Long CTG•CAG Repeat Sequences Markedly Stimulate Intramolecular Recombination*

Marek Napierala, Pawel Parniewski¶, Anna Pluciennik‡, and Robert D. Wells§

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

¶Center for Microbiology and Virology
Polish Academy of Sciences
106 Lodowa Street
93-232 Lodz, Poland

‡Present Address: Department of Biochemistry
Duke University Medical Center
150 Nanaline H. Duke Bldg., Research Drive
Durham, NC 27710
Running title:

Frequency of Intramolecular Recombination Between Triplet Repeats
ABSTRACT

Previous studies have shown that homologous recombination is a powerful mechanism for generation of massive instabilities of the myotonic dystrophy CTG•CAG sequences. However, the frequency of recombination between the CTG•CAG tracts has not been studied. Herein, we performed a systematic study on the frequency of recombination between these sequences using a genetic assay based on an intramolecular plasmid system in E. coli. The rate of intramolecular recombination between long CTG•CAG tracts oriented as direct repeats was extraordinarily high; recombinants were found with a frequency exceeding 12%. Recombination occurred in both RecA+ and RecA- cells, but was ~2 - 11 times higher in the recombination proficient strain. Long CTG•CAG tracts recombined ~10 times more efficiently than non-repeating control sequences of similar length. The recombination frequency was 60 fold higher for a pair of (CTG•CAG)_{165} tracts compared with a pair of (CTG•CAG)_{17} sequences. The CTG•CAG sequences in orientation II (CTG repeats present on a lagging strand template) recombine ~2 – 4 times more efficiently than tracts of identical length in the opposite orientation relative to the origin of replication. This orientation effect implies the involvement of DNA replication in the intramolecular recombination between CTG•CAG sequences. Thus, long CTG•CAG tracts are hotspots for genetic recombination.
INTRODUCTION

Genetic instabilities (expansions and deletions) of simple repeating sequences are important in the life cycles of both prokaryotic (1) and eukaryotic (2) cells. This fundamental mechanism of mutagenesis has been found in mycoplasma, bacteria, yeast, mammalian cell cultures, and in humans. In mycoplasma and bacteria, these genetic polymorphisms are the basis for phase variations, which control the expression of genes (3-7). In humans, the expansions and deletions of simple repeating sequences are closely tied to the etiologies of cancers (8-12) as well as hereditary neurological diseases (reviewed in 2).

The general mechanism accepted for all of these instabilities is slipped-strand mispairing which allows mismatching of neighboring repeats and, depending on the strand orientation, enables the insertion or deletion of repeats during DNA polymerase-mediated duplication (reviewed in 2,13). The enzymatic machineries involved include DNA replication and repair (nucleotide excision repair, methyl-directed mismatch repair, DNA polymerase III proofreading) (reviewed in 2,13). Biochemical and genetic studies showed also that expansions and deletions of the TRS sequences occur in vivo by homologous recombination (14,15). These investigations, carried out in a two-plasmid system, demonstrated that the expansion mechanism is principally gene conversion rather than unequal crossing-over (15).
In the present work, we cloned two triplet repeat tracts in the same plasmid and used an intramolecular assay to study the recombinational properties of the CTG•CAG sequences (Fig. 1). Intramolecular recombination systems have been widely used to investigate the mechanism of the recombination processes (16-24) as well as to establish the recombinational properties of different DNA sequences, including microsatellites (25).

It has been shown that recombination between TRS tracts can lead to repeat expansion (14,15,26). Evaluation of the frequency of recombination between CTG•CAG tracts provides important information about the cellular mechanisms of instability, relative to replication and repair.

Herein, we have developed the first genetic assay for monitoring the frequencies of intramolecular recombination between CTG•CAG tracts in *E. coli*. Interestingly, long CTG•CAG repeat sequences from myotonic dystrophy are preferred sites for intramolecular recombination. In a companion paper (27), we have established a genetic assay for monitoring the recombination frequency of the CTG•CAG repeat tracts in an intermolecular system.
EXPERIMENTAL PROCEDURES

**Parent Plasmids**— pRW3244, pRW4026, pRW3246 and pRW3248 were the parent plasmids containing (CTG•CAG)$_n$ tracts used for these experiments; these pUC19 NotI derivatives contain the (CTG•CAG)$_n$ tracts cloned into the *Hinc*II site of the polylinker (28-30). For nomenclature of the TRS, CTG•CAG designates a duplex sequence of repeating CTG, which may also be written TGC or GCT; CAG, the complementary strand, may also be written as AGC or GCA. The orientation is 5’ to 3’ for both designations of the antiparallel strands.

pRW3244 contains (CTG•CAG)$_{17}$, pRW4026 contains the (CTG•CAG)$_{67}$, pRW3246 contains (CTG•CAG)$_{98}$ and pRW3248 contains (CTG•CAG)$_{175}$ sequence. The (CTG•CAG)$_{175}$ sequence is not a pure CTG•CAG tract but contains two G to A interruptions at repeats 28 and 69 (30); all other TRS are pure (not interrupted). All of these sequences have non-repeating human flanking sequences (19 and 41 bp) outside the repeated tract. pRW3815 (Ohshima and Wells, unpublished) is a pUC18 NotI derivative and contains the (GTC•GAC)$_{79}$ tract. These plasmids were maintained in *E. coli* HB101 (Life Technologies, Inc.) ($mrB$, $mmr$, $hsdS20$, ($r^c$, $m^z$), $recA1$, $supE44$, $ara14$, $galK2$, $lacY1$, $proA2$, $rplS20$, ($Sm^R$), $xyl5$, $λ^-$, $leuB6$, $mtl^{-1}$). The (CTG•CAG)$_n$ and (GTC•GAC)$_n$ sequences were subcloned into pBR322.

**Cloning of (CTG•CAG)$_n$ and (GTC•GAC)$_n$ Sequences into pBR322**— The general strategy of this investigation involved recloning of the (CTG•CAG)$_n$ and
(GTC•GAC)_n sequences from pUC19_NotI and pUC18_NotI derivatives (28), respectively, into pBR322. Fragments containing the CTG•CAG and the GTC•GAC TRS were prepared from these plasmids by digesting the pUC19_NotI or pUC18_NotI derivatives with EcoRI and HindIII (New England Biolabs, Inc.) followed by filling-in the recessed 3’ termini with 0.1 U of the Klenow fragment of E. coli DNA polymerase I (U. S. Biochemical Corp.) and the four dNTPs (0.1 mM each). In the case of pRW4806 (a pUC19_NotI derivative harboring a tract of 165 uninterrupted CTG•CAG repeats), the insert was prepared by Alul digestion. The blunt-ended DNA fragments were used for cloning to obtain plasmids containing the TRS tracts in both orientations relative to the unidirectional ColE1 origin of replication. The digested DNA was electrophoresed in a 7% polyacrylamide gel, stained with ethidium bromide and the bands containing the triplet repeat fragment were excised. The DNA was eluted from the excised bands, purified by phenol-chloroform extraction and precipitated with ethanol (31). The vector was prepared by digesting pBR322 with EcoRI and HindIII followed by filling-in the recessed 3’ termini as described earlier. The vector and the insert were mixed at a molar ratio of approximately 1:10 and ligated for 14 h at 16°C by the addition of 20 U of T4 DNA ligase (United States Biochemical Corp.). The ligation mixture was ethanol precipitated and transformed into E. coli HB101 by electroporation (at voltage 2.5 kV, cuvette size 0.2 mm) and plated on LB agar plates containing 100 µg/ml ampicillin. Plasmid DNA was isolated from individual transformants by the Wizard Plus Miniprep DNA Purification System (Promega). Clones containing the CTG•CAG repeats in orientations I and II (defined in 28,30) were obtained.
and characterized by restriction mapping. The inserts cloned into the 
EcoRI/HindIII site of pBR322 are referred to as “X inserts” (Fig. 2). The pBR322 
derivatives containing a single CTG•CAG sequence (the “X insert”) were 
subsequently used to clone the second TRS tract (CTG•CAG or GTC•GAC) into 
the PvuII site at position 2064 of the pBR322 backbone (“Y insert”, Fig. 2). The 
same experimental approach was used to clone the second TRS insert, except 
that after ligation the reaction mixture was subjected to PvuII digestion to 
eliminate plasmids lacking the insert. This strategy enabled the construction of a 
family of plasmids harboring two homologous TRS tracts oriented as direct 
repeats or inverted repeats as well as plasmids containing non-homologous 
repeats (Fig. 2).

All plasmids were characterized by restriction mapping (to determine the 
orientation and length of the cloned TRS) and dideoxy sequencing of both 
strands with Thermo Sequenase Radiolabeled Terminator Cycle Sequencing Kit 
(United States Biochemical Corp.). The sequencing reactions were carried out 
according to the manufacturer’s recommendations using the following pBR322 
specific primers: pBR322EcoRI, GTATCACGAGGCCCT which 3’-terminates at 
the pBR322 map position 4347 (New England Biolabs, Inc.); pBRHR, 
GCGTTAGCAATTTAACTGTGAT which 3’-terminates at the pBR322 map 
position 49 (Genosys Inc.); pBRPF, GCTTCACGACCACGCTGAT which 3’-
terminates at the pBR322 map position 2052 (Genosys Inc.); pBRPR, 
GTCAGAGGTTTACCCGTCAT which 3’-terminates at the pBR322 map position 
2087 (Genosys Inc.). The products of the sequencing reactions were analyzed
on 6% Long Ranger gels (FMC BioProducts) containing 7.5 M urea in the glycerol tolerant gel buffer (United States Biochemical Corp.). The gels were dried and exposed to x-ray film.

Cloning of Non-repeating DNA Sequences into pBR322—Two different non-repeating sequences were used as controls in this study: the 564 bp fragment of \( \lambda \) phage DNA (\( \text{HindIII} \) fragment from nucleotide position 36895 to 37459) and the 354 bp fragment of the human DMPK gene (part of the exon 7 and intron 7) (32-35). pRW4804 and pRW4805 were constructed by digestion of \( \lambda \) phage DNA with \( \text{HindIII} \) and cloning one of the released restriction fragments (564 bp) into the \( \text{HindIII} \) and \( \text{PvuII} \) sites of pBR322. Thus, the two plasmids, pRW4804 and pRW4805 harbor direct and inverted repeats, respectively (Fig. 2).

