Expansion and Deletion of Triplet Repeat Sequences in Escherichia coli Occur on the Leading Strand of DNA Replication*

Ravi R. Iyer and Robert D. Wells‡

From the Center for Genome Research, Institute of Biosciences and Technology, Texas A & M University, Department of Biochemistry and Biophysics, Texas Medical Center, Houston, Texas 77030-3303

Expansions and deletions of triplet repeat sequences that cause human hereditary neurological diseases were previously suggested to be mediated by the formation of DNA hairpins on the lagging strand during replication. The replication properties of CTG-CAG, CGG-CCG, and TTC-GAA repeats were studied in Escherichia coli using an in vivo phagemid system as a model for continuous leading strand synthesis. The repeats were substantially deleted when the CTG, CGG, and GAA repeats were the templates for rolling circle replication from the f1 phage origin. The deletions may be mediated by hairpins formed by these repeat tracts. The distributions of the deletion products of the CTG-CAG and CGG-CCG tracts indicated that hairpins of discrete sizes mediate deletions during complementary strand synthesis. Deletions during rolling circle synthesis are caused by larger hairpins of specific sizes. Thus, most deletion products were of defined lengths, suggesting a preference for specific hairpin intermediates. Small expansions of the CTG-CAG and CGG-CCG repeats were also observed, presumably due to the formation of CTG and CGG hairpins on the nascent complementary strand. Since rolling circle replication has been established in vitro as a model for leading strand synthesis, we conclude that triplet repeat instability can also occur on the leading strand of DNA replication.

The genetic instability of three triplet repeat sequences (TRS), CTG-CAG, CGG-CCG, and TTC-GAA, has been shown to result in approximately 12 hereditary neurological diseases (1–3) including myotonic dystrophy (4), Kennedy’s disease (5), fragile X disease (6), and Friedreich’s ataxia (7). These diseases are inherited in a non-Mendelian fashion by a phenomenon called “anticipation,” which is characterized by an increase in severity and a decrease in age of onset from one generation to the next. The anomalous expansion of TRS has been identified as the molecular basis for anticipation (1). Triplet repeat tracts have been shown (19–21) to adopt compact secondary structures in vitro and in vivo. In vitro expansion and deletion of the CTG-CAG repeats is mediated by hairpin formation by the CTG repeats on the lagging strand during DNA replication. This model was based on two findings. First, single-stranded phage replication from the filamentous f1 and M13 viral origins was characterized both in vitro and in vivo (19–21). In these phages, the unidirectional synthesis of the Watson and the Crick strands of the DNA takes place by different mechanisms. First, the single-stranded (+)-strands are converted to the double-stranded replicative form (RF) by complementary strand synthesis. Second, the RF is replicated by a rolling circle mechanism to yield (+)-strands. Remarkably, both steps occur by continuous strand synthesis and are mediated by the absence of discontinuous Okazaki fragments. Therefore, due to the absence of any step analogous to lagging strand synthesis, rolling circle replication has been exploited in in vitro replication studies as a model to simulate the synthesis of the leading strand (22, 23).

We proposed (9) that the in vivo expansion and deletion of the CTG-CAG repeats is mediated by hairpin formation by the CTG repeats on the lagging strand during DNA replication. This model was based on two findings. First, single-stranded TRS have been shown (24–28) to adopt compact secondary structures in vitro (see below). Second, replication-dependent deletions between direct repeats occur due to secondary structure formation preferentially on the discontinuous lagging strand (29), because the lagging strand template is more likely to transiently be in a single-stranded state compared with the leading strand. However, the role of the leading strand in deletion formation has not been well studied. Kang et al. (9) could not distinguish between TRS instabilities on the leading strand from those on the lagging strand, because the two strands are synthesized concurrently in the CoE1 replication system. Therefore, we utilized the filamentous phagemid system (30) to dissect the replication fork and focus on the continuous leading strand synthesis of TRS in vivo.

A substantial body of evidence (25, 31, 32) indicates that the genetic instability of TRS is derived from their intrinsically biophysical properties. In vitro studies on short single-stranded oligonucleotides containing TRS showed their capacity to adopt
Continuous Strand DNA Synthesis of Triplet Repeat Sequences

Purification and Characterization of Single-stranded DNA—E. coli NM522 was transformed with the appropriate plasmid by electroporation and plated on LB Agar plates (1% Bacto-tryptone, 0.5% yeast extract, 1% NaCl, 1.5% agar) containing 100 μg/ml ampicillin. Single-stranded phagemids were isolated from infected E. coli colonies prepared by electroporation and then inoculated into 1 ml of LB (1% Bacto-tryptone, 0.5% yeast extract, 1% NaCl) and grown at 37 °C with shaking at 250 rpm. When the cells had grown to an absorbance (600 nm) of 0.2 units, the tubes were inoculated with helper phage M13K07 at a multiplicity of infection of 100 (50). The single-stranded DNA was then isolated by standard dideoxy chain termination sequencing (50). Control cultures of E. coli NM522 uninfected with M13K07 were unable to grow in the presence of 100 μg/ml kanamycin, thus validating the use of kanamycin selection for the continued presence of M13K07.

Propagation of Triplet Repeat-containing Phagemids in E. coli—Phagemids containing undeleted TRS tracts were prepared as described previously (10) and transformed into competent E. coli NM522 cells, which were preinfected with M13K07 helper phage and maintained by kanamycin selection. The transformation mixture was inoculated into 10-ml LB tubes (containing ampicillin and kanamycin, both at 100 μg/ml) at a cell density of 10^7 cells/ml. The cultures were grown at 37 °C with shaking at 250 rpm. When the cultures reached an absorbance (600 nm) of 1.0 unit (20–24 h), an aliquot was inoculated into a fresh tube of 10 ml LB (with kanamycin and ampicillin as before) at a final dilution of 1 × 10^{-7}. The original culture was harvested, and double-stranded RF DNA was isolated by standard alkaline lysis procedures (50). The cultures were then maintained in log phase by repeated recultivation (defined in the figure legends as number of recultivations). The phagemid RF DNA was stored at -70°C.

