The mouse Ms6-hm microsatellite consists of a tandem array of the pentamer \(d(CAGGG)\). This microsatellite is extremely hypervariable, showing a germ line mutation rate of 2.5%/gamete. The mechanism responsible for this instability is not known. The ability to form intranand structures is a conserved feature of many hypervariable sequences, and it has been suggested that the formation of such structures might account for instability by affecting DNA replication, repair, or recombination. Here we show that this microsatellite is able to form intranand structures as well. Under physiological conditions, the Ms6-hm microsatellite forms a hairpin as well as two different unusual intranand tetraplexes. The hairpin forms in the absence of monovalent cation and contains G-A, G-C, and G-G base pairs in a 1:1:1 ratio. In the presence of K\(^+\), a tetraplex is formed in which the adenines are unpaired and extrahelical, and the cytosines are involved in C-C\(^+\) pairs. In Na\(^+\), a tetraplex forms that contains C-C\(^+\) pairs, with the adenines being intrahelical and hydrogen-bonded to guanines. Tetraplex formation in the presence of Na\(^+\) requires both cytosines and adenines and might reflect the altered internal dimensions of this tetraplex, perhaps resulting from the ability of the C-C\(^+\) pairs to become intercalated in this sequence context. Our demonstration of the stabilization of tetraplexes by hydrogen bonding between adenines and guanines expands the hydrogen-bonding possibilities for tetraplexes and suggests that the category of sequences with tetraplex-forming potential may be larger than previously appreciated.

Long arrays of tandemly repeated sequences are common in eukaryotes and, in humans, make up \(~10\%\) of the genome (1). These sequences show varying degrees of polymorphism, with some arrays being stable, whereas others show frequent intergenerational changes that involve the gain or loss of a large number of repeat units. The mouse microsatellite locus Ms6-hm on chromosome 4 is an example of a highly unstable tandem array (2). This microsatellite, which contains repeats of the pentamer \(d(CAGGG)\), shows a high rate of both germ line and somatic mutations (3), with germ line mutation rates of 2.5%/gamete and the number of repeats ranging from 200 to >1000, making it one of the most unstable mouse loci thus far described (3). The mechanism responsible for this instability is not known, but is of wider interest since, in humans, unstable tandem arrays are responsible for a group of genetic diseases known as the triplet expansion disorders (4–6).

We have previously shown that like a number of other hypervariable tandem arrays, the Ms6-hm microsatellite forms a set of strong blocks to DNA synthesis in vitro in the presence of K\(^+\) (7). It has been suggested that instability might result from repeated strand slippage in the repeat tract during replication (8), and as such, these blocks might play a role in promoting this process. These blocks do not form in the presence of cations such as Rb\(^+\), Cs\(^+\), Li\(^+\), or NH\(_4\)^+. They occur at the 3\(^\prime\)-end of the G-rich strand and are template concentration-independent and strand-specific. They are abolished when guanines in the template are replaced with 7-deazaguanine, and the guanines in this sequence are completely protected from DMS\(^1\) modification in the presence of K\(^+\) (7). Thus, the properties of the Ms6-hm microsatellite are consistent with intranand tetraplex formation.

However, unlike other tetraplex-forming sequences, which form a block to DNA synthesis only in the presence of K\(^+\) (7), the Ms6-hm microsatellite blocks DNA synthesis also in the presence of Na\(^+\). This suggested to us that the underlying structure responsible for this block might differ in some fundamental way from previously described tetraplexes. We show here that, in fact, the Ms6-hm microsatellite forms two different tetraplexes, one in the presence of K\(^+\) and a different one that is seen only in the presence of Na\(^+\). In addition, this microsatellite also forms an unusual hairpin in the absence of cations. Our observations expand the category of sequences we would expect to form tetraplexes and might be relevant in particular to a series of previously identified sequences with tetraplex-forming potential that contain adenines. These sequences include the telomeresc of a number of eukaryotes (15), the hypervariable minisatellite in the human insulin promoter (16), and the immunoglobulin switch regions (17).

