4202–4209 Nucleic Acids Research, 1995, Vol. 23, No. 20

CGG repeats associated with DNA instability and chromosome fragility form structures that block DNA synthesis *in vitro*

K. Usdin* and K. J. Woodford

Section on Genomic Structure and Function, National Institute of Diabetes and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892-0830, USA

Received May 15, 1995; Revised and Accepted September 6, 1995

ABSTRACT

A large increase in the length of a CGG tandem array is associated with a number of triplet expansion diseases, including fragile X syndrome, the most common cause of heritable mental retardation in humans. Expansion results in the appearance of a fragile site on the X chromosome in the region of the CGG array. We show here that CGG repeats readily form a series of barriers to DNA synthesis *in vitro*. These barriers form only when the $(CGG)_n$ strand is used as the template, are K^+-dependent, template concentration-independent, and involve hydrogen bonding between guanines. Chemical modification experiments suggest these blocks to DNA synthesis result from the formation of a series of intrastrand tetraplexes. A number of lines of evidence suggest that both triplet expansion and chromosome fragility are the result of replication defects. Our data are discussed in the light of such evidence.

INTRODUCTION

Large scale instability of CGG repeats is associated with the triplet expansion mutations fragile X syndrome (1,2), the most common cause of inherited mental retardation in humans, FraxE mental retardation (3) and a mutation at FraxF (4). The instability of these repeats is related to the number of uninterrupted CGGs (5) and expansion results in the appearance of a fragile site at that locus. Fragile sites, visible as poorly staining regions or gaps in chromatin, are most prominent when the cells are starved for the precursors of DNA synthesis. These sites are prone to breakage and illegitimate recombination (6). At least one autosomal fragile site is also associated with a large CGG tract (7). It has been suggested that both triplet expansion (8) and expression of the fragile site (9) are intrinsic properties of the $(CGG)_n$ tract. As yet no exchange of flanking markers has been seen that would support a recombination-based mechanism for triplet expansion; however, the fact that both point mutations and the gain or loss of repeat units occur predominantly at the 3′-end of the CGG tract (8) suggests that triplet expansion may result from a replication defect.

In this paper we show that CGG tracts are inherently capable of forming a set of barriers to DNA synthesis *in vitro*. Our structural data suggest that the barriers are a series of K^+-dependent tetrahelical DNA structures that form readily under physiological conditions. The potential ramifications of these blocks to DNA synthesis are discussed.

MATERIALS AND METHODS

Oligonucleotides

Oligomers were synthesized on an ABI Model 381A oligonucleotide synthesizer. 5′-Dimethoxytrityl-*N*2-dimethylaminomethylidene-7-deaza-2′-deoxyguanosine, 3′-[(2-cyanoethyl)-(*N*,*N*-diisopropyl)]phosphoramidite was used to synthesize oligonucleotides containing 7-deazaguanine (Glen Research). Oligonucleotides were purified by denaturing gel electrophoresis, end-labeled with T4 polynucleotide kinase and [γ-^{32}P]ATP and purified by NENSORB column chromatography (NEN).

Template construction

M13mp18 clones containing CGG-rich repeats were constructed and single-stranded bacteriophage DNA purified as described previously (10). A double-stranded PCR fragment was synthesized from a *supF*-containing plasmid using a primer pair whose 3′-ends are homologous to the 5′- and 3′-ends of the *Escherichia coli supF* gene, respectively, Xstring4 [5′-GTACGAATT$(CGG)_{12}(CGg)_4$CTCGAGTCAACGTAACACTTT-3′ (g = 7-deazaguanine)] and supFR1 (5′-GATCGAATTCGTCGACATGGTGGTGGGGGAA-3′) (10). This PCR product was then purified by gel electrophoresis and tested for its ability to arrest DNA synthesis using the primer SupFR4 (5′-ATGCTTTTACTGGCCTGCT-3′) as described below.

Assay for DNA synthesis arrest sites

Primer extensions were carried out in the presence or absence of various cations together with either the four deoxynucleotides or a mixture of deoxynucleotides and a single dideoxynucleotide. The positions of the arrest sites were mapped at single nucleotide resolution by reference to dideoxynucleotide-mediated chain terminations or, in the case of reactions carried out without

* To whom correspondence should be addressed

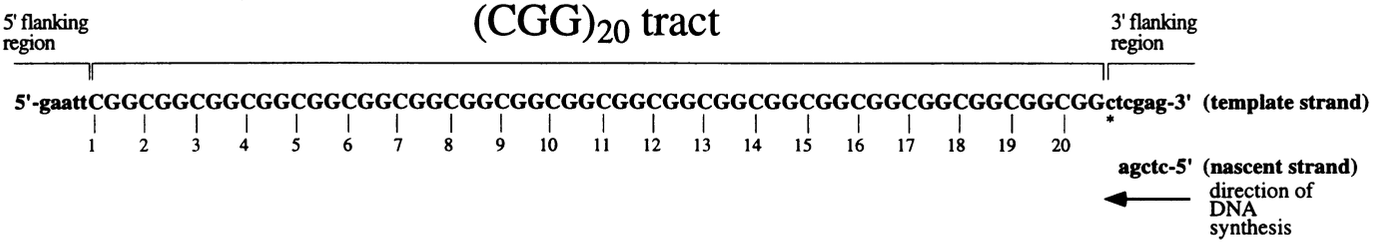

Figure 1. Region of $(CGG)_{20}$-containing template indicating the triplet numbering scheme used throughout this manuscript, the direction of DNA synthesis and the sequences of part of both the 5′ and 3′ flanking regions (lower case).

dideoxynucleotides, to a sequencing ladder generated using the same primer/template combination.

