Proc. Natl. Acad. Sci. USA
Vol. 92, pp. 5199–5203, May 1995
Biophysics

Hairpins are formed by the single DNA strands of the fragile X triplet repeats: Structure and biological implications

(fragile X triplets/structure determination by two-dimensional NMR/hairpins of the C-rich strands/asymmetric expansion/*de novo* methylation)

XIAN CHEN*†, S. V. SANTHANA MARIAPPAN*, PAOLO CATASTI*†, ROBERT RATLIFF‡, ROBERT K. MOYZIS‡, ALI LAAYOUN§, STEVEN S. SMITH§, E. MORTON BRADBURY†¶, AND GOUTAM GUPTA*‖

*Theoretical Biology and Biophysics, T-10, MS-K710, †Life Sciences Division, LS-2, MS 880, ‡Center for Human Genome Studies, Los Alamos National Laboratory, Los Alamos, NM 87545; §Department of Cell and Tumor Biology, City of Hope National Medical Center, 1500 East Duarte Road, Duarte, CA 91010; and ¶Department of Biological Chemistry, School of Medicine, University of California, Davis, CA 95616

Communicated by Alfred G. Redfield, Brandeis University, Waltham, MA, December 22, 1994 (received for review August 9, 1994)

ABSTRACT **Inordinate expansion and hypermethylation of the fragile X DNA triplet repeat, $(GGC)_n \cdot (GCC)_n$, are correlated with the ability of the individual G- and C-rich single strands to form hairpin structures. Two-dimensional NMR and gel electrophoresis studies show that both the G- and C-rich single strands form hairpins under physiological conditions. This propensity of hairpin formation is more pronounced for the C-rich strand than for the G-rich strand. This observation suggests that the C-rich strand is more likely to form hairpin or "slippage" structure and show asymmetric strand expansion during replication. NMR data also show that the hairpins formed by the C-rich strands fold in such a way that the cytosine at the CpG step of the stem is C·C paired. The presence of a C·C mismatch at the CpG site generates local flexibility, thereby providing analogs of the transition to the methyltransferase. In other words, the hairpins of the C-rich strand act as better substrates for the human methyltransferase than the Watson–Crick duplex or the G-rich strand. Therefore, hairpin formation could account for the specific methylation of the CpG island in the fragile X repeat that occurs during inactivation of the *FMR1* gene during the onset of the disease.**

Simple tandemly repeated DNA sequences are interspersed in both transcribed and nontranscribed regions of chromosomes (1–3). The hypothesis (4) that the unusual DNA structures adopted by these repeats principally determine their specific functions is gaining strength. We have previously described the unusual hairpin structures (5, 6) adopted by a variety of repetitive DNA sequences. Here, we show by NMR and gel electrophoresis that the individual strands from the fragile X triplet repeats, $(GGC)_n \cdot (GCC)_n$, form intramolecular hairpins under physiological conditions. In these hairpins, the number of Watson–Crick G·C pairs is maximized in the stem through the formation of G·G or C·C mispairs flanked by G·C pairs (Fig. 1). As shown below, these hairpins provide structural basis for three major phenomena associated with the fragile X syndrome (3, 4): (*i*) the site-specific fragility, (*ii*) the amplification of the repeat (especially the preferential expansion of the C-rich strand), and (*iii*) the hypermethylation of the CpG island adjacent to the fragile X gene, *FMR1*.

MATERIALS AND METHODS

Gel Electrophoresis. Oligonucleotides were fully denatured by heating at 95°C for 2 min in 5 mM Tris/1 mM EDTA buffer, pH 7.5, containing 5 mM or 200 mM NaCl, followed by incubation at 55°C for 10 min and gradual cooling to room temperature. The samples were equilibrated at room temperature for 10 min and loaded on a preequilibrated gel [20% polyacrylamide in 0.6× TBE buffer (1× TBE buffer = 0.09 M Tris borate, pH 8.3/2 mM EDTA)] and then electrophoresed at 4°C at 75 V. The gel was stained with ethidium bromide (10 μg/ml).