The exon7/intron7 fragment of the human DMPK gene used for construction of pRW4871 and pRW4873 was obtained by PCR amplification of the sequence from the human genomic DNA. The PCR reaction was carried out in a volume of 20 \( \mu \)l containing 50 ng of genomic DNA, 1.5 mM MgCl\(_2\), 50 mM KCl, 10mM Tris/HCl (pH 8.3), 200 \( \mu \)M of each dNTP and 0.2 U of \text{PfuTurbo} DNA polymerase (Stratagene). The PCR primers DM7F, GGCTCGAGACTTCATTCAGC and DM7R, TAGATGGGCACAGAGCAGGT were used at the concentration 1 \( \mu \)M. Amplification on a PCR System 9700 (Applied Biosystems) involved 35 cycles: 20s/95\(^{\circ}\)C, 20s/58\(^{\circ}\)C and 40s/72\(^{\circ}\)C. Polyacrylamide gel electrophoresis purified PCR product was phosphorylated using 2mM ATP and 5U of T4 polynucleotide kinase (New England Biolabs, Inc.) and cloned into the \( \text{HindIII} \) and \( \text{PvuII} \) sites of
pBR322. pRW4871 as well as pRW4873 contain homologous sequences oriented as direct repeats; however, the orientations of the pairs of inserts are opposite in these two plasmids.

Cloning of the Green Fluorescence Protein Gene (GFP) into pBR322

Derivatives—pBR322 and pBR322 derivatives containing direct, inverted, non-homologous repeats and non-repeating DNA sequences were digested with EcoRV and Eagl (positions 185 and 939 on the pBR322 map, respectively) to remove the 754 bp DNA fragment of the vector backbone. The digested plasmids were purified by 5% acrylamide gel electrophoresis as described earlier and ligated to the GFPuv gene (36). The GFPuv gene was obtained by digestion of the pGFPuv (Clontech Laboratories, Inc.) with PvuII and Eagl (positions 56 and 1078 on the pGFPuv map, respectively). After ligation and transformation into E. coli HB101, transformants were screened using a long-wave length UV lamp. The cells carrying plasmids with the GFPuv gene emitted a strong green fluorescence.

The GFP cassette from pGFPuv contains the GFPuv variant of the green fluorescent protein gene inserted in frame with the lacZ initiation codon from pUC19 so that a β-galactosidase-GFPuv fusion protein is expressed from the lac promoter in E. coli.

Conditions of Bacterial Growth for Recombination Studies—For determinations of recombination properties, plasmids containing TRS tracts were
electrophoresed in 1% agarose gels and bands corresponding to the supercoiled plasmids were excised from the gels, transferred into dialysis tubes and electroeluted (31). To avoid DNA damage, plasmid purifications were performed without ethidium bromide staining and UV irradiation of DNA. In all experiments, only gel-purified, supercoiled plasmid DNA was used for transformation of the appropriate E. coli strains. To assure the identical conditions for experiments with all plasmids studied, a large batch of the competent cells was prepared for each set of experiments and the transformations were always done in parallel. The transformants were cultured, harvested and analyzed under the same conditions. The following E. coli strains were used: AB1157 (37) as a parent of the recombination deficient strain JC10289 (thr-1, ara-14, leuB6, Δ(gpt-proA)62, lacY1, tsx-33, glnV44(AS), galK2, λ, rac-, hisG4(Oc), rfbD1, mgl-51, Δ(recA-srl)306, srlR301::Tn10, rpsL31(strR), kgdK51, xylA5, mtl-1, argE3(Oc), thi-1). Strains were obtained from the E. coli Genetic Stock Center, Yale University, New Haven, CT. In the population experiments, the transformation mixture was inoculated into 10-ml LB tubes containing 100µg/ml ampicillin at a cell density of 10^2 cells/ml. The cultures were grown at 37°C with shaking at 250 rpm. At late log phase (A_{600} ~1.0 units), the cells were harvested, and the plasmid DNA was isolated as described above and analyzed by restriction digestion.

To determine the frequency of recombination, plasmids harboring the GFPuv gene were transformed into the appropriate E. coli strain, plated onto LB plates containing 100 µg/ml ampicillin and incubated for 16 hr at 37°C. The frequency of recombination was measured as the ratio of the number of white colonies to the
The total number of viable cells. The white as well as a representative number of fluorescent colonies were inoculated into 10 ml of LB medium (containing ampicillin at 100 µg/ml). After overnight growth, the plasmids were isolated and subjected to the restriction and DNA sequencing analyses. The statistical analyses were performed using SigmaStat v. 2.03.

This genetic assay enabled the detection and quantitation of the recombination events which occurred directly after transformation of the parental plasmids into the host cells. In order to detect those recombination events that took place at a later stage of colony formation, the recombination product would have to outgrow the parental plasmid molecules (that are present in a large excess at the moment of the recombination event). Consequently, the recombinant plasmid should have a tremendous replication advantage over the parental plasmids. However, this can be easily ruled out by the results of copy number analyses (see Results).

In addition, the white and the fluorescent colonies are stable. Randomly selected fluorescent colonies (350 total) were inoculated into one bulk culture, mixed and then plated on plates containing ampicillin. After overnight growth, no white colonies were observed among ~2x10^5 colonies screened. The same experiment was repeated for the white colonies which revealed no fluorescent colony formation in ~10^5 white colonies analyzed. These results indicate that the “color of the colony” (i.e. recombination status of the plasmid) is established at the earliest stage of the colony formation and masking or overgrowing of the cells.
to alter the apparent color (e.g. fluorescent cells by the white ones or vice versa) is highly unlikely.

_Determination of E. coli Growth Rates and Plasmid Copy Numbers_—To ensure that results of the population experiments are not biased by the growth advantage of cells containing recombination products over the cells containing parental plasmids, the doubling time of _E. coli_ cells harboring either recombination substrates or the recombination products with CTG•CAG tracts of different lengths and orientations was established. The determination of the doubling time and plasmid copy numbers as well as the recombination studies were carried out under identical conditions of bacterial growth (10-ml LB tubes containing 100µg/ml ampicillin, 37°C with shaking at 250 rpm.). In each case, approximately $10^2-10^3$ cells/ml were used to start the cultures. Aliquots of 10 µl were withdrawn at every 30 – 60 min for approximately 8 h, diluted in LB and subsequently plated on agar plates without ampicillin. The growth curves were prepared using SigmaPlot 2000 v. 6.10 and the doubling time was calculated as described previously (38).

To exclude the possibility of the replicative advantage of recombination products over the parental plasmids, the copy numbers of these plasmids were determined as described earlier (39,40); the size of the _E. coli_ genome of 4,639 Kbp (41) was used for these calculations. The quantitative analyses of plasmid and genomic DNAs separated by agarose gels were performed using FluorChem v. 3.04 (Alpha Innotech Corp.).
Agarose and Polyacrylamide Gel Analyses of Recombination Products—In order to analyze the products of intramolecular recombination between repeating sequences, the isolated DNAs were linearized with AflIII and labeled by end-filling with the Klenow fragment of *E. coli* DNA polymerase I and \( \alpha^{-32}\text{P}]dATP. The labeled DNAs were separated on 1% agarose gels in TAE (40 mM Tris acetate, 1 mM EDTA, pH 8) buffer, the gels were dried and exposed to x-ray film. The instabilities of the TRS tracts of the recombination products were determined using *Sphi*/BamHI digestion followed by end-labeling as described above. The products were resolved in 5 - 7 % polyacrylamide gels in TAE buffer. The lengths of the CTG•CAG inserts were calculated as described earlier (29). The primary structures of more than 35 individual recombination products were determined by direct DNA sequencing of one or both DNA strands.
RESULTS

Intramolecular System to Study Recombination Between (CTG•CAG)ₙ

Sequences – We used an intramolecular plasmid system to study recombination between TRS tracts, where two homologous repetitive sequences are located on the same plasmid molecule and are separated by nonhomologous intervening sequences.

Two homologous TRS tracts present on the same replicon can be oriented relative to each other as direct or inverted repeats (Fig. 1). The term “orientation” is used in this study to define the relative directionality between two recombining homologous sequences (direct and inverted repeats). The terms “orientation I” and “orientation II” refer to the orientation of the TRS sequences relative to the origin of replication; for example, for the plasmids containing (CTG•CAG)ₙ tracts in orientation I, the CTG repeat is in the leading strand template whereas, for the plasmids harboring (CAG•CTG)ₙ tracts, in orientation II, the CTG repeat is in the lagging strand template (28-30,42,43).

The recombination frequency as well as the types of final products of the intramolecular recombination event strongly depend on the relative orientation of the recombining sequences (Fig. 1) (18,21,24,44). The recombination event between direct repeats may lead to the deletion of one of the homologous tracts and any intervening sequences between the repeats (20,21,44-47). In the case of homologous TRS tracts, the intervening sequence separating the repeats will be also deleted. However, due to their repetitive nature, two homologous
CTG•CAG tracts can align and hybridize with each other in several different frames (the number of frames equals the number of repeats divided by 3). As a result of possible different alignments of the CTG•CAG sequences, the length of TRS tracts in the recombination products may vary from the minimum length required for recombination to occur to the maximum length determined by the size of both recombining homologous sequences. The intervening sequence separating the two TRS tracts is inviable since it lacks an origin of replication as well as the ampicillin resistance gene and therefore will be lost during cell division.

The recombination event between inverted repeats (Fig. 1, right panel) can lead to the inversion of the intervening sequence between the homologous repeats (44,48). This will result in an inversion of the direction of the ampicillin resistance gene and an inversion of the origin of replication. However, other types of products of intramolecular recombination between inverted repeats such as head-to-head inverted dimers have also been described previously (18,24).

*Plasmids Containing Direct and Inverted Repeats* - Intramolecular plasmid systems have been used widely for investigating the mechanisms of recombination and the influence of different factors on this process (16-19,21-24). We used this system to investigate recombination between CTG•CAG repeats in *E. coli*. For this study, we constructed and characterized a family of pBR322 derivatives (Fig. 2). Various lengths of CTG•CAG repeats (17, 67, 98, 165 and 175) were cloned into the EcoRI/HindIII and *Pvu*II sites of pBR322. Two
homologous TRS tracts inserted in both orientations (I and II) as direct and inverted repeats (Fig. 2, left and center columns) were separated by approximately 2,000 bp of the intervening sequence (Fig. 2, dotted region) and 2,300 bp of the intervening sequence harboring the unidirectional replication origin and the ampicillin resistance gene. Introduction of the X TRS insert into the EcoRI/HindIII site of pBR322 inactivated the tetracycline resistance gene (49). The cloning of the Y TRS insert into the PvuII site destroyed the rop gene of pBR322, which mediates the activity of RNAI. The latter resulted in an elevated copy number of the plasmids (50).

