Small Slipped Register Genetic Instabilities in *Escherichia coli*
in Triplet Repeat Sequences Associated with HereditaryNeurological Diseases*

(Received for publication, February 10, 1998, and in revised form, April 27, 1998)

Robert D. Wells‡§, Pawel Parniewski‡, Anna Plucieniuk¶, Albino Bacolla‡, Robert Gellibolian‡¶,and Adam Jaworski¶

From the ‡Institute of Biosciences and Technology, Center for Genome Research, Department of Biochemistryand Biophysics, Texas A & M University, Texas Medical Center, Houston, Texas 77030-3303 and the ¶Center of Microbiologyand Virology, Polish Academy of Sciences, 93-232 Lodz, Lodowa 106, Poland

Genetic instability investigations on three triplet repeatsequences (TRS) involved in human hereditaryneurological diseases (CTG-CAG, CGG-CCG, and GAATTC) revealed a high frequency of small expansions ordeletions in 3-base pair registers in *Escherichia coli*. Thepresence of G to A polymorphisms in the CTG-CAGsequences served as reporters for the size and locationof these instabilities. For the other two repeat sequences,length determinations confirmed the conclusions foundfor CTG-CAG. These studies were conducted in strainsdeficient in methyl-directed mismatch repair or nucleotideexcision repair in order to investigate the involvementof these postreplicative processes in the geneticinstabilities of these TRS. The observation that smalland large instabilities for (CTG-CAG)�¾ fall into distinctsize classes (1–8 repeats and approximate multiples of41 repeats, respectively) leads to the conclusion thatmore than one DNA instability process is involved. Theslippage of the complementary strands of the TRS isprobably responsible for the small deletions and expansionsin methyl-directed mismatch repair-deficient andnucleotide excision repair-deficient cells. A model isproposed to explain the observed instabilities via strandmisalignment, incision, or excision, followed by DNA synthesisand ligation. This slippage-repair mechanismmay be responsible for the small expansions in type 1hereditary neurological diseases involving polyglutaminetrimers. Furthermore, these observations may relateto the high frequency of small deletions versus alof larger instabilities observed in lymphoblastoid cells from myotonic dystrophy patients.

Neurogenetic diseases including myotonic dystrophy, fragile X syndrome, Huntington’s disease, spinobulbar muscular atrophy, spinocerebellar ataxia type 1, and Friedreich’s ataxia result from expanded triplet repeat sequences (TRS)¹ CTG-CAG, CGG-CCG, or GAA-TTC within their genes (1–3). Also, instabilities (expansions and deletions) of TRS have been associated with limb developmental diseases, including human synpolydactyly, hypodactyly in mice, and hereditary nonpolyposis colon cancer (3–6). The earlier age of onset and the increased severity of the neurological diseases through family pedigrees (clinically referred to as anticipation) are influenced by the lengths of the TRS. Long tracts of TRS are unstable and show repeat size polymorphisms in successive generations and in different tissues. In addition to observations in humans, TRS instabilities have been demonstrated in *Escherichia coli* (7–15), yeast (16, 17), transgenic mice (18), and cultured cells from patients (19, 20) and are influenced by factors including host strain genotypes (11, 12), DNA replication (7–9, 11–13), methyl-directed mismatch repair (10), growth conditions (8, 9), transcription (8), and nucleotide excision repair.² Simple repetitive sequences are known to cause misalignment-mediated DNA synthesis errors that give rise to instabilities (22–25). These deletions and expansions are thought to be due to the formation of unusual DNA secondary structures that can cause frame shift (small slipped register) mutations during DNA synthesis (26). Mono- and dinucleotide repeats including Z-DNA-forming sequences readily show sequence length polymorphisms, presumably due to multiple slippages in templates (22, 24, 25, 27, 28). Numerous prior investigations demonstrated the occurrence of DNA instabilities caused by inactivation of the methyl-directed mismatch repair (MMR) or nucleotide excision repair system (NER) system (29–41). Much less is known about the involvement of these repair systems in the instability of TRS (10, 42, 43). The mechanism of TRS expansion that is responsible for the hereditary neurological diseases has not been fully elucidated. However, the mechanism responsible for large expansions (scores or hundreds of TRS), such as for CTG-CAG in myotonic dystrophy and CGG-CCG for the fragile X syndrome, is DNA replication errors caused by the slipped register of the DNA complementary strands along with transient formation of hairpin loops and DNA polymerase pausing with primer relocation and strand elongation (11, 12, 16, 17, 44–46). Also, genetic recombination is a robust and efficient way to achieve large expansions of CTG-CAG.³ Small expansions are also extremely important for certain types of inherited neurological diseases that involve polyglutamine expansions that are encoded by CTG-CAG repeats. Normal gene products seem to tolerate a rather wide variation in the size of a polyglutamine tract (between 10–35 glutamines...
without any detectable adverse effects beyond a threshold of 35–40 glutamines) in five of the eight known diseases. However, when the tracts are beyond this threshold, pathological properties are observed. For all of the eight diseases known to share this mechanism, a strong inverse correlation exists between the length of the polyglutamine tract and the age of onset of clinical symptoms; for each added glutamine residue beyond the threshold, on average, a 1.5–2-year earlier onset age is found (48). Furthermore, in lymphoblastoid cell lines from myotonic dystrophy patients, two types of mutations of the expanded CTG-CAG repeat alleles were detected: frequent mutations that showed small changes in the repeat size and relatively rare mutations with large changes in the CTG-CAG repeat. We believe that the large changes in repeat size observed in this human cell line are due to replication or recombination as elucidated in simpler cells (11, 12, 16, 17, 44, 46). However, the molecular mechanism responsible for the small instabilities has not been identified and may be rather different from that involved with the large instabilities.