DNA polymerases tested include Taq DNA polymerase (Gibco-BRL), Sequenase™, a modified T7 DNA polymerase (United States Biochemicals Inc.), Klenow fragment of *Escherichia coli* DNA polymerase I (Gibco BRL) and T4 DNA polymerase (Boehringer-Mannheim). Reaction mixtures contained 50 mM Tris–HCl, pH 9.3, and 2.5 mM $MgCl_2$, 13 fmol template, 0.4 pmol ^{32}P-end-labeled primer, 1 U polymerase and either 100 μM all four dNTPs for primer extensions or 5 μM dNTPs and 7-deaza-dGTP and either 8.3 μM ddGTP, 150 μM ddATP, 300 μM ddTTP or 100 μM ddCTP in the case of dideoxynucleotide-containing reactions, in a total of 6 μl. In some reactions gene 32 protein (Pharmacia) was added to a final concentration of 160 μg/ml. The primer used for the analysis of $(GCC)_n$-containing templates was the M13 sequencing primer (5′-CGCCAGGGTT-TTCCCAGTCACGAC-3′) and the primer used for analysis of the $(CGG)_n$-containing templates was supFR4. Different monovalent cations were added to each reaction to a final concentration of 40 mM. The reactions using Taq polymerase were then subjected to 30 cycles of denaturation, annealing and extension (30 s at 94°C, 30 s at 55°C and 30 s at 72°C) in a Perkin-Elmer-Cetus PCR machine. Reactions with thermolabile polymerases were carried out for 5 min at 37°C. Reactions were terminated by the addition of one half volume of a stop buffer containing 95% (v/v) formamide, 10 mM EDTA, pH 7.6, 0.1% xylene cyanol and 0.1% bromophenol blue and the reaction products analyzed on 5.5% sequencing gels.

Gel electrophoresis

Radiolabeled oligonucleotide at a concentration of 2.5 μM was heated to 94°C for 5 min to eliminate any intermolecular associations. A 20 μl aliquot was then loaded on a 10% polyacrylamide gel containing 0.5× TBE buffer and, where indicated, 7 M urea or 40 mM of the indicated cation. Samples were then electrophoresed at room temperature at 10 V/cm until the bromophenol blue had migrated 7 cm. The gels were then dried and autoradiographed.

Chemical modification

Radiolabeled oligonucleotide at a concentration of 2.5 μM in 20 μl of a 50 mM Tris–HCl, pH 9.0, 2 mM $MgCl_2$ solution was heated to 94°C for 1 min to remove any pre-existing structure. The samples were then incubated at 37 or 55°C for 5 min in the presence or absence of 40 mM KCl. Samples were then treated with either 2.5% dimethylsulfate (DMS) for 1 min at room temperature or 2% bromoacetaldehyde (BAA) for 5 min at 37°C. BAA-treated samples were extracted twice with ether and precipitated with 1 ml butanol. These samples were then resuspended in 20 μl 50 mM Tris–HCl, pH 9.0, 2 mM $MgCl_2$ and treated with DMS as described above. Both sets of samples were then mixed with 20 μl 2 M pyrrolidine and incubated at 90°C for 15 min. The samples were precipitated with 1 ml butanol, washed several times with 70% ethanol and resuspended in 10 μl buffer containing 95% (v/v) formamide, 10 mM EDTA, pH 7.6, 0.1% xylene cyanol and 0.1% bromophenol blue. Three microliters of each sample were analyzed on 12% sequencing gels.

RESULTS

When DNA synthesis is carried out on a single-stranded template containing a $(CGG)_{20}$ tract (see Fig. 1) a series of DNA synthesis arrest sites is visible, either as truncated polynucleotide chains in primer extension reactions (Fig. 2) or as non-dideoxynucleotide-mediated chain terminated products in reactions containing dideoxynucleotides (Fig. 3) when KCl is present in the reaction mix. The first arrest site is seen opposite the C residue immediately 3′ of the $(CGG)_{20}$ tract in the template (C*, see Fig. 1) and the other arrest sites are seen opposite successive Cs in the template (i.e. at C_{20}, C_{19}, C_{18}, etc. using the numbering scheme illustrated in Fig. 1). The same pattern of DNA synthesis arrest is seen when double-stranded templates are used (data not shown). These arrest sites constitute a significant impediment to DNA synthesis, reducing the amount of full-length product produced (Fig. 2). This is not likely to be due to an effect of K^+ on the polymerase itself, since polymerization is unaffected elsewhere on the template and in templates lacking a CGG tract no difference is seen in the amount of high molecular weight material generated in the presence and absence of K^+ (10). Little if any DNA synthesis arrest is seen in the absence of monovalent cation or when NaCl, NH_4Cl, RbCl or CsCl are used in place of KCl (Fig. 3; unpublished observations), indicating that it is the K^+ cation, rather than the Cl^- anion, that is responsible for the KCl-dependent arrest of DNA synthesis. Increasing the amount of dNTPs in the reaction mix has no effect on the appearance of arrest sites (results not shown).

The occurrence of these DNA arrest sites is not polymerase specific, since they are seen with a number of different DNA polymerases, including modified T7 DNA polymerase (Sequenase™), the Klenow fragment of *E.coli* DNA polymerase I, Taq polymerase, AMV reverse transcriptase and T4 DNA polymerase (Figs 2–4; Usdin and Woodford, unpublished observations). These sites are not an artifact of heating and cooling, since