RESULTS

Instability of (CTG+CAG)_{75} Depends on the Orientation of the Sequence with Respect to the f1 Origin of Replication—In vivo growth of plasmids containing CTG-CAG triplet repeats in E. coli was performed by digesting pGEM3zf^+ and pGEM3zf~ with EcoRI and BspMI, followed by filling in the cohesive ends as described above. The linearized, blunt ended vector was purified by elution from a 1% agarose gel. The vector and insert were mixed and ligated by the addition of 10 units of T4 DNA ligase for 16 h at 16 °C. The ligation mixture was then electroporated into E. coli. coli NM522 by electroporation. Plasmid DNA was isolated from individual transformants and characterized by restriction mapping and sequencing of both insert strands.

Experimental Procedures

Strains and Help Phage—E. coli host strain NM522 (F laclacZlacI5 proAB lacI5 proE thi lac-proAB thi supE44 metB5 (rpsL15 gyrB21) [lacI5 proE thi lac-proAB thi supE44 metB5]) and helper phage M13K07 (Kam5) were purchased from Promega Corp. E. coli NM522 is proficient for homologous recombination.

Cloning Vectors, Oligonucleotides, and Probes—The phagemid cloning vectors pGEM3zf^+ and pGEM3zf~ (3199 bp), which carry the f1 origin of replication in addition to the CoEl origin, were purchased from Promega Corp. The f1 origin is in opposite orientations in pGEM3zf+ and pGEM3zf~. Sequences containing CAG, CGG, and TTC triplet repeats that are prone to form hairpins are present in the template for rolling circle replication. Because f1 replication is template for rolling circle replication. Because f1 replication is orientation-dependent behavior of these repeats in replication (9). The ability of the CCG repeats to form secondary structures in vitro also differs significantly from the complementary CCG repeats (25, 27). Whereas short CCG repeats form hairpins (42, 43) or tetraplexes (38, 39), the CCG repeats exclusively form hairpins (36, 44, 45). However, there is not a consensus in the literature regarding the relative stabilities of the structures formed by the CCG and the CCG repeats. The secondary structures formed by the TRS tracts have been shown to impede the progression of DNA polymerases in vitro (39, 46, 47) and in vivo (48).

Herein, we show that instability of the CTG-CAG repeats in an in vivo single-stranded phage replication system depends on the orientation of the repeats with respect to the f1 replication origin. Substantial deletions were observed when the CTG repeats that are prone to form hairpins are present in the template for rolling circle replication. Because f1 replication is characterized by the absence of a discontinuous lagging strand, our data show that deletion of TRS tracts can occur on the leading strand of DNA replication in E. coli. We also observe expansion of CGG-CCG and CTG-CAG repeats during complementary strand synthesis in vivo.
**Phagemids containing triplet repeats**

Inserts containing triplet repeat sequences of different lengths were cloned into the phagemid vectors pGEM3Zf+ and pGEM3Zf−. Orientation is defined with respect to the double-stranded unidirectional ColE1 origin of replication. Hence, sequences in orientation I have the CAG, CCG, or TTC tracts comprising the lagging strand template. Sequences in orientation II have CTG, CGG, or GAA tracts comprising the lagging strand template. The number of repeats in the (GAA/TTG)26 and (CTG/CTG)27 tracts has an error of ±5 repeats. In all other cases, the exact repeat number has been determined by DNA sequencing. Replication of the phagemids from the f1 origin was confirmed by the purification of single-stranded DNA from the phagemid particles in the culture supernatant. The single-stranded DNA was characterized by sequencing.

| Phagemid name | Triplet repeat | Orientation |
|---------------|----------------|-------------|
| pGEM3Zf+ clones | | |
| pRW3711 | (CTG)27ACT | I |
| pRW3711 | (CTG-CAG)175 | I |
| pRW3539 | (CTG-CAG)175 | II |
| pRW3511 | (CTG-GCC)42 | II |
| pRW3517 | (CTG-GCC)42 | II |
| pRW3543 | (TTC-GAA)30 | II |
| pRW3545 | (TTC-GAA)30 | II |
| pGEM3Zf− clones | | |
| pRW3712 | (CTG-CAG)175 | II |
| pRW3540 | (CTG-CAG)175 | II |
| pRW3512 | (CTG-GCC)42 | II |
| pRW3518 | (CTG-GCC)42 | II |
| pRW3544 | (TTC-GAA)30 | II |
| pRW3546 | (TTC-GAA)30 | II |

**TABLE I**

Phagemids containing triplet repeats

E. coli has revealed the involvement of DNA replication in the instabilities associated with these sequences (9). It was hypothesized that expansions and deletions occur due to formation of hairpin structures by CTG repeats on the lagging strand template (28). This was based on the earlier finding that the discontinuous synthesis of Okazaki fragments on the lagging strand increases its probability of existing in a single-stranded state relative to the leading strand template, thereby resulting in preferential mutagenesis of the lagging strand (29). The replication of these sequences in a system where leading and lagging strands can be dissected would therefore be an attractive way to test this hypothesis. Mechanistically, the replication from the filamentous phage f1 origin can be distinguished into two stages (19–21). First, the single-stranded (+)- strand template is converted to double-stranded RF by a continuous complementar strand synthesis step. Second, the double-stranded RF is replicated via a rolling circle replication step to yield single-stranded (+)-strands. The rolling circle replication is also continuous and is analogous to leading strand synthesis (22, 23). The absence of a discontinuous strand synthesis step involving Okazaki fragments is a characteristic feature of replication from the f1 origin (51).

In order to delineate the instabilities that occur due to the formation of hairpin structures during rolling circle replication (and therefore, by extension, during leading strand synthesis) from those that occur due to similar structures during complementary strand synthesis, we established an in vivo phagemid replication system for TRS. The phagemids pGEM3Zf+ and pGEM3Zf− were used, which carry the origin of replication from the filamentous phage f1 oriented oppositely in the two phagemids. This facilitates the replication of the top strand by pGEM3Zf− and the bottom strand by pGEM3Zf+ when rolling circle replication is induced. Replication from the f1 origin can be induced by the use of a helper phage such as M13K07, which infects E. coli host strains that have conjugal F pili (52). The host strain NM522 was chosen because it carries an F factor and thus has the F pili (53). When grown in E. coli NM522 in the presence of helper phage M13K07, the phagemids are replicated from the f1 origin by rolling circle and complementary strand synthesis (54). pRW3711 and pRW3712 (Fig. 1) contain (GCT)175ACT-(GCT)175ACT(GCT)175 (referred to as (CTG-CAG)175 for convenience) cloned into the polylinker of pGEM3Zf+ and pGEM3Zf−, respectively. We propagated phagemids pRW3711 and pRW3712 in log phase in E. coli NM522 in the presence of the helper phage M13K07 as described under “Experimental Procedures.” To confirm that the phagemids were being replicated from the f1 origin, the single-stranded (+)-strands were purified from the supernatant and characterized by dideoxy chain termination sequencing (data not shown). The E. coli NM522 cultures carrying pRW3711 and pRW3712 were maintained in log phase growth by repeated recultivation. After each recultivation, the cultures were harvested, and the double-stranded phagemid DNA were isolated. The DNAs were digested, end-labeled, and analyzed on 1% agarose and 5% polyacrylamide gels. The agarose gels were blotted onto nylon membranes and hybridized to a radiolabeled probe containing 15 repeats of CTG in order to verify the identities of the putative deletion products. It was observed that the CTG-containing probe hybridized to the band containing the full-length TRS as well as to the bands of shorter lengths (data not shown). An analysis of the deletion products was performed by electrophoresing the digested DNA on polyacrylamide gels (Fig. 2A).