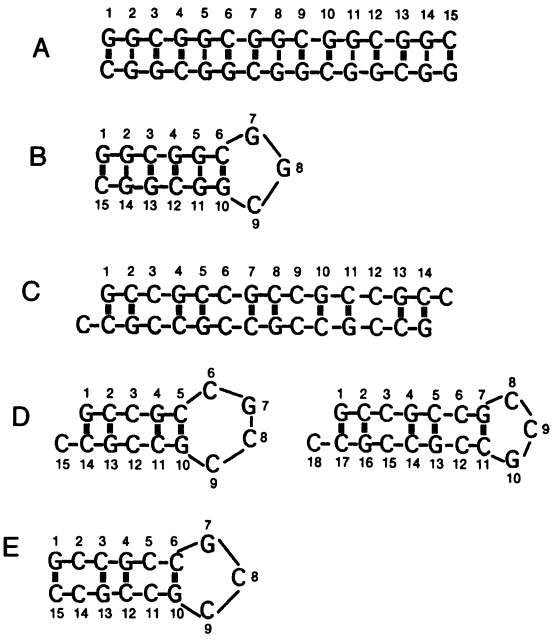

FIG. 1. Schematic representations of the self-assembled structure of $(GGC)_5$ duplex (*A*), $(GGC)_5$ hairpin (*B*), $(GCC)_5$ slipped duplex (*C*), $(GCC)_5$ and $(GCC)_6$ hairpins (*D*), and blunt $(GCC)_5$ hairpin (*E*). Note that in the hairpins of the G- and C-rich strands, the central mismatched G·G or C·C pair in the stem is surrounded by two Watson–Crick G·C pairs.

Methylation Assay. DNA methylation involving tritium incorporation was carried out, following the procedure of Smith *et al.* (6). A mixture containing 8 μM DNA and 8 μM tritiated *S*-adenosylmethionine was preincubated in the reaction buffer [50 mM Hepes, pH 7.0/50 mM NaCl/2 mM dithiothreitol/75 μM spermine/10% (vol/vol) glycerol] for 30 min at 37°C. The reaction was initiated with DNA methyltransferase and the reaction product was then recovered by precipitation with trichloroacetic acid (TCA). The rate of methylation was determined by measuring tritium incorporation into the TCA-insoluble DNA.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: NOE, nuclear Overhauser enhancement; NOESY, NOE spectroscopy; DQF-COSY, double-quantum filtered correlated spectroscopy.

‖To whom reprint requests should be addressed.

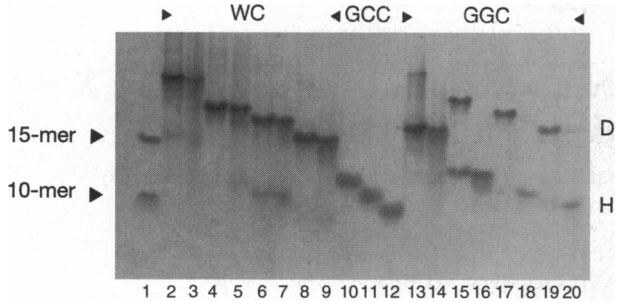

FIG. 2. Electrophoretic mobilities in nondenaturing polyacrylamide gel of single-stranded repeats $(GGC)_n$, $(GCC)_n$, and Watson–Crick (WC) heteroduplex in presence of NaCl at low and high concentrations (5 mM and 200 mM, respectively). Lane 1 shows the 15- and 10-mer DNA duplex markers. The positions of unimolecular hairpin conformation (i.e., hairpin H) and the self-complementary duplexes (D) with 15 and 10 Watson–Crick base pairs are also marked in the gel as 15- and 10-mer, respectively. For $(GGC)_n$ [lanes 14 (n = 11), 16 (n = 7), 18 (n = 6), and 20 (n = 5)], the hairpin is the predominant species in 5 mM NaCl. With increasing salt concentration (200 mM NaCl), the equilibrium shifts toward the $(GGC)_n$ duplex [lanes 13 (n = 11), 15 (n = 7), 17 (n = 6), and 19 (n = 5)]. The population of hairpin increases with increasing repeat number, n. As can be seen, the hairpin is still the dominant conformer for n = 11 even in 200 mM NaCl. Also note that the odd repeat number gives higher percentage of hairpin than the next even number [$(GGC)_5 > (GGC)_6$]. Lanes 10, 11, and 12 show the hairpin conformations of $(GCC)_n$ for n = 7, 6, and 5, respectively. Because $(GCC)_n$ remains in the hairpin conformation at low and high salt, the gel data are shown only for the low salt. Lanes 2, 4, 6, and 8 show the gel data for Watson–Crick duplexes, $(GGC)_n \cdot (GCC)_n$, for n = 11, 7, 6, and 5, respectively, in 200 mM NaCl. Lanes 3, 5, 7, and 9 show the gel data for Watson–Crick duplexes, $(GGC)_n \cdot (GCC)_n$, for n = 11, 7, 6, and 5, respectively, in 5 mM NaCl. Note that lanes 2–9 tend to show a small population of $(GCC)_n$ hairpins.

