Non-B DNA Conformations Formed by Long Repeating Tracts of Myotonic Dystrophy Type 1, Myotonic Dystrophy Type 2, and Friedreich’s Ataxia Genes, Not the Sequences per se, Promote Mutagenesis in Flanking Regions*

Received for publication, April 24, 2006 Published, JBC Papers in Press, June 21, 2006, DOI 10.1074/jbc.M603888200

Marzena Wojciechowska, Marek Napierala, Jacquelynn E. Larson, and Robert D. Wells

From the Institute of Biosciences and Technology, Center for Genome Research, Texas A&M University System Health Science Center, Houston, Texas 77030

The expansions of long repeating tracts of CTG-CAG, CCTG-CAGG, and GAA-TTC are integral to the etiology of myotonic dystrophy type 1 (DM1), myotonic dystrophy type 2 (DM2), and Friedreich’s ataxia (FRDA). Essentially all studies on the molecular mechanisms of this expansion process invoke an important role for non-B DNA conformations which may be adopted by these repeat sequences. We have directly evaluated the role(s) of the repeating sequences per se, or of the non-B DNA conformations formed by these sequences, in the mutagenic process. Studies in *Escherichia coli* and three types of mammalian (COS-7, CV-1, and HEK-293) fibroblast-like cells revealed that conditions which promoted the formation of the non-B DNA structures enhanced the genetic instabilities, both within the repeat sequences and in the flanking sequences of up to ~4 kbp. The three strategies utilized included: the *in vivo* modulation of global negative supercoil density using topA and gyrB mutant *E. coli* strains; the *in vivo* cleavage of hairpin loops, which are an obligate consequence of slipped-strand structures, cruciforms, and intramolecular triplexes, by inactivation of the SbcC protein; and by genetic instability studies with plasmids containing long repeating sequence inserts that do, and do not, adopt non-B DNA structures *in vitro*. Hence, non-B DNA conformations are critical for these mutagenesis mechanisms.

Segments of certain genomes can be structurally polymorphic (1–5). The structural transitions of DNA regions from the right-handed B-form to at least eleven non-B conformations is a striking feature of the expanded microsatellite sequences associated with a number of human hereditary neurological diseases. Long repeating tracts of CTG-CAG, CCTG-CAGG, and GAA-TTC can adopt strand-slipped structures with hairpin loops, triplexes, tetraplexes, cruciforms, and sticky DNA under appropriate conditions. The expansions of these sequences are related to the etiology of myotonic dystrophy type 1 (DM1), myotonic dystrophy type 2 (DM2), and Friedreich’s ataxia (FRDA), respectively. Virtually all studies on the molecular mechanisms of the expansion and deletion processes invoke these non-B DNA conformations as crucial elements (reviewed in Ref. 1). The quasi-stable non-B DNA conformations may impede or block replication (6) and/or transcription (4, 7–9).

Non-B DNA structure-prone sequences are unstable and recombinogenic, and are a common source of instability leading to gross DNA rearrangements (10–15). Recent reports have shown that non-B structure-forming DNA from natural sources such as long CTG-CAG repeats (11, 14), short sequences expected to adopt Z- or H-DNA (12, 13), as well as the poly (R-Y) tract from human *PKD1* gene (10) exhibited mutagenic effects when replicated in mammalian and bacterial cells. Double-stranded breaks (DSBs) were accumulated at and around the repeating sequences and error-prone repair pathways were proposed to be involved in the formation of gross DNA rearrangements (16). The presence of long repeating sequences caused the formation of expansions and deletions within the repeat tracts (reviewed in Ref. 1) and promoted gross deletions, insertions, point mutations, inversions, and other rearrangements several kb distal from the site of the repeat tracts (10–13, 15, 16).

Global negative supercoil density (17) acts in concert with local transient waves of topological changes (18, 19) generated by replication and/or transcription, and both have a critical influence on the formation and stability of non-B DNA structures *in vivo* (reviewed in Ref. 2). A higher level of negative supercoiling destabilized long CTG-CAG, CCG-CGG, and GAA-TTC repeats in *Escherichia coli* (17) and lowered the viability of the host cells in the presence of the 2.5-kb poly(R-Y) tract from the human *PKD1* gene (20). Thus, the mutagenic potential of specific repeating DNA sequences is modulated by the intracellular conditions that affect the propensities of the non-B DNA conformations *in vivo*.

---

*This work was supported by Grants NS37554 and ES11347 from the National Institutes of Health, the Robert A. Welch Foundation, and Seek a Miracle (Muscular Dystrophy Association). 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.

1 To whom correspondence should be addressed. Tel.: 713-677-7651; Fax: 713-677-7689; E-mail: rwells@ibt.tamhsc.edu.

2 The abbreviations used are: DM1, myotonic dystrophy type 1; DSB, double-stranded break; TRS, triplet repeat sequence; RS, repeat sequence; IPTG, isopropyl β-D-thiogalactoside; LM-PCR, linker-mediated PCR; NHEJ, non-homologous end-joining; CFU, colony-forming unit; FACS, fluorescence-activated cell sorter; SSA, single-strand annealing; GFP, green fluorescent protein; CMV, cytomegalovirus; nt, nucleotide; DM2, myotonic dystrophy type 2; FRDA, Friedreich’s ataxia.
Non-B DNA Structures Promote Mutagenesis

Three independent strategies were employed herein to investigate the role of non-B DNA conformations in the genetic instabilities caused by repeating tri- and tetranucleotide sequences within the repeat tracts and in their flanking regions. First, the in vivo negative supercoiling density was altered in E. coli with topA and gyrB mutants to modulate the torsion on the DNA, which determines the stability of non-B DNA structures; high levels of negative supercoiled density stabilize the underwound non-B DNA conformations and thus induce greater genetic instabilities, as expected (17). Second, the extent of genetic instabilities in the DM2 repeats and their flanking sequences was investigated in the presence and absence of the E. coli SbcC protein which cleaves hairpin loop structures and thereby abolishes mutagenesis; when the SbcC protein was inactive, the DM2 repeats were inert in their mutagenic capacity, thus, strengthening the concept of the role of hairpin loops and slipped structures in the mutagenic process. Third, comparative instability studies were conducted with long repeats of CAA-TTG relative to the other repeat sequences; prior investigations (21, 22) revealed the inability of the CAA sequences to adopt non-B DNA structures. Indeed, this TRS was shown to be inert in the mutagenic capacity of its small-molar derivatives (21) (Fig. 1A, right panel).

The myotonic dystrophy type 1 (DM1) triplet repeat tract (CTG-CAG)n was obtained from pRW5302 (11) by excision with EcoRI/EagI (the recessed 5’ terminus of the fragment was filled-in with dNTPs using the Klenow fragment of E. coli DNA polymerase I) (U.S. Biochemical Corp.), and was used for cloning the repeats in both orientations (27). The pEGFP-C1 vector was digested with EcoRI/Smal or EcoRI/Xhol (the recessed 5’ terminus of the latter fragment was filled-in as described above) and then used to clone the insert in orientations I and II, respectively. The (CTG-CAG)198 tract was excised form pRW5301 (11) and, for cloning in orientation II, proceeded as described above for the 175 mer, whereas, for orientation I, the EcoRI/HindIII filled-in fragment (11) was cloned into the Smal site of the pEGFP-C1 vector.

The myotonic dystrophy type 2 (DM2) tetranucleotide repeats (CCTG-TTC)n (where n = 114 and 200) were obtained from two pCR2.1TOPO derivatives (28). To prepare plasmids harboring the (CCTG-CAGG)114 insert in orientations I (29), the parental plasmid was cleaved with Xhol/BamHI and HindIII/SmaI and utilized to clone the inserts in orientations I and II, respectively. Restriction fragments containing the (CCTG-CAGG)100 repeat were excised from their derivative plasmids by HindIII/Xhol and HindIII/EcoRV and were used for cloning into the pEGFP-C1 in orientations I and II, respectively. The vector digested with HindIII/Xhol and HindIII/Smal was ligated with the inserts to obtain clones in orientations I and II, respectively.