Upon quantitation of the deletion products, a substantial difference was observed in the stability of the TRS between pRW3711 and pRW3712 in the presence of helper phage M13K07 (Fig. 2C). In the case of pRW3711, the full-length (175 repeats) TRS had been completely deleted to a TRS tract 20–30 repeats in length by the third recultivation. On the other hand, the TRS tract in pRW3712 had up to 55% undeleted full-length TRS remaining even after five recultivations in log phase. When pRW3711 and pRW3712 were propagated in E. coli NM522 in the absence of M13K07 helper phage, there was no significant deletion of their TRS tracts in both cases even after five recultivations (data not shown).

These experiments show that the difference in stability between pRW3711 and pRW3712 is dependent on the f1 origin, because the two plasmids are identical in all respects except for the orientation of the f1 origin. Furthermore, since the instabilities are observed only in the presence of helper phage, replication from the f1 origin is required for the stability differences in the TRS tracts between these two plasmids.

**Instability of (CTG-CAG), Onset of the Length of the TRS Tract**—The severity and the lower age of onset of triplet repeat diseases have been correlated with an increase in the length of the TRS tract in certain genes in patients (1–3, 28, 55). The
biochemical properties of CTG such as their ability to bind nucleosomes (56, 57), to block the progression of DNA polymerases (46, 47), and to circularize due to their inherent flexibility (58, 59) have been shown to be dependent upon their length. Furthermore, the stability of CTG in plasmids in E. coli is also dependent on the length of the repeat tract (9–11, 60). In order to determine the effect of CTG length on their genetic instability in our phagemid replication system, plasmids pRW3111 and pRW3121 (46) were used, which contain (GCT)27ACT(GCT)102 (referred to as (CTG)130 for convenience) cloned into pGEM3Zf+ and pGEM3Zf−, respectively (Table I). These plasmids were propagated in E. coli NM522 in log phase, and the RF DNA was isolated after each recultivation. The CTG tract was excised from the RF DNA using EcoRI and HindIII, labeled with [α-32P]dATP, separated on 5% acrylamide gels, and exposed to x-ray film. The lanes numbered 1–5 contain DNA isolated from cultures harvested after 1–5 recultivations. The arrow indicates the band containing the full-length triplet repeat. The deletion products migrate in the region encompassed by the open box. All (CTG-CAG)n-containing restriction fragments also contain 116 bp of nonrepetitive flanking sequence. A shows the analysis of the deletion products from the growth of pRW3711 and pRW3712, which contain (CTG-CAG)175. The difference in the electrophoretic mobilities of the TRS-containing fragments is due to the difference in size (see legend to Table I). B shows a similar analysis of deletion products from the growth of pRW3311 and pRW3312, which carry (CTG-CAG)130. C, the extents of the instabilities of (CTG-CAG)130 and (CTG-CAG)175 were measured by exposing the dried 5% acrylamide gels from the recultivation experiments with pRW3111 (○) (top curve), pRW3121 (○) (next to top curve), pRW3711 (○), and pRW3712 (○) (shown in Fig. 2, A and B) to a Molecular Dynamics PhosphorImager screen followed by scanning. The amount of radioactivity (as estimated by the signal intensity) in the band corresponding to the full-length TRS was measured as a proportion of the total radioactivity in the lane below the band. This was taken as the percentage of molecules in the sample that contained an undeleted TRS tract. The percentage of undeleted TRS was plotted on the y axis against the number of recultivations. The data were computed as an average of three separate experiments. The error bars indicate the S.D. values. The curves were drawn by connecting the points manually on the program Canvas 5.0 (Deneba Software Inc.) using the Bezier curve tool.

![Figure 2](image_url)
the (CTG-CAG)130 tract in pRW3111 and pRW3121 showed the following properties. First, the (CTG-CAG)130 tract was almost completely stable (<10% formation of deletions) even after five recultivations when replicated from the f1 origin. This is in sharp contrast to the instability observed in the case of pRW3711 and pRW3712. Second, there was no difference between the stabilities observed for pRW3111 and pRW3121, as observed for pRW3711 and pRW3712. This clearly shows that the in vivo processes responsible for the deletion of the (CTG-CAG)135 tract do not destabilize the shorter (CTG-CAG)130 tract. Thus, the length of the CTG-CAG tract determines its instability in this filamentous phage replication system.

Reversing the Orientation of (CTG-CAG)n in the pGEM3Zf+ and pGEM3Zf− System Results in the Reversal of the Instabilities—The differences observed between the stabilities of pRW3711 and pRW3712 could be attributed to the fact that the rolling circle template was composed of CTG repeats in pRW3711 and CAG repeats in pRW3712. Therefore, hairpins formed by CTG in the rolling circle template could be bypassed during DNA replication, thus resulting in substantial deletion formation. However, there was also the possibility that derivatives of pGEM3Zf+ (pRW3711) were more unstable than derivatives of pGEM3Zf− (pRW3712) due to an inherently greater instability of sequences cloned in pGEM3Zf+ than those cloned in pGEM3Zf−. In order to conclusively show that the deletions are directed not by the vector sequence but by the sequence of the triplet repeat that forms the template strand (CTG or CAG), we inverted the orientation of the triplet repeat tract in pRW3711 and pRW3712. Upon the inversion of the triplet repeat orientation, the rolling circle template of the pGEM3Zf+ derivative would contain the CAG repeats, while that of the pGEM3Zf− derivative would contain the CTG repeats. We hypothesized, therefore, that the pGEM3Zf+ derivative would be more stable in the in vivo filamentous phage replication system than the pGEM3Zf− derivative.