MATERIALS AND METHODS

Template Preparation—Oligodeoxyribonucleotides containing an EcoRI site at the 5\(^\prime\)-end with the sequence of interest followed by 15 bases of the 5\(^\prime\)-end of the Escherichia coli supF gene were synthesized using standard phosphoramidite chemistry. These oligodeoxyribonucleotides were then used together with a primer that contained an EcoRI site and 15 bases complementary to the 3\(^\prime\)-end of the supF gene to generate a PCR fragment as described previously (9, 14). For PCR fragments containing inosine or 7-deazadenine, diTP or 7-deaza-dATP was substituted for the appropriate dNTP in the PCR. For the construction of plasmid templates, the PCR product was digested with EcoRI, purified by gel electrophoresis, and then cloned into the plasmid pMS189A as described previously (9, 14). Plasmids were propagated in E. coli MBM7070, isolated by alkaline lysis, and purified by CsCl density gradient centrifugation.

The abbreviations used are: DMS, dimethyl sulfate; PCR, polymerase chain reaction; DEPC, diethyl pyrocarbonate; CAA, chloroacetaldehyde;

*The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.
†These authors contributed equally to this work.
‡Present address: Div. of Bone and Mineral Diseases, Jewish Hospital of St. Louis, St. Louis, MO 63110.
¶To whom correspondence should be addressed: NIH, Bldg. 8, Rm. 202, 8 Center Dr. MSC 0830, Bethesda, MD 20892-0830. Tel.: 301-496-2189; Fax: 301-402-0053; E-mail: ku@helix.nih.gov.

\(9\) The abbreviation used are: DMS, dimethyl sulfate; PCR, polymerase chain reaction; DEPC, diethyl pyrocarbonate; CAA, chloroacetaldehyde.
Intrastrand Structures in a Hypervariable Microsatellite

gradient centrifugation according to standard procedures. PCR fragments that were used directly as templates in tetraplex assays were purified on 5% polyacrylamide gels.

Tetraplex Assay—Primer was labeled at the 5'-end with [γ-32P]ATP (3000–6000 Ci/mmol; NEN Life Science Products) using T4 polynucleotide kinase (Epistem Technologies Corp., Madison, WI) and a buffer containing 50 mM Tris-HCl (pH 8.0) and 10 mM MgCl2. Reaction mixtures (in a total volume of 6 μl) contained 0.2–2 nM template, 0.16 nM primer, 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), or ddTTP (0.6 mM)), 50 mM Tris-HCl (pH 9.3), 2.5 mM MgCl2, 5 units of Taq polymerase (Life Technologies, Inc.), and, where indicated, 50 mM monovalent cation. Reaction mixtures were subjected to PCR using 30 cycles of 30 s at 95 °C, 30 s at 55 °C, and 30 s at 72 °C. The reaction was terminated by the addition of 0.5 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% bromphenol blue, and the reactions were heated at 90 °C for 5 min prior to electrophoresis on a 6.5% polyacrylamide sequencing gel.

Chemical Modification of the (CAGGG)_n-containing Oligodeoxyribonucleotide—DMS was purchased from Sigma, and diethyl pyrocarbonate (DEPC) and chloroacetaldehyde (CAA) were purchased from Fluka (Buchs, Switzerland). These reagents were used for chemical modification of the oligodeoxyribonucleotide without further purification. End-labeled deoxynucleotides (1–5 ng) were heated in Tris/EDTA buffer for 1 min at 90 °C and then diluted into 40 μl of the reaction buffer. Reactions carried out at pH 5.5 used a buffer consisting of 50 mM concentration of the acetate salt of the indicated cation (pH 5.5) and 2 mM MgCl2. Reactions at other pH values used a buffer consisting of 50 mM Tris adjusted to the appropriate pH with HCl and either no added cations or 2 mM MgCl2 or 2 mM MgCl2 and a 50 mM concentration of the specified monovalent cation as either the glutamate or the chloride salt (the anion has no effect on tetraplex formation (14)). Reactions were heated for 30 s at 95 °C and then incubated for 5–60 min at 55 °C. One microliter of DMS, 4 μl of DEPC, or 4 μl of CAA was added to each tube. These reactions were allowed to proceed at room temperature for 1 min (DMS) or 20 min (DEPC and CAA). Reactions were terminated by repeated precipitation with butanol. CAA-treated templates were treated with DMS, precipitated with butanol, and then heated for 30 min at 90 °C in 1 mM piperidine. DMS-treated samples were treated either directly with piperidine or with sodium acetate (pH 5.2) at 90 °C for 4 min or 60% hydrazine for 30 s at room temperature before piperidine cleavage. DEPC-treated samples were treated with piperidine without further modification. These reactions were precipitated twice with 1 ml of 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, and 0.05% bromphenol blue; denatured for 2 min at 90 °C; and subjected to electrophoresis on a 20% sequencing gel at 3500 V until the bromophenol blue was 5 cm from the bottom of the gel. Gels were covered with plastic wrap and exposed to x-ray film overnight at −70 °C. Gels were scanned and analyzed by NIH Image (Version 1.58).

