Double-strand breaks in the myotonic dystrophy type 1 and the fragile X syndrome triplet repeat sequences induce different types of mutations in DNA flanking sequences in *Escherichia coli*

Beata Kosmider and Robert D. Wells*

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

Received May 2, 2006; Revised July 28, 2006; Accepted August 7, 2006

ABSTRACT

The putative role of double-strand breaks (DSBs) created *in vitro* by restriction enzyme cleavage in or near CGG•CCG or CGT•CAG repeat tracts on their genetic instabilities, both within the repeats and in their flanking sequences, was investigated in an *Escherichia coli* plasmid system. DSBs at TRS junctions with the vector generated a large number of mutagenic events in flanking sequences whereas DSBs within the repeats elicited no similar products. A substantial enhancement in the number of mutants was caused by transcription of the repeats and by the absence of recombination functions (*recA*, *recBC*). Surprisingly, DNA sequence analyses on mutant clones revealed the presence of only single deletions of 0.4–1.6 kb including the TRS and the flanking sequence from plasmids originally containing (CGG•CCG)43 but single, double and multiple deletions as well as insertions were found for plasmids originally containing (CGT•CAG)n, where n = 43 or 70. Non-B DNA structures (slipped structures with loops, cruciforms, triplexes and tetraplexes) as well as microhomologies are postulated to participate in the recombination and/or repair processes.

INTRODUCTION

The genetic instabilities of TRS are involved in the etiology of approximately 20 hereditary neurological diseases including myotonic dystrophy type 1 (DM1), fragile X syndrome (FRAX) and Huntington’s disease [reviewed in Ref. (1–3)]. Whereas large expansions of TRSs are most frequently observed in patient materials related to FRAX and DM1, deletion events of CGG•CCG and their flanking sequences were also found in clinical cases (1–3) as well as intergenerational contractions of the CTG•CAG repeats in DM1 patients (4–10). The molecular mechanisms responsible for these instabilities have been elucidated in simple genetically tractable systems, such as bacteria or yeast [reviewed in Ref. (2,11)] and later studies in mammalian systems have been important in revealing cell-specific factors. Using these strategies, DNA replication, recombination (especially gene conversion), and DNA repair (especially methyl-directed mismatch repair and nucleotide excision repair) based mechanisms have been implicated. Furthermore, the capacity of the simple repeating sequences to adopt non-B DNA conformations (including slipped structures, cruciforms, DNA unwinding elements, tetraplexes, triplexes and sticky DNA [reviewed in Ref. (1,2,12)] have been implicated in virtually all of the proposed molecular mechanisms. These non-B DNA structures, no doubt, serve to impede the progression of DNA replication and/or transcription complexes (13–17) and, thereby promote the genetic instabilities of the repeat sequences and, thus, to lead to pathogenesis (18).

Prior investigations on the molecular mechanisms of the genetic instabilities in model systems (bacteria, yeast and cell culture) have revealed the expansions and/or deletions within the TRS (11,18–23) with no alterations in the DNA sequences, which flank the unstable repeats. However, recent studies (13,14) show that long tracts of CTG•CAG, CCTG•CAGG, GAA•TTC, and an intronic poly(R•Y) sequence from the PKD1 gene do cause the formation of gross deletions, inversions and other rearrangements involving the sequences which flank the repeats. DSBs in the DNA are intimately involved in the repair and/or recombination processes, which effect these gross deletions and genomic rearrangements. The DSBs may originate from *in vitro* cleavage by restriction enzymes (24–29), radiation of biological materials (30) and drugs *in vivo* (31). DSBs are the primary lesions responsible for mutagenesis, chromosomal rearrangements and cell death (32). These DSBs may...
be repaired by a mechanism in *Escherichia coli* that generates products identical to those identified in mammalian systems (33, 34) in the absence of DNA homology. All organisms have developed pathways for repairing DSBs. In eubacteria, DSBs are repaired mainly by the RecBCD family of enzymes (35). Presumably, RecBC acts within a complex of other enzymes, including RecA and other proteins (36–38). The rejoining of DNA ends in *E. coli* by transforming cultures with linearized plasmids has been studied previously (27–29) and the most common product results from the ligation of the DNA ends to restore the original plasmid. Furthermore, deletions were also observed (27, 29).

Herein, we have investigated the role of DSBs within or near CGG or CTG repeat tracts in an *E. coli* plasmid system. Instabilities within the TRS as well as the flanking sequences were monitored. Surprisingly, we found different types of mutagenic patterns for these two repeat sequences. For the former sequence, only single deletions were observed but DSB repair at the latter sequence induces single, double and multiple deletions as well as inversions.

**MATERIALS AND METHODS**

### *E. coli* strains

All cloning was done in *E. coli* HB101 (F−, Δ(gpt-proA)62, leuB6, glnV44, ara-14, ga/K2, lacY1, Δ(mcrC-mrr)), rpsL20, (StrR), xyl-5, mtl-1, recA13) (14). Experiments were carried out in the following four *E. coli* strains: parental KMBL1001 (no known mutations) (39); parental AB1157 (thr-1, ara-14, leuB6, Δ(gpt-proA)62, lacY1, tss-33, gsr, glnV44(AS), ga/K2, l, recA13), mtl-1, rpsL31, kgdK51, xylA5, mtl-1, argE3(Oc), thi-1) (13); mutant recBC− (JC5519 = AB1157 + recB21, recC22) (25); mutant recA− (JC10287 = AB1157 + Δ(recA-srl)306) (25).