Hence, we constructed phagemids pRW3539 and pRW3540 (Fig. 3), where the (CAG-CTG)175 tract was cloned into pGEM3Zf+ and pGEM3Zf−, respectively, in orientation II (see “Experimental Procedures”). The phagemids were then propagated in E. coli NM522 in log phase for five successive recultivations and the RF DNA was analyzed by restriction digestion, end labeling, and electrophoresis through 5% polyacrylamide gels (Fig. 4A). Upon quantification of these gels, it became evident that pRW3540 is more unstable than pRW3539 (Fig. 4B) when replicated from the f1 origin. Thus, while the (CTG-CAG)175 tract in orientation I is less stable in pRW3711 than in pRW3712 (Fig. 2), the (CAG-CTG)175 tract in orientation II is more stable in pRW3539 than in pRW3540 (Fig. 4). Thus, a reversal of the orientation of the CTG-CAG tract relative to the f1 origin results in the reversal of the instabilities.

This experiment clearly shows that the instability of the repeat tract in this filamentous phage replication system is determined by the triplet repeat composition of the template strand during rolling circle and complementary strand synthesis stages. The CTG repeat has a higher propensity to adopt fold-back structures than the CAG repeat (33, 34). However, methodology does not exist to enable the analysis of hairpin loops in vivo (25). In this replication system, instability of the CTG-CAG tract is high when the CTG tract is the template during rolling circle replication. Therefore, our data implicate the rolling circle step of the phagemid replication cycle to be where fold-back structures are formed by the CTG repeat. This results in the deletion of the CTG-CAG repeat tract observed in pRW3711 and pRW3540.

Figure 3. Orientation of (CAG-CTG)175 in recombinant phagemids. Phagemids pRW3539 and pRW3540 contain (CAG-CTG)175 cloned in the opposite orientation (orientation II) compared with pRW3711 and pRW3712 (Fig. 1). Both pRW3539 and pRW3540 contain the TRS in the same orientation (II) with respect to the ColE1 origin but in opposite orientations with respect to the f1 origin. When replication is initiated at the f1 origin, pRW3539 yields a plus strand that contains CTG repeats, whereas pRW3540 yields a plus strand that contains CAG repeats.

Figure 4. In vivo instability of (CAG-CTG)175. A, pRW3539 and pRW3540 were analyzed as in the legend to Fig. 2A. B, the instabilities of pRW3539 (■) and pRW3540 (▲) were quantitated as described in the legend to Fig. 2C from three separate recultivation experiments. The data treatment is also as described in the legend to Fig. 2C.
Instabilities of (CGG-CCG)_{81} Depend on the Orientation of the f1 Origin and on the Length of the Repeat Tract—In addition to the CTG-CAG repeats, the instabilities of the CGG-CCG repeats (12) and the TTC-GAA repeats (13, 61) have been shown to be dependent on their orientations relative to the ColE1 unidirectional origin of duplex plasmid replication. Shimizu et al. (12) propagated CGG-CCG repeat tracts cloned in pUC19 in a variety of E. coli host strains and observed a clear effect of the orientation of the repeats relative to the origin of replication on their instability. They showed that deletions predominate when the CGG repeat occurs on the template of the lagging strand. This is consistent with the in vitro evidence that CGG repeats can adopt fold-back structures (43).

In order to determine if the f1 orientation affects the stabilities of these sequences, we investigated the properties of pRW3517 and pRW3518 (Table I) in the filamentous phage replication system. These phagemids were constructed by cloning a tract of (CGG-CCG)_{81} originally from the human FRAX A gene from patients with the fragile X disease (12) into pGEM3Zf^{+} and pGEM3Zf^{-}. The two phagemids were propagated in E. coli NM522 in log phase for five recultivations in the presence of helper phage M13K07. The RF DNA was isolated and analyzed as described above. The data treatment is also as described above. The full-length TRS tract containing (CGG-CCG)_{81} is represented by ●. The deletion products identified from pRW3517 were (CGG-CCG)_{20} (●) and (CGG-CCG)_{6} (□). C, the instabilities of pRW3518 were quantitated as described above. The data treatment is also as described above. The full-length TRS tract containing (CGG-CCG)_{81} is represented by ■. The deletion products identified from pRW3518 were (CGG-CCG)_{52} (●), (CGG-CCG)_{45} (□), (CGG-CCG)_{39} (□), (CGG-CCG)_{34} (□), and (CGG-CCG)_{27} (○). We estimate that the product sizes are ±1 repeat unit.

**Fig. 5.** In vivo instability of (CGG-CCG)$_{81}$. A, pRW3517 and pRW3518 (Table I) were analyzed as described in the legend to Fig. 2A. All (CGG-CCG)$_{81}$-containing restriction fragments also contain 76 bp of nonrepetitive flanking sequence. B, the instabilities of pRW3517 were quantitated as follows. The extents of the instabilities of (CGG-CCG)$_{81}$ were measured by exposing the dried 5% acrylamide gels from the recultivation experiments with pRW3517 (shown in Fig. 5A) to a Molecular Dynamics PhosphorImager screen followed by scanning. The amount of radioactivity (as estimated by the signal intensity) in the band corresponding to the (CGG-CCG)$_{81}$ tract (■) was measured as a proportion of the total radioactivity in the lane below the band. This was taken as the percentage of molecules containing an undeleted TRS tract. The relative amounts of each of the deletion products of (CGG-CCG)$_{81}$ were also determined by measuring the signal intensity for a band as a proportion of the cumulative signal intensity for all the deletion products in the lane. The deletion products that constituted >4% of total TRS at any stage during the growth were identified and plotted. The percentage of total TRS constituted by each deletion product was plotted on the y axis against the number of recultivations. The curves were drawn as in Fig. 2C. The deletion products identified from pRW3517 were (CGG-CCG)$_{52}$ (●) and (CGG-CCG)$_{6}$ (□). C, the instabilities of pRW3518 were quantitated as described above. The data treatment is also as described above. The full-length TRS tract containing (CGG-CCG)$_{81}$ is represented by ■. The deletion products identified from pRW3518 were (CGG-CCG)$_{52}$ (●), (CGG-CCG)$_{45}$ (□), (CGG-CCG)$_{39}$ (□), (CGG-CCG)$_{34}$ (□), and (CGG-CCG)$_{27}$ (○). We estimate that the product sizes are ±1 repeat unit.
undeleted in pRW3517, as much as 42% of the TRS in pRW3518 was undeleted at the same stage. After six recultivations, there was no detectable undeleted TRS in pRW3517, but 30% of the TRS was undeleted in pRW3518. Thus, the orientation of the f1 replication origin has a substantial influence on the stability of the (CGG-CCG)\textsubscript{61} tract. Furthermore, the phagemid pRW3517 contains the CGG repeat in the minus strand, which forms the template for rolling circle replication. It is therefore likely that the CGG repeats form fold-back structures on the minus strand during rolling circle replication, resulting in a shortened TRS tract.