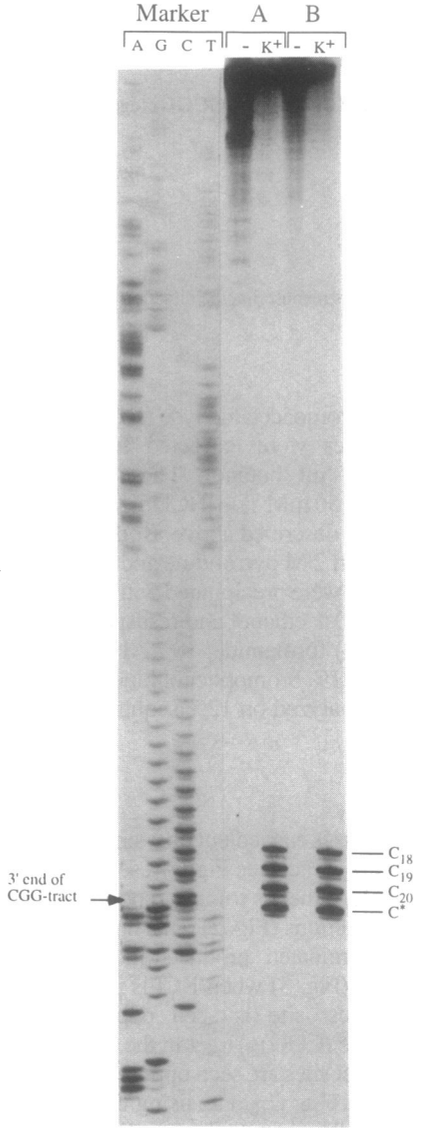

Figure 2. Arrest of DNA synthesis by a $(CGG)_{20}$ tract. Primer extension on single-stranded templates containing 20 consecutive CGG triplets without (–) and with (K^+) 40 mM KCl and either Sequenase™ (**A**) or Klenow polymerase (**B**). The primer used was SupFR4. The reactions were carried out as described in Materials and Methods. The marker is a sequencing reaction carried out with the same primer on the same template in the absence of K^+. The positions of the DNA synthesis arrest sites on the template sequence are indicated with a letter C and a subscript numeral. The letter identifies the base opposite which DNA synthesis is stopped. The number identifies in which CGG-repeat this base is located. Numbering of repeats begins at the 5′ most repeat on the template, with 20 being the 3′ most repeat in the $(CGG)_{20}$ tract. The template contains an additional C residue in the flanking sequence immediately following the 20th repeat. This C is designated by an asterisk (C*).

they are also seen when the templates are incubated at 37°C before addition of KCl. Neither the use of gene 32 protein nor the full T4 DNA primase–helicase replication complex eliminated these arrest sites (Nossal and Usdin, unpublished observations). Arrest of DNA synthesis is unlikely to be due to sequence context, since the same sites are seen with $(CGG)_{20}$ tracts cloned in different orientations and into different vectors, nor are they related to their distance from the priming site, since the same sites are also observed when a primer located 124 bases further 3′ on the template is used (data not shown). DNA synthesis arrest sites are not visible when the complementary strand is used as a template (Fig. 3B) and clones containing truncated versions of the $(CGG)_n$ tract in which $n = 8$ or that contain fewer than 13 uninterrupted CGG repeats show no such arrest sites (Fig. 4), indicating that arrest of DNA synthesis is not due to vector sequences or some artifact created at the junction of the vector and the CGG tract.

Templates containing 20 CGG repeats show at least eight sets of discernible K^+-dependent DNA arrest sites, depending on reaction conditions (Figs 3–5), while templates containing 16 CGG repeats show four arrest sites (Fig. 4). Longer exposures do not reveal any additional bands above background. These data suggest that 13 repeats are the minimum number required to cause DNA synthesis arrest under these conditions, since the last detectable arrest site is found at C_{14}, the first base 3′ of the 13th triplet. This is consistent with our finding that a template containing $(CGG)_{12}CGC(CGG)_2$, i.e. a maximum of 12 uninterrupted CGGs, shows no K^+-dependent arrest sites (Fig. 4).

These arrest sites can still be seen in samples that are incubated at 75°C for 5 min before addition of a thermostable polymerase, indicative of their thermal stability (Fig. 5). A template containing the sequence $(CGG)_{13}AGG(CGG)_6$ also formed barriers to DNA synthesis at 75°C (Fig. 5) and no significant difference between these templates is seen even after prolonged incubation at 85°C (data not shown), indicating that, in contrast to mutations that disrupt the $(NGG)_n$ motif, mutations that preserve this motif are still capable of forming stable DNA synthesis arrest sites. A small amount of K^+-independent pausing is seen in the middle of the $(CGG)_n$ tract at elevated temperatures. This we attribute to a combination of a weak K^+-independent structure, perhaps a triplex, and problems associated with the polymerase at these temperatures.

The N7 position of some guanines is necessary for formation of DNA synthesis arrest sites

A template in which some guanine residues were replaced with 7-deazaguanine was synthesized as described in Materials and Methods. This template shows no K^+-dependent arrest sites, while a control template produced the normal set of DNA arrest sites (Fig. 6). Since the N7 of 7-deazaguanine is unable to participate in hydrogen bonding, this result suggests that the structure(s) causing DNA synthesis arrest contain hydrogen bonds that involve the N7 of guanines. Since Watson–Crick base pairs do not involve the N7 position of guanine, presumably these are non-Watson–Crick interactions.

A $(CGG)_{20}$-containing oligonucleotide forms a folded DNA structure that is stabilized by K^+

Arrest of DNA synthesis was independent of template concentration over a wide range and was visible at template concentrations <1 nM (data not shown), suggestive of an intrastrand structure. Increased mobility of oligonucleotides on non-denaturing gels is diagnostic of an intramolecular structure, while reduced mobility is typical of intermolecular interactions. An 90mer containing 20 CGG repeats migrates in a manner consistent with its size on denaturing gels (Fig. 7B). However, in non-denaturing gels run at room temperature this oligonucleotide migrates faster than M1, an oligonucleotide 69 bases long which does not show any