### Plasmids

The interrupted (CGG or CTG)n strands are the templates for lagging strand synthesis, the TRS containing fragments from the former sequence had 7 bp of non-repetitive flanking sequences at the Acc65I site and 15 bp at the HincII site. The pGFPT vector digested with Acc65I and Stul because Acc65I and BsiWI have compatible sticky ends and HincII and Stul have blunt ends. On the other hand, for the plasmid preparations with the inserts in orientation II, where the (CGG)n or (CTG)n strands are the templates for leading strand synthesis, the TRS containing fragments from the former sequence had 7 bp of non-repetitive flanking sequences at the Accl65 site and 15 bp at the HincII site. The pGFPT vector was digested with BsiWI and Stul because Acc65I and BsiWI have compatible sticky ends and HincII and Stul have blunt ends. On the other hand, for the plasmid preparations with the inserts in orientation II, where the (CGG)n or (CTG)n strands are the templates for lagging strand synthesis, the TRS containing fragments from the parental digested plasmids pRW4531, pRW4525 and pRW4523 were released by Acc65I and HincII digestions. These inserts had, in addition to the 43 CGG or CTG repeats in pRW5503, the 43 CTG or CGG repeats in pRW5509, 13 bp of non-repetitive flanking sequences at the Acc65I site and 15 bp at the HincII site. The pGFPT vector was digested with BsiWI and Stul because Acc65I and BsiWI have compatible sticky ends and HincII and Stul have blunt ends. On the other hand, for the plasmid preparations with the inserts in orientation II, where the (CGG)n or (CTG)n strands are the templates for lagging strand synthesis, the TRS containing fragments from the parental digested plasmids pRW4531, pRW4525 and pRW4523 were released by XbaI and SmaI digestions. In addition to the 43 CGG or CTG repeats in pRW5504, the 43 CTG or CGG repeats in pRW5512 and the 70 CGG or CTG repeats in pRW5511 and the 70 CTG or CGG repeats in pRW5509, 13 bp of non-repetitive flanking sequences at the Acc65I site and 15 bp at the HincII site. The pGFPT vector was digested with BsiWI and Stul because Acc65I and BsiWI have compatible sticky ends and HincII and Stul have blunt ends. On the other hand, for the plasmid preparations with the inserts in orientation II, where the (CGG)n or (CTG)n strands are the templates for lagging strand synthesis, the TRS containing fragments from the parental digested plasmids pRW4531, pRW4525 and pRW4523 were released by XbaI and SmaI digestions. In addition to the 43 CGG or CTG repeats in pRW5504, the 43 CTG or CGG repeats in pRW5512 and the 70 CTG or CGG repeats in pRW5511, these inserts had 7 bp of non-repetitive flanking sequences at the XbaI site and 12 bp at the SmaI site. These fragments were individually ligated to the pGFPT vector digested with SpeI and Stul since XbaI and SpeI have compatible sticky ends and SmaI and Stul have blunt ends.

The parental digested plasmids pRW4531, pRW4525 and pRW4523 and the linearized pGFPT vector were purified by electrophoresis on a 6% polyacrylamide gel.
gel in TAE buffer [40 mM Tris–acetate and 1 mM EDTA (pH 8.0)]. The gel was stained with ethidium bromide and the corresponding bands containing the TRS and the linearized vector were excised. The DNA was then eluted from the excised bands, purified by phenol/chloroform extractions and precipitated with ethanol (43). Each of the purified inserts and the linearized vector were mixed at approximately a 10:1 molar ratio and ligated for 16 h at 16°C with 20 U of T4 DNA ligase in the presence of 1 mM ATP. Subsequently, the procedure was carried out as described earlier for the preparation of pGFPTΔE. The plasmids from single CFUs were isolated using Wizard Plus Minipreps DNA Purification System (as before). For individual clones, the lengths of the TRS were determined by digestion with MfeI and ApaI. The inserts were released and these fragments were then radiolabeled with [α-32P]dATP (43) and 1 U of the Klenow fragment of E.coli polymerase I was added; the fragments were then resolved on a 6% native polyacrylamide gel. The gel was dried, exposed to X-ray film and then analyzed using a PhosphorImager (Storm 820 with Image Quant software, Molecular Dynamics).

The confirmation of the orientation of the CGG·CCG repeats in the sequencing reaction was very difficult because of the repetitive CG rich arrays, which cause extensive strand slippage and hairpin formation (40,42,44). Therefore, to improve the efficiency of this reaction, the plasmids harboring the CGG·CCG repeats were first run on a 1% agarose gel and the bands were excised using a DNA Gel Extraction Kit (Millipore Corp., Bedford, MA). Due to the high resolving power of gel electrophoresis, any other interference in the sequencing reaction was minimized. Subsequently, the DNA was ethanol-precipitated and selected clones containing the full-length restriction fragments harboring the undeleted 43 CGG·CCG and 43 or 70 CTG·CAG repeats were also sequenced on both strands. The forward and reverse primers GGATCCGTTCAACTAGCAG and CAAGCTGTGACCGTCTCCG, respectively, were used. The locations of the forward primer were the same for pRW5503, pRW5504, pRW5511, pRW5512, pRW5509 and pRW5510 at map position 808. However, the positions of the reverse primers for both pRW5503 and pRW5504 were at 1320, for pRW5504 and pRW5512 at 1325, for pRW5509 at 1407 and pRW5510 at 1409. The concentrations of the primers and DNA employed and the cycle sequencing conditions were as stated above and sequenced in the same institute.