The Friedrich’s ataxia sequences (GAA-TTC)n (where n = 60 and 150), were excised from pRW3804 and pRW3805 (30), respectively, by Stul/BssHII cleavage (the recessed 5’ termini of the fragments were filled-in as described above) and were used for blunt-end ligation to the vector that was linearized with SmaI.

The CAA-TTG sequence that was found to be inert in the formation of hairpin structures (9, 21), was prepared as follows. Two complementary synthetic oligonucleotides (CAA)33 and (TTC)33 (Sigma Genosys) were purified by denaturing gel electrophoresis, eluted from the gel and ethanol-precipitated (31). The oligonucleotides were phosphorylated using 0.2 mm ATP and T4 polynucleotide kinase (U.S. Biochemical Corp.) and annealed as described (32). After 12 h of ligation conducted at 16 °C, the reaction mixture was subjected to MfeI digestion to remove part of the enhancer region, the TATA box was digested with EcoRI/BamHI and HindIII/SmaI and utilized to clone the inserts in orientations I and II, respectively.

The CAA-TTG sequence that was found to be inert in the formation of hairpin structures (9, 21), was prepared as follows. Two complementary synthetic oligonucleotides (CAA)33 and (TTC)33 (Sigma Genosys) were purified by denaturing gel electrophoresis, eluted from the gel and ethanol-precipitated (31). The oligonucleotides were phosphorylated using 0.2 mm ATP and T4 polynucleotide kinase (U.S. Biochemical Corp.) and annealed as described (32). After 12 h of ligation conducted at 16 °C, the reaction mixture was subjected to MfeI digestion to remove part of the enhancer region, the TATA box was digested with EcoRI/BamHI and HindIII/SmaI and utilized to clone the inserts in orientations I and II, respectively.

The CAA-TTG sequence that was found to be inert in the formation of hairpin structures (9, 21), was prepared as follows. Two complementary synthetic oligonucleotides (CAA)33 and (TTC)33 (Sigma Genosys) were purified by denaturing gel electrophoresis, eluted from the gel and ethanol-precipitated (31). The oligonucleotides were phosphorylated using 0.2 mm ATP and T4 polynucleotide kinase (U.S. Biochemical Corp.) and annealed as described (32). After 12 h of ligation conducted at 16 °C, the reaction mixture was subjected to MfeI digestion to remove part of the enhancer region, the TATA box was digested with EcoRI/BamHI and HindIII/SmaI and utilized to clone the inserts in orientations I and II, respectively.

The CAA-TTG sequence that was found to be inert in the formation of hairpin structures (9, 21), was prepared as follows. Two complementary synthetic oligonucleotides (CAA)33 and (TTC)33 (Sigma Genosys) were purified by denaturing gel electrophoresis, eluted from the gel and ethanol-precipitated (31). The oligonucleotides were phosphorylated using 0.2 mm ATP and T4 polynucleotide kinase (U.S. Biochemical Corp.) and annealed as described (32). After 12 h of ligation conducted at 16 °C, the reaction mixture was subjected to MfeI digestion to remove part of the enhancer region, the TATA box was digested with EcoRI/BamHI and HindIII/SmaI and utilized to clone the inserts in orientations I and II, respectively.

The CAA-TTG sequence that was found to be inert in the formation of hairpin structures (9, 21), was prepared as follows. Two complementary synthetic oligonucleotides (CAA)33 and (TTC)33 (Sigma Genosys) were purified by denaturing gel electrophoresis, eluted from the gel and ethanol-precipitated (31). The oligonucleotides were phosphorylated using 0.2 mm ATP and T4 polynucleotide kinase (U.S. Biochemical Corp.) and annealed as described (32). After 12 h of ligation conducted at 16 °C, the reaction mixture was subjected to MfeI digestion to remove part of the enhancer region, the TATA box was digested with EcoRI/BamHI and HindIII/SmaI and utilized to clone the inserts in orientations I and II, respectively.

The CAA-TTG sequence that was found to be inert in the formation of hairpin structures (9, 21), was prepared as follows. Two complementary synthetic oligonucleotides (CAA)33 and (TTC)33 (Sigma Genosys) were purified by denaturing gel electrophoresis, eluted from the gel and ethanol-precipitated (31). The oligonucleotides were phosphorylated using 0.2 mm ATP and T4 polynucleotide kinase (U.S. Biochemical Corp.) and annealed as described (32). After 12 h of ligation conducted at 16 °C, the reaction mixture was subjected to MfeI digestion to remove part of the enhancer region, the TATA box was digested with EcoRI/BamHI and HindIII/SmaI and utilized to clone the inserts in orientations I and II, respectively.

The CAA-TTG sequence that was found to be inert in the formation of hairpin structures (9, 21), was prepared as follows. Two complementary synthetic oligonucleotides (CAA)33 and (TTC)33 (Sigma Genosys) were purified by denaturing gel electrophoresis, eluted from the gel and ethanol-precipitated (31). The oligonucleotides were phosphorylated using 0.2 mm ATP and T4 polynucleotide kinase (U.S. Biochemical Corp.) and annealed as described (32). After 12 h of ligation conducted at 16 °C, the reaction mixture was subjected to MfeI digestion to remove part of the enhancer region, the TATA box was digested with EcoRI/BamHI and HindIII/SmaI and utilized to clone the inserts in orientations I and II, respectively.

The CAA-TTG sequence that was found to be inert in the formation of hairpin structures (9, 21), was prepared as follows. Two complementary synthetic oligonucleotides (CAA)33 and (TTC)33 (Sigma Genosys) were purified by denaturing gel electrophoresis, eluted from the gel and ethanol-precipitated (31). The oligonucleotides were phosphorylated using 0.2 mm ATP and T4 polynucleotide kinase (U.S. Biochemical Corp.) and annealed as described (32). After 12 h of ligation conducted at 16 °C, the reaction mixture was subjected to MfeI digestion to remove part of the enhancer region, the TATA box was digested with EcoRI/BamHI and HindIII/SmaI and utilized to clone the inserts in orientations I and II, respectively.

The CAA-TTG sequence that was found to be inert in the formation of hairpin structures (9, 21), was prepared as follows. Two complementary synthetic oligonucleotides (CAA)33 and (TTC)33 (Sigma Genosys) were purified by denaturing gel electrophoresis, eluted from the gel and ethanol-precipitated (31). The oligonucleotides were phosphorylated using 0.2 mm ATP and T4 polynucleotide kinase (U.S. Biochemical Corp.) and annealed as described (32). After 12 h of ligation conducted at 16 °C, the reaction mixture was subjected to MfeI digestion to remove part of the enhancer region, the TATA box was digested with EcoRI/BamHI and HindIII/SmaI and utilized to clone the inserts in orientations I and II, respectively.
repeats, respectively, were present on the leading strand template from the bidirectional SV40 origin of replication. The controls pEGFP-C1 and pRW5329 were devoid of any long repeat tracts.

The DM1 repeat tracts have 19 and 41 bp of nonrepetitive human flanking sequences 5' and 3' of the tracts, respectively (27). The (CTG-CAG)_{10} sequence was pure (uninterrupted), whereas the (CTG-CAG)_{1725} carried two G → A interruptions at repeats 28 and 69 (27). The (CCTG-CAGG)_{14} and the (CCTG-CAGG)_{100} contained a single G → T interruption at repeat 12 that gives the sequence (CCTG)_{14},CCTT(CCTG)_{18}, where n = 102 and 188, respectively (29). The sequence of the (GAA-TTC)_{60} insert was (GAA)_{60},GAAAAAGAAAA(GAA)_{10}, which was named (GAA-TTC)_{60} for simplicity. The (GAA-TTC)_{14} run was (GAA)_{144},GAG(GAA)_{2}. Both repeating tracts were flanked by 34 bp and 102 bp of the human FRDA gene (33).

Fig. 1A, left side shows the intact pCAMV plasmids that harbored the repeat tracts cloned in both orientations, and downstream of the EGFP reporter gene. These constructs were used to prepare the set of plasmids with the inactive CMV promoter (Fig. 1A, right panel).