The instability of the CGG-CCG repeats is dependent on the length of the tract in fragile X patients (6). We showed previously (12) that CGG-CCG repeat tracts are unstable when cloned in plasmids in E. coli in a length-dependent manner. Short tracts containing up to 10 repeats were completely stable, whereas tracts containing more than 24 repeats were increasingly unstable when cloned in pUC19 and propagated in E. coli DH5α. Therefore, we studied the effect of the length of the CGG-CCG repeat sequence on its instability in the filamentous phagemid replication system \textit{in vivo}. pRW3511 and pRW3512 (Table I) were constructed by cloning (CGG-CCG)\textsubscript{6} into pGEM3Zf\textsuperscript{+} and pGEM3Zf\textsuperscript{−}, respectively. Characterization of the double-stranded RF after propagation in E. coli NM522 in the presence of M13K07 over five recultivations revealed that the TRS in both pRW3511 and pRW3512 was highly stable (>95% undeleted) even after five recultivations (data not shown). Also, there was no f1 orientation-dependent difference in the stabilities of pRW3511 and pRW3512. This behavior of the short CGG-CCG tracts is in agreement with the observations of Shimizu et al. (12).

**TRS Deletion Occurs Predominantly during Rolling Circle Replication**—We have observed a substantial influence of replication from the f1 origin on the genetic instability of CTG-CAG and CGG-CCG repeat sequences. For the CTG-CAG repeats (Fig. 1), pRW3711 contains CTG in the (−)-strand and is genetically less stable than pRW3712, wherein CAG is in the (−)-strand (Fig. 2C). For the CGG-CCG repeats (Table I), pRW3517, in which CGG is in the (−)-strand, is more unstable (Fig. 5B) than pRW3518, which carries CCG in the (−)-strand (Fig. 5C). The greater instability of the TRS tracts in the phagemids that contain CTG and CCG repeats in the (−)-strand rolling circle template suggests the following. First, \textit{in vitro} experiments have shown (33, 34, 37) that CTG repeats have a higher propensity to form hairpins than CAG repeats. Therefore, we propose that the rolling circle template is the location for the formation of hairpins by the CTG repeats, resulting in substantial deletions (Fig. 6, left). Also, based on \textit{in vivo} observations, several workers have suggested (12, 48, 62) that CGG repeats of physiologically relevant lengths are more likely to form fold-back structures than CCG repeats. Hence, it can be concluded that the deletions of the CGG-CCG repeats also take place on the rolling circle template due to CGG-containing hairpins. Second, the full-length (CTG-CAG)\textsubscript{175} tract in pRW3712 (Fig. 2) and the (CGG-CCG)\textsubscript{61} tract in pRW3518 (Fig. 5) also exhibit some instability. For the CTG-CAG repeats, the instabilities of pRW3712 are mediated by the formation of hairpins by the CTG repeats during the synthesis of the complementary strand (Fig. 6, right). The deletion of the CGG-CCG repeats in pRW3518 can be mediated by CGG hairpins during the synthesis of the complementary strand. However, it is also possible that some of these deletions are mediated by CGG hairpins on the rolling circle template. Our experimental system does not allow us to distinguish between deletions that occur due to CGG hairpins from those that occur due to CGG hairpins.

In summary, since the phagemid replication system has been widely investigated (22, 23) as a model for leading strand synthesis, our \textit{in vivo} data clearly show that the deletion of TRS can occur on the leading strand of DNA replication. Furthermore, since the synthesis of the complementary strand is continuous, the deletion events observed during this step also support the leading strand deletion model.

**Different Length Hairpins Mediate TRS Deletion by Bypass DNA Synthesis on the Rolling Circle and on the Complementary Strand**—We observed that the direction of f1 replication has a substantial effect on the stability of CTG-CAG and CGG-CCG repeats as evidenced by the quantitation of the extents of instabilities. The products of the instabilities of pRW3517 and pRW3518, which contain a (CGG-CCG)\textsubscript{61} tract, were further
characterized. For pRW3517, the quantitation of the TRS instability is shown in Fig. 5B. The full-length (CGG-CCG)151 gave rise to two major products containing (CGG-CCG)20 and (CGG-CCG)9. The product containing six repeats appears after one recultivation and increases to constitute 22% of total TRS by the second recultivation. On the other hand, the product containing 20 repeats appears only after the third recultivation but constitutes 97% after six recultivations. Fig. 7A (left) shows that the (CGG-CCG)6 and the (CGG-CCG)30, can arise from the (CGG-CCG)151, by the formation of long hairpins containing ~37 and 30 CGG repeats, respectively, in the stem of the hairpin. These deletions probably take place during the conversion of the double-stranded RF to the single-stranded (+) strand by rolling circle replication. Thus, the instability of the (CGG-CCG)151 tract in pRW3517 is mediated by long hairpins that yield relatively few products.

In contrast, an analysis of the deletion products of pRW3518 (Fig. 5C) showed that the (CGG-CCG)18 was deleted to tracts containing 52, 45, 39, 34, and 27 repeats. The full-length (CGG-CCG)151 is gradually deleted until it constitutes approximately 30% of total TRS after six recultivations. The loss of the (CGG-CCG)151 coincides with the appearance of tracts containing the (CGG-CCG)5 and the (CGG-CCG)19 by the third recultivation. Fig. 7B shows that the primary deletion products are double-stranded RFs containing 52 and 39 repeats, which can arise due to the formation of CGG hairpins containing ~14 and 20 repeats, respectively, in the stems. The loss of the (CGG-CCG)32 from the third recultivation to the fourth (Fig. 5C) corresponds to the appearance of a product containing (CGG-CCG)45, suggesting that the (CGG-CCG)15 is a deletion product of the (CGG-CCG)32. This secondary deletion presumably takes place when the (CGG)6 in the single-stranded (+) strand synthesized from the double-stranded RF forms a small hairpin containing three repeats in the stem (Fig. 7B). This hairpin can be bypassed during complementary strand synthesis, resulting in an RF containing (CGG-CCG)15.

