NGG-triplet repeats form similar intrastrand structures: implications for the triplet expansion diseases

K. Usdin*

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

Received March 10, 1998; Revised and Accepted July 14, 1998

ABSTRACT

Tandem repeats of certain trinucleotides show extensive intergenerational instability in humans that is associated with a class of genetic disorders known as the Triplet Expansion Diseases. This instability is thought to be a consequence of the formation of intrastrand structures, including hairpins, triplexes and tetraplexes, by the tandem repeats. I show here that CGG-repeats which are associated with this group of diseases, and AGG- and TGG-repeats which are not currently known to be, form several intrastrand structures including tetraplexes. In all cases the tetraplexes have the same overall conformation in which all the G residues are involved in G4-tetrads. CGG-repeats also form stable hairpins, but AGG- and TGG-repeats do not form hairpins of comparable stability. However, since tetraplexes can be thought of as folded hairpins, many of the properties ascribed to disease-associated triplets that form hairpins, may apply to these sequences as well. The fact that AGG- and TGG-repeats are not currently associated with any triplet expansion disease suggests either that the ability to adopt an intrastrand folded structure is not sufficient for expansion, or that other diseases associated with such triplets might remain to be identified.

INTRODUCTION

Tandem arrays of trinucleotides (triplet repeats) are abundant in eukaryote genomes, and a subset of these repeats is associated with the Tandem Expansion Disorders. These are neurological or neuromuscular disorders that result from an increase (expansion) with the Triplet Expansion Disorders. These are neurological or eukaryote genomes, and a subset of these repeats is associated with this group of diseases, and AGG- and TGG-repeats which are not currently known to be, form several intrastrand structures including tetraplexes. In all cases the tetraplexes have the same overall conformation in which all the G residues are involved in G4-tetrads. CGG-repeats also form stable hairpins, but AGG- and TGG-repeats do not form hairpins of comparable stability. However, since tetraplexes can be thought of as folded hairpins, many of the properties ascribed to disease-associated triplets that form hairpins, may apply to these sequences as well. The fact that AGG- and TGG-repeats are not currently associated with any triplet expansion disease suggests either that the ability to adopt an intrastrand folded structure is not sufficient for expansion, or that other diseases associated with such triplets might remain to be identified.

MATERIALS AND METHODS

Oligonucleotides

The oligodeoxyribonucleotides used in this study are listed in Table 1. Oligodeoxyribonucleotides containing a mixture of A, C, G, T or U were purchased from Integrated DNA Technologies, Inc. (Coralville, IA) or Gibco BRL (Gaithersburg, MD). Oligonucleotides to be used in chemical modification or gel mobility experiments were purified by polyacrylamide gel electrophoresis. Oligonucleotides (25 ng) were end-labeled with [γ-32P]ATP (Dupont-NEN, 3000–6000 Ci/mmol) using T4 polynucleotide kinase (Epicentre Technologies, Inc.), and a buffer containing 50 mM Tris–HCl, pH 9.3, and 10 mM MgCl2 (*GG)*, where * indicates an abasic site, was prepared by treatment of 5′ end-labeled (UGG)20 with uracil DNA glycosylase (Gibco BRL, Gaithersburg, MD) according to the supplier’s instructions. Complete removal of uracil residues was assessed by heating the sample to 90°C for 10 min followed by electrophoresis on a 20% sequencing gel. The gel was covered with plastic film and exposed to X-ray film overnight at −70°C. A 10 base product was

*Correspondence should be addressed to: Building 8, Room 202, National Institutes of Health, 8 Center Dr MSC 0830, Bethesda, MD 20892-0830, USA.
Tel: +1 301 496 2189; Fax: +1 301 402 0053; Email: ku@helix.nih.gov
Figure 1. Model for the (CGG)$_{20}$ tetraplex. This tetraplex contains a mixture of G$_4$-tetrads and C•C$^+$ pairs in a 2:1 ratio. G$_4$-tetrads in the tetraplex are indicated by the gray parallelograms. The hydrogen bonding scheme for the G$_4$-tetrads is shown alongside the tetraplex. The hydrogen bonds in the C•C$^+$ pairs are shown by the dotted lines. The choice of which cytosine in each pair was shown to be protonated was made arbitrarily. The hydrogen bonding scheme for G$_4$-tetrads and C•C$^+$ pairs is shown alongside the tetraplex.

observed corresponding to the number of bases 5′ of the first expected abasic residue in the oligodeoxyribonucleotide.