Construction of the pGFPT Derivatives—For experiments conducted in E. coli, pGFPT containing the GFP reporter gene was used as a cloning vector (10). The (CCTG-CAGG)_{n} (n = 114 and 200) obtained from pCR2.1TOPO derivatives (28, 29) were cloned into the vector proximal to the 3'-end of the reporter cassette (Fig. 1, panel B). Both inserts were excised from their derivative plasmids by cleavage with EagI/EcoRV or SpeI/EcoRV and used for cloning in orientations I and II, respectively. The pGFPT vector was digested with EagI/Stul or SpeI/AfIII (the recessed 5' terminus of the latter fragment was filled-in as described above), and was used to clone the repeat sequences in orientations I and II, respectively. Fig. 1, panel B, depicts a list of plasmids that were used in the procaryotic part of the studies. As a control, the pGFPT vector with no repeat sequences was used (10).

Characterization of Cloned Products—The inserts, which were the products of the digestions described above, were purified by native PAGE, ligated with appropriate vectors, and electroporated into E. coli HB101 (31). An aliquot of the cells harboring the pGFPT derivatives were plated on LB agar plates containing 100 μg/ml ampicillin (Ap), whereas transformation mixtures of ligated products with the pEGFP-C1-backbones were spread on kanamycin-containing agar plates (50 μg/ml). In both experiments, single colonies were picked and grown in LB media supplemented with 10% fetal bovine serum (Invitrogen) for 4 days. The episomal DNAs were isolated using alkaline lysis (Promega, Wizard Plus Miniprep DNA purification system). The episomal DNAs were then isolated by alkaline lysis (Promega, Wizard Plus Miniprep DNA purification system). To eliminate plasmids that were not replicated in COS-7 cells, DpnI digestion was performed overnight at 37 °C with 40 units of the enzyme. The plasmids were then transformed into E. coli HB101 and plated on LB plates containing kanamycin (50 μg/ml). The effectiveness of the DpnI cleavage was assessed by digestion of the parental plasmids, which were not introduced into COS-7 cells. As a control cleavage, the plasmids harvested from the CV-1 and HEK-293 cells were subjected to the enzyme. The absence of colonies on LB plates confirmed the complete fragmentation of the plasmids by DpnI. In the experiments conducted in the cells lacking the T antigen, DNA harvested was immediately used for transformation of E. coli HB101.

The plasmids depicted in Fig. 1, panel B were electroporated into the E. coli cells and growth was conducted for five successive recultivations. During the experiments, aliquots of cells were spread on LB-agar plates after the 1st, 3rd, and 5th growth cycles to determine the fluorescent status of single cells. Studies were performed when transcription from the lacZ promoter-operator was stimulated by addition of isopropyl β-D-thiogalactoside (IPTG) into the medium as well as when transcription was silenced by the LacI repressor (34). The details of these studies were described (11).

Determination of DSBs in Mammalian Cells by LM-PCR—Plasmids were isolated from the mammalian cells, phenol-chloroform purified and treated with the E. coli DNA polymerase I Klonef fragment plus dNTPs (New England Biolabs) to fill-in the plasmid ends. Complementary oligonucleotides LM1 (5'-CTCACCAGGAAACGTGTTGAAAGTTAAAGAT-TCT-3') and LM2 (5'-AGCATTTTACTTTGAC-3') were annealed together to prepare a linker that contained one end with a 3'-overhang and one blunt end. The filled-in DNA isolated from mammalian cells and linker were ligated, purified, and used as templates in PCR reactions performed at 95 °C for 30 s (denaturation), 65 °C for 30 s (annealing), and 72 °C for 3 min. 30 s (extension), for 25 cycles. LM3 primer (5'-GTGAAAGTTAAAGTGCT-3') was used to amplify the DNA fragments. The PCR products were separated on 1% agarose gel followed by elution, and re-amplification. The pure homogenous products were then cloned into the pDrive Cloning Vector (Qiagen) and sequenced in order to determine the position of the double strand breaks.

Genetic Rearrangements Monitored by Individual Colony Analyses—The analyses were performed to determine if the non-B structure-forming CTG-CAG, CCTG-CAGG, and GAA-TTC repeat sequences are mutation hot-spots in mammalian and bacterial cells. Additionally, we used plasmids (pRW5615 and pRW5616) containing a CAA-TTC sequence...
Non-B DNA Structures Promote Mutagenesis

that does not form hairpin structures (9, 21, 35). The vectors devoid of any long direct repeats were utilized as controls. Single colony analyses enabled determination of the fraction of mutated clones and facilitated characterization of the individual products of rearrangement events in the cells with various genetic backgrounds.

For both vectors (Fig. 1, panels A and B), the RS were cloned in close proximity to the 3’-end of the reporter cassette, so that mutations affecting the repeats could extend into the flanking regions causing the lack of expression of the reporter genes. We analyzed randomly chosen colonies, because in the mammalian cells transfected with the shuttle plasmids, the FACS analyses did not give satisfactory results on the changes in the EGFP fluorescence level during the cell growth (data not shown), and because of the lack of expression of the eukaryotic GFP gene in E. coli cells that prevented quantitative analysis of fluorescent/white colonies on the agar plates after transformation. Isolation of plasmids from ~100 single colonies was carried out; each experiment was repeated three times. Restriction mapping and sequencing analyses of rearranged clones were performed to determine the locations and types of mutations. The primers used were forward primers (5’-CCCTGAGACAGGAACCC-AAC-3’, 5’-CGTTCGAGCCTACGCTGAT-3’, 5’-ATAAG-GCCGACGGGTC-3’, and 5’-GCTCGCAACCCAAAGTCCG-3’), and reverse primers (5’-CCCTGTAAGCGCGCAATTAG-3’, 5’-GATCTTTGGCAAGCCCTAGGCC-3’, and 5’-CGT-TGAGTCCGTTCTTTAATAG-3’) (Sigma Genosys).

In E. coli, the fluorescent status of the single colonies was determined after spreading of the aliquots of the cell cultures on LB plates containing IPTG. The fraction of non-fluorescent CFUs was calculated (11), and only white colonies were used for further characterization. Restriction mapping of the isolated plasmids confirmed that loss of fluorescence was caused by mutations in the GFP reporter gene. Such clones were characterized in detail by DNA sequence analyses using primers described previously (11). The fraction of the rearrangements caused by the repeat sequences in mammalian cells was calculated by subtracting the background of mutations obtained in E. coli HB101. The bacterial cells were transfected with the RS-containing plasmids and grown on agar plates for 12 h. No rearrangements were observed after individual colonies analyses for all plasmids used. Statistical analyses were performed using SigmaStat version 2.03.

RESULTS

Approach to Study Mutagenesis in Mammalian and Bacterial Cells—We investigated the relationship between the ability of long tracts of the CTG-CAG, CCTG-CAGG, GAA-TTC, and CAA-TTG repeats to form unorthodox DNA conformations and their mutagenic capacity. Previous studies in E. coli showed that the non-B DNA structure-forming CTG-CAG repeats caused gross deletions in sequences flanking the tracts (11). To determine if the secondary structures, which are formed by the repeated DNAs and not the sequences themselves, cause the mutagenesis, we extended our investigations in E. coli by using strains with specific genetic backgrounds that would be expected to be informative regarding DNA conformational properties. Also, we conducted experiments in mammalian cells.

In bacterial cells, plasmids were used that contained the long CCTG-CAGG tract (Fig. 1B) with a well defined capacity to adopt non-B DNA structures (29). This tract was cloned in close proximity to and downstream from the GFP reporter gene; we searched for non-fluorescent mutants during the long-term growth of E. coli cells. Determination of the correlation between the negative superhelical density that influences the stability of secondary structures and mutagenesis in sequences surrounding the repeat tract was achieved by using strains harboring mutations in the topA and gyrB genes (17). Also, the effect of the processing of DNA secondary structures formed by the repeats in relation to mutagenesis was assessed in the sbcC mutant strain and compared with its parental AB1157 background (23).