Fig. 5C further shows that deletion products containing (CGG-CCG)34 and (CGG-CCG)27 appear late in the growth. Whereas the (CGG-CCG)18 constitutes 13% of total TRS after four recultivations and then diminishes to 5% after six recultivations, the (CGG-CCG)19 arises after four recultivations and finally accounts for 42% after six recultivations. The appearance of the (CGG-CCG)19 coincides with the reduction of both the (CGG-CCG)30 and the (CGG-CCG)11, suggesting that the (CGG-CCG)19 is formed by the deletion of the (CGG-CCG)20 and the (CGG-CCG)11 via different hairpin intermediates containing three or five CGG repeats in the hairpin stem (Fig. 7B). In agreement with previous work (10, 11), we observed that the absolute size of the deletion products of pRW3517 and pRW3518 varied slightly from experiment to experiment, but the overall pattern of deletions was reproducible.

Thus, the deletion of the (CGG-CCG)15 tract in pRW3518 is mediated by a variety of CGG hairpins of different sizes on the complementary strand, which gives rise to numerous deletion products. In sharp contrast, we observe only two major deletion products from pRW3517 as discussed earlier. For the CTG-CAG repeat, we observed differences between the deletion products that arose from pRW3711 and pRW3712 (Fig. 2A). The (CTG-CAG)175 tract in pRW3711 was deleted to major products containing approximately 10 and 22 repeats, presumably mediated by long CTG hairpins containing ~75–82 repeats in the stem during rolling circle replication. On the other hand, we identified at least five major deletion products of pRW3712 (Fig. 2A) containing (CTG-CAG)170, (CTG-CAG)143, (CTG-CAG)130, (CTG-CAG)110, and (CTG-CAG)92. The formation of these products is mediated by CTG hairpins of size ranging from 2 to 70 repeats during complementary strand synthesis.

Therefore, we propose that different length hairpin loops mediate the deletion processes on the rolling circle template and on the complementary strand.

(CTG-CAG)150 and (CGG-CCG) Expand during Replication in Vivo—In addition to deletions, we also detect expansions of the CGG-CCG and the CTG-CAG repeats. For the CGG-CCG repeats, an analysis of the instability products of pRW3517 (Fig. 5B) showed that a tract containing (CGG-CCG)6 appears early in the growth, constitutes 22% of total TRS by the second recultivation, and then disappears to less than 1% of total TRS after six recultivations. Since a total of 76 bp of non-TRS DNA flank the CGG-CCG repeats, it is possible to identify deletion products that contain down to zero repeats. We found that the disappearance of the (CGG-CCG)6 does not coincide with the appearance of smaller deletion products. However, it does correspond to the increase in (CGG-CCG)20, which constitutes 97% of total TRS after six recultivations. Therefore, it is likely that a substantial portion of the (CGG-CCG)6 may have arisen due to an expansion of the (CGG-CCG)6. The right side of Fig. 7A shows that such an expansion may be mediated by a six- or seven-repeat CGG hairpin formed on the newly synthesized complementary strand. In contrast, no putative expansions were observed for the (CGG-CCG)12 tract in pRW3518 (Fig. 5, A and B). For the CTG-CAG repeat, Fig. 2A shows that the (CTG-CAG)175 in pRW3711 is deleted to a 10-repeat tract. The (CTG-CAG)150 decreases to less than 1% of total TRS by the fourth recultivation and appears to contribute to the increase in the (CTG-CAG)20, which accounts for 95% of total TRS after five recultivations. This suggests a possible expansion of the (CTG-CAG)150 to a tract containing (CTG-CAG)20 due to the formation of a CTG hairpin on the nascent complementary strand. The length of the stem of this hairpin intermediate could be five triplet repeats. We were unable to detect any species that could have arisen by expansion from pRW3712 (Fig. 2A).

Thus, although deletions are the predominant instabilities observed, expansions of the CGG-CCG and CTG-CAG repeats were also detected. Whereas essentially all of the (CGG-CCG)6 and the (CTG-CAG)15 were presumably expanded to larger products, the length increases due to these putative expansions were relatively modest. The inability to detect similar putative expansions from pRW3712 and pRW3518 may be due to the absence of sufficiently short TRS tracts that would be prone to expand.

Length and Orientation-dependent Instabilities of (TTC-GAA)150—An expansion of the TTC-GAA repeat in the human frataxin gene results in the autosomal recessive neurological disease Friedreich’s ataxia (7). The TTC-GAA repeats were shown (13, 61) to be unstable, depending on their orientation relative to the unidirectional ColE1 replication origin in pUC19 and pSPL3 vectors when propagated in various E. coli host strains. The TTC-GAA repeats have a propensity to form supercoiled-dependent put·pur·pyr triple helical structures (61). However, the role of these triplexes, if any, in the genetic instability of TTC-GAA sequences is uncertain (13, 61, 63).

In order to understand the effect of the f1 replication on these repeats, we constructed two phagemids, pRW3545 and pRW3546 (Table I), by cloning a tract of (TTC-GAA)150 from the human frataxin gene (13) into pGEM3Zf+ and pGEM3Zf−, respectively (see “Experimental Procedures”). The phagemids were propagated in log phase in E. coli NM522 for five recultivations, and the RF DNA was digested, end-labeled, and analyzed on 5% polyacrylamide gels (Fig. 8A). Quantitative analyses of these gels showed that the (TTC-GAA)150 repeat in pRW3546 is less stable than in pRW3545 (Fig. 8B). Once again, it
is clear that the orientation of the f1 origin has an influence on the stability of the triplet repeat tract. Ohshima et al. (13) showed that if the GAA repeats comprise the template of the lagging strand when replicated from the unidirectional ColE1 origin in E. coli, the plasmid is significantly destabilized. This is in contrast to the opposite orientation, where TTC repeats...
All (TTCGAA)70 were analyzed as described in the legend to Fig. 2 of nonrepetitive flanking sequence. It was previously suggested (9) that the genetic instability of long tracts of CTG-CAG sequences in duplex plasmids in E. coli is mediated by hairpin formation on the lagging strand template during DNA replication. We postulated (9) that the lagging strand as the location of hairpin formation by the CTG repeats based on earlier work (29) that showed that, during DNA replication, palindromic sequences are more likely to be involved in the genetic instability of these repeats cannot be ruled out.