Table 1. Oligodeoxyribonucleotides used in this study

<table>
<thead>
<tr>
<th>Oligodeoxyribonucleotide</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>(AGG)$_{20}$</td>
<td>d(GTACGAATTC)(AGG)$_{20}$CTCGAGCTACGTAACACTTTT</td>
</tr>
<tr>
<td>(TGG)$_{20}$</td>
<td>d(GTACGAATTC)(TGG)$_{20}$CTCGAGCTACGTAACACTTTT</td>
</tr>
<tr>
<td>(CGG)$_{20}$</td>
<td>d(GTACGAATTC)(CGG)$_{20}$CTCGAGCTACGTAACACTTTT</td>
</tr>
<tr>
<td>(UGG)$_{20}$</td>
<td>d(GTACGAATTC)(UGG)$_{20}$CTCGAGCTACGTAACACTTTT</td>
</tr>
<tr>
<td>(GG)$_{30}$</td>
<td>d(GTACGAATTC)(GG)$_{20}$CTCGAGCTACGTAACACTTTT</td>
</tr>
<tr>
<td>AMP2</td>
<td>d(GGCGCACAGGAAATTTGGA)</td>
</tr>
<tr>
<td>supFR1</td>
<td>d(GATCGAATTCGTCGACATGTGTTGGGGA)</td>
</tr>
<tr>
<td>supFR4</td>
<td>d(TGTTTTTACTGCGCTGCT)</td>
</tr>
<tr>
<td>T$<em>2$G$</em>{34}$</td>
<td>d(GTACGAATTCGTCGACATGTGTTGGGGA)</td>
</tr>
<tr>
<td>C$_4$</td>
<td>d(GTACGAATTCGTCGACATGTGTTGGGGA)</td>
</tr>
<tr>
<td>2xGGGC</td>
<td>d(GTACGAATTCGTCGACATGTGTTGGGGA)</td>
</tr>
</tbody>
</table>

*An abasic site.

Clone construction

Oligonucleotides were cloned into the plasmid pMS189A as previously described (12,16). Plasmids were replicated in Escherichia coli MBM7070, isolated by alkaline lysis, and purified by CsCl gradient centrifugation according to standard procedures.

Preparation of 7-deazaguanine and 7-deazaadenine containing templates

Templates containing 7-deazaguanine or 7-deazaadenine were prepared by PCR amplification of plasmids containing the (AGG)$_{20}$ sequence using the primers AMP2 and supFR1 which flank the repeat. The primer binding sites are located ~500 bases apart, the precise distance depending on the template, with the repeat being located about half way between the two primer binding sites. Reaction mixtures containing 10 ng of plasmid template DNA containing the repeat of interest; 1 µM each of AMP2 and supFR1; 2.5–5 U Taq polymerase (Gibco BRL); 50 mM Tris–HCl, pH 8.0; 10 mM MgCl$_2$; 100 µM of each dNTP. Templates containing 7-deazaguanine were prepared by using 7-deaza-dGTP instead of dGTP in the reaction mix. Templates containing 7-deazaadenine were prepared by using 7-deaza-dATP in place of dATP in the reaction. The reaction mixes were then overlaid with a drop of mineral oil and subjected to 30 cycles of 30 s at 95°C, 30 s at 55°C and 30 s at 72°C in a Perkin-Elmer-Cetus Thermal Cycler. The PCR products were purified on a 5% polyacrylamide gel and used as templates in the tetraplex assay described below.

Intrastrand tetraplex assay

Sequences were tested for the ability to block DNA synthesis reactions as follows (17). Reaction mixtures (total volume 6 µl) contained 0.2–2 nM of plasmid or PCR template containing the repeat of interest, 0.16 nM of the 5′ end-labeled supFR4, 10 µM dNTPs, one of the following dideoxynucleotides at the concentration indicated in parentheses: ddATP (0.3 mM), ddGTP (0.017 mM), ddCTP (0.2 mM), ddTTP (0.6 mM), 50 mM Tris–HCl (pH 9.3), 2.5 mM MgCl$_2$, 5 U of Taq polymerase (Gibco BRL) and, where indicated, 50 mM of monovalent cation. Reaction mixtures were subjected to 30 rounds of heating and cooling in a Perkin-Elmer PCR machine for 30 s at 95°C, 30 s at 55°C and 30 s at 72°C. The reaction was terminated by the addition of one half volume of stop buffer containing 95% (v/v) formamide, 10 mM EDTA (pH 9.5), 10 mM NaOH, 0.1% xylene cyanol and 0.1% bromophenol blue, and the mixtures heated at 90°C for 5 min prior to electrophoresis on a 6.5% polyacrylamide sequencing gel.