Figure 1. Diagram of plasmids and scheme of study. All plasmids contain interrupted CGG·CCG or CTG·CAG repeats with EcoRV and EcoRI recognition sites (in bold) neighboring and inside the TRS, respectively. The inserts (25) were cloned into pGFPTΔE harboring a transcription terminator cassette (13). The control pGFPTΔE contains no TRS. In orientation I, the (CGG)n or (CTG)n strand is the template for leading strand synthesis; in orientation II, the (CGG)n or (CTG)n strand is the template for lagging strand synthesis (25,40). Vertically cross-hatched arrow, ampicillin resistance gene (AmpR); horizontally cross-hatched arrow, unidirectional pUC19 origin of replication (Ori); dotted box, transcription terminator cassette (Ter); open arrow, lacZ promoter-operator (Pr); solid gray bar, lacZ–GFP fusion gene; solid black bar, triplet repeats. The CGG·CCG or CTG·CAG tracts were cloned into pGFPTΔE vector. DSBs were introduced by digestion of these plasmids with restriction enzymes. Linearized plasmids were immediately transformed into E.coli and the transformation mixtures were spread on LB agar plates. Individual white CFUs were analyzed by biochemical analyses and the sequences of the mutant clones were determined to evaluate if the breakpoints were flanked by non-B DNA conformations.
Preparation of linear DNAs

Supercoiled DNA from the control pGFPTΔE and pRW5503, pRW5504, pRW5511, pRW5512, pRW5509 and pRW5510 were prepared by the Maxiprep procedure for 1 l cultures and the DNA was purified by cesium chloride/ethidium bromide centrifugation overnight (43). Linear DNA was obtained by digestion of these plasmids with EcoRV or EcoRI at the unique sites neighboring and inside the TRS, respectively (Figure 1). The control pGFPTΔE vector was digested with StuI and ApaI to induce DSBs, which correspond to a similar distance between these recognition sites and the GFP gene as in plasmids containing DSBs and digested with EcoRV and EcoRI, respectively. We found no white CFUs after E.coli transformation with linear control pGFPTΔE as well as with circular vector (data not shown). Linearized vector and the plasmids harboring the CCG•CCG or CTG•CAG repeats were first run on a long 1% agarose gel using the DNA Gel Extraction Kit (as described before). Subsequently, the bands of linear DNAs were excised from the agarose gel and eluted in the Spectra/Por® regenerated cellulose dialysis membranes (Spectrum Laboratories Inc., Rancho Dominguez, CA) using the modified TAE electrophoresis buffer [40 mM Tris-acetate and 0.1 mM EDTA (pH 8.0)] from the DNA Gel Extraction Kit (as before). Immediately prior to collecting the buffer containing linear DNA, the electrode polarity was reversed for 1 min to loosen any DNA bound to the membrane (28). Subsequently, the DNA was purified by phenol/chloroform extractions and then precipitated with ethanol.

The conditions of complete digestion of all plasmids were determined by long 1% agarose gel where only a single band of linear DNA was observed (24–26). As a second control to determine the purity of the plasmids, each sample was rerun on both agarose gel and polyacrylamide gels. In the latter case, DNAs were radiolabeled with [α-32P]dATP, exposed to X-ray film and analyzed using a PhosphorImager (as before). We found no contamination with circular DNA within the assay detection limit (data not shown) (25). Our previous studies (25) were more sensitive to any possible contamination by residual circular DNA than this investigation due to the strategies involved; thus we have confidence in the purity of our linearized plasmids. All linear plasmids were immediately used for E.coli transformations (29).

Transformation of supercoiled and linear plasmids containing CCG•CCG or CTG•CAG tracts into E.coli

Supercoiled pGFPTΔE, pRW5503, pRW5504, pRW5511, pRW5512, pRW5509 and pRW5510 were transformed by electroporation as described above into each of the E.coli strains with different genetic backgrounds (JC5519, JC10287, KMBl1001 and AB1157) to evaluate the role of the host strains in the mutagenic process. To determine the fluorescent status of the colonies, the transformation mixtures were spread on LB agar plates containing 100 μg/ml Amp and isopropyl β-D thiogalactoside (IPTG) at a final concentration of 2 mM for studies in the presence of lacZ transcription. Parallel experiments were also carried out in the absence of lacZ transcription but including 100 μg/ml kanamycin (Kan) where the E.coli strains were first transformed with the plasmid pIQ-kan as described previously (13,14,21) (B. Kosmider and R.D. Wells, manuscript in preparation). By using this cloning strategy, the TRS are located downstream of the lacZ promoter and therefore transcription through this repeat tract was inhibited. However, we found no white CFUs in the presence and in the absence of lacZ transcription using supercoiled plasmids containing (CCG•CCG)43, (CTG•CAG)43 and (CTG•CAG)70 (data not shown) and hence this result serves as an additional control.

Subsequently, linearized pGFPTΔE, pRW5503, pRW5504, pRW5511, pRW5512, pRW5509 and pRW5510 (Figure 1) digested with EcoRV or EcoRI were transformed into E.coli as described above in the presence and in the absence of lacZ transcription. The viable green (intact GFP gene and flanking sequences) and white (mutated GFP gene and flanking sequences) CFUs were screened and counted from the transformation mixtures; however, the total number of CFUs observed was dramatically reduced (by ~50- to 100-fold), after E.coli transformation with the linear plasmids in comparison with their supercoiled forms (data not shown), as reported (28,29,45). In all experiments, the CFUs from the linearized control pGFPTΔE were fluorescent green, indicating a functional GFP reporter gene (Table 1). Therefore, all white CFUs found in this assay must have been induced by the DSB repair events in plasmids containing the TRS. These CFUs were re-streaked three times on LB agar plates with Amp and IPTG (as described earlier) for experiments in the presence of lacZ transcription and with Amp and Kan (as before) in the absence of lacZ transcription to confirm a persistent white phenotype (13,14) (B. Kosmider and R.D. Wells, manuscript in preparation). Subsequently, the plasmids from these white CFUs were isolated by alkaline lysis using a Wizard Plus MiniPreps DNA Purification System for the sequencing reactions to determine the DNA breakpoints (Supplementary Table 1). Three or more independent experiments were performed separately for each supercoiled plasmid or for the plasmids digested with EcoRV or EcoRI, both in the presence and in the absence of lacZ transcription.