Gel assay for cytosine 5-methylation of the G-rich or C-rich strand and Watson–Crick duplex was carried out with the repeat number n = 7. For the bacterial methyltransferase *Sss* I, DNA methylation reactions were performed according to the instructions of the supplier, New England Biolabs. The DNA samples (8 μM) were preequilibrated and annealed in 1× NEB (50 mM NaCl/10 mM MgCl2/10 mM Tris·HCl, pH 7.5/1 mM dithiothreitol). Four units of *Sss* I and 160 μM *S*-adenosylmethionine were added and the mixture was incubated at 37°C for 2 hr. The same treatment was used for DNA substrates (8 μM) of the human methyltransferase in 4× Hepes buffer (50 mM NaCl/10 mM dithiothreitol/400 mM Hepes, pH 7.5). Equimolar concentrations of the complementary strands of G- or C-rich strands were added to the solution after the reaction was complete. This was followed by treatment with protease K to inactivate and digest the methyltransferase. The solution then was heated to 95°C and slowly cooled to anneal the Watson–Crick duplex. The proteins were removed by extraction with phenol and the DNA duplex was recovered by precipitation with ethanol. To determine the relative methylation on the target strands (G- or C-rich strands and Watson–Crick duplexes), the DNA samples were resuspended in a 20-μl solution containing 1× NEB and the methyl-sensitive restriction enzyme *Bso*FI (8 units) and were incubated at 55°C for 2 hr. The restriction enzyme *Bso*FI cuts double-stranded DNA with a recognition site of 5′-GCNGC-3′, but not the methylated form of the site. The digested products were analyzed by gel electrophoresis.

RESULTS

Identification of the Hairpins by Gel Electrophoresis. Theoretically, at neutral pH, the two individual strands of the fragile X repeat can form either a mismatched homoduplex or a monomeric hairpin (Fig. 1). The duplex and the stem of the corresponding hairpin for G- and C-rich strands involve G·G and C·C pairs, respectively. Note that the hairpin of a given sequence [i.e., $(GGC)_n$ or $(GCC)_n$] should have half the length of but approximately the same cross-section as the duplex of length n (Fig. 1). Therefore, the duplex is expected to have slower gel mobility than the corresponding hairpin. Fig. 2 shows the electrophoretic mobilities of G- and C-rich single strands along with the marker duplex sequences of defined lengths in a nondenaturing polyacrylamide gel. The gel data show the presence of two structural forms of the G-rich strands—hairpin and duplex. For a given sequence, the relative population of the hairpin or the duplex depends on the repeat number (n) and the DNA and salt concentrations. For low n, the hairpin of the G-rich strands is the predominant conformation at low salt (5 mM NaCl) and the duplex is the major conformer at high salt (200 mM NaCl). With increasing n, the equilibrium gradually shifts toward the hairpin structure within the salt concentration range of 5–200 mM NaCl. For n = 11, the hairpin is the predominant conformation at all salt concentrations. The gel data of C-rich strands show only hairpins at both 5 and 200 mM NaCl. Thus the C-rich strands more readily form hairpins than the G-rich strands. The thermodynamic preference for the individual G- and C-rich strands to form hairpins as a function of n is determined by how easily the fragile X duplex, $(GGC)_n \cdot (GCC)_n$, can undergo transitions to two unimolecular hairpins (Fig. 1). This tendency of the G-rich strand not to form a hairpin for low repeat numbers at physiological salt concentrations implies that the C-rich strand does not dissociate from the Watson–Crick paired duplex to form a hairpin at low n. This is primarily because the formation of hairpins by the individual G- and

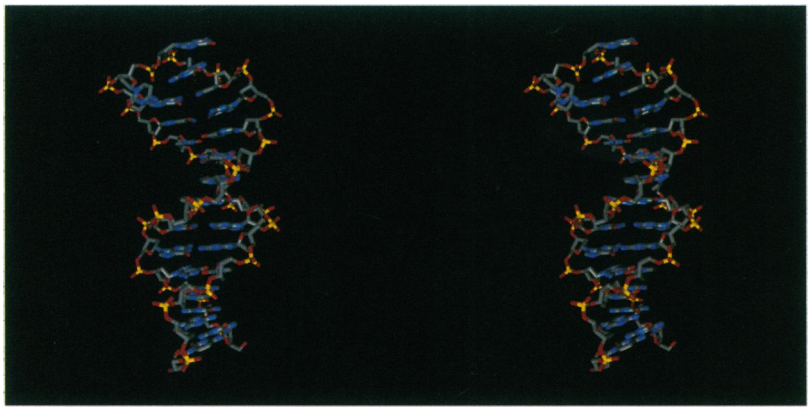

FIG. 3. Stereo pair of a representative energy-minimized structure of the $(GCC)_5$ hairpin consistent with the NMR data. Analysis of NOESY data (200 and 500 ms of mixing time) of $(GCC)_5$ with the aid of full-matrix NOESY simulation resulted in a set of 135 distance constraints.