Conformational analysis of oligodeoxyribonucleotides

Oligodeoxyribonucleotides were heated to 90°C for 5 min and then subjected to electrophoresis at 6 V/cm on a 10% polyacrylamide gel containing either 1× TBE alone, or with either 7 M urea or 100 mM KCl added. In the case of the KCl-containing gels, 100 mM KCl was added to the running buffer as well. Gels were covered with plastic wrap and exposed to X-ray film at –70°C.

Chemical modifications of NGG-containing oligodeoxyribonucleotides

End-labeled oligodeoxyribonucleotides (1–5 ng per reaction) were heated in TE buffer for 1 min at 90°C. They were then incubated either in a buffer containing 40 µl of 50 mM Tris–HCl pH 9.3, with either no added cations, or 2 mM MgCl$_2$ with or without 50 mM KCl or NaCl as indicated, or in a buffer containing 2 mM MgCl$_2$ and either 50 mM sodium acetate (pH 6.0) or 50 mM potassium acetate (pH 6.0) as indicated. Reactions were then heated for 30 s at 95°C, 30 s at 55°C and then incubated for 60 min at 72°C. One microliter of DMS, 2 µl of
DMS reactions were carried out for 1 min at room temperature. 20 mM KMnO₄ or 4 µl of DEPC was added to each tube. The DMS reactions were carried out for 20 min at 37°C. Both reactions were terminated by precipitation with 1 ml butanol. Reaction mixtures containing KMnO₄ were incubated for 5 min at 37°C before termination of the reaction by the addition of 8 µl of 2.5 M β-mercaptoethanol. These reactions were then precipitated twice with 1 ml butanol. The precipitates were then resuspended in 1 M piperidine and cleavage at modified residues effected by incubation at 90°C for 30 min. Samples were precipitated twice with 1 ml butanol, dried under vacuum, dissolved in 20 µl of 42.5% (v/v) formamide, 5 mM EDTA (pH 9.5), 5 mM NaOH, 0.05% xylene cyanol, 0.05% bromphenol blue, denatured for 2 min at 90°C and the cleavage products resolved by electrophoresis on a 20% sequencing gel. Gels were covered with plastic wrap and exposed to X-ray film at –70°C.

RESULTS

There are a variety of biophysical and biochemical methods that can be used to examine intrastrand DNA structures including electrophoretic mobility shift analysis (10), the formation of cation-specific blocks to DNA synthesis (12,16–18), sensitivity to single strand-specific enzymes (10) and modification by specific chemical probes (9,12,16,19,20). A combination of these methods was used to analyse the intrastrand structures formed by CGG-, TGG- and AGG-repeats.

AGG-repeats form K⁺-dependent blocks to DNA synthesis

We have previously shown that both CGG-repeats and TGG-repeats form K⁺-specific blocks to DNA synthesis that are diagnostic of the formation of intrastrand tetraplexes (12,18). Potassium ions are thought to be uniquely able to stabilize tetraplexes because they are small enough to fit inside the central cavity of the tetraplex, yet large enough to coordinate the eight O₆ oxygen atoms in successive tetrads (21). On templates containing either CGG- or TGG-repeats premature chain termination is seen opposite successive non-G bases at the 3’ end of the repeat tract on the template as summarized in Figure 2. Here I show that templates containing the sequence (AGG)₂₀ form two classes of blocks to DNA synthesis (Fig. 3). One class of blocks is seen when Mg²⁺ is the sole cation in the reaction, while the second class is K⁺ dependent. The same pattern of premature chain termination is seen with a variety of polymerases, and is independent of temperature over the range 37–85°C. This pattern is seen whether templates are incubated for 30 s at any one of these temperatures before the polymerase is added, or if 30 cycles of denaturation and annealing/extension are used (data not shown). The latter result suggests that both sets of blocks form rapidly, and do not require extended incubations at high temperature or repeated cycles of heating and cooling. Both classes of blocks are independent of template concentration over a wide range, are only seen when the template is the purine-rich strand, and occur even in the absence of the complementary strand (Fig. 3, left panel). These blocks do not form when the template adenines are replaced with 7-deaza-adenine (c⁷A) (Fig. 3, right hand panel). 7-deaza-adenine is an isosteric analog of adenine that contains a carbon atom rather than a nitrogen atom at the 7 position, and consequently also has an altered π-electron shell. The effect of c⁷A suggests either that the N⁷ position is involved in hydrogen bonding in the structure that blocks DNA synthesis, or that stacking energy contributions from adenines are important for its stability (22).

