease in length, which translates as 9–10 bp, or 3 repeats, contained in an average slip-out. With the low end of the range of contour lengths for the $R = 1.25$ major population of S-DNA being 493 Å, this represents a 35% change in length, which translates as a sum total of ~93 bp, or up to 31 repeats, contained in slipped out DNAs. DNA sequences involved in slipped mispairing may occur as a single event or as multiple slipped out regions in a single S-DNA molecule, as previously suggested (5).

Micrographs of the slower migrating $R = 2.0$ minor S-DNA population (see Fig. 4, hollow arrow) showed that most of the molecules contained unusual structures, many with large slip-outs (Fig. 7). Some of the structures contained single or multiple branches, reminiscent of three-way junctions (Fig. 7A), three-way junctions with aggregated DNA (possibly random coiled/single-stranded regions) at the junction points (Fig. 7B), molecules with both loops and branches (Fig. 7C), cruciform-like structures (Fig. 7D), θ-like structures (Fig. 7E) and bi-lobed slip-outs (Fig. 7F). These results demonstrate that the $R = 2.0$ minor S-DNA population is composed of a heterogeneous population of distinct monomeric DNA structures composed of relatively long segments of the $(CTG)_{50}\cdot(CAG)_{50}$ repeat tract or structures formed by interaction of two distant loops. The multiple and longer branches are consistent with the slower elctrophoretic mobility of this DNA population.

EM analysis of a plasmid containing $(CTG)_{255}\cdot(CAG)_{255}$ is shown in Figure 8. To facilitate mapping, plasmids were linearized with *Nde*I, which places the repeat tract 265 bp from one end of the 3513 bp linear DNA molecule (Fig. 8A). The linear non-treated DNAs appeared as long molecules with smooth contours, expected of random sequence B-DNA (14; data not shown). Following denaturation/renaturation, 9% of the DNA molecules contained visible single or multiple branches, each of which mapped to the triplet repeat tract (Fig. 8B–F). Each structure involved ≥15 repeats (this is the limit of EM resolution). Many of the molecules contained double loops of equal (Fig. 8B) or differing size. Some contained two opposed branches similar to cruciform structures (Fig. 8C) or two displaced branches (Fig. 8D). The branches appeared at different positions along the repeat tract (Fig. 8E, see both molecules) and were composed of varying lengths of repeats (Fig. 8E and F, see both molecules). As suggested from the polyacrylamide gel mobility pattern shown in Figure 2, the EM results confirm that a heterogeneous population of distinct structures is formed in the long reannealed repeat tracts. The presence of these looped and branched structures in trinucleotide repeat-containing DNAs is dependent upon the process of denaturation/renaturation and they are not observed in DNAs composed of random sequence DNAs (data not shown).

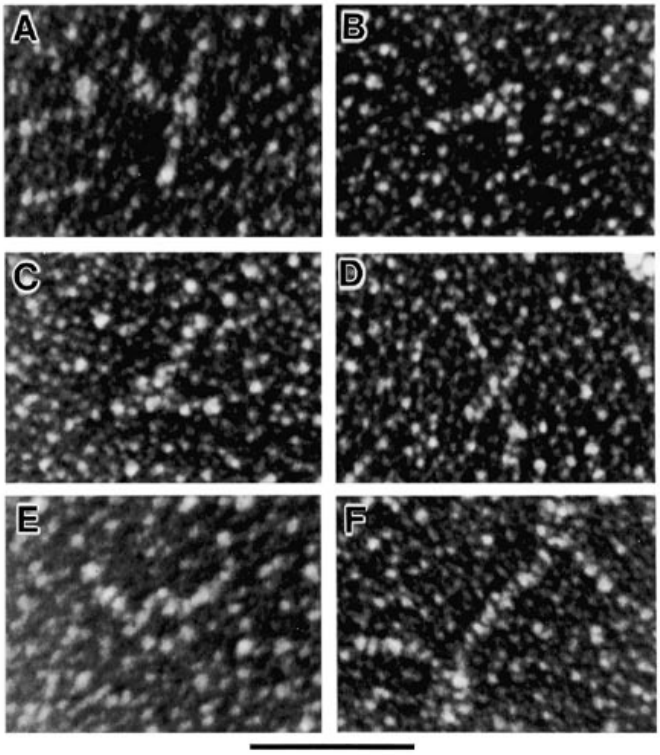

Figure 7. EM analysis of the DM $(CTG)_{50}\cdot(CAG)_{50}$ gel-purified $R = 2.0$ minor S-DNA population. The *Hin*dIII–*Eco*RI $R = 2.0$ minor S-DNA population (see Fig. 5A for map) was gel purified and analyzed by EM. (**A–F**) Micrographs of individual $R = 2.0$ major S-DNAs shown in reverse contrast. Bar 55 nm. See text for detailed descriptions.