As part of our ongoing program to elucidate the molecular events involved in genetic instabilities as related to human neurological diseases, we have investigated small slipped register genetic instabilities. These small expansions and deletions found in E. coli were studied optimally in the absence of certain repair functions (methyl-directed mismatch repair or nucleotide excision repair), since these activities would be expected to recognize the looped structures formed in the slipped TRS conformation.

EXPERIMENTAL PROCEDURES

Plasmids—All plasmids used in these experiments containing repeating (CTG-CAG) inserts are shown in Fig. 1. The inserts may also be designated (TGC-CAG) or (GCT-AGC). pRW3248 is a derivative of pUC19 NotI and was described previously (10). The plasmid contains (CTG)_175 as the leading strand template for replication, termed orientation I. This sequence is not homogenous but contains two G to A interruptions at repeats 28 and 69. pRW3297, pRW3296, and pRW3294 are derivatives of pUC19 containing different lengths of (CTG-CAG) in orientation I and were constructed and characterized in this laboratory. Four pRW3296, and pRW3294 are derivatives of pUC19 containing different lengths of (CTG-CAG) in orientation I and were constructed and characterized in this laboratory.4

4 K. Ohshima, unpublished results.
logarithmic phase ($A_{600} = 0.7$) (9) in LB medium containing 100 $\mu$g/ml of ampicillin, and plasmids were isolated. These TRS instabilities were experienced in the primary MMR-proficient and -deficient strains rather than in the secondary cloning in E. coli HB101; only one of the 25 clones analyzed from the wild type cells (Tables I–III) showed any small instability. The number of triplet repeat expansions and the positions of the interruptions in these clones were determined by DNA sequencing using the dyeoxy chain termination method and by restriction analyses. The same technique was used to analyze the repeat composition of (CTG-CAG)$_{75}$ from pRW3297, pRW3294 and pRW3296 were very stable and contained only nonepanded (CTG-CAG) inserts.

Analysis of the Repeat Composition of (CTG-CAG)$_{75}$ from pRW3248

Molecular Biological Aspects—Grown in NER-proficient and -deficient Strains—The plasmids with full-length (CTG-CAG)$_{75}$ insert as well as those with deletions in the (CTG-CAG)$_{75}$ insert were obtained as follows. E. coli parental NER-proficient strain and the uvrA, uvrB, and uvrA uvrB mutants harboring pRW3248 were grown in LB media containing ampicillin (100 $\mu$g/ml) for approximately 60 generations. These conditions were shown to cause TRS instabilities. The cells were harvested, and the DNA containing both full-length and deleted products was isolated and used to transform E. coli HB101. Cells were then spread onto ampicillin plates, and DNA from single colonies was isolated. Only those preparations that contained a homogeneous length of the CTG-CAG insert, as determined by cleavage with EcoRI and HindIII, were selected for sequence analyses.

Calculation of the Length of Deletion Products of (GAA-TTC)$_{176}$ from pRW3808—E. coli AB1157 harboring pRW3808 was propagated overnight in LB medium with ampicillin (100 $\mu$g/ml). The plasmid was isolated, purified, and cleaved with EcoRI and PstI to release the TRS-containing fragment. The restriction products were analyzed by 10% PAGE in TAE buffer at 3 V/cm at room temperature along with 1-kilobase pair DNA size markers (Life Technologies) as well as the EcoRI/PstI restriction fragments derived from pRW3832 and pRW3821 containing (GAA-TTC)$_{15}$ and (GAA-TTC)$_{25}$, respectively. The negatives of the ethidium bromide-stained gels were quantitated by densitometry (300S, Molecular Dynamics, Inc.). 506 and 517 bp size markers as well as EcoRI/PstI fragments containing (GAA-TTC)$_{176}$ (full-length product from pRW3808), (GAA-TTC)$_{15}$, (GAA-TTC)$_{57}$ (710 bp), and (GAA-TTC)$_{73}$ (1710 bp) were used to compose the standard curve by plotting the log of number of bp versus the distance migrated from the origin of the gel. The 1-kilobase pair size marker alone could not be used to construct a standard curve because of the anomalous migration of the TRS tracts (50).

