Proc. Natl. Acad. Sci. USA
Vol. 92, pp. 11019–11023, November 1995
Genetics

Mismatch repair in *Escherichia coli* enhances instability of $(CTG)_n$ triplet repeats from human hereditary diseases

ADAM JAWORSKI*, WILLIAM A. ROSCHE, ROBERT GELLIBOLIAN, SEONGMAN KANG, MIHO SHIMIZU, RICHARD P. BOWATER, RICHARD R. SINDEN, AND ROBERT D. WELLS†

Center for Genome Research, Institute of Biosciences and Technology, Texas A&M University, Texas Medical Center, 2121 W. Holcombe Boulevard, Houston, TX 77030-3303

Communicated by Paul Modrich, Duke University Medical Center, Durham, NC, August 7, 1995 (received for review May 8, 1995)

ABSTRACT **Long CTG triplet repeats which are associated with several human hereditary neuromuscular disease genes are stabilized in ColE1-derived plasmids in *Escherichia coli* containing mutations in the methyl-directed mismatch repair genes (*mutS*, *mutL*, or *mutH*). When plasmids containing $(CTG)_{180}$ were grown for about 100 generations in *mutS*, *mutL*, or *mutH* strains, 60–85% of the plasmids contained a full-length repeat, whereas in the parent strain only about 20% of the plasmids contained the full-length repeat. The deletions occur only in the $(CTG)_{180}$ insert, not in DNA flanking the repeat. While many products of the deletions are heterogeneous in length, preferential deletion products of about 140, 100, 60, and 20 repeats were observed. We propose that the *E. coli* mismatch repair proteins recognize three-base loops formed during replication and then generate long single-stranded gaps where stable hairpin structures may form which can be bypassed by DNA polymerase during the resynthesis of duplex DNA. Similar studies were conducted with plasmids containing CGG repeats; no stabilization of these triplets was found in the mismatch repair mutants. Since prokaryotic and human mismatch repair proteins are similar, and since several carcinoma cell lines which are defective in mismatch repair show instability of simple DNA microsatellites, these mechanistic investigations in a bacterial cell may provide insights into the molecular basis for some human genetic diseases.**

Molecular genetic investigations have provided important new insights into the etiology of several human hereditary diseases, including myotonic dystrophy, Kennedy disease, spinocerebellar ataxia type I, Huntington disease, dentatorubral-pallidoluysian atrophy, and fragile X and E syndromes (reviewed in refs. 1–3). The genes associated with these diseases contain heritable unstable triplet repeat sequences (CTG in the first five cases and CGG in the latter two). The repeats can be either within the genes or in their 5′ or 3′ untranslated regions. These triplet repeats undergo expansion (to thousands of repeats in some cases) associated with a disease phenomenon called anticipation in which the penetrance is increased in successive generations. The molecular basis for this mutation, which exhibits non-Mendelian genetic properties, is unknown but may be due to slippage of the direct repeat triplets (3–5), which may be promoted by DNA secondary structures during replication (3, 6). The importance of this expansion mutation has led us to investigate the biological features of these unique DNA sequences.

One *Escherichia coli* mismatch repair (MMR) system is the methyladenine-directed postreplication process involving MutS, MutL, MutH, and other proteins (reviewed in refs. 7–10). Their coordinated actions lead to mismatch-dependent nicking of unmethylated progeny strands at a hemimethylated GATC, the subsequent degradation by exonucleases of the newly synthesized DNA containing mismatches, and the resynthesis of DNA. MMR genes are conserved in bacteria and higher organisms. Yeast MSH2 and human hMSH2 are analogous to the *E. coli* MutS protein, whereas yeast MLH1 and PMS1 and human hMLH1, hPMS1, and hPMS2 correspond to the bacterial MutL protein. A number of studies suggest a functional similarity between the postreplication repair pathway in prokaryotes and eukaryotes (7–10). A broad range of substrates is recognized, including single-base substitutions, displaced dinucleotides, small heteroduplex loops, and small loops and hairpins. Recently, the correction of larger loops (up to five or more unpaired bases) by human cell extracts (11) and the specific binding of human MSH2 to loops as large as 14 nucleotides (12) were reported.

Growing evidence indicates direct association between the mutation of human genes encoding MMR proteins and susceptibility to inherited cancer (reviewed in refs. 7–12). The direct relationship between DNA microsatellite hypervariability, somatic MMR defects, and tumor development has been suggested for colon carcinomas related to hereditary nonpolyposis colorectal cancer and perhaps for many other sporadic tumors (13–16).