DISCUSSION

This paper extends a previous observation that denaturing and renaturing DNA containing long tracts of (CTG)·(CAG) or (CGG)·(CCG) triplet repeats results in formation of alternative DNA secondary structures with reduced electrophoretic mobility in polyacrylamide gels (5). In this report we present data addressing ideas only hypothesized in our initial report (5). Specifically we demonstrate that: (i) the slow migrating species are in fact slipped-strand DNAs formed by out of register base pairing between complementary strands; (ii) the homoduplex nature of the slipped structures; (iii) the expected structural heterogeneity of slipped structures. Additionally we give a range of lengths of possible slipped out repeats. By denaturing and renaturing DNAs with equal lengths of repeats in the complementary strands we show unequivocally that S-DNA can form with slipped out regions on both complementary strands. Two populations of S-DNAs are formed for $(CTG)_{50}\cdot(CAG)_{50}$-containing plasmids; the first having single or multiple short slipped out regions (≤15 repeats each, the limit of EM resolution) that appear as kinks in the electron microscope; the second slower migrating population containing visible branches, loops (involving 15–50 repeats) and three-way and four-way junctions expected of larger slipped-strand structures (or from the interaction of two distant loops). Several groups have shown that single-stranded CTG, CAG, CGG or

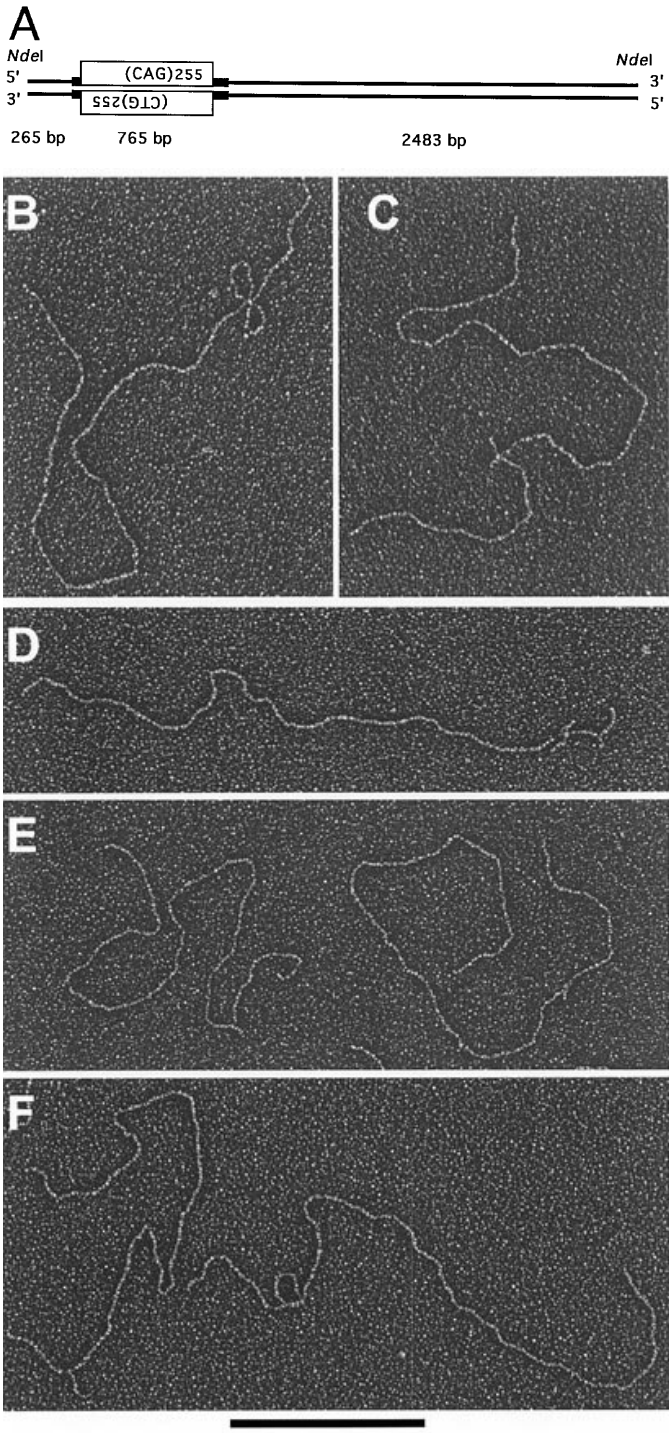

Figure 8. EM analysis of DM $(CTG)_{255}\cdot(CAG)_{255}$ S-DNAs. (**A**) Map of the DM $(CTG)_{255}\cdot(CAG)_{255}$ *Nde*I-linearized plasmid. Human non-repetitive flanking sequences (positions 357–375 and 391–433, as in 7) are in bold. The thin lines represent plasmid vector sequences. The map is drawn to scale. DNAs were analyzed without purification. (**B–F**) Electron micrographs of denatured/renatured $(CTG)_{255}\cdot(CAG)_{255}$ *Nde*I-linearized plasmid shown in reverse contrast. Bar 170 nm. See text for detailed descriptions.

CCG oligonucleotides can form intrastrand hairpins (18–22). The small kinks and the larger branches in S-DNA observed by EM may be hairpins of varying sizes. Both S-DNA populations migrate as a series of heterogeneous bands, consistent with slipped out regions located randomly throughout the repeat tract. This observation is confirmed by EM mapping of kinks at random locations in the repeat tract. Interestingly, the predominant $R = 1.25$ $(CTG)_{50}\cdot(CAG)_{50}$ S-DNA population contained single or multiple short slip-outs. Petruska *et al.* have presented evidence that for a given length of a single-stranded CTG (or CAG) oligonucleotide multiple short hairpin structures rather than a single large hairpin may form (21), while Gao and colleagues demonstrated that as few as two to three repeats can form hairpins (19).