For wild type cells, the majority of these large deletion products were stabilized and DNA preparations were used to construct a standard curve because of the anomalous migration of the TRS tracts (50). In the range from 1173 to 506 bp, the curve was fit by a quadratic function. The parameters of the standard curve were then used to determine the number of bp ($x$) of the unknown (GAA-TTC)$_{176}$ containing fragments by solving for the distance $x = (b < b + 4ac - 4ac + b^2)/2c$, where $a$ is the intercept, $y$ is the relative migration, and $b$ and $c$ are the constants of $x$ and $x^2$, respectively.

Analysis of Expansion and Deletion Products of (CCG-CCG)$_{34}$ from pRW3024—The methodologies and pRW3024 are described elsewhere in this section and in the legend to Fig. 5.

General Techniques—DNA preparations and agarose and polyacrylamide gel electrophoresis were performed according to standard laboratory protocols (52). Plasmids were prepared by alkaline lysis of 1 liter of culture and purified by CsCl/ethidium bromide centrifugation overnight. Also, purification of 10-ml cultures was performed using the standard alkaline lysis miniprep procedure. Restriction digestes were performed following the manufacturer’s instructions. The length analyses were performed by electrophoresis through 5 or 7% polyacrylamide gels in TAE (40 mM Tris acetate, 1 mM EDTA, pH 8.0 or TBE (90 mM Tris borate, 1 mM EDTA, pH 8.3) buffers. The gels were stained with ethidium bromide and photographed. For resolution, the SacI/HindIII bands corresponding to the inserts containing the original or the deleted TRS were eluted from the gel and ligated into the SacI- and HindIII-digested vector DNA. The ligation mixture was transformed into E. coli HB101 by electroporation. The transformants were selected on LB agar plates containing ampicillin (100 $\mu$g/ml). The positions of the interruptions in the clones were analyzed by dyeoxy sequencing of one or both strands with Sequenase (version 2.0; U.S. Biochemical Corp.) using M13 primers (New England Biolabs).

RESULTS

Expansions and Deletions in (CTG-CAG)$_{75}$ in Wild Type and MMR Mutants—In an effort to further explore the molecular processes involved in instability of (CTG-CAG)$_{75}$, we have analyzed the sequences of a family of deletion products generated from pRW3248 (Fig. 1). The plasmid contains the myotonic dystrophy (CTG-CAG)$_{75}$ sequence with two G to A polymorphisms at the 28th and the 69th repeats. These interruptions serve as valuable markers. Since DNA sequence analyses were required for these determinations, we do not have a quantitative genetic assay that facilitated the study of a large number of clones. The deletion products were generated in E. coli lacking MutL or MutS, or MutH and in the isogenic wild type strain. The studies described herein were conducted in the three MMR-deficient strains, since this repair system can recognize and correct small heterologous loops containing up to four bases (53). Such loops might form during replication of CTG-CAG and, if not repaired, lead to small expansions and deletions. Prior investigations of TRS instabilities in MMR-proficient and -deficient cells also enabled the study of deletions (10), but we were unable to identify the nature of the mispairs that served as MMR substrates.

The MMR-deficient as well as the wild type strains were transformed with pRW3248. The cells were grown for approximately 100 generations. Plasmids were isolated and cleaved with restriction enzymes to release the triplet repeat containing inserts. Bands corresponding to fragments with fewer than 175 triplet repeats were recloned into pUC19 NolI. The number of triplet repeats and the positions of the G to A interruptions in 43 clones were determined by PAGE and DNA sequencing. These data are presented in Fig. 2, and the detailed compositions of the clones analyzed are shown in Table I. In general, the results are presented in both the graphical and tabular forms, since the figures more clearly show the locations of the deletions, whereas the tables show the sizes and locations of the small instabilities in repeat numbers. The G to A interruptions in pRW3248 originally located at the 28th and at the 69th repeats (Fig. 1) were found in the MMR-deficient, but not the wild type, E. coli strains at these or nearby positions. These observations show that, besides the large deletions, small deletions (clones 17, 25, 27, 31, 32, and 35; Table I) and small expansions (clones 17–19, 26, 32, and 39) occurred within the TRS-containing inserts. For two clones, another G to A polymorphism close to the G to A interruption at position 28 was observed (clones 17 and 32; Table I). This can be explained by slippage events that took place at or very close to repeat number 28. The size of the small deletions was 1 triplet repeat unit, and the sizes of small expansions varied from 1 to 8 triplet repeats.

