The Myotonic Dystrophy Type 1 Triplet Repeat Sequence Induces Gross Deletions and Inversions*

The capacity of (CTG-CAG)\textsubscript{n} and (GAA-TTC)\textsubscript{n} repeat tracts in plasmids to induce mutations in DNA flanking regions was evaluated in *Escherichia coli*. Long repeats of these sequences are involved in the etiology of myotonic dystrophy type 1 and Friedreich's ataxia, respectively. Long (CTG-CAG)\textsubscript{n} (where \( n = 98 \) and 175) caused the deletion of most, or all, of the repeats and the flanking GFP gene. Deletions of 0.6–1.8 kbp were found as well as inversions. Shorter repeat tracts (where \( n = 0 \) or 17) were essentially inert, as observed for the (GAA-TTC)\textsubscript{176}-containing plasmid. The orientation of the triplet repeat sequence (TRS) relative to the unidirectional origin of replication had a pronounced effect, signaling the participation of replication and/or repair systems. Also, when the TRS was transcribed, the level of deletions was greatly elevated. Under certain conditions, 30–50% of the products contained gross deletions.

DNA sequence analyses of the breakpoint junctions in 47 deletions revealed the presence of 1–8-bp direct or inverted homologies in all cases. Also, the presence of non-B folded conformations (i.e. slipped structures, cruciforms, or triplexes) at or near the breakpoints was predicted in all cases. This genetic behavior, which was previously unrecognized for a TRS, may provide the basis for a new type of instability of the myotonic dystrophy protein kinase (*DMPK*) gene in patients with a full mutation.

Myotonic dystrophy type 1 (DM1) is an autosomal dominant neuromuscular disease that exhibits a high incidence (1:8000) and shows frequent mortality in affected infants (1). An unstable region on chromosome 19q13.3 was discovered as the genetic basis of DM1. A polymorphic locus was found to be larger in DM1 patients (1–3), because of substantial expansions of a CTG-CAG repeat tract in the 3'-untranslated region of the myotonic dystrophy protein kinase (*DMPK*) gene (1). As many as 3000 repeats (9000 bp) have been found in some patients, expanded from the normal range of 5–37 repeats.

DM1 displays a non-Mendelian inheritance pattern. The molecular mechanisms responsible for this genetic instability have been extensively investigated in recent years in bacteria, yeast, cell culture, and mouse systems (reviewed in Refs. 1–4). DNA replication (5–9), repair (10–13), and recombination (14–16) are involved, probably acting in concert with other factors/processes, such as single-strand DNA-binding proteins (17) and transcription (9, 18). Also, the long CTG-CAG repeat tract can adopt an unusual flexible and writhed conformation (19), which may promote the formation of slipped structures (9, 20, 21) with a transiently formed, quasi-stable, long CTG sequence along with an unpaired and unstacked long CAG complementary strand. These types of preferential single-strand stabilities and DNA conformational behaviors are integral to the interpretation of the genetic instability effects of TRS orientation relative to the direction of DNA replication (1–4, 9).

A 2.5-kbp poly(purine-pyrimidine) tract from the human polycystic kidney disease gene 1 (*PKD1*), which is known to form triplexes, slipped structures, and other non-B DNA conformations (22), induced long deletions and other instabilities in plasmids that were manifested by mismatch repair and, in some cases, transcription. The breakpoints occurred at or near the predicted non-B DNA conformations. Distance measurements also indicated a significant proximity of alternating purine-pyrimidine and oligo(purine-pyrimidine) tracts to breakpoint junctions in 222 gross deletions and translocations, respectively, involved in human diseases. In 11 of these deletions, which were analyzed in detail, the breakpoints were explicable by non-B DNA structure formation. Hence, Bacolla et al. (21) concluded that alternative DNA conformations trigger genomic rearrangements through recombination-repair activities. Also, substantial literature is growing on the role of non-B DNA conformations involving low copy repeats in genomic rearrangements (deletions, inversions, duplications, translocations, etc.) associated with human diseases (reviewed in Ref. 23).

Here we have substantially extended these studies by exploring the capacity of CTG-CAG and GAA-TTC repeat tracts of various lengths, extents of interruptions (polymorphisms), and orientations to serve as mutagens, as a function of transcription. Most surprisingly, long CTG-CAG repeat tracts promoted the formation of inversions and long deletions (0.6–1.8 kbp) that removed part or all of the repeats as well as the flanking GFP reporter gene. This behavior, if found in humans, implies that the *DMPK* gene flanking the long CTG-CAG repeats in patients may be subject to deletions and rearrangements, thus stabilizing the DMPK protein.

**EXPERIMENTAL PROCEDURES**

*Escherichia coli* Strains—Strains JTT1 and AB1157 were obtained from the *E. coli* Genetic Stock Center, Yale University, New Haven, CT.

* This work was supported by National Institutes of Health Grants NS37554 and ES11847, the Robert A. Welch Foundation, the Muscular Dystrophy Association, Friedreich’s Ataxia Research Alliance, and the Seek-a-Miracle Foundation. 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 1784 solely to indicate this fact.

‡ To whom correspondence should be addressed: Center for Genome Research, Institute of Biosciences and Technology, Texas A & M University System Health Science Center, Texas Medical Center, Houston, Texas 77030.

1 The abbreviations used are: DM1, myotonic dystrophy type 1; TRS, triplet repeat sequence; CFUs, colony-forming units; IPTG, isopropyl \( \beta \)-thiogalactoside; DSB, double-strand breaks; Ap, ampicillin.

Received for publication, September 10, 2004, and in revised form, October 13, 2004
Published, JBC Papers in Press, October 15, 2004, DOI 10.1074/jbc.M410427200

Marzena Wojciechowska, Albino Bacolla, Jacquelynn E. Larson, and Robert D. Wells‡

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

This paper is available on line at http://www.jbc.org
The *E. coli* strains KMB1001 and CS5428 were obtained from Dr. Nora Goosen (Leiden Institute of Chemistry, The Netherlands). The *E. coli* strains JCC510 and KA796 were kind gifts from Dr. Benoist Michel (Institut National de la Recherche Agronomique, France) and Dr. R. Schaeper (NIEHS, National Institutes of Health, Research Triangle Park, NC), respectively. The *E. coli* strain RW118 was obtained from Dr. Roger Woodgate (NICH, National Institutes of Health, Bethesda, MD). Table I shows the relevant nomenclature and genetic features of these strains.

**Parental Plasmids**—The (CTG/CAG) 

This fragment was filled in with 0.1 unit of the Klenow fragment of DNA polymerase I (U.S. Biochemical Corp.) and the four dNTPs (NTPs). The reactions were carried out according to the manufacturer's recommendations (U.S. Biochemical Corp.). The primers for sequencing the regions upstream reporter gene were obtained from Sigma Genosys (The Woodlands, TX). The forward primer at position 52 of pRW5301 was GCAGCTGGCACGACAGGTT-C. The reverse primers at position 1509 or 1846 of the same plasmid were CAAGCTGTGACCGTCTCCG and CAGGGTTATTGTCTCATG, respectively. In some cases, it was necessary to use a primer at the origin of replication. This forward primer at position 3390 was GCT-GCCGCTGGATCCGACTAC-GCCGG) and AP2 (CGAGTCGGCTCAGCCGTGCGCCG) (Genosys). The products of the sequencing reactions were analyzed on 6% Long Ranger Gels (FMC BioProducts), containing 7.5 M urea, in the glycerol-tolerant gel buffer (U.S. Biochemical Corp.). The gels were dried and exposed to x-ray film. The superhelical forms of the DNA containing the undeleted CAG/CTG and GAA/TTT repeats in orientations I and II (5, 25) were used for all subsequent experiments. The pGFP vector was used as a control (Fig. 1).