Sequence specificity for S-DNA formation

Theoretically, denaturation and renaturation of almost any tandemly repeated sequence can lead to out of register mispairings, resulting in slipped-strand structures. The relative structural stability of the slipped DNAs may be dependent upon the length and sequence of the repeat unit as well as the number of repeats in the tract. Belotserkovskii and Johnston (23,24) have shown that denaturation and renaturation of the dinucleotide repeats $(GA)_{37}\cdot(TC)_{37}$ and $(CA)_{30}\cdot(TG)_{30}$ resulted in a minor but reproducible amount of novel products. These products all migrated in polyacrylamide as dimers, trimers, tetramers and higher order aggregates. The mispaired aggregates of the dinucleotide repeats are likely to be formed by mispairing of multiple (>2) strands (24); this has been confirmed by EM analysis of these DNAs (25). Denaturation and renaturation of these dinucleotide repeats did not result in formation of S-DNAs (23,24; C.E.Pearson and R.R.Sinden, unpublished results), suggesting that not all tandemly repeated sequences are capable of stable slipped-strand structure formation. An analysis of the sequence dependence of various mono-, di-, tri- and tetranucleotide repeat tracts on S-DNA formation is in progress and will be the subject of a later paper.

Heterogeneity of S-DNA

EM analysis of S-DNAs confirmed the identity and heterogeneous nature of slipped-strand DNAs. First, in all samples studied by EM the triplet repeat-containing molecules appeared as duplex monomers and not aggregated multimeric complexes as observed for denatured and renatured dinucleotide repeats $(GA)_{37}\cdot(TC)_{37}$ and $(CA)_{30}\cdot(TG)_{30}$ (23–25). The observed single or multiple kinks in the majority of the anomalously migrating slipped DNAs could be due to bulge structures caused by unpaired/slipped out triplets. The kinks in the $R = 1.25$ major S-DNA population are reminiscent of kinks observed for DNAs containing bulged or unpaired/deleted bases, which are also known to migrate anomalously slowly (13,14). Since the majority of these molecules contained only kinks with no visible three- or four-way junctions and the contour length measurements are decreased by three to 31 repeats, many of the slipped out regions in most of the DM S-DNA are relatively short.