For wild type cells, the majority of these large deletion products lost either one or both of the G to A interruptions (clones 4–16, Table I and Fig. 2; clone 58 and 59, Table III and Fig. 4). Hence, the more unstable region of the insert is the 5’-end, which contains the G to A interruptions.

Similar experiments were performed using pRW3297, pRW3296, and pRW3294 containing (CTG-CAG)$_{25}$, (CTG-CAG)$_{55}$, and (CTG-CAG)$_{75}$, respectively. Plasmids were transformed into MMR-proficient strains, as well as mutS and mutL strains, and the cells were grown for approximately 100 generations. The isolated plasmids were characterized by PAGE and DNA sequencing. The repeat compositions of eight clones of (CTG-CAG)$_{75}$ are shown in Table II and Fig. 3. Again, the positions of the two G to A interruptions were found not only at their original positions of 28 and 59 but also at new positions, suggesting that small deletions (clones 46, 47, and 50) and the expansion (clone 46) occurred within the triplet repeat region. In the case of clone number 46, the position of the G to A interruption was shifted from repeat 59 to repeat 58. This indicates the deletion of one triplet repeat between the two interruptions (Table II, deletion region b), but since the overall number of triplet repeats in this clone remained the same (73 repeats), the expansion of one triplet repeat must have taken
place also. Hence, both an expansion and a deletion of one triplet repeat took place at different locations in the insert.
pRW3296 and pRW3294 containing (CTG\textsubscript{z}CAG)\textsubscript{57} and (CTG\textsubscript{z}CAG)\textsubscript{42} (Fig. 1) were substantially more stable when grown in wild type, mutL, and mutS strains than pRW3297, and no changes in their lengths or in the positions of the G to A interruptions were observed (data not shown).

In both sets of experiments, none of the clones obtained from the wild type strain underwent the small expansions or deletions. However, the frequency of the instability events in MMR-deficient strains (the number of clones with small expansions and deletions divided by the total of 34 clones analyzed for the MMR mutants) was very high (38%).

Genetic Instabilities of (CTG\textsubscript{z}CAG)\textsubscript{n} in Wild Type and NER Mutants—In another set of experiments, the repeat composition of (CTG\textsubscript{z}CAG)\textsubscript{175} clones from pRW3248 (Fig. 1) as grown in NER-proficient and -deficient strains was analyzed. The nucleotide excision repair enzyme complex can recognize and preferentially bind to bubble and loop regions in duplex DNA (54). The involvement of this repair system in the instability of long CTG-CAG sequences is under current investigation.\textsuperscript{2} pRW3248 was transformed into NER-proficient strains as well as the uvrA or uvrB mutants and the uvrA uvrB double mutant, and the cells were grown for approximately 60 generations. The isolated plasmids were characterized by PAGE and DNA sequencing. These data are graphically presented in Fig. 4, and the detailed composition of 30 clones is shown in Table III. As seen for plasmids propagated in the MMR-deficient strains, the positions of the G to A polymorphisms in some of these clones were changed and clustered around positions 28 and 69. However, besides the large deletions (clones 57–59, 73–75, 80, and 81) and the expansion (clone 57), only small deletions were observed (clones 56, 70–72, and 79; Table III), whereas no small expansions were found. Hence, the mechanisms of recognition and repair of the slipped structures formed by the TRS may be different for methyl-directed mismatch repair and nucleotide excision repair (34, 39).

The frequency of small deletions of CTG-CAG from the wild type strain (one out of eight clones analyzed) was not substantially different than that from the uvrA uvrB double mutant (one out of six clones analyzed). However, the absence of UvrA or UvrB affects the frequency of small deletions in different
ways. The uvrB mutant yielded a higher proportion of small deletions (three out of 10 clones analyzed) than the uvrA mutant (none of the clones analyzed). These data show a different mode of action for UvrA and UvrB on the CTG-CAG triplet repeat sequence.

In conclusion, we propose that the small expansions and deletions observed in the uvrB mutant are due to a different mechanism than those observed in the uvrA mutant. This suggests that UvrA and UvrB have distinct roles in the repair of CTG-CAG triplet repeat sequences.