In mammalian cells, we studied the mutagenic capacity of the unstable long repeats of CTG-CAG, CCTG-CAGG, and GAA-TTC (Fig. 1A) for which the propensities for transient conformational changes in vivo have been described (1, 36). Also, the CAA-TTG sequence that does not form stable hairpin structures (9, 21) was used. The repeats were cloned into the shuttle vector in both orientations relative to the SV40 origin and were positioned downstream of the EGFP reporter gene. Determination of the pathways and the factors involved in mutagenesis in the mammalian cells was performed in three types of cell lines (COS-7, CV-1, and HEK-293). The primate COS-7 cells carried out replication of the SV40 plasmids, whereas the other two cell lines, lacking the T-antigen, were not able to support replication of these plasmids. By construction of two sets of plasmids with either the intact or the truncated CMV promoter, we analyzed the influence of transcription on the mutagenesis in the flanking DNA.

Negative Supercoil Density Modulates Mutagenesis in E. coli—Modulation of the stability of DNA secondary conformations by in vivo changes in the level of negative supercoil density (−σ) and its correlation with genetic instabilities within the long tracts of CTG-CAG, CCTG-CAGG, and GAA-TTC sequences was recently described (17). Herein, we focused our investigations on determining if the level of −σ influences mutagenesis outside the repeats in flanking sequences by using the genetically unstable and structure-forming CCTG-CAGG tract (29). In recultivation experiments, we modulated the in vivo superhelical turns of the RS-containing plasmids by utilization of strains harboring mutations in the topA and gyrB genes. The DM2 repeats expressed a substantial mutagenic potential which depended on the length of the tract, its orientation relative to the origin of replication, transcription status, and the activities of the DNA topology enzymes (Table 1). In the presence of transcription in the topA mutant strain, we observed that an elevation of the level of −σ, which stabilizes non-B DNA conformations (17), caused a ~2–3-fold increase of the fraction of mutants in comparison with the parental ITT1 strain. This effect was observed for both the 114-mer and 200-mer of the repeats and for both orientations. In the SD7 strain, lowering of the negative superhelical density of the plasmids harboring the CCTG-CAGG tract caused the opposite effect and, in the presence of transcrip-
Non-B DNA Structures Promote Mutagenesis

A. Mammalian Cells

Active Transcription

Inactive Transcription

B. E. coli Cells

FIGURE 1. Plasmids used in this study. Panel A shows a list of the plasmids with pEGFP-C1 backbone used in mammalian cells. Panel B depicts the plasmids, which were derivatives of pGFP (10) that were used in E. coli. In both the eukaryotic (A) and prokaryotic (B) studies, the repeat sequences were cloned proximal and downstream of the reporter genes EGFP and GFP, respectively, in both orientations. Orientations I and II were defined under "Experimental Procedures." Panel A, left side shows the constructs with intact CMV promoter, used in the experiments with active transcription, whereas the right side depicts plasmids created after excision of 250 bp of the promoter to turn off transcription from the CMV promoter. In E. coli, the transcription was activated by adding IPTG into the LB media or silenced by co-transformation with plasmid expressing the CCTG-CAGG tract in all E. coli strains used. This effect was previously found also for the long CTG-CAG repeats (11). Determination of the supercoil densities of the plasmids grown in these three strains revealed the expected alterations in −α (data not shown). Note that these studies can only be conducted in E. coli, not in mammalian cells, because of the unavailability of characterized mutant cells (17). Hence, our results show that the in vivo modulation of the DNA secondary structures localized at the CCTG-CAGG tracts influenced the level of mutagenesis both within the repeats and in flanking sequences.

SbcC Repair Protein Affects Mutagenesis in E. coli—Hairpin structures that arise as intermediates of the replication and transcription processes, and impede the progression of the polymerases (9), can be recognized and cleaved in vivo by SbcC and SbcD proteins (25). Herein, we determined if the processing of secondary structures formed at the CCTG-CAGG arrays (29) influenced mutagenesis in sequences flanking the repeats using the parental AB1157 and mutant sbc strains. In the parental strain, we found that the RS-containing plasmids destabilized the repeat tracts and mutagenized the flanking DNA, and in the presence of transcription the fraction of white CFUs varied from 0.01 to 0.25. The deleterious capacity of the CCTG-CAGG tract was more pronounced for the shorter tract (Table 2). In the absence of transcription, we observed an abolition or a dramatic reduction of the fraction of non-fluorescent colonies. Interestingly, no white CFUs were found in the mutant strain for all plasmids analyzed, either in the presence or absence of transcription. Therefore, the presence of the SbcC repair protein that recognized and processed secondary structures (hairpins) formed at the CCTG-CAGG repeats triggered events resulting in the pronounced genetic instability in the RS and the flanking DNA.

Lack of Mutagenic Effect of CAA-TTG Compared with Other Repeat Tracts in COS-7 Cells—The capacity of four repeating triplet and tetranucleotide sequences cloned into pEGFP-C1 to induce mutations in their flanking regions in COS-7 cells was
Non-B DNA Structures Promote Mutagenesis

TABLE 1
Effect of negative superhelical density on mutagenesis in E. coli

| Length of DM2 repeats | E. coli strain | Active transcription orientation | Inactive transcription orientation |
|----------------------|----------------|----------------------------------|----------------------------------|
|                      |                | I                                | II                               | I                                | II                               |
| 114                  | RS2            | 0.002                            | 0.66                             | 0.00                             | 0.001                            |
| (topA)               |                | 18/8,506                         | 6,714/10,205                     | 0/9,678                          | 13/10,100                        |
| JT1                  |                | 9/9,103                          | 1,875/9,447                      | 0/10,100                         | 0/9,388                          |
| (Parental)           |                | 0/9,677                          | 2,111/10,239                     | 0/9,932                          | 0/10,510                        |
| SD7                  |                | 0.20                             | 0.30                             | 0.00                             | 0.00                            |
| (topA gyrb)          |                | 5,021/9,911                      | 8,711/11,007                     | 0/11,021                         | 19/9,105                        |
| 200                  | RS2            | 0.51                             | 0.80                             | 0.00                             | 0.002                            |
| (topA)               |                | 1,958/10,005                     | 2,750/9,350                      | 0/9,987                          | 0/7.765                         |
| JT1                  |                | 0.20                             | 0.30                             | 0.00                             | 0.00                            |
| (Parental)           |                | 357/10,956                       | 1,557/9,642                      | 0/10,761                         | 0/10,611                        |

TABLE 2
Effect of the processing of hairpin structures on mutagenesis in E. coli

| Length of DM2 repeats | E. coli strain | Active transcription orientation | Inactive transcription orientation |
|----------------------|----------------|----------------------------------|----------------------------------|
|                      |                | I                                | II                               | I                                | II                               |
| 114                  | AB1157         | 0.01                             | 0.18                             | 0.00                             | 0.001                            |
| (Parental)           |                | 120/12,112                       | 2,198/12,230                     | 0/9,912                          | 9/8,596                          |
| 114                  | PF2070         | 0.00                             | 0.00                             | 0.00                             | 0.00                            |
| (sbcC)               |                | 0/10,105                         | 0/9,103                         | 0/10,910                         | 0/9,392                         |
| 200                  | AB1157         | 0.25                             | 0.25                             | 0.00                             | 0.002                            |
| (Parental)           |                | 3,236/12,762                     | 2,743/11,008                     | 0/9,200                          | 20/9,754                        |
| 200                  | PF2070         | 0.00                             | 0.00                             | 0.00                             | 0.00                            |
| (sbcC)               |                | 0/9,578                          | 0/9,546                         | 0/8,910                          | 0/10,752                        |

investigated. The ability of long CTG-CAG, CCTG-CAGG, and GAA-TTC repeats to adopt transient conformational changes that block DNA polymerase progression is well documented (1), whereas the repeating CAA-TTG sequence was shown to have no DNA helix-coil transition, consistent with the absence of a slipped structures (21), and no in vitro DNA polymerase pausing was detected at the repeats (9). Spiro et al. (35) showed that the CAA-TTG repeats had a low mutation rate, which was identical in frequency to a random DNA sequence with no secondary structure. These results strongly suggest that CAA-TTG repeats do not form the same types, if any, of secondary structure. These results are strongly indicative that the capacity of a RS to adopt slipped structures with hairpin loops is a requirement for eliciting mutagenesis in sequences flanking the RS.

