Genetic Instabilities in (CTG-CAG) Repeats Occur by Recombination*

(Received for publication, November 18, 1998, and in revised form, May 12, 1999)

John P. Jakupciak‡ and Robert D. Wells§

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

The expansion of triplet repeat sequences (TRS) associated with hereditary neurological diseases is believed from prior studies to be due to DNA replication. This report demonstrates that the expansion of (CTG-CAG) in vivo also occurs by homologous recombination as shown by biochemical and genetic studies. A two-plasmid recombination system was established in Escherichia coli with derivatives of pUC19 (harboring the ampicillin resistance gene) and pACYC184 (harboring the tetracycline resistance gene). The derivatives contained various triplet repeat inserts ((CTG-CAG), (CGG-CCG), (GAA-TTC), (GTC-GAC), and (GTG-CAC)) of different lengths, orientations, and extents of interruptions and a control non-repetitive sequence. The availability of the two drug resistance genes and of several unique restriction sites on the plasmids enabled rigorous genetic and biochemical analyses. The requirements for recombination at the TRS include repeat lengths >30, the presence of CTG-CAG on both plasmids, and recA and recBC. Sequence analyses on a number of DNA products isolated from individual colonies directly demonstrated the crossing-over and expansion of the homologous CTG-CAG regions. Furthermore, inversion products of the type [(CTG)13(CAG)67][(CTG)67(CAG)13] were isolated as the apparent result of “illegitimate” recombination events on intrahelical pseudoknots. This work establishes the relationships between CTG-CAG sequences, multiple fold expansions, genetic recombination, formation of new recombinant DNA products, and the presence of both drug resistance genes. Thus, if these reactions occur in humans, unequal crossing-over or gene conversion may also contribute to the expansions responsible for anticipation associated with several hereditary neurological syndromes.

Genetic instabilities (expansions and deletions) of triplet repeat sequences (TRS) (CTG-CAG), (CGG-CCG), or (GAA-TTC)) are a hallmark of certain hereditary neurological diseases (1, 2). Numerous workers in human genetics have proposed DNA replication, gene conversion, recombination, and related processes as the mechanism(s) responsible for these alterations in repeat sequence lengths. Subsequent in vivo studies in genetically tractable systems (1, 3) and in vitro investigations (4) have demonstrated expansions and deletions during DNA replication, probably by slipped strand misalignment due to preferential hairpin formation of TRS. Similar molecular studies in vivo on gene conversion and recombination are lacking.

Several human genetic studies on patient materials reported haplotype analyses, especially related to myotonic dystrophy (DM) and the fragile X syndrome, which implicated gene conversion and/or unequal crossing-over (types of recombination) to genetic instabilities. In the first report, Korneluk and co-workers (5, 6) proposed unequal crossing-over (5) and gene conversion (6) as the mechanisms responsible for the expansions and deletions observed in the (CTG-CAG) mutation during DM transmission. This conclusion was derived from haplotype analyses of six polymorphic markers in the DM region. The TRS was presumed to be the site of the discontinuous gene conversion events.

Second, Nelson and collaborators (7) investigated the loss of mutation at the FMR1 locus through multiple exchanges between maternal X chromosomes. They described a daughter of a female carrier who inherited the fragile X premutation chromosome based on haplotype analyses using flanking markers. The (CGG-CGG) repeat sequence and the intragenic polymorphic marker FMRb showed the normal maternal alleles, whereas two other intragenic markers showed the risk haplotype. Since the other intragenic markers are located between the markers (CGG-CGG) and FMRb, this results in patches of normal and fragile X sequence in the FMR1 gene of the daughter and was explained on the basis of gene conversion. Likewise, prenatal diagnosis of the fragile X syndrome showed a loss of mutation owing to a double recombinant or gene conversion at the FMR1 locus (8).

Third, Brown et al. (9) investigated reverse mutations in the fragile X syndrome as well as founder effects. Based on these haplotype analyses of nearby markers to the (CGG-CGG) repeats, revertants were discovered in a small percentage of the premutation carrier offspring. Gene conversion (9) and recombination (10) were proposed as the responsible mechanisms. Fourth, homologous recombination involving unequal pairing of sister chromatids leading to the formation of a four-stranded synaptic structure was suggested by Wieringa and co-workers (11) as the mechanism to explain the (CTG-CAG) expansion responsible for gonosomal mosaicism in DM patients. Fifth, Warren (12) interpreted the work of Olsen and co-workers (13) on the polyalanine expansion in synpolydactyly to result from unequal crossing-over in the HOXD13 protein gene and the authors agreed (12). Sixth, Kidd and co-workers (14) studied haplotype analyses of the DM locus on a worldwide basis with

* This work was supported by National Institutes of Health Grants GM52982 and NS37554 and the Robert A. Welch Foundation. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ Part of the Genetics Graduate Program at Texas A&M University.
§ To whom correspondence should be addressed: Institute of Biosciences and Technology, Center for Genome Research, Texas A&M University, Texas Medical Center, 2121 W. Holcombe Blvd., Houston, TX 77030. Tel.: 713-677-7651; Fax: 713-677-7689; E-mail: rwell@ibt.tamu.edu.

The abbreviations used are: TRS, triplet repeat sequences; DM, myotonic dystrophy; bp, base pair(s); kb, kilobase pair(s); amp, ampicillin; tet, tetracycline; RB, recombinant band.
emphasize on the implications for the evolution of modern humans and the origin of the DM mutations. Several patterns of haplotype variation and linkage disequilibria were explained on the basis of gene conversion events such as unequal sister chromatid recombination.

For these six cases, it was presumed that the TRS were the sites for the recombination (or gene conversion) events. However, for the majority of the instabilities in TRS involved in human neurological diseases, evidence was not available to specifically identify the molecular mechanism(s) of the processes (i.e., replication, recombination, or repair) (reviewed in Ref. 1). Some workers (15–17) have supported the concept that recombination is not involved in expansion.

During the mapping of the genes for these neurological disorders, substantial linkage analyses were performed utilizing flanking markers, some quite near the repeats. Since the exchange of flanking markers was not generally found (15, 16), simple homologous recombination, as a general mechanism for expansion, has not been favored. Recently, experiments in yeast have tried to elucidate the independent roles of the RAD50–55 family of genes, which are responsible for various pathways of recombination. CTG-CAG sequences were shown (18) to be susceptible to strand breaks, and the occurrence of double-strand breaks was length-dependent. Also, experiments with yeast rad27 strains had augmented instability of the TRS. RAD27 encodes a nuclease involved in Okazaki fragment processing. It is known that the majority of errors that accumulate during crossing-over with or without exchange of flanking sequences is due to the presence of the TRS (19). Herein, we demonstrate that gene conversion or unequal crossing-over with or without exchange of flanking sequences is a powerful mechanism for (CTG-CAG) expansion in Escherichia coli.

EXPERIMENTAL PROCEDURES

Plasmids—The plasmids used in these experiments are derivatives of the unidirectionally replicating pUC19 and pACYC184. The cloning and characterization of plasmids containing the repeat CTG, CAG, and GTC-GAC repeating sequences were described (23–26, 28–30). pRW4100 and pRW4110 were constructed by digesting pRW3481 and pRW3463, respectively, with PvuII and recloning the fragments containing the triplet repeats into the pRSI1 sites of pACYC184. pRW4115 was constructed by digesting pRW3822 with EcoRI and EcoRV and recloning the fragment containing the triplet repeat into the EcoRI and BstEII sites of pACYC184. pRW4105 and pRW4106 were constructed by digesting bacteriophage λ DNA with HindIII and recloning the 564-bp fragment into the HindIII site of pUC19 or by end-filling to give blunt ends and then recloning into the pRSI1 sites of pACYC184. All of the CTG-CAG sequences cloned into pUC19 contain no interruptions but the (CTG-CAG)17, cloned in pACYC184 (pRW3239) contains two G-to-A polymorphisms at the 28th and at the 69th repeat, respectively. The 16 bp of human sequence at the proximal end relative to the interruptions and 43 bp distal to the interruptions. pRW3041 containing (CGG-CCG)27, with two interruptions was constructed as follows: a fragment containing (CGG-CCG)34 was isolated from RN2 (48) by digestion by SmaI and HindII and was then inserted into the pRSI1 sites of pACYC184. pACT-2 was purchased from CLONTECH Laboratories, Inc.

Bacterial Strains—The following E. coli strains were used: AB1157 (47) as a parent of the recombinogenic deficient strains; JC10289 (thr-1, ara-14, leuB6, Dggt-proA62, lacY1, rss-33, glnV44A58), galK2, l, rac, hisC40(4), rfdB1, mgl-51, DrecA-srl306, srl301;Tu10, rpsL31(trrR), hkgK51, yxiA5, mit-1, argE3(Oc), thi-1; and JC5519 (thr-1, ara-14, leuB6, Dggt-proA62, lacY1, rss-33, glnV44A58), galK2, l, rac, hisC40(4), rfdB1, recC22, recB21, rpsL31(trrR), hkgK51, yxiA5, mit-1, argE3(Oc), thi-1. All strains were obtained from the E. coli Genetic Stock Center, Yale University, New Haven, CT.

Standard Genetic Techniques—Plasmid preparation, agarose gel, and polyacrylamide gel electrophoreses were carried out according to standard laboratory protocols (20). Transformations were performed by electroporation (48, 49). For the cotransformation experiments, each strain was cotransformed with a mixture of the two appropriate plasmids; the plasmids were originally grown in AB10289 which is the recA- strain. Forty microliters of washed cells of each strain (5 × 107 cells/ml) was initially transformed with 1 μl of the supercoiled DNA (0.5 μg/ml) listed in Fig. 1. For the experiments involving cotransformation, conjugation plasmids were prepared for electroporation, and the transformation mixture contained various combinations of the pUC19 and pACYC184 derivatives. The supercoiled plasmid volume was equally divided between the two test plasmids. A voltage of 2000 was delivered for 4.1 to 5.8 ms. The cuvette size was 0.2 mm.

Cotransformants were selected on LB agar plates containing ampicillin and tetracycline (75 μg/ml) and tetracycline (12 μg/ml) and grown for 4–16 h at 37 °C. Individual colonies were selected for culture and were grown to mid-logarithmic phase (0.3–0.9, 4–16 h) at 37 °C in LB media containing ampicillin (75 μg/ml) and tetracycline (12 μg/ml) under aerobic conditions.

Plasmid purification and gel electrophoresis and analysis were conducted as described (23–26, 28–30). The plasmid products obtained from recA- strains were quantitated by staining the agarose gels with ethidium bromide and photographed. The amount of DNA in the RB region (defined in Fig. 2) versus the total DNA in the gel was calculated and used as background and subtracted from the analyses involving recA+ strains. The plasmid inserts and flanking sequences were characterized by dideoxy sequencing on both strands with Sequenase (version 2.0). The pACYC184 primers, purchased from Genosys Inc., were the following: primer 4244 (ACGGTCTTCTTTAAAAG-GCCG) which 3′-terminates at map position 95; primer 4245 (CGTCATGCTAGCAAGACAGGAG) which 3′-terminates at map position 522. The PACT-2 primer was purchased from CLONTECH Laboratories, Inc., and was primer GALA AD (ACCACTACAATGGATG) which 3′-terminates at map position 5155. Restriction mapping reactions and ligase reactions were conducted as described (21).

Genetic Analyses—The frequency of the survival of amp-tet colonie due to the presence of the TRS was calculated by counting the number of colonies on LB agar plates containing ampicillin and tetracycline using standard microbiological techniques. pUC19 derivatives and pACYC184 derivatives were transformed into E. coli strains and grown at 37 °C. The number of viable cells was determined by growth in the presence of streptomycin (20 μg/ml). The number of mutant cells was determined by growth in the presence of ampicillin and tetracycline. E. coli AB1157 and JC10289 were washed and diluted to a concentration of 5 × 107 cells/ml and cotransformed with combinations of plasmids as indicated: for example, pUC19 + pRW3239; pRW3080 + pACYC184; or pRW3080 + pRW3239.

RESULTS

Interplasmid Recombination—To evaluate the potential role of homologous recombination in the expansion of TRS, a two-plasmid system was established in E. coli. (The term recombination is used in a general sense and includes gene conversion, unequal crossing-over, and sister chromatid exchange.) For our study, one family of plasmids (Fig. 1) was a derivative of pUC19 (a diminutive form of pBR322) that contains the unidirectional ColE1 origin of replication and harbors the ampicillin resistance gene (20). The other family of plasmids was derived from pACYC184 (21) that harbors the tetracycline resistance gene. A computer search revealed that little or no sequence identity exists between these two vectors; only single copies of identical tracts that were 48, 30, 23, 20, 16, and 12 bp in length were found along with several copies of 9–5 bp segments. Thus, the nonidentical nature of these two vectors enabled our focus on the potential effects of cloned tracts of different TRS on recombination. Prior investigations (22) revealed the stable cotransformation of derivatives of these two plasmids.

To analyze the recombination behavior of these plasmids (Fig. 1), experiments were conducted in three E. coli strains that are isogenic but differ in their recombination capacity. To test the recombination capacity of CTG-CAG repeat sequences,
compared with controls, each strain (AB1157, JC10289, and JC5519) was transformed with the parental control vectors pUC19 and pACYC184 that lack TRS. As a second control, each strain was cotransformed with the vector pACYC184 along with various pUC19 derivatives that contained different lengths and types of TRS. Third, the vector pUC19 was cotransformed along with pRW3239 (which contains (CTG\(^z\)CAG)\(_{175}\)). Finally, the three strains were cotransformed with various pUC19 derivatives that contained different lengths and types of TRS along with pRW3239. The only segments of the non-homologous plasmids that had identical sequences were the repeating tracts of (CTG\(^z\)CAG), (GTC\(^z\)GAC), (GTG\(^z\)CAC), (GAA\(^z\)TTC), (CGG\(^z\)CCG), and a fragment of phage \(\lambda\) DNA.

A strength of this experimental approach is the capacity to select for cells harboring recombinant DNA products that contain both drug resistances and to assay rigorously these products for expansions by restriction mapping and/or DNA sequencing since they contain unique recognition sites. Since the two vectors have different replication origins and copy numbers, exhaustive control studies, described above, were performed followed by analyses of either the supercoiled or linearized DNA which provided confidence of the products formed by recombination. For the 11 control experiments, described above, the CTG-CAG tracts were unchanged in length and/or deletions occurred, as expected (28–32). Thus, the differences in the origins of replication and the copy numbers per se did not

| TRS    | Plasmid designation | n  | TRS    | Plasmid designation | n  |
|--------|---------------------|----|--------|---------------------|----|
| None   | pUC19               | 0  | None   | pACYC184           | 0  |
| CTG-CAG| pRW2163             | 13 | CTG-CAG| pRW3239           | 175|
| CTG-CAG| pRW2180             | 30 | CTG-CAG| pRW3036           | 36 |
| CTG-CAG| pRW3036             | 36 | CTG-CAG| pRW3047           | 47 |
| CTG-CAG| pRW3080             | 80 | CTG-CAG| pRW3216           | 100|
| GTC-GAC| pRW3415             | 34 | GTC-GAC| pRW4110           | 54 |
| GTC-GAC| pRW3462             | 47 | CGG-CGG| pRW3041           | 81 |
| GTC-GAC| pRW3463             | 54 | CGG-CGG| pRW3683           | 49 |
| CGG-CGG| pRW3017             | 17 | CGG-CGG| pRW3867           | 60 |
| CGG-CGG| pRW3683             | 49 | CGG-CGG| pRW3311           | 81 |
| CGG-CGG| pRW3848             | 140| GTG-CAC| pRW4100           | 140|
| GTG-CAC| pRW3482             | 150| GAA-TTC| pRW4115           | 150|
| GAA-TTC| pRW3822             | 150| λ phage fragment| pRW4106          | —  |
| 564 bp| λ phage fragment     | —  | 564 bp| λ phage fragment   | —  |

**Fig. 1. Plasmids used for transformation of *E. coli*.** Plasmids in the left column are derivatives of pUC19 and in the right column are derivatives of pACYC184. The types of sequences and lengths of the TRS are listed under column n. For example, pRW3036 is a pUC19 derivative that contains a pure insert of (CTG\(^z\)CAG)\(_{36}\). All CTG-CAG inserts cloned into pUC19 are completely homogeneous (i.e. no interruptions). All other plasmids have TRS that contain interruptions as described previously (23–26, 28–30), except for pRW3017 which has a pure stretch of (CGG\(^z\)CCG)\(_n\). The longest stretch of uninterrupted repeating sequences is listed under column n. All TRS are in orientation I (23–26, 28–30). All plasmids have SacI (S) and EcoRI (E) sites common, but the pUC19 derivatives also contain a PstI(P) site and a NotI site on each side of the polylinker, whereas all pACYC184 plasmids contain a BstI(\(S^*\)) site. The ampicillin resistance gene (Amp) is designated as a checkered box, and the filled box on pACYC184 designates the tetracycline resistance gene (Tet). The origins of replication that are unique for the two different plasmids are shown by vertically and diagonally striped boxes. The shaded box is the polylinker region. pUC19 is a high copy number plasmid (20) with \(\sim 500\) copies per cell, whereas pACYC184 has a copy number of \(\sim 10\) per cell. The plasmids are not drawn to scale.
affect the TRS stability.

**Effects of TRS Sequences on Recombination as Analyzed with Superoiled DNAs**—Fig. 2 shows a typical gel electrophoragram of supercoiled plasmids as isolated from *E. coli* AB1157, which is recA⁺. The plasmids used for transformation and cotransformations are listed above each lane and are depicted on the right. The 1st 5 lanes are controls, and the last lane contains a new product and complete loss of the original transforming DNAs.

Considering the singly transformed cells, lane *a* shows the 1.8-kbp band (monomer supercoiled DNA), whereas the other bands are dimers, trimers, and other multimeric forms of pRW3080. The formation of multimers of plasmids in recA cells was reported previously (47). Since each plasmid is entirely homologous with itself, intraplasmid recombination in a recA⁺ background is expected. Lane *b* shows the monomer supercoiled form of pRW3239. Its lower copy number precluded the observation of dimers and higher multimers. Note that the size relationship between these control plasmids is as expected.

Considering the cotransformed cells, lane *c* shows the DNA isolated from *E. coli* AB1157 cotransformed with the vectors pUC19 and pACYC184. This lane shows the supercoiled starting materials, as expected; monomer supercoiled pUC19, at 1.7 kbp, migrated slightly faster than the larger pUC19 derivative, pRW3080 (which contains (CTG-CAG)₈₀ shown in lane *a*). Likewise, the dimer of pUC19, the band at 3.9 kbp, migrated slightly faster than the dimer of pRW3080 (at 4.1 kbp). Since the supercoiled vectors were recovered unchanged in size, no recombination occurred between these two DNAs. pACYC184, although not visible on this gel due to its lower copy number, is present since our media contained tetracycline (as well as ampicillin). Lane *d* shows the results of the cotransformation of pUC19 and pRW3239; the supercoiled monomer of pUC19 is the band at 1.7 kbp (identical to lane *c*). The next largest band in lane *d* is at 3.6 kbp which is the monomer of pRW3239 and comigrates with the molecules found in lane *b*, as expected. Thus, when one plasmid harbors a TRS, the interaction between the two plasmids is no different than when both vectors, which both lack TRS tracts, are cotransformed into recA⁺ cells. The identity of the other bands found in lane *d* is interpreted as for the data in lanes *a* and *c*. Lane *e* shows the results of cotransformation of pRW3080 and pACYC184. The 1.8-kbp band comigrates with the DNA in lane *a*, as expected. The DNAs found above 1.8 kbp comigrate with the DNAs found in lanes *a*–*d*, as described above. Hence, even though the recA system was intact, no new products were detected that could be ascribed to recombination. These results are as expected and are identical to the data obtained from the two-plasmid system when conducted in the 11 control studies with isogenic, but recombination-deficient, strains.

Lane *f* shows the extraordinary result of the formation of large recombinant bands from recA⁺ cells cotransformed with pRW3080 and pRW3239. A loss of the monomer forms of both plasmids and the formation of DNA (~55% of the total DNA) that migrates in the recombinant band (RB) region is revealed. Hence, comparison of the products in lanes *c*–*e* with the products found in lane *f* (the cotransformation with both plasmids containing the TRS) clearly demonstrates the loss of monomer forms of the starting plasmids and formation via recombination of new recombinant products, which contain both plasmids in various ratios. In addition, homodimers and other homomultimers are visible, as expected, and in some cases represent up to ~45% of the total DNA. The same experiment conducted in isogenic recombination-deficient strains failed to yield RB and the loss of the plasmid monomers (see “Requirement for Recombination Genes”). The general features revealed in Fig. 2 have been found in at least 100 other similar determinations.

Not only is the formation of the very long RB products (~18–40 kbp) extraordinary, but the complete loss of the pRW3080 (and probably also the not visible pRW3239) starting material was dramatic. This result was found even for very short culture times (less than 4 h). Hence, the reaction in *vivo* to form the RB from the appropriate (CTG-CAG) containing plasmids was very powerful. Furthermore, lane *f* does not contain a smear of DNA products suggesting that the formation of a discrete series of plasmids may have been triggered by secondary structures within the triplet repeat sequences themselves, as observed for other instabilities (23, 30, 50).

Whereas it is not possible to propagate the pUC19 derivative pRW3080 alone, nor the pACYC184 derivative pRW3239 alone, in the presence of both ampicillin and tetracycline (since they do not possess both drug resistances), we found that isolation of the recombinant product (Fig. 2, lane *f*) and subsequent retransformation into recA⁺ cells (JC10289) was effective, and the RB was stable. This result was expected since the RB (18–40 kbp) contains both drug resistances. Since the same number of colonies (~10%) was found when RB, retransformed into JC10289, was grown on only amp, or only tet, or both amp and tet, we conclude that the transformants are not due to a
FIG. 3. Analysis of transformation and cotransformation products in the linear form. A. agarose gel electrophoretic analysis of the linearized DNA products. The DNAs described in Fig. 2 were linearized as follows and analyzed on 0.8% agarose gels. Lane a shows pUC19
minority form. Parenthetically, whereas JC10289 contains Tn10:rfl, and hence expresses a low level of tetracycline resistance, our studies were performed in the presence of 12 μg/ml tetracycline; JC10289 will not grow at tetracycline concentration >2 μg/ml.

**Effects of TRS Sequences on Recombination as Analyzed with Linearized DNAs—** Since analyses of the supercoiled DNA products revealed the apparent recombination between plasmids containing CTG-CAG tracts, the material in the RB region was characterized further. Fig. 3A shows a typical gel electropherogram of linearized plasmid products as isolated from E. coli AB1157 transformed or cotransformed with the designated plasmids. Lanes a and b contain only pUC19 derivatives; lanes c and d contain only pACYC184 derivatives, and lanes e and f contain the mixture of two plasmids. Lane a shows the result of cells transformed with pUC19 and the product digested with *Pst*I. The appearance of a single band at ~2700 bp is as expected. When pRW3080 (which contains (CTG-CAG)_{175}) was grown in AB1157 and digested with *Pst*I, the products migrated at approximately 2950 bp (Fig. 3A, lane b); the products migrated slightly slower than pUC19 (lane a) due to the presence of the TRS insert. Interestingly, some small deletions were observed in the (CTG-CAG)_{175} tract and appear as a smear below 2950 bps, as expected (23).

Unlike the results found in *recA*− strains that yielded monomer forms of the plasmid (23–25, 30), the use of *recA*+ strains yielded monomer and multimer forms of plasmids. When multimers of a single type of plasmid were linearized, the result was the presence of only one band, which migrates to its known plasmid length, as expected. Hence, changes in the plasmid sequence would be distinguishable from the original plasmid. However, no such changes were detected when using AB1157 cells. Obviously, both plasmids (pUC19 and pRW3080) were cultured in the presence of ampicillin since the absence of the gene encoding tetracyclin resistance would preclude the propagation in the presence of tetracycline.

Considering the pACYC184 derivatives, lanes c and d show the *Sac*II-linearized products of pACYC184 and of pRW3239; distinct bands were observed at 4244 and 4450 bp, respectively. The sharpness of these two bands is due to the fact that larger plasmids stabilize TRS (1). Also, the faintness of the bands reflects the lower copy number of pACYC184. Obviously, these plasmids were grown only in the presence of tetracycline since the absence of the gene encoding ampicillin resistance would preclude the use of both antibiotics.

For the cotransformation experiments, pRW3080 and pRW3239 (both contain (CTG-CAG) sequences) were cultured in recombination-proficient cells in the presence of both ampicillin and tetracycline. After linearization of the recombinant products with *Pst*I (Fig. 3B) (which only cleaves the pUC19 derivative), the products are the 3.1-kbp unit length for pRW3080 cleaved out of the recombinant band (Fig. 3A, lane e). Due to recombination initiated in the CTG-CAG tracts between the two otherwise non-homologous plasmids, expansion products are detected since this 3.1-kbp DNA is slightly larger than pRW3080 (lane b). pRW3080 cleaved out of RB by *Pst*I is the major product in Fig. 3A, lane e, due to its high copy number, even in RB. In addition, two faint products are observed at 11 kbp, which represents 1 unit of pRW3080 linked to 2 units of pRW3239, and at 19–40 kbp which represents 1 unit of pRW3080 linked with multiple units of pRW3239 as diagrammed in Fig. 3B. The 7-kbp band representing 1 unit of pRW3080 and 1 unit of pRW3239 is not visible on this gel but was repeatedly observed on other analyses. Interestingly, the supercoiled monomer form of pRW3239 starting material which cannot be linearized by *Pst*I became fully integrated into the recombinant bands.

When the cotransformation of AB1157 with pRW3080 and pRW3239 was linearized with a restriction enzyme (*Sac*II) that solely cuts the pACYC184 derivative (Fig. 3B), several distinct products that differed in size from those products obtained by *Pst*I linearization were observed (Fig. 3A, lane f). This is consistent with the result of one type of plasmid recombined with one or more plasmids of the second type (Fig. 3B). The DNA at 4.4 kbp is linearized pRW3239 and the product at 10 kbp represents 1 pRW3239 unit linked to 2 pRW3080 units. In addition, broad bands in the range of 19–40 kbp were observed which represent recombinant bands consisting of 1 unit of pRW3239 with multiple units of pRW3080. Furthermore, an interesting facet of the analysis of lane f is the total absence of the high copy number supercoiled monomeric (or nicked/linear) pRW3080 starting material that cannot be linearized by *Sac*II (migrating at 1.8 kbp, Fig. 2, lane a); hence, this DNA became fully integrated into the recombinant bands. We presume that the same is true for pRW3239, but its low copy number precludes this observation. The 4.4-kbp band is assigned as linear pRW3239 because *Sac*II only linearizes pRW3239 and *Pst*I only linearizes the 3.1-kbp pRW3080. *Sac*II will not linearize pRW3080 so the supercoiled form will be present in the various products depicted in Fig. 3B. *Pst*I will not linearize pRW3239 so its supercoiled form will be present in the various products depicted in Fig. 3B.

Hence, this analysis also reveals the robust formation of RB from the appropriate (CTG-CAG) containing plasmids. The general features revealed in Fig. 3A were found in at least 100 other similar analyses. Whereas recombination may occur at any time during the growth of the cultures, “jackpots” have not been observed in our studies.

**RB Is Not a Catenate—** The possibility was considered that the recombinant band, which is presumably circular, in Fig. 2, lane f, might exist as cationed DNA (interlinked plasmids) rather than a recombinated product with the cationed DNA migrating slower than either of the substrate plasmid monomers. To address this question, studies were conducted on the linearized with *Pst*I; lane b shows the *Pst*I-linearized products of pRW3080 (containing (CTG-CAG)_{175}); lane c shows the *Sac*II-linearized product of pACYC184; lane d shows the *Sac*II-linearized product of pRW3239 (which contains (CTG-CAG)_{175}); lane e shows the *Pst*I-linearized products of AB1157 cotransformed with pRW3080 and pRW3239; and lane f shows the *Sac*II-linearized products of AB1157 cotransformed with pRW3080 and pRW3239. The DNA size markers were a 1-kbp ladder (Life Technologies, Inc.); the sizes of these bands are shown at the side of the figure. The figures on the right of the gel represent the linearized molecular structures; the top two figures represent linearized cointegrants containing e units of pUC19 (lane f) or m units of pACYC184 (lane e). The open boxes represent the CTG-CAG tracts. B, indication of interplasmid recombination events. The thick lines represent a pUC19 derivative, and the cross-hatched segments represent pRW3239. The boxes on each plasmid represent the homologous TRS. The unique *Pst*I and *Sac*II sites are shown; a single EcoRI site is present on both vectors. The arrows designate the locations of the primers used for sequencing. The figure is not drawn to scale. When *E. coli* AB1157 was cotransformed with the plasmids, the TRS align for strand invasion and exchange. A single crossover event is depicted (upper right). The lower left structures depict two possible recombinated plasmids, but a large number of multimers containing different numbers of the two plasmid units in different orientations can also exist. No restriction was placed on the amount of recombination or the extent of mutimer formation that can occur. If the recombinated plasmid population is linearized at a restriction site (i.e., *Pst*I) unique for the pUC19 derivatives, linear forms will be generated where the two ends of one pUC19 derivative flank one or more copies of pRW3239. Alternatively, if the RB is linearized with *Sac*II which is unique for pRW3239, one or more copies of the pUC19 derivative will be flanked by portions of a single copy of pRW3239.
DNA products using single site restriction enzymes. The products in Fig. 3A, lanes e and f, showed that the DNA did not exist as catenated products. Digestion with unique restriction enzymes resulted in the absence of circular monomer plasmids and in the appearance of 7 kbp and larger products that represented one digested plasmid connected to one or more undigested plasmids. If the DNA had been catenated, then use of the unique restriction enzymes (PstI and SacII as well as three other enzymes) (Fig. 3B) would have released the monomer supercoiled pRW3080 that would have migrated as a 1.8-kbp product such as that found in Fig. 2, lane a. Hence, our data show that the TRS are involved in genetic recombination events rather than in the formation of catenanes.

Survival of Amp-Tet’ Colonies—The frequency of occurrence of observing amp-tet’ colonies was determined genetically when E. coli AB1157 was cotransformed with plasmids that harbored various TRS as well as non-repetitive DNA. More than 41 prior studies revealed a direct correlation between the number of amp-tet-resistant colonies observed in a cotransformation experiment, the biochemical presence of RB in agarose gel electrophoretic determinations, and the 1–5-fold expansions of the CTG-CAG sequences. Cells that were cotransformed with pRW3080 (which contains (CTG-CAG)$_{175}$) and with pRW3239 (which contains (CTG-CAG)$_{175}$) had an ~90-fold enhanced survival compared with the average of the other pairs of plasmids. Alternatively, for the control studies, we found only background levels of RB, no expansion of the TRS, and low numbers of colonies. The types of sequences investigated were as follows: the DM (CTG-CAG) repeat; the sequence isomer (GTC-GAC) (24); a TRS that is not known to be associated with a disease (GTC-GAC) (24); the Friedreich’s ataxia sequence (GAA-TTC) (24, 51); the fragile X sequence (CGG-CCG) (25); a 564-bp HindIII bacteriophage A fragment; as well as two non-homologous sequence mixtures.

Interestingly, it may be noted that, as an infrequent event, we did observe the recombination of the fragile X sequences as measured by a loss of monomer plasmid and the formation of RB. However, the lengths of CGG-CCG tracts cloned into the plasmids were shorter than those of CTG-CAG, and this may contribute to the lower frequencies. Unfortunately, (CGG-CCG) tracts are extremely unstable in E. coli (25) rendering their study in our two-plasmid recombination system less clear to interpret than the results with the (CTG-CAG) sequences as well as the other six control sequences. Interestingly, prior replication-based studies (29) revealed the facile expansion of CTG-CAG compared with the nine other TRS. In summary, the robust recombination observed in this two-plasmid system is dependent on the presence of CTG-CAG in both DNAs.

Effect of TRS Sequence—Biochemical studies were also conducted to test the requirement for identical TRS sequences for homologous recombination events. Table I shows the results of replacing the (CTG-CAG)$_{10}$ tract in pRW3080 with various lengths of the sequence isomer (GTC-GAC) (24) or (CGG-CCG) (25, 26) in cotransformation experiments with pRW3239. The amount of DNA observed in the recombinant band region was at background level in all of these cases. Hence, the TRS tracts in the recombining plasmids must contain the same sequences. The combination of (CTG-CAG) with the sequence isomer (GTC-GAC) could, in principle, form parallel-stranded structures (27), but this is ineffective in this system. Also, in principle, GTC-GAC and CGG-CCG could form paired structures with 2/3 Watson-Crick pairs and 1/3 incorrect pairs; this is also ineffective in recombination. Thus, 33% of AC and GT oppositions create a non-recombinogenic pair of plasmids.

Furthermore, similar investigations were conducted with pUC19 and pACYC184 derivatives which both contain tracts of the fragile X sequence CGG-CCG. When pRW3311 (which contains (CGG-CCG)$_{175}$) was cotransformed with pRW3041 (a pACYC184 derivative that contains (CGG-CCG)$_{175}$), a less pronounced amount of RB was formed compared with CTG-CAG. Further experiments are in progress to determine if this is an effect of the length of the TRS, the lower capacity of CGG-CCG, compared with CTG-CAG, to form looped hairpin structures (25, 45) or other factors.

Effect of Length of TRS—The effect of TRS length on the formation of recombinant products was determined by maintaining (CTG-CAG)$_{175}$ in pRW3239 constant whereas the length of the same sequence in the pUC19 derivative was varied. Fig. 2, lane f, presents typical data that show that up to 63% of the DNA was found in the RB region of an agarose gel when pRW3080 and pRW3239 were cotransformed into recombiant-proficient cells. The data collected using other lengths of (CTG-CAG) in the pUC19 vector are summarized in Fig. 4. pRW3080 was replaced with pUC19 derivatives that contain different lengths of (CTG-CAG) and were cotransformed with pRW3239. The data revealed a sinusoidal curve with a threshold for the appearance of RB at 30 repeats. When the repeat lengths were 30 or shorter, monomer supercoiled DNA was recovered. However, when the pUC19 plasmids containing 36, 47, 80, or 100 repeats were cotransformed with pRW3239, no monomer supercoiled DNA was found, but rather the majority of the DNA was in the RB region (Fig. 2, lane f).

Hence, these results (Fig. 4) correlate strongly with the phenomena observed in myotonic dystrophy patients which demonstrate a repeat threshold where no symptoms are observed below the break point of 30 repeats, whereas progressively more severe symptoms occur above the threshold (1, 2). By analogy in E. coli, repeat lengths of 30 and less were stable and transmitted from one cell division to the next with high fidelity, whereas lengths above 36 repeats were unstable transmitted (expanded up to 5-fold their original lengths (see below)). This threshold is similar to the length observed in humans for the transmission of expanded alleles (11).

Requirement for Recombination Genes—All prior investigations from this laboratory (23–25, 28–30) on genetic instabilities of TRS were in recA’ E. coli, usually strain HB101. The
The percentage of DNA in the RB region as compared with the total amount of DNA on the agarose gel electrophoretic analyses was determined for different lengths of CTG-CAG inserts in pUC19 for the three strains of *E. coli*. The cotransformation experiments, as performed in the wild type strain (AB1157), were repeated in the recombinant-deficient strains JC10289 (recA<sup>+</sup>) and JC5519 (recB<sup>−</sup> C<sup>−</sup>). The *E. coli* strains were cotransformed with pRW3239 (containing (CTG-CAG)<sub>175</sub>) and with one of the following plasmids: pUC19 (containing (CTG-CAG)<sub>0</sub>); pRW2163 (containing (CTG-CAG)<sub>30</sub>); pRW2180 (containing (CTG-CAG)<sub>36</sub>); pRW3036 (containing (CTG-CAG)<sub>60</sub>); pRW3047 (containing (CTG-CAG)<sub>13</sub>); pRW3080 (containing (CTG-CAG)<sub>80</sub>); or pRW3216 (containing (CTG-CAG)<sub>100</sub>). The DNA was isolated and electrophoresed through 1.1% agarose gels in TAE buffer. The gels were photographed, and the negatives were scanned to quantitate the amounts of DNA. The standard deviation for *E. coli* rec<sup>A</sup> was ±17% and was ±2% for the other 14 experiments.

**Table II**

| Length of CTG-CAG repeats | DNA in RB region % |
|---------------------------|--------------------|
| *E. coli* rec<sup>A</sup> (AB1157) | 80 55 |
| *E. coli* rec<sup>A</sup> (JC10289) | 0 0 |
| 0 13 |
| 30 2 |
| 36 2 |
| 47 3 |
| 80 3 |
| 100 7 |
| *E. coli* rec<sup>B</sup> C<sup>−</sup> (JC5519) | 0 0 |
| 0 13 |
| 30 1 |
| 36 2 |
| 47 4 |
| 80 5 |
| 100 1 |

As a control, when rec<sup>A</sup> cells were cotransformed, no expansion was observed by sequence analysis. Therefore, the presence of RecA directly affects the genetic stability of (CTG-CAG). As a further control, if pRW3036 was propagated in AB1157, without pRW3239 and only in the presence of ampicillin (Table III), no expansion of the TRS was found by DNA sequence analysis, as expected. In this case, homodimers, homotrimers, etc., were formed (as for pUC19 and pRW3080 in Fig. 2) due to recombination of pRW3036 with itself, but the TRS was not expanded. Similar results (no formation of RB and no expansions of the TRS) were found by restriction mapping for pRW3080 alone and for pRW2180 by sequence analysis. It is well known (23–25, 28–30) that short repeats are transmitted beyond 80 due to the lack of markers in the uninterrupted repeat tracts.

As a control, when rec<sup>A</sup> cells were cotransformed, no expansion was observed by sequence analysis. Therefore, the presence of RecA directly affects the genetic stability of (CTG-CAG). As a further control, if pRW3036 was propagated in AB1157, without pRW3239 and only in the presence of ampicillin (Table III), no expansion of the TRS was found by DNA sequence analysis, as expected. In this case, homodimers, homotrimers, etc., were formed (as for pUC19 and pRW3080 in Fig. 2) due to recombination of pRW3036 with itself, but the TRS was not expanded. Similar results (no formation of RB and no expansions of the TRS) were found by restriction mapping for pRW3080 alone and for pRW2180 by sequence analysis. It is well known (23–25, 28–30) that short repeats are transmitted accurately and do not readily undergo alterations in repeat length. In all sequence determinations, the vector sequences (pUC19 and pACYC184) were unchanged.

In summary, numerous DNA sequence analyses and restriction mapping studies demonstrated the presence of expansions of 1–5-fold in length.

**Fate of Point Mutations after Recombination Events**—pRW3239 contains two point mutations that serve as useful molecular markers to follow the fate of individual TRS through the recombination events. If these interruptions in the TRS are directly involved in the exchange (or used as repair templates (Fig. 5)) of CTG-CAG sequences, they might move from one plasmid location to the other. pRW3080 (which contains an uninterrupted repeating sequence of (CTG-CAG)<sub>80</sub>) was cotransformed with pRW3239 in *E. coli* AB1157, and the recombinant product was linearized with *Pst*I and gel-purified to give the 7-kbp linearized band (not visible in Fig. 3A but repeatedly detected in other similar experiments), and the eluted band...
was ligated and retransformed into recA<sup>+</sup> cells (AB1157). Individual colonies were selected, and the recombinant DNA was purified and sequenced using a pUC19 primer (1211) (Fig. 5). Note that the CTG-CAG sequence in pRW3239 will not be analyzed when the pUC19 primers are used. DNA sequence analyses revealed the presence of one of the G-to-A interruptions (which occur at positions 28 and 69 of the (CTG-CAG)<sub>z</sub> tract in pUC19) in the TRS flanking the pUC19 derivative. Interestingly, the G-to-A interruption from the pACYC184-derived sequence was observed at approximately repeat number 127. Thus, the only way this result could have occurred was for recombination to take place between the CTG-CAG originally in pRW3239 with the pUC19-derived pRW3080 (data not shown).

A second case of transfer of a point mutation was found by sequencing the product from a cotransformation of pRW3216 (containing (CTG-CAG)<sub>100</sub>) and pRW3239 using the pACYC184 primer (4245). This primer hybridizes to the distal end from the original location of the two G-to-A interruptions. This product revealed a G-to-A interruption at position 26 at the distal end. Since sequence analyses of pRW3216 done prior to its use in the two-plasmid recombination system showed the absence of G-to-A mutations, this "new" interruption must be the result of two or more crossover events (gene conversion). Hence, this clone must be derived from a multimer of the recombinant structure (Fig. 3B, lower left) that contains several tandem copies of the pUC19 plasmids. These two cases of the exchange of point mutations are rare.

Interestingly, these results directly demonstrate the recombination-based expansion without the exchange of flanking sequences. Thus, the expansion and the presence of an interruption in the TRS flanking the pUC19 vector sequence is the result of gene conversion or crossing over (Fig. 5).

**Table III**

| Plasmids | Length of (CTG-CAG) tract in pUC19 derivative cotransformed with pRW3239 | Length of repeats in pUC19 derivative products | Increase in length of (CTG-CAG) |
|----------|-------------------------------------------------|---------------------------------|------------------|
| pRW3036  | 36                                              | 175                             | 139              |
| pRW3047  | 47                                              | 117                             | 70               |
| pRW3080  | 80                                              | 220                             | 140              |
| pRW2180  | 30                                              | 30                              | 0                |
| pRW3036  | 36                                              | 36                              | 0                |
| pRW3080  | 80                                              | 80                              | 0                |

Recombination and Genetic Instabilities

Inversions within CTG-CAG—In addition to the length polymorphisms and the exchange of point mutations, a plasmid (pRW4444) was isolated that had a switch in the type of repeating sequence in the two complementary strands. This unique deletion product of a (CTG-CAG)<sub>z</sub> sequence must have undergone a recombination event to yield [((CTG)<sub>67</sub>(CAG)<sub>13</sub>]<sup>z</sup> as the repeat sequence in the isolated product. SURE cells were transformed with pRW4404 (which contains (CTG-CAG)<sub>100</sub> cloned into the SmaI-EcoRI sites of pACT2). DNA sequence analyses of the isolated product plasmid (pRW4444) revealed [((CTG)<sub>13</sub>(CAG)<sub>67</sub>]<sup>z</sup> as the repeat sequence and restriction mapping confirmed the presence of a PstI site (CTG-CAG) at the center of the sequence inversion. Hence, this unusual result directly demonstrates the occurrence of an inversion event, the change of orientation of the sequence relative to outside markers, which is likely due to the formation of a slipped-strand structure (52) with staggered single-stranded loops which became rehybridized to form intrahelical pseudoknots, theta shape, figure eight, and bow-shaped structures (Fig. 6) (34). These unorthodox conformations may exist <i>in vivo</i> due to the facile slippage of the CTG and CAG complementary strands relative to each other. A type of "illegitimate" recombination event must have occurred across the four-stranded intersection of one of these rehybridized structures to generate the [((CTG)<sub>13</sub>(CAG)<sub>67</sub>]<sup>z</sup> product (Fig. 6, bottom).

A second example of this type of recombination event that elicited a switch in the repeating sequence of the two complementary strands was found in another colony isolated from the recultivation of pRW4404 in SURE cells (as described above). The DNA sequence of the isolated products (pRW4445) revealed [((CTG)<sub>13</sub>(CAG)<sub>67</sub>]<sup>z</sup> as the undeleted but rearranged TRS insert. We presume that this product was formed by the mechanisms described above (Fig. 6) for pRW4444.

In summary, these DNA sequence analyses provide direct biochemical evidence for recombination between the TRS, confirming the genetic observations.

**DISCUSSION**

This report describes the direct demonstration that recombination mediates expansion and contraction of CTG-CAG repeats. The expansion events are dependent on the presence of long CTG-CAG sequences in the two-plasmid recombination system and require recombination-proficient cells to give frequent, severalfold expansions. Recombination was proven genetically and biochemically by the following: (a) the presence of
Recombination and Genetic Instabilities

Fig. 5. A model for expansion of CTG-CAG repeats mediated by recombination. Each strand of the CTG-CAG repeat participating in recombination is shown (the TRS in pUC19 and pACYC are open boxes). The dots in the insert in pACYC184 represent the G-to-A mutations at positions 28 and 69. The heavy solid lines represent the pUC19 vector, and the thinner cross-hatched vector is pACYC184. Two possible mechanisms are proposed. Homologous recombination between the two TRS on the left side gives an exchange of the G-to-A interruption from one plasmid to the other sequence with exchange of flanking sequences. An alternative mechanism (right side) shows that a double-strand break occurs within the CTG-CAG tract (18, 19, 35, 54, 55) and exchange at the broken ends forms two Holiday-like junctions separated by the distance "k." DNA repair synthesis (dashed lines) restores the sequence with (CTG-CAG). Synthesis occurs on both strands resulting in the expansion of pRW3080. Depending on the extent of branch migration, different size expansions will be formed. Resolution of the junctions leaves the flanking sequences unaltered. The structures are processed as in unequal crossing-over between sister strands. Misalignments between the strands as well as other intermediates involving single-stranded loops that are displaced, melted, and/or slipped could lead to expansion and explain the formation of new products. Adapted from Ref. 42.

Both amp and tet resistances in the recombinant products; (b) the formation of long cointegrant DNAs; (c) the expansion of (CTG-CAG) tracts by DNA sequencing and by restriction mapping for the longest tracts; (d) the transfer of G-to-A polymorphisms from the TRS in the pACYC184 derivative to the TRS in the pUC19 derivatives; and (e) inserts with strand inversions (i.e. [(CTG)\textsubscript{i}](CAG)\textsubscript{m}][(CTG)\textsubscript{m}](CAG)\textsubscript{i}]. These results are in stark contrast to prior investigations in recombination-deficient E. coli and in yeast where expansions were substantially less frequent than deletions by a ratio of approximately 1:100 (23, 28–32). For recombination-proficient cells, the ratio of expansions to deletions is as high as 100:1. This conclusion is derived from more than 50 cotransformation experiments of pRW3239 with any one of four pUC derivatives where the lengths of the PstI-SacI fragments containing the TRS were analyzed for expansions and deletions by polyacrylamide gel electrophoresis. When a plasmid (pRW3036) containing (CTG-CAG)\textsubscript{36} was propagated alone in recA\textsuperscript{−} cells, multimers were observed as expected (47, 53), but the TRS length was unaltered and no RB was observed. However, when this plasmid was cotransformed with a pACYC184 derivative containing (CTG-CAG)\textsubscript{175}, RB was observed by gel electrophoresis and substantial expansion of the 36 repeat tract was found in the amp-tet\textsuperscript{−} colonies indicating the involvement of recombination in expansion. Similar results were found when pRW3080 or pRW3047 replaced pRW3036. The reason why no recombination-based expansions and no RB were observed when the pUC19 derivatives containing (CTG-CAG) sequences were propagated alone is uncertain but may be due to a some property of a second non-homologous plasmid for initiating recombination in the presence of both tet and amp. Neither vector contains a χ site.

Both amp and tet resistances in the recombinant products; (b) the formation of long cointegrant DNAs; (c) the expansion of (CTG-CAG) tracts by DNA sequencing and by restriction mapping for the longest tracts; (d) the transfer of G-to-A polymorphisms from the TRS in the pACYC184 derivative to the TRS in the pUC19 derivatives; and (e) inserts with strand inversions (i.e. [(CTG)\textsubscript{i}](CAG)\textsubscript{m}][(CTG)\textsubscript{m}](CAG)\textsubscript{i}]. These results are in stark contrast to prior investigations in recombination-deficient E. coli and in yeast where expansions were substantially less frequent than deletions by a ratio of approximately 1:100 (23, 28–32). For recombination-proficient cells, the ratio of expansions to deletions is as high as 100:1. This conclusion is derived from more than 50 cotransformation experiments of pRW3239 with any one of four pUC derivatives where the lengths of the PstI-SacI fragments containing the TRS were analyzed for expansions and deletions by polyacrylamide gel electrophoresis. When a plasmid (pRW3036) containing (CTG-CAG)\textsubscript{36} was propagated alone in recA\textsuperscript{−} cells, multimers were observed as expected (47, 53), but the TRS length was unaltered and no RB was observed. However, when this plasmid was cotransformed with a pACYC184 derivative containing (CTG-CAG)\textsubscript{175}, RB was observed by gel electrophoresis and substantial expansion of the 36 repeat tract was found in the amp-tet\textsuperscript{−} colonies indicating the involvement of recombination in expansion. Similar results were found when pRW3080 or pRW3047 replaced pRW3036. The reason why no recombination-based expansions and no RB were observed when the pUC19 derivatives containing (CTG-CAG) sequences were propagated alone is uncertain but may be due to a some property of a second non-homologous plasmid for initiating recombination in the presence of both tet and amp. Neither vector contains a χ site.

The plasmid copy number in the single and dual transformations was analyzed. The number of copies of the singly transformed DNAs was ~50 to 1 for pRW3080 and pRW3239, respectively. However, when the two DNAs were cotransformed and selected for both tet and amp resistance, RB was found as expected, and the ratio of the pRW3080 and 3239 in RB was ~20 to 1. These data are the average of five experiments on the EcoRI-linearized DNAs after agarose gel electrophoretic fractionation (both DNA contain a single EcoRI site). Hence, the replication of recombinant multimers in RB (Fig. 3B, lower left) may increase the copy number of the pRW3239 component due to the dominance of the ColE1 origin in pRW3080. It is noteworthy that cotransformation of pRW3080 (or pRW3036 or 3047) with pRW3239 into recombination-deficient cells (JC10289 or JC5519) results in no formation of RB and no expansion of the CTG-CAG tracts in the pUC derivatives as revealed by restriction mapping.

The extent of expansions observed in this system was severalfold the original lengths. The point of recombination initiation lay within the CTG-CAG repeat sequences since no recom-
combination occurred in other regions of the plasmids as observed by restriction mapping. DNA sequencing revealed that when changes in the size of the plasmids did occur, the changes were within the TRS, not in the vectors. Although alterations in repeat length might have occurred by replication-based events including repair, all such events were controlled because the strains were equally competent to carry out replication-based instability reactions (23, 28–30). Hence, this two-plasmid recombination system may be considered as a primitive model of two eucaryotic chromosomes that harbor various alleles of CTGzCAG.

The expansion of CTG-CAG by recombination without the exchange of flanking markers can be explained by the double-strand break repair model (Fig. 5). The simple repeating nature of the CTG-CAG sequence in the presence of RecA may cause the pairing of sites for the alignment of the double-strand break gap into a homologous template. The ends of the break may be displaced followed by strand invasion and subsequent DNA synthesis (separated by length k) to extend the chains followed by resolution to generate an expanded CTGzCAG. This model directly explains our results found for the recombinant product of the cotransformation of pRW3239 (containing two G-to-A interruptions) and pRW3080 (containing an uninterrupted tract of (CTGzCAG)80); the product had a G-to-A interruption transferred to the CTG-CAG tract in pRW3080. Thus, the template for DNA repair or a portion of the exchanged DNA tract contained the G-to-A interruption and the expansion occurred without the exchange of flanking sequences (gene conversion).

The DNA replication fork stalls when it encounters a CTG-CAG sequence (3, 4) which can result in double-stranded breaks (18, 19, 35, 54, 55). These gaps provide binding sites for RecA which processes the DNA ends via recombinational repair. Premature termination and replication fork collapse requires recombinational repair to continue. Thus, it is likely that a complex interrelationship exists between replication and recombination functions in vivo. Considering the data described herein, examples were observed of expansions both with and without exchange of flanking sequences. Hence, both simple homologous recombination and gene conversion (Fig. 5) were observed in E. coli as found in patients (see Introduction). CTG-CAG tracts longer than 30 units are effective sites for homologous recombination since the recovery of both drug resistance markers occurred at only ~1% of the frequency in the absence of the TRS. The length of the CTG-CAG tract of approximately 90 bp as a minimum for efficient recombination is reminiscent of the threshold observed for expansion from the normal to the premutation stage (1, 2) in DM. Also, this length is in good agreement with prior determinations on the required extent of sequence identity (50–75 bp) for homologous recombination in E. coli (36, 37), especially considering that substantially different types of sequences were studied in rather disparate systems. Longer CTG-CAG tracts that are flexible and writhed (26) have a propensity to slip (38, 52) which may initiate recombination. Prior work (39–43) revealed the recombinogenic properties of simple, direct repeat sequences.

These studies also revealed the requirement for a perfect homologous CTG-CAG sequence in an antiparallel arrangement on both vectors. Attempts to achieve recombination be-
between CTG-CAG and GTC-GAC failed. Hence, parallel DNA (27) is not a substrate for this recombination system. Likewise, correct Watson-Crick pairing to the extent of 66% is ineffective since recombination between CTG-CAG and CGG-CGG failed also. Furthermore, the expansions of the CTG-CAG repeats in the two-plasmid system were independent of the orientations of the TRS in the vectors. 2

TRS instability by gene conversion (unequal crossing-over) is a robust process, and thus, this mechanism along with DNA replication may contribute to the length polymorphisms observed in human diseases. CTG-CAG seems to have special properties for recombination. Whereas the reason for this behavior is uncertain, prior investigations of nucleosome assembly (44), expansion by replication (29), conformational flexibility and writhing (26, 38), capacity for adopting hairpin loops (44), and susceptibility for double-strand breaks in vivo (18) revealed its unorthodox character. A prior review has summarized the molecular similarities between studies in humans and E. coli related to hereditary neurological diseases (45). We have no evidence on the extent to which, if at all, recombination is responsible for TRS expansions in humans. However, E. coli has been a useful model to investigate the molecular processes responsible for other events related to these instabilities (45, 56).

Since CTG-CAG repeats are recombinogenic and since genetic recombination utilizes enzyme systems different from replication, new avenues for therapeutic intervention strategies in human hereditary neurological diseases may be developed using somatic cell gene therapy.

Acknowledgments—We thank Dr. Robert Gellibolian for the preparation and characterization of pRW3041, Dr. Adam Jaworski for cloning pRW4404 and -4444, and Dr. John Wilson for advice.

REFERENCES

1. Wells, R. D., and Warren, S. T. (eds) (1998) Genetic Instabilities and Hereditary Neurological Diseases, Academic Press, Inc., San Diego.
2. Paulson, H. L., and Fischbeck, K. H. (1996) Ann. Rev. Neurosci. 19, 79–107.
3. Samadashvily, G. M., Raca, G., and Mirkin, S. M. (1997) Nat. Genet. 17, 284–304.
4. Ohshima, K., and Wells, R. D. (1997) J. Biol. Chem. 272, 16798–16806.
5. Tsilfidis, C., Markenzie, A. E., Mettler, G., Barcelo, J., and Korneluk, R. G. (1992) Nat. Genet. 1, 192–195.
6. O’Hoy, K. L., Tsilfidis, C., Mahadevan, M. S., Neville, C. E., Barcelo, J., Hunter, A. G. W., and Korneluk, R. G. (1993) Science 259, 809–810.
7. Van den Ouweland, A. M. W., Deelen, W. H., Kunst, C. B., Uziel, M.-L. G., Nelsen, D. L., Warren, S. T., Oostra, B. A., and Halley, J. J. (1994) Hum. Mol. Genet. 3, 1823–1827.
8. Losekoot, M., Hoogendoorn, E., Olmer, R., Jansen, C. C. A. M., Oosterwijk, J. C., Van den Ouweland, A. M. W., Halley, D. J. J., Warren, S. T., Willemsen, R., Oostra, B. A., and Bakker, E. (1997) J. Med. Genet. 34, 924–926.
9. Brown, W. T., Houck, G. E., Jr., Ding, X., Zhong, N., Nolin, S., Glicksman, A., Dobkin, C., and Jenkins, E. C. (1996) Am. J. Med. Genet. 64, 287–292.
10. Zhong, N., Kajanoja, E., Smiths, B., Pietrofesa, J., Curley, D., Wang, D., Ju, W., Nolin, S., Dobkin, C., Rynnanen, M., and Brown, W. T. Am. J. Med. Genet. 64, 226–233.
11. Jansen, G., Willems, P., Coerwinkel, M., Nillesen, W., Sneets, H., Vits, L., Howeler, C., Brunner, H., and Wieringa, B. (1994) Am. Soc. Hum. Genet. 54, 575–585.
12. Warren, S. T. (1997) Science 275, 408–409.

2 J. P. Jakupcic and R. D. Wells, manuscript in preparation.