The instability of TTC-GAA repeats in patients of Friedreich's ataxia has been shown to be dependent on the length of the repeat tract (7, 64). Also, the TTC-GAA repeats are unstable in plasmids in E. coli in a length-dependent manner (13). These workers found that whereas (TTC-GAA)70 cloned in pUC19 was completely stable when propagated in E. coli SURE cells, tracts containing (TTC-GAA)150 and (TTC-GAA)270 were highly unstable under the same conditions. Therefore, we studied the effect of the length of the TTC-GAA repeat on its instability in the filamentous phagemid replication system in vivo. pRW3543 and pRW3544 (Table I) were constructed containing (TTC-GAA)70 cloned in pGEM3ZF+ and pGEM3Zf−, respectively. Characterization of the double-stranded RF after propagation in E. coli NM522 in the presence of M13K07 over five recultivations revealed that both pRW3543 and pRW3544 showed the following. First, there was no difference in the stabilities of pRW3543 and pRW3544 over five recultivations (data not shown). Second, even after the fifth recultivation, the (TTC-GAA)70 tract was stably maintained (<5% deleted) (data not shown). This behavior of the shorter TTC-GAA repeat tract is similar to that observed for the (CTG-CAG)130 tract in pRW3111 and pRW3121 (Fig. 2). This is in agreement with the general paradigm of length-dependent instability of TRS (1).

Interestingly, the deletion patterns of the TTC-GAA repeat do not appear to be different for pRW3545 and pRW3546 (Fig. 8A). This is in sharp contrast to the product distribution observed for CTG-CAG and the CCG-CCG repeats discussed earlier. The deletion products of the (TTC-GAA)150 in both pRW3545 and pRW3546 are numerous, suggesting that a variety of small hairpins or slipped structures are responsible for this behavior.

**DISCUSSION**

**Replication-based Deletion of Triplet Repeats**—Our data show that CTG-CAG, CCG-CCG, and TTC-GAA repeats are expanded and deleted during filamentous phage replication in vivo. It was previously suggested (9) that the genetic instability of long tracts of CTG-CAG sequences in duplex plasmids in E. coli is mediated by hairpin formation on the lagging strand template during DNA replication. We postulated (9) that the lagging strand as the location of hairpin formation by the CTG repeats based on earlier work (29) that showed that, during DNA replication, palindromic sequences are more likely to form secondary structures on the lagging strand template than on the leading strand template due to misalignment. Once formed, these secondary structures can be bypassed by the DNA replication complex, resulting in deleted progeny strands.

The discontinuous synthesis of the Okazaki fragments on the lagging strand involves several steps that include new primer synthesis, polymerase cycling to the 3'-OH terminus of the new primer, synthesis of the nascent lagging strand, and termination of Okazaki fragment synthesis (22). Thus, from the moment the DNA is unwound by the helicase until the completion of Okazaki fragment synthesis, the lagging strand template is in the single-stranded state, providing an opportunity for the formation of secondary structures.

In contrast, on the leading strand, the DNA polymerase closely follows the DNA helicase, which unwinds the DNA. In vitro studies on leading strand synthesis (65) showed that the rapid movement of the replication fork depends on protein-protein interactions between the α-subunit of DNA polymerase III and the DnaB helicase. Also, synthesis of the leading strand does not require the presence of SSB (66). Therefore, it was proposed (22, 65) that little or no leading strand template exists freely in the single-stranded state between the DNA helicase and the DNA polymerase.

**Model Systems for the Analysis of the Leading Strand**—Leading strand synthesis has been studied using in vitro replication systems in which rolling circle replication of a tailored duplex DNA template is sustained by T4 (23) or E. coli (66, 67) replication proteins. The advantage of using rolling circle replication to study the leading strand lies in the ability to prevent
Okazaki fragment synthesis in these systems and thus facilitate the study of the leading strand alone. As opposed to the replication of isometric phages like ΦX174, wherein rolling circle replication occurs concurrently with Okazaki fragment synthesis, the filamentous phages f1, fd, and M13 synthesize the (+)-strand by rolling circle replication and the (−)-strand by complementary strand synthesis (19). Therefore, a phagemid replication system that uses an f1 origin is appropriate to dissect the replication fork and investigate the continuous leading strand synthesis in vivo.

A key difference between the f1 and the E. coli replication systems is that these processes utilize different helicases. In E. coli, the hexameric DnaB helicase unwinds the duplex as it tracks along the lagging strand template in the 5′ to 3′ direction in concert with the leading strand polymerase to which it is physically coupled (65, 68). In contrast, f1 replication utilizes the dimeric E. coli Rep helicase, which unwinds the duplex DNA as it tracks along the (−)-strand in the 3′ to 5′ direction (20, 51, 68). Since the Rep helicase and the polymerase function independently on (−)-strand during f1 replication (69), it is possible for substantial single-stranded regions to be created.

Hairpin Formation on the Leading Strand Template—We observed frequent deletions when (CTG)175 constituted the rolling circle template. However, the phagemid was substantially more stable when (CAG)130 was present in the rolling circle template. In order to rule out the possibility that other factors such as plasmid-flanking sequence may have been responsible for the observed differences, we switched the orientations of theTRS in the phagemids pGEM3Zf1 and pGEM3Zf2. The experiments done with these switched phagemids confirmed our earlier conclusion that the presence of (CTG)175 in the rolling circle template was deleterious. We also found that this behavior was dependent on the length of the CTG repeat tract, because (CTG-CAG)130 was completely stable in this system.