When Mg²⁺ is the only cation present, (AGG)₂₀-containing templates produce multiple blocks to DNA synthesis in the middle of repeat tract (indicated by the open arrows in Fig. 3). The position and properties of these blocks are consistent with the formation of a purine-purine:pyrimidine triplex during DNA synthesis as has been previously observed in tracts of alternating adenines and guanines (23). In such a triplex the unoccupied 3’ end of the repeat on the purine-rich template strand folds back and hydrogen bonds with bases in the 5’ half of the newly copied purine-rich tract. The copied region remains hydrogen bonded to the nascent pyrimidine-rich strand via Watson–Crick base pairs. Of the 16 possible base triplets of the purine-purine:pyrimidine type, those involving A*G:T and G*G:C are the most stable (24,25). A triplex containing A*G:T and G*G:C triplets in a 1:2 ratio is thus the most likely triplex to be formed by this sequence. In A*G:T triplets, the adenine that is hydrogen bonded to the thymine in the nascent strand also acts as an N⁷ donor for the second adenine in a reversed Hoogsteen hydrogen bond. This hydrogen bonding arrangement could account for the inability of c⁷A-substituted templates to form these blocks to DNA synthesis.

In the presence of both Mg²⁺ and K⁺ the purine-rich templates produce an additional set of blocks to DNA synthesis. This time the

![Figure 2](https://academic.oup.com/nar/article-abstract/26/17/4078/1177721/4080)

**Figure 2.** Summary of DNA synthesis arrest sites on (CGG)₂₀ and (TGG)₂₀-containing templates. The position of the K⁺-dependent blocks to DNA synthesis are marked by black arrows. No other blocks to DNA synthesis are seen. The flanking bases have no effect on the pattern of DNA synthesis arrest and for this reason are simply indicated by the letter ‘N’. The position of the non-G base in the repeat is indicated.
Figure 3. DNA synthesis arrest assay on templates containing \((\text{AGG})_{20}\). (A) DNA synthesis arrest by the \((\text{AGG})_{20}\) sequence on either single-stranded templates containing adenines or on PCR-generated templates in which all the adenines were replaced by 7-deaza-adenine (c\(^7\)A) was tested as described in the Materials and Methods section, in the absence of added monovalent cation (0), or in the presence of 50 mM of the indicated cation. T, C, G and A indicate the bases on the template strand. The location of the \((\text{AGG})_{20}\) tract within the template is indicated by the bracket on the right hand side of the figure. The A residues in the template are labeled, with A\(_1\) being the 5'-most, and A\(_{58}\) being the 3'-most adenine in the repeat array. The white arrows mark the major monovalent cation-independent blocks to DNA synthesis, and the black arrows the K\(^+\)-specific blocks. The pattern of DNA synthesis arrest is summarized in (B). The same numbering convention as (A) has been used. The flanking bases which do not affect the pattern of DNA synthesis arrest are indicated by the letter ‘N’.

the premature chain terminations occur at the 3' end of the G-rich strand of the repeat tract, once again opposite the A residues in successive triplets (marked by the filled arrows in Fig. 3). These blocks are not seen in the absence of monovalent cation, or in the presence of Li\(^+\), Na\(^+\), Rb\(^+\) or Cs\(^+\) (Fig. 3). The properties of these blocks are consistent with the formation of intrastrand tetraplexes. A similar series of K\(^+\)-dependent blocks to DNA synthesis is seen on templates containing the sequence \((\text{UGG})_{20}\) (data not shown). The similarity between the pattern of K\(^+\)-dependent premature chain termination on the \((\text{AGG})_{20}\)-containing template as summarized in Figure 3B, and that of the \((\text{CGG})_{20}\)- and \((\text{TGG})_{20}\)-containing templates (Fig. 2), suggests that the underlying structures that are responsible for the K\(^+\)-specific blocks to DNA synthesis are similar.

The chemical reactivities of the CGG-, TGG- and AGG-tetraplexes are similar

Dimethylsulfate (DMS) methylates the N\(^7\) of guanine which is not involved in Watson–Crick base pairing (26). Therefore protection of G residues from this reagent is diagnostic of the formation of non-Watson–Crick base interactions. Diethylpyrocarbonate (DEPC) carboxyethylates A residues at the N\(^7\) position particularly when the A residues are unpaired or in the syn conformation (27). Potassium permanganate oxidizes 5,6-carbon–carbon double bond of unpaired T residues (28).