Several defined ordered sequence DNAs (dosDNAs) can adopt non-B-DNA structures under appropriate conditions (negative supercoil density, ionic strength, etc.). For example, repeating Pur-Pyr sequences [i.e., $(G\text{-}C)_n$ or $(A\text{-}C)_n$] form left-handed Z-DNA, Pur·Pyr tracts [i.e., $(A\text{-}G)_n$ or $(G)_n$] form triplexes, and inverted repeats adopt cruciform structures (3, 17). A broad range of biophysical, chemical, and enzymatic probes, and electron microscopy determinations show that CTG and CGG repeats exist in non-B conformations (R.D.W., R.G., M.S., S. Amirhaeri, S.K., K. Ohshima, J. Larson, Y. Fu, T. Caskey, and B. Oostra, unpublished work; R.R.S., P. Chastain, E. Eichler, S.K., D. Nelson, S. Levene, and C. Pearson, unpublished work). Expansions and deletions of CTG repeats have been detected in *E. coli* (6). One likely mechanism for the instability of triplet repeats involves slippage of the template and progeny strands during DNA replication (3–6).

Here, we report that mutations in *mutS*, *mutL*, and *mutH* in *E. coli* cause an increase in stability of long CTG repeats. The instability observed in MMR-proficient cells may be the result of the generation of long single-strand regions and the formation of DNA secondary structures which promote deletions.

MATERIALS AND METHODS

Plasmids. pRW3247 and pRW3248 containing $(CTG)_{180}$ triplet repeat sequences were described previously (6). The $(CTG)_{180}$ insert of pRW3248 was cloned with the CTG in the leading strand of the unidirectionally replicating plasmid pUC19*Not*I (orientation I). pRW3247 contains the $(CTG)_{180}$

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

Abbreviations: MMR, mismatch repair; VSP, very short patch repair.

*Present address: Department of Genetics of Microorganisms, University of Lodz, Banacha 12/16, Lodz, Poland.
†To whom reprint requests should be addressed.

insert in the opposite orientation, with the CTG sequence in the lagging strand of the pUC19*Not*I vector (orientation II). Both plasmids contain 16 bp of human sequence proximal to the $(CTG)_n$ repeats and 43 bp distal to the $(CTG)_n$. The nearest *dam* recognition sequences (GATC) are 32 bp 5′ of the $(CTG)_{180}$ insert and 6 bp downstream on the 3′ side of this sequence. Fourteen additional GATC *dam* recognition sites are present in the pUC19*Not*I vector. pRW3244 and pRW3246 were used as size markers for gel electrophoresis and contain $(CTG)_{17}$ and $(CTG)_{100}$, respectively, cloned in the same vector at the same location, both in orientation I (6).

The construction of pRW3306 containing 160 CGG triplet repeats was as follows: A DNA fragment which contains $(CGG)_{80}$ was isolated from pTM10 (gift of B. A. Oostra, Erasmus University, The Netherlands) by digestion with *Bst*UI and *Hae* III. The insert was ligated to generate multimers by using T4 DNA ligase. A head-to-tail dimer of $(CGG)_{80}$ was inserted into the *Hin*cII site of the polylinker of pUC19. The insert contains 5 bp of the human flanking sequence proximal to the *Hin*dIII site and at the junction of the two blocks of $(CGG)_{80}$. The CGG repeat sequences in pTM10, which were derived from the cDNA of fragile X patients, contain polymorphisms of the perfect repeat at the 12th repeat (AGG) and at the 73rd repeat (CAG). The nearest *dam* methylation sites are 10 bp upstream of the CGG tract and 1452 bp 3′ of this sequence. The insert was cloned so that the leading strand contains the CGG triplets and the lagging strand contains the CCG sequence. All inserts were sequenced by using dideoxynucleotide sequencing (Sequenase).

Bacterial Strains. We used the following *E. coli* MMR mutator phenotype strains: KA796 (*ara, thi, Δpro–lac*) as a parent of the MMR strains; NR8039, like KA796, but *mutH101*; NR8040, like KA796, but *mutL101*; and NR8041, like KA796, but *mutS101*. All strains were $recA^+$ and were obtained as the kind gift of R. Schaaper (the National Institute of Environmental Health Science, Research Triangle Park, NC).

To study the influence of the RecA protein on the frequency of mutations of $(CTG)_n$ and $(CGG)_n$ sequences in the mutator cell lines, we placed the $recA^-$ allele into mutator $recA^+$ strains (NR8039, NR8040, and NR8041) by means of phage P1 transduction, creating strains WAR1 (*mutH*), WAR2 (*mutL*), and WAR3 (*mutS*). Another set of $recA^+$ strains used was the double mutants containing a second mutation in the *dam* gene. They are NK7510 (*dam-3, mutL103*::Tn*5, kan*^{r}, *str*^{r}, W3110 background) and NK7511 (*dam-3, mutS104*::Tn*5, kan*^{r}, *str*^{r}, W3110 background), gift of P. Modrich (Duke University Medical Center, Durham, NC). We also investigated the stability of the human $(CTG)_n$ and $(CGG)_n$ sequences in isogenic $recA^-$ mutator cell lines: RH4320, *strA, lacZ* (ICR36), *trp-540*?, *thi-*?, *recA1, mut-25* (*mutL*); RH4321, *strA, lacZ* (ICR36), *trp-540*?, *thi-*?, *recA1* (*mutS3*); and RH4448, *strA, lacZ* (ICR36), *trp-540*?, *thi-*?, *recA* (*mutH*). These strains were a gift from M. Radman (Institut Jacques Monod, Paris). Question marks regarding genotypes are as designated by M. Radman.