DNA sequencing

To detect DNA junctions in mutant clones, plasmids isolated from white CFUs were sequenced using primers for the regions downstream and upstream of the GFP gene as well as the CCG•CCG or CTG•CAG repeats. However, the determinations of the sequences of the mutant clones in plasmids containing the CCG•CCG repeats was very difficult (40,42,44); therefore, we first used the DNA Gel Extraction Kit to improve the sequencing reactions. The forward and reverse primers of the white mutants of pRW5503, pRW5511, pRW5504, pRW5512, pRW5509 and pRW5510 were GCCAGGGGGAAAAGCGCTG and CAGGTTATTGTTCAG, respectively, with their positions at 3203 and 1658 for the first two plasmids; for the next two, the map positions were at 3208 and 1663 and for pRW5509 were at 3240 and 1744 and finally for pRW5510, were at 3242 and 1746, respectively. In some cases, additional forward and reverse primers, CACTGCTATCCCTGAGTTCTG...
Table 1. The mutagenic effect of the DSBs induced by EcoRV digestion at CGG•CCG or CTG•CAG tracts on the number of white CFUs.

<table>
<thead>
<tr>
<th>E.coli strain</th>
<th>pGFPTΔE</th>
<th>(CCG•CCG)43</th>
<th>(CTG•CAG)43</th>
<th>(CTG•CAG)70</th>
</tr>
</thead>
<tbody>
<tr>
<td>JC519</td>
<td>0</td>
<td>0.3820*</td>
<td>0.1794*</td>
<td>0.0188</td>
</tr>
<tr>
<td>(recBC−)</td>
<td>0/382</td>
<td>0/440</td>
<td>34/89</td>
<td>4/119</td>
</tr>
<tr>
<td>JCI10287</td>
<td>0</td>
<td>0.3281*</td>
<td>0.1238*</td>
<td>0.0016</td>
</tr>
<tr>
<td>(recA−)</td>
<td>0/395</td>
<td>0/489</td>
<td>21/64</td>
<td>3/110</td>
</tr>
<tr>
<td>KMBL1001</td>
<td>0/571</td>
<td>0/608</td>
<td>156/478</td>
<td>2/445</td>
</tr>
<tr>
<td>AB1157</td>
<td>0/633</td>
<td>0/725</td>
<td>101/491</td>
<td>2/471</td>
</tr>
</tbody>
</table>

The pGFPTΔE control vector containing no repeats or plasmids harboring (CCG•CCG)43, (CTG•CAG)43 or (CTG•CAG)70 were transformed into four E.coli strains with different genetic backgrounds (see ‘Materials and Methods’). The occurrence of white CFUs (bold font) was calculated as the ratio of the number of white colonies to the total number of viable cells (white and green) (ratios underneath the bold decimal fractions). The composite results for orientations I and II for each TRS length in at least three independent experiments are shown. The statistically significant differences between the white CFUs for each TRS in comparison with the control pGFPTΔE in all strains and the presence of lacZ transcription were calculated using the z-test. The data which show significant differences are marked with asterisks.

Results

Strategy of investigation

Numerous studies (24–31) describe the repair of DSBs introduced into plasmids but only three reports (24–26) considered the influence of DSBs introduced by restriction enzymes in plasmids harboring TRS. Moreover, the influence of these DSBs on the instability within the TRSs was determined but their effects on flanking DNA sequences were not analyzed in E.coli. Hence, we have evaluated the effect of DSBs which were introduced by restriction enzymes at or inside CGG•CCG or CTG•CAG repeating tracts on non-repetitive flanking DNA sequences in plasmids in E.coli. The experiments utilized the vector pGFPTΔE, which harbored the reporter GFP gene. The inserts were created to contain a unique EcoRI recognition site inside the CGG•CCG or CTG•CAG repeats in order to induce DSBs (25). These sequences also have a unique EcoRV recognition site at the junction of the vector with the TRS; therefore, this EcoRV site was utilized to linearize the plasmids to generate molecules with the repeat tracts at the ends after DSB inductions (Figure 1), recA−, recBC− and parental strains were used (‘Materials and Methods’) to explore the genetic background involved in the repair process. Also, we conducted studies in the absence and in the presence of lacZ transcription (Table 1), which enables an analysis of DSB repair events and the effects of both DSB repair and transcription, respectively. Transcription was induced by 2 mM IPTG and was inhibited by co-transformation with the pI1-kan plasmid, which expressed a variant lacI repressor (see ‘Materials and Methods’); white and green CFUs, respectively, were counted.

Interestingly, no white CFUs, indicative of a functional lacI repressor, were found in any experiments after transformation with plasmids containing the TRS cleaved with EcoRI (data not shown). A z-test was used to analyze all statistically significant differences including between the fractions of white CFUs detected after DSB induction at the TRS by EcoRV and the control, which was
Table 2. The effect of orientation of the CGG•CCG or CTG•CAG tracts on the number of white CFUs in the presence of lacZ transcription after formation of DSBs by EcoRV

<table>
<thead>
<tr>
<th>E.coli strain</th>
<th>(CGG•CCG)_{43} Orientation I</th>
<th>Orientation II</th>
<th>(CTG•CAG)_{43} Orientation I</th>
<th>Orientation II</th>
<th>(CTG•CAG)_{70} Orientation I</th>
<th>Orientation II</th>
</tr>
</thead>
<tbody>
<tr>
<td>JC5519</td>
<td>0.3333</td>
<td>0.5882</td>
<td>0.0156</td>
<td>0.0545</td>
<td>0.1764</td>
<td>0.1860</td>
</tr>
<tr>
<td>(recBC−)</td>
<td>24/72</td>
<td>10/17</td>
<td>1/64</td>
<td>3/55</td>
<td>9/51</td>
<td>8/43</td>
</tr>
<tr>
<td>JC10287</td>
<td>0.3170</td>
<td>0.3478</td>
<td>0.0147</td>
<td>0.0476</td>
<td>0.1020</td>
<td>0.125</td>
</tr>
<tr>
<td>(recA−)</td>
<td>13/41</td>
<td>8/23</td>
<td>1/68</td>
<td>2/42</td>
<td>5/49</td>
<td>4/32</td>
</tr>
<tr>
<td>KMBL1001</td>
<td>0.0957*</td>
<td>0.6114*</td>
<td>0.0041</td>
<td>0.0049</td>
<td>0.0051*</td>
<td>0.0227*</td>
</tr>
<tr>
<td>AB1157</td>
<td>0.0890*</td>
<td>0.3768*</td>
<td>0.0034</td>
<td>0.0054</td>
<td>0.0740</td>
<td>0.1393</td>
</tr>
</tbody>
</table>

The orientations of the inserts were explained in the legend to Figure 1. The conditions of the assays and the method of calculation of the occurrence of white CFUs (bold font) were presented in the legend to Table 1. The z-test was used to show the statistically significant differences between orientations I and II for each TRS and are designated with asterisks.