At pH 9.0 in the presence of K\(^+\), or at pH 6.5 in the absence of K\(^+\), the second G in each repeat of the sequence \((\text{CGG})_{20}\) shows some protection from DMS modification (Fig. 4, first and third lanes). In contrast, the first G in each repeat shows little if any DMS protection with the extent of DMS modification being comparable to that seen for guanines outside the repeat. This pattern of modification is consistent with the formation of a hairpin containing a mixture of G•C and G•G base pairs as previously described (7). The first G in the repeat participates in the G•C pair and since this base pairing scheme does not involve the N\(^7\) position, this residue is sensitive to DMS modification. The second G in each repeat is involved in a Hoogsteen G•G pair with a guanine on the opposite side of the hairpin. In this G•G pair each G acts as an N\(^7\) donor 50% of the time, and thus shows partial
(AGG)$_{20}$- and (TGG)$_{20}$-repeats show no DMS protection of guanines in the absence of cation (Fig. 5A and B, lane 0). Strong DMS protection is seen for all guanines in the (TGG)$_{20}$ tract in the presence of a combination of K$^+$ and Mg$^{2+}$ with a lesser amount of protection being seen in the presence of Na$^+$ and Mg$^{2+}$, or Mg$^{2+}$ alone. Strong DMS protection is seen for AGG-repeats in the presence of Mg$^{2+}$ alone, or in the presence of Mg$^{2+}$ and either Na$^+$ or K$^+$ (Fig. 5). Since in general, the efficacy of these cations at stabilizing tetraplexes is K$^+ >$ Na$^+ >$ Mg$^{2+}$ (21,30), these data suggest that the stability of these tetraplexes decreases in the order (AGG)$_{20}$ > (TGG)$_{20}$ > (CGG)$_{20}$. Unlike CGG-repeats, the DMS protection of these repeats is independent of pH (data not shown). When very short incubation times are used (∼30 s) before DMS modification of the (AGG)$_{20}$ oligodeoxyribonucleotide, protection is seen only when K$^+$ is present [data not shown]. This supports the idea that the nature of the cation affects not only the stability of the tetrplex but also its rate of formation. No change in the DMS protection pattern of CGG-repeats is seen with shorter incubation times (12).

In the presence of K$^+$, i.e., when the TGG-tract forms a tetraplex, the T residues are as reactive with KMnO$_4$ as they are in the absence of any cation (Fig. 5A). The accessibility of these residues to modification under all conditions suggests that they are unpaired even in the tetraplex. Similarly, both in the presence and absence of cations, the A residues in the (AGG)$_{20}$-repeat are equally reactive with DEPC, showing a level of modification comparable to A residues outside the repeat (Fig. 5B). This suggests that these A residues are also unpaired in the tetraplex.

These patterns of chemical reactivity were seen at oligodeoxyribonucleotide concentrations similar to the template concentrations used in the DNA synthesis arrest assay. They are independent of over the same range of temperatures (37–85°C; data not shown), and are seen in buffers identical to those used for the DNA synthesis arrest assay. Moreover, at short incubation times (30 s), the cation specificities of both the blocks to DNA synthesis and the chemical modification pattern diagnostic of tetraplex formation were the same. Since DNA synthesis arrest was seen even on single-stranded templates and required only a single round of heating and cooling before addition of the polymerase, i.e., conditions analogous to those used in the chemical modification assay, it is reasonable to assume that the tetraplexes detected by these two methods are the same. The ability of c$^+$/A to abolish the K$^+$-dependent blocks to DNA synthesis by the AGG-repeat is not inconsistent with the hypothesis that the adenines in the AGG-tetraplex responsible for these blocks are unpaired, since c$^+$/A not only lacks an nitrogen at the 7 position but also has an altered π-electron shell. This affects not only the ability of the analog to act as an N$^7$ donor, but also negatively impacts helix stabilizing interactions such as base stacking (22). For example, incorporation of c$^+$/A into Watson–Crick double-stranded DNA lowers the $T_m$ of the duplex in spite of the fact that the N$^7$ of adenine does not participate in a hydrogen bond in a Watson–Crick A•T base pair (22).