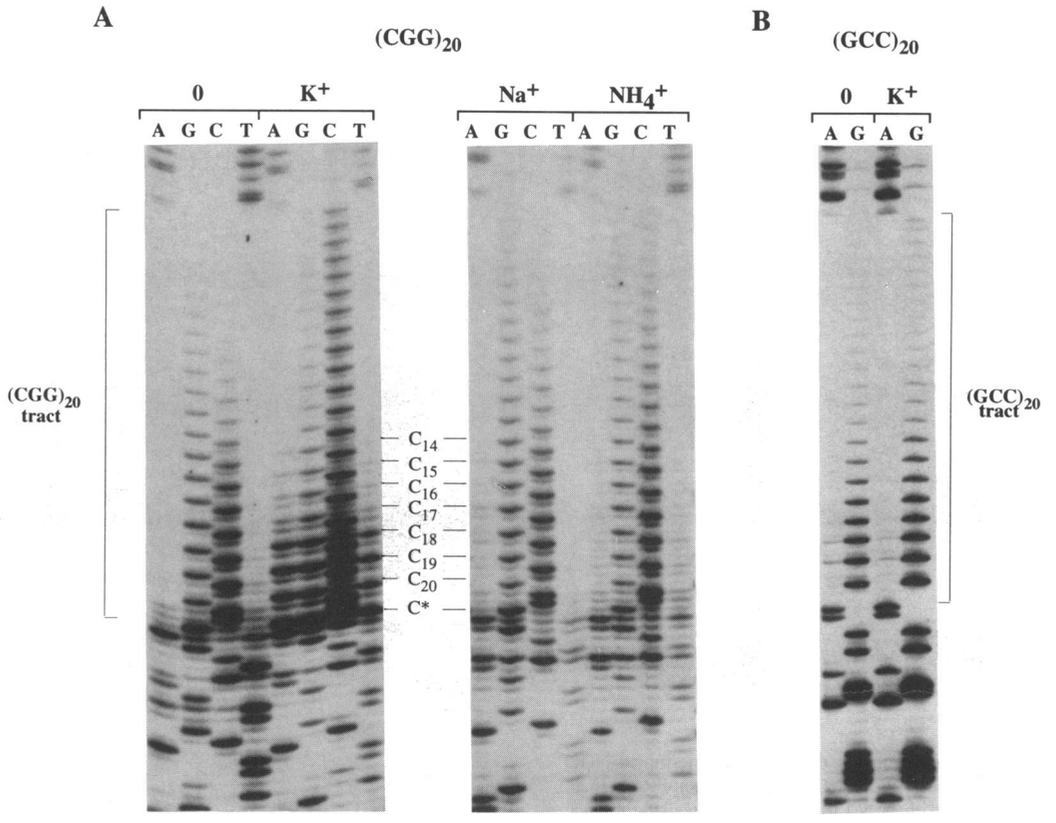

Figure 3. Arrest of DNA synthesis is K^+-specific and strand-specific. Primer extension reactions were carried out using Taq polymerase and dideoxynucleotides as described in Materials and Methods in the presence or absence of 40 mM of the indicated monovalent cation. The bases corresponding to the DNA arrest sites are indicated alongside the gel. (**A**) Analysis of CGG-containing template. (**B**) Analysis of the complementary GCC-containing template.

anomalous mobility relative to double-stranded DNA markers. The anomalous mobility is not likely to be due to the flanking sequences, since the M1 oligonucleotide has the same flanking sequences. The fastest mobility is seen when 40 mM KCl is present in the gel. The mobility of the oligonucleotide is slower in the presence of 40 mM LiCl and is slower still without added monovalent cation. A similar trend of mobility shift is seen with an oligonucleotide containing the same flanking sequence by with the CGG repeats replaced by $(T_2T_7)_4$, a telomere-like sequence known to form a very stable tetraplex (Weitzmann, Woodford and Usdin, manuscript in preparation). Methylation of the guanines by DMS eliminates the anomalous migration, indicating that the anomalous mobility is due to the formation of a structure involving hydrogen bonding at the N7 positions of guanines (data not shown).The behavior of the $(CGG)_{20}$-containing oligonucleotide during electrophoresis is consistent with an intramolecular folded structure whose compactness or stability is highest in KCl. No bands consistent with intermolecular structures were seen.

The $(CGG)_{20}$ oligonucleotide is protected from modification by BAA and DMS

The $(CGG)_{20}$-containing oligonucleotide was treated with BAA and DMS. BAA alkylates adenosines at N1 and N6 and cytosines at N3 and N4 (11), positions normally involved in base pairing interactions. Secondary modification of BAA-treated molecules by DMS or formic acid followed by β-elimination by pyrrolidine results in specific base cleavage. In the case of DMS-treated samples this cleavage is at all guanines as well as the cytosines that reacted with BAA. In the case of formic acid cleavage occurs at all purines as well as at BAA-reactive cytosines. BAA-dependent cleavage is thus diagnostic for cytosines that are unpaired. DMS reacts specifically with the N7 position of guanines on unmodified DNA and DMS protection is thus diagnostic of guanine hydrogen bonding arrangements that involve the N7 position.

Of all the cytosines in the CGG-tract only C11 was reactive with BAA, irrespective of whether secondary modification was carried out with DMS (Fig. 8) or formic acid (data not shown). The same pattern of BAA reactivity was seen both in the presence and absence of K^+ and no cleavage at cytosines is visible without BAA treatment (see DMS reactions in Fig. 8). Our data thus indicate that, with the exception of C11, which is unpaired in both the presence and absence of K^+, the cytosines are protected from BAA modification. When the oligonucleotide was treated with DMS in the presence of K^+, the guanine residues are protected from methylation, showing significantly less cleavage than the no K^+ control (Fig. 8). The DMS protection in the presence of K^+ indicates that the N7 position of each guanine in this tract is involved in a K^+-dependent hydrogen bonded structure. The almost complete protection from DMS modifica-