Table 3. The orientation effect of the CGG•CCG or CTG•CAG tracts on the number of white CFUs in the absence of lacZ transcription after formation of DSBs by EcoRV

<table>
<thead>
<tr>
<th>E.coli strain</th>
<th>(CGG•CCG)_{43} Orientation I</th>
<th>Orientation II</th>
<th>(CTG•CAG)_{43} Orientation I</th>
<th>Orientation II</th>
<th>(CTG•CAG)_{70} Orientation I</th>
<th>Orientation II</th>
</tr>
</thead>
<tbody>
<tr>
<td>JC5519</td>
<td>0.1333</td>
<td>0.2083</td>
<td>0</td>
<td>0.0833</td>
<td>0.0166</td>
<td>0.0232</td>
</tr>
<tr>
<td>(recBC−)</td>
<td>4/30</td>
<td>10/48</td>
<td>0/41</td>
<td>1/12</td>
<td>1/60</td>
<td>1/43</td>
</tr>
<tr>
<td>JC10287</td>
<td>0.1111</td>
<td>0.1515</td>
<td>0</td>
<td>0.0338</td>
<td>0</td>
<td>0.0196</td>
</tr>
<tr>
<td>(recA−)</td>
<td>8/72</td>
<td>5/33</td>
<td>0/62</td>
<td>2/59</td>
<td>0/25</td>
<td>1/51</td>
</tr>
<tr>
<td>KMBL1001</td>
<td>0.0529</td>
<td>0.0841</td>
<td>0</td>
<td>0.0033</td>
<td>0.0047</td>
<td>0.0053</td>
</tr>
<tr>
<td>AB1157</td>
<td>0.0183*</td>
<td>0.0714*</td>
<td>0</td>
<td>0.0035</td>
<td>0.0072</td>
<td>0.0106</td>
</tr>
</tbody>
</table>

The orientations of the inserts were explained in the legend to Figure 1. The conditions of the assays and the method of calculation of the occurrence of white CFUs (bold font) were presented in the legend to Table 1. The asterisks represent the statistically significant differences (z-test) between orientations I and II.

linearized pGFPTΔE (Table 1). Additionally, we also determined the number of white CFUs separately for orientations I and II of the TRS (Tables 2 and 3). In orientation I, the (CGG)_{n} or (CTG)_{n} strand is the template for leading strand synthesis and in orientation II, these strands are the templates for lagging strand synthesis. Plasmids from the white CFUs induced by DSB induction by EcoRV at the TRS were isolated for DNA sequence determinations to characterize the breakpoints (Supplementary Table 1). Surprisingly, we observed different mutagenic patterns for the DSB repair products in plasmids containing the CGG•CCG or the CTG•CAG sequences (Figure 3; Supplementary Tables 1 and 2).

DSBs introduced inside the TRS by EcoRI do not induce white CFUs

We previously reported (25) that DNA cleavage by EcoRI inside the repeat tract caused instabilities within the TRS; however, its influence on flanking sequences was not studied. Therefore, we determined if this digestion would enhance the induction of mutations in neighboring genes. Experiments were carried out in the presence and in the absence of lacZ transcription where the transformation mixture was spread on agar plates containing IPTG and Amp or Kan and Amp, respectively, (see ‘Materials and Methods’). In control studies with linearized pGFPTΔE, no white CFUs were found. Interestingly, we also detected no white CFUs after transformation with plasmids containing (CGG•CCG)_{43}, (CTG•CAG)_{43} or (CTG•CAG)_{70} repeats (data not shown) digested with EcoRI within the tracts (Figure 1). These results are in agreement with the observations that DSBs between two direct repeats increase the rate of recombination between the repeats with the concomitant deletion of the intervening sequences (51,52). In previous studies (25,26,53,54), single-strand annealing was proposed to explain the TRS shortening induced by DSBs introduced inside the tracts; however, no prior studies were conducted on mutations in the flanking DNA sequences in E.coli. Thus, these results are in accord with prior studies (25).

DSBs introduced by EcoRV in CGG•CCG repeats induce more white CFUs than in CTG•CAG

Different types of TRSs and their lengths had a pronounced effect on repeat instabilities after DSB induction; (11,22–26,40,51) however, mutations in the flanking DNA sequences were not analyzed previously in bacterial cells. Therefore, experiments were carried out herein to determine if DSBs induced by EcoRV at (CGG•CCG)_{43}, (CTG•CAG)_{43} or (CTG•CAG)_{70} tracts would have an influence on the promotion of mutations in flanking DNA sequences in the presence and in the absence of lacZ transcription (see ‘Materials and Methods’). The linearized plasmids pRW5503, pRW5504, pRW5511, pRW5512, pRW5509 or pRW5510 (Figure 1) were transformed into recA−, recBC− and the parental strains and the transformation mixtures were spread on agar plates. The total number of viable CFUs (white and green) were low, as reported previously after E.coli transformation with linearized DNA cleaved by restriction enzymes...
No white CFUs were found after transformation of all strains with the control-linearized pGFPTAE in the presence and in the absence of lacZ transcription (Table 1). Hence, all white CFUs which were detected represent single mutagenic events from the repair of DSBs introduced into the plasmids at the TRSs.

To summarize the data in Table 1, we found, first, a higher number of white CFUs was observed after DSBs were induced at the (CGG•CCG)_{43} tract than for the (CTG•CAG)_{43} or (CTG•CAG)_{70}. Second, these values were higher for the CTG•CAG sequence containing 70 repeats than for the shorter tract harboring 43 repeats. Third, the highest fraction of white CFUs was detected in the recA− and recBC− strains in comparison with the parental strains. Fourth, a higher number of white CFUs were found in the presence than in the absence of lacZ transcription. Linear DNA has to convert into a circular supercoiled form in order for transcription and replication to start. Transcription and transcription-replication collisions should be downstream from the circularization step. Thus, our results indicate that the type and length of the TRS has an influence on the DSB repair involving flanking DNA sequences.

The number of white CFUs detected after DSB induction by EcoRV at the TRS may be underestimated because the mutant clones with deletions in the other direction on the plasmid involving ampicillin resistant and/or the origin of replication genes will not survive (13,14,55). Note that in comparison with the parental strains. Fourth, a higher number of white CFUs were found in the presence than in the absence of lacZ transcription. Linear DNA has to convert into a circular supercoiled form in order for transcription and replication to start. Transcription and transcription-replication collisions should be downstream from the circularization step. Thus, our results indicate that the type and length of the TRS has an influence on the DSB repair involving flanking DNA sequences.

The orientation dependent effect of the CGG•CCG and CTG•CAG repeat tracts in supercoiled plasmids transformed into E.coli on the induction of instability within the TRSs was previously reported (14,40,42,56–58). Herein, we extend these analyses by evaluating the effects of the presence of the TRS at the DSBs and by measuring the mutagenic consequences in the sequences flanking the TRSs. In the presence of lacZ transcription, DSBs introduced at the TRSs caused the lowest numbers of viable CFUs in recA− and recBC− strains which was also demonstrated in studies of the transformation efficiency of the linearized pBR322 plasmid (28). Simultaneously, the highest numbers of white CFUs were detected in these strains (Table 1). This is in agreement with our previous work where the lowest instability within the TRS after DSB induction was also found in recA− and recBC− strains; however, their effect on flanking DNA sequences was not analyzed in this report (25).

**DSBs in TRSs induce mutations in an orientation dependent manner**

The orientation dependent effect of the CGG•CCG and CTG•CAG repeat tracts in supercoiled plasmids transformed into E.coli on the induction of instability within the TRSs was previously reported (14,40,42,56–58). Herein, we extend these analyses by evaluating the effects of the presence of the TRS at the DSBs and by measuring the mutagenic consequences in the sequences flanking the TRSs. In the presence of lacZ transcription, DSBs introduced at TRSs promoted a higher number of white CFUs in orientation II than in orientation I (Table 2). In the absence of lacZ transcription (Table 3), higher numbers of white CFUs were in general also found in orientation II than in orientation I. However, for the (CGG•CCG)_{43} tract, differences in the numbers were detected only for AB1157. Finally, after introduction of DSBs into the (CTG•CAG)_{70} and (CTG•CAG)_{43} tracts, no statistically significant values were found for orientations II and I in all bacterial strains.

Thus, we conclude that the DSBs induced at the TRSs have an orientation dependent effect on the mutagenic events in neighboring genes in the presence and in the absence of lacZ transcription. This behavior may be caused by a proximity effect of the presence of the TRSs joined to the GFP in orientation II (Figure 1) whereas, in orientation I, the TRS is distal to the reporter gene in the linear DNA. It is also possible that the number of white CFUs in the plasmids with the insert in orientation I was underestimated because the TRS was close to the ampicillin resistance gene which could cause plasmid loss (13,55).

**The types of TRSs and their orientations influence the nature of mutagenic products**

The types of mutagenic events in the flanking sequences induced by linearized DNA containing different types of TRSs were analyzed. Restriction mapping was carried out for 33 deleted clones; each clone represents a single mutagenic event (see ‘Materials and Methods’). These products were derived from the six plasmids shown in Figure 1. The sizes of the deletion events for plasmids digested with EcoRV at the CGG•CCG sequence were in the range of 0.4–1.6 kb (Figure 2A and B). For clones harboring the CTG•CAG tracts, deletions were from 1.0 to 1.7 kb (Figure 2D–F).

Interestingly, the DSBs in the two TRS sequences induced different mutagenic patterns in the repair products (Figure 2). First, when DSBs were introduced into the CGG•CCG sequence, both small (<1.0 kb) and large (≥1.0 kb) deletions were detected but only large deletions were found at the CTG•CAG repeats. Second, only single deletions were detected (Figure 2A and B) for all mutant clones derived from plasmids harboring the CGG•CCG sequence. On the other hand, the DSBs at the CTG•CAG tracts induced a broad spectrum of mutagenic events involving single deletions in 45% of the cases, double deletions in 10% of cases and multiple deletions in 10% of cases. Third, only DSBs introduced into the CTG•CAG sequences caused DNA inversions which were found in 35% of the cases. These different types of mutations found for CGG•CCG and for CTG•CAG tracts suggest the involvement of different sequence and/or structure specific proteins in the repair pathways of DSBs, especially since we did not find any white CFUs after E.coli transformation with the linearized pGFPTAE control (Table 1). The influence of the repeat sequence on the types of products formed is surprising since scores of prior studies [reviewed in (1–3)] have revealed different extents of behaviors (e.g. deletions-expansions or repair-recombination reactions) but not different types of processes (to elicit different types of products).

In summary, we found an influence of the type, orientation and the length of the repeat tracts on the induction of the mutagenic events (Figure 3). For (CGG•CCG)_{43}, only single deletions were identified in both orientations. For (CTG•CAG)_{43} and (CTG•CAG)_{70}, single, double and multiple deletions as well as inversions were detected (summarized in Supplementary Table 2). Moreover, for the former tract, single and double deletions were identified in both orientations and for the latter sequence single deletions were found in orientation I. In the other cases of the CTG•CAG sequence containing 43 and 70 repeats, mutagenic events were found only for orientation II. Additionally,
<table>
<thead>
<tr>
<th>E. coli strain</th>
<th>Mutant clone</th>
<th>Size of Deletions (kbp)</th>
<th>(CGG•CCG)$_{43}$</th>
<th>Amp$^R$</th>
<th>Ori</th>
<th>Ter</th>
<th>GFP</th>
<th>Number of repeats remaining</th>
</tr>
</thead>
<tbody>
<tr>
<td>pRW5503 (I)</td>
<td>JC5519</td>
<td>1.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>JC10287</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>14 + 22</td>
</tr>
<tr>
<td></td>
<td>JC10287</td>
<td>1.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>18 + 22</td>
</tr>
<tr>
<td></td>
<td>KMBL1001</td>
<td>1.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>KMBL1001</td>
<td>1.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>16 + 22</td>
</tr>
<tr>
<td></td>
<td>KMBL1001</td>
<td>1.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>AB1157</td>
<td>0.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>AB1157</td>
<td>1.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>13</td>
</tr>
<tr>
<td>pRW5504 (II)</td>
<td>JC5519</td>
<td>0.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>JC10287</td>
<td>1.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>KMBL1001</td>
<td>1.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>KMBL1001</td>
<td>1.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>AB1157</td>
<td>1.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>pRW5511 (I)</td>
<td>JC5519</td>
<td>1.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>JC10287</td>
<td>1.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>KMBL1001</td>
<td>1.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>AB1157</td>
<td>1.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>pRW5512 (II)</td>
<td>JC5519</td>
<td>1.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>JC5519</td>
<td>1.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>JC5519</td>
<td>1.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>JC10287</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>37</td>
</tr>
<tr>
<td></td>
<td>JC10287</td>
<td>1.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>KMBL1001</td>
<td>1.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>AB1157</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>39</td>
</tr>
</tbody>
</table>
a length dependence was also found because double deletions were observed between 0 and 20% of the cases for (CTG\textsuperscript{C15}CAG)\textsubscript{43} and (CTG\textsuperscript{C15}CAG)\textsubscript{70} in orientation II, respectively. Similar correlations were observed for inversions where 57.1 and 60% of the cases were detected in orientation II for CTG\textsuperscript{C15}CAG harboring 43 and 70 repeats, respectively. The orientation effect may be explained by the presence of CTG\textsuperscript{C15}CAG tracts attached to the GFP gene in orientation II (Figure 1). Hence, the TRSs had a stronger mutagenic effect where the TRSs were in close proximity to the GFP reporter gene (Discussion). Three types of DSB repair products were found (Discussion); first, plasmids where the entire TRS was deleted, second, plasmids with shortened TRSs where the EcoRI recognition site inside the tract was deleted and third, plasmids where the TRS was shortened but the EcoRI recognition site was preserved (Figure 4).

**DNA sequences and putative non-B DNA structures at breakpoints of mutant clones**

Previous studies revealed that mutagenic events coincided with non-B DNA structures (12–14). Therefore, we analyzed the mutagenic spectra after \textit{E.coli} transformation with linear DNAs containing the TRSs. The data are presented in the Supplementary Data. We found that the breakpoints coincided with sequences that may form cruciforms, triplexes, tetraplexes and slipped structures. Also, microhomologies between the breaks were found. A general model to describe these deletion events is presented (Figure 4).

**DISCUSSION**

The role of DSBs proximal to CTG\textsuperscript{C15}CAG or CGG\textsuperscript{C15}CCG repeats in plasmids on the genetic instabilities within the TRS or in flanking sequences was investigated. The DSBs were created \textit{in vitro} by restriction endonuclease cleavage near the center of the TRSs or at the junctions of the TRSs and the vector. Unexpectedly, we discovered that different types of mutagenic events occurred at the TRSs at their junctions with the vector for the CGG\textsuperscript{C15}CCG and CTG\textsuperscript{C15}CAG repeats. Only single deletions were found for the plasmids containing CGG\textsuperscript{C15}CCG whereas single, double and multiple deletions as well as inversions were found after DSB repair in plasmids harboring the CTG\textsuperscript{C15}CAG tracts. The occurrence of mutant clones was higher in recombination-deficient mutant strains than in the parental strains. Also, when active...
transcription occurred across the TRSs, the mutagenic effect was substantially stimulated.

Figure 4 shows a model for the mutagenic effects in flanking sequences which were observed after induction of DSBs into the plasmids by EcoRV cleavage at the TRSs. Several types of non-B DNA structures may occur in linearized DNA that may facilitate recombination events. Studies with short oligonucleotides demonstrated that repeating tracts of CTG, CAG, CGG and CCG single-strands can form intranudplex hairpin structures (60–63). The presence of inverted repeats at the ends of linear DNA promoted the formation of hairpins (64,65) and subsequently caused DNA rearrangements (65). These consequences were suggested to be the result of fraying and/or unwinding at the two free ends of DSBs, thereby, making them suitable substrates for repair and/or recombination enzymes (24,28,45). Linear DNA with terminal inverted repeats may also be initially processed by endonucleases or helicase/nucleases which may promote rearrangements (65).

The length of the TRS influences its stability and mutagenicity (14,23,40,66). More white CFUs were found for (CTG•CAG)70 than for (CTG•CAG)43. This shows that the longer repeat sequences were deleted with a higher frequency and/or more rapidly than the shorter tracts. Moreover, it is known that the DSB repaired DNAs (after recircularization) would be stably replicated and any repeat alterations would have occurred predominantly through erroneous DSB repair (24). Alternatively, it is also possible, even likely, that the number of white CFUs was underestimated in general because of possible mutations in the opposite direction on the plasmid (in the ampicillin resistance gene); since cells harboring these mutations are inviable, their quantitation was not possible (13,55).