### Table I

| Clone | Number of repeats in the insert | Number of pure repeats between interruptions | Position of interruption(s) | Observed changes in repeat composition |
|-------|-------------------------------|---------------------------------------------|-----------------------------|---------------------------------------|
|       |                               |                                              |                            | Expansions | Deletions |
| 1     | ~150                          | 40                                          | 28, 69                      | 4          | -25       |
| 2     | 105                           | 40                                          | 28, 69                      | 5          | 70        |
| 3     | 101                           | 40                                          | 28, 69                      | 4          | 74        |
| 4,5   | >70                           | 28                                          |                            | 10         | 105       |
| 6     | 45                            |                                              |                            | 105        |           |
| 7     | 41                            |                                              |                            | 134*       |           |
| 8     | 40                            |                                              |                            | 135*       |           |
| 9     | 39                            |                                              |                            | 136*       |           |
| 10    | 35                            |                                              |                            | 139*       |           |
| 11-14 | 32                            |                                              |                            | 143*       |           |
| 15    | 27                            |                                              |                            | 148*       |           |
| 16    | 26                            |                                              |                            | 149*       |           |
| mutL  | 17                            | 136                                         | 28, 32, 72                 | 4          | 1         | 42 |
| 18    | 133                           | 48                                          | 28, 77                      | 8          | 50        |
| 19    | 89                            | 46                                          | 28, 75                      | 6          | 92        |
| 20    | 81                            |                                              |                            |            | 94        |
| 21    | 65                            |                                              |                            |            | 110       |
| 22    | 51                            |                                              |                            | 124*       |           |
| 23    | 49                            |                                              |                            | 126*       |           |
| 24    | 43                            |                                              |                            | 132*       |           |
| mutS  | 25                            | >136                                        | 28, 68                      | 39         | 1 <39     |
| 26    | 88                            | 46                                          | 28, 75                      | 6          | 93        |
| 27    | 69                            | 1                                           | 27                         |            | 105       |
| 28    | 69                            |                                              |                            |            | 106       |
| 29    | 67                            |                                              |                            |            | 108       |
| 30    | 50                            |                                              |                            |            | 125       |
| 31    | 49                            |                                              |                            | 125        |           |
| mutH  | 32                            | 140                                         | 28, 32, 72                 | 3, 39      | 1         | 38 |
| 33, 34| >136                          | 40                                          | 28, 69                      | 1          | <39       |
| 35    | 133                           |                                              |                            |            | 41        |
| 36, 37| 68                            | 39                                          | 28, 68                      |            |           |
| 38    | 65                            |                                              |                            |            |           |
| 39    | 65                            |                                              |                            |            |           |
| 40    | 61                            |                                              |                            |            |           |
| 41    | 60                            |                                              |                            |            |           |
| 42    | 58                            |                                              |                            |            |           |
| 43    | 49                            |                                              |                            |            |           |

### Table II

| Clone | Strain  | Number of repeats in the insert | Number of pure repeats between interruptions | Position of interruptions | Observed changes in repeat composition |
|-------|---------|---------------------------------|---------------------------------------------|---------------------------|---------------------------------------|
|       |         |                                 |                                              |                           | Expansions | Deletions |
|       |         |                                 |                                              |                           | a          | b          | c |
| 44    | Wild type| 73                              | 28, 59                                      | 30                        | 30         |           |
| 45    | mutS    | 73                              | 28, 59                                      | 30                        | 30         |           |
| 46    | mutL    | 73                              | 28, 58                                      | 29                        | 1          | 1          |
| 47    | mutL    | 66                              | 27, 57                                      | 29                        | 29         |           |
| 48    | mutL    | 64                              | 19, 50                                      | 30                        | 30         |           |
| 49    | mutL    | 63                              | 28, 59                                      | 30                        | 30         |           |
| 50    | mutL    | 56                              | 12, 42                                      | 29                        | 16         | 1          |
| 51    | mutL    | 55                              | 10, 41                                      | 30                        | 30         |           |

Eight clones harboring full-size or deleted inserts were isolated from the MMR-proficient strain (wild-type), as well as the mutS and mutL mutants. The size and location of the deletions and the expansion are shown in the sixth column, where a represents the region from the 1st to the 28th triplet repeat, b from the 29th to 59th, and c from the 60th to 73rd. Deletions and the expansion occurred within the regions specified (numbers under a, b, or c).
deletions are probably due to the slippage of the complementary strands in register by one or a few triplet repeat units (Fig. 8).

**Expansions and Deletions in (CGG-CCG)n**—Expansion and deletion analyses were also conducted on CGG-CCG fragile X sequences (6–32 units in length) (14) that lacked interruptions. Investigations with (CGG-CCG)24 showed that expansions and deletions could be resolved as distinct bands on polyacrylamide gel electrophoresis. Fig. 5 shows a densitometric scan of the bands from PAGE analyses on (CGG-CCG)24. Each peak represents a band differing from its neighbor by one triplet repeat unit, corresponding to the range of (CGG-CCG)24 to (CGG-CCG)7. A nearly linear relationship in the migration exists for these species (the relationship is actually quadratic but appears linear over this small range of lengths). Also, expansion products were visually observed on this gel but could not be resolved by densitometric scans; these expansion products maintained the triplet repeat periodicity for up to approximately 45 TRS. In summary, this family of expansion and deletion products must have been derived by slippage of the register of the CGG-CCG complementary strands by multiples of 3 bp. Although these sequences did not bear reporter interruptions that were important for the work described above with CTG-CAG for unraveling the molecular mechanism, the high resolution of the analytical system for these CGG-CCG lengths warrants the conclusion of DNA slippage. Eichler et al. (55) have hypothesized microsatellite slippage of (CGG-CCG)n sequences from fragile X patients as intermediates in the expansion to full mutations.