**Transformation of TRS-containing Plasmids into *E. coli*—** The primers for cloning the fragments containing the (GAA/TTC) repeats, which were radiolabeled with [α-32P]dATP, and electrophoresed on a 7% native polyacrylamide gel. The gels were then analyzed using a PhosphorImager (Storm 820, Amersham Biosciences).

Conditions of Bacterial Growth for the Screening of White Colonies upon Transcription Activation or Repression—The parental *E. coli* strains were used to evaluate the capacity of the TRS to induce mutations in the upstream reporter gene. To identify single green and white colonies ("white CFUs"), the liquid culture from growth cycles 1, 3, and 5 were spread on LB agar plates containing Ap (150 μg/ml) and IPTG (2 mM), and the plates were incubated at 37°C. By using a long wavelength UV lamp, green and white colonies were counted. In each experiment, 100 white colonies were isolated from the 1st, 3rd, and 5th growth cycles were re-streaked three times on LB agar plates (Ap and IPTG as before) to be certain that the loss of fluorescence was a permanent phenotype (22). Approximately 30 white colonies were transferred to LB liquid medium and grown at 37°C, and the DNA was isolated. This DNA was used for restriction analyses, DNA sequencing, and re transformation into new competent *E. coli* cells to confirm the white phenotype. The primers for sequencing the regions upstream reporter gene were obtained from Sigma Genosys (The Woodlands, TX). The forward primer at position 28 (gift from Dr. Richard P. Bowater, University of East Anglia, UK) to turn off transcription (21). For the purpose of these studies, the *E. coli* strains (Table I) were first transformed with pBlu-kan expressing the repressor and then subsequently with the appropriate plasmids harboring the CTG/CAG and GAA/TTT repeats (Fig. 1). The transformants were spread on LB agar plates containing Ap (150 μg/ml) and kanamycin (50 μg/ml), and the recultivation assay was conducted as described earlier. After the 1st, 3rd, and 5th growth cycles, the cultures were plated on LB agar plates containing Ap (150 μg/ml), kanamycin (50 μg/ml), and IPTG (2 mM) and incubated at 37°C. After the green and white colonies were counted and the white ones re-streaked (three times), the liquid cultures were prepared, and DNA was isolated as described previously. The utilization of different parental strains enabled a preliminary survey of the potential role of genetic backgrounds on the mutagenic process.

**Sequencing Primers**—The primers for sequencing the regions upstream of the GFP gene and downstream of the CTG/CAG as well as the GAA/TTT repeat sequences were obtained from MWG Biotec (High Point, NC) or from Sigma Genosys (The Woodlands, TX). The forward primer at position 52 of pRW5301 was GCGATCCGAGCCAGT-GCCGGT-TCC. The reverse primers at position 1509 or 1846 of the same plasmid were CAAAGTGTGACGCTTCCG and CAGGTTATATTCTCAG, respectively. In some cases, it was necessary to use a primer at the origin of replication. This forward primer at position 3390 was GCT-TCCAGGGGAAACGCCGTC. To detect breakdowns of the primers of pGFP vector, the DNA receptor gene was cloned into the pGFP vector. The plasmids were used at a concentration of 10 pmol/μl and 200 ng/μl, respectively, for the sequencing reactions.

**DNA Sequencing**—The DNA isolated from the 47 mutant white colonies was sequenced from both strands by using the primers described above. The DNA was sequenced in the Molecular and Human Genetics.
Deletions and Inversions Induced by Long CTG·CAG Tracts

| TRS     | Plasmid     | Number of triplet repeats | Orientation |
|---------|-------------|---------------------------|-------------|
| CTG+CAG | pRW5300     | 17                        | I           |
|         | pRW5308     | 17                        | II          |
|         | pRW5301     | 98                        | I           |
|         | pRW5305     | 98                        | II          |
|         | pRW5302     | 175                       | I           |
|         | pRW5309     | 175                       | II          |
| GAA·TTC | pRW5303     | 9                         | II          |
|         | pRW5304     | 176                       |             |
|         | pGFP        | none                      |             |

Fig. 1. Plasmids used in this study. All plasmids were derivatives of pGFP (named pRW3619 in Ref. 21), which contains a transcription terminator cassette cloned into the Sapl site of pGFPuv (Clontech). The CTG·CAG-containing fragments of different lengths (solid gray segment) were cloned into the EcoRl and Eagl recognition sites of the pGFP (for details see “Experimental Procedures”). Orientations I and II were defined by the presence of CTG or CAG repeats, respectively, on the leading strand template for DNA replication. With the exception of the (CTG·CAG)$_n$ repeat sequence, which is not a pure tract because it contains two G to A interruptions at repeats 28 and 69 (4), the CTG·CAG tracts of length 17 and 98 are pure (uninterrupted). The GAA·TTC containing fragments were also cloned downstream of the GFP gene. The tracts are uninterrupted triplet repeats. Cross-hatched arrow, pUC19 origin of replication; box with large X, transcription terminator sequence, short black arrow, lacZ promoter-operator (Pr); solid black segment, lacZ·GFP fusion gene; long open arrow, ampicillin resistance gene; E, EcoRI; A, Eagl recognition sites.

Sequencing Core at the Baylor College of Medicine, Houston. The DNA of the mutants was analyzed by cycle sequencing using a GeneAmp PCR System 9700 and the ABI 3700 Sequencer. Cycle sequencing conditions were as follows: initial denaturation at 96 °C for 10 min, and 25 cycles of heating (96 °C, 10 s), annealing (50 °C, 30 s), and elongation (60 °C, 4 min).

RESULTS

Strategy of Study—The GFP gene served as a reporter to study the influence of the triplet repeat tracts (CTG·CAG)$_n$ and (GAA·TTC)$_n$ on mutations in flanking sequences. All the TRS were cloned into the region of the vector that was adjacent and downstream to the GFP gene (Fig. 1), so that repair of the non-B DNA structures, which may form at the repeat sequences (21), could be analyzed.

The (CTG·CAG)$_n$-containing fragments (where $n = 17, 98$, and 175) were cloned either in orientation I or II, relative to the origin of replication; for the plasmids containing (CTG·CAG)$_n$, tracts in orientation I, the CTG repeat is in the leading strand template, whereas for the plasmids in orientation II, the CTG repeat is in the lagging strand template (5, 24, 25, 29, 30). The (GAA·TTC)$_n$ repeat sequences (where $n = 9$ and 176) were cloned into the pGFP vector only in orientation II, where the GAA repeats are in the lagging strand template; orientation I is less genetically stable than orientation I (31, 32). As a control, the pGFP vector with no repeat sequences was used (Fig. 1). We performed the experiments either with transcription activation or repression and conducted five successive re-cultivation steps. To activate transcription, experiments were performed in the presence of 2 mM IPTG, whereas co-transformation with pL9·kan ensured that the lacI9 repressor inhibited GFP transcription from the lacZ promoter (22, 28).

Fraction of White CFUs Depends on the Type and Length of the TRS Sequence—The genetic instability of the TRS depends on the length of the repeat tracts (12, 25). To determine whether repeat tracts of different lengths influenced mutations in sequences flanking the repeats, we transformed the E. coli strains (Table I) with the nine plasmids listed in Fig. 1, harboring (CTG·CAG)$_n$, or (GAA·TTC)$_n$ repeat sequences. To identify single colonies, liquid cultures at the end of the 1st, 3rd, and 5th re-cultivations were spread on IPTG-containing agar plates, and the number of green and white colonies was counted. To verify the fluorescent status of the cells used to start each re-cultivation, the transformation mixture was immediately plated on LB plates, and the green fluorescence of the cells was determined. The cells used for each re-cultivation assay contained plasmids with a functional GFP reporter gene, because all colony-forming units (“CFUs”) were fluorescent. Therefore, all white colonies arose during the re-cultivation growths of the cells (Fig. 1). The fraction of white CFUs was calculated as the ratio of the number of white colonies to the total number of viable cells (green and white).

Several “parental” E. coli strains that are genotypically different were studied. This diversity may influence the cellular behavior. To investigate a possible role of DNA repair in TRS-induced mutagenesis, while taking into account the genotypic variability, we examined whether the presence of the TRS increased the mutations in six different parental E. coli strains, all proficient in the four main repair pathways (methyl-directed mismatch repair, nucleotide excision repair, transcription-coupled repair, and base excision repair). The results of screening for white colonies, when plasmids contained (CTG·CAG)$_n$ tracts of different lengths ($n = 0, 17, 98$, and 175) and (GAA·TTC)$_{176}$, are shown in Table II. The data revealed that the loss of fluorescence of the GFP reporter gene depends on the presence of the CTG·CAG tract, because the fraction of white CFUs increased with the length of the repeat tract. The total number of CFUs analyzed ranged from 18,700 to 68,237, with an average of 34,718 CFUs in each experiment. Table II is a summation of all data for plasmids in both orientations and with and without transcription, in order to present the composite global results.

For all six strains, no white CFUs were found (Table II) with plasmids that lacked the TRS. When plasmids contained (CTG·CAG)$_{17}$, mutants were found in KMBL1001 and KA796 strains at a frequency of 0.0004 and 0.001, respectively. A distinct increase in the fraction of mutants was found when longer CTG·CAG sequences, either pure or interrupted, were present. For plasmids harboring the uninterrupted (CTG·CAG)$_{17}$, the highest fraction of white CFUs (0.31) was found when the DNA was propagated in KMBL1001. A slightly lower fraction of mutants was formed in the JJC510 and JTT1 strains (Table II), whereas in KA796, the fraction of white CFUs reached a level of 0.001. There was no effect on the loss of fluorescence when the re-cultivation assays were conducted after transformation of AB1157 and RW118 strains with plasmids containing the uninterrupted (CTG·CAG)$_{17}$ sequence (Ta-
Deletions and Inversions Induced by Long CTG·CAG Tracts

### Table I

| Strain          | Relevant genotypes                                                                 | Name               | Ref. |
|-----------------|-----------------------------------------------------------------------------------|--------------------|------|
| HB101           | F·, Δgpt-proA62, leuB6, glnV44, ara-14, gaiK2, lacY1, Δ(mcrC-mrr), rpsL20, (Str+) | Cloning strain     | 15   |
| KMBL1001        | No known mutations                                                                | Parental strain 1  | 22   |
| AB1157          | thr-1, ara-14, leuB6, Δ(gpt-proA62, lacY1, tex-33, gyrA, glnV44(As), gaiK2, lac, rac, hisG4(Or), rpsD1, mig-51, rpsL31, kgdK51, xylA5, mtl-1, argE3(Oc), thi-1 | Parental strain 2  | 75   |
| JJC510          | GY7486; λ [psi(sfa·lacZ) CI-ind·Δlac-pro, rpsL]                                    | Parental strain 3  | 22   |
| JTT1            | gal-25, λ·, pyrF287, for-1, rpsL195(STR), ic1R7(ConSt), trpR72(Am)                  | Parental strain 4  | 22   |
| RW118           | thi-1, thr-1, araD139, lacY1, argE3(Oc), Δ(gpt-proA62, mtl-1, xyl-5, rpsL31, tex-33, supE44, galK2(Or), hisG4(Oc), rpsD1, kgdK51, sulA211) | Parental strain 5  | 76   |
| KA796           | (ara, thi, Δpro-lac)                                                               | Parental strain 6  | 77   |
| CS5428          | KMBL1001 except wva:cam                                                            | ΔUvrA              | 22   |

### Table II

#### Mutagenesis induced by CTG·CAG and GAA/TTC repeat sequences in E. coli DNA repair-proficient cells

The strategy to determine the fraction of white CFUs was described under “Experimental Procedures.” Briefly, after transformation, the E. coli cells were plated onto LB plates containing Ap and IPTG and incubated at 37°C. The fluorescence of the colonies harboring a functional GFP gene was determined by exposure to a long wavelength UV lamp. The fraction of white CFUs was calculated by dividing the number of white mutants by the total number of viable colonies (green and white). The fractions shown as zero (Tables II–IV) represent data where no white CFUs at all were found, even in multiple repeat experiments. The data represent the combined results of three or more independent experiments consisting of the five-step re-cultivation protocol. Data for each length of the CTG·CAG repeats represent the composite results obtained for orientations I and II, as well as when transcription was both turned on and off. Data for GAA/TTC (orientation II) were obtained under the same experimental conditions. The fraction of white CFUs (bold font) is shown as mutant/total CFUs.

#### Parental E. coli strains

| Strain   | Length of CTG-CAG | Length of GAA/TTC, 176 |
|----------|-------------------|------------------------|
|          | 0                 | 17                     | 98                     | 175                       |
| KMBL1001 | 0                 | 0.0004                 | 0.31                   | 0.13                      | 0                       |
|          | 0/32,205          | 26/57,747              | 21,388/68,237          | 8,152/54,741              | 0/46,987                |
| JJC510   | 0                 | 0                      | 0.279                  | 0.231                     | 0                       |
|          | 0/23,870          | 0/36,975               | 15,313/54,777          | 14,433/62,458             | 0/35,551                |
| JTT1     | 0                 | 0                      | 0.23                   | 0.143                     | 0                       |
|          | 0/29,150          | 0/33,925               | 7,095/30,211           | 5,515/38,396              | 0/43,334                |
| KA796    | 0                 | 0                      | 0.001                  | 0.001                     | 0                       |
|          | 0/27,203          | 20/18,700              | 27,285/507            | 94/28,783                 | 0/37,310                |
| AB1157   | 0                 | 0                      | 0                      | 0.133                     | 0                       |
|          | 0/25,492          | 0/24,278               | 0/33,496              | 4,165/31,325              | 0/39,711                |
| RW118    | 0                 | 0                      | 0.116                  | 0                        | 0                       |
|          | 0/21,387          | 0/23,254               | 0/24,383              | 2,961/25,452              | 0/29,004                |
Deletions and Inversions Induced by Long CTG-CAG Tracts

Effect of TRS insert orientation on the fraction of white CFUs for plasmids containing (CTG-CAG)_n sequences (n = 17, 98, and 175)

The orientation dependence of the repeat sequences and the loss of fluorescence from the GFP reporter gene are presented as the fraction of white CFUs. Growth cycles of *E. coli* cells harboring plasmids with the repeat sequences in both orientations were conducted in the presence of IPTG, after the bacterial cells were transformed with the TRS-containing plasmids (Table I). In order to visualize green colonies, IPTG was included in the agar plates in all cases (see "Experimental Procedures"). The data shown are the summation of the results for each length of the TRS in orientations I and II. The fraction of white CFUs (bold font) is shown as mutant/total CFUs.

| Parental E. coli strains | (CTG-CAG)_{17} | (CTG-CAG)_{98} | (CTG-CAG)_{175} |
|-------------------------|----------------|----------------|-----------------|
|                         | Orientation I | Orientation II | Orientation I | Orientation II | Orientation I | Orientation II |
| KMBl1001                | 0             | 0.0016         | 0.229          | 0.459          | 0             | 0.365          |
|                         | 0/41,217      | 26/16,530      | 9,961/43,349   | 11,427/44,888  | 0/32,401      | 8,152/22,340   |
| JJC510                  | 0             | 0              | 0.179          | 0.356          | 0             | 0.561          |
|                         | 0/19,119      | 0/17,856       | 11,038/30,940  | 4,275/23,837   | 0/36,751      | 14,433/25,707  |
| JTT1                    | 0             | 0              | 0.214          | 0.0013*        | 0             | 0.272          |
|                         | 0/17,864      | 0/16,061       | 7,085/22,539   | 10/7,672       | 0/18,156      | 5,515/20,240   |
| KA796                   | 0             | 0.0017         | 0.0015         | 0              | 0             | 0.008          |
|                         | 0/6,732       | 20119,968      | 27/17,204      | 0/11,303       | 0/17,056      | 9/11,727       |
| AB1157                  | 0             | 0              | 0              | 0              | 0             | 0.219          |
|                         | 0/10,780      | 0/13,498       | 0/23,211       | 0/10,285       | 0/12,311      | 4,165/19,014   |
| RW118                   | 0             | 0              | 0              | 0              | 0             | 0.196          |
|                         | 0/11,856      | 0/11,398       | 0/14,715       | 0/9,668        | 0/10,387      | 2,961/15,065   |

* For orientation II of (CTG-CAG)_{98} in JTT1, a small fraction of mutants, as compared with orientation I, may be linked to a reduced viability of the cells harboring pRW5305 due to plasmid loss and deletion events affecting the origin of replication and/or the ampicillin gene.

Transcription through the long CTG-CAG sequences increases the fraction of white CFUs

For experiments with inactive transcription, *E. coli* cells harboring pK4-kan were transformed with the designated plasmids and grown in LB medium without IPTG. Parallel experiments with active transcription were conducted in the presence of IPTG, after the bacterial cells were transformed with the TRS-containing plasmids (Table I). In order to visualize green colonies, IPTG was included in the agar plates in all cases (see "Experimental Procedures"). The data shown are the summation of the results for each length of the TRS in orientations I and II. The fraction of white CFUs (bold font) is shown as mutant/total CFUs.

| Parental E. coli strains | (CTG-CAG)_{17} | (CTG-CAG)_{98} | (CTG-CAG)_{175} |
|-------------------------|----------------|----------------|-----------------|
|                         | IPTG | No IPTG | IPTG | No IPTG | IPTG | No IPTG |
| KMBl1001                | 0.00086 | 0 | 0.593 | 0.00059 | 0.282 | 0.00019 |
|                         | 26/30,263 | 27/484 | 21,372/41,096 | 16/27,141 | 8,147/28,871 | 5/25,870 |
| JJC510                  | 0 | 0 | 0.563 | 0 | 0.389 | 0.00012 |
|                         | 0/20,279 | 0/16,696 | 15,313/27,201 | 0/27,576 | 14,430/37,020 | 3/25,438 |
| JTT1                    | 0 | 0 | 0.432 | 0.00029 | 0.204 | 0 |
|                         | 0/20,197 | 0/13,728 | 7,091/16,395 | 4/13,816 | 5,515/20,734 | 0/11,362 |
| KA796                   | 0.0021 | 0 | 0.00177 | 0 | 0.0064 | 0 |
|                         | 209/429 | 0/9,274 | 27/15,239 | 0/13,268 | 9/14,645 | 0/14,138 |
| AB1157                  | 0 | 0 | 0 | 0 | 0.209 | 0.00035 |
|                         | 0/11,953 | 0/12,325 | 0/19,391 | 0/14,105 | 4,161/19,903 | 4/11,422 |
| RW118                   | 0 | 0 | 0 | 0 | 0.188 | 0 |
|                         | 0/12,563 | 0/10,691 | 0/12,445 | 0/11,938 | 2,961/12,217 | 0/13,235 |

when pRW5305 containing 98 repeats was re-cultivated in *E. coli* JJC510 (0.356 versus 0.179). Most surprisingly, when growth was conducted in JTT1 cells, we detected the reverse effect, i.e., a higher fraction of white CFUs was formed for (CTG-CAG)_{175} in orientation I (Table III). In this case, there were 7,985 white CFUs out of 22,599 CFUs in orientation I (0.314) and only 10 white mutants out of 7,672 CFUs (0.0013) for orientation II (Table III). This anomalous result is probably linked to the inviability of the cells harboring pRW5305, because a considerably smaller number of CFUs was observed on LB plates. Hence, this low frequency of white CFUs may have been due to plasmid loss (22) or deletion events affecting the ampicillin gene and/or the replication origin.

The orientation dependence on the fraction of white CFUs was extreme for the longer interrupted (CTG-CAG)_{98} repeats (Table III). When the re-cultivations of the cells harboring plasmids containing this longer tract cloned in orientation I (pRW5302) were conducted, no white CFUs were detected in any of the six strains. Alternatively, the same tract in orientation II showed a large number of white mutants. The highest frequency was found in JJC510 (14.433 white CFUs out of 25,707 total CFUs) (0.561). A somewhat lower fraction was found in KMBl1001 cells (8,152 white CFUs out of 22,340 total viable cells) (0.365) (Table III). A similar fraction of deleted mutants was detected in strains RW118 and AB1157 (2,961 white CFUs out of a total of 15,065 CFUs (0.196) and 4,165 out of 19,014 (0.219), respectively). The lowest fraction (94 out of 11,727) (0.008) was counted in KA796 (Table III). For the shorter (CTG-CAG)_{17} in KMBl1001 and KA796, white CFUs were found only for orientation II in a ratio of 0.0016 and 0.0017, respectively.

In summary, these data show that the orientation, as well as the length of the CTG-CAG sequence, are important factors that influence the fraction of white CFUs.

Transcription through the CTG-CAG Sequence Increases the Fraction of White CFUs—Transcription has been shown to...
induce mutations that invoke DNA repair and recombination (33–38). Induction of transcription in long CTG/H18528CAG repeats contained on plasmids in E. coli revealed an increase in the frequency of deletions within the repeat tract (18). Therefore, we tested whether active transcription through the DM1 TRS influenced the fraction of mutations in sequences flanking the repeats. Table IV shows the composite data on the distribution of white CFUs found for plasmids containing CTG/H18528CAG sequences in orientations I and II, when experiments were conducted in the presence (IPTG) or the absence (no IPTG) of transcription through the repeats.

Propagation of plasmids harboring (CTG/H18528CAG)98 in the absence of transcription (no IPTG) gave rise to no, or very few, white CFUs (Table IV). A very small fraction of white CFUs was found in KMBL1001 and JTT1 strains (0.00059 and 0.00029, respectively); 16 white CFUs out of 27,141 total CFUs and 4 out of a total of 13,816, respectively, were found. For the longer interrupted (CTG-CAG)$_{98}$ in the absence of transcription, a small fraction of mutants was observed only in KMBL1001, JJC510, and AB1157 (Table IV). Five mutants out of 25,870 total CFUs were found in KMBL1001; in JJC510 and AB1157, the white CFUs comprised 3 out of a total of 25,438 and 4 out of 11,422 under transcription repression by pIQ-kan (28). No white CFUs were observed in any of the six strains with (CTG/CAG)$_{17}$ in the absence of transcription.

In the presence of transcription, the CTG-CAG repeat sequences caused a significant elevation in the fraction of the white CFUs in a length-dependent manner (Table IV). For (CTG-CAG)$_{17}$, a small fraction of white CFUs appeared in the KMBL1001 and KA796 strains, and as the length increased, more strains showed a response (Table IV). The highest fraction of deleted mutants was observed for both the pure and interrupted CTG-CAG tracts in KMBL1001 and JJC510. Plasmids harboring (CTG-CAG)$_{98}$ showed a substantial fraction of mutants when replicated in JTT1 (Table IV). The lowest fraction of white CFUs for both the 98 and 175 CTG-CAG repeat sequences was found when plasmids were cultivated in KA796. Alternatively, studies on the influence of transcription on the

### Table IV: Deletions of pRW5301

| E. coli Strain | Mutant Clone | Size of Deletions (kbp) | Amp$^R$ | On | Ter | GFP | (CTG-CAG)$_{98}$ | Number of repeats remaining |
|----------------|--------------|-------------------------|--------|----|-----|-----|------------------|---------------------------|
| KMBL1001      | 1,2,3        | 1.4                     |        |    |     |     |                  | 7                         |
| KMBL1001      | 4            | 1.2                     |        |    |     |     |                  | 55                        |
| KMBL1001      | 5            | 0.8                     |        |    |     |     |                  | 94                        |
| KMBL1001      | 6            | 0.9                     |        |    |     |     |                  | 30                        |
| KMBL1001      | 7            | 1.0                     |        |    |     |     |                  | 7                         |
| KMBL1001      | 9            | 0.6                     |        |    |     |     |                  | 21                        |
| KMBL1001      | 19           | 1.4                     |        |    |     |     |                  | 20                        |
| JTT1          | 20,22,25     | 1.5                     |        |    |     |     |                  | 0                         |
| JTT1          | 21           | 1.6                     |        |    |     |     |                  | 0                         |
| JTT1          | 23,27        | 1.8                     |        |    |     |     |                  | 0                         |
| JTT1          | 24           | 1.4                     |        |    |     |     |                  | 10                        |
| JTT1          | 26,31        | 1.4                     |        |    |     |     |                  | 0                         |
| JJC510        | 28,29,30     | 1.5                     |        |    |     |     |                  | 0                         |

### Table IV: Deletions of pRW5305

| E. coli Strain | Mutant Clone | Size of Deletions (kbp) | Amp$^R$ | On | Ter | GFP | (CTG-CAG)$_{98}$ | Number of repeats remaining |
|----------------|--------------|-------------------------|--------|----|-----|-----|------------------|---------------------------|
| KMBL1001      | 32,33,34,37,38 | 1.4                   |        |    |     |     |                  | 17                        |
| KMBL1001 *    | 35           | 1.4                     |        |    |     |     |                  | 17                        |
| KMBL1001 *    | 36           | 1.3                     |        |    |     |     |                  | 35                        |

### Fig. 2

Sequences of deletion mutants (white CFUs) derived from restriction maps of pRW5301, pRW5305, pRW5309, and pRW5304. Transcription was activated by the introduction of 2 mM IPTG into the LB media, whereas its repression was obtained by co-transformation with pIQ-kan. The first column for each part (A–D) shows the E. coli strains, which harbored the plasmids. The second column lists the names of the individual mutant clones. The third column shows the sizes of the deletions, and the fourth column is a schematic representation of the deletions. The open spaces between the segments indicate the location of the mutations. The last column shows the number of triplet repeats that are retained in the deleted plasmids. All clones were obtained in the presence of transcription, except those in the last two rows of part B (asterisks). All sequences of the “white mutants” in B and C had inversions that include the CTG-CAG repeat tract and short sequences downstream and upstream of the TRS. The open boxes mark the regions of the inversions. Amp$^R$, ampicillin resistance gene; Ori, ColE1 unidirectional origin of replication; Ter, transcription terminator cassette; GFP, green fluorescent protein gene.
deletion behaviors of (GAA-TTC)\textsubscript{176} in orientation II in all six parental E. coli strains revealed no mutagenic response. Indeed, in a total of 232,097 CFUs, no mutants were found. The one exception was in the ΔUvrA strain, where white CFUs were found only in the presence of transcription (data not shown).

Hence, the long CTG-CAG sequences exerted their mutagenic character through a process associated with transcription. The fraction of mutants formed upon transcription activation was considerably greater than in the presence of replication alone. Thus, we conclude that the deletions detected herein were the result of repair-dependent reactions, which were enhanced by transcription.

**Types and Locations of Mutations**—Restriction mapping and DNA sequence analyses of the white CFUs were performed to evaluate the alterations within the repeat tracts and flanking sequences. Analyses of 47 clones revealed that all mutants contained a nonfunctional GFP gene. These clones were characterized in detail. Twenty one were from pRW5301, 7 from pRW5305, 12 from pRW5309, and 7 from pRW5304 (Fig. 2). All mutations were large deletions; also derivatives of pRW5305 and pRW5309 contained inversions (Fig. 2, B and C, boxed regions). More than one clone was found with identical mutations from individual transformations of the plasmids harboring the repeat tracts.

For pRW5301 containing (CTG-CAG)\textsubscript{98} in orientation I (Fig. 2A), all white CFUs contained deletions ranging from 0.6 to 1.8 kbp. We found 15 clones with a single deletion (clones 1–4, 20–23, and 25–31), with one break always mapping near the terminator cassette (Fig. 1), and the second either inside the CTG-CAG repeat tract (clones 1–4) or downstream of the tract (clones 20–23 and 25–31). Therefore, these 15 mutant clones had lost the entire GFP reporter gene. Moreover, we found six clones that had two deletions (clones 5–7, 9, 19, and 24): one within the repeat tract and the other affecting the reporter gene. Five of these clones (all but clone 24) had a small segment of the reporter gene remaining. In addition, different numbers of residual CTG-CAG repeats were found as follows: 10 clones (mutants 1–7, 9, 19, and 24) had 7–94 CTG-CAG repeats remaining, whereas 11 (clones 20–23 and 25–31) lacked all repeats (Fig. 2A).

All mutant derivatives of pRW5305 and pRW5309 underwent both deletions and inversion reactions (Fig. 2, B and C, boxed regions). All clones had two deletion events, one affecting the GFP gene and the other the repeat sequences. The retained part of the GFP gene was 40 and 115 bp for plasmids containing (CTG-CAG)\textsubscript{98} and (CTG-CAG)\textsubscript{175}, respectively. Also, all clones had only a few CTG-CAG repeats remaining, which were in the inverted orientation. In fact, all mutants had two additional breaks outside the repeat tract within the EcoRI and EagI recognition sites; the repair of these breaks led to the inversion events giving rise to the sequences at positions 1010–
Deletions and Inversions Induced by Long CTG-CAG Tracts

The 1st column lists the names of the individual mutant clones. The 2nd and 3rd columns show the first and second breakpoints of each rearrangement site; uppercase letters indicate the nucleotides (nts) that were retained, and lowercase letters indicate nts that were deleted. The nts that were homologous between the first and second breakpoints are underlined. In the 2nd and 3rd columns, the sequences read from the 5' to the 3' ends of the top strand. For the inversions, the direction of the reading is the same, but the bottom strand (boldface type) is shown, whereby the reading proceeds from the high to low numbers of the plasmids. The 4th column shows the sequences of the rearranged products. Sequences that were joined in an inverted orientation are shown in boldface (parts B and C). The last column lists the breakpoint positions; the numbers indicate the map number of the last 3' and first 5' retained nts.

| A. Deletions of pRW5301 | 1st breakpoint | 2nd breakpoint | Junctions | Breakpoint positions |
|-------------------------|----------------|----------------|-----------|----------------------|
| 1–3                     | CGGATGCgctt    | gctgcTGCTGCT   | CGGATGCgctt | 3753/1332 |
| 4                       | TATCTGggttt    | gctgcTGCTGCT   | TATCTGggttt | 3723/1190 |
| 5–7                     | GAGCTGgctt     | agatccCAACGA   | CACCAGCAACGA | 138920 |
| 9                       | GTAATACaatt    | ctataCTTCCA    | GTAATACaatt | 195/584  |
| 19                      | TAAcctg        | ttacAGCAT      | TAAcctg    | 3596/977 |
| 5–7, 9, 19              | CTCTGCTgctg    | gctgcTGCTGCT   | CTCTGCTgctg | Inside repeat tract |
| 20                      | GCGGAGaagcga   | gageCGACCT     | GCGGAGaagcga | 3642/1397 |
| 21, 22, 25              | GCGGAGaagcga   | gageCGACCT     | GCGGAGaagcga | 3653/1448 |
| 23, 27                  | GCGGCCgctg     | aagctTATTTC    | GCGGCCgctg | 3709/1785 |
| 24                      | GCGCAdacgac    | gctgcTACCCCG   | GCGCAdacgac | 3637/1020 |
| 26, 31                  | GTGAGCGCaacgc  | gctgcTACCCCG   | GTGAGCGCaacgc | 97/1449  |
| 28–30                   | GCCTTCgTtAtA   | ctatgGCCTGCT   | GCCTTCgTtAtA | 371/1441 |

| B. Deletions of pRW5305 | 1st breakpoint | 2nd breakpoint | Junctions | Breakpoint positions |
|-------------------------|----------------|----------------|-----------|----------------------|
| 32–38                   | GGTAACaagtc    | gattacACATGG   | GGTAACaagtc | 3602/977 |
| TgatgAGCTT              |                |                | GGGATGCTGCT | 1010/1417 |
| CAGCCAgacag             | cagacACACAG    | CAGCCAgacag    | CAGCCAgacag | Inside repeat tract |
| CAAGCTAAATCCA           | TTCGGCGGCGC    | TTCGGCGGCGC    | TTCGGCGGCGC | 1010/1425 |

| C. Deletions of pRW5309 | 1st breakpoint | 2nd breakpoint | Junctions | Breakpoint positions |
|-------------------------|----------------|----------------|-----------|----------------------|
| 39–50                   | GCGGCCCaatc    | caaatGCCCC     | GCGGCCCaatc | 79/902  |
| GAAATCCaattg            | gatgcTACCCCG   | GAAATCCaattg   | 1012/1721 |
| CAGCCAgacag             | cagacACACAG    | CAGCCAgacag    | CAGCCAgacag | Inside repeat tract |
| GGGGCCGacag             | ggagccTGAGGG   | GGGGCCGacag    | GGGGCCGacag | 1076/1376 |

| D. Deletions of pRW5304 | 1st breakpoint | 2nd breakpoint | Junctions | Breakpoint positions |
|-------------------------|----------------|----------------|-----------|----------------------|
| 10–13, 16               | GCCTTCgctgc    | ttcgcgTTCCTC   | CCGGTCTCTCCTC | 4285/1763 |
| 14                      | GCCTTCgctgc    | ttcgcgTTCCTC   | CCGGTCTCTCCTC | 4285/1789 |
| 15                      | GCCTTCgctgc    | ttcgcgTTCCTC   | CCGGTCTCTCCTC | 4285/1745 |
| 10–16                   | AAGGGGGATCTgctg | AAGGGGGATCTgctg | AAGGGATCTgctg | 2160/2164 |

1417 of pRW5305 and 1076–1721 of pRW5309 (Fig. 2, B and C). The reason why derivatives of pRW5305 and pRW5309 contain inversions, in addition to deletions, is unclear but may be due to the strategy of their cloning. Even though all clones had the same breakpoint junctions after the terminator cassette and within the GFF gene, they represent independent mutation events because they were found in separate transformations. Hence, these regions must be hot spots for recombination in orientation II.

For the long GAAATTC repeat sequence, the repaired products of mutations revealed the occurrence of one large deletion of 2.0 kbp (Fig. 2D). Restriction mapping of the DNA from seven white mutants clones (10–16) showed that one break always occurred within 2 bp from the replication origin region, and the second inside the repeat tract. Therefore, these seven clones had lost the entire GFF reporter gene and a considerable part of the repeat sequence. The number of repeats remaining varied from 15 to 23 (Fig. 2D). Clones 10–13 and 16 had 17 repeats remaining, whereas clones 14 and 15 revealed 15 and 23 repeats, respectively. Furthermore, all repaired products of pRW5304 had an additional small deletion localized downstream of the repeat tract, which removed one copy of a 4-bp GATC tandem repeat. Because these three clones were found from a single transformation, it is conceivable that they were derived from a common event.

Sequence Features at the Breaks—The ability of CTG-CAG repeats to adopt quasi-stable folded secondary structures is well established (1, 6, 19, 39, 40). Non-B DNA structures are susceptible to strand breaks, either single or double, within the repeat tract (41, 42). The breaks appear to be repaired by the RecA-dependent homologous recombination pathway (43). Repair of the breaks can lead to instability of the repeat tract and, moreover, cause deletions of the sequences flanking the repeats (41, 44). The propensity of long CTG-CAG sequences to induce breaks in an adjacent gene, which were subsequently repaired, was determined.

We analyzed the positions of the breaks and the sequences at the junctions for the 47 mutant white CFUs. We also inspected the sequences flanking the breakpoints for any direct, inverted, or mirror repeats capable of forming slipped structures, cruciforms, or triplexes (1, 9, 39), respectively.

Analyses of the sequences at the breaks revealed the existence of short homologies, from one to eight nucleotides (Table V, A–D). Breakpoints did not occur at random positions but were within specific repeat sequences able to adopt nonorthodox DNA conformations (21). Moreover, sequences flanking the breaks revealed the existence of repeat motifs. For example, in the vicinity of the break at position 4285 in the repaired products of pRW5304, three copies of an eight-nucleotide motif of direct repeats GGCCTTTT were detected (Fig. 3A). This panel shows an example of the non-B DNA structures at breakpoints of deleted pRW5304 (clones 10–16 in Table V, D). A few mutant clones of pRW5301 (clones 5–7) (Table V, A) had breaks mapped at positions 138 and 920 of the vector part of the plasmid. These sites were also near three copies of direct (CAG, TACC, GGC, TTA) and mirror (CATC) repeat motifs (data not shown). Fig. 3, B–D, shows examples of the presumptive structures at the breakpoints for clones 4–7 and 19. The presence of these sequences near the breakpoints may have played a role in the
Fig. 3. Putative non-B DNA structures at breakpoints. A shows deleted clones 10–13 and 16 of pRW5304; B–D present deleted clones 19, 5–7, and 4, respectively, of pRW5301. The structures were predicted from the DNA sequences flanking the breakpoints of deletions and are examples, since alternative conformations are also possible. The nucleotides in gray boxes indicate homologous sequences at the breakpoints identified by DNA sequencing (Table V). The arrows show the sequences that were deleted, and the boldface numbers designate the positions of the breaks (given the homologies, these numbers are arbitrary). The bold dashed lines between non-B DNA structures designate the deleted intervening sequences in A, C, and D. B, the boldface arrow designates a possible folding of the DNA with strand exchanges.
deletion events (21). Characterization of a few additional deletions confirmed the association of deletion breakpoints with sites of non-B DNA structures. The non-B DNA conformations may have served as substrates for the repair machinery that generates long deletions (21).

Mutants, products of DNA repair, derived from pRW5305 and pRW5309, harboring long CTG-CAG sequences in orientation II, were deletions and also revealed inversion events (Table V, B and C). Clones 32–50 had two deletions affecting 1.2–1.4 kbp of the plasmids. For deleted derivatives of pRW5309, one mutation event occurred between positions 7 (in the vector DNA) and 902 (in the GFP gene) (Table V, C), which were joined together. The next break mapped at position 1012 within the EcoRI recognition site and was followed by bp 1721 located inside the second EcoRI restriction site of the opposite strand, hence leading to the inversion. Homologous GAATTC sequences were present at both breakpoints (Table V, C). The end point of the inverted fragment (bp 1076) located in the first EagI site continued with bp 1736 mapped inside the second EagI recognition site of the opposite strand. These breakpoints had GCGGCG homologous tracts. The other breaks, which occurred within the CTG-CAG sequence, caused deletions of the repeat tract.

For the mutants of pRW5305, a similar repair behavior was found. A 1.1-kbp deletion occurred between positions 3602 (in the vector) and 977 (in the GFP gene) (Table V, B) that were joined together. A third break was found 33 nucleotides downstream of this site at position 1010, which was followed by bp 1417 of the opposite strand causing the inversion. A homologous GAATT tract was present at the breaks. The end point of the inversion (bp 1010) continued with bp 1425 of the opposite strand, at sites where the inversion revealed a homologous GC pair (Table V, B). Furthermore, the CTG-CAG sequence, which was found in the inverted orientation, had lost a considerable number of the repeats.

In summary, the presence of the long CTG-CAG repeats promotes the formation of multiple breaks in sequences flanking the repeats. Their repair occurred between motifs that
Deletions and Inversions Induced by Long CTG-CAG Tracts

shared homology of a few nucleotides. In all cases, the positions of the breakpoints were near or within specific repeat sequences capable of forming non-B DNA structures. Thus, weakened and/or distorted base pairs in the unorthodox DNA conformations (1, 9, 45) probably served as substrates for the generation of large deletions and rearrangements (inversions).

DISCUSSION

Long tracts of the myotonic dystrophy (CTG-CAG)ₙ repeats promote inversions and deletions of 0.6–1.8 kbp of the repeats along with a portion, or all, of the flanking GFP gene. A large number of prior genetic instability studies with the TRS (1–26, 29, 30, 40–57) revealed expansions or deletions within the TRS, but no alterations were observed in the flanking and nonrepeating sequences. Thus, this remarkable mutagenic behavior was not recognized previously. Also, the effect of the orientation of the TRS insert relative to the unidirectional origin of plasmid replication was dramatic, and active transcription across the CTG-CAG tracts greatly stimulated the formation of deletions. It was not uncommon to observe 30–50% of all colonies with gross deletions. The DNA sequences of the breakpoint junctions in 47 deletions revealed the presence of short (1–8 bp) direct or inverted repeat homologies, and the presence of slipped structures, cruciforms, or triplexes at or near the breakpoints was predicted in all cases. Hence, we propose that the slipped strand (1–16, 18, 21, 22, 24–26, 29, 30, 40–54, 58) and/or the flexible and writhed (19) conformations of long CTG-CAG repeat tracts promote the formation of rearrangements.

The length of the CTG-CAG tract has a pronounced effect on the capacity of plasmids to promote gross deletions. If n = 0 or 17, essentially no mutants were observed. Alternatively, for the longer tracts (n = 98 and 175), substantial deletions were found. The role of TRS length on the capacity to adopt non-B DNA structures has been established for CTG-CAG repeats (1–6, 8–16, 18, 22, 24, 25, 29, 30, 39–49, 51–53, 58) and for GAA-TTC repeats (31, 32, 55, 59–61). Whereas the exact role of the non-B conformations adopted by the CTG-CAG repeats in the deletion process remains to be clarified, the distinct effect of repeat length in triggering these mutagenic reactions strongly suggests a role for the overall DNA topology rather than the sequence alone.

Most interestingly, long tracts of GAA-TTC were inert in promoting the formation of deletions in the work described herein. This TRS was shown to adopt triplex as well as sticky DNA conformations (31, 32, 55, 59–61); triplexes were demonstrated to cause the site-specific introduction of DNA damage in eukaryotic cells (62, 63). Thus, the molecular basis of the mutational impotency of long GAA/TTC repeats found in our studies remains to be clarified.² It is possible that the relatively long sequences flanking the GAA-TTC repeats (see “Experimental Procedures”) could contain deletions that were not detected by our assay conditions. However, the large range of deletion lengths promoted by long CTG-CAG tracts makes this possibility unlikely.

Long CTG-CAG repeats in orientation II were much more prone to promote gross deletions and inversions than in orientation I. In fact, for the longest CTG-CAG repeat (n = 175), deletions and inversions were only observed in orientation II but not in orientation I; as found for the effect of length, the host cell strain had an influence. This effect of insert orientation is diagnostic for an involvement of replication repair in the genetic instability behavior. Although this effect was first seen in plasmids harboring CTG-CAG repeats in E. coli (5, 6), it has been repeatedly observed in a wide range of studies in yeast, cell cultures, and mice (1–4, 9, 30, 41, 42, 44, 45, 47, 58, 64). This behavior is due to the preferential capacity of the CTG repeat-containing strand on the lagging strand template to adopt hairpin loop structures (compared with the less stable CAG repeat strand), which serve as an impediment for replication fork progression at the repeats and thereby enable the induction of double-strand breaks at the stalled fork. Also, the (CTG-CAG)ₙ₂₅ insert with two G to A interruptions was less mutagenic than the shorter but uninterrupted 98 repeat tract (Table II). Numerous other examples have been found of the highly disruptive effect of interruptions on genetic instabilities (10, 14, 15, 56, 57, 65–67).

Active transcription of the TRS caused an increase in the formation of gross deletions by several orders of magnitude. This dramatic effect reveals the important consequences of transcription as a biological process in mutagenesis, which has been reviewed extensively (9, 33–37, 51, 68–72). Virtually every process that exposes the single strands of DNA also destabilizes triplet repeats, including transcription (18, 51), replication (5, 52), recombination (14–16, 41, 53), and DNA repair (10–12, 54). When transcription occurs on a DNA segment that is simultaneously being replicated or contains lesions, which need to be repaired, transient changes occur in the DNA topology (9, 69, 73). As the negatively supercoiled DNA facilitates strand separation, it is vulnerable to metabolic attacks on the single-stranded regions leading to both mutagenic and recombinogenic lesions (14, 39, 69). Because transcription generates a high level of negatively supercoiled DNA and thereby promotes the formation of underwound non-B conformations, it is possible that the TRS-induced mutations were caused by these conformations at the repeat tracts. We demonstrated that the sequences at the breakpoints of the deletions for all 47 mutants could adopt supercoiled-dependent non-B conformations, in agreement with prior studies (21, 23).

Prior investigations revealed (21) that the highly unusual 2.5-kbp poly(purine-pyrimidine) sequence from intron 21 of the human PKD1 gene induced long deletions and other instabilities in plasmids that were mediated by mismatch repair and transcription. Other prior studies showed that this 2.5-kbp R-Y tract forms non-B DNA structures (22). For 11 deletions, which were analyzed in detail, the breakpoints could be explained by the formation of non-B DNA conformations. This work proposed that alternative DNA conformations (but not the sequences per se) promote genomic rearrangements through recombination-repair activities. The work described herein, demonstrating that long CTG-CAG repeat tracts also trigger the formation of large deletions and inversions and are greatly stimulated by transcription, provides a substantial extension of the original observations (21, 22) and establishes a clear role of transcription. Although transcription through the long CTG-CAG tracts (12, 18) is known to enhance its instability (via deletions), the different conditions of bacterial growth and the strains used did not previously allow detection of the gross deletions and inversions.