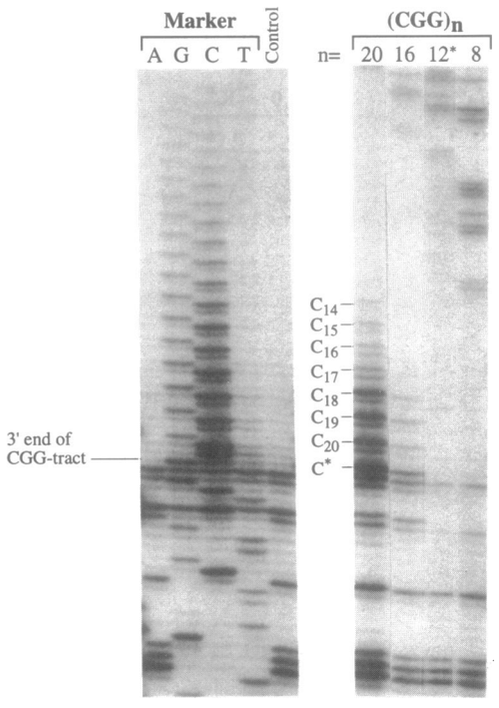

Figure 4. DNA arrest is dependent on the number of uninterrupted NGG triplets. Marker, sequencing ladder of M13(CGG)$_{20}$ using supFR4 primer. Control, primer extension reaction on the M13(CGG)$_{20}$ template in the absence of KCl. DNA synthesis by Taq polymerase in the presence of dideoxy-dATP and 40 mM KCl on various (CGG)$_n$ templates, where n = 8–20. From left to right, M13(CGG)$_{20}$, M13(CGG)$_{16}$, M13(CGG)$_{12}$CGC(CGG)$_2$ and M13(CGG)$_8$. SupFR4 was used as the primer in all reactions.

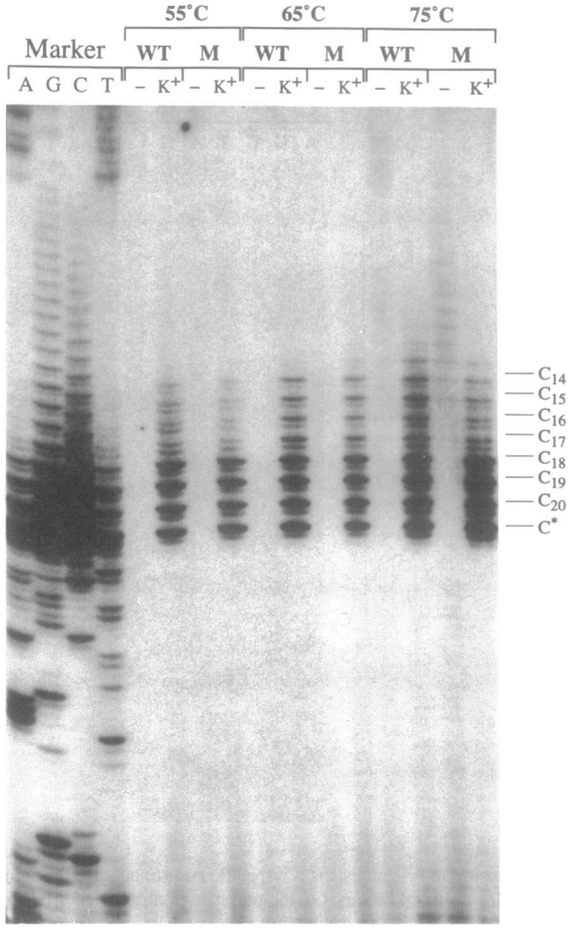

Figure 5. The barriers to DNA polymerase can be observed after incubation at elevated temperatures. Primer extension by Taq polymerase of the M13(CGG)$_{20}$ (WT) and M13(CGG)$_{13}$AGG(CGG)$_6$ (M) templates at various temperatures in the absence and presence of 40 mM KCl. The templates were incubated at 55°C for 30 s and then equilibrated at the indicated temperature for 5 min prior to addition of the polymerase. Primer extension was carried out at this same temperature. The marker is a sequencing ladder of the M13(CGG)$_{20}$ template produced in the presence of 40 mM KCl.

tion indicates that this structure involves guanine tetrads in which the guanines all act as N7 donors.

DISCUSSION

Templates containing a number of CGG repeats form a set of K^+-dependent barriers to DNA synthesis *in vitro*. These blocks to DNA synthesis are seen only when the CGG-rich strand serves as the template and are located at the 3′-end of the CGG tract. The fact that these barriers to DNA synthesis are located at this end rules out a triplex-like structure formed between the template and the nascent strand, since in these structures arrest of DNA synthesis occurs in the middle of the G-rich tract (12). Since these barriers are seen in the absence of the complementary strand, they are not due to interstrand triplex formation between the CGG-rich strand and its complement. The template concentration independence, together with the absence of slowly migrating species in gels of oligonucleotides containing CGG repeats, suggest that they are not the result of intermolecular interactions.

The fact that the migration of CGG tracts in non-denaturing gels is faster than predicted from its molecular weight, particularly when KCl is present, together with the observation that this anomalous migration is eliminated when guanines in the oligonucleotide are methylated by DMS, indicates that the CGG tract forms a K^+-stabilized intrastrand structure that involves the N7 position of at least some guanines. 7-Deazaguanine substitution experiments confirm that arrest of DNA synthesis requires the availability of the N7 position of guanine, ruling out a hairpin containing only G-C base pairs (13), since in a G-C pair N7 of guanine is not involved in hydrogen bonding. Since blocking of just four of the 32 (12.5%) guanines in the sequence (CGG)$_{16}$ is sufficient to abolish the arrest sites, it is unlikely that the block to DNA synthesis is due to a hairpin containing only G-G base pairs or a mixture of G-C and G-G base pairs. In such structures only 50 and 33% of the guanines, respectively, would have to be N7 donors.