RESULTS

Intrastrand tetraplexes form when four G-rich regions on a single strand interact to form a series of tetrads. Successive tetrads form a hollow stem that is bounded by loops formed by the bases between the G-rich regions (Fig. 1, L1, L2, and L3). K+ ions are generally the most effective monovalent cations at stabilizing these structures. It is thought that this is due either to an optimal fit of K+ inside the stem of the tetraplexes such that the cation is able to coordinate the 0° atoms of guanines in adjacent tetrads (18) or to its relative ease of dehydration (19). We have recently demonstrated that the ability to block DNA synthesis in the presence of K+ is a diagnostic feature of intrastrand tetraplexes (11), presumably because the K+-stabilized tetraplexes form barriers to DNA polymerase progression. We have now found ions other than K+ to be effective at producing a block to DNA synthesis in any of the other tetraplex-forming sequences that we have tested (7, 9–13), so our observation of the ability of Na+ to cause a block to DNA synthesis on templates containing the mouse Ms6-hm repeats (7) is somewhat surprising. However, since the properties of the arrest sites are all consistent with tetraplex formation, our working hypothesis was that the structural basis of the blocks to DNA synthesis seen in this region was the formation of one or more intrastrand tetraplexes with unusual properties.

Since the major sites of polymerase arrest occurred at the 3'-end of the repeat array (summarized graphically in Fig. 2; also see lanes marked WT in Fig. 8) (7), we conclude that the major structure formed in the presence of both cations involves all eight repeats. A lesser amount of polymerase pausing is seen immediately 3' of the fifth, sixth, and seventh repeats from the 5'-end in the presence of K+, but not Na+ (7). These products represent DNA synthesis arrest occurring at tetraplexes involving fewer repeats. The fact that a template containing only five CAGGG repeats produces only one weak K+-dependent chain termination product and no Na+-dependent ones (data not shown) is consistent with this idea. In addition to the larger number of polymerase stops, the amount of K+-dependent polymerase pausing at the 3'-end of the repeat tract is greater than the amount of Na+-dependent pausing (see lanes marked WT in Fig. 8) (7). Two explanations are possible: either CAGGG repeats form a single type of tetraplex with a higher stability in K+; or two types of tetraplex form, one in the presence of K+ and a different, somewhat less stable tetraplex in the presence of Na+.

We used DMS, CAA, and DEPC to analyze the structures formed by eight d(CAGGG) repeats as well as a variety of base-substituted variants of this sequence. DMS methylates the N7 of guanine, which is not involved in Watson-Crick base pairing (20). Protection of guanines from this reagent is thus diagnostic of the formation of non-Watson-Crick base interactions. DMS also modifies adenosines and cytosines, albeit to a lesser extent, with the rate of these reactions varying with the particular configuration of the base. Chloroacetaldehyde reacts with the N3 and N6 of adenine and the N3 and N4 of cytosine (21), positions involved in hydrogen bonding in Watson-Crick A:T and G:C pairs, respectively. The first step in the reaction of CAA with DNA is believed to be the alkylation of the endocyclic nitrogen (21), and N6-substituted adenosines and N4-substituted cytidines are reactive with CAA (22). It is thus feasible that cytosines that are hydrogen-bonded at N4 but not N3 and that adenosines in which the N6 but not the N1 proton is involved in a hydrogen bond will also be reactive with CAA. DEPC decarboxylates adenosines and, to a lesser extent, guanines (23). This modification occurs at adenosine N6 and N7 and guanine N7 and is seen most readily when these bases are unpaired or in an
The third guanine in each repeat shows partial protection of the third guanine in the first repeat and of DMS protection (data not shown). Substitution of A7, A17, A27, or the C at position 41 with a T greatly reduces the amount of guanine reactivity of the second guanine residue, whereas the first repeat is involved in a G-A pair, and the third guanine is protected, albeit to a lesser extent. Replacement of the C at position 41 with a T greatly reduces the amount of guanine protection for certain guanines both in the absence of monovalent cation (or in the presence of Li+ to a lesser extent. Replacement of the C at position 41 with a T greatly reduces the amount of guanine protection for certain guanines both in the absence of monovalent cation (or in the presence of Li+) and in the presence of K+ (Fig. 3). Since unpaired guanines or guanines in Watson-Crick G-C pairs are not protected from DMS modification, this indicates that secondary structures involving non-Watson-Crick base interactions are formed under all of these conditions. The patterns of DMS protection are different in Li+, K+, and Na+ (Fig. 3, lanes marked L, K, and N), indicating that at least three different secondary structures are formed by the mouse Ms6-hm repeat.