A higher fraction of white CFUs was found in orientation II than in orientation I after DSBs were introduced at the TRSs, probably because of the proximity of the TRS to the GFP. This result is likely due to the stronger effect of the repeat tracts on this gene when the repeats are close to GFP. When the repeat tracts are at the distal end to the GFP (orientation I), a looping of the DNA is apparently required to enable the TRS and this gene to interact in order to effect the recombination-repair step. Moreover, multiple deletions and inversions were found only after DSB repair of plasmids containing the CTG•CAG sequence in orientation II. In a related study (B. Kosmider and R. D. Wells, manuscript in preparation), we found gross deletions and complex DNA rearrangements induced by the CGG•CCG tract. However, the results presented herein were obtained from a different system, under different assay conditions and for a shorter CGG•CCG tract.

Prior investigations demonstrated that when transcription is initiated on plasmids there is continuous requirement for

![Figure 3. Influence of the type and orientation of TRS on mutagenic patterns of DSB repair. The percentages of DSB repair products involving single, double and multiple deletions as well as inversions identified in the mutant clones are shown for plasmids harboring: (A) (CGG-CCG)43, in (B) (CTG-CAG)43 and in (C) (CTG-CAG)70. The orientations of the TRSs were explained in the legend to Figure 1. White bars indicate orientation I, whereas filled bars show orientation II. The very small percentages near zero are actually zero but are shown as larger values on this Figure to enable visualization. The data for this Figure was derived from Figure 2. Percentages are shown for each type of product for a given orientation.](https://academic.oup.com/nar/article-abstract/34/19/5369/3111956)
topological constraint, since subsequent cutting with a restriction endonuclease abolishes or substantially reduces this process (67–69). However, a higher number of mutant clones (white CFUs) were detected in the presence than in the absence of *lacZ* transcription for plasmids harboring CTG·CAG and CGG·CCG repeats in all strains. This enhancement of the mutagenicity of the TRSs may be due to: first, replication and transcription complex collisions since replication is faster in *E.coli* (70). Second, transcription modulates genetic instabilities by influencing the formation of non-B DNA structures within the TRSs (70–72). Third, this effect could be dependent on other associated processes, such as translation (11) and not on transcription alone. Previous work showed that the proportion of clones containing deletions in recBCΔ and recAΔ strains was higher than in the parental strains after transformation with linearized pBR322 (28,29,65). This result was explained by the fact that RecBCD is the major enzyme responsible for the transformability of the linearized DNA (65). We found that rec mutations decreased the occurrence of transformations, in general, and increased the fraction of white CFUs in comparison with the parental strains. The highest number of white CFUs was found in recBCΔ and recAΔ strains for both TRSs. However, the number of deleted clones in rec mutant strains increased with increasing CTG·CAG tract length and was higher for the CGG·CCG than for the CTG·CAG sequence. In recBCΔ mutants, the repair pathway is severely inhibited (73), these mutants exhibit very low viability, and the RecA-dependent (or a related) pathway is involved in the absence of RecBCD proteins. It is also well established that recA mutants are hypersensitive to DNA damage (74). Furthermore, null mutations in either the recA, recB or recC genes virtually eliminate DNA break repair (74). Thus the RecBCD pathway has a central role in the repair of linear DNA in *E.coli* (75).

Non-B DNA structures (slipped structures with hairpin loops, cruciforms, triplexes, tetraplexes and sticky DNA, etc.) are important factors in the generation of mutagenic
events (1,2,13). Our sequencing data show that DSB repair at CTG•CAG repeat tracts induced breaks at multiple sites. Furthermore, the sites of the mutagenic events in the deleted clones’ originally harboring CGG•CCG or CTG•CAG repeats coincided with the non-B DNA structures. Moreover, the presence of alternative structures formed by CGG•CCG and CTG•CAG tracts is well established and endonucleolytic cleavage in the DNA can be generated in the secondary structures [reviewed in Ref. (1.2,12)]. Furthermore, the presence of slipped structures within the repeat-containing end was previously reported (15,59) and a polarized/directional deletion effect was observed in COS1 cells (24).

This in vitro and in vivo study was conducted to better understand the role of DSBs at specific loci in genetic instabilities. Our eventual goal is to understand the molecular processes of the genetic instabilities in humans as related to hereditary neurological diseases; however, to date, most mechanistic studies like the present contribution have, of necessity, been conducted in genetically tractable model systems (E.coli and yeast) (2,3).

Different repair products obtained for the CGG•CCG or CTG•CAG tracts indicate the need to study the role of other repeat sequences involved in DSB repair. This will allow a detailed analysis of the mechanism(s) by which each TRS may affect this process. Obviously, future investigations in mammalian systems are required to delineate the relevance of these findings to human diseases. If these mechanisms are operative in humans and if site specific DSBs can be induced in vivo in DNA, this approach could give new insights into the molecular mechanisms of instabilities and possibly serve as a basis for developing therapeutic strategies for the future.

These results show that different types of repair products are induced by DSBs at CGG•CCG or CTG•CAG repeat tracts. Thus, the two repeat sequences must utilize alternative mechanisms of DNA repair and involve different sequence and/or structure specific proteins in the processing of the cleaved DNA ends. Future studies will focus on the identification of the specific proteins or enzymatic systems involved in our proposed model.

SUPPLEMENTARY DATA
Supplementary Data are available at NAR online.

ACKNOWLEDGEMENTS
This work was supported by grants from the National Institutes of Health (ES11347), and the Robert A. Welch Foundation. The authors thank Drs Albino Bacolla and Leslie S. Son for helpful discussions and Jacquelynn E. Larson for assistance. Funding to pay the Open Access publication charges for this article was provided by NIH.

Conflict of interest statement. None declared.

REFERENCES