Coggins and colleagues (26,27) observed by EM a variety of highly knotted structures in reannealed minisatellite DNAs. They proposed a model of ‘paired loops’ in which the slipped out regions resulting from slipped misalignment would hybridize with each other forming ‘underwound loops’. The bow ties and figures of eight we observed may be similar to these paired loops. The knotted structures observed by Coggins (26,27) differed from those we observe in that our structures are composed of microsatellite sequences with a repeat unit length of 3 nt, whereas they were looking at minisatellite repeat units ranging from 17 to

37 nt long. Furthermore, the structures we have characterized exhibited a surprisingly high degree of thermostability, while those of the longer repeat units were not biophysically stable (26).

In addition to slipped homoduplex (S-DNA) structures, we have described slipped intermediate heteroduplex DNA (SI-DNA) of the form $(CTG)_{30}\cdot(CAG)_{50}$ and $(CTG)_{50}\cdot(CAG)_{30}$ in which the strands have different numbers of repeats (4). Analysis of SI-DNAs by polyacrylamide gel electrophoresis revealed a series of heteroduplex-specific products which appeared as two distinct major slower migrating products, representing the $(CTG)_{30}\cdot(CAG)_{50}$ and $(CTG)_{50}\cdot(CAG)_{30}$ heteroduplexes, as well as a series of fainter closely spaced products. Both of the sister heteroduplex SI-DNAs migrated slightly slower than the $R = 2.0$ $(CTG)_{50}\cdot(CAG)_{50}$ S-DNA. The relative slow migration of the SI-DNAs containing 20 repeat slip-outs is consistent with slow migration of the $(CTG)_{50}\cdot(CAG)_{50}$ $R = 2.0$ S-DNAs containing large loops and slip-outs.

The $(CTG)_n\cdot(CAG)_n$ S-DNAs and SI-DNAs induced by strand denaturation and renaturation are remarkably stable (5,4). Single-stranded CTG, CAG, CGG or CCG repeats can form intrastrand hairpins. Branch migration of slipped out trinucleotide repeats would require not only the breaking and reforming of base pairs at the junction point but also at every base pair within the slipped out hairpin. Branch migration of slipped out trinucleotide repeats may be severely impeded by intrastrand base pairs within the slipped out repeats. Thus for either long intrastrand slip-outs or long out of register interstrand duplex regions the activation energy for breakage may be so high that complete renaturation takes forever. In this fashion the remarkable stability of S-DNAs may be due to both kinetic as well as thermodynamic factors.

Possible biology of S-DNA

Genetic analyses of triplet repeat disease loci in patients DNA indicate that there is a close relationship between the number of repeats and both the genetic stability, the age of onset and the severity of the disease symptoms (1,7,28,29). We observed both an increase in the relative amounts of new products and the complexity of the pattern of new bands for the reannealed DM fragments as the number of repeats increased. This indicates a direct relationship between the number of repeats and the propensity to form slipped structures. Interestingly, the inflection point where the propensity to form S-DNA increased occurred between 30 and 50 repeats, which are repeat lengths typical of normal individuals and intermediate allele lengths found in protomutation DM individuals who display mild disease symptoms (28,29). The ability to form multiple slipped isomers increases the possibility that different slipped intermediates may be stabilized and this may result in a higher chance of repeat length changes.

It is evident from the results presented here that in the chromosome, where the number of repeats in each strand are presumably the same, slipped mispairing in the triplet repeat tracts would result in homoduplex slipped structures (S-DNA). Alternatively, if these structures form at a replication fork during DNA replication they could result in slipped intermediates (SI-DNA), where the number of repeats in the nascent strand differs from that of the template strand. The remarkable stablility of these structures *in vitro* suggests that either, or both of these possible structures may participate in expansion of triplet repeats in cells.

ACKNOWLEDGEMENTS

This work was supported in part by funds from the Institute of Biosciences and Technology, an NIH grant (GM31819) to J.G. and an NIH grant (GM52982) to R.R.S.