As controls, plasmids with non-repeating homologous sequences instead of the CTG•CAG repeats were constructed (Fig. 2). Two different non-repeating DNA fragments were used: the 564 bp fragment of bacteriophage λ DNA and the 354 bp fragment of exon 7/ intron 7 of the human DMPK gene (see Experimental Procedures for details). In addition, pBR322 derivatives containing one CTG•CAG tract (the X insert) and its isomeric GTC•GAC sequence (51) (the Y insert) were constructed (Fig. 2, right column) as controls for non-homologous TRS tracts in one plasmid.

All plasmids were maintained in E. coli HB101, which is RecA-. Previous studies showed that intramolecular plasmid recombination is not dependent on the function of the recA gene product (18-20,23). Thus, even the propagation of the plasmids in E. coli HB101 to obtain working stocks of plasmids can cause DNA rearrangements due to RecA independent recombination. In addition to the recombination events, cultivation of E. coli harboring plasmids with TRS tracts
leads to the genetic instability of repeating sequences manifested predominantly as deletion products (28-30,51,52). This applies mainly to the long uninterrupted CTG•CAG sequences such as (CTG•CAG)_{67} or longer (28,30,52). To eliminate the possibility of transformation by plasmids containing either large rearrangements caused by recombination (e.g. dimers, substantial deletions, duplications) or smaller deletions within the TRS tracts due to replication errors, all plasmids were subjected to extensive agarose-gel purification. The purity and sequence integrity of DNA was determined before transformation using restriction analyses and DNA sequencing. Only plasmids that met the above-mentioned criteria were subsequently used for transformation experiments.

Recombination Between Direct Repeats – The recombination behavior of the plasmids shown in Fig. 2 was studied in two *E. coli* strains that differed in their recombination capacity: AB1157 (parent) and JC10289 (RecA\(^{-}\)). For all plasmids, both single colony analyses and population experiments were performed (Experimental Procedures). The plasmids isolated from *E. coli* AB1157 and JC10289 (Fig. 3, lanes indicated by “+” and “-”, respectively) were analyzed by *Afl*III digestion. The unique *Afl*III recognition site is located between the origin of replication and the Y TRS insert, about 60 bp from the origin of replication. Thus, large rearrangements such as dimerization or deletion of the DNA segment between homologous sequences, which may result from recombination, can be detected. Restriction analyses of plasmids containing direct repeats (pRW4815, pRW4817, pRW4819, pRW4821, pRW4823,
pRW4825 and pRW4804) isolated from RecA⁺ and RecA⁻ E. coli showed bands of 4,500 – 5,500 bp in size, corresponding to the starting DNA (co-migrating on agarose gel with the plasmids used for transformation; Fig. 3A, lanes C) and shorter DNA fragments at approximately 2,500 – 3,000 bp. As revealed by restriction analyses of plasmids isolated from single colonies and DNA sequencing of several clones, the shorter fragments (at 2,500 - 3,000 bp) correspond to the recombination products between direct repeats, which harbor only one stretch of CTG•CAG repeats and lack the intervening sequence separating the two homologous TRS tracts. The same type of recombination products was observed in the case of pRW4804 containing the homologous non-repeating sequences (Fig. 3A).

On the other hand, the products of intramolecular deletion between homologous sequences have never been detected for plasmids harboring inverted repeats (pRW4816, pRW4818, pRW4820, pRW4822, pRW4824 and pRW4826) nor for plasmids containing CTG•CAG tracts and their isomeric, non-homologous sequence (pRW4830, pRW4831, pRW4832 and pRW4833; Figs. 3B and C).

Thus, we conclude that the predominant products of recombination between directly repeated CTG•CAG tracts are intramolecular deletions. Moreover, quantitative analyses of the data presented in Fig. 3A, obtained using Phosphorimager scanning of radioactively-labeled restriction fragments, showed that the amount of the recombination products was strongly dependent on the length of the CTG•CAG sequence. In the case of pRW4815 and pRW4817, both
containing (CTG\textbullet CAG)_{17}, the recombination product constituted \(~1\% of the total DNA isolated. For plasmids harboring (CTG\textbullet CAG)_{67} (pRW4819 and pRW4821) and (CTG\textbullet CAG)_{98} (pRW4823 and pRW4825), the recombination products accounted for \(~15\% and \~30\% of the total DNA, respectively. The plasmids containing 98 CTG\textbullet CAG repeats (410 bp of homologous sequence including human myotonic dystrophy flanking sequences and a fragment of the pUC19 poly linker) showed approximately 20 – 30 times higher propensity of recombination product formation when compared to the non-repeating 564 bp phage \lambda DNA (Fig. 3A, compare pRW4823 and pRW4825 with pRW4804).

Quantitative analyses of the data presented in Fig. 3A also revealed that the amount of the recombination products depends on the orientation of the CTG\textbullet CAG sequence relative to the origin of replication, with orientation II being more recombination-prone than orientation I (Fig. 3A). Also, this was confirmed later by using a genetic assay to study the frequency of recombination between the direct repeats (see below). The influence of the CTG\textbullet CAG orientation on the recombination frequency suggests an important involvement of replication mechanisms such as polymerase pausing and induction of DNA nicks in intramolecular recombination between the CTG\textbullet CAG sequences.

The products of intramolecular deletion between the direct repeats were detected in both RecA\textsuperscript{+} and RecA\textsuperscript{-} strains; however, \textit{E. coli} JC10289 (RecA\textsuperscript{-}) exhibited a lower recombination propensity than the isogenic RecA\textsuperscript{+} cells (Fig 3A, compare lanes indicated by “+” with “-”). These results are in agreement with previous studies (19,45) showing that intramolecular plasmid recombination does
occur efficiently independent of the recA gene function, although the presence of RecA increases the frequency of this process.

Furthermore, the spectrum of recombination products is different for plasmids containing short stretches of CTG•CAG as compared to the long tracts. Digestion of the recombination products from pRW4815 and pRW4817 (containing (CTG•CAG)_{17}) with AflIII showed a single band (within the resolution of the agarose gel) but recombination between the homologous sequences harboring 67 and 98 CTG•CAG repeats gave a set of products spanning a distance of at least 500 bp (Fig. 3A). This effect might be due to the higher instability of the longer TRS tracts present in the recombination products; however, it is more likely that the size variability of the CTG•CAG tracts in the recombination products increases with the length of the recombining homologous sequences.

**CTG•CAG Length, Orientation and Number of Tracts Does Not Affect the Doubling Time of E. coli Nor the Plasmid Copy Number** – The quantitation of the data obtained from the population experiments could be strongly biased, at least in principle, by different physiologies of bacterial cells containing recombination products versus cells harboring parental plasmids. Two major processes might cause enrichment of the recombinant plasmids in the cells and therefore influence the outcome of the experiments. First, the cells containing smaller plasmids with one CTG•CAG tract may have a growth advantage over cells harboring the larger parental plasmids with two TRS tracts. Second, the differences in size and the number of TRS present in the replicon could influence
the replicative advantage (copy number) of one type of plasmid over the other. Thus, studies were conducted to evaluate the magnitude of these potential influences.

The growth curves of *E. coli* AB1157 and JC10289 host strains harboring plasmids with two of the longest, uninterrupted TRS tracts studied (pRW4863 and pRW4865, with two (CTG•CAG)\textsubscript{165} tracts in orientations I and II, respectively) were compared to the growth curves of bacteria harboring recombination products with single TRS tracts of 21, 58 and \~92 CTG•CAG repeats. The doubling time (t\textsubscript{2}) calculated during the exponential phase of growth was almost identical for *E. coli* AB1157 harboring pRW4863 (t\textsubscript{2AB}=21.7±1.3 min.), pRW4865 (t\textsubscript{2AB}=22.0±1.4 min.) and for bacteria that harbored recombination products (22.7±1.4 min., 21.8±1.8 min. and 23.7±2.1 min. for plasmids containing the single tracts of 21, 58 and 92 repeats, respectively). The doubling time of *E. coli* JC10289 was lengthened by 15 – 25% for all plasmids studied. Approximately 20% difference in t\textsubscript{2} between AB1157 and JC10289 is observed regardless of the presence of the plasmid. The doubling time of bacteria harboring non-repeating DNA sequences (pRW4804) was \~5 – 10% shorter than for *E. coli* harboring plasmids with CTG•CAG repeats. There was no statistical difference in t\textsubscript{2} between pRW4804 (19.7±1.8 min.) and the recombination product of this plasmid (20.9±0.8 min). Thus, under our experimental conditions, no growth advantage of cells harboring the recombination products over cells harboring the recombination substrates was observed. These results are in agreement with the previous findings (30) that cells harboring plasmids with a shorter TRS tract
((CTG•CAG)_{17}) do not have a growth advantage over cells containing plasmids with (CTG•CAG)_{175} tract, so long as the cultures were maintained in the exponential phase of growth (even for several generations). Alternatively, the growth advantage was pronounced after E. coli passed through the stationary phase (30), which are conditions never employed in our studies.

In order to determine the difference in the replication propensities of plasmids with one or two TRS tracts, the copy numbers of pRW4865 (with two (CTG•CAG)_{165} tracts), pRW4815 (with two (CTG•CAG)_{17} tracts) and the recombination product (with one CTG•CAG sequence of 21 repeats) were analyzed. Using the detergent lysis method (39,40) we observed that the plasmid copy number is ~10% higher for the recombinant plasmid (148±9 copies per genome) than for pRW4865 (134±7 copies per genome) or pRW4815 (129±7 copies per genome) in E. coli AB1157. Thus, the copy number of the plasmid harboring a pair of short (CTG•CAG)_{17} tracts is very similar to the copy number of the plasmid carrying two long 165 repeats tracts. We can also conclude that the difference in copy numbers between recombination substrates and products is negligible.

It should be noted that recombination products (~3 Kbp plasmids) do not form multimeric forms (dimers, trimers etc.) with a high efficiency while maintained in E. coli AB1157 (RecA^+). However, recombination substrates are capable of forming large amounts of multimeric forms in the recombination proficient cells. Considering the oligomeric states of the plasmids, 134 copies of pRW4865 and 129 copies of pRW4815 account for 283 and 277 monomer-equivalents,
respectively. Thus, the copy number as well as the number of monomer-equivalents does not depend on the length of CTG•CAG tracts (in the range of 17 to 165 repeats). Determination of the copy number of the same plasmids in JC10289 (RecA*) revealed that both recombination substrates and products are maintained in *E. coli* at almost the same copy number.