**Figure 4.** Chemical modification of oligodeoxyribonucleotides containing (CGG)$_{20}$. Oligodeoxyribonucleotides containing the (CGG)$_{20}$ sequence were treated with DMS at pH 9.0 and pH 6.0 in the presence of 2 mM Mg$^{2+}$ and with (K) or without (O) 50 mM K$^+$ as described in the Materials and Methods. The bracket demarcates the CGG-repeat. The C residues in the repeat are indicated. The numbering refers to the position of the C within the CGG-tract with C$_1$ being the 5′-most base in the tract and C$_{38}$ the 3′-most C residue.

 Protection from methylation by DMS. In contrast, in the presence of K$^+$ at neutral pH or below, all the G residues show strong protection from DMS modification consistent with the formation of intrastrand tetraplexes in which all guanines are obligatory N$^7$ donors (Fig. 4, second lane). The fact that DMS protection of the guanines in the CGG-tract increases as the pH is lowered suggests that the underlying DNA structure responsible for this protection is stabilized by protonation. Given that the pH$_a$ of cytosines is much closer to neutral than that of guanines (29), the protonated bases are in all probability cytosines. Moreover, since cytosines are protonated at N$^3$ (29), a residue which serves as a proton acceptor in a Watson–Crick G•C pair, the pH sensitivity suggests that the tetraplex does not involve hydrogen bonds between guanines and cytosines. Protonated cytosines may stabilize the tetraplex by virtue of the additional hydrogen bonds provided by the formation of C•C$^+$ pairs, via increased stacking energy contributions, or both.

Oligodeoxyribonucleotides containing the sequence (NGG)$_{20}$ where N = A, C, T, U or an abasic site show similar mobilities on acrylamide gels containing K$^+$

This data suggests that interactions between guanines provide the major stabilizing force in tetraplexes. To test this hypothesis directly, the tetraplex-forming ability of an oligodeoxyribonucleotide
Figure 5. Chemical modification of oligodeoxyribonucleotides containing (TGG)$_{20}$ or (AGG)$_{20}$. Oligodeoxyribonucleotides containing either (TGG)$_{20}$ or (AGG)$_{20}$ were treated with DMS, and either KMnO$_4$ or DEPC in the absence of cation (0), in the presence of 2 mM Mg$^{2+}$ (M) or 2 mM Mg$^{2+}$ and either 50 mM K$^+$ (K) or 50 mM Na$^+$ (N) as described in the Materials and Methods. The bracket demarcates the NGG-repeat. The T residues and A residues in each repeat tract are indicated with N$_1$ being the 5$'$-most nucleotide in the array, and N$_{58}$ being the non-guanine base in the last (3$'$-most) triplet.

(*GG)$_{20}$ which contained abasic residues in place of the non-G bases was examined. Since abasic oligodeoxyribonucleotides are sensitive to piperidine treatment, and cause termination of DNA synthesis, it is not possible to measure tetraplex formation by either the extent of DMS protection or by the amount of K$^+$-dependent DNA synthesis arrest. However since oligodeoxyribonucleotides that form intramolecular structures show a higher mobility in polyacrylamide gels than would be expected from their molecular weights and oligonucleotides capable of tetraplex formation have the highest mobility in the presence of K$^+$ (31), it was possible to use electrophoretic mobility shift analysis to examine the tetraplex-forming potential of all the NGG-containing oligodeoxyribonucleotides including (*GG)$_{20}$ and the oligonucleotide (UGG)$_{20}$ from which it was derived.

Electrophoresis was carried out on 10% polyacrylamide gels containing Tris-borate (pH 8.3) alone, or together with either 7 M urea or 100 mM KCl. Two oligodeoxyribonucleotides (M1, a 60mer, and M2, a 69mer) which do not form intramolecular folded structures under these conditions were used as markers.