When the long structure-prone repeats of the DM1, DM2, and FRDA genes were present in plasmids replicated in COS-7 cells, a significant fraction of the recovered molecules had deletions in the regions flanking the repeats (Table 3). On average, the level of mutagenesis induced by all three repeating sequences was ~5–6-fold greater than the control vectors that were devoid of the repeats. By examination of individual colonies, we showed the following: (a) for the CTG-CAG sequence, a higher mutagenic capacity was expressed by the shorter and uninterrupted tract containing 98 repeats than for the 175 mer, and the mutagenic potential was orientation and transcription-dependent; the interrupted 175 mer showed a significantly higher (p < 0.001) fraction of mutants when transcription was active, but the orientation effect was detected only in the absence of transcription; (b) for the DM2 sequence, a significantly higher (p < 0.001) fraction of mutants was obtained for the 200-mer than for 114-mer; the mutagenic effect was elevated in orientation II (when the CAGG repeats were present on the leading strand template) and in the presence of transcription for both lengths used (Table 3); (c) the (GAA-TTC)60 transfection of primate cells were grown for 4 days, plasmids were isolated and introduced into E. coli HB101. The analyses of the DNAs obtained from single colonies which were picked at random since no screen was available enabled the characterization of individual events that occurred in the COS-7 cells and the determination of the fraction of the rearranged clones.

Table 3 shows that the cloned long CAA-TTG sequences exhibited only a background level of mutagenesis. The fraction of mutants found for the interrupted tract of the 99 mer approximated the level detected for the control vector which was void of the long direct repeats. Similarly, the shorter and uninterrupted 66 mer did not give a significantly higher fraction of mutants in comparison with the control molecule. Hence, these results agree with the data described above indicating that the capacity of a RS to adopt slipped structures with hairpin loops is a requirement for eliciting mutagenesis in sequences flanking the RS.
tract showed a higher level of mutagenesis in comparison with the 150-mer (p < 0.001) in the presence and absence of transcription; an orientation effect (with orientation II being more mutagenic) was observed for both lengths used only when transcription was active (p < 0.001) (Table 3).

Hence, these results show that only non-B DNA structure-forming long CTG-CAG, CCTG-CAGG, and GAA-TTC repeats were effective in elevating the mutagenesis in DNA flanking regions over the background of the control DNAs. Also, the orientation of the repeats relative to the origin of replication and transcription played crucial roles in influencing the mutagenic capacity.

Effect of Inactivation of Replication on the Mutagenesis in Mammalian Cells—CTG-CAG expansions did not occur in human DM1 fibroblasts in the absence of cellular proliferation (37). Because the types of mutations induced in non-dividing cells can be different from those induced in dividing cells, we analyzed the mutation spectrum of the CTG-CAG, CCTG-CAGG, and GAA-TTC tracts after inactivation of replication. Primate CV-1 and human HEK-293 cells were used that did not express the T-antigen and could not support SV40 replication of the shuttle plasmids.

A set of molecules with the repeats cloned in orientation II (Fig. 1A) was chosen for the experiments conducted as described for COS-7 cells ("Experimental Procedures"). Individual colony analyses showed that the fraction of clones with large scale deletions encompassing the regions surrounding the RS was ~5-fold lower in CV-1 and HEK-293 cells than in COS-7. The average fraction of mutants for non-replicated plasmids was 5.9 × 10^-5 whereas those propagated in COS-7 reached a level of 27.2 × 10^-5 (compare Tables 3 and 4). This result demonstrates that the lack of replication, a process that can facilitate formation of non-B DNA conformations and DSB inductions, caused a significant impairment of the mutagenic potential of the DM1, DM2, and FRDA repeats.

Table 4 shows the results of single colony studies of plasmids isolated from the CV-1 and HEK-293 cells. Differences (p < 0.001) were found on comparing the level of mutagenesis between the mammalian cell lines for all three sequences analyzed. The stimulatory effect of transcription on the elevated fraction of mutants was observed. In CV-1 cells, significant differences (p < 0.001) were obtained when compared with the fraction of mutants under the conditions of active and inactive transcription for all sequences analyzed. In the HEK-293 cell line, a similar effect was observed for the (CTG-CAG)_90, (CCTG-CAGG)_200, and (GAA-TTC)_200 repeats. However, there was no influence of transcription on the level of mutagenesis caused by (GAA-TTC)_90 (Table 4); this effect may be caused by the different types of non-B DNA conformations adopted by these sequences (1, 4, 16, 29, 38).

Thus, in the absence of SV40 replication, a higher mutagenic capacity of the long repeats was found when transcription was active. This behavior was observed also during active replication of the plasmids in COS-7 cells.

Classification of Mutation Products Generated in the Presence of the Long Direct Repeats—Characterization of products of individual rearrangement events by restriction mapping and DNA sequencing analyses showed that the RS-containing plasmids iso-

### Table 3

| Repeat sequence | Length | Active transcription orientation | Inactive transcription orientation |
|-----------------|--------|---------------------------------|----------------------------------|
|                 |        | I                               | II                               |
|                 |        | Fraction of deleted mutants     |                                  |
|                  |        |                                |                                  |
| No repeats      | 0      | 0.05 (14/300)                  | ND*                             |
| CAA-TTG         | 66     | 0.03 (9/300)                   | 0.03 (8/300)                    |
| CTG-CAG         | 98     | 0.18 (53/297)                  | 0.3 (92/302)                    |
| CCTG-CAGG       | 114    | 0.1 (29/300)                   | 0.22 (66/300)                   |
| CCTG-CAGG       | 200    | 0.32 (97/300)                  | 0.39 (147/300)                 |
| GAA-TTC         | 60     | 0.12 (35/300)                  | 0.4 (123/305)                   |
| GAA-TTC         | 150    | 0.07 (22/300)                  | 0.28 (84/300)                   |

* ND, experiments were not performed.

### Table 4

| Repeat sequence | Length | CV-1                     | HEK-293                    |
|-----------------|--------|--------------------------|----------------------------|
|                 |        | Active transcription     | Inactive transcription     |
|                 |        | orientation              | orientation              |
| No repeats      | 0      | 0.01 (3/299)             | 0.01 (3/300)              |
| CAA-TTG         | 66     | 0.07 (20/300)            | 0.05 (10/298)             |
| CTG-CAG         | 98     | 0.11 (32/300)            | 0.04 (44/299)             |
| CCAA-TTC        | 60     | 0.07 (21/299)            | 0.04 (14/300)             |

The plasmids containing (CTG-CAG)_90, (CCTG-CAGG)_200, and (GAA-TTC)_200, repeats cloned in orientation II (Fig. 1A) were introduced into the CV-1 and HEK-293 fibroblast-like cells that did not express the large Tumor antigen and could not support replication of the SV40 plasmids. The growth was conducted as described ("Experimental Procedures"). The fraction of rearranged DNA was determined after transformation of E. coli HB101 with episomal DNAs isolated 4 days after transfection. The experiments were performed when transcription from the CMV promoter was active or was silenced (by removal of part of the promoter). The data shown for each length of the repeats represents the combined results of three separate experiments where randomly chosen colonies were characterized. The fraction of rearranged mutants (bold font) was calculated and is shown as the ratio of the number of deleted clones to the total number of isolated clones for each plasmid.
lated from *E. coli*, and mammalian cells were either full-length (starting plasmids) or mutated molecules. Instabilities (deletions and expansions) were found within the RS and either in the upstream or downstream flanking regions (Fig. 2) in a significant fraction of the rearranged clones (Tables 1–4). Whereas bacterial mutants contained only simple large deletions, mammalian mutants also had complex rearrangements, which included deletions, inversions, insertions, duplications, and point mutations.

In *E. coli*, deletions spanned from 1.1 to 2.0 kbp of the plasmids; exemplary data is shown in Fig. 2A. DNA was sequenced from 23 non-fluorescent colonies; in 80% of the cases identical mutations were found in two or more individual clones (e.g. clones 28–30). In the case of DM2 sequence, mutations were found in both upstream and downstream DNA flanking the CCTG–CAGG tract. In 5/23 (22%) of the clones, the entire DM2 tract was deleted, whereas 88% of mutants had from 3 to 120 repeats remaining. Healed junctions in the upstream flanking region in 61% of clones were found at or in the vicinity of the transcription terminator cassette, but 39% (9/23) of the mutants had the junctions within the GFP reporter gene (e.g. clones 32 and 33) (Fig. 2A). The location of the second mutation junction was within the CCTG–CAGG tract (e.g. clones 28–30, 35) or downstream of it (clones 26, 27, 31). In these cells, 39% of the clones had 1–5 nt of homology at the repaired junctions of the healed ends; the remaining 61% did not contain any microhomology. These data differ somewhat from our previous results (11), where all CTG–CAG mutants harbored homologous nucleotides at the healed junctions.