These experiments showed clearly that parental plasmids with long repeat tracts do not have a replication disadvantage compared to the recombination products. Moreover, due to the approximately two times higher amount of monomer-equivalents present in the recombination proficient cells carrying plasmids with two TRS tracts, the frequency of recombination events leading to the formation of the smaller plasmids (calculated from the data presented in Fig. 3A) may be under rather than overestimated.

In summary, these data show only negligible effects of both the replicative advantage of recombinant plasmids over parental plasmids and the growth advantage of cells containing smaller plasmids with one TRS tract on the outcome of the population experiments shown in Fig. 3. On the other hand, the data obtained from the biochemical approach must be interpreted cautiously due to their lower precision, sensitivity and statistical significance compared to the genetic assay. Therefore, we established a genetic assay for determining the frequency of intramolecular recombination between two TRS tracts (see below).

*Recombination Between Inverted Repeats* – Experiments with plasmids containing the inverted CTG•CAG repeats were carried out under identical
conditions as described above for plasmids harboring direct repeats. In contrast to the plasmids containing direct repeats, when the two TRS tracts were oriented as inverted repeats, no products of intramolecular deletions were ever observed (Fig. 3B). This result was expected (18,24,44).

The predicted product of recombination between inverted repeats is a simple intramolecular inversion as shown in Fig. 1 (right panel) (44,48). More than two hundred colonies from recombination studies with pRW4816, pRW4818, pRW4820, pRW4822, pRW4824, pRW4826 and pRW4805 along with DNAs isolated from the population experiments were analyzed in both RecA+ and RecA- strains. NheI/AatII digestion was used to identify the inversion products. These two sites flank the X TRS and the analysis of the starting DNA should give rise to two fragments that are ~550 – 800 bp and 4100 – 4350 bp long (the size depends on the number of CTG•CAG repeats present). If the products of intramolecular inversion due to recombination are formed, two new bands should be detected on agarose gels (~2450 – 2700 bp and 2200-2450 bp in length). Unexpectedly (44,48), intramolecular inversions were never detected (data not shown).

Intramolecular recombination between inverted repeats can also lead to the formation of a head-to-tail dimer with complex DNA rearrangements (18,24). This kind of recombination product would be easily detected by AflIII digestion as well as by electrophoresis of supercoiled plasmid DNA on agarose gels. None of those two approaches showed formation of such recombination products.
Several factors can explain the failure of detection of recombination between inverted repeats. The lower frequency of recombination between inverted repeats in comparison to direct repeats was reported previously (44); therefore, this process may not be detectable by our radiolabelling methods. Even in the case of non-repeating sequences (pRW4804 and pRW4805), the products of recombination between direct repeats were detected in contrast to those of inverted repeats. Using biochemical methods (restriction digestion and radioactive labeling), we were able to detect products of the recombination events occurring with a frequency of $\geq 10^{-4}$. Furthermore, it is possible that the intrinsic properties of the TRS sequences (e.g. to form stable DNA structures (2), pause DNA polymerases (51,53,54) or cause the double-strand breaks (55-58)) favor a specific recombination/repair pathway, which in our experimental conditions strongly promotes recombination between direct repeats and/or inhibits inverted repeat recombination.

Frequency of Intramolecular Recombination Between Direct Repeats

Depends on the Length of the CTG•CAG Tracts – The green fluorescent protein (GFP) gene was cloned into the region of the plasmids that underwent deletion during recombination (for details see Experimental Procedures). The GFP cassette contains the GFPuv variant of the green fluorescent protein, which is expressed in E. coli under the control of the lac promoter (36). The GFP emits strong green fluorescence when irradiated with long wave-length UV light. The detection of the fluorescence does not require exogenous substrates or cofactors.
and is completely independent of the genetic background of bacterial host cells (59,60). Plasmids containing the GFP gene separating two TRS tracts were transformed into *E. coli* AB1157 and JC10289. In all experiments the transformations were performed with a large excess of cells to DNA molecules so that transformation should have occurred by a single plasmid molecule (61). The white colonies are formed only when the incoming plasmid undergoes recombination immediately after the transformation, leading to the loss of the GFP gene located between the direct repeats. When the incoming plasmid is established in the host cell and replicates several times, the expression of the GFP gene leads to fluorescent colony formation. The frequency of recombination was measured as the ratio of the number of white colonies to the total number of viable cells (Fig. 4A).

Sixteen plasmids containing the GFP cassette were constructed and used in our experiments (Figs. 4B and C). The two major factors found to influence the recombination frequency between direct repeats were the length of the TRS tract and the orientation of the CTG•CAG sequence relative to the origin of replication.

Fig. 4B shows that plasmids containing short, (CTG•CAG)₁₇ tracts in orientation I recombined in *E. coli* AB1157 with the frequency ~5 times lower than plasmids harboring non-repeating sequences. However, simple inversion of the orientation (pRW4817gfp) caused a statistically significant (P<0.001), 4 fold increase in recombination propensity of the (CTG•CAG)₁₇ sequences. Thus, the plasmids containing the (CTG•CAG)₁₇ tracts in orientation II recombined with a
frequency similar to non-repeating, homologous DNA fragments which were 200 – 400 bp longer.

The effect of length of the homologous TRS regions on the recombination frequency was dramatic. For CTG•CAG tracts in orientation I, lengthening the recombining sequences to 67, 98 and 165 repeats increased the rate of recombination 6.5, 8 and 60 times, respectively, in comparison to (CTG•CAG)17. A similar effect of the length of the recombining sequences was observed also for plasmids harboring TRS tracts in orientation II (Figs. 4B and 5). Hence, the long CTG•CAG tracts have a much higher propensity for recombination than shorter tracts; moreover, the frequency of recombination between the longest (CTG•CAG)165 sequences studied (pRW4863gfp and pRW4865gfp) was approximately 7 to 10 times higher (P<0.001) than the frequency observed for non-repeating sequences of comparable length (pRW4804gfp). In addition, the level of recombination between the 564 bp long λ phage DNA fragments was only slightly higher (11.9×10⁻³) than for the 354 bp DMPK gene fragments (9.2 – 10.4×10⁻³). This statistically insignificant difference (P=0.07) suggests that the frequency of recombination does not depend on the length of the recombining fragments in the case of non-repeating DNA sequences. These results are in agreement with previous studies (21) showing that the frequency of recombination between non-repeating DNA sequences (fragments of the tetracycline resistance gene) oriented as direct repeats increases as the length of the homologous sequences increases from 14 to 100 bp. Further lengthening of
the repeats (up to 854 bp) had little or no effect on the recombination frequency (21).

It should be pointed out that the rate of intramolecular recombination between long CTG•CAG tracts was extraordinarily high; recombinants were found with a frequency 1.5 – 12.6% (for plasmids containing 98 and 165 CTG•CAG repeats). Therefore, the recombination products could be easily detected and visualized in plasmids isolated from population experiments, as shown in Fig. 3A.

**CTG•CAG Tracts in Orientation II are More Susceptible to Recombination**– Although CTG•CAG tracts stimulate recombination in both orientations, a pronounced orientation-dependence of the frequency was observed. The frequencies of recombination were 4, 2 and 3.5 times higher for plasmids containing (CTG•CAG)₁₇, (CTG•CAG)₆₇, and (CTG•CAG)₉₈ in orientation II, respectively, than for plasmids harboring repeats of the same length but in orientation I (P<0.001). Surprisingly, in the case of the longest tracts studied ((CTG•CAG)₁₆₅), the orientation dependence was found to be the opposite. However, the frequency of recombination for both pRW4863gfp and pRW4865gfp was much higher than for plasmids harboring shorter tracts. The plasmid containing (CTG•CAG)₁₆₅ in orientation I showed a higher recombination propensity than plasmids with the (CTG•CAG)₁₆₅ in the orientation II (126×10⁻³ versus 81×10⁻³). The reason for this finding is uncertain but we believe that the instability of the CTG•CAG tracts contributes to this behavior. Long, uninterrupted CTG•CAG sequences (even containing 67 or 98 repeats) are extremely unstable
in plasmids cultivated in \textit{E. coli}. In addition, the CTG•CAG repeats in orientation II undergo deletions with a much higher rate than those in orientation I. Both pRW4863gfp and pRW4865gfp harbor the very long (CTG•CAG)$_{165}$ sequences; thus, it is essentially impossible to stably maintain them in \textit{E. coli}. Although, the preparation of pRW4863gfp (orientation I) used in these experiments contained only 5 – 10% deletions, the pRW4865gfp preparation (orientation II) contained approximately 30 – 35% deletions (estimated by restriction digestion followed by DNA labeling and calculated as the total amount of deletions from both TRS inserts). This difference in the TRS stability is likely responsible for the apparent lower frequency of recombination observed between the two (CTG•CAG)$_{165}$ inserts in orientation II. Thus, if this extreme level of instability was not encountered for the DNAs with orientation II, we anticipate that the frequencies of \( \sim 180 \times 10^{-3} \) would have been observed, amounting to a 36 fold enhancement compared to the shortest CTG•CAG tracts.

As expected, the formation of white colonies due to recombination was not detected for plasmids harboring inverted repeats (pRW4820gfp and pRW4822gfp) as well as for plasmids containing a CTG•CAG tract in the same plasmid as the isomeric GTC•GAC repeats (pRW4830gfp and pRW4831gfp, Fig. 4C). In the case of pBR322gfp, which contains no homologous repeating sequences, a single white colony was found (\( \sim 30,000 \) colonies were screened) and the restriction analysis of DNA isolated from that colony showed the existence of a point mutation (1 or 2 nucleotide deletion) within the GFP gene, which was obviously a sporadic event.
Intramolecular Recombination is Independent of RecA – Intramolecular recombination experiments with plasmids containing CTG•CAG repeats were done in both recombination proficient and deficient *E. coli* strains. In contrast to the significant reduction of the intermolecular recombination frequency by *recA* gene knockout (14,15), intramolecular plasmid recombination is known to proceed efficiently in *recA* deficient strains (19,45).

Similar to the results obtained with *E. coli* AB1157, the recombination frequency in JC10289 (*recA*) was strongly dependent on the length and orientation of the recombining CTG•CAG tracts (Figs. 4B and 5). The *recA* gene inactivation reduced the overall rates of intramolecular recombination by 2 to 11 fold in comparison to the isogenic recombination proficient *E. coli* cells. In the case of plasmids containing shorter CTG•CAG tracts (17 and 67 repeats), the effect of *recA* deletion was modest (2 – 4 fold decrease in recombination frequency). However, a stronger, 3 to 11 fold reduction in the recombination rate was detected for plasmids harboring (CTG•CAG)98 and (CTG•CAG)165. These results are in agreement with the previous studies, where the effect of *recA* mutation on intramolecular recombination varied from zero to a 40 fold decrease in frequency and was predominantly dependent on the length of the recombining sequences (reviewed in 17).