**Expansions and Deletions from (GAA-TTC)n**—Similar investigations were conducted on the Friedreich’s ataxia TRS (GAA-TTC)176 (50) in *E. coli* AB 1157. Expansions and deletions derived from pRW3808 were analyzed by polyacrylamide gel electrophoresis. Fig. 6 shows a densitometric scan of the bands from PAGE. It was possible to visualize 25 distinct molecular species ranging from 176 to 14 triplet repeats. Similar data were obtained in each of the nine isolates investigated (data not shown). For fragments containing 142–14 triplet repeats, the peaks representing bands differ from each other by multiples of 3 bp ranging between 2 and 10 triplet repeats. A smear above the full-length (GAA-TTC)176 fragment (to the left of peak 176) most likely represents expansion products that could not be resolved by the densitometric analysis. Thus, the results of these instability investigations are in complete agreement with the results described above with CTG-CAG and CGG-CCG.

---

**FIG. 3.** Graphic representation of the (CTG-CAG)73 full-length and deletion products from pRW3297 listed in Table II. The original locations of the G to A interruptions at repeats 28 and 59 are indicated by the vertical dotted lines, whereas their positions in the deleted inserts are shown by vertical ticks.

**FIG. 4.** Graphic representation of (CTG-CAG)175 full-length and deletion products from pRW3248 listed in Table III. The original locations of the G to A interruptions at repeats 28 and 69 are indicated by the vertical dotted lines. The locations of the inserts with no interruptions cannot be precisely determined and are shown to the right (designated by asterisks).
Genetic Instabilities of Triplet Repeats

TABLE III

Thirty clones harboring full-sized and deleted inserts were isolated from the NER-proficient strain (wild type) as well as the uvrA, uvrB, and uvrA uvrB strains (see “Experimental Procedures”). The quality of the sequencing data enabled the precise determination of the number of repeats except for the cases designated otherwise (clones 52–55, 60–69, 73, 74, and 76–78). However, in all cases, the location of the G to A interruptions was unambiguous. The fourth column lists the number of CTG units between two or more interruptions. The size and location of the expansions and deletions is shown in the fifth column, where a represents the region from the 1st to the 28th triplet repeat, b from the 29th to 69th, and c from the 70th to 175th. An expansion and the deletions occurred within the specified regions (numbers under a, b, or c) as well as in the region b plus c (numbers between letters b and c) and a to c (numbers with asterisks).

| Clone  | Number of repeats in the insert | Position of interruption(s) | Number of pure interruptions between interruptions | Observed changes in repeat composition |
|--------|---------------------------------|-----------------------------|-----------------------------------------------|--------------------------------------|
|        |                                 |                             |                                               |                                      |
| Wild type |                                |                             |                                               |                                      |
| 52–55  | -175                            | 28, 69                      | 40                                            |                                      |
| 56     | 172                             | 28, 66                      | 40                                            |                                      |
| 57     | 134                             | 28, 69, 110                 | 40, 40                                        |                                      |
| 58     | 62                              | 28                          | 40                                            |                                      |
| 59     | 25                              | 28                          | 40                                            |                                      |
| uvrA   |                                 |                             |                                               |                                      |
| 60–65  | -175                            | 28, 69                      | 40                                            |                                      |
| uvrB   |                                 |                             |                                               |                                      |
| 66–69  | -175                            | 28, 69                      | 40                                            |                                      |
| 70     | 174                             | 27, 68                      | 40                                            |                                      |
| 71     | 173                             | 26, 67                      | 40                                            |                                      |
| 72     | 170                             | 26, 67                      | 40                                            |                                      |
| 73, 74 | -120                            | 28, 69                      | 40                                            |                                      |
| 75     | 84                              | 28                          | 40                                            |                                      |
| uvrA uvrB|                               |                             |                                               |                                      |
| 76–78  | -175                            | 28, 69                      | 40                                            |                                      |
| 79     | 170                             | 28, 69                      | 40                                            |                                      |
| 80     | 67                              | 28, 69                      | 40                                            |                                      |
| 81     | 30                              | 28                          | 40                                            |                                      |