Figure 6. Electrophoretic mobility of (NGG)$_{20}$-containing oligodeoxyribonucleotides. Oligodeoxyribonucleotides containing either (CGG)$_{20}$, (AGG)$_{20}$, (TGG)$_{20}$, (UGG)$_{20}$ and a derivative of (UGG)$_{20}$ known as (*GG)$_{20}$ which lacks the uracil bases containing 1x TBE either alone, or with either 7 M urea or 100 mM KCl. Two oligodeoxyribonucleotides (M1, a 60mer, and M2, a 69mer) which do not form intramolecular folded structures under these conditions were used as markers.
The sequences (CGG)\textsubscript{20}, (TGG)\textsubscript{20} and (AGG)\textsubscript{20} form a number of different interstrand and intrastrand structures. CGG-repeats form stable hairpins, analogous to those formed by a subset of the disease associated repeats (6,8). Oligodeoxynucleotides containing the other NGG-triplet repeats migrate slightly faster than expected on non-denaturing polyacrylamide gels in the absence of K\textsuperscript{+}, and a small amount of P1 nucleosome protection is observed in the repeat. However, since the CGG-hairpin shows a much larger electrophoretic mobility shift, and much stronger P1 nucleosome protection of stem bases, it would suggest that if hairpins are formed by these other repeats they are much weaker. AGG-repeats also form triplexes of the purine*purine:pyrimidine type by interaction with the nascent strand during DNA synthesis. By analogy with other triplex-forming sequences, AGG-repeats would also be expected to also form both purine*purine:pyrimidine triplexes and pyrimidine*purine:pyrimidine triplexes involving the purine-rich strand and its complement, analogous to the structures reported for GAA-repeats which undergo expansion in Friedreich’s ataxia (14). In addition to these structures, all NGG-repeats form tetrplexes. A number of lines of evidence suggest that these tetrplexes are all similar. All three repeats show the same pattern of K\textsuperscript{+}-dependent DNA synthesis arrest, and all of the guanines in the repeat are protected from modification by DMS indicating that the N\textsuperscript{2} atom of each G residue is involved in hydrogen bonding. The fact that tetrplex formation by the CGG-repeat is favored by low pH is consistent with a tetrplex stabilized by hydrogen bonds involving protonated bases, presumably cytosines. Since neither G•C base pairs or G•C•G•C tetrads are favored by cytosine protonation, these data support the hypothesis that the CGG-tetrplex is stabilized by C•C•C•C pairs. Similarly, the fact that neither the adenines in the AGG-tetrplex nor the thymines in the TGG-tetrplex are hydrogen bonded, supports structures in which all the guanines are involved in G4-tetrads. Moreover, G4-tetrads are the major stabilizing forces in these tetrplexes since removal of the non-G base does not eliminate the ability of these sequences to form a K\textsuperscript{+}-specific folded structure, and tetrplex structures that maximize the number of G4-tetrads seem to be the most stable. This lends support to the model for CGG-tetrplexes in which G4-tetrads and the sequence (T\textsubscript{3}G\textsubscript{7})\textsubscript{4}. The panel labeled (T\textsubscript{3}G\textsubscript{7})\textsubscript{4} shows the pattern of DNA synthesis arrest on a template containing the sequence (T\textsubscript{3}G\textsubscript{7})\textsubscript{4}. The panel labeled C\textsubscript{4} shows the data for the template in which four guanines in the sequence (T\textsubscript{3}G\textsubscript{7})\textsubscript{4} have been replaced with cytosines in such a way as to disturb only one G4-tetrad. The templates labeled 2xGC\textsubscript{4}GC\textsubscript{4} have four guanines replaced in such a way as to disrupt two G4-tetrads and has the potential to form two G•C•G•C tetrads. These templates were used in the DNA synthesis arrest assay in the presence (K\textsuperscript{+}) or absence (0) of 50 mM K\textsuperscript{+}. The lane markers T, C, G and A indicate the bases on the template strand. The brackets delineate the tetrplex-forming sequence.

**DISCUSSION**

The sequences (CGG)\textsubscript{20}, (TGG)\textsubscript{20} and (AGG)\textsubscript{20} form a number of different interstrand and intrastrand structures. CGG-repeats form stable hairpins, analogous to those formed by a subset of the disease associated repeats (6,8). Oligodeoxynucleotides containing the other NGG-triplet repeats migrate slightly faster than expected on non-denaturing polyacrylamide gels in the absence of K\textsuperscript{+}, and a small amount of P1 nucleosome protection is observed in the repeat. However, since the CGG-hairpin shows a much larger electrophoretic mobility shift, and much stronger P1 nucleosome protection of stem bases, it would suggest that if hairpins are formed by these other repeats they are much weaker. AGG-repeats also form triplexes of the purine*purine:pyrimidine type by interaction with the nascent strand during DNA synthesis. By analogy with other triplex-forming sequences, AGG-repeats would also be expected to also form both purine*purine:pyrimidine triplexes and pyrimidine*purine:pyrimidine triplexes involving the purine-rich strand and its complement, analogous to the structures reported for GAA-repeats which undergo expansion in Friedreich’s ataxia (14).