GM30 and GM31 (18), which are the wild-type very short patch repair $(VSP)^+$ and VSP^- strains, respectively, were kindly provided by M. G. Marinus (University of Massachusetts Medical School).

Standard Genetic Techniques. Plasmid DNA preparations, agarose gels, and polyacrylamide gel electrophoreses were carried out according to standard laboratory protocols (19). Electroporation was conducted with homogeneous undeleted supercoiled DNA, purified from an agarose gel, a procedure that ensures a pure population of plasmids containing full-length $(CTG)_n$ and $(CGG)_n$ repeats. Transformants were selected on LB agar plates (19) containing ampicillin (100 μg/ml).

Media and Growth Conditions. LB medium (19) containing ampicillin (100 μg/ml) was used for growth of transformants. Two types of experiments were carried out to assay for the genetic stability of $(CTG)_n$ and $(CGG)_n$ triplet repeats: analysis of single transformants and studies on populations of transformants. Plasmid DNAs for transformations were purified by agarose gel electrophoresis, after isolation from *E. coli* HB101. Single colonies (6–12) of each transformant were grown in 10 ml of LB medium to the end of logarithmic phase (OD_{650} = 0.8–0.9) at 37°C under aerobic conditions. The number of generations of cells was estimated by counting the number of colonies on LB agar plates at the beginning and the end of the growth of the cultures. To keep the cultures growing logarithmically for up to 120 generations of the cells, cultures were diluted 10^6-fold with fresh medium. Dilution of the growing cultures was repeated six to eight times to study the kinetics of deletion during these experiments. Analyses of populations were done in a similar way, but 1000–2000 transformants were used for inoculation of 100 ml of LB medium, and cells were grown under intensive aeration.

Analysis of Stability of $(CTG)_n$ and $(CGG)_n$ Sequences. The stability of $(CTG)_n$ and $(CGG)_n$ inserts in both orientations was analyzed by determining the sizes of plasmids on 1.75% agarose gels in Tris/borate/EDTA (TBE) buffer (19) and by analysis of the sizes of restriction fragments (*Sac* I–*Hin*dIII) containing the triplet repeats on 8% polyacrylamide gels or 1.75% agarose gels in TBE buffer. The reduction of the sizes of the supercoiled DNA as well as the restriction fragments, due to the loss of triplet repeats, was determined. The gels were stained with ethidium bromide and photographed. The amount of DNA in each deletion product was determined by quantitating the monomer plasmid region of the negatives with a densitometer (Molecular Dynamics). The total area under the curve was used as the total amount of DNA present, and the percentage of the plasmid that had not undergone deletion was determined by dividing the area under the curve representing the full-length material by the total area under the curve.

RESULTS

Mutation of *E. coli* MMR Genes Reduces Deletions in $(CTG)_{180}$ Repeats. The possible involvement of the *E. coli* MMR system in the destabilization of $(CTG)_n$ was investigated biochemically by using isogenic $recA^+$ strains with mutations in *mutS, mutL*, or *mutH*. Each strain was transformed with agarose gel-purified plasmid containing $(CTG)_{180}$ cloned in either orientation with respect to the direction of replication (pRW3248 and pRW3247). Several transformants of each strain were grown under identical conditions in parallel cultures, and the plasmid DNA was purified and then analyzed on an agarose gel. Significant differences in the stability of the $(CTG)_{180}$ sequence in orientation I, where the CTG strand is the leading strand of replication, were observed in the parent strain (KA796) and the MMR^- strains (Fig. 1). In the MMR^+ cells, where the repair pathway is functional, cells accumulated deletion products with increasing numbers of generations. After 100–120 generations, only 20–30% of the plasmid contained the full-length $(CTG)_{180}$ sequence (Fig. 1, parent). In contrast, after the same number of generations in MMR^- strains, the fraction of the plasmid containing full-length $(CTG)_{180}$ was significantly higher (Fig. 1). Plasmids from *mutS, mutL*, and *mutH* strains contained full-length repeat lengths in 60–85% of the plasmids.