The Monovalent Cation-independent Structure—In the absence of monovalent cation or in the presence of Li+, there is partial protection of the third guanine in the first repeat and of the first and third guanines in the remaining repeats (Fig. 3). The third guanine in each repeat shows ~50% of the DMS reactivity of the second guanine residue, whereas the first guanine is also protected, albeit to a lesser extent. Replacement of the C at position 41 with a T greatly reduces the amount of DMS protection (data not shown). Substitution of A7, A17, A27, and A37 with N6-methyladenine, an adenine analog that contains a methyl group at N6, or nebularine, a purine base that lacks the N6 amino group, abolishes the partial DMS protection of the guanines (Fig. 4), indicating that hydrogen bonding of some or all of these adenines is involved in the structure responsible for this protection. These data are consistent with the formation of a hairpin in which the first guanine in each repeat is involved in a G-A pair, the second guanine is involved in a Watson-Crick G-C base pair, and the third guanine is hydrogen-bonded to the third guanine in a repeat on the opposite side of the hairpin (Fig. 5). In a G-G pair, each of these guanines acts as an N7 donor 50% of the time, thus accounting for the ~50% reduction in DMS reactivity of these bases. A base pairing scheme in which guanines in the G-A pairs are obligatory N7 donors should afford their complete protection from DMS modification. Therefore, the partial protection of the first guanine in each repeat could result from either an oscillation between a G-A pair that involves guanine N7 and one that does not or a mixture of the two types of base pairs. A G-syn-AH "anti" base pair is the only G-A pair involving the normal tautomeric forms of these bases in which guanine N7 participates in a hydrogen bond; however, this scheme requires that the adenine be protonated, and since we observed a similar pattern of DMS protection over the pH range 5.5–10, this finding is somewhat surprising. Possibilities for G-A pairs that do not involve guanine N7 include a G-anti-A-syn pair, a G-anti-A-anti pair, or a sheared G-A pair.

Fig. 2. Pattern of DNA synthesis arrest on the (CAGGG)8 template. The repeat is shown in boldface, and the bases within the repeat are numbered starting from the 5’-end of the top strand. Each repeat unit is indicated by a bracket and numbered 1–8 starting from the 5’-end. The black arrows indicate the K+-dependent blocks to DNA synthesis, and the gray arrows indicate the Na+-dependent ones. The length of the arrows indicates the relative amounts of premature chain termination in each case. This figure summarizes our previously published data (7).

Fig. 3. DMS modification of an oligodeoxyribonucleotide containing the (CAGGG)8 tract. An oligodeoxyribonucleotide containing the Ms6-hm repeat was modified with DMS as described under “Materials and Methods.” The labeled L, K, and N indicate the products for reactions carried out in the presence of Li+, K+, and Na+ respectively. Each repeat unit is marked by a bracket and numbered 1–8 starting from the 5’-end. The sequence of a representative repeat unit is shown alongside unit 2. The DMS-reactive bases are shown in boldface with the identity of the base and its position in the repeat tract indicated alongside. The black arrow indicates the DMS-hyperreactive guanine that is seen in the presence of K+ and is diagnostic of the formation of the K+-tetraplex.
DMS-reactive, there was a single DMS-hyperreactive guanine in the 3'-flanking region (Fig. 3, arrow) that was not seen in the presence of either Li⁺ or Na⁺. Hyperreactivity of this base would be consistent with its N⁷ atom being in a particularly exposed conformation, as it might be at the junction of the tetraplex and a region of single strandedness, and appears to be a common feature of K⁺-dependent tetraplexes.² This pattern of DMS modification was independent of pH over the range 5.5–10.

The same pattern of DMS protection, including the hyperreactive guanine in the flanking region, was seen when A⁷, A¹⁷, A²⁷, and A³⁷ of the Ms6-hm repeat were replaced with N⁶-methyladenine (m⁶dA) or nebularine were modified with DMS as described under “Materials and Methods.” The lanes marked L, K, and N indicate the products for reactions carried out in the presence of Li⁺, K⁺, and Na⁺, respectively. Each repeat unit is indicated by a bracket and numbered 1–8 starting from the 5'-end. The sequence of a representative repeat unit is shown alongside unit 2. The DMS-reactive bases in the unsubstituted oligonucleotide in the presence of Na⁺ are shown in boldface with the identity of the base and its position in the repeat tract indicated alongside. The positions of the N⁶-methyladenine and nebularine residues are indicated with asterisks and filled circles, respectively. The black arrows indicate the DMS-hyperreactive guanine that is seen in the presence of K⁺ and is diagnostic of the formation of the K⁺ tetraplex.

² K. J. Woodford, M. N. Weitzmann, and K. Usdin, manuscript in preparation.
isosteric analog of adenine, 7-deazaadenine, has little effect on either the K**1**-dependent arrest of DNA synthesis or the pattern of K**1**-dependent DMS protection (data not shown). However, replacement of all adenines with 7-deazaadenine eliminates the K**1**-dependent arrests. This observation is not necessarily inconsistent with the interpretation that the adenines are unpaired since the presence of a carbon rather than a nitrogen at position 7 of 7-deazaadenine not only abolishes the ability of this analog to act as an N7 donor, but also produces an altered π-electron shell that can affect other interactions that may stabilize the tetraplex in other ways (25).

The adenines in the repeat were uniformly reactive with CAA in the presence of K**1** at both high (Fig. 7), and low pH values. Given the DEPC hyperreactivity of the adenines, the CAA data are consistent with these bases being unpaired. At high pH, the cytosines are CAA-reactive (Fig. 7), but at low pH, they are not (data not shown). This suggests that the cytosines form pH-sensitive hydrogen bonds at low pH. Since the adenines are unpaired at all pH values and since G-C pairs or G-C-G-C tetrads are not favored by protonation, this suggests that the K**1** tetraplex contains C-C**+** pairs at low pH. The reason for the CAA reactivity of the cytosines at high pH is less clear-cut. One interpretation is that the cytosines are unpaired under these conditions. However, replacement of C**6** and C**36** with thymines reduces the extent of polymerase arrest in the presence of K**1** (Fig. 8, center panel). This would be consistent with a role of hydrogen bonding involving these cytosines even at higher pH values. Moreover, replacement of C**6**, G**16**, C**26**, and C**36** with thymines enables the tract to block DNA polymerase as effectively as the unsubstituted repeat (Fig. 8, right panel). These results suggest that hydrogen bonds form between C**6** and G**16** and between C**26** and C**36**. The C**6**-to-T**6** and C**36**-to-T**36** substitutions would disrupt both C-C pairs, without providing significant compensating hydrogen bonding, whereas the substitution of all four cytosines in the stem would make the formation of hydrogen bonds between thymines possible and thus restore stability to the structure. In neutral C-C pairs, the N**4** amino group of one cytosine is involved in a hydrogen bond with the N**3** of the other cytosine, but since this base pair only involves a single hydrogen bond, it is likely that each cytosine acts as an N**3** acceptor 50% of the time, with the lifetime of each conformation being shorter than the reaction time of the structure with CAA (26). The CAA reactivity of cytosines at pH values above neutral could thus be reconciled with the effect of
Intrastrand Structures in a Hypervariable Microsatellite

Fig. 8. Effect of different C-to-T substitutions on the stability of the two tetraplexes. Plasmid clones that contained a C-to-T substitution at C6 and C26, or at C6, C16, C26, and C36 were constructed as described under “Materials and Methods.” These templates were used together with the unsubstituted template in the tetraplex assay. The assay was conducted in the absence of added monovalent cation (0), in the presence of 50 mM KCl, or in the presence of 50 mM Na+ (NaCl). The lane markers T, C, G, and A indicate the bases on the template strand. The bracket on the left demarcates the repeat array. WT, wild type.

thymine substitutions if it is assumed that the cytosines are involved in neutral C-C pairs in which each cytosine oscillates between a conformation in which it acts as an N3 donor and one in which it acts an N3 acceptor. The cytosines would be vulnerable to CAA modification either when they are acting as N3 acceptors or during the transition from N3 acceptor to N3 donor. All the available data therefore lend support to a model for the K+-tetraplex shown in Fig. 9, in which all the guanines are involved in G4 tetrads; the adenines are extrahelical; and C6, C16, C26, and C36 are involved in neutral C-C pairs at non-acidic pH values and in hemiprotonated C-C+ pairs at low pH.