In addition to these structures, all NGG-repeats form tetrplexes. A number of lines of evidence suggest that these tetrplexes are all similar. All three repeats show the same pattern of K\textsuperscript{+}-dependent DNA synthesis arrest, and all of the guanines in the repeat are protected from modification by DMS indicating that the N\textsuperscript{2} atom of each G residue is involved in hydrogen bonding. The fact that tetrplex formation by the CGG-repeat is favored by low pH is consistent with a tetrplex stabilized by hydrogen bonds involving protonated bases, presumably cytosines. Since neither G•C base pairs or G•C•G•C tetrads are favored by cytosine protonation, these data support the hypothesis that the CGG-tetrplex is stabilized by C•C•C•C pairs. Similarly, the fact that neither the adenines in the AGG-tetrplex nor the thymines in the TGG-tetrplex are hydrogen bonded, supports structures in which all the guanines are involved in G4-tetrads. Moreover, G4-tetrads are the major stabilizing forces in these tetrplexes since removal of the non-G base does not eliminate the ability of these sequences to form a K\textsuperscript{+}-specific folded structure, and tetrplex structures that maximize the number of G4-tetrads seem to be the most stable. This lends support to the model for CGG-tetrplexes in which G4-tetrads and the sequence (T\textsubscript{3}G\textsubscript{7})\textsubscript{4}. The panel labeled (T\textsubscript{3}G\textsubscript{7})\textsubscript{4} shows the pattern of DNA synthesis arrest on a template containing the sequence (T\textsubscript{3}G\textsubscript{7})\textsubscript{4}. The panel labeled C\textsubscript{4} shows the data for the template in which four guanines in the sequence (T\textsubscript{3}G\textsubscript{7})\textsubscript{4} have been replaced with cytosines in such a way as to disturb only one G4-tetrad. The templates labeled 2xGC\textsubscript{4}GC\textsubscript{4} have four guanines replaced in such a way as to disrupt two G4-tetrads and has the potential to form two G•C•G•C tetrads. These templates were used in the DNA synthesis arrest assay in the presence (K\textsuperscript{+}) or absence (0) of 50 mM K\textsuperscript{+}. The lane markers T, C, G and A indicate the bases on the template strand. The brackets delineate the tetrplex-forming sequence.

**Figure 7.** The stability of tetrplexes with the potential to form C•C mispairs or G•C•G•C tetrads. Three different templates were used in these experiments. The potential tetrplexes formed by the different templates are shown below the data for each template. The gray parallelograms represent G4-tetrads. Potential G•C•G•C tetrads are shown as open parallelograms. The panel labeled (T\textsubscript{3}G\textsubscript{7})\textsubscript{4} shows the pattern of DNA synthesis arrest on a template containing the sequence (T\textsubscript{3}G\textsubscript{7})\textsubscript{4}. The panel labeled C\textsubscript{4} shows the data for the template in which four guanines in the sequence (T\textsubscript{3}G\textsubscript{7})\textsubscript{4} have been replaced with cytosines in such a way as to disturb one G4-tetrad. The template labeled 2xGC\textsubscript{4}GC\textsubscript{4} has four guanines replaced in such a way as to disrupt two G4-tetrads and has the potential to form two G•C•G•C tetrads. These templates were used in the DNA synthesis arrest assay in the presence (K\textsuperscript{+}) or absence (0) of 50 mM K\textsuperscript{+}. The lane markers T, C, G and A indicate the bases on the template strand. The brackets delineate the tetrplex-forming sequence.
C+C* pairs exist in a 2:1 ratio as shown in Figure 1. It also suggests that a similar structure is formed by AGG-tetraplexes and TGG-tetraplexes where pairs of G4-tetrads alternate with unpaired adenines or thymines. Since these tetraplexes form very rapidly under physiologically reasonable conditions, it is possible that they might be able to form in cells anytime the DNA is unpaired for even a short period of time. The stability of each tetraplex is likely to have an effect on the probability of its formation and its longevity in vivo. AGG-repeats show somewhat less K+-dependent DNA polymerase pausing than the other NGG-repeats under the conditions of the assay used here [Fig. 3. and (12,18)]. This could be due either to the relative instability of AGG-tetraplexes or to the competition between tetraplexes that are able to block the polymerase, and other DNA conformations that can be adopted by this sequence. However, since the efficacy of cations at stabilizing tetraplexes is K+ > Na+ > Mg2+ (21,30), the fact that AGG-tetraplex formation is efficient in the presence of Mg2+ alone, while a lesser amount tetraplex in the presence of Mg2+ alone reflects a decreased

REFERENCES