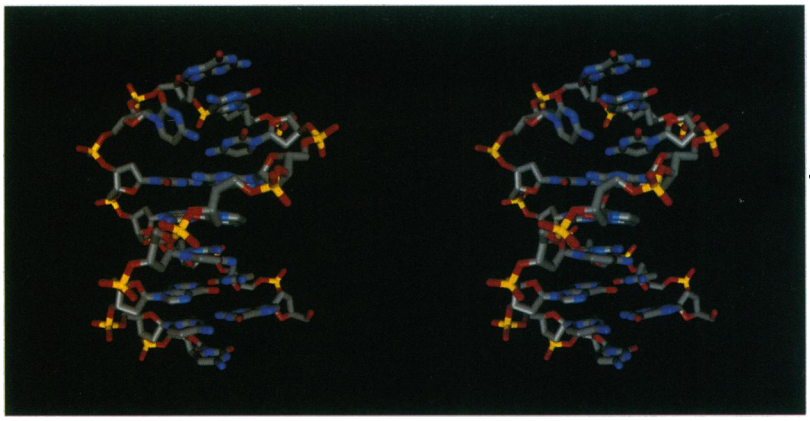

FIG. 4. A stereo pair of the hairpin model of $(GGC)_9$ in which the stem satisfies the NMR constraints of the $[(GGC)_4]_2$ duplex. Analysis of NOESY data (200 and 500 ms mixing time) of $[(GGC)_4]_2$ with the aid of full-matrix NOESY simulation resulted in a set of 163 average interproton distances for the 12-base-pair-long duplex.

C-rich strands is energetically less favorable than formation of the Watson–Crick duplex $(GGC)_n{\cdot}(GCC)_n$. The biological occurrence of repeats which are all longer than $n = 11$ relieves this constraint. Nonetheless, the observation that for appropriate n the Watson–Crick duplex $(GGC)_n{\cdot}(GCC)_n$ can undergo a transition to the hairpins formed by the individual single strands is of great importance, since these hairpins provide molecular descriptions of the "slippage" structures proposed to be replication intermediates required for strand expansion (see Fig. 7).

Structural Characterization of the Hairpins by NMR. Knowledge of the precise stereochemistry of the hairpins, including the exact base-base interactions in the stem and loop, is important in understanding their biological role. Extensive one- and two-dimensional (1D/2D) NMR experiments were carried out to obtain the stereochemical details of the hairpin structures formed by the individual G- and C-rich strands of the fragile X repeat. The nature of base-pairing patterns was identified by monitoring (*i*) temperature-dependent imino proton profiles and (*ii*) nuclear Overhauser enhancements (NOEs) from the imino protons in 1D/2D NOE experiments in ^1H_2O:^2H_2O (9:1, vol/vol). Glycosidic torsions and sugar puckers were deduced by interpreting the NOE intensities and the coupling constants derived from double-quantum filtered correlated spectroscopy (DQF-COSY). A set of average interproton distances for pairwise interactions was obtained by performing full-relaxation-matrix simulation with the NOE spectroscopy (NOESY) intensities at 500 and 200 ms of mixing time in conjunction with the linked-atom least squares refinement (7). The starting structure was constructed from these average interproton distances. The interproton distance constraints as well as the base-pairing constraints that were consistent with the 2D NMR data were included as harmonic potentials in restrained molecular dynamics and energy minimization (5, 7). Detailed NMR studies will be published elsewhere.

C-Rich Strand. Fig. 3 shows the stereo pair of a representative energy-minimized structure of the $(GCC)_5$ hairpin. All nucleotides in the hairpin adopt C2′-*endo*, *anti* conformation as evidenced by the NOESY and DQF-COSY data. In this structure, the 5′-C in the CpG step of the stem is C·C paired and not G·C paired. The C·C pair most likely involves a single hydrogen bond between N4-H and N3. This leads to two possibilities in which either of the two cytosines can act as a proton donor (participation of N4-H) or an acceptor (participation of N3). In a 400-ps unrestrained molecular dynamics simulation, we observed that the C3·C12 pair in the $(GCC)_5$ hairpin (Fig. 1) can undergo local periodic sliding motions between the two degenerate hydrogen-bonding states without violating local or distant NOE constraints. Such a sliding motion makes cytosines in the C·C pairs intrinsically more flexible than cytosines in the G·C pairs. Previously Gueron *et al.* (8) showed that mismatch pairs are more susceptible to base-pair open-closure dynamics than are the canonical Watson–Crick pairs. In addition, weaker intra- and internucleotide NOESY cross-peaks at the C·C pairs of the $(GCC)_5$ and $(GCC)_6$ hairpins indicate the presence of local flexibility.