The Na+-dependent Structure—In the presence of Na+, there is strong protection of most of the guanines from DMS modification that is diagnostic of tetraplexes (Fig. 3). However, in contrast to what is seen in K+, two guanines (G20 and G26) and a cytosine (C21) were reactive with DMS at neutral pH and above, suggesting that these bases are located in the loops of the Na+-tetraplex. Consistent with this observation is the fact that substitution of G20 with thymine has no effect on the amount of Na+-dependent DNA polymerase arrest (data not shown). The location of these bases in the loops rules out a tetraplex with a conformation similar to that of the K+-tetraplex. Similarly, it rules out a tetraplex in which the adenines are aligned with one another so as to contain a mixture of G4 tetrads and G-C pairs or G-C-G-C tetrads. Rather, these data are consistent with a structure with the same overall conformation as the one shown in Fig. 9 (right). The reactivity of G16, G20, and C21 was reduced at low pH. This reduction, without changes elsewhere in the tetraplex, suggests that although the overall conformation of the structure is relatively pH-independent, there may be some pH-dependent conformational changes in the loops.

Substitution of A7, A17, A27, and A37 with either N6-methyladenine or nebularine abolished the DMS protection of stem guanines completely (Fig. 4). The effect of N6-methyladenine would be consistent either with a role for the N6 amino group in hydrogen bonding or with an effect of the bulky methyl group on steric hindrance. The nebularine data support the first hypothesis. All the adenines in the repeat were DEPC-reactive, with A32 being hyperreactive (Fig. 6). The same pattern of reactivity of these bases was seen at all pH values. This reactivity indicates that adenine N7 is not involved in hydrogen bonding. Replacement of A7 and A37 with 7-deazaadenine abolishes the Na+-dependent arrests and the Na+-dependent DMS protection pattern (data not shown). However, since all the adenines are reactive with DEPC, this is probably due to the negative effect of the altered π-electron shell of the 7-deazaadenine analog on stacking rather than the inability to act as an N7 donor (25, 27). The same two guanines that were DMS-reactive (G10 and G20) were also DEPC-reactive at high pH (Fig. 6), supporting the assignment of these bases to the loops of the tetraplex. The DEPC reactivity of these guanines is reduced at low pH, consistent with pH-sensitive conformational changes occurring in these loops.

In the presence of Na+ at non-acidic pH values, C21, A22, C31, and A32 are CAA-reactive (Fig. 7), consistent with the location of these bases in the loops as shown in Fig. 9. In contrast, A7, C6, A17, C16, A12, C26, A27, and C36 are protected from CAA modification. The protection of adenines is consistent with our N6-methyladenine and nebularine data and supports the contention that the stem adenines are involved in hydrogen bonds. Moreover, it suggests that the hydrogen bonding involves adenine N3 as well as the N6 proton since the first site of CAA attack is thought to be the endocyclic nitrogen. The formation of G-syn-AH-anti pairs could account for both the DMS protection of the guanines and the chemical reactivity of the adenines. Since the DMS protection of most of the guanines remains constant over the pH range 5.5–10, our interpretation, if correct, would suggest that the adenines in this sequence can be protonated even at relatively high pH values. This would be consistent with our model for the hairpin formed by this sequence in the presence of Li+. The formation of A-anti-G-Anti-Anti-G anti-tetrads in which the N7 of each guanine acts as an acceptor for the amino proton of the second adenine in each tetrad would also be consistent with our data. However, this tetrad is thought to be improbable based on stereochemical considerations (28). The protection of stem cytosines from modification by CAA (Fig. 7) demonstrates that cytosines are also
hydrogen-bonded in the Na\(^{+}\) tetraplex probably in C-C pairs.