Restriction analysis of plasmid DNA isolated after about 60 generations from the parent KA796 and the MMR^- strains indicated that the deletions occurred within the $(CTG)_{180}$ triplet repeat sequence (Fig. 2, lanes 2). Deletion products containing short CTG sequences (about 20 and 60 repeats) are frequent in the parental strain. The digestion patterns of DNA

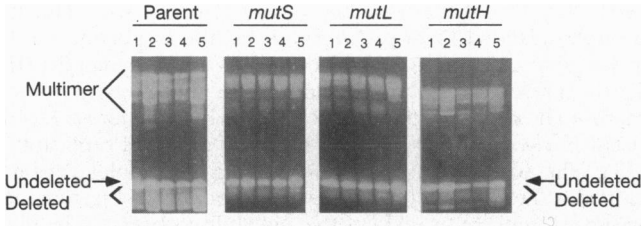

FIG. 1. Analysis of stability of CTG triplet repeats from single transformants as a function of cell growth. The figure shows agarose gel electrophoresis of supercoiled pRW3248 which contains $(CTG)_{180}$ in orientation I isolated from cultures of the parent, *mutS, mutL*, and *mutH* strains grown for about 40, 60, 80, 100, and 120 generations of cells (lanes 1–5, respectively). A constant aliquot of DNA purified from equal-size cultures was added to each lane. Because of variable yields, the amount and size of the deleted products relative to the undeleted DNA should be analyzed rather than a comparison of the amount of deletions from one lane to another. The upper bands are multimers of the plasmids, which are normally found in *recA*$^+$ cells.

from *mutS, mutL*, and *mutH* strains showed fewer deletions than in the parental MMR$^+$ strain, and the deletion products contained longer CTG sequences (100–140 repeats) (Fig. 2). That the patterns of the products found for the three mutant strains were not the same may be expected, since the biochemical pathways could be blocked at different steps. No deletions were observed for pRW3244 containing the $(CTG)_{17}$ insert in any strain, consistent with previous results (6).

To determine if the variability is due to differences between individual cultures, a preliminary fluctuation analysis (20, 21) was performed, utilizing our biochemical assay. We compared the parent and *mutS* strains from cultures growing from single cells containing pRW3248. As found in Fig. 1, the MMR$^+$ strain showed more and larger deletions than the *mutS* strain.

Deletion Analysis in Cell Populations. If the instability of triplet repeat sequences is high, one might expect high variability between individual cells. Genetic methods (20, 21) can be used to study this variation, but they cannot be directly applied to this biochemical assay of instability. To account for variations among individuals without examining large numbers of cultures, a population of cells (representing an innoculum of 1000–2000 cells) was examined. After about 80 generations of growth in the MMR$^+$ strain, the $(CTG)_{180}$ tract in orientation I had undergone deletions to yield products containing 17–20 triplet repeats in 85–90% of pRW3248 molecules (Fig. 3). Under the same conditions, the MMR$^-$ strains showed a much lower

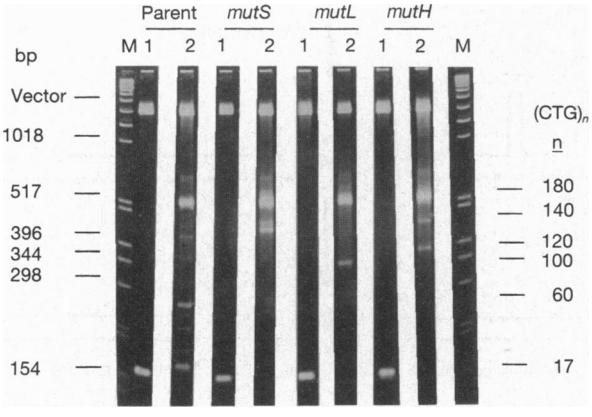

FIG. 2. Polyacrylamide gel electrophoretic analyses of *Sac* I–*Hin*dIII-digested DNAs containing CTG repeats grown in parent and mutator cell lines. Lanes 1, pRW3244 containing $(CTG)_{17}$; lanes 2, pRW3248 containing $(CTG)_{180}$ grown in the parent KA796 strain and in the mutator cell lines (*mutS, mutL*, and *mutH*, respectively). Lanes M, 1-kb marker. The sizes of the restriction fragments along with their numbers of triplet repeats are shown.