G-Rich Strands. The structural details of the hairpins formed by the G-rich strand are also important for understanding the expansion of this strand during replication. However, complete structure determination of these hairpins by NMR was not possible, since we found that the G-rich strand exists predominantly in the duplex state for a range of DNA concentrations (0.5–2 mM in strand) for $n \geq 11$. Since the stem of the hairpin and the duplex are conformationally similar, it is noteworthy to mention some of the salient features of the duplex: (*i*) $(GGC)_4$, $(GGC)_5$, and $(GGC)_6$ form a six-base-pair-long structural repeat

$$G1^{anti}\text{-}G2^{anti}\text{-}C3^{anti}\text{-}G4^{anti}\text{-}G5^{syn}\text{-}C6^{anti}\text{-}G7^{anti}$$
$$C1^{anti}\text{-}G2^{syn}\text{-}G3^{anti}\text{-}C4^{anti}\text{-}G5^{anti}\text{-}G6^{anti}\text{-}C7^{anti}.$$

(*ii*) Two symmetric O6···H-N1 hydrogen bonds are present in the $G^{anti}{\cdot}G^{syn}$ pairing. (*iii*) The cytosines in the two CpG sequences of this repeat are distinguished by the presence of $G^{anti}{\cdot}G^{syn}$ on the 5′ end of C. The (5′-3′) sequential H2″(C)–H8(G) NOE at the CpG fragment is weak when G^{anti} is the 5′ neighbor of C (i.e., C3pG4) and the same NOE is totally absent when G^{syn} is the 5′ neighbor of C (i.e., in C6pG7). This (5′-3′) sequential H1′–H8 and H6–H8 NOEs are also absent at the two CpG segments of the repeat. This type of intrastrand NOE discontinuity suggests a local structural discontinuity at the CpG step due to the formation of adjacent $G^{anti}{\cdot}G^{syn}$ pairs. We used a molecular modeling approach to construct the hairpin

Table 1. Rates of methylation of the cytosines in the CpG step of triplet repeats

Methyl-transferase	Repeat (*n*)	Methylation rate, fmol/min		
		GCC	GGC	WC
Sss I	5	—	—	—
	6	—	2.6	2.0
	7	—	24	23
	11	7.1	30	21
Human	5	3.9	1.1	1.6
	6	12	1.7	4.9
	7	23	9.5	14
	11	290	27	54

All values are averages of four measurements. A dash indicates too low to measure. Hairpins contain half the potential methylation sites present in the Watson–Crick (WC) duplex.

structures of the G-rich strands. The stem of the hairpin is constructed on the basis of the NMR data of the duplex and then the two arms of the stem are connected by energetically stable loop segments. Fig. 4 shows the stereo pair of the proposed energy-minimized hairpin model of $(GGC)_9$ in which the stem structure is consistent with the NMR data of $[(GGC)_4]_2$ duplex.

In summary, the NMR and gel data show that both $(GCC)_n$ and $(GGC)_n$ strands can form hairpin structures under physiological conditions. Once formed, these monomeric hairpins may be susceptible to nuclease attack. Nucleases with apparent specificity for single-stranded loop regions have been reported in yeast (9). Hence, this tandemly repeated DNA sequence may become fragile simply due to its propensity to form hairpins.

Site-Specific Methylation of the Hairpins by Human and Bacterial Methyltransferases. Apart from being a site prone to cleavage, the fragile X repeat is also associated with the *FMR1* gene (3). In fragile X syndrome, the cytosines at the CpG sites become hypermethylated. This methylation may play a role in inactivating the *FMR1* gene (10). We used the bacterial methyltransferase (*Sss* I) and the methyltransferase from human placenta to study the methylation of the G-rich and C-rich strands under conditions in which unimolecular hairpins were the primary species.