The specific requirement for K^+ is difficult to rationalize in terms of a simple hairpin. Not only do the binding constants of alkali metal ions to the phosphate groups in DNA decrease slightly with increasing metal ion radius, but neither G-C hairpins nor G-G hairpins are K^+-dependent (Weitzmann, Woodford and Usdin, manuscript in preparation). On the other hand, the K^+ specificity of these arrest sites is reminiscent of the ionic requirements of tetraplexes. The ion specificity of tetraplexes is thought to be related to the size of the cavity created by a series of consecutive guanine tetrads. The ionic radius of K^+ is such that it is both small enough to fit inside the cavity and large enough to be able to form an octahedral coordination complex with the O6 oxygen atoms in adjacent tetrads, thus stabilizing the tetraplex

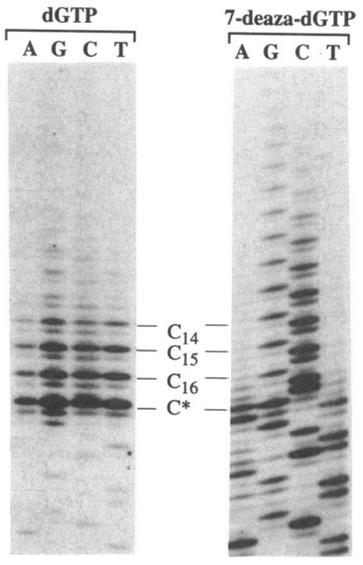

Figure 6. The N7 position of guanines is required for formation of the block to DNA synthesis. A PCR fragment containing a $(CGG)_{16}$ tract in which the N7 positions of a guanine in four adjacent triplets at the 3′-end were blocked was synthesized by PCR using a oligonucleotide containing four 7-deazaguanine residues as described in Materials and Methods. A template synthesized using an unsubstituted oligonucleotide was synthesized and purified in parallel. The control fragments containing no 7-deazaguanine and the 7-deazaguanine-substituted fragments were then used as templates for DNA synthesis by Taq polymerase in the presence of KCl as described in Materials and Methods.

(14). The most parsimonious explanation of the sequence requirements of the arrest sites, their intrastrand nature, the requirement for the N7 position of guanines and the K^+ ion specificity is the formation of some sort of intrastrand tetrahelical structure.

This is consistent with our chemical modification data. In both the presence and absence of K^+ the 11th cytosine in the $(CGG)_{20}$ tract is strongly reactive with BAA, a reagent specific for unpaired cytosines (and adenines), while the other cytosines are either unreactive or poorly reactive. This suggests that most of the cytosines in the CGG tract are involved in hydrogen bonding or are otherwise protected from BAA modification, both in the presence and absence of K^+. In the absence of K^+ an almost uniform strong reactivity of the CGG tract is seen to DMS, a reagent specific for the N7 position of guanines. However, in the presence of K^+ most of the guanines in the CGG tract are almost completely protected from methylation by DMS. The position of the BAA-reactive cytosine in the middle of this tract seen in the absence of K^+ would be consistent with its position in the central loop of a hairpin; the lack of reactivity of the remaining cytosines indicates that they are hydrogen bonded. This structure may account for the trace amount of premature chain termination sometimes seen in the absence of added monovalent cation.

In the presence of K^+ a transition occurs to a structure in which the guanines are almost completely protected from DMS modification. This strong protection is diagnostic of a tetraplex structure. The BAA protection of the cytosines seen without K^+ is preserved. The unpaired cytosine in the 11th triplet could be fitted into the central loop of the tetraplex, while the residues that would be located in the two flanking loops are apparently not reactive. This lack of reactivity may be due to interactions

A

OLIGO	SEQUENCE	SIZE (bases)
1	5'-gtacgaatt.....$(CGG)_{20}$...ctcgagtcaacgtaacacttt-3'	90
2	5'-gtacgaatt.....$(T_2G_7)_4$...ctcgagtcaacgtaacacttt-3'	66
M1	5'-gtacgaattctaga..$(C)_{35}$.....tcgagtcaacgtaacacttt-3'	69
M2	5'-ataatgcatacgatttaggtgacactatagaataccagctgtggtggtgaagctgtagcc-3'	60

B

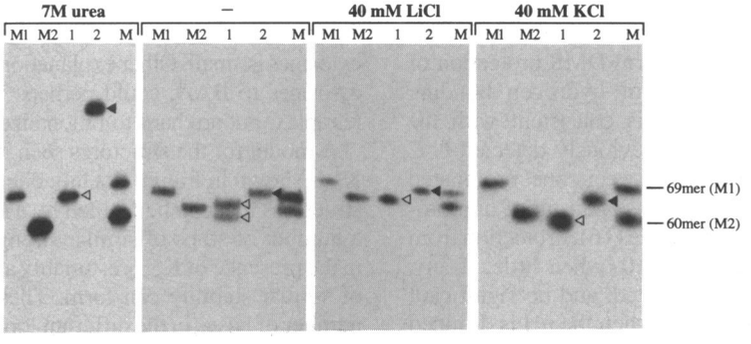

Figure 7. A $(CGG)_{20}$ oligonucleotide migrates anomalously on non-denaturing gels. The $(CGG)_{20}$-containing oligonucleotide was electrophoresed alongside various marker oligonucleotides on 10% polyacrylamide gels containing 0.5× TBE buffer. (**A**) Sequences of the oligonucleotides tested. The labeling scheme in this panel is the same as the labeling scheme in each of the gels shown in (B). M1 and M2 are marker oligonucleotides. Note that oligo 1, oligo 2 and the M1 oligonucleotide contain the same flanking sequences. (**B**) Gels containing 7 M urea, no additions (–), 40 mM LiCl or 40 mM KCl as indicated. The arrowheads indicate the bands showing variable gel mobility. Open arrowhead, the $(T_2G_7)_4$-containing oligonucleotide (oligo 1); filled arrowhead, the $(CGG)_{20}$-containing oligonucleotide (oligo 2). The lane labeled M contains an equimolar mixture of the M1 and M2 oligonucleotides.