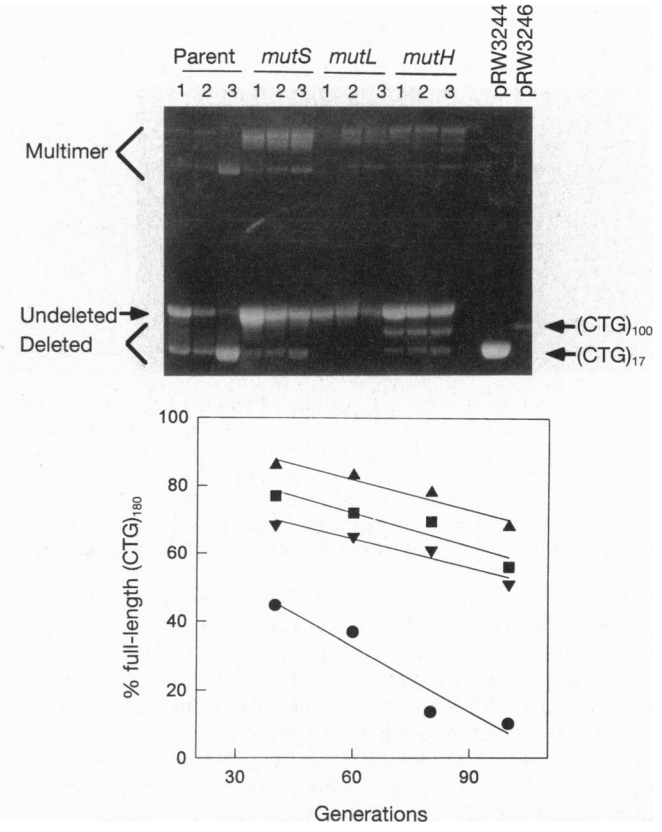

FIG. 3. Analysis of stability of $(CTG)_n$ triplet repeats from populations of transformants of pRW3248. (*Upper*) Lanes 1, 2, and 3 are agarose gel electrophoretic analyses of supercoiled DNA isolated from parent, *mutS, mutL*, and *mutH* cultures grown for 40, 60, and 80 generations, respectively. Supercoiled pRW3244 containing $(CTG)_{17}$ and pRW3246 containing $(CTG)_{100}$, respectively, were isolated from *E. coli* HB101 as supercoiled plasmid size markers. The positions of the undeleted and the deleted plasmids are designated. The bands at the top of the gels are multimers of the plasmids. The gels were quantitated; the standard deviation for three independent experiments was 5–10%. (*Lower*) Kinetics of deletion of $(CTG)_{180}$ of pRW3248 determined from the population of transformants as a function of cell growth; parent (●), *mutS* (■), *mutL* (▲), and *mutH* (▼).

accumulation of deletions. DNA isolated from the *mutL, mutS*, and *mutH* strains typically showed deletions in 10–25% of plasmid molecules after 80 generations of growth.

Influence of the Orientation of $(CTG)_n$ on Instability in the MMR$^-$ Strains. Kang *et al.* (6) showed that the stability of long $(CTG)_n$ sequences in *E. coli* is orientation dependent and that the frequency of deletions is higher when the CTG sequence is replicated in the lagging strand of the pUC19*Not*I vector. To test the influence of the *E. coli* MMR system on the stability of $(CTG)_n$ in this orientation (orientation II), we transformed pRW3247 into the MMR$^-$ strains. There was little difference in the fraction of full-length $(CTG)_{180}$ in MMR$^+$ and MMR$^-$ strains as a function of the number of generations in population studies or single-transformant analysis. However, the restriction analysis of DNA isolated from a population study after 40, 60, and 80 generations of growth indicated that there was a difference in the products of deletion, comparing the MMR$^+$ and MMR$^-$ strains. In MMR$^+$ cells, with an increasing number of generations, the $(CTG)_{180}$ sequence was reduced to two short $(CTG)_n$ tracts which contained approximately 17 and 40 repeats (deletion of about 100–140 triplet repeats) (Fig. 4). In the *mutS, mutL*, and *mutH* strains, more deletion products of different sizes and, in general, longer $(CTG)_n$ tracts were observed than in the MMR$^+$ strain (Fig. 4). In the *mutH* strain,

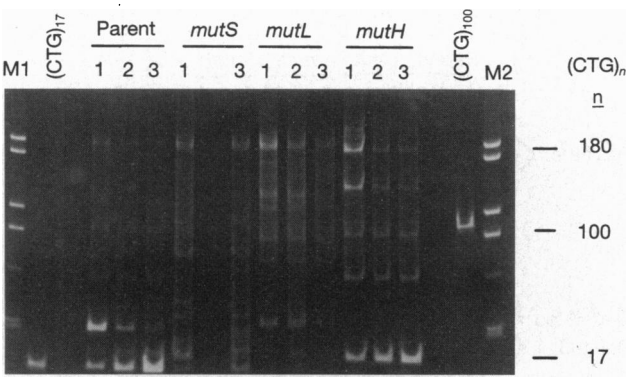

FIG. 4. Polyacrylamide gel analysis of *Hin*dIII–*Sac* I restriction fragments of pRW3247. Lanes 1, 2, and 3 show *Hin*dIII–*Sac* I restriction fragments of pRW3247 isolated from parent, *mutS, mutL*, and *mutH* strains grown for about 40, 60, and 80 generations, respectively. M1 is the 1-kb marker, and M2 is the marker of pBR322 digested with *Msp* I endonuclease. The lanes designated $(CTG)_{17}$ and $(CTG)_{100}$ are *Hin*dIII–*Sac* I digests of pRW3244 and pRW3246, respectively.

major, distinct deletion products were observed containing about 20, 60, 100, and 140 triplet repeats.

In summary, we conclude that the *E. coli* MMR system can significantly enhance the deletions of the $(CTG)_n$ triplet repeats *in vivo* for sequences cloned in orientation I. Inactivation of this system leads to a pronounced decrease in the frequency and size of deletions of the triplet repeats. In orientation II, while there is little effect on the frequency, deletions are smaller in MMR^- strains.

Stability of $(CGG)_n$ Triplet Repeats in *E. coli* MMR Mutants. More than 90% of the CGG insert in pRW3306 was deleted when grown for 60 generations in either MMR^- or MMR^+ strains, although the nature of these deletions was different. In the MMR^+ cells, deletion of about 90 and 130 repeats occurred preferentially, while in *mutS* and *mutL* cells, deletions of 108–112 repeats predominated (data not shown).

Effect of *recA* and VSP on Triplet Stability. No differences were observed in the stability of the $(CTG)_{180}$ or $(CGG)_{160}$ insert in three sets of MMR^- *recA*$^+$ and MMR^- *recA*$^-$ cells (see *Materials and Methods*), suggesting no role of the RecA recombination system in instability (data not shown). Moreover, there was no effect of the VSP pathway (22, 23) on the stability of CTG and CGG sequences. The patterns of deletion for pRW3248 [$(CTG)_{180}$] or pRW3306 [$(CGG)_{160}$] were the same when grown in GM30 (VSP^+) and GM31 (VSP^-).

DISCUSSION

The fundamental biology of CTG and CGG repeats is being investigated in a genetically well-defined organism because of the importance of the instability (expansions and deletions) of triplet repeats in human hereditary diseases. When cloned in plasmids near the ColE1 origin, the frequency of deletions of long CTG and CGG repeats was very high in wild-type *E. coli*, which contain a proficient MMR system, especially when CTG is the lagging strand template. Inactivation of the *E. coli* methyl-directed MMR system by mutations in the *mutS, mutL*, and *mutH* genes significantly decreased the frequency of deletions, as assayed biochemically, when the CTG sequence was the leading-strand template. The mechanism of deletions in the wild-type and MMR^- strains may be different, since large distinct deletions (120–160 repeats) were found in MMR^+ cells, whereas deletions of about 40–60 repeats were observed in MMR^- strains.

ColE1 unidirectional replication is initiated by leading-strand synthesis by DNA polymerase I of at least 200–400 bp of DNA from a long RNA primer. During the initial phase of replication by polymerase I, the lagging template strand forms a single-stranded D loop. After dissociation of polymerase I, subsequent elongation is carried out by DNA polymerase III. In the presence of chloramphenicol or mutations in polymerase III, strand switching, presumably by polymerase I, can occur *in vivo* up to 600–900 bp from the origin of replication (24). The plasmids used herein contain the triplet repeats cloned at the pUC19 polylinker, which is about 400 bp from the origin of replication. Thus, the instability observed in this system may result from the potential for a strand switch by DNA polymerases and/or from replication errors at DNA secondary structures formed within the initial D loop.

Insertion into the multiple cloning sites in pUC-based vectors apparently provides a particularly unstable system for studying the stability of triplet repeats. When inserted further from the origin in pUC19 or when inserted into other plasmids, triplet repeats are considerably more stable (6). The differential stability of triplet repeats due to origin proximity may be one factor that contributes to triplet repeat instability in humans if different replication origins are used early in development or in differentiated somatic cells.

CTG triplet repeats appear to be most unstable when the CTG strand is maintained as a single-stranded region. Perhaps the formation of DNA secondary structures such as a $(CTG)_n$ hairpin (Fig. 5) within these regions is responsible for replication errors leading to deletion of triplet repeats. The instability of the $(CTG)_n$ tract in the leading strand in MMR^+ cells may result from the generation of a long single-strand gap