Instability of the CTG•CAG Sequences in the Recombination Products – To study the instability of the CTG•CAG tract resulting from intramolecular
recombination, plasmids were isolated from white colonies and analyzed by agarose gel electrophoresis. The electrophoretic migration of recombinants showed that all plasmids (~400 DNA samples) isolated from white colonies lost a significant portion (~2Kbp) of the vector backbone (Fig. 6A). Other types of rearrangements were not observed. In order to further characterize the structure of the recombinants, a total of 37 individual recombination products, representing plasmids harboring TRS of different lengths, were subjected to the DNA sequencing. The sequence analyses revealed that all recombination products studied harbored only one TRS tract flanked by the human myotonic dystrophy sequences and fragments of the polylinker (Fig. 6B). The smallest recombination product analyzed carried only 13 triplet repeats (Fig. 6B), whereas the largest entirely sequenced TRS tract contained 165 CTG•CAG repeats. This expansion product was created by a recombination event between two (CTG•CAG)\textsubscript{98} inserts. In the case of recombinants containing very short CTG•CAG inserts, we were able to analyze, using a single sequencing reaction, the entire TRS tract along with the human flanking sequences and segments of the polylinker (Fig. 6B). In addition, pBR322 sequences that were originally separated in the parental plasmids by a distance exceeding 2 Kbp (E and P sites in Fig. 6B) could also be detected. Thus, these results proved that recombination occurred between two homologous TRS inserts resulting in the deletion of the intervening sequence.

In order to analyze the size of the TRS inserts in a large number of recombinants, plasmids were subjected to SphI/BamHI restriction digestion followed by end-labeling and polyacrylamide gel electrophoresis (Fig. 7).
Although the parental plasmids contain seven recognition sites for *SphI/BamHI*, the recombinants harbor unique *SphI* and *BamHI* restriction sites (the remaining five are lost during recombination along with the 2 Kbp of intervening sequence (Fig. 6B)). Therefore, digestion by these restriction enzymes splits the recombinant plasmids into an ~2.3 Kbp pBR322 fragment (identical for all plasmids studied containing the origin of replication and the ampicillin resistance gene; Fig. 6B) and the CTG•CAG-containing inserts (Fig. 7A-D).

At least 40 individual isolates from each recombination experiment were analyzed by *SphI/BamHI* restriction digestion. Fig. 7 shows the polyacrylamide gel analyses of typical data obtained for pRW4815gfp, pRW4819gfp, pRW4823gfp and pSF3gfp (all plasmids contain the CTG•CAG tracts in orientation I). There were no significant differences in the overall TRS size distributions in the recombination products from plasmids harboring CTG•CAG sequences in orientations II and I.

In the case of pRW4815gfp and pRW4817gfp, the progenitor (starting length (CTG•CAG)$_{17}$) was retained in more than 80% of the recombination products analyzed (Figs. 7A and 8). Only 5 isolates harbored longer tracts with (CTG•CAG)$_{23}$ being the longest (Fig. 7A, lane 12) and 7 clones contained deleted products (12 repeats was the shortest). These data imply that the homologous, non-repeating flanking sequences may be an important factor for recombination between short repeats (particularly (CTG•CAG)$_{17}$).

The most interesting recombinational behavior was found for the plasmids harboring (CTG•CAG)$_{67}$ sequences (Figs. 7B and 8). Analyses of the
pRW4819gfp and pRW4821gfp recombination products (Fig. 7B) revealed that more than 30% contained expanded CTG•CAG tracts. Deletions and retentions of the progenitor insert length were detected in ~30% and 40% of analyzed clones, respectively (Fig. 8). The smallest CTG•CAG tract found among the recombination products had 15 repeats and the longest, expanded CTG•CAG tract contained 130 repeats.

For plasmids containing 98 or 165 CTG•CAG repeats (pRW4823gfp, pRW4825gfp, pRW4863gfp and pRW4865gfp), only ~5% of the isolates maintained the size of the progenitor sequence; the majority of the clones analyzed (70 – 80%) harbored deleted CTG•CAG sequences (Figs. 7C and 8). Thus, these sequences were very prone to deletions and expansions were observed infrequently. The longest TRS tract found in the recombination products had 172 and 289 CTG•CAG repeats for pRW4823gfp and pRW4863gfp, respectively.

Hence, we conclude that the intramolecular recombination between CTG•CAG repeats results in the genetic instabilities of the TRS tracts.

Effect of Interruptions – It was previously demonstrated that as little as 2.8% of heterology might reduce the frequency of recombination more than 1000 fold between non-repeating sequences in an E. coli plasmid system (24). The GFP gene-containing derivative of pSF3 (Fig. 2) was used to analyze the influence of sequence interruptions on the frequency of intramolecular recombination between TRS tracts. pSF3gfp harbors two CTG•CAG inserts in orientation I, 98
and 175 repeats in length. The longer tract contains two G to A interruptions at repeats 28 and 69 (Experimental Procedures) but the other tract has no interruptions. Note that all TRS tracts studied herein (Fig. 2) containing 165 or fewer repeats are uninterrupted whereas (CTG•CAG)$_{175}$ is interrupted (Experimental Procedures). The frequency of recombination for pSF3gfp was ~2 fold higher than for pRW4823gfp containing two (CTG•CAG)$_{98}$ tracts in orientation I (both with no interruptions) and ~4 fold lower than the recombination frequency observed in the case of pRW4863gfp ((CTG•CAG)$_{165}$). In fact, the insert harboring 175 repeats with two interruptions contains a tract of 106 pure CTG•CAG repeats which is long enough to efficiently recombine with the (CTG•CAG)$_{98}$ sequence. This result showed that the presence of these two interruptions has no influence on frequency of intramolecular recombination between the CTG•CAG repeats.

Although the presence of these two interruptions had no influence on the frequency of recombination, their effect on the length of the CTG•CAG tracts in the recombination products was pronounced (Figs. 7D and 8). Restriction analyses of the recombination products revealed that ~60% of clones contained CTG•CAG inserts of 175 repeats or longer, whereas, for pRW4823gfp (containing two uninterrupted (CTG•CAG)$_{98}$ inserts), less than 30% of the recombinants had 98 or more CTG•CAG repeats (Figs. 7C and D)

Similar findings regarding the influence of interruptions on intermolecular recombination between CTG•CAG sequences were described recently (14,15). The results obtained using a two-plasmid system demonstrated that both
multiple-fold expansions and increase of frequency of recombination were
observed when one of the recombining sequences (usually cloned into pACYC)
contained an interrupted CTG•CAG tract and the other of the recombining
plasmids harbored an uninterrupted CTG•CAG tract. The presence of two
interruptions in each of the recombining CTG•CAG sequences reduced the
frequency recombination as well as inhibited the formation of recombination-
mediated TRS expansions (14,15).
DISCUSSION

Previous studies showed that repetitive sequences, such as interspersed Alu repeats (62) as well as tandemly repeated minisatellites (63,64) and microsatellites (25,65-68), exhibit a higher recombination capacity than non-repeating DNA tracts. Also, it has been postulated that TRSs, such as CTG•CAG repeats, may be recombination hot spots (69,70).

The principal conclusions from our studies are the following: first, long CTG•CAG microsatellites are preferred sites of intramolecular recombination in E. coli. The frequency of recombination between two directly repeated CTG•CAG tracts is up to ~10 times higher than between two non-repeating sequences (λ DNA and DMPK DNA) of similar length in recombination proficient cells. Second, when the TRS tracts are oriented as inverted repeats, no products of homologous recombination were observed. Third, the effect of length of the homologous CTG•CAG tracts on the recombination frequency is dramatic. We found that increasing the length of the homologous sequences from 17 to 165 CTG•CAG repeats showed a 60-fold increase in the recombination frequency between direct repeats. This effect was similar for TRS tracts in orientation I and in orientation II. Fourth, a pronounced orientation-dependence in the frequency was observed, although directly repeated CTG•CAG tracts recombined efficiently in both orientations relative to the origin of replication. TRS tracts present in orientation II (CTG repeats on the lagging strand template) are more susceptible
to recombination. Fifth, intramolecular recombination between CTG•CAG tracts was observed in both parental and recA E. coli strains, but the frequency of this process was elevated by 2 to 11 fold in the recombination proficient cells. This effect is dependent on the length of the recombining sequences. Sixth, intramolecular recombination between CTG•CAG tracts led to high genetic instability (deletions and expansions) of the repeating sequences.

Several features of the TRS tracts may contribute to their recombinogenic behavior. During the recombination process, two CTG•CAG repeat tracts can hybridize with each other in many registers; in contrast, the two control, non-repeating sequences (564 bp λ DNA and 354 bp DMPK DNA) can align only in one frame. The number of possible alignments between two homologous TRS tracts increases with the number of repeats present in the recombining sequences, which can have an influence on the kinetics of the synapsis step of homologous recombination.

In contrast to the extremely frequent intramolecular events between direct repeats, we were unable to detect any products of recombination between the inverted repeats. Thus, recombination between head-to-head oriented TRS was reduced by at least 100 fold compared to the head-to tail oriented CTG•CAG inserts. These results are consistent with previous data from studies on plasmid (44) and chromosome (71,72) recombination in prokaryota, yeast (73), mammalian chromosomes (74) and mammalian extrachromosomal elements.
A dramatic difference in the recombination frequency between direct and inverted repeats was also observed for site-specific recombination systems such as the Tn3 resolvase (76-78). For site-specific recombination systems, the orientation-dependence was attributed to the geometry of the DNA (78). Studies on recombination between inverted repeats in the *Salmonella typhimurium* chromosome showed that the inversion process depended predominantly on the chromosomal localization of the head-to-head oriented repeats (71). Some DNA sequences separating the recombining repeats were shown to be permissive for recombination while others did not stimulate recombination (nonpermissive) (71,79). The most comprehensive studies in *E. coli* plasmid systems to resolve the orientation-dependence revealed that homologous recombination between inverted repeats occurs predominantly via nonconservative pathways (44,80). Thus, homologous recombination leads to the formation of linear, inviable recombinants from circular plasmid substrates harboring inverted repeats (44). We favor the idea, suggested earlier for recombination between non-repeating sequences in yeast (73), that there is more than one efficient recombination pathway leading to the high frequency of intramolecular deletions observed for directly repeated CTG•CAG sequences. In contrast, only a completely conservative event (reciprocal exchange) can cause the formation of the predicted inversion between two homologous TRS tracts in the head-to-head orientation.
Intramolecular recombination depends on the relative orientation of recombining TRS inserts to each other (direct and inverted tracts) as well as on the orientation of the pair of the direct repeats relative to the origin of replication. A significantly higher frequency of recombination was observed for the pair of CTG•CAG repeats in orientation II (when the CTG repeats are on the lagging strand template). These results confirm a tight connection between formation of the stable secondary structures by CTG•CAG repeats, replication arrest and recombination. A higher propensity for recombination between CTG•CAG tracts in orientation II is in agreement with the formation of more stable hairpin structures by CTG repeats. Furthermore, the arrest of the replication fork progression occurred in vivo, predominantly when the CTG sequence was present on the lagging strand template (in orientation II) (54).