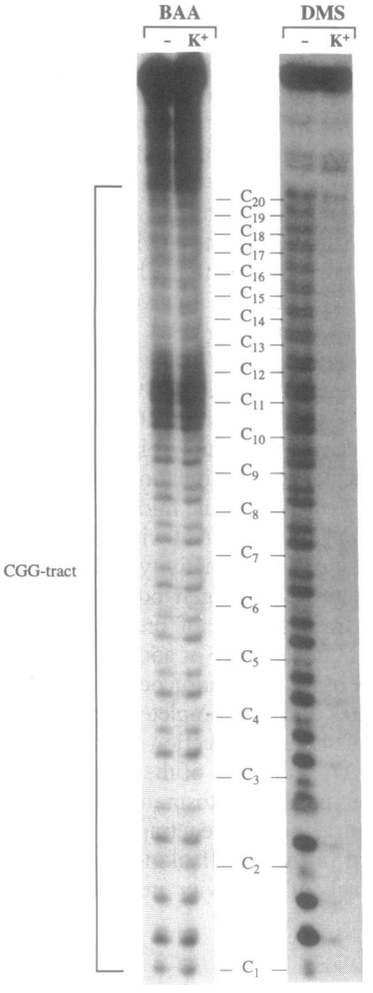

Figure 8. Chemical modification of a $(CGG)_{20}$-containing oligonucleotide. The $(CGG)_{20}$-containing oligonucleotide was prepared and treated with DMS and BAA as described in Materials and Methods. The C residue of each triplet in the $(CGG)_{20}$ tract is labeled based on the numbering scheme shown in Figure 1.

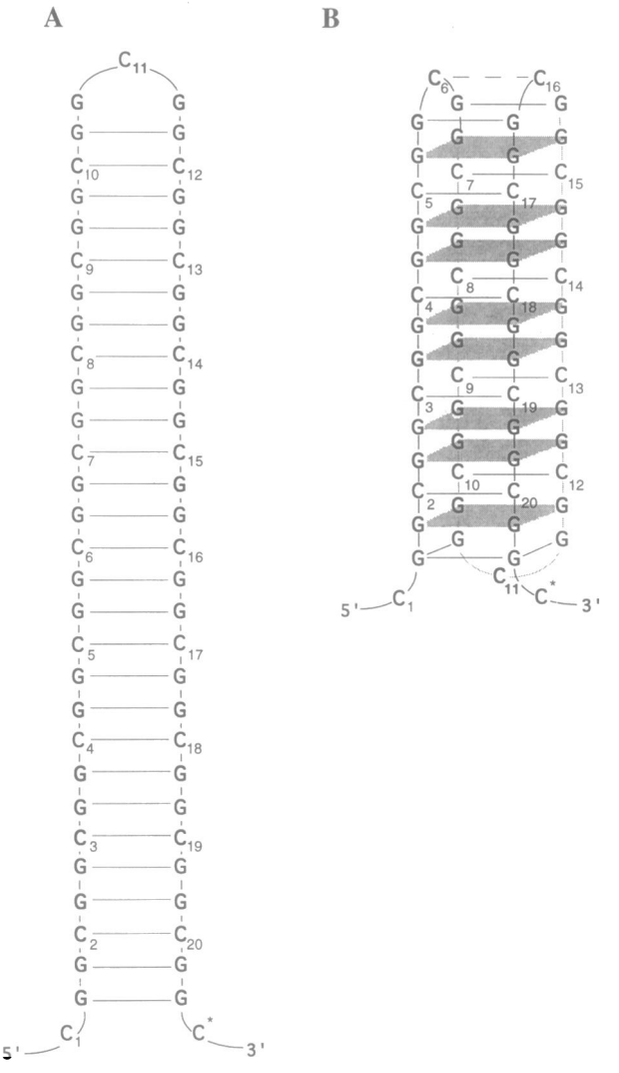

Figure 9. Models for the K^+-independent (**A**) and K^+-dependent (**B**) structures formed by a $(CGG)_{20}$ tract. Only the two structures most consistent with our data are shown. Tetrads are indicated as gray parallelograms. Hypothetical base interactions amongst loop bases are shown as dotted lines. Different loop sizes, loop orientations and tetrad numbers are possible for the tetraplex.

between these loops, which would be located on the same side of the tetrad stem.

A mixture of G-C base pairs and G quartets or tetrads in this structure is not likely, since almost uniform DMS protection of guanines is observed. A combination of hydrogen bonding between cytosines (15) and G tetrads is consistent with the observed chemical modification data. Previously detected C-C pairs in solution involve interaction between one protonated cytosine and one neutral one, although base pairing of unprotonated cytosines is also theoretically possible (16). Protection from BAA modification is seen even at pH 9.0, when little, if any, protonation of cytosines would be expected, and no significant increase in DNA synthesis arrest occurs when the pH is dropped to 6.5. While NMR experiments have demonstrated that some protonation of cytosines is seen at pH 8.0 (17), it is also possible that the hydrogen bonding interaction involves two neutral cytosines. Since the substitution of an A for a C in the $(CGG)_{20}$ tract does not affect DNA synthesis arrest at high pH, it seems likely that the contribution to tetraplex stability of pairing between cytosines is small. Other explanations, such as inaccessibility of the cytosines to BAA, could perhaps be invoked in the case of the tetraplex, but are hard to rationalize in the case of the hairpin.

A model for the structures seen in the presence and absence of K^+ is shown in Figure 9. Only one of the possible K^+-dependent structures that can be formed by a $(CGG)_{20}$ tract is shown. Since a number of stops of similar strength are seen for this sequence in the presence of K^+, presumably a number of different structures of similar stability can form. These structures may vary in the number of bases in the different loops, the orientation of the loops and in the number of CGG triplets involved. The precise relationship between the K^+-independent structure and the K^+-dependent ones is not known at this time. Given some similarities between the two structures, it is tempting to speculate that the K^+-independent structure represents a folding intermediate in the pathway of formation of the K^+-dependent tetraplex, but

such conclusions await a more careful kinetic analysis of the folding reaction.

While this work was in progress it was reported that oligonucleotides containing seven or more CGG triplets could form stable intermolecular structures, presumably tetraplexes, since the stochiometry of the structure was tetramolecular and the guanines were protected from DMS modification (18). It was suggested that longer sequences could form intramolecular tetrahelical structures. The data presented here are consistent with this idea. Tetraplex formation by the CGG-containing strand provides a rationale for the strand-specific nature of the arrest sites. The complementary strand is much less G-rich and therefore is unable to form the same type of structure and thus does not block DNA synthesis in the same way.