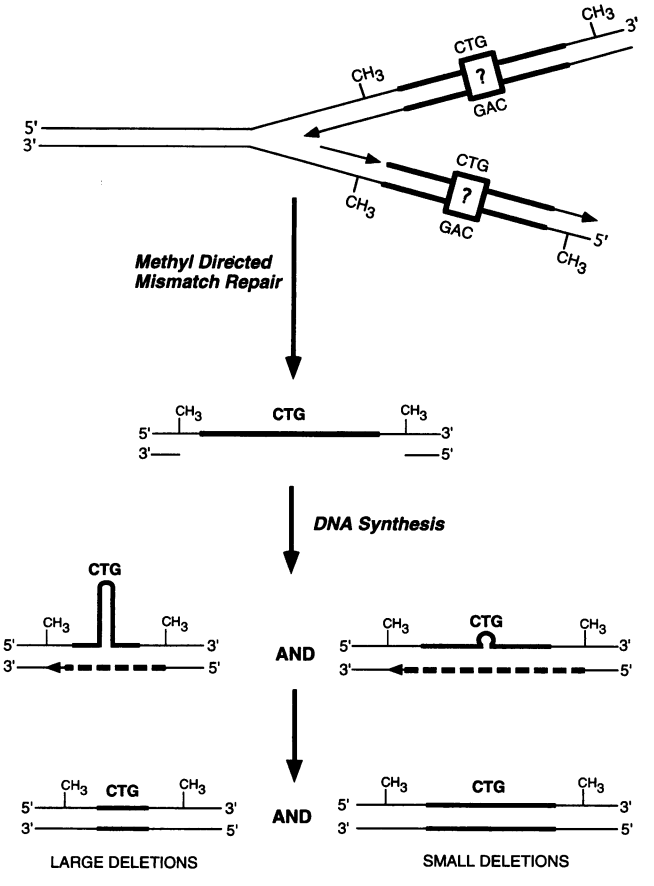

FIG. 5. Model for the involvement of *E. coli* MMR proteins in the enhancement of destabilization of $(CTG)_n$ triplet repeats *in vivo*. The bold boxes containing question marks represent the non-B-DNA structures or small loops recognized by the MMR system. The occurrence of long CTG repeats in a single strand during mismatch repair (in orientation I as shown) or during replication of the lagging strand in orientation II can lead to DNA secondary structures that may promote deletions.

between the *dam* recognition site and the site of mismatch recognized by the *MutHLS* system (Fig. 5). The MMR system may be activated by a three-base loop generated by replication misalignment by a single triplet. The 600-bp-long single-strand template containing $(CTG)_{180}$ may be more prone to form a stable hairpin structure or other DNA secondary structure *in vivo* than shorter sequences [i.e., $(CTG)_{17}$] which are stable in all strains (6). Large hairpin loops may be deleted by slipped misalignment during the resynthesis of duplex DNA. Inactivation of the methyl-directed MMR system, conditions that should prevent generation of large gaps in the leading strand, led to a significant decrease in large deletions. In MMR^- cells, the formation of longer products and smeared deletion products may be explained by numerous short slippages of DNA polymerase during replication (3–5). In the unstable orientation II, the CTG strand is the lagging-strand template and there was little influence of the MMR system on the stability of CTG repeats. When present as an unreplicated lagging strand, the CTG strand is predominately single stranded. Presumably, the presence of a single strand containing CAG repeats does not lead to deletions as frequently as the presence of a single strand containing CTG repeats.

We have employed a biochemical assay to detect large deletions of $(CTG)_n$ or $(CCG)_n$ tracts in MMR^+ and MMR^- *E. coli*. Using this assay, we detect fewer deletions in the repeats when the plasmids containing the repeats are propagated in MMR^- bacteria. Our current model of triplet repeat instability states that large deletions occur when the $(CTG)_n$-containing strand is single stranded and folds into a non-B-DNA structure. When the CTG repeat is in the lagging strand (6), or there is a loss of functional single-strand DNA-binding protein (W.A.R., A.J., S.K., S. Kramer, J. Larson, D. Giedvoc, R.D.W., and R.R.S., unpublished work), more single-strand DNA should be present during DNA replication, increasing the probability of formation of a non-B-DNA structure. In MMR^+ cells, small slippages will activate repair, producing a single-stranded region that has the opportunity to fold into a hairpin. The loss of MMR function would decrease the opportunity for these large deletions, which are detected in our biochemical assay. Genetic assays in *E. coli* and yeast (25, 26) clearly demonstrated increased instability of dinucleotide repeats in MMR^- strains. In these cases, mutations were detected in the absence of repair of misalignment. We cannot address the frequency of small slippage events with our assay, but they may occur at elevated levels, consistent with the results in *E. coli* and yeast (25, 26) to activate the repair function in the MMR^+ cells. In summary, our results may not be inconsistent with the increased genetic instability reported in yeast and *E. coli* because we are measuring different types of mutations.

The types of mismatches formed during replication of the long CTG repeats are uncertain. In general, *E. coli* MMR proteins cannot recognize mispairs longer than four bases, but the correction of small heterologous loops of up to four bases has been reported (27). Presumably, three base loops formed during replication by misalignment of the template and its corresponding progeny strand would be recognized by the MutS and MutL proteins. Larger misalignments would not likely be recognized by these proteins. It is not known if MutS and MutL would recognize mismatches within a $(CTG)_n$ or $(CAG)_n$ hairpin stem. Unlike their effects on the CTG sequence, CGG repeats were not stabilized by MMR^- strains. The reason for this is unclear but may be due to the rate of formation and the stability of hairpins in CTG and CGG triplet repeats as designated by question marks in Fig. 5.

Mutations of the *E. coli mutS* and *mutL* genes as well as their analogs in yeast, mammalian, and human cell lines result in a few-hundred-fold enhancement in the frequency of mutations, depending on the nature of the sequence being studied (7–10, 26, 28). Our demonstration that the bacterial MMR system is an important component determining both the frequency and types of deletions in CTG repeats suggests that eukaryotic MMR proteins, which are similar to the *E. coli* enzymes, may play an analogous role. Several endometrial and colorectal carcinoma cell lines which are defective in MMR show instability of simple DNA microsatellites (11–15). Elucidation of the genetic and biochemical mechanisms of repeat instability in a bacterial cell may provide useful clues for human intervention strategies.