In mammalian cells, among the 61 clones that were sequenced, 48 (79%) underwent large scale deletions and 13 (21%) were products of more complicated and multistep rearrangements. The latter events were very uncommon (2/15) for the control molecules with no long
repeats and for the unreplicated RS-harboring plasmids. For the RS-containing plasmids, the simple deletions spanned from 1.0 to 3.2 kbp in length, and encompassed the region between the pUC and SV40 origins (Fig. 2B). One healed junction present in the majority (94% (45/48)) of the clones was located upstream of the RS and mapped within the CMV promoter (e.g., clones 1–4, 6) or upstream at the bacterial origin (clones 8, 11, 12, and 13). The remaining 6% of clones had their junctions within the EGFP gene (clone 5). The second healed termini (23%) were localized within the RS (e.g., clones 1, 5, 9, and 13) or downstream of this region. 27% (13/48) of the mutants had a second healed junction that mapped at the SV40 origin (clones 3, 4, 10, 14, and 16). These molecules were able to replicate in COS-7 cells since the core of the origin remained intact (data not shown).

Complex rearrangement products underwent large scale single or double deletions affecting from 1.8 to more than 3.1 kbp of the DNA; additionally, inversions of 34 to ~200 bp as well as short duplications of several bp were found (Fig. 3). In 7 (54%) cases out of 13 analyzed, we found healed junctions from deletions and/or inversions at one of the mutation hot-spot regions of the CMV promoter or the SV40 origin (e.g., clones 19, 20 and 24). For the clone 21, two independent inversion events were found in addition to a 2.9-kbp deletion and short (3 and 4 bp) duplications. The terminus sequence of the pUC ori proximal to the CMV promoter was mutated in three (23%) of the sequenced clones. Deletions and inversions were found which affected that region also (clone 25). In three mutants (e.g., clones 22 and 23), the DNA immediately flanking the repeat array also underwent inversions in addition to deletions. Identical locations, to the bp, of the junctions of the healed ends (clones 22 and 23) were found in 2/61 of analyzed clones.

FIGURE 3. Examples of schematic representations of the healed junctions of clones with complex rearrangements. Panels A–D depict mutants obtained in COS-7 cells, whereas E and F, respectively, were from CV-1 and HEK-293. Panel A presents clone 19, the product of rearrangements of pRW5327; panel B shows clone 20, a derivative of pRW5335; panel C depicts clone 21, which was a mutant of pRW5332; panel D represents rearrangement products of clones 22 and 23, derivatives of pRW5333 and pRW5326, respectively; panels E and F show clones 24 and 25, derivatives of pRW5611 and pRW5606, respectively. For all clones, the white (open) bars indicate the deleted regions; filled bars, remaining segments of the plasmids; striped bars, inverted sequences; blue bars, present duplications. The numbers shown on the top of each segment depict the lengths of the rearrangements (bp), and horizontal lines indicate the location of the mutations. The homologous sequences at the healed junctions are shown in red font, whereas the surrounding nt that were deleted and retained are in black font as lowercase and uppercase letters, respectively.
Non-B DNA Structures Promote Mutagenesis

Healed junctions of products of the simple deletions showed the presence of microhomologies of 1–6 nt in 80% of the cases (Fig. 2B), whereas all but one of the complex rearrangement mutants had up to 13-nt homologies at the junctions (Fig. 3). Hence, the sequence homology at the junctions suggested that non-B DNA conformation-mediated DSBs could be repaired by NHEJ and/or SSA pathways (39).

In summary, these data show that the products of mutagenesis caused by the long CTG-CAG, CCTG-CAGG, GAA-TTC, and CAA-TTG repeats were both simple and complex rearrangements. Also, deletions of the tracts extended into the downstream and upstream non-repeating flanking regions.

DNA Healed Junctions Correlate with Regions Abundant in Perfect Direct, Inverted, and Mirror Repeat Sequences—The healed junctions of the mutants were accumulated predominantly in three hot-spot regions (Fig. 4). The areas comprised the long repeat tracts of the DM1, DM2, and FRDA genes and segments of DNA located either upstream or downstream of the RS. The DNA sequences found within the hot-spots have potential to fold into non-B DNA conformations because of their composition, length, and repetitive character (1, 3, 4, 40). Hot-spot A encompassed the CMV promoter and the upstream region that included the pUC origin. This mutation hot-spot was rich in a variety of long (up to 21 bp) direct (ACGGTAA, AATGGCCCGGCTGGC, CCCCAATTGAGTCTAAT, GGCCTTT, TTGCGAGTACATCAA), inverted (TTGCTGGCCTGGC, CCCCATTGACGTCAATG, AAAACGCCAGCAA), and mirror (ATTTTT, CCTTT, CCCCAGC, TTGGC) repeats that were present in variable numbers of copies. Hot-spot B included the long tracts of CTG-CAG, CCTG-CAGG, or GAA-TTC direct repeats. The tracts were flanked by inverted (GAAATCGGCCCTT) and direct (GGATCCAC, ACTCTCCC, GAAAT) repeats that included the mutation endpoints in several of the rearranged clones. The third hot-spot C region was comprised of the SV40 origin of replication and contained two copies of a 72 bp of direct repeat, three copies of a 21-bp direct repeat tract, and 11 bp of an inverted sequence that was duplicated. DSBs and healed junctions at hot-spot C were found mainly for mutants that underwent replication in COS-7 cells.

For the non-replicated RS-containing plasmids and for the control molecules (Fig. 1A), we found the accumulation of healed junctions mostly within hot-spot regions A and B (Fig. 4).

Therefore, in mammalian cells, the termini of gross rearrangements were mapped to the repeat-rich areas, where direct, inverted, and mirror repetitive sequences were overrepresented. Such regions of specific compositions are known to be predisposed to DNA conformational alterations, resulting in the stimulation of mutagenesis. These features were also found in E. coli (10, 11), strongly implying a direct mutagenic role of non-B DNA conformations in both prokaryotic and eukaryotic types of cells. This mutagenic capacity is likely driven by an increased susceptibility to DSB formation within the repetitive sequences and their subsequent error-prone repair (10, 13, 16).

Non-B DNA Structure-forming Direct Repeats Induced DSBs in Mammalian Cells—Plasmids were isolated from the mammalian cells (COS-7, CV-1, or HEK-293) 4 days after transfection and were subjected to LM-PCR to determine the presence of DSBs. Results of the reactions conducted with the plasmids harboring the RS revealed distinct LM-PCR products corresponding to the broken DNA molecules (Fig. 5). No products of this type were obtained for the CCTG-CAGG-containing plasmid that was not introduced into the mammalian cells (Fig. 5, compare lanes 1–5 with the control lane 6). The DSBs were mapped within distinct regions of the plasmids by sequencing the LM-PCR products (Fig. 4, pink bars). The long tri- and tetranucleotide repeat sequences and their immediate adjacent regions were hot-spots for generating the DSBs mostly for the plasmids that underwent replication in COS-7 cells. Additionally, the CMV promoter and the neighboring upstream DNA as well as the SV40 origin had accumulated DSBs. Interestingly, we detected a hot-spot region for the RS-induced DSBs at the sequence of the bacterial origin that was distal to the CMV promoter (Fig. 4, pink bars). Deletions of that segment removed virtually the entire sequence of the pUC ori, therefore preventing replication of these molecules in E. coli.