Replacement of C\(^6\), C\(^{16}\), C\(^{26}\), and C\(^{36}\) with thymines eliminates the Na\(^{+}\)-dependent DNA synthesis arrest site completely (Fig. 8), as does replacement of A\(^7\), A\(^{17}\), A\(^{27}\), and A\(^{37}\) (data not shown). The fact that replacement of stem adenines or cytosines with thymines abolishes the Na\(^{+}\) tetraplex, but not the K\(^{+}\) tetraplex, indicates that both cytosines and adenines are required for the Na\(^{+}\) structure. Model sequences with the potential to form A-G pairs or A-G-A-G tetrads do not show any evidence for tetraplex formation in the presence of Na\(^{+}\), either at neutral or acidic pH values (data not shown). This suggests that hydrogen bonds between guanines and adenines alone do not account for the Na\(^{+}\)-dependent tetraplex. Likewise, model sequences with the potential to form C-C or C-C pairs also do not form tetraplexes in the presence of Na\(^{+}\) either at high or low pH\(^{2}\) suggesting that such pairs are also not sufficient to allow formation of the tetraplex in the presence of Na\(^{+}\). One explanation for the Na\(^{+}\) effect could be that the hydrogen-bonding arrangement between adenines and guanines favors intercalation of the hemiprotonated C-C pairs. The closer proximity of metal ion-binding sites that result from this intercalation (30) may allow the smaller Na\(^{+}\) cation to stabilize this type of tetraplex. If this hypothesis is correct, it would suggest that the Na\(^{+}\) structure is unusual not only because it is stabilized by hydrogen bonds of both a typical G\(_4\) tetrad-based tetraplex and an i-motif structure in which the constituent C-C pairs are intercalated (31).

**DISCUSSION**

We have described a set of folded structures that form on the purine-rich strand of the tandem repeat found at the mouse Ms6-hm locus. These structures include a hairpin containing a mixture of G-G, G-C, and G-A pairs and two tetrplexes, one formed in the presence of K\(^{+}\) containing a mixture of G\(_4\) tetrads, extrahelical adenines, and neutral C-C pairs and one formed in the presence of Na\(^{+}\) containing G\(_4\) tetrads, hemiprotonated C-C pairs, and either G-syn-AH\(^{−}\)-anti pairs or G-anti-A-anti-G-anti-A-anti tetrad.

These tetraplexes are seen at physiologically reasonable ionic strengths and form very rapidly in solution. The Na\(^{+}\)-K\(^{+}\) effect resembles a switch in which the nature of the major cation determines the ultimate configuration of the tetraplex. Whether this "switch" has any biological consequences remains to be determined. This switch is different from the one reported by Sen and Gilbert (32), which involved an oscillation between intermolecular tetraplexes formed in the presence of Na\(^{+}\) and intramolecular tetraplexes that formed preferentially in the presence of K\(^{+}\), but serves to underscore how fluctuations in the cationic environment of DNA might elicit a change in the behavior of a particular region of DNA. The observation that a different hydrogen-bonding interaction is seen on lowering the pH raises the possibility of an additional switch that is pH-sensitive. The observation that adenines can participate in tetrads is relevant not only for the structure formed by the Ms6-hm locus, but also for other regions of eukaryote genomes since the ability of adenines to contribute to tetraplex stability increases the number of sequences with the potential to form stable tetraplexes. In particular, the potential contribution of adenines might be germane to the structures of other tetraplex-forming regions that contain adenines, e.g. the (T\(_6\)A\(_G\))\(_6\) repeats that comprise the telomeres of all vertebrates (15), the Dictyostelium telomere (G\(_6\)AGAGAG\(_G\))\(_6\) (33), the polymorphic repeats in the immunoglobulin switch regions (G\(_4\)AGCTG\(_4\))\(_n\) (17), the hypervariable sequence found in the D4S43 locus (G\(_4\)AG\(_G\)AAG\(_A\))\(_n\) (34), and the hypervariable sequence in the human insulin promoter (ACAGGGGTGTT-GGGG)\(_n\) (16). Since there is a growing body of evidence, both direct (35–37) and indirect (29, 38–41), for the formation of tetraplexes in vivo, the ability of additional sequences to form tetraplexes might have important biological consequences.
REFERENCES

1. Willard, H. F. (1989) Genome 31, 737–744
2. Kelly, R., Bulfed, G., Collick, A., Gibbs, M., and Jeffreys, A. J. (1989) Genom- ics 5, 844–856
3. Kelly, R., Gibbs, M., Collick, A., and Jeffreys, A. J. (1991) Proc. R. Soc. Lond. B Biol. Sci. 245, 235–245
4. Ashley, C. T., and Warren, S. T. (1995) Annu. Rev. Genet. 29, 703–728
5. Bates, G., and Lehrach, H. (1994) Bioessays 16, 277–284
6. Caskey, C. T., Pizzuti, A., Fu, Y. H., Fenwick, R. J., and Nelson, D. L. (1992) Science 256, 784–789
7. Weitzmann, M. N., Woodford, K. J., and Usdin, K. (1997) J. Biol. Chem. 272, 9517–9523
8. Sinden, R. R., and Wells, R. D. (1992) Curr. Opin. Biotechnol. 3, 612–622
9. Usdin, K., and Woodford, K. J. (1995) Proc. Natl. Acad. Sci. U. S. A. 92, 4935–4939
10. Usdin, K. (1998) Nucleic Acids Res. 26, 4078–4085
11. Weitzmann, M. N., Woodford, K. J., and Usdin, K. (1996) J. Biol. Chem. 271, 20958–20964
12. Howell, R., Woodford, K. J., Weitzmann, M. N., and Usdin, K. (1997) J. Biol. Chem. 272, 5208–5214
13. Howell, R., and Usdin, K. (1996) Mol. Biol. Evol. 14, 144–155
14. Woodford, K. J., Howell, R. M., and Usdin, K. (1994) J. Biol. Chem. 269, 27929–27935
15. Moyzis, R. K., Buckingham, J. M., Cram, L. S., Dani, M., Deaven, L. L., Jones, M. D., Mayne, J., Ratliff, R. L., and Wu, J. R. (1988) Proc. Natl. Acad. Sci. U. S. A. 85, 6622–6626
16. Bell, G. I., Karam, J. H., and Rutter, W. J. (1981) Proc. Natl. Acad. Sci. U. S. A. 78, 5759–5763
17. Mills, F. C., Brooker, J. S., and Camerini-Otero, R. D. (1990) Nucleic Acids Res. 18, 7305–7316
18. Pinnavaia, T. J., Marshall, C. L., Mettler, C. M., Fisk, C. L., Miles, H. T., and Becker, E. D. (1978) J. Am. Chem. Soc. 100, 3625–3627
19. Hud, N. V., Smith, P. W., Anet, F. A. L., and Feigon, J. (1996) Biochemistry 35, 15383–15390

Intrastrand Structures in a Hypervariable Microsatellite 30749

20. Maxam, A. M., and Gilbert, W. (1980) Methods Enzymol. 65, 499–560
21. Leonard, N. J. (1984) CRC Crit. Rev. Biochem. 15, 125–199
22. Barnet, J. R., Secrist, Jr., A. B., and Leonard, N. J. (1972) Biochem. Biophys. Res. Commun. 46, 597–604
23. Leonard, N. J., McDonald, J. J., Henderson, R. E., and Reichmann, M. E. (1971) Biochemistry 10, 3335–342
24. Lee, J., Evans, D. H., and Morgan, A. R. (1986) Nucleic Acids Res. 8, 4305–4320
25. Seela, F., Berg, H., and Rosemeyer, H. (1989) Biochemistry 28, 6193–6198
26. Boulard, Y., Cognet, J. A., and Fazakerley, G. V. (1997) J. Mol. Biol. 268, 331–347
27. Seela, F., and Kehne, A. (1987) Biochemistry 26, 2232–2238
28. Murchie, A. I., and Lilley, D. M. (1997) EMBO J 16, 993–1001
29. Weisman-Shomer, P., and Fry, M. (1994) Biochem. Biophys. Res. Commun. 205, 305–311
30. Berger, I., Egli, M., and Rich, A. (1996) Proc. Natl. Acad. Sci. U. S. A. 93, 12116–12121
31. Ahmed, S., Kintanar, A., and Henderson, E. (1994) Nat. Struct. Biol. 1, 83–88
32. Sen, D., and Gilbert, W. (1990) Nature 344, 410–414
33. Hardin, C. C., Henderson, E., Watson, T., and Prosser, J. K. (1991) Biochemistry 30, 4460–4472
34. Horn, G. T., McClatchey, A. I., Richards, B., MacDonald, M. E., and Gusella, J. F. (1991) Nucleic Acids Res. 19, 4772
35. Nagesh, N., and Chatterji, D. (1995) Biochem. Biophys. Methods 30, 1–8
36. Hammond-Kosack, M. C., Kilpatrick, M. W., and Docherty, K. (1992) J. Mol. Endocrinol. 9, 221–225
37. Hammond-Kosack, M. C., Kilpatrick, M. W., and Docherty, K. (1993) J. Mol. Endocrinol. 10, 121–126
38. Frantz, J. D., and Gilbert, W. (1995) J. Biol. Chem. 270, 9413–9419
39. Liu, Z., and Gilbert, W. (1994) Cell 77, 1083–1092
40. Schierer, T., and Henderson, E. (1994) Biochemistry 33, 2240–2246
41. Weisman-Shomer, P., and Fry, M. (1993) J. Biol. Chem. 268, 3306–3312