Since CTG repeats have been proposed to form stable secondary structures (34), we suspected that the observed deletions occurred due to hairpin formation in the CTG tract on the rolling circle template. To confirm that the rolling circle template was the location for the instabilities, we studied the behavior of (CGG-CGG)13 in this system and found that the presence of the (CGG)13 on the rolling circle template gave substantially more deletions than when (CGG)13 was present in the same location. Since it was postulated (12, 48, 62) that under in vivo conditions and at physiologically relevant lengths, CGG was more likely than CCG to adopt stable secondary structures, we concluded that the rolling circle template was indeed the location for a majority of the deletions. Therefore, by extrapolation, these data show that deletions can occur by hairpin formation on the leading strand. Further, our observation that plasmids carrying TTC-GAA repeats are more unstable when the GAA repeats are present in the rolling circle template supports our contention (13) that the GAA repeats have a higher propensity to adopt compact secondary structures than the TTC repeats.

Based on genetic and biochemical studies on filamentous phage genome organization, it was proposed that the minus strand was ~95% responsible for transmission of genetic information (70, 71). Therefore, the deletion and expansion events that occur on the (+)-strand may be more difficult to detect in the in vivo phagemid replication system. It is possible that the higher instability of TRS due to hairpin formation observed on the (−)-strand reflects this inherent bias. However, since the (+) and the (−)-strands are replicated by continuous DNA synthesis, the deletions and expansions observed during both steps of phagemid replication strongly support the leading strand deletion model.

Polymerase Pausing May Facilitate Hairpin Formation on the Leading Strand—We observe leading strand deletions although it was suggested (9) that secondary structures are less likely to form on this strand due to the absence of significant stretches of single-stranded DNA. In vivo (48) and in vitro (39, 46, 47) experiments show that long tracts of TRS can block the progression of DNA polymerases. In contrast, Hiasa and Marrians (72) studied the in vitro bidirectional and rolling circle replication of CTG-CAG repeats from the oriC replication origin of E. coli and did not detect any impediment to the progression of the replication forks through the triple repeat tract. However, these workers could not rule out the possibility that replication fork stalling did occur at low frequencies that caused deletions and expansions in vivo that were undetectable in their in vitro assays.

Therefore, we speculate that for the in vivo f1 replication system the TRS are able to stall the leading strand polymerase but are unable to block the progression of the Rep helicase on the (−)-strand. Since the polymerase and the Rep helicase are not physically coupled, they could function independently of each other (69). Hence, any retardation of the polymerase by the TRS would have no impact on the progression of the helicase, and a substantial region of single-stranded leading template is created. This single-stranded region would have ample opportunity to adopt secondary structures that could be bypassed by the polymerase, resulting in a deleted TRS tract.

Unlike the rolling circle template, the (−)-strand is single-stranded. Therefore, deletions would be expected to be frequent when CTG or CCG repeats constitute the complementary strand. However, in this case we observe only a modest instability. This could be because SSB rapidly binds to the tail of the (−)-strand immediately after its synthesis by rolling circle replication and removes the secondary structures on the (−)-strand (51), including those formed by the TRS (15). The easy access of SSB to the secondary structures on the (−)-strand results in their expedient removal, thus substantially obviating the deletion process. This agrees with the observations of Roche et al. (15) that the SSB protein enhances the stability of CTG-CAG repeats on plasmids in E. coli. These workers proposed that this behavior was due to the ability of SSB to destabilize the hairpins formed by the CTG repeats on the lagging strand during DNA replication.

Different Length Hairpins Mediate Rolling Circle and Complementary Strand Deletions—Our results for the CGG-CGG and the CTG-CAG repeats show that the deletion process is different during rolling circle replication than during complementary strand synthesis. For the CGG-CGG repeats, when (CGG)13 was the template for rolling circle replication, the TRS was deleted to short tracts containing six and 20 repeats. These deletions were probably caused by long CGG hairpins that contained 30–40 repeats in their duplex stems. The relatively few deletion products observed indicate that these long hairpins are quite homogenous. Alternately, the deletions are formed early in the replication of the phagemids and eventually outgrow the longer TRS tracts (10, 11). In contrast, when the (−)-strand contains (CGG)n, a variety of deletion products ranging from 27 to 52 repeats are observed. These products arose from primary deletions of the (CGG)n and then secondary deletion of the primary products. We believe that the mediators of this process are hairpin structures composed of CGG repeats that vary in size from two to 20 repeats in the stem. Rolling circle replication of the (CTG)n template resulted in extremely few deletion products that were 10 and 20 repeats in length, perhaps caused by the formation of 75–80-repeat-long CTG hairpins. Replication of the (−)-strand resulted in several deletion products that contained 32, 100, 135, 145, and 170
peats can also take place during the phagemid replication cycle
hairpins on the nascent lagging strand during DNA replication.

E. coli
Replication-mediated expan-
apparently smaller hairpin intermediates that mediate them
explain not only the lower extent of deletions but also the
ever, at a low frequency, small hairpins or slipped structures
circle and hence prevent the formation of long hairpins. How-
strand as soon as it is synthesized and reeled off the rolling
circle template. These deletion products covered almost the

Interestingly, the behavior of the TTC
Assortment of TTC or GAA secondary structures on both the
Thus, we propose that these deletions were mediated by an

REFERENCES
1. Wells, R. D., and Warren, S. T. (eds) (1998) Genetic Instabilities and Hereditary Neurological Diseases, Academic Press, San Diego, CA
2. La Spada, A. R. (1997) Brain Pathol. 7, 943–963
3. Parniewski, P., Jaworski, A., and Bacolla, A. (1999) J. Biol. Chem. 274, 14588–14596
4. Mitas, M. (1997) Nucleic Acids Res. 25, 2245–2253
5. Gao, X., Huang, X., Smith, G. K., and Zhang, M. (1998) in Genetic Instabilities and Hereditary Neurological Diseases (Wells, R. D., and Warren, S. T., eds) pp. 647–676, Academic Press, Inc., San Diego, CA
6. Mitas, M., Yu, A., Dill, J., Kamp, T. J., Chambers, E. J., and Haworth, I. S. (1995) J. Biol. Chem. 270, 1050–1059
7. Chen, X., Mariappan, S. V. S., Catarsi, P., Bradbury, R. D., and Gupta, G. (1998) Nucleic Acids Res. 26, 816–823
8. Nadel, Y., Weisman-Shomer, P., and Fry, M. (1995) J. Biol. Chem. 270, 298–305
9. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) Molecular Cloning: A

Acknowledgments—We thank Drs. R. P. Bowater, P. Parniewski, A. Jaworski, and A. Bacolla for helpful discussions and Dr. Kenneth J. Marians for critical comments on the manuscript.