DISCUSSION

The slippage of the complementary strands of simple repeating DNA sequences is known to be important for genetic instabilities. The capacity of these sequences to adopt non-B DNA conformations that promote misalignment-mediated DNA synthesis errors can cause frame shift mutations during DNA synthesis (22–26). The expansion events involved in the etiology of hereditary neurological diseases fall into two categories (type 1 and type 2) (1). For relatively small TRS expansions for the type 1 diseases involving expanded glutamine tracts in the target proteins, the mechanisms proposed herein for small slipped register expansions and deletions (SSED) may be responsible. However, for the massive expansions involved in the type 2 diseases such as myotonic dystrophy and fragile X, a different mechanism involving replication (7–12, 16, 56) or recombination3 is likely. Other workers (57) have also suggested the involvement of two different DNA repair mechanisms for correction of “small” and “large” lengths of repeats. The slippage of DNA complementary strands was hypothesized to be the mechanism responsible for the in vitro synthesis of long DNA polymers of repeating nucleotide sequences (25, 58, 59), slipped strand mispairing mutagenesis (60–62), the genetic hypermutability of dinucleotide repeat sequences in mismatch repair-deficient cells related to hereditary non-polyposis colon cancer (4, 63, 64), and the intrinsic genetic instabilities in TRS associated with several human neurological diseases (3, 11, 12, 46).

Herein, we demonstrated that CTG-CAG, CGG-CCG, and GAA/TTC sequences undergo small expansions and deletions caused by slippage events in vivo. For wild type cells, we found large deletions at the 5'-ends of the inserts that contain the G to A interruptions. We believe this is due to the formation of G-T and C-A mispairs (Fig. 7) as a consequence of complementary strand slippage. These mispairs trigger the mismatch repair system to cause the removal of the progeny strand. The resulting long single-stranded CTG tract forms a hairpin struc-

FIG. 5. Expansion and deletion products found for (CGG-CCG)<sub>12</sub>. The DNA was prepared by digestion of pRW3024 in the unstable orientation (CCG in the leading strand) with Sca1/HinDIII, and the resulting fragments were electrophoresed by PAGE at 4 °C (14). Inset, a densitometric scan of the PAGE analysis (shown above) of the fragments derived from (CGG-CCG)<sub>12</sub>. Numbers above the peaks represent the number of triplet repeats in the most abundant species, and each peak is a fragment differing from its neighbor by one repeating unit. The full figure shows the number of triplet repeats as a function of distance (in cm) migrated from the top of the gel. This relationship is generally graphed on a semilogarithmic scale, but since we are concentrating on only a small portion of the gel, we find that the measurements are essentially linear in the range from 24 to 7 repeats with a correlation coefficient of r<sup>2</sup> = 0.993. Individual expansion products (to the left of peak 24) are visualized on the photograph of the gel, but could not be resolved by the densitometric analysis.
naturally occurring CTA tract is similar to, but shorter than, the (CTG versus (position 5 combination of these two pathways (step 1 in Fig. 8) (10, 71). High levels of negative supercoiling will promote DNA slippage. Additionally, CTG-CAG and CGG-CGG are reported to be cleaved by topoisomerases (reviewed in Ref. 69). Finally, the UvrB-UvrC complex was shown to efficiently re-

FIG. 6. Deletion products found for (GAA\textsubscript{TTC})\textsubscript{176} from pRW3808. The experiment was conducted and analyzed as described under “Experimental Procedures.” Inset, densitometric scan of the restriction fragments derived from pRW3808 analyzed on PAGE showing the intensity (in pixels) of the fragments containing the full-length (GAA\textsubscript{TTC})\textsubscript{176} and the deleted products. The length of the restriction fragments (in bp) was calculated as reported under “Experimental Procedures.” The estimated number of triplet repeats for some of the peaks is indicated. This calculation takes into account the fact that long (GAA\textsubscript{TTC})\textsubscript{n} triplet repeat-containing fragments migrate approximately 15% faster than expected (50). The acrylamide gel analysis is shown above the tracing; the gel analysis and the tracing do not precisely correspond, since the tracing is a derivative of the original scan obtained by the formula shown under “Experimental Procedures” for conversion of the lengths to repeat units. In the full figure, each peak in the inset, which corresponded to a distinct band on the gel, was plotted as a function of its distance from the gel origin. The data were fit by a quadratic function ($r^2 = 0.9975$).