Thus, these results show that the rearrangements observed in the presence of the long CTG-CAG, CCTG-CAGG, or
Non-B DNA Structures Promote Mutagenesis

The molecular mechanisms responsible for the genetic instabilities of repeating tri- and tetranucleotide DNA sequences related to several human hereditary neurological diseases have been investigated (reviewed in Refs. 1, 3, 4, 16, 36, 40–43). Virtually all studies have postulated roles for non-B DNA conformations in these instabilities; however, the work described herein may be among the first to directly address the relationship between the non-B DNA structures of the repeats (slipped structures with hairpin loops, cruciforms, triplexes, sticky DNA, etc.) and their mutagenic potential. Three different strategies were utilized for this investigation: first, we studied the mutagenic potential of the repeat sequences (CTG-CAG, CCTG-CAGG, and GAA-TTC), which have repeatedly been postulated to adopt non-B DNA structures in mammalian cells (1, 5, 29, 40, 44–46), as well as control plasmids and the CAA-TTG-containing molecules, which are inert in their capacity to form unorthodox DNA conformations (9, 21, 22, 35). Only the repeating sequences which can fold into non-B DNA structures stimulated the formation of rearrangements at regions flanking the repeat sequences in a substantial fraction of the molecules. This effect was replication-dependent and was more pronounced when transcription was active through the RS. Second, we studied the roles of topoisomerase I and gyrase B in E. coli mutant cells to elicit different levels of intracellular negative supercoil density; these analyses revealed that high negative supercoil density enhances the mutagenic potential of the CCTG-CAGG repeats. Again, active transcription was a principal factor which influenced mutagenesis. The role of in vivo negative supercoil density in stabilizing non-B DNA structures, which thus modulates the genetic instabilities within the RS, is well known but no prior investigations were conducted on the effects in sequences flanking the RS (1, 17). Third, we evaluated the involvement of the SbcC protein on the formation of gross deletions in E. coli. The inactivation of the SbcC nuclease that recognizes and cleaves hairpin conformations abolished the mutagenesis in the DNA flanking the DM2 repeat tract. Thus, all three strategies strongly support the concept of specific repeating DNA sequences that fold into quasi-stable non-B DNA conformations and influence the instabilities, thus, causing mutagenesis in the mammalian and E. coli cells.

Genomic rearrangements have been widely studied and are a common outcome of instabilities (1, 10, 11, 16), but the molecular mechanisms of the rearrangements, which are fundamental for understanding a family of human diseases (reviewed in Ref. 16), are poorly understood. Genomic rearrangements include gross deletions, insertions, inversions, duplications, point mutations, and related genetic events. Recent studies (10–13, 16) in prokaryotic and eukaryotic model systems documented that DNA regions in the vicinity of specific non-B DNA-forming sequences were affected by mutations, and error-prone repair of DSBs formed at and around such conformations was postulated. Non-B DNA structures, such as triplexes or left-handed Z-DNA, which are potentially very common in the human genome (1, 12, 16, 47–49), are highly mutagenic in mammalian cells (12, 13). Also, introduction of a 2.5-kb poly(RY) tract (10), which forms a variety of non-B DNA structures (20, 50, 51) mutagenized flanking DNA regions. Furthermore, prior investigations (14, 15) in mammalian cells showed instabilities in sequences that flank long CTG-CAG tracts as a result of DSBs and recombination repair. However, related studies in E. coli (11) revealed the mutagenic potential of the DMI-repeating sequence as linked to its capacity to adopt non-B DNA structures.

We postulate that the transient formation of unorthodox DNA conformations, which were adopted by certain types of repeating sequences (CTG-CAG, CCTG-CAGG, and GAA-TTC) in underwound negatively supercoiled plasmids caused the instabilities within the repeats (reviewed in Refs. 1 and 36) and also played important roles in the stimulation of the rearrangements in DNA flanking sequences outside the repeat tracts. In vitro studies (9) revealed no blockage of DNA polymerases at the repeating tract of CAA-TTG, and no supercoil-dependent structural transitions on two-dimensional gels; these results are consistent with lack of non-B DNA structures adopted by the CAA-TTG tract. Other prior investigations (9, 21, 22) revealed that long CAA-TTG sequences were inert in adopting slipped structures with hairpin loops. Interestingly, this sequence lacking the propensity for hairpin structure formation had an expansion rate identical to the frequency found for random, and non-B DNA structure-adopting sequences (22, 35). Hence, we conclude that repeating sequences that can adopt non-B DNA structures are mutagenic, whereas the CAA-TTG tract, which does not fold into unorthodox conformations, is not.

Two other experimental strategies were utilized to evaluate the role of non-B DNA structures in mutagenesis. First, studies with the mutant SbcC E. coli strain, which lacks the capacity to recognize and cleave hairpin structures, revealed the pronounced effect of this pathway on the mutagenic capacity of CCTG-CAGG repeat sequences. No white CFUs were found in

GAA-TTC repeats were associated with error-prone repair of DSBs. The breaks were detected in sequences of the repeats (hot-spots A-C) that adopt non-B DNA structures.

**DISCUSSION**

The plasmids depicted in Fig. 1 were isolated from the COS-7, CV-1, and HEK-293 cells and subjected to LM-PCR to detect the presence of DSBs ("Experimental Procedures"). Lanes 1–4 show LM-PCR products of pRW5257, pRW5330, pRW5608, and pRW5246, respectively, obtained from the COS-7 cells; lane 5 corresponds to the LM-PCR products of pRW5253 isolated from CV-1 cells. As a control, pRW5327 not introduced into the mammalian cells was subjected to LM-PCR (lane 6).
Non-B DNA Structures Promote Mutagenesis

the mutant strain for all plasmids analyzed either in the presence or absence of transcription. The formation of hairpin loops is one of the consequences of slipped structures as well as cruciforms. Second, the role of negative supercoil density in stabilizing non-B DNA structures is well known (17) and, thus, was investigated as a modulator of mutagenesis. In addition to global supercoil density, local transient supercoiling induced by transcription may influence the stability of certain non-B DNA conformations, and has recently been found to affect the genetic instability within the CTG-CAG, CGG-GCC, and GAA-TTC repeats (17). The likelihood of formation of underwound unorthodox DNA structures that increases with the extent of negative supercoiling was also shown to interfere with the viability of E. coli containing a plasmid harboring a 2.5-kb poly(R-Y) tract (20). Herein, the mutagenic potential of the long tract of CCTG-CAGG repeats in a plasmid was sensitive to changes in the level of negative supercoil density, and was concomitantly affected by both the increase of supercoiling and the stimulation of transcription. Therefore, in vivo changes in DNA topology by either global or local modulation of supercoil density influenced the structure-induced DNA mutagenesis, as predicted. In summary, all three experimental strategies are consistent with the interpretation that non-B DNA structures are potent stimulators of mutagenesis.

Replication and transcription that create local changes in DNA topology, and facilitate the formation of structural obstacles for the progression of DNA and RNA polymerases (7, 36) were key factors in influencing the formation of the gross rearrangements in mammalian and E. coli cells. Alleviation of the arrested polymerase machineries often results in the formation of DSBs (16, 42), and DNA sequences that undergo transient conformational changes from the right-handed B form to non-B DNA structures such as cruciforms, slipped structures with loops, triplexes, sticky DNA, etc., have been shown to be the preferred sites for DSBs (10–13,16). In fact, the DSBs can be mapped in some cases to the thermodynamically strained regions of these unorthodox conformations (3, 10). The repair of these breaks is facilitated by terminal homologies of several bp (10, 11, 13, 52), suggesting that error-prone NHEJ and/or SSA repair pathways may be involved in the processing of the broken DNA (39). Hence, the formation of quasi-stable non-B DNA structures at specific repeat sequences may trigger a series of subsequent events resulting in rearrangements.

In some rare cases of FRAXA-affected humans (53, 54), instabilities of the CGG-CCG repeats that include full mutations along with deletions of DNA regions both upstream and downstream of the repeats were found. Hitherto, to our knowledge, no similar deletion events in the sequences flanking the expanded repeats of the DM1, DM2, and FRDA genes have been detected. However, because of our results on the mutagenic potential of the RS in mammalian and bacterial cells, it may be important to analyze the DNA regions flanking the RS in humans. Future studies may provide significant information on the role of these genomic regions in the disease etiologies.

Acknowledgments—We thank Dr. Albino Bacolla for constructive suggestions on the manuscript.

REFERENCES

1. Wells, R. D., and Ashizawa, T. (2006) Genetic Instabilities and Neurological Diseases, 2nd Ed., Elsevier-Academic Press, San Diego, CA
2. Sinden, R. R. (1994) DNA Structure and Function, Academic Press, San Diego, CA.
3. Wells, R. D., Dere, R., Hebert, M. L., Napierala, M., and Son, L. S. (2005) Nucleic Acids Res. 33, 3785–3798
4. Bowater, R. P., and Wells, R. D. (2001) Prog. Nucleic Acids Res. Mol. Biol. 66, 159–202
5. Wells, R. D., and Harvey, S. C. (1988) Unusual DNA Structures, Springer-Verlag, New York
6. Samadashwily, G. M., Racu, G., and Mirkin, S. M. (1997) Nat. Genet. 17, 298–304
7. Mirkin, E. V., and Mirkin, S. M. (2005) Mol. Cell Biol. 25, 888–895
8. Mochmann, L. H., and Wells, R. D. (2004) Nucleic Acids Res. 32, 4469–4479
9. Ohshima, K., Kang, S., Larson, J. E., and Wells, R. D. (1996) J. Biol. Chem. 271, 16773–16783
10. Bacolla, A., Jaworski, A., Larson, J. E., Jakupciak, J. P., Chuzhanova, N., Abeyesinghe, S. S., O’Connell, C. D., Cooper, D. N., and Wells, R. D. (2004) Proc. Natl. Acad. Sci. U. S. A. 101, 14162–14167
11. Wojciechowska, M., Bacolla, A., Larson, J. E., and Wells, R. D. (2005) J. Biol. Chem. 280, 941–952
12. Wang, G., Christensen, L. A., and Vasquez, K. M. (2004) Proc. Natl. Acad. Sci. U. S. A. 101, 2677–2682
13. Wang, G., and Vasquez, K. M. (2004) Proc. Natl. Acad. Sci. U. S. A. 101, 13448–13453
14. Marcardier, J. L., and Pearson, C. F. (2003) J. Biol. Chem. 278, 33848–33856
15. Meservy, J. M., Sargent, R. G., Iyer, R. R., Chan, F., McKenzie, G. J., Wells, R. D., and Wilson, J. H. (2003) Mol. Cell Biol. 23, 3152–3162
16. Bacolla, A., and Wells, R. D. (2004) J. Biol. Chem. 279, 47411–47414
17. Napierala, M., Bacolla, A., and Wells, R. D. (2005) J. Biol. Chem. 280, 37366–37376
18. Liu, L. F., and Wang, J. C. (1987) Proc. Natl. Acad. Sci. U. S. A. 84, 7024–7027
19. Rahmouni, A. R., and Wells, R. D. (1989) Science 246, 358–363
20. Bacolla, A., Jaworski, A., Connors, T. D., and Wells, R. D. (2001) J. Biol. Chem. 276, 18597–18604
21. Richard, G. F., Goellner, P. W., Liu, C., Arnheim, N., and Petes, T. D. (1999) Proc. Natl. Acad. Sci. U. S. A. 96, 1504–1509
22. Moore, H., Greenwell, P. W., Liu, C., Arnheim, N., and Petes, T. D. (1999) Proc. Natl. Acad. Sci. U. S. A. 96, 1504–1509
23. Oussatcheva, E. A., Hashem, V. I., Zou, Y., Sinden, R. R., and Potaman, V. N. (2001) J. Biol. Chem. 276, 30878–30884
24. Connelly, J. C., Kirkham, L. A., and Leach, D. R. F. (1996) Proc. Natl. Acad. Sci. U. S. A. 93, 7979–7984
25. Connelly, J. C., de Leau, E. S., and Leach, D. R. (1999) Nucleic Acids Res. 27, 1039–1046
26. Napierala, M., Parniewski, P. P., Pluciennik, A., and Wells, R. D. (2002) J. Biol. Chem. 277, 34087–34100
27. Wang, G., Jaworski, A., Ohshima, K., and Wells, R. D. (1995) Nat. Genet. 10, 213–218
28. Liquori, C. L., Ricker, K., Moseley, M. L., Jacobsen, J. F., Kress, W., and Naylor, S. L., Day, J. W., and Ranum, L. P. (2001) Science 293, 864 – 867
29. Dere, R., Napierala, M., Ranum, L. P., and Wells, R. D. (2004) J. Biol. Chem. 279, 41715–41726
30. Ohshima, K., Montermini, L., Wells, R. D., and Pandolfo, M. (1998) J. Biol. Chem. 273, 14588–14595
31. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) Molecular Cloning A Laboratory Manual, 2nd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
32. Sobczak, K., de Mezer, M., Michlewski, G., Krol, J., and Krzyzosiak, W. J. (2003) Nucleic Acids Res. 31, 5469–5482
33. Vetcher, A. A., Napierala, M., Iyer, R. R., Chastain, P. D., Griffith, J. D., and Wells, R. D. (2002) J. Biol. Chem. 277, 39217–39227
34. Bowater, R. P., Shen, D., and Lilley, D. M. (1994) Biochemistry 33, 9266–9275
35. Spiro, C., Pelletier, R., Rolfsmeier, M. L., Dixon, M. J., Lahue, R. S., Gupta, G., Park, M. S., Chen, X., Mariappan, S. V., and McMurray, C. T. (1999) Mol. Cell 4, 1079–1085
36. Wells, R. D., and Warren, S. T. (eds) (1998) Genetic Instabilities and Hereditary Neurological Diseases, Academic Press, San Diego, CA
37. Yang, Z., Lau, R., Marcadier, J. L., Chitayat, D., and Pearson, C. E. (2003) Am. J. Hum. Genet. 73, 1092–1105
38. Vetcher, A. A., Napierala, M., and Wells, R. D. (2002) J. Biol. Chem. 277, 39228–39234
39. Bzymek, M., and Lovett, S. T. (2001) Proc. Natl. Acad. Sci. U. S. A. 98, 8319–8325
40. Sinden, R. R., Pearson, C. E., Potaman, V. N., and Ussery, D. W. (1998) in Genes and Genomes (Verma, R. S., ed) Vol. 5A, 1st Ed., pp. 1–141, JAI Press Inc., Greenwich, CT
41. Sinden, R. R., Potaman, V. N., Oussatcheva, E. A., Pearson, C. E., Lyubchenko, Y. L., and Shlyakhtenko, L. S. (2002) J. Biosci. 27, 53–65
42. Wang, G., and Vasquez, K. M. (2006) Mutat. Res., in press
43. Pearson, C. E., Edamura, K. N., and Cleary, J. D. (2005) Nat. Rev. Genet. 6, 729–742
44. Wells, R. D. (1988) J. Biol. Chem. 263, 1095–1098
45. Sinden, R. R. (1999) Am. J. Hum. Genet. 64, 346–353
46. Sinden, R. R., and Wells, R. D. (1992) Curr. Opin. Biotechnol. 3, 612–622
47. Schroth, G. P., and Ho, P. S. (1995) Nucleic Acids Res. 23, 1977–1983
48. Feigon, J., Wang, A. H., van der Marel, G. A., van Boom, J. H., and Rich, A. (1985) Science 230, 82–84
49. Bacolla, A., Collins J. R., Gold, B., Chuzhanova, N., Yi, M., Stephens, R. M., Stefanov, S., Olsh, A., Jakupciak, J. P., Dean, M., Lempicki, R. A., Cooper, D. N., and Wells, R. D. (2006) Nucleic Acids Res. 34, 2663–2675
50. Van Raay, T. J., Burn, T. C., Connors, T. D., Petri, L. R., Germino, G. G., Klinger, K. W., and Landes, G. M. (1996) Microb. Comp. Genom. 1, 317–327
51. Blaszak, R. T., Potaman, V., Sinden, R. R., and Bissler, J. J. (1999) Nucleic Acids Res. 27, 2610–2617
52. Sargent, R. G., Brenneman, M. A., and Wilson, J. H. (1997) Mol. Cell Biol. 17, 267–277
53. Nichol Edamura, K., and Pearson, C. E. (2005) Hum. Genet. 118, 301–304
54. Fan, H., Booker, J. K., McCandless, S. E., Shashi, V., Fleming, A., and Farber, R. A. (2005) Am. J. Med. Genet. A 136, 214–217