Fig. 4 presents a model for the mechanisms of formation of the products described in Fig. 2. The DSB, close to or within specific sequences capable of adopting non-B DNA conformations, may induce repair by the single or multiple deletion pathways. Repair of DSBs occurring between sequences with direct or inverted homologies at the breakpoints caused inversions and deletions of part or all of the repeat tracts along with some flanking DNA. For example, four mutants (Fig. 2, clones 1–4), and 11 other DNAs (Fig. 2, clones 20–23 and 25–31) are products of repair events that could have been formed by the left and right side mechanisms, respectively, of the single deletion pathway (see Fig. 4). However, clone 7 is a typical exam-

² M. Wojciechowska and R. D. Wells, unpublished data.
ple of a product formed by the multiple deletions pathway. Also, all derivatives of pRWS305 and pRWS309 as well as a few clones derived from pRWS301 (clones 5, 6, 9, 19, and 24), which had retained flanking DNA downstream of the repeats, were also derived by the multiple deletions mechanism.

Because long repeat tracts of CTG-CAAG induce gross deletions and inversions in flanking genes, the consequences of this expanded sequence in DM1 patients with a full mutation may be profound. If the same type of behavior is found in humans as observed herein, substantial deletions or other rearrangements may occur near the 3′-untranslated region causing a deletion at the carboxyl terminus of the DMPK protein. Alternatively, this process may cause a proteolytic lability of DMPK. Although this kinase has been studied extensively from biochemical, immunological, and regulatory standpoints (reviewed in Ref. 74), little or no data are available on its integrity in full mutation patients. If DMPK is labilized in patients, this novel genetic process may be responsible, at least in part, for the disease pathology.

REFERENCES

1. Wells, R. D., and Warren, S. T. (eds) (1998) Genetic Instabilities and Hereditary Neurological Diseases, Academic Press, San Diego

2. La Spada, A. R. (1997) Brain Pathol. 7, 943–963

3. Cummings, C. J., and Zoghbi, H. Y. (2000) Annu. Rev. Genomics Hum. Genet. 1, 281–328

4. Cummings, C. J., and Zoghbi, H. Y. (2000) Hum. Mol. Genet. 9, 909–916

5. Kang, S., Jaworski, A., Ohshima, K., and Wells, R. D. (1995) Nat. Genet. 10, 213–218

6. Kang, S., Ohshima, K., Jaworski, A., and Wells, R. D. (1996) J. Mol. Biol. 258, 543–547

7. Kerenzine, M. J., Goodman, M. F., and Petruska, J. (2002) J. Biol. Chem. 277, 41379–41389

8. Iyer, R. R., and Wells, R. D. (1999) J. Biol. Chem. 274, 3865–3877

9. Bowater, R. P., and Wells, R. D. (2001) Proc. Nucleic Acids Res. Mol. Biol. 66, 159–202

10. Parniewski, P., Jaworski, A., Wells, R. D., and Bowater, R. P. (2000) J. Mol. Biol. 299, 865–874

11. Oussatcheva, E. A., Hashem, V. I., Zou, Y., Sinden, R. R., and Potaman, V. N. (2002) Annu. Rev. Genet. 36, 501–535

12. Bowater, R. P., Jaworski, A., Larson, J. E., Parniewski, P., and Wells, R. D. (2004) J. Biol. Chem. 279, 15861–15867

13. Schumacher, S., Fuchs, R. P., and Bichara, M. (1998) J. Biol. Chem. 273, 30878–30884

14. Pluciennik, A., Iyer, R. R., Parniewski, P., and Wells, R. D. (1999) Nucleic Acids Res. 27, 616–623

15. Napierala, M., Dere, R., Vetcher, A., and Wells, R. D. (2004) J. Biol. Chem. 279, 28386–28397

16. Jakupciak, J. P., and Wells, R. D. (2000) Proc. Natl. Acad. Sci. U. S. A. 97, 1513–1518

17. Zou, Y., Hashem, V. I., Sinden, R. R., and Potaman, V. N. (2002) Annu. Rev. Genet. 36, 501–535

18. Bowater, R. P., Jaworski, A., Larson, J. E., Parniewski, P., and Wells, R. D. (1999) Annu. Rev. Genet. 33, 865–874

19. Napierala, M., Parniewski, P., Pluciennik, A., and Wells, R. D. (2002) J. Biol. Chem. 277, 33653–33660

20. Jakupciak, J. P., and Wells, R. D. (1999) J. Biol. Chem. 274, 3865–3877

21. Bowater, R. P., and Wells, R. D. (2001) Proc. Nucleic Acids Res. Mol. Biol. 66, 159–202

22. Parniewski, P., Bacolla, A., Jaworski, A., and Wells, R. D. (1999) Nucleic Acids Res. 27, 616–623

23. Schumacher, S., Fuchs, R. P., and Bichara, M. (1998) J. Biol. Chem. 273, 30878–30884

24. Bowater, R. P., Chen, D., and Lilley, D. M. (1994) J. Mol. Biol. 258, 543–547

25. Bacolla, A., Jaworski, A., Larson, J. E., Harvey, S. C., Stollar, B. D., and Wells, R. D. (1997) J. Biol. Chem. 272, 1263–1267

26. Grasselli, G., Marziali, A., Cuccu, D., Agostini, F., and Sancesario, G. (1999) J. Biol. Chem. 274, 23468–23479

27. Jakupciak, J. P., and Wells, R. D. (1999) J. Biol. Chem. 274, 23468–23479

28. Bowater, R. P., Chen, D., and Lilley, D. M. (1994) J. Mol. Biol. 258, 543–547

29. Maurer, D. J., O'Callaghan, B. L., and Livingston, D. M. (1996) J. Biol. Chem. 271, 1754–1760

30. Sakamoto, N., Larson, J. E., Iyer, R. R., Montermini, L., Pandolfo, M., and Wells, R. D. (1997) J. Biol. Chem. 272, 16717–16722

31. Schumacher, S., Fuchs, R. P., and Bichara, M. (1998) J. Biol. Chem. 273, 11990–11996

32. Sakamoto, N., Montermini, L., Pandolfo, M., and Wells, R. D. (1995) J. Biol. Chem. 270, 1057–1065

33. Jakupciak, J. P., and Wells, R. D. (1999) J. Biol. Chem. 274, 23468–23479

34. Bowater, R. P., Chen, D., and Lilley, D. M. (1994) J. Mol. Biol. 258, 543–547

35. Bacolla, A., Gellibolian, R., Shimizu, R., Amirhaeri, S., Kang, S., Ohshima, K., Larson, J. E., Harvey, S. C., Stollar, B. D., and Wells, R. D. (1997) J. Biol. Chem. 272, 16717–16722

36. Gellibolian, R., and Bacolla, A. (1998) in Genetic Instabilities and Hereditary Neurological Diseases (Wells, R. D., and Warren, S. T., eds) pp. 561–583, Academic Press, San Diego

37. Bacolla, A., Jaworski, A., Larson, J. E., Jakupciak, J. P., Chuzhanova, N., Abey singhe, S. H., O'Connell, C. D., Cosper, D. N., and Wells, R. D. (2004) Proc. Natl. Acad. Sci. U. S. A. 101, 14162–14167

38. Bacolla, A., Jaworski, A., Conners, T. D., and Wells, R. D. (2001) J. Biol. Chem. 276, 18597–18604

39. Bacolla, A., and Wells, R. D. (2004) J. Biol. Chem. 279, 47411–47414

40. Pluciennik, A., Iyer, R. R., Parniewski, P., and Wells, R. D. (2000) J. Biol. Chem. 275, 28386–28397

41. Bowater, R. P., Rosche, W. A., Jaworski, A., Sinden, R. R., and Wells, R. D. (1996) J. Mol. Biol. 264, 82–96

42. Ohshima, K., Montermini, L., Wells, R. D., and Pandolfo, M. (1998) J. Biol. Chem. 273, 14588–14595

43. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

B. Wieringa, personal communication.