There is both direct and indirect evidence for tetraplex formation *in vivo* (19,20). We would predict that tetraplex formation would occur rapidly any time the CGG-rich strand was unpaired. DNA replication might provide such an opportunity, since large regions of the genome may be single-stranded at this time (21). Factors such as the rate of replication, the presence of binding proteins that sequester the complementary strand, and whether the CGG tract is on the lagging versus the leading strand of DNA synthesis may all affect the chances of tetraplex formation and therefore DNA synthesis arrest. It is possible that these factors vary with the developmental stage and may only be observed *in vivo* at specific times or in specific cell types.

These structures may promote triplet expansion by preventing progression of the polymerase during DNA replication, thus favoring repeated strand slippage, or by stabilizing the strand slippage intermediate. It is possible that the blocks to DNA synthesis provided by such structures may also account for chromosome fragility, since delayed replication, which would be associated with the formation of such structures, could account for the appearance of the chromatin at the fragile site. Structures such as these are likely to be stabilized by methylation of cytosines (15). This could explain why 5-azacytidine, an inhibitor of DNA methyltransferase, reduces fragile site expression (22), while methionine, a precursor of the methyl donor in this reaction, S-adenosylmethionine, induces expression (23).

The number of repeats necessary to form a barrier to DNA synthesis is within the normal range for the fragile X locus and is not generally associated with triplet expansion. However, recent evidence suggests that even alleles with repeat numbers in the normal range may occasionally give rise to alleles that show a large increase in repeat number (24). It is also possible that *in vivo* more repeats are necessary for structure formation or that repeated strand slippage at successive arrest sites is required. In addition, at least with respect to chromosome fragility, there is a continuum of expression of the fragile site that ranges from very low levels in normal individuals to high levels of expression in affected individuals (25).

Alleles containing more than 24 uninterrupted CGGs (8) are over-represented in fragile X chromosomes and alleles containing two or more AGG interruptions are rarely if ever found in fragile X chromosomes (5). While our data suggest that the contribution of cytosines to the overall structure is small, it is possible that *in vivo* the contribution of cytosines to tetraplex stability is more important, particularly if the structure formed by the pure CGG tract is only marginally stable.

REFERENCES

1 Fu,Y.H., Kuhl,D.P., Pizzuti,A., Pieretti,M., Sutcliffe,J.S., Richards,S., Verkerk,A.J., Holden,J.J., Fenwick,R.G., Warren,S.T., Oostra,B.A., Nelson,D.L. and Caskey,C.T. (1991) *Cell*, **67**, 1047–1058.
2 Kremer,E.J., Pritchard,M., Lynch,M., Yu,S., Holman,K., Baker,E., Warren,S.T., Schlessinger,D., Sutherland,G.R. and Richards,R.I. (1991) *Science*, **252**, 1711–1714.
3 Knight,S.J., Voelckel,M.A., Hirst,M.C., Flannery,A.V., Moncla,A. and Davies,K.E. (1994) *Am. J. Hum. Genet.*, **55**, 81–86.
4 Hirst,M.C., Barnicoat,A., Flynn,G., Wang,Q., Daker,M., Buckle,V.J., Davies,K.E. and Bobrow,M. (1993) *Hum. Mol. Genet.*, **2**, 197–200.
5 Snow,K., Tester,D.J., Kruckeberg,K.E., Schaid,D.J. and Thibodeau,S.N. (1994) *Hum. Mol. Genet.*, **3**, 1543–1551.
6 Glover,T.W. and Stein,C.K. (1988) *Am. J. Hum. Genet.*, **43**, 265–273.
7 Nancarrow,J.K., Kremer,E., Holman,K., Eyre,H., Doggett,N.A., Le Paslier,D., Callen,D.F., Sutherland,G.R. and Richards,R.I. (1994) *Science*, **264**, 1938–1941.
8 Kunst,C.B. and Warren,S.T. (1994) *Cell*, **77**, 853–861.
9 Nussbaum,R.L., Airhart,S.D. and Ledbetter,D.H. (1986) *Am. J. Med. Genet.*, **23**, 457–466.
10 Woodford,K.J., Howell,R.M. and Usdin,K. (1994) *J. Biol. Chem.*, **269**, 27029–27035.
11 Kohwi,Y. and Kohwi-Shigematsu,T. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 3781–3785.
12 Baran,N., Lapidot,A. and Manor,H. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 507–511.
13 Wells,R.D. and Sinden,R.R. (1993) In Davies,K.E. and Warren,S.T. (eds), *Genome Rearrangement and Stability*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 107–138.
14 Sen,D. and Gilbert,W. (1990) *Nature*, **344**, 410–414.
15 Hardin,C.C., Corregan,M., Brown,B.2. and Frederick,L.N. (1993) *Biochemistry*, **32**, 5870–5880.
16 Aboul-ela,F., Koh,D., Tinoco,I.J. and Martin,F.H. (1985) *Nucleic Acids Res.*, **13**, 4811–4824.
17 Sklenar,V. and Feigon,J. (1990) *Nature*, **345**, 836–838.
18 Fry,M. and Loeb,L.A. (1994) *Proc. Natl. Acad. Sci. USA*, **91**, 4950–4954.
19 Hammond-Kosack,M.C., Kilpatrick,M.W. and Docherty,K. (1992) *J. Mol. Endocrinol.*, **9**, 221–225.
20 Liu,Z. and Gilbert,W. (1994) *Cell*, **77**, 1083–1092.
21 Kunkel,T.A. (1992) *J. Biol. Chem.*, **267**, 18251–18254.
22 Mixon,J.C. and Dev,V.G. (1983) *Am. J. Hum. Genet.*, **35**, 1270–1275.
23 McDermott,A., Walters,R., Howell,R.T. and Gardner,A. (1983) *J. Med. Genet.*, **20**, 169–178.
24 Macpherson,J.N., Bullman,H., Youings,S.A. and Jacobs,P.A. (1994) *Hum. Mol. Genet.*, **3**, 399–405.
25 Nussbaum,R.L. and Ledbetter,D.H. (1986) *Annu. Rev. Genet.*, **20**, 109–145.