The bacterial methyltransferase, *Sss* I, appears to have a strict requirement for a Watson–Crick paired CpG as found in the Watson–Crick duplex and the G-rich hairpin (structure **1** in Fig. 6). These two forms are methylated by *Sss* I at similar rates (Table 1). The substrate efficiency of the G-rich strand with repeat number 11 is 1.5 times higher than that of the Watson–Crick duplex. However, in the C-rich strand the CpG dinucleotides are disrupted by a C·C mispair. The substrate efficiency of the C-rich hairpins for *Sss* I is 1/4 that of the Watson–Crick duplex (Table 1). Fig. 5*A* shows (in a gel assay) the protection against *Bso*FI subsequent to methylation by *Sss* I. The restriction enzyme *Bso*FI has the duplex recognition site GCNGC, where N can be any nucleotide and the cytosines are unmethylated. Upon methylation of the cytosines, the same sequence becomes resistant to *Bso*FI. Note that the G-rich strand shows the highest protection against *Bso*FI (Fig. 5*A*) because the recognition site in this strand, GCGGC, gets methylated at the cytosines by *Sss* I.

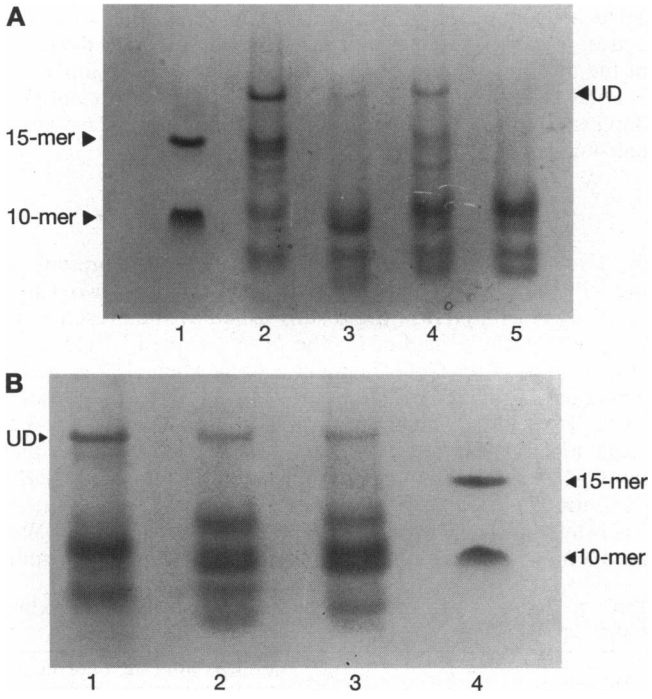

FIG. 5. Gel assay for detecting methylation in the CpG step of the G-rich and C-rich single strands and the Watson–Crick heteroduplex. (*A*) DNA methylation was catalyzed by bacterial methyltransferase *Sss* I. Lanes: 1, DNA markers, 10-mer and 15-mer duplex; 2, *Bso*FI digest of methylated G-rich strands; 3, *Bso*FI digest of methylated C-rich strands; 4, *Bso*FI digest of methylated Watson–Crick duplexes; 5, *Bso*FI digest of unmethylated Watson–Crick duplexes (restriction control). UD, undigested. (*B*) DNA methylation was catalyzed by human methyltransferase. Lanes: 1, DNA markers; 2, *Bso*FI digest of methylated C-rich strands in presence of the unmethylated complementary G-rich strand; 3, *Bso*FI digest of G-rich strands in presence of the unmethylated complementary C-rich strand; 4, *Bso*FI digest of methylated Watson–Crick duplexes.

In contrast, as shown in Table 1, compared with the Watson–Crick duplex the enzymatic activity of the human methyltransferase is 6-fold stimulated by the presence of the C·C mispair at the CpG site (11, 12); see structure **2** in Fig. 6. As shown in Fig. 5*B*, a weak protection against *Bso*FI cleavage is observed in the C-rich strand subsequent to methylation by the human methyltransferase even when tritium incorporation is quite significant. This is because the human enzyme methylates the C-rich strand in such a way that the *Bso*FI recognition site is altered at the nonessential nucleotide (i.e., GCmCGC).

Structural Basis of CpG Methylation Inside the Fragile X Repeat. Bacterial and human methyltransferases operate by the same mechanism (13, 14). For each enzyme, initial sequence-specific recognition is followed by nucleophilic attack at C6 to form a nonplanar dihydrocytosine intermediate which cannot be accommodated by the structure of a B-DNA helix (11). In the case of the bacterial enzyme *Hha* I, the crystal structure of the intermediate (15) is known. This crystal structure suggests that the enzymes have evolved to accommodate the nonplanar intermediate by rotating it to an extrahelical position within the active site of the enzyme. Therefore,

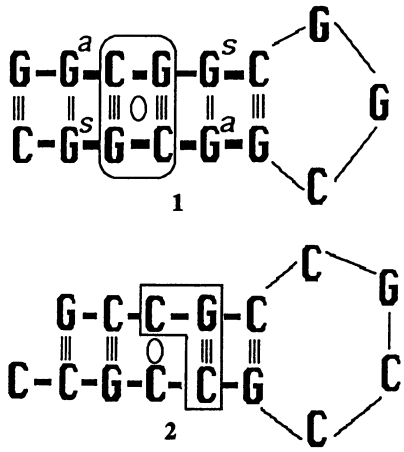

FIG. 6. The structural features of the recognition elements of bacterial and human methyltransferases. The hairpin of the G-rich strand (structure **1**) has the recognition sequence

$$5'\text{-}G^{anti}\text{-}C^{anti}\text{-}G^{anti}\text{-}G^{syn}\text{-}3'$$
$$3'\text{-}G^{syn}\text{-}G^{anti}\text{-}C^{anti}\text{-}G^{anti}\text{-}5'$$

for the bacterial enzyme *Sss* I (two potential methylation sites in the sequence are shown in bold). Note that the Watson–Crick paired CpG site is flanked by $G^{anti}\cdot G^{syn}$ mismatches which introduce structural discontinuity at the CpG step and flexibility at the CpG link. The recognition element for *Sss* I is boxed. The C-rich hairpin (structure **2**) provides the recognition sequence

$$5'\text{-}C^{anti}\text{-}C^{anti}\text{-}G^{anti}\text{-}3'$$
$$3'\text{-}G^{syn}\text{-}C^{anti}\text{-}C^{anti}\text{-}5'$$

for the human enzyme. The recognition element for the human enzyme is boxed. Note that the cytosines in the CpG step are involved in C·C pairs which offer a great deal of flexibility to the cytosines in the CpG step. The twofold symmetry in the recognition sequences of the bacterial and human enzymes is marked with an ellipse.

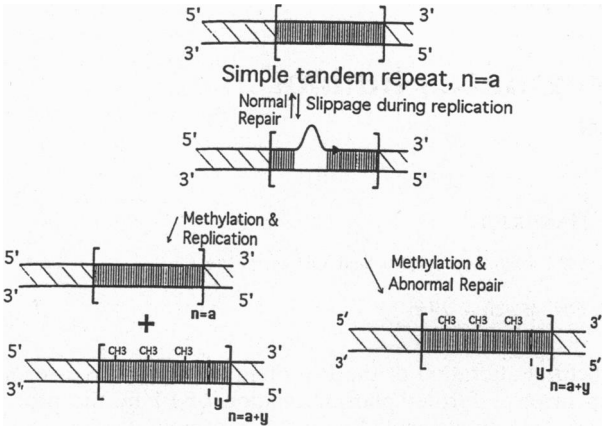

FIG. 7. Slippage during replication is proposed. The slippage is due to the formation of the hairpins by the G- and C-rich strands shown respectively as structures **1** and **2** (Fig. 6). The amount of expansion which can occur due to abnormal repair of a second round of replication depends on the size of the hairpin—i.e., the longer the size of hairpin (or the slippage), the longer the size (y) of expansion. From our studies, it appears that the C-rich strand of the fragile X repeat is most likely to get methylated first, and then in the next round of replication the complementary G-rich strand will also be methylated by the mechanism of maintenance or methyl-directed methylation (6, 16, 17).

any structure that satisfies the sequence recognition requirement and also possesses sufficient flexibility at the CpG site to allow the C to be rotated out would be a better substrate for the methyltransferase than the Watson–Crick duplex. Note that the NMR-derived structure of the C-rich strand hairpin (Figs. 3 and 6) satisfies this requirement because of the presence of the flexible C·C pair at the CpG that allows considerable freedom for the C to flip in and out of the structure. Similarly, in the stem of the hairpins of G-rich strands (Figs. 4 and 6), the presence of G^{syn}·G^{anti} pair adjacent to the CpG site causes local flexibility such that the C can flip in and out of the structure with relative ease.