FIG. 7. Model for generation of large deletions at the 5’-end of the interrupted CTG-CAG tracts.

ture (Fig. 7), which is bypassed during DNA synthesis to give large deletions of this region. Alternatively, in the repair-deficient cells, a different model is proposed for small slipped register expansions and deletions in vivo (Fig. 8). The top molecule represents a (CTG-CAG)\textsubscript{14} tract that contains the naturally occurring CTA-TAG interruption at position 7. This tract is similar to, but shorter than, the (CTG-CAG)\textsubscript{n} fragments on which SSED was observed. Following strand dissociation, slippage produces two hairpins, one composed of CTTGCTG, the other of CAGCAG. Misalignment also splits the polymorphic CTA-TAG, so that now CTA pairs with CAG and TAG opposes CTG, which gives rise to two mismatched base pairs, A-C and G-T, respectively. Thus, resolution of this structure into a regular B-helix requires the repair of hairpin loops as well as of mismatches. The left pathway proposes that an incision is inflicted opposite to each loop (arrows), which, followed by DNA repair synthesis and ligation to fill-in the gaps, leads to expansion. Repair of the mismatched base pairs may be accomplished on either strand, but only the use of the bottom strand template results in a shift to the right of the original interruption (position 9 versus 7).

The pathway on the right envisions cleavage of the hairpin loops at their base, followed by ligation; this generates a two-triplet repeat deletion. Again, only repair of the mismatched bases by using the bottom strand as a template results in a concomitant shift to the left of the CTA-TAG polymorphism (position 5 versus 7). Similar results may also be obtained by a combination of these two pathways (e.g. an incision opposite to the CAG loop plus an excision of the CTG loop) if a second round of replication is included in the overall reaction.

The small instabilities were found essentially only in the repair-deficient strains. In this study, we attribute the higher frequency of large deletions, in the wild type cells in the regions containing the interruptions, with the formation of long hairpins in the CTG strand. Such hairpins may form after the "long patch" removal of the CAG strand by the MMR system acting on the slipped DNA structures. It is possible that the increased SSED observed herein in the MMR-deficient strains results from the processing of similar slipped DNA structures by activities that do not involve such long patch intermediates. Indeed, it has been shown that heteroduplex DNA is a substrate for the nicking activity of endonucleases such as deoxyuridine 5’-endonuclease (65–68). Also, hairpin structures were reported to be cleaved by topoisomerases (reviewed in Ref. 69). Finally, the UvrB-UvrC complex was shown to efficiently repair an 11-base noncomplementary region in a duplex DNA if a 3’-nick was introduced (70). Our data in E. coli with MMR+ cells are in agreement with studies reported in pmsI and msh2 yeast mutants (43).

The slippage of complementary strands containing long TRS leading to the formation of loops is promoted by any process involving strand separation (step 1 in Fig. 8) (10, 71). High levels of negative supercoiling will promote DNA slippage. Additionally, CTG-CAG and CGG-CGG are known to be more flexible and writhed than random DNA and, as a consequence, accumulate abnormally high levels of supercoil density (7, 56). The interaction of DNAs with helix tracking enzymes such as DNA or RNA polymerases introduces high levels of superhelical domains, which may lead to localized denaturation, which
will promote the slippage of TRS. Prior theoretical investigations (47) showed that slippage is favored by longer rather than shorter TRS, and hence, the sequence lengths associated with the disease states are more prone to further genetic instabilities than the lengths found in normal individuals. DNA replication may or may not be involved in the strand dissociation step (step 1 in Fig. 8).

Small instabilities are not restricted to products that were originally isolated as deletions from the parent plasmids. Indeed, in a previous study (12) where the large replication-dependent expansion products of (CTG)$_n$ (CAG)$_n$, the vertical line is CTG-CAG, a full circle is CTA, a full square is TAG, and the vertical line at position 7 is CTA-TAG. The opposing carets show non-Watson-Crick oppositions within a repetitive unit whose top and bottom sequence occupy a different position within the series. Note that identical shifts in the positions of the interruptions would be observed if SSED occurred to the left of the interruptions. In summary, slippage of the TRS complementary strands in vivo, followed by DNA repair steps, may be responsible for the small expansions and deletions observed in type 1 human hereditary neurological diseases. The availability of genetically and biochemically tractable systems in E. coli and yeast provide optimism for elucidating the molecular details of these processes.

REFERENCES
1. Paulson, H. L., and Fischbeck, K. H. (1996) Annu. Rev. Neurosci. 19, 79–107
2. Campuzano, V., Monterini, L., Molot, M. D., Pianese, L., Cossée, M., Cavalcanti, F., Monros, E., Rodius, F., Duclós, F., Monticelli, A., Zara, F., Cani, J., Koutnikova, H., Bidichandani, S. I., Geller, C., Brice, A., Trouillas, P., De Michele, G., Filla, A, De Frutos, R., Palau, F., Patel, P. I., Di Donato, S., Mandel, J.-L., Coccoza, S., Koenig, M., and Pandolfo, M. (1996) Science 271, 1425–1427
3. Wells, R. D., and Warren, S. T. (eds.) (1998) Genetic Instabilities and Hereditary Neurological Diseases, Academic Press, Inc., San Diego, CA
4. Fischel, R., Lescoe, M. K., Rao, M. R. S., Copeland, N. G., Jenkins, N. A.,