We thank J. E. Larson for technical assistance. This work was supported by grants from the National Science Foundation (DMB-9103942) and the Robert A. Welch Foundation to R.D.W. and by grants from the National Institutes of Health (ES5508) to R.R.S. and (GM52982) to R.D.W. and R.R.S.

1. Nelson, D. L. (1993) in *Genome Analysis: Genome Rearrangement and Stability*, eds. Davies, K. E. & Warren, S. T. (Cold Spring Harbor Lab. Press, Plainview, NY), Vol. 7, pp. 1–24.
2. Willems, P. J. (1994) *Nat. Genet.* **8,** 213–215.
3. Wells, R. D. & Sinden, R. R. (1993) in *Genome Analysis: Genome Rearrangement and Stability*, eds. Davies, K. E. & Warren, S. T. (Cold Spring Harbor Lab. Press, Plainview, NY), Vol. 7, pp. 107–138.
4. Richards, R. I. & Sutherland, G. R. (1994) *Nat. Genet.* **6,** 114–116.
5. Eichler, E. E., Holden, J. J. A., Popovich, B. W., Reiss, A. L., Snow, K., Thibodeau, S. N., Richards, C. S., Ward, P. A. & Nelson, D. L. (1994) *Nat. Genet.* **8,** 88–94.
6. Kang, S., Jaworski, A., Ohshima, K. & Wells, R. D. (1995) *Nat. Genet.* **10,** 213–218.
7. Modrich, P. (1995) *Science* **266,** 1959–1960.
8. Kolodner, R. D. & Alani, E. (1994) *Curr. Opin. Biotech.* **5,** 585–594.
9. Karran, P. & Bignami, M. (1994) *BioEssays* **16,** 833–839.
10. Cleaver, J. E. (1994) *Cell* **76,** 1–4.
11. Umar, A., Boyer, J. C. & Kunkel, T. A. (1994) *Science* **266,** 814–816.
12. Fishel, R., Ewel, A., Lee, S., Lescone, M. K. & Griffith, J. (1994) *Science* **266,** 1403–1405.
13. Shibata, D., Peinado, M. A., Jonov, Y., Malkhosyan, S. & Perucho, M. (1994) *Nat. Genet.* **6,** 237–281.
14. Nicolaides, N. C., Papadopulos, N., Liu, B., Wei, Y., Carter, K. C., Ruben, S. M., Rosen, C. A., Haseltine, W. A., Fleischman, R. D., Fraser, C. M., Adams, M. D., Venter, J. C., Dunlop, M. G., Hamilton, S. R., Peterson, G. M., de la Chapelle, A., Vogelstein, B. & Kinzler, K. W. (1994) *Nature (London)* **371,** 75–80.
15. Parson, R., Li, G.-M., Longley, M. J., Fang, W.-H., Papadopulos, N., Jen, J., de la Chapelle, A., Kinzler, K. W., Vogelstein, B. & Modrich, P. (1993) *Cell* **75,** 1227–1236.
16. Umar, A., Boyer, J., Thomas, D. C., Nguyen, D. C., Risinger, J. I., Boyd, J., Ionov, Y., Perucho, M. & Kunkel, T. A. (1994) *J. Biol. Chem.* **269,** 14367–14370.
17. Sinden, R. R. (1994) *DNA Structure and Function* (Academic, San Diego).
18. Marinus, M. G. (1973) *Mol. Gen. Genet.* **127,** 47–55.
19. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab. Press, Plainview, NY), 2nd Ed.
20. Davis, B. D., Dulbecco, R., Eisen, H. N. & Ginsberg, H. S. (1990) *Microbiology* (Lippincott, Philadelphia), 4th Ed., pp. 124–125.
21. Luria, S. E. & Delbrück, M. (1943) *Genetics* **28,** 491–511.
22. Lieb, M. & Rehmat, S. (1995) *J. Bacteriol.* **177,** 660–666.
23. Modrich, P. (1991) *Annu. Rev. Genet.* **25,** 229–253.
24. Backman, K., Betlach, M., Boyer, H. W. & Yanofsky, S. (1979) *Cold Spring Harbor Symp. Quant. Biol.* **43,** 69–76.
25. Levinson, G. & Gutman, G. A. (1987) *Nucleic Acids Res.* **15,** 5323–5338.
26. Strand, M., Prolla, T. A., Liskay, R. M. & Petes, T. D. (1993) *Nature (London)* **365,** 274–276.
27. Parker, B. O. & Marinus, M. G. (1992) *Proc. Natl. Acad. Sci. USA* **80,** 1730–1734.
28. Aquilina, G., Hess, P., Branch, P., MacGeoch, C., Casciano, I., Karran, P. & Bignami, M. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 8905–8909.