Intramolecular recombination between directly repeated CTG•CAG tracts also occurs efficiently in a RecA` E. coli. However, the presence of the recA gene product increases, in a length-dependent manner, the rate of intramolecular deletion by 2 – 11 fold in comparison to the isogenic RecA` cells. Previous studies (reviewed in 17) revealed the capacity of RecA` cells to affect intramolecular recombination. Hence, we conclude that at least two types of recombination pathways, RecA dependent as well as RecA independent, are responsible for recombination between CTG•CAG tracts in E. coli. Intramolecular deletions between direct repeats can occur by a RecA independent single-strand annealing pathway (21,47,81,82). Previous studies
also showed that these types of recombination products could result from slippage during DNA replication (22,47,82). Therefore, we cannot exclude that replication misalignment may be partially responsible for intramolecular deletions between direct repeats observed in our study. In addition, RecA dependent mechanisms such as crossing-over and half crossing-over (44,80) may contribute significantly to the high frequency of recombination observed in the case of long CTG•CAG sequences.

Recent data showed that recombination between TRS tracts leads to large scale deletions and expansions within tandem repeat sequences at high frequency (14,15,27). We found also that intramolecular recombination between long, uninterrupted CTG•CAG tracts is a source of great instability of the TRS tracts. As expected (14,15), recombination had no significant influence on the stability of short inserts, containing 17 CTG•CAG repeats. In addition, the presence of the G to A interruptions in one of recombining sequences reduced the length variability of the CTG•CAG tracts observed in recombination products. We suggest that the interruptions disturb the homogeneity in the CTG•CAG repeat units and thus decrease the number of possible alignments between the two recombining TRS tracts, therefore leading to the genetic stabilization of the repeating sequence.

This work, as well as the accompanying study on the frequency of intermolecular recombination between long CTG•CAG sequences (27), shows
their very high recombination potential in *E. coli*. Their recombination hot spot characteristics and their capacity to expand by recombination (14,15,26,27) may be responsible for the genome instabilities observed in humans. Several cases of the involvement of recombination processes in TRS expansions in humans were described (reviewed in 14,83). Also, taking into account the statistical overrepresentation of trinucleotide microsatellites in eukaryota (84), the frequent recombination events between TRS tracts may be a source of mutations (deletions and inversions) leading to genetic diseases. In addition, recombination between TRS may have an important evolutionary role (70) by promoting rearrangements of genetic information within different loci leading to the formation of novel genes.
REFERENCES

1. van Belkum, A., Scherer, S., van Alphen, L., and Verbrugh, H. (1998) Microbiol. Mol. Biol. Rev. 62, 275-293

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

3. De Bolle, X., Bayliss, C. D., Field, D., van de Ven, T., Saunders, N. J., Hood, D. W., and Moxon, E. R. (2000) Mol. Microbiol. 35, 211-222

4. Ren, Z., Jin, H., Whitby, P. W., Morton, D. J., and Stull, T. L. (1999) J. Bacteriol. 181, 5865-5870

5. Lysnyansky, I., Sachse, K., Rosenbusch, R., Levisohn, S., and Yogev, D. (1999) J. Bacteriol. 181, 5734-5741

6. Glew, M. D., Baseggio, N., Markham, P. F., Browning, G. F., and Walker, I. D. (1998) Infect. Immun. 66, 5833-5841

7. Liu, L., Dybvig, K., Panangala, V. S., van Santen, V. L., and French, C. T. (2000) Infect. Immun. 68, 871-876

8. Vogelstein, B., and Kinzler, K. W. (1998) The Genetic Basis of Human Cancer, McGraw-Hill U.S.A. publishers
9. Leach, F. S., Nicolaides, N. C., Papadopoulos, N., Liu, B., Jen, J.,
Parsons, R., Peltomaki, P., Sistonen, P., Aaltonen, L. A., Nystrom-Lahti,
M., Guan, X. Y., Zhang, G. J., Meltzer, P. S., Yu, J. W., Kao, F. T., Chen,
D. A., Cerosaletti, D. M., Fournier, R. E. K., Todd, S., Lewis, T., Leach, R.
J., Naylor, S. L., Green, J., Jass, J., Watson, P., Lynch, H. T., Trent, J. M.,
de la Chapelle, A., Kinzler, K. W., and Vogelstein, B. (1993) Cell 75, 1215-
1225

10. Thibodeau, S. N., Bren, G., and Schaid, D. (1993) Science 260, 816-819

11. Lengauer, C., Kinzler, K. W., and Vogelstein, B. (1998) Nature 396, 643-
649

12. Loeb, K. R., and Loeb, L. A. (2000) Carcinogenesis 21, 379-385

13. Bowater, R. P., and Wells, R. D. (2000) Prog. Nucleic Acid Res. Mol. Biol.
66, 159-202

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

15. Jakupciak, J. P., and Wells, R. D. (2000) J. Biol. Chem. 275, 40003-40013

16. Lovett, S. T., Gluckman, T. J., Simon, P. J., Sutera, V. A. J., and Drapkin,
P. T. (1994) Mol. Gen. Genet. 245, 294-300

17. Lovett, S. T., Drapkin, P. T., Sutera, V. A. J., and Gluckman-Peskind, T. J.
(1993) Genetics 135, 631-642
18. Bi, X., and Liu, L. F. (1996) *Proc. Natl. Acad. Sci. U. S. A.* **93**, 819-823

19. Bi, X., and Liu, L. F. (1994) *J. Mol. Biol.* **235**, 414-423

20. Bi, X., Lyu, Y. L., and Liu, L. F. (1995) *J. Mol. Biol.* **247**, 890-902

21. Bi, X., and Liu, L. F. (1996) *J. Mol. Biol.* **256**, 849-858

22. Feschenko, V. V., and Lovett, S. T. (1998) *J. Mol. Biol.* **276**, 559-569

23. Morag, A. S., Saveson, C. J., and Lovett, S. T. (1999) *J. Mol. Biol.* **289**, 21-27

24. Lyu, Y. L., Lin, C. T., and Liu, L. F. (1999) *J. Mol. Biol.* **285**, 1485-1501

25. Murphy, K. E., and Stringer, J. R. (1986) *Nucleic Acids Res.* **14**, 7325-7340

26. Cemal, C. K., Huxley, C., and Chamberlain, S. (1999) *Gene* **236**, 53-61

27. Pluciennik, A., Iyer, R. R., Napierala, M., Larson, J. E., Filutowicz, M., and Wells, R. D. (2001) *J. Biol. Chem.* (submitted) (accompanying manuscript)

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

29. Pluciennik, A., Iyer, R. R., Parniewski, P., and Wells, R. D. (2000) *J. Biol. Chem.* **275**, 28386-28397
30. Bowater, R. P., Rosche, W. A., Jaworski, A., Sinden, R. R., and Wells, R. D. (1996) *J. Mol. Biol.* **264**, 82-96

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

32. Mahadevan, M., Tsilfidis, C., Sabourin, L., Shutler, G., Amemiya, C., Jansen, G., Neville, C., Narang, M., Barcelö, J., O’Hoy, K., Leblond, S., Earle-MacDonald, J., de Jong, P. J., Wieringa, B., and Korneluk, R. G. (1992) *Science* **255**, 1253-1255

33. Fu, Y.-H., Pizzuti, A., Fenwick, R. G., Jr., King, J., Rajnarayan, S., Dunne, P. W., Dubel, J., Nasser, G. A., Ashizawa, T., de Jong, P. J., Wieringa, B., Korneluk, R. G., Perryman, M. B., Epstein, H. F., and Caskey, C. T. (1992) *Science* **255**, 1256-1258

34. Brook, J. D., McCurrach, M. E., Harley, H. G., Buckler, A. J., Church, D., Aburatani, H., Hunter, K., Stanton, V. P., Thirion, J. P., Hudson, T., Sohn, R., Zemelman, B., Snell, R. G., Rundle, S. A., Crow, S., Davies, J., Shelbourne, P., Buxton, J., Jones, C., Juvonen, V., Johnson, K., Harper, P. S., Shaw, D. J., and Housman, D. E. (1992) *Cell* **68**, 799-808

35. Mahadevan, M. S., Amemiya, C., Jansen, G., Sabourin, L., Baird, S., Neville, C. E., Wormskamp, N., Segers, B., Batzer, M., and Lamerdin, J. (1993) *Hum. Mol. Genet.* **2**, 299-304
36. Crameri, A., Whitehorn, E. A., Tate, E., and Stemmer, W. P. (1996) *Nat. Biotech.* **14**, 315-319

37. Bachmann, B. J. (1987) *Derivations and Genotypes of Some Mutant Derivatives of Escherichia coli K-12* (Neidhardt, F. D., ed., Ed.), Am. Soc. for Microbiol., Washington, DC

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

39. Chiang, C. S., and Bremer, H. (1988) *Plasmid* **20**, 207-220

40. Lin-Chao, S., and Bremer, H. (1986) *Mol. Gen. Genet.* **203**, 143-149

41. Blattner, F. R., Plunkett, G., 3rd, Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J. D., Rode, C. K., Mayhew, G. F., Gregor, J., Davis, N. W., Kirkpatrick, H. A., Goeden, M. A., Rose, D. J., Mau, B., and Shao, Y. (1997) *Science* **277**, 1453-1474

42. Maurer, D. J., O’ Callaghan, B. L., and Livingston, D. M. (1996) *Mol. Cell. Biol.* **16**, 6617-6622

43. Freudenreich, C. H., Stavenhagen, J. B., and Zakian, V. A. (1997) *Mol. Cell. Biol.* **17**, 2090-2098

44. Yamamoto, K., Takahashi, N., Fujitani, Y., Yoshikura, H., and Kobayashi, I. (1996) *Genetics* **143**, 27-36
45. Laban, A., and Cohen, A. (1981) *Mol. Gen. Genet.* **184**, 200-207