REFERENCES

1 Pearson,C.E. and Sinden,R.R. (1998) In Wells,R.D. and Warren,S.T. (eds), *Genetic Instabilities and Hereditary Neurological Diseases*. Academic Press, New York, NY, in press.
2 Richards,R.I. and Sutherland,G.R. (1992) *Cell*, **70**, 709–712.
3 Fishel,R. and Kolodner,R.D. (1995) *Curr. Biol.*, **5**, 382–395.
4 Pearson,C.E., Ewel,A., Acharya,S., Fishel,R. and Sinden,R.R. (1996) *Hum. Mol. Genet.*, **6**, 1117–1123.
5 Pearson,C.E. and Sinden,R.R. (1996) *Biochemistry*, **35**, 5041–5053.
6 Hanvey,J.C., Shimizu,M. and Wells,R.D. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 6292–6296.
7 Mahadevan,M., Tsilfidis,C., Sabourin,L., Shutler,G., Amemiya,C., Jansen,G., Neville,C., Narang,M., Barceló,J.M., O'Hoy,K., LeBlond,S., Earle-MacDonald,J., DeJong,P.J., Wieringa,B. and Korneluk,R.G. (1992) *Science*, **255**, 1253–1255.
8 Chastain,P.D., Eichler,E.E., Kang,S., Nelson,D.L, Levene,S.D. and Sinden,R.R. (1995) *Biochemistry*, **34**, 16125–16131.
9 Bowater,R.P., Rosche,W.A., Jaworski,A., Sinden,R.R., and Wells,R.D. (1996) *J. Mol. Biol.*, **264**, 82–96.
10 Pearson,C.E., Ruiz,M.T., Price,G.B. and Zannis-Hadjopoulos,M. (1994) *Biochemistry*, **33**, 14185–14196.
11 Pearson,C.E., Zannis-Hadjopoulos,M., Price,G.B. and Zorbas,H. (1994) *EMBO J.*, **14**, 1571–1580.
12 Griffith,J.D. and Christiansen,G. (1978) *Annu. Rev. Biophys. Bioengng*, **7**, 19–35.
13 Hsieh,C.-H. and Griffith,J.D. (1989) *Proc. Natl Acad. Sci. USA*, **86**, 4833–4837.
14 Wang,Y.-H. and Griffith,J. (1991) *Biochemistry*, **30**, 1358–1363.
15 Delwart,E.L., Shpaer,E.G., Louwagie,J., McCutchan,F.E., Grez,M., Rubsamen-Waigmann,H. and Mullins,J.I. (1993) *Science*, **262**, 1257–1261.
16 Lilley,D.M.J. (1995) *Proc. Natl. Acad. Sci. USA*, **92**, 7140–7142.
17 Kang,S., Jaworski,A., Oshima,K. and Wells,R.D. (1995) *Nature Genet.*, **10**, 213–218.
18 Gacy,A.M., Goellner,G., Juranic,N., Macura,S. and McMurray,C.T. (1995) *Cell*, **81**, 533–540.
19 Zheng,M., Huang,X., Smith,G.K., Yang,X. and Gao,X. (1996) *J. Mol. Biol.*, **264**, 323–336.
20 Mariappan,S.,S.V.S., Garcia,A.E. and Gupta,G. (1996) *Nucleic Acids Res.*, **24**, 775–783.
21 Petruska,J., Arnheim,N. and Goodman,M.F. (1996) *Nucleic Acids Res.*, **24**, 1992–1998.
22 Mitas,M (1997) *Nucleic Acids Res.*, **25**, 2245–2253.
23 Belotserkovskii,B.P. and Johnston,B.H. (1996) *Science*, **271**, 222–223.
24 Belotserkovskii,B.P. and Johnston,B.H. (1996) In Sarma,R.H. and Sarma,M.H. (eds), *Biological Structure and Dynamics, Proceedings of the Ninth Conversation*. Adenine Press, The University, Albany, NY. Vol. 2, pp. 157–164.
25 Gaillard,C. and Strauss,F. (1994) *Science*, **264**, 433–436.
26 Coggins,L.W. and O'Prey,M. (1989) *Nucleic Acids Res.*, **17**, 7417–7426.
27 Coggins,L.W., O'Prey,M. and Akhter,S. (1992) *Gene*, **121**, 279–285.
28 Barceló,J.M., Mahadevan,M.S., Tsilfidis,C., MacKenzie,A.E. and Korneluk,R.G. (1993) *Hum. Mol. Genet.*, **2**, 705–709.
29 Yamagata,H., Miki,T., Sakoda,S.-I., Yamanaka,N., Davies,J., Shelbourne,P., Kubota,R., Takenaka,S., Nakagaha,M., Ogihara,T. and Johnson,K. (1994) *Hum. Mol. Genet.*, **3**, 819–820.