DISCUSSION

On the basis of the structural and enzymological data described above, we propose a simple mechanism to explain the molecular events associated with fragile X syndrome. Fig. 7 describes a slippage model suggested by the strong tendency of the triplet repeats to form hairpins. In this scheme, the G- and the C-rich strands independently form hairpin structures. During replication, the growing chain can easily slip or slip and slide to form a hairpin which makes a part of the already replicated template again available for replication. This would result in expansion of the repeat in the next replication cycle. Our studies also predict that the C-rich strand is more likely to form slippage structure and get preferentially expanded during *in vitro* replication. Ordinarily repair through loop excision would preserve the sites intact although this might promote chromosomal breaks. Unrepaired loops would be substrates for *de novo* methylation (Fig. 7). Erroneous repair against the longer strand in the heteroduplex could also result in repeat expansion. The tendency for expansion is dependent on repeat number, since the tendency for hairpin formation is also dependent on repeat number. This tendency has been referred to as a dynamic mutation (18). Stability of repeats shorter than $n = 50$ and instability of repeats longer than $n = 50$ may reflect the tendency of the repair systems to become saturated when loop formation in the region occurs at high frequency. Finally, our studies suggest that methylation of the C-rich strands is likely to be a consequence of hairpin formation. However, we cannot rule out the possibility that methylation of loops might cause site-specific fragility (10). Moreover, it should be added that the maintenance of methylation as shown in Fig. 7 is associated with late replication of the fragile X repeat (16). This might lead to delayed chromosome condensation or gap formation resulting in site-specific fragility.

We thank Dr. Angel E. Garcia for his help during various stages of this work. We thank Drs. Neville Kallenbach of New York University and Christian Burks of Los Alamos National Laboratory for suggestions and corrections on the manuscript. We are grateful to Dr. Cliff Unkefer for giving us access to the 500-MHz Bruker-AMX NMR spectrometer at Chemical Science and Technology 4 (CST-4). This work was supported by Los Alamos National Laboratory Grant XL-77 and the Human Genome Project of the Office of Health and Environmental Research of the Department of Energy and by Grant 0388 from the Smokeless Tobacco Research Council, Inc.

1. Moyzis, R. K., Torney, D. C., Meyne, J., Buckingham, J. M., Wu, J. R., Burks, C., Sirotkin, K. M. & Goad W. B. (1989) *Genomics* **4,** 273–289.
2. Grady, D. I., Ratliff, R. L., Robinson, D. L., McCanlies, E. C., Meyne, J. & Moyzis, R. K. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 1695–1699.
3. Oberié, I., Rousseau, F., Heitz, D., Kretz, C., Devys, D., Hanauer, A., Boué, J., Bertheas, M. F. & Mandel, J. L. (1991) *Science* **252,** 1097–1102.
4. Sinden, R. R. & Wells, R. D. (1992) *Curr. Opin. Biotechnol.* **3,** 612–622.
5. Catasti, P., Gupta, G., Garcia, A. E., Ratliff, R., Hong, L., Yau, P., Moyzis, R. K. & Bradbury, E. M. (1993) *Biochemistry* **33,** 3819–3830.
6. Smith, S. S., Laayoun, A., Lingerman, R. G., Baker, D. J. & Riley, J. (1994) *J. Mol. Biol.* **243,** 143–151.
7. Gupta, G., Garcia, A. E. & Hiriyanna, K. T. (1993) *Biochemistry* **32,** 948–960.
8. Gueron, M., Kochoyan, M. & Leroy, J. (1987) *Nature (London)* **328,** 89–92.
9. Symington, L. S. & Kolodner, R. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 7247–7251.
10. Nancacarrow, J., Kramer, E., Holman, K., Eyre, H., Doggett, N. A., LaPeaslier, D., Callen, D. F., Sutherland, G. R. & Richards, R. I. (1994) *Science* **264,** 1938–1941.
11. Smith, S. S., Kan, J. L. C., Baker, D. J., Kaplan, B. E. & Dembek, P. (1991) *J. Mol. Biol.* **217,** 39–51.
12. Smith, S. S., Hard, T. A. & Baker, D. J. (1987) *Nucleic Acids Res.* **15,** 6899–6916.
13. Osterman, D. G., Depillis, G. D., Wu, J. C., Matsuda, A. & Santi, D. V. (1988) *Biochemistry* **27,** 5204–5210.
14. Smith, S. S., Kaplan, B. E., Sowers, L. C. & Newman, E. M. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 4744–4748.
15. Klimasanskas, S., Kumar, S., Roberts, R. J. & Cheng, X. (1994) *Cell* **76,** 357–369.
16. Laird, C. D., Jaffe, E., Karpen, G., Lamb, M. & Nelson, R. (1987) *Trends Genet.* **3,** 274–281.
17. Riggs, A. D. (1975) *Cytogenet. Cell Genet.* **14,** 9–25.
18. Richards, R. I. & Sutherland, G. R. (1994) *Nat. Genet.* **8,** 114–116.