46. Bierne, H., Ehrlich, S. D., and Michel, B. (1997) *EMBO J.* **16**, 3332-3340

47. Bzymek, M., and Lovett, S. (2001) *Genetics* **158**, 527-540

48. Ennis, D. G., Amundsen, S. K., and Smith, G. R. (1987) *Genetics* **115**, 11-24

49. Jaworski, A., Blaho, J. A., Larson, J. E., Shimizu, M., and Wells, R. D. (1989) *J. Mol. Biol.* **207**, 513-526

50. Stewart, G. S., Lubinsky-Mink, S., Jackson, C. G., Cassel, A., and Kuhn, J. (1986) *Plasmid* **15**, 172-181

51. Ohshima, K., Kang, S., Larson, J. E., and Wells, R. D. (1996) *J. Biol. Chem.* **271**, 16773-16783

52. Jaworski, A., Rosche, W. A., Gellibolian, R., Kang, S., Shimizu, M., Bowater, R. P., Sinden, R. R., and Wells, R. D. (1995) *Proc. Natl. Acad. Sci. U. S. A.* **92**, 11019-11023

53. Kang, S., Ohshima, K., Shimizu, M., Amirhaeri, S., and Wells, R. D. (1995) *J. Biol. Chem.* **270**, 27014-27021

54. Samadashwily, G. M., Raca, G., and Mirkin, S. M. (1997) *Nat. Genet.* **17**, 298-304
55. Freudenreich, C. H., Kantrow, S. M., and Zakian, V. A. (1998) *Science* **279**, 853-856

56. Jankowski, C., Nasar, F., and Nag, D. K. (2000) *Proc. Natl. Acad. Sci. U. S. A.* **97**, 2134-2139

57. Richard, G. F., Dujon, B., and Haber, J. E. (1999) *Mol. Gen. Genet.* **261**, 871-882

58. Richard, G. F., Goellner, G. M., McMurray, C. T., and Haber, J. E. (2000) *EMBO J.* **19**, 2381-2390

59. Suarez, A., Guttler, A., Stratz, M., Staendner, L. H., Timmis, K. N., and Guzman, C. A. (1997) *Gene* **196**, 69-74

60. Inouye, S., Ogawa, H., Yasuda, K., Umesono, K., and Tsuji, F. I. (1997) *Gene* **189**, 159-162

61. Hanahan, D., Jessee, J., and Bloom, F. R. (1991) *Methods Enzymol.* **204**, 63-113

62. Morgan, N. V., Tipping, A. J., Joenje, H., and Mathew, C. G. (1999) *Am. J. Hum. Genet.* **65**, 1330-1341

63. Wahls, W. P., Wallace, L. J., and Moore, P. D. (1990) *Cell* **60**, 95-103

64. Jeffreys, A. J., Barber, R., Bois, P., Buard, J., Dubrova, Y. E., Grant, G., Hollies, C. R., May, C. A., Neumann, R., Panayi, M., Ritchie, A. E., Shone,
A. C., Signer, E., Stead, J. D., and Tamaki, K. (1999) Electrophoresis 20, 1665-1675

65. Treco, D., and Arnheim, N. (1986) Mol. Cell. Biol. 6, 3934-3947

66. Blaho, J. A., and Wells, R. D. (1989) Prog. Nucleic Acid Res. Mol. Biol. 37, 107-126

67. Gendrel, C. G., Boulet, A., and Dutreix, M. (2000) Genes Dev. 14, 1261-1268

68. Dutreix, M. (1997) J. Mol. Biol. 273, 105-113

69. Skowronski, J., Plucienniczak, A., Bednarek, A., and Jaworski, J. (1984) J. Mol. Biol. 177, 399-416

70. Young, E. T., Sloan, J. S., and Van Riper, K. (2000) Genetics 154, 1053-1068

71. Segall, A., Mahan, M. J., and Roth, J. R. (1988) Science 241, 1314-1318

72. Mahan, M. J., and Roth, J. R. (1988) Genetics 120, 23-35

73. Prado, F., and Aguilera, A. (1995) Genetics 139, 109-123

74. Liskay, R. M., and Stachelek, J. L. (1986) Proc. Natl. Acad. Sci. U. S. A. 83, 1802-1806
75. Kitamura, Y., Yoshikura, H., and Kobayashi, I. (1990) *Mol. Gen. Genet.* 222, 185-191
76. Grindley, N. D., and Reed, R. R. (1985) *Annu. Rev. Biochem.* 54, 863-896
77. Watson, M. A., Boocock, M. R., and Stark, W. M. (1996) *J. Mol. Biol.* 257, 317-329
78. Droge, P., and Cozzarelli, N. R. (1989) *Proc. Natl. Acad. Sci. U. S. A.* 86, 6062-6066
79. Mahan, M. J., and Roth, J. R. (1991) *Genetics* 129, 1021-1032
80. Takahashi, N. K., Yamamoto, K., Kitamura, Y., Luo, S. Q., Yoshikura, H., and Kobayashi, I. (1992) *Proc. Natl. Acad. Sci. U. S. A.* 89, 5912-5916
81. Kuzminov, A. (1999) *Microbiol. Mol. Biol. Rev.* 63, 751-813
82. Bzymek, M., and Lovett, S. T. (2001) *Proc. Natl. Acad. Sci. U. S. A.* 98, 8319-8325
83. Jakupciak, J. P., and Wells, R. D. (2000) *IUBMB Life* 50, 355-359
84. Cox, R., and Mirkin, S. M. (1997) *Proc. Natl. Acad. Sci. U.S.A.* 94, 5237-5242
**FOOTNOTES**

*This research was supported by grants GM52982, NS37554, and ES11347 from the National Institutes of Health, the Robert A. Welch Foundation, and the Friedreich’s Ataxia Research Alliance.

§To whom correspondence should be addressed: Center for Genome Research, Institute of Biosciences and Technology, Texas A&M University, Texas Medical Center, 2121 W. Holcombe Blvd. Houston, TX 77030-3303; Tel.: (713) 677-7651; Fax: (713) 677-7689; Email: rwells@ibt.tamu.edu

The abbreviations used are: TRS, trinucleotide repeat sequence(s); bp, base pair(s); GFP, green fluorescent protein; DMPK, dystrophy myotonica-protein kinase gene.
ACKNOWLEDGEMENTS

We thank Drs. John Wilson, Marcin Filutowicz, Ravi R. Iyer and John P. Jakupciak for helpful discussions and Liliana E. Harris for critically reading the manuscript.
LEGENDS TO FIGURES

Fig. 1. **The products of intramolecular recombination between CTG•CAG tracts oriented as direct repeats or as inverted repeats.** The homologous recombination between direct repeats (left panel) leads to the formation of a smaller plasmid containing only one CTG•CAG tract; the DNA fragment which originally separated the two TRS tracts (shown as a dotted area) is deleted. This deleted fragment is inviable due to the absence of an origin of replication and the ampicillin resistance gene and will therefore be lost. In the case of the inverted repeats (right panel), the recombination event between two homologous sequences leads to the inversion of the sequence separating the repeats shown here as an inversion of the direction of the replication origin and the ampicillin resistance gene. The ampicillin resistance gene (Amp) is designated as a white arrow. The gray arrow shows the orientation of the unidirectional origin of replication (ori). A portion of the CTG•CAG tracts is black and a second portion has a white background to illustrate the location and the consequences of the recombination events.

Fig. 2. **Plasmids used in this study.** All plasmids are derivatives of pBR322 and contain two inserts (X and Y) oriented as direct repeats (left column) or inverted repeats (central column). Control plasmids harboring two non-homologous TRS are shown in the right column. TRS inserts as well as a non-repeating DNA sequences were cloned into the HindIII/EcoRI (X TRS insert) or PvuII (Y TRS insert) sites of pBR322 (for details, see Experimental Procedures).
The inserts containing 17, 67, 98 and 165 CTG•CAG repeats as well as the (GTC•GAC)_{79} insert are homogeneous (i.e. are perfect repeating sequences and contain no interruptions). The (CTG•CAG)_{175} sequence present in pSF3 and pSF4 is not a pure CTG•CAG tract but contains two G to A interruptions at repeats 28 and 69. The actual sequences of the leading strand templates of the TRS inserts are shown for all plasmids. Thus, CTG•CAG and CAG•CTG inserts correspond to orientation I and orientation II, respectively (28-30). The ampicillin resistance gene (Amp) is designated as a white arrow. The gray arrow shows the approximate position and direction of the origin of replication (ori).

Fig. 3. **Restriction analyses of products of the intramolecular recombination events between direct or inverted repeats or non-homologous sequences.** Plasmids were isolated from *E. coli* AB1157 (RecA^+^) and JC10289 (RecA^-^) cultures that were grown until the late log phase. The DNA was linearized with *Afl*III, end-labeled and electrophoretically separated through 1% agarose gels in TAE buffer. The starting material (the plasmids used for transformation of the *E. coli* strains) is indicated as “C”. Recombination proficient and deficient strains are shown as “+” or “−”, respectively. The 1-kilobase pair DNA ladder (Life Technologies, Inc.) was used as a size marker and the sizes of these bands are indicated (left sides of gels). Panel A, the results of *Afl*III digestion of plasmids containing direct repeats. Brackets designate the full-length plasmids (~5 Kbp) as well as the products of intramolecular deletion due to recombination between the repeated tracts (at ~3 Kbp). Panel B, *Afl*III
digestion products of plasmids containing the inverted repeats. Panel C shows the AflIII digestion products of control plasmids containing non-homologous TRS.

Fig. 4. **Frequency of recombination between direct repeats.** The strategy used for determining the recombination frequencies is shown in panel A. The plasmids containing the TRS sequences shown in Fig. 2 were modified as described under Experimental Procedures. Briefly, a 0.8 Kbp fragment of the vector sequence separating the repeats was replaced by the green fluorescent protein gene (GFP). The designations of the plasmids are identical to those shown in Fig. 2 except for the “gfp” suffix. After transformation, the *E. coli* cells were plated onto LB plates containing ampicillin (100 µg/ml) and incubated for 16 hr at 37 °C. The fluorescence of the colonies harboring the nonrecombined plasmids was detected by exposing the colonies to a long-wave UV lamp. Panel B, the recombination frequency was measured as the ratio of the number of white colonies to the total number of viable cells (both fluorescent green and white colonies). For each plasmid, three or more independent experiments were performed and at least 7,000 colonies were counted, except for pRW4863gfp and pRW4865gfp where ~2000 colonies were counted. The frequency was calculated as the mean of the data collected from all experiments. R represents relative frequency of recombination and is calculated relative to the frequency of recombination observed for pRW4815gfp harboring a pair of (CTG•CAG)₁₇ inserts. Panel C, the frequency of white colony formation in experiments conducted with control plasmids in *E. coli* AB1157.
Fig. 5. **The effect of length and orientation on frequency of intramolecular recombination between direct repeat tracts.** Circles represent the plasmids containing the CTG•CAG cloned in orientation I (pRW4815gfp, pRW4819gfp, pRW4823gfp and pRW4863gfp); diamonds represent the plasmids containing the CTG•CAG cloned in orientation II (pRW4817gfp, pRW4821gfp, pRW4825gfp and pRW4865gfp); the triangles represent the plasmids containing the non-repeating DNA sequences (pRW4804gfp, pRW4871gfp and pRW4873gfp). Data from experiments performed in *E. coli* AB1157 are shown on the main graph; the inset shows the data from *E. coli* JC10289 (RecA−). The standard deviations are shown by the error bars. The homologous sequences are composed of the CTG•CAG repeats as well as the human flanking sequences and segments of the polylinker.

Fig. 6. **Analysis of products of intramolecular recombination between CTG•CAG repeats.** Panel A, agarose gel analysis of the plasmid DNAs isolated from the single colonies from *E. coli* AB1157 after transformation with pRW4823gfp. Lanes 1 – 6, DNA isolated from white colonies; lanes 7 – 12, plasmids isolated from fluorescent colonies. The sizes of bands of the supercoiled DNA ladder (Sc) are shown on the left. Panel B, sequence analysis of an intramolecular recombination product harboring 13 CTG•CAG repeats. A schematic diagram of the recombinant plasmid is shown on the right. The recognition sites for *BamH*I and *Sph*I used to determine the sizes of TRS tracts in
the recombination products (Fig. 7) are indicated. $E$ and $P$ indicate the cloning sites for the X TRS insert (EcoRI) and the Y TRS insert (PvuII), respectively. The numbers in parentheses represent the original positions of these restriction sites on the pBR322 map.

**Fig. 7.** The influence of CTG•CAG length and sequence interruptions on recombination-mediated instability. The plasmids were isolated from white colonies and were digested with BamHI/SphI to release the TRS containing inserts. Labeled DNA fragments were separated by 5 – 7% polyacrylamide gel electrophoresis to determine the lengths of the CTG•CAG sequences. Each panel shows the analysis of approximately 30 individual colonies. A 1 Kbp ladder and TRS size standard (M) were used to determine the lengths of the CTG•CAG tracts. The TRS size standard (bands identified by arrows) contains four BamHI/SphI fragments containing 17, 67, 98 and 175 CTG•CAG repeats. Panel A, products of recombination between two (CTG•CAG)$_{17}$ tracts (pRW4815gfp). Panel B, products of recombination between two (CTG•CAG)$_{67}$ tracts (pRW4819gfp). Panel C, products of recombination between two (CTG•CAG)$_{98}$ tracts (pRW4823gfp). Panel D, products of recombination between two TRS tracts containing (CTG•CAG)$_{175}$ and (CTG•CAG)$_{98}$ (pSF3gfp).

**Fig. 8.** Frequency of the CTG•CAG repeat expansions and deletions mediated by intramolecular recombination. The lengths of the TRS-containing fragments (shown in Fig. 7) were measured and the numbers of CTG•CAG...
repeats were calculated as described earlier (29). Black bars = deletions; gray bars = retention of size of progenitor sequence; white bars = expansions.

The orientation of the CTG•CAG tracts relative to the origin of replication in the parental plasmids had no influence on TRS size distribution of the recombination products. Therefore, data obtained for plasmids harboring the same number of triplet repeats but present in different orientations were combined and plotted together in a single bar. Each bar represents the data collected from the analysis of approximately 90 clones.
FIG. 1

Direct repeats

Inverted repeats
### Direct repeats

- **X insert**: (CTG•CAG)_{17}
- **Y insert**: (CTG•CAG)_{17}

### Inverted repeats

- **X insert**: (CTG•CAG)_{17}
- **Y insert**: (CTG•CAG)_{17}

### Control plasmids

- **X insert**: (CTG•CAG)_{165}
- **Y insert**: (CTG•CAG)_{165}

| Plasmid   | X insert          | Y insert          | X TRS insert       | Y TRS insert       |
|-----------|-------------------|-------------------|--------------------|--------------------|
| pRW4815   | (CTG•CAG)_{17}    | (CTG•CAG)_{17}    | (GAC•GTC)_{79}     |                   |
| pRW4817   | (CAG•CTG)_{17}    | (CAG•CTG)_{17}    | (GTC•GAC)_{79}     |                   |
| pRW4819   | (CTG•CAG)_{67}    | (CTG•CAG)_{67}    | (GAC•GTC)_{79}     |                   |
| pRW4820   | (CAG•CTG)_{67}    | (CAG•CTG)_{67}    | (GTC•GAC)_{79}     |                   |
| pRW4821   | (CTG•CAG)_{67}    | (CTG•CAG)_{67}    | (GAC•GTC)_{79}     |                   |
| pRW4822   | (CAG•CTG)_{67}    | (CAG•CTG)_{67}    | (GTC•GAC)_{79}     |                   |
| pRW4823   | (CTG•CAG)_{98}    | (CTG•CAG)_{98}    | (GAC•GTC)_{79}     |                   |
| pRW4824   | (CAG•CTG)_{98}    | (CAG•CTG)_{98}    | (GTC•GAC)_{79}     |                   |
| pRW4825   | (CTG•CAG)_{98}    | (CTG•CAG)_{98}    | (GAC•GTC)_{79}     |                   |
| pRW4826   | (CAG•CTG)_{98}    | (CAG•CTG)_{98}    | (GTC•GAC)_{79}     |                   |
| pSF4      | (CTG•CAG)_{175}   | (CAG•CTG)_{98}    | (GAC•GTC)_{79}     |                   |
| pRW4805   | (CTG•CAG)_{175}   | (CAG•CTG)_{98}    | (GAC•GTC)_{79}     |                   |
| pRW4808   | 564 bp λ phage DNA | 564 bp λ phage DNA |                   |                   |
| pRW4804   | 564 bp λ phage DNA | 564 bp λ phage DNA |                   |                   |
| pRW4807   | 354 bp DMPK DNA   | 354 bp DMPK DNA   |                   |                   |
| pRW4871   | 354 bp DMPK DNA   | 354 bp DMPK DNA   |                   |                   |
| pRW4873   | 354 bp DMPK DNA   | 354 bp DMPK DNA   |                   |                   |

(both inserts are inverted compared to pRW4871)
FIG. 3

A

|            | pRW4815 | pRW4817 | pRW4819 | pRW4821 | pRW4823 | pRW4825 | pRW4804 |
|------------|---------|---------|---------|---------|---------|---------|---------|
| 1 Kb       | C       | C       | C       | C       | C       | C       | C       |
| 1 Kb       | +       | -       | -       | -       | +       | -       | -       |

B

|            | pRW4816 | pRW4818 | pRW4820 | pRW4822 | pRW4824 | pRW4826 |
|------------|---------|---------|---------|---------|---------|---------|
| 1 Kb       | C       | C       | C       | C       | C       | C       |
| 1 Kb       | +       | -       | -       | -       | +       | -       |

C

|            | pRW4830 | pRW4831 | pRW4832 | pRW4833 |
|------------|---------|---------|---------|---------|
| 1 Kb       | C       | C       | C       | C       |
| 1 Kb       | +       | +       | +       | +       |
A

![Diagram of plasmid transformation and frequency of recombination](image)

Frequency of recombination = \( \frac{\text{number of white colonies}}{\text{total number of colonies}} \)

B

| Plasmid     | Inserts                      | bp of homology | Recombination frequency (x10^{-3}) | \( \text{AB1157 (recA^+)} \) | \( \text{JC10289 (recA^-)} \) |
|-------------|------------------------------|----------------|------------------------------------|-----------------------------|-----------------------------|
| pRW4815gfp  | (CTG • CAG)_{17}            | 167            |                                    | 2                          | 1                           |
| pRW4817gfp  | (CAG • CTG)_{17}            | 167            |                                    | 8                          | 4                           |
| pRW4819gfp  | (CTG • CAG)_{67}            | 317            |                                    | 13                         | 6                           |
| pRW4821gfp  | (CAG • CTG)_{67}            | 317            |                                    | 23                         | 12                          |
| pRW4823gfp  | (CTG • CAG)_{98}            | 410            |                                    | 16                         | 8                           |
| pRW4825gfp  | (CAG • CTG)_{98}            | 410            |                                    | 60                         | 31                          |
| pRW4863gfp  | (CTG • CAG)_{165}           | 611            |                                    | 126                        | 61                          |
| pRW4865gfp  | (CAG • CTG)_{165}           | 611            |                                    | 81                         | 41                          |
| pRW4804gfp  | 564 bp \( \lambda \) DNA   | 564            |                                    | 12                         | 6                           |
| pRW4871gfp  | 354 bp DMPK DNA             | 354            |                                    | 9                          | 5                           |
| pRW4873gfp  | 354 bp DMPK DNA             | 354            |                                    | 10                         | 5                           |

FIG. 4
| Plasmid       | X TRS insert          | Y TRS insert          | Orientation of repeat insert | Frequency of white colonies ($\times 10^{3}$) |
|--------------|----------------------|----------------------|-----------------------------|---------------------------------------------|
| pRW4825gfp   | (CAG•CTG)$_{98}$     | (CAG•CTG)$_{98}$     | direct                      | 60                                          |
| pRW4820gfp   | (CTG•CAG)$_{98}$     | (CAG•CTG)$_{98}$     | inverted                    | <0.1                                        |
| pRW4822gfp   | (CAG•CTG)$_{98}$     | (CTG•CAG)$_{98}$     | inverted                    | <0.1                                        |
| pRW4830gfp   | (CTG•CAG)$_{67}$     | (GAC•GTC)$_{67}$     | sequence isomer             | <0.1                                        |
| pRW4831gfp   | (CTG•CAG)$_{67}$     | (GTC•GAC)$_{98}$     | sequence isomer             | <0.1                                        |
| pBR322gfp    | —                    | —                    | N.A.                        | 0.03                                        |
Long CTG\textsuperscript{+}CAG repeat sequences markedly stimulate intramolecular recombination
Marek Napierala, Pawel Parniewski, Anna Pluciennik and Robert D. Wells

J. Biol. Chem. published online June 3, 2002

Access the most updated version of this article at doi: 10.1074/jbc.M202128200

Alerts:
  • When this article is cited
  • When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts