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

ABSTRACT

The putative role of double-strand breaks (DSBs) created *in vitro* by restriction enzyme cleavage in or near CGG•CCG or CTG•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 $(CGG \cdot CCG)_{43}$ but single, double and multiple deletions as well as insertions were found for plasmids originally containing $(CTG \cdot CAG)_n$ (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

*To whom correspondence should be addressed. Tel: +1 713 677 7651; Fax: +1 713 677 7689; Email: rwells@ibt.tamhsc.edu

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•CCG or CTG•CAG 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, (Str^R), 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, tsx-33, gsr$^-$, glnV44(AS), ga/K2, λ^-, rac$^-$, hisG4(Oc), rfbD1, mgl-51, rpsL31, kdgK51, xylA5, mtl-1, argE3(Oc), thi-1) (13); mutant *recBC*$^-$ (JC5519 = AB1157 + *rec*B21, *rec*C22) (25); mutant *recA*$^-$ (JC10287 = AB1157 + Δ(*rec*A-srl)306) (25).

Plasmids

The interrupted (CGG•CCG)$_{43}$, (CTG•CAG)$_{43}$ and (CTG•CAG)$_{70}$ tracts were from the parental plasmids pRW4531, pRW4525 and pRW4523 (25), respectively. Due to the high instability of the CGG•CCG tract (40–42) during the TRS construction from the four single synthetic oligonucleotides (25), the longest sequence obtained contained 43 repeats; the numbers of repeats for the CTG•CAG sequence were 43 and 70. These tracts were created to contain a unique EcoRI recognition site inside the TRS and a unique EcoRV recognition site at the junctions of the CGG•CCG or CTG•CAG tracts with the vectors, which were used to induce DSBs *in vitro*.

The pGFPT vector (previously called pRW3619) (13), which contains a transcription terminator cassette was employed for all cloning. However, pGFPT has an EcoRI recognition site downstream of the *GFP* gene; therefore, to induce DSBs only within the TRS, we destroyed this restriction recognition site as follows. The vector was cleaved with EcoRI (all restriction enzymes used in this study were purchased from New England Biolabs Inc., Beverly, MA) and the 5′ overhang was filled-in with the Klenow fragment of *E.coli* DNA polymerase I (U.S. Biochemical Corp.) and the product was ligated for 16 h at 16°C with 20 U of

T4 DNA ligase (U.S. Biochemical Corp.) in the presence of 1 mM ATP. The ligation mixture was ethanol-precipitated and transformed into *E.coli* HB101 by electroporation (2.5 kV, cuvette size 0.2 mm) (43) in a Stratagene Electroporator 1000 and cells were spread onto Luria–Bertani (LB) agar plates containing 100 µg/ml ampicillin (Amp). The transformants were screened using a model 3-6000 Fotodyne hand-held long wavelength ultraviolet (UV) lamp and single colony-forming units (CFUs) with a functional *GFP* gene were used to inoculate a liquid culture (in 10 ml LB medium at 37°C at a shaking rate of 250 r.p.m.). Subsequently, the plasmids from these CFUs were isolated by alkaline lysis using a Wizard Plus Minipreps DNA Purification System (Promega) according to the manufacturer's recommendations. DNA was digested with EcoRI and analyzed on a 1% agarose gel to identify clones lacking this GAATTC site. The undigested clone was called pGFPTΔE where the destroyed EcoRI site was also confirmed by DNA sequencing using the forward primer GGATCCGTTCAACTAGCAG at position 808 and the reverse primer CAAGCTGTGACCGTCTCC at 1165 in the plasmid map. All primers used in this study were purchased from Sigma Genosys, The Woodlands, TX. For the sequencing reactions, the primers and DNA were used at a concentration of 10 pmol/µl and 200 ng/µl, respectively. Cycle sequencing conditions were as follows: initial denaturation at 96°C for 5 min, 25 cycles of heating (96°C, 10 s), annealing (50°C, 30 s) and elongation (60°C, 4 min). The DNA was sequenced in the Molecular and Human Genetics Sequencing Core at the Baylor College of Medicine, Houston, TX. A GeneAMP PCR System 9700 and an ABI 3700 Sequencer were used.
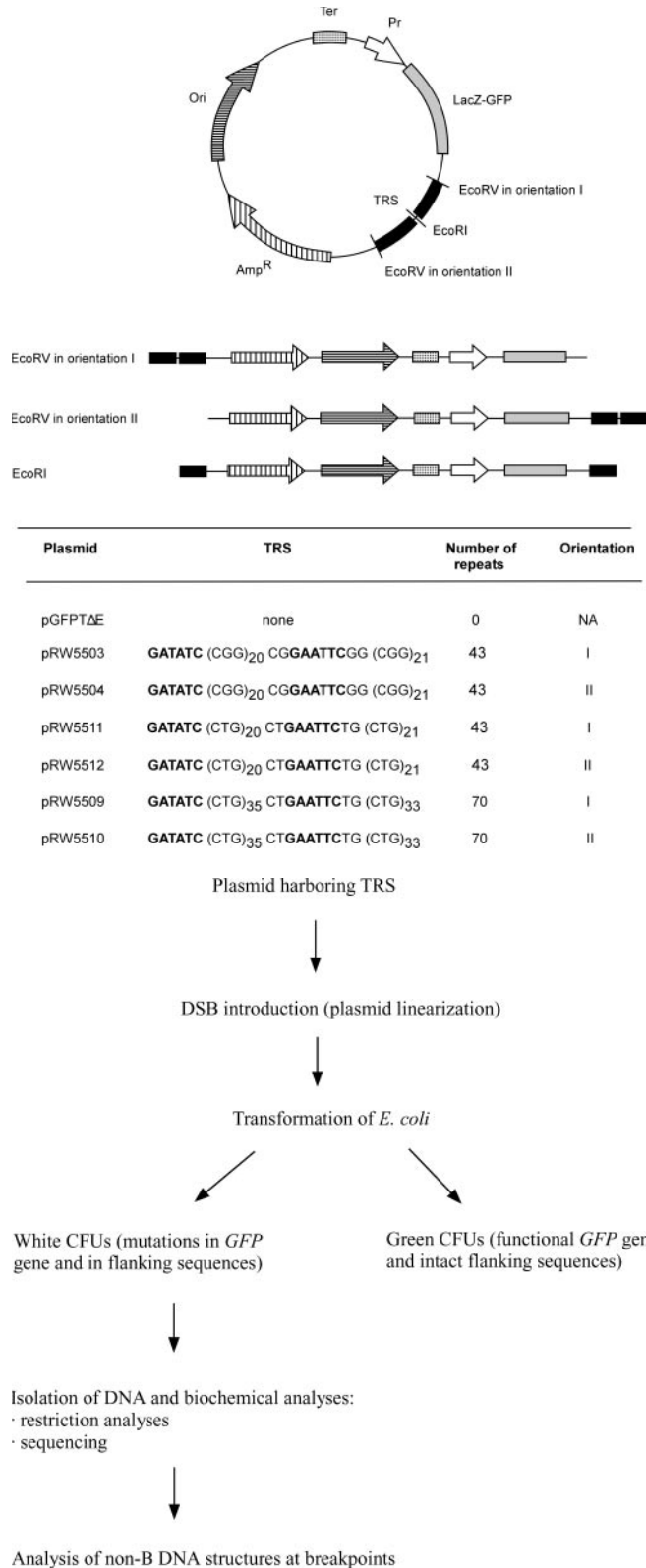
pRW5503 and pRW5504 containing CGG•CCG and pRW5511, pRW5512, pRW5509 and pRW5510 harboring CTG•CAG repeats (Figure 1) were prepared and characterized as follows. To obtain inserts in orientation I, where the (CGG)$_n$ or (CTG)$_n$ strands are the templates for leading strand synthesis, the TRS containing fragments from pRW4531, pRW4525 and pRW4523 (25) were released by Acc65I and HincII digestions. These inserts had, in addition to the 43 CGG•CCG repeats in pRW5503, the 43 CTG•CAG repeats in pRW5511 and the 70 CTG•CAG repeats in pRW5509, 13 bp of non-repetitive flanking sequences at the Acc65I site and 15 bp at the HincII site. The pGFPTΔE vector was digested with BsiWI and StuI because Acc65I and BsiWI have compatible sticky ends and HincII and StuI 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 pRW4531, pRW4525 and pRW4523 (25) were released by XbaI and SmaI digestions. In addition to the 43 CGG•CCG repeats in pRW5504, the 43 CTG•CAG repeats in pRW5512 and the 70 CTG•CAG repeats in pRW5510, 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 pGFPTΔE vector digested with SpeI and StuI since XbaI and SpeI have compatible sticky ends and SmaI and StuI have blunt ends.

The parental digested plasmids pRW4531, pRW4525 and pRW4523 (25) and the linearized pGFPTΔE vector were purified by electrophoresis on a 6% polyacrylamide

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

Plasmid	TRS	Number of repeats	Orientation
pGFPTΔE	none	0	NA
pRW5503	**GATATC** (CGG)$_{20}$ CG**GAATTC**GG (CGG)$_{21}$	43	I
pRW5504	**GATATC** (CGG)$_{20}$ CG**GAATTC**GG (CGG)$_{21}$	43	II
pRW5511	**GATATC** (CTG)$_{20}$ CT**GAATTC**TG (CTG)$_{21}$	43	I
pRW5512	**GATATC** (CTG)$_{20}$ CT**GAATTC**TG (CTG)$_{21}$	43	II
pRW5509	**GATATC** (CTG)$_{35}$ CT**GAATTC**TG (CTG)$_{33}$	70	I
pRW5510	**GATATC** (CTG)$_{35}$ CT**GAATTC**TG (CTG)$_{33}$	70	II

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 (Amp^R); 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 using biochemical analyses and the sequences of the mutant clones were determined to evaluate if the breakpoints were flanked by non-B DNA conformations.

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-^{32}P]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 pRW5511 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.

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 TRS 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 CGG•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-^{32}P]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 CGG•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 Amp and

50 µg/ml kanamycin (Kan) where the *E.coli* strains were first transformed with the plasmid pI^Q-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 (CGG•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 CGG•CCG or CTG•CAG repeats. However, the determinations of the sequences of the mutant clones in plasmids containing the CGG•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 GCTTCCAGGGGGAAACGCCTG and CAGGGTTATTGTCTCAGT, 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, CCTGCGTTATCCCCTGATTCTG

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.

E.coli strain	pGFPTΔE Transcription		(CGG•CCG)$_{43}$ Transcription		(CTG•CAG)$_{43}$ Transcription		(CTG•CAG)$_{70}$ Transcription	
	On	Off	On	Off	On	Off	On	Off
JC5519	**0**	**0**	**0.3820***	**0.1794***	**0.0336***	**0.0188**	**0.1808***	**0.0194**
(*recBC*⁻)	0/382	0/440	34/89	14/78	4/119	1/53	17/94	2/103
JC10287	**0**	**0**	**0.3281***	**0.1238***	**0.0272***	**0.0165**	**0.1111***	**0.0131**
(*recA*⁻)	0/395	0/489	21/64	13/105	3/110	2/121	9/81	1/76
KMBL1001	**0**	**0**	**0.2845***	**0.0694***	**0.0044**	**0.0018**	**0.0152***	**0.0050**
	0/571	0/608	136/478	28/403	2/445	1/532	7/460	2/398
AB1157	**0**	**0**	**0.2057***	**0.0403***	**0.0042**	**0.0019**	**0.1021***	**0.0095***
	0/633	0/725	101/491	15/372	2/471	1/512	29/284	4/418

The pGFPTΔE control vector containing no repeats or plasmids harboring (CGG•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 in the presence and absence of *lacZ* transcription were calculated using the *z*-test. The data which show significant differences are marked with asterisks.

and GCATATGGTGCACTCTC, respectively, were used. For the first two plasmids, their positions were at 3378 and 1432; for next two DNAs at 3383 and at 1457 and for pRW5509 at map positions 3465 and at 1529 and for pRW5510 at 3467 and at 1531. The concentrations of the primers and DNA and the cycle sequencing conditions were as stated above and were sequenced in the same institute.

Analysis of non-B DNA structures at breakpoints in the deleted plasmids

Mutant clones with single, double and multiple deletions as well as with inversions were found in the sequencing data (Supplementary Table 1). Previous studies have shown that different types of repetitive elements coincide with DNA breakpoints, which may mediate mutagenic events (13,14,46–48) (B. Kosmider and R.D. Wells, manuscript in preparation). Therefore, we analyzed the sequences of the mutant clones to evaluate if the breakpoints were flanked by non-B DNA conformations. Supplementary Figure 1 shows that we found the presence of direct repeats, which may form slipped structures, mirror repeats capable of folding into triplex conformations, a tetraplex, a structure formed by a recombination event between sequences with loops, and inverted repeats known to form cruciforms. The latter was determined by the base pairing beyond the Mfold program at http://www.bioinfo.rpi.edu/applications/mfold/old/dna/form1.cgi (49) to include G•T pairs (49,50). The sequences of 150 bp downstream and 150 bp upstream from the breakpoints were analyzed.

Statistical analyses

The statistical *z*-test containing the Yates correction for continuity with the program SigmaStat 2.0 (SPSS Inc., Chicago, IL) was applied. We used the *z*-values to obtain $P(\alpha)$, the probability of incurring a type I error. To be significant, the $P(\alpha)$ value at $P < 0.05$ [$P(\alpha)_{0.05}$] had to exceed 0.800 (13). The statistical significance of mutagenic events was analyzed for TRS in all *E.coli* strains used in this study. First, the numbers of white CFUs were compared with the data for the control pGFPTΔE (Table 1). Second, the numbers of white CFUs in orientation I were compared with orientation II for each

TRS length in the presence (Table 2) and in the absence of *lacZ* transcription (Table 3).

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 DNAs 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 pI^Q-kan plasmid, which expressed a variant *lacI* allele overproducing the *lacZ* repressor (see 'Materials and Methods'); white and green CFUs, respectively, were counted.

Interestingly, no white CFUs, indicative of a functional loss of the neighboring *GFP* gene, 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

E.coli strain	(CGG•CCG)$_{43}$ Orientation I	Orientation II	(CTG•CAG)$_{43}$ Orientation I	Orientation II	(CTG•CAG)$_{70}$ Orientation I	Orientation II
JC5519	**0.3333**	**0.5882**	**0.0156**	**0.0545**	**0.1764**	**0.1860**
(*recBC*⁻)	24/72	10/17	1/64	3/55	9/51	8/43
JC10287	**0.3170**	**0.3478**	**0.0147**	**0.0476**	**0.1020**	**0.125**
(*recA*⁻)	13/41	8/23	1/68	2/42	5/49	4/32
KMBL1001	**0.0957***	**0.6114***	**0.0041**	**0.0049**	**0.0051***	**0.0227***
	29/303	107/175	1/243	1/202	1/196	6/264
AB1157	**0.0890***	**0.3768***	**0.0034**	**0.0054**	**0.0740**	**0.1393**
	26/292	75/199	1/288	1/183	12/162	17/122

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

E.coli strain	(CGG•CCG)$_{43}$ Orientation I	Orientation II	(CTG•CAG)$_{43}$ Orientation I	Orientation II	(CTG•CAG)$_{70}$ Orientation I	Orientation II
JC5519	**0.1333**	**0.2083**	**0**	**0.0833**	**0.0166**	**0.0232**
(*recBC*⁻)	4/30	10/48	0/41	1/12	1/60	1/43
JC10287	**0.1111**	**0.1515**	**0**	**0.0338**	**0**	**0.0196**
(*recA*⁻)	8/72	5/33	0/62	2/59	0/25	1/51
KMBL1001	**0.0529**	**0.0841**	**0**	**0.0033**	**0.0047**	**0.0053**
	10/189	18/214	0/232	1/300	1/211	1/187
AB1157	**0.0183***	**0.0714***	**0**	**0.0035**	**0.0072**	**0.0106**
	4/218	11/154	0/227	1/285	1/137	3/281

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

(28,29,33,45). No white CFUs were found after transformation of all strains with the control-linearized pGFPTΔE 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 the presence and in the absence 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 largest 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
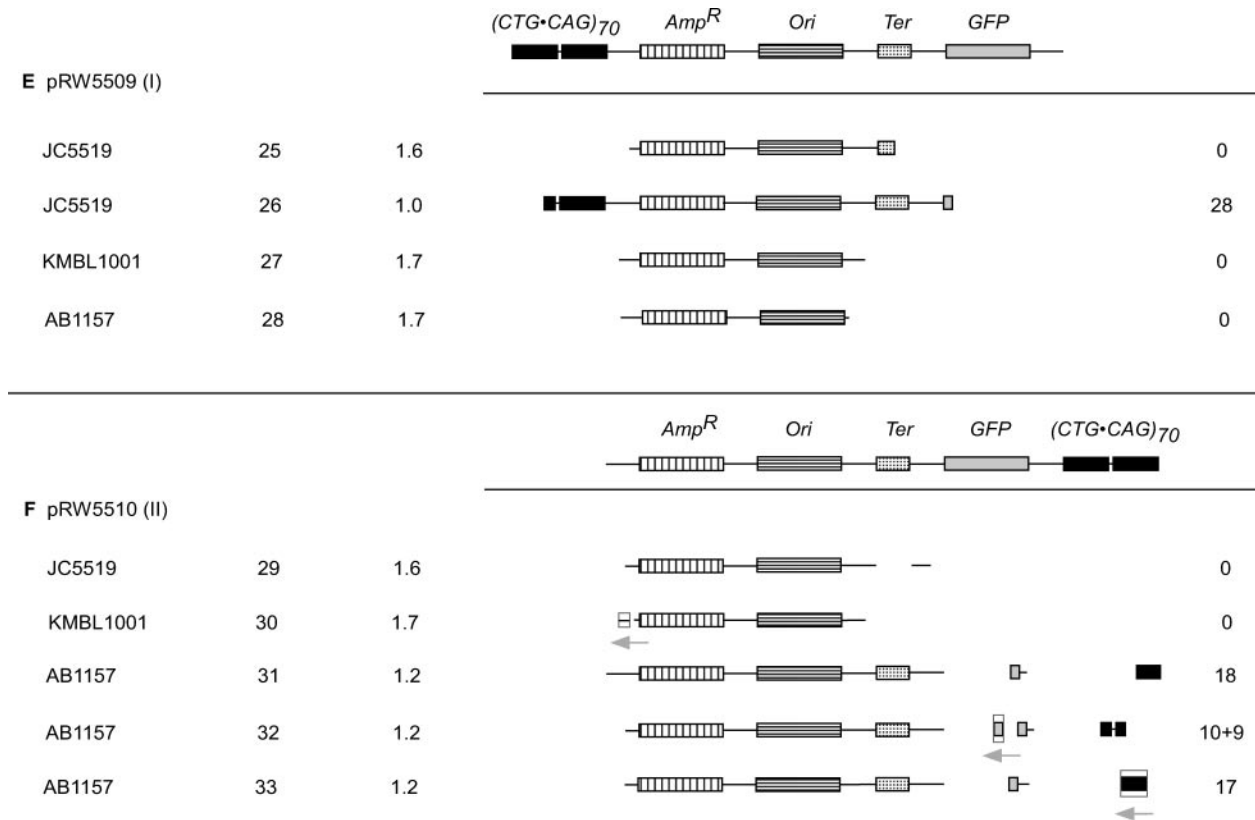
The types of mutagenic events in the flanking sequences induced by linearized DNA containing different types of TRSs were analyzed. Restriction mapping was carried out for 33 deleted clones; each clone represents a single mutagenic event (see 'Materials and Methods'). These products were derived from 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 repeats 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 pGFPTΔE 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,

E. coli strain	Mutant clone	Size of Deletions (kbp)		Number of repeats remaining

A pRW5503 (I)

			(CGG·CCG)₄₃ / AmpR / Ori / Ter / GFP	
JC5519	1	1.6		0
JC10287	2	1.0		14 + 22
JC10287	3	1.1		18 + 22
KMBL1001	4	1.2		11
KMBL1001	5	1.1		16 + 22
KMBL1001	6	1.3		0
AB1157	7	0.4		0
AB1157	8	1.2		13

B pRW5504 (II)

AmpR / Ori / Ter / GFP / (CGG·CCG)₄₃

JC5519	9	0.6		0
JC10287	10	1.6		0
KMBL1001	11	1.2		0
KMBL1001	12	1.5		0
AB1157	13	1.5		0

C pRW5511 (I)

(CTG·CAG)₄₃ / AmpR / Ori / Ter / GFP

JC5519	14	1.1		19
JC10287	15	1.2		0
KMBL1001	16	1.4		0
AB1157	17	1.5		0

D pRW5512 (II)

AmpR / Ori / Ter / GFP / (CTG·CAG)₄₃

JC5519	18	1.2		0
JC5519	19	1.3		0
JC5519	20	1.4		0
JC10287	21	1.0		37
JC10287	22	1.4		0
KMBL1001	23	1.5		0
AB1157	24	1.0		39

Figure 2. Restriction maps of the mutant clones. The first column shows the *E.coli* strains which harbored the appropriate plasmids; the second column lists the number designations of individual mutant clones generated in the presence of *lacZ* transcription; the third column presents the sizes of the deletions (kb); the fourth column shows a schematic representation of the locations of the deletions (presented as open spaces between the DNA segments) and inversions, which are indicated in open boxes with an arrow below to indicate the opposite orientations relative to the origin of replication. In orientation I, the DSB introduced by EcoRV is before the repeat tract (A, C, E) whereas in orientation II it is after the TRS (B, D, F); the fifth column shows the numbers of CGG·CCG or CTG·CAG repeats remaining. The schematic linear plasmids are shown with the symbols described in the legend to Figure 1. The inserts with CGG·CCG tracts containing 43 repeats in orientations I and II are shown in (**A** and **B**), respectively, the CTG·CAG sequences with 43 repeats in orientations I and II are presented in (**C** and **D**), respectively and the CTG·CAG tracts containing 70 repeats in orientations I and II are shown in (**E** and **F**), respectively.

a length dependence was also found because double deletions were observed between 0 and 20% of the cases for (CTG•CAG)$_{43}$ and (CTG•CAG)$_{70}$ 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•CAG harboring 43 and 70 repeats, respectively. The orientation effect may be explained by the presence of CTG•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 *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•CAG or CGG•CCG repeats in plasmids on the genetic instabilities within the TRS or in flanking sequences was investigated. The DSBs were created *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•CCG and CTG•CAG repeats. Only single deletions were found for the plasmids containing CGG•CCG whereas single, double and multiple deletions as well as inversions were found after DSB repair in plasmids harboring the CTG•CAG tracts. The occurrence of mutant clones was higher in recombination-deficient mutant strains than in the parental strains. Also, when active
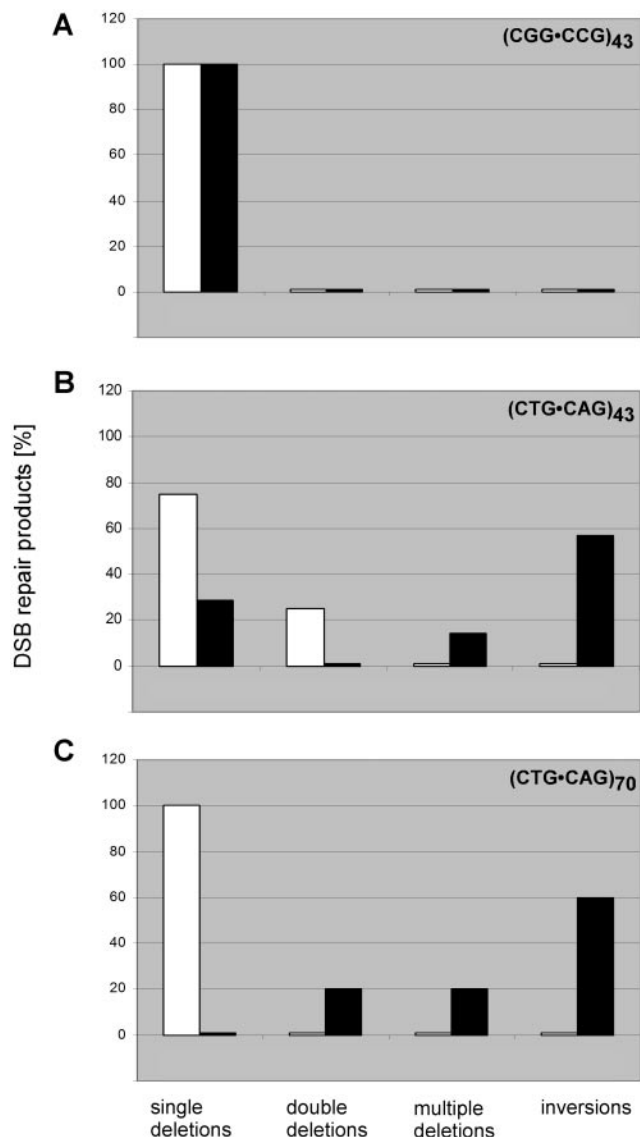
Figure 3. Influence of the type and orientation of TRS on mutagenic patterns of DSB repair. The percentages of DSB repair products involving single, double and multiple deletions as well as inversions identified in the mutant clones are shown for plasmids harboring: (**A**) $(CGG \cdot CCG)_{43}$, in (**B**) $(CTG \cdot CAG)_{43}$ and in (**C**) $(CTG \cdot CAG)_{70}$. 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.

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 at the repeat sequences (1,2,13). Since the DSBs were caused before the transformation, their resulting repeat ends may be susceptible to various types of alterations: first, nuclease degradation; second, end-fraying, which can lead to (a) increased single-strand degradation of the repeat and (b) the formation of unusual DNA structures that are recognized and processed by repair and/or recombination enzymes and third, hairpin formation which would cause reiterative synthesis (59). The proteins involved are unknown at present but their actions result in deletions (24), depending on the type of TRS. For the EcoRV cleaved molecules where the *GFP* reporter gene is distal from the TRS, we propose a looping back of the DNA structure to enable the interactions of non-B DNA structure-containing sequences to facilitate the mutagenic events. For all white CFUs, the *GFP* gene was not functional and, therefore, all or some of this gene was deleted. Figure 4 shows that the plasmid products contain: (A) no TRS; (B) a shortened TRS where the EcoRI site within the tract was also deleted; and (C) plasmids containing a part of the TRS with the EcoRI site retained.

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 intrastrand duplex 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 $(CTG \bullet CAG)_{70}$ than for $(CTG \bullet CAG)_{43}$. 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
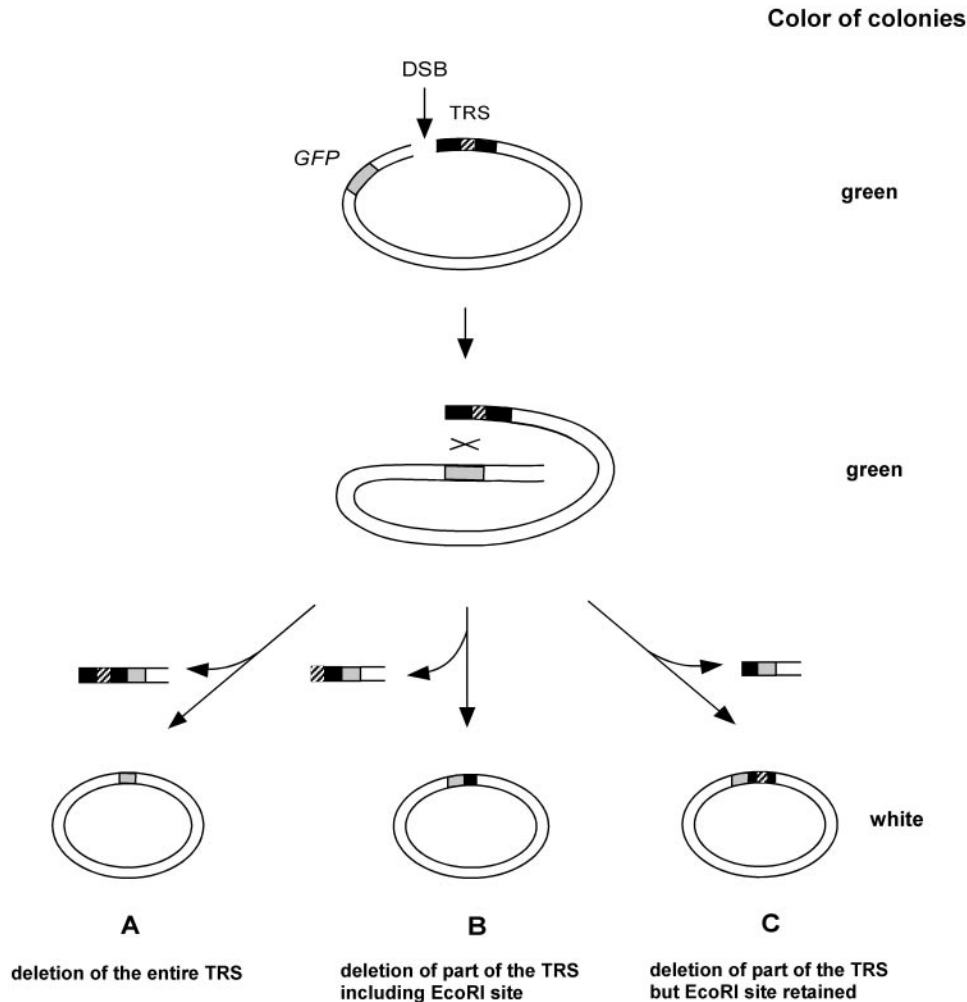
Figure 4. Model of mutagenic events triggered by DSBs introduced by EcoRV at CGG·CCG or CTG·CAG repeats. The filled and gray boxes represent the TRS with interruptions (EcoRI recognition site) and the *GFP* gene, respectively. The ends of the TRS created by DSBs may form alternative DNA conformations which may be processed by structure and/or repeat sequence specific proteins and/or nucleases. The homologous nucleotides between the breakpoints may serve as substrates for DSB repair which leads to the deletion events. All sequenced mutant clones had no functional reporter *GFP* gene (white CFUs) and contained: in (**A**), no TRS because of the deletion of the entire repeat tracts with a part of *GFP* gene; in (**B**), the shortened TRS where both the interruption (EcoRI recognition site), a part of the TRS and the *GFP* gene was deleted and, in (**C**), a portion of the TRS with the interruption but the EcoRI site was retained.

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 *recBC⁻* and *recA⁻* 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 *recBC⁻* and *recA⁻* 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 *recBC⁻* 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 of 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

Conflict of interest statement. None declared.

REFERENCES

1. Wells,R.D., Dere,R., Hebert,M.L., Napierala,M. and Son,L.S. (2005) Advances in mechanisms of genetic instability related to hereditary neurological diseases. *Nucleic Acids Res.*, **33**, 3785–3798.
2. Wells,R.D. and Ashizawa,T. (eds) (2006) *Genetic Instabilities and Neurological Diseases*. Academic Press, San Diego, CA.
3. Gatchel,J.R. and Zoghbi,H.Y. (2005) Diseases of unstable repeat expansion: mechanisms and common principles. *Nature Rev. Genet.*, **6**, 743–755.
4. O'Hoy,K.L., Tsilfidis,C., Mahadevan,M.S., Neville,C.E., Barcelo,J., Hunter,A.G. and Korneluk,R.G. (1993) Reduction in size of the myotonic dystrophy trinucleotide repeat mutation during transmission. *Science*, **259**, 809–812.
5. Shelbourne,P., Winqvist,R., Kunert,E., Davies,J., Leisti,J., Thiele,H., Bachmann,H., Buxton,J., Williamson,B. and Johnson,K. (1992) Unstable DNA may be responsible for the incomplete penetrance of the myotonic dystrophy phenotype. *Hum. Mol. Genet.*, **1**, 467–473.
6. Brunner,H.G., Jansen,G., Nillesen,W., Nelen,M.R., de Die,C.E., Howeler,C.J., van Oost,B.A., Wieringa,B., Ropers,H.H. and Smeets,H.J. (1993) Brief report: reverse mutation in myotonic dystrophy. *N. Engl. J. Med.*, **328**, 476–480.
7. Ashizawa,T., Anvret,M., Baiget,M., Barcelo,J.M., Brunner,H., Cobo,A.M., Dallapiccola,B., Fenwick,R.G., Jr, Grandell,U., Harley,H. *et al.* (1994) Characteristics of intergenerational contractions of the CTG repeat in myotonic dystrophy. *Am. J. Hum. Genet.*, **54**, 414–423.
8. Abbruzzese,C., Costanzi Porrini,S., Mariani,B., Gould,F.K., McAbney,J.P., Monckton,D.G., Ashizawa,T. and Giacanelli,M. (2002) Instability of a premutation allele in homozygous patients with myotonic dystrophy type 1. *Ann. Neurol.*, **52**, 435–441.
9. Hunter,A.G., Jacob,P., O'Hoy,K., MacDonald,I., Mettler,G., Tsilfidis,C. and Korneluk,R.G. (1993) Decrease in the size of the myotonic dystrophy CTG repeat during transmission from parent to child: implications for genetic counselling and genetic anticipation. *Am. J. Med. Genet.*, **45**, 401–407.
10. Harley,H.G., Rundle,S.A., MacMillan,J.C., Myring,J., Brook,J.D., Crow,S., Reardon,W., Fenton,I., Shaw,D.J. and Harper,P.S. (1993) Size of the unstable CTG repeat sequence in relation to phenotype and parental transmission in myotonic dystrophy. *Am. J. Hum. Genet.*, **52**, 1164–1174.
11. Bowater,R.P., Jaworski,A., Larson,J.E., Parniewski,P. and Wells,R.D. (1997) Transcription increases the deletion frequency of long CTG•CAG triplet repeats from plasmids in *Escherichia coli*. *Nucleic Acids Res.*, **25**, 2861–2868.
12. Bacolla,A. and Wells,R.D. (2004) Non-B DNA conformations, genomic rearrangements, and human disease. *J. Biol. Chem.*, **279**, 47411–47414.
13. Bacolla,A., Jaworski,A., Larson,J.E., Jakupciak,J.P., Chuzhanova,N., Abeysinghe,S.S., O'Connell,C.D., Cooper,D.N. and Wells,R.D. (2004) Breakpoints of gross deletions coincide with non-B DNA conformations. *Proc. Natl Acad. Sci. USA*, **101**, 14162–14167.
14. Wojciechowska,M., Bacolla,A., Larson,J.E. and Wells,R.D. (2005) The myotonic dystrophy type 1 triplet repeat sequence induces gross deletions and inversions. *J. Biol. Chem.*, **280**, 941–952.
15. Pearson,C.E., Edamura,K.N. and Cleary,J.D. (2005) Repeat instability: mechanisms of dynamic mutations. *Nature Rev. Genet.*, **6**, 729–742.
16. Pelletier,R., Krasilnikova,M.M., Samadashwily,G.M., Lahue,R. and Mirkin,S.M. (2003) Replication and expansion of trinucleotide repeats in yeast. *Mol. Cell. Biol.*, **23**, 1349–1357.
17. Ohshima,K. and Wells,R.D. (1997) Hairpin formation during DNA synthesis primer realignment *in vitro* in triplet repeat sequences from human hereditary disease genes. *J. Biol. Chem.*, **272**, 16798–16806.
18. Bowater,R.P. and Wells,R.D. (2001) The intrinsically unstable life of DNA triplet repeats associated with human hereditary disorders. *Prog. Nucleic Acid Res. Mol. Biol.*, **66**, 159–202.
19. Balakumaran,B.S., Freudenreich,C.H. and Zakian,V.A. (2000) CGG/CCG repeats exhibit orientation-dependent instability and orientation-independent fragility in *Saccharomyces cerevisiae*. *Hum. Mol. Genet.*, **9**, 93–100.
20. Edamura,K.N. and Pearson,C.E. (2005) DNA methylation and replication: implications for the 'deletion hotspot' region of FMR1. *Hum. Genet.*, **118**, 301–304.
21. Mochmann,L.H. and Wells,R.D. (2004) Transcription influences the types of deletion and expansion products in an orientation-dependent manner from GAC·GTC repeats. *Nucleic Acids Res.*, **32**, 4469–4479.
22. Bowater,R.P., Rosche,W.A., Jaworski,A., Sinden,R.R. and Wells,R.D. (1996) Relationship between *Escherichia coli* growth and deletions of CTG·CAG triplet repeats in plasmids. *J. Mol. Biol.*, **264**, 82–96.

23. Parniewski,P., Jaworski,A., Wells,R.D. and Bowater,R.P. (2000) Length of CTG.CAG repeats determines the influence of mismatch repair on genetic instability. *J. Mol. Biol.*, **299**, 865–874.

24. Marcadier,J.L. and Pearson,C.E. (2003) Fidelity of primate cell repair of a double-strand break in a (CTG){middle dot}(CAG) tract: effect of slipped DN1 structures. *J. Biol. Chem.*, **278**, 33848–33856.

25. Hebert,M.L., Spitz,L.A. and Wells,R.D. (2004) DNA double-strand breaks induce deletion of CTG•CAG repeats in an orientation-dependent manner in *Escherichia coli*. *J. Mol. Biol.*, **336**, 655–672.

26. Hebert,M.L. and Wells,R.D. (2005) Roles of double-strand breaks, nicks, and gaps in stimulating deletions of CTG•CAG repeats by intramolecular DNA repair. *J. Mol. Biol.*, **353**, 961–979.

27. Garaev,M.M., Bobkov,A.F., Bobkova,A.F., Kalinin,V.N., Smirnov,V.D., Khudakov Yu,E. and Tikchonenko,T.I. (1982) The site-specific deletion in plasmid pBR322. *Gene*, **18**, 21–28.

28. Conley,E.C. and Saunders,J.R. (1984) Recombination-dependent recircularization of linearized pBR322 plasmid DNA following transformation of *Escherichia coli*. *Mol. Gen. Genet.*, **194**, 211–218.

29. Conley,E.C., Saunders,V.A. and Saunders,J.R. (1986) Deletion and rearrangement of plasmid DNA during transformation of *Escherichia coli* with linear plasmid molecules. *Nucleic Acids Res.*, **14**, 8905–8917.

30. Leloup,C., Garty,G., Assaf,G., Cristovao,A., Breskin,A., Chechik,R., Shchemelinin,S., Paz-Elizur,T., Livneh,Z., Schulte,R.W. *et al.* (2005) Evaluation of lesion clustering in irradiated plasmid DNA. *Int. J. Radiat. Biol.*, **81**, 41–54.

31. Dar,M.E. and Jorgensen,T.J. (1995) Deletions at short direct repeats and base substitutions are characteristic mutations for bleomycin-induced double- and single-strand breaks, respectively, in a human shuttle vector system. *Nucleic Acids Res.*, **23**, 3224–3230.

32. Wang,Y., Putnam,C.D., Kane,M.F., Zhang,W., Edelmann,L., Russell,R., Carrion,D.V., Chin,L., Kucherlapati,R., Kolodner,R.D. *et al.* (2005) Mutation in Rpa1 results in defective DNA double-strand break repair, chromosomal instability and cancer in mice. *Nature Genet.*, **37**, 750–755.

33. King,J.S., Valcarcel,E.R., Rufer,J.T., Phillips,J.W. and Morgan,W.F. (1993) Noncomplementary DNA double-strand-break rejoining in bacterial and human cells. *Nucleic Acids Res.*, **21**, 1055–1059.

34. King,J., Fairley,C. and Morgan,W. (1998) The joining of blunt DNA ends to 3′-protruding single strands in *Escherichia coli*. *Nucleic Acids Res.*, **26**, 1749–1754.

35. Singleton,M.R., Dillingham,M.S., Gaudier,M., Kowalczykowski,S.C. and Wigley,D.B. (2004) Crystal structure of RecBCD enzyme reveals a machine for processing DNA breaks. *Nature*, **432**, 187–193.

36. Kowalczykowski,S.C., Dixon,D.A., Eggleston,A.K., Lauder,S.D. and Rehrauer,W.M. (1994) Biochemistry of homologous recombination in *Escherichia coli*. *Microbiol. Rev.*, **58**, 401–465.

37. Kuzminov,A. (1999) Recombinational repair of DNA damage in *Escherichia coli* and bacteriophage lambda. *Microbiol. Mol. Biol. Rev.*, **63**, 751–813.

38. Churchill,J.J., Anderson,D.G. and Kowalczykowski,S.C. (1999) The RecBC enzyme loads RecA protein onto ssDNA asymmetrically and independently of chi, resulting in constitutive recombination activation. *Genes Dev.*, **13**, 901–911.

39. Bacolla,A., Jaworski,A., Connors,T.D. and Wells,R.D. (2001) PKD1 unusual DNA conformations are recognized by nucleotide excision repair. *J. Biol. Chem.*, **276**, 18597–18604.

40. Shimizu,M., Gellibolian,R., Oostra,B.A. and Wells,R.D. (1996) Cloning, characterization and properties of plasmids containing CGG triplet repeats from the FMR-1 gene. *J. Mol. Biol.*, **258**, 614–626.

41. Eichler,E.E., Holden,J.J., Popovich,B.W., Reiss,A.L., Snow,K., Thibodeau,S.N., Richards,C.S., Ward,P.A. and Nelson,D.L. (1994) Length of uninterrupted CGG repeats determines instability in the FMR1 gene. *Nature Genet.*, **8**, 88–94.

42. Hirst,M.C. and White,P.J. (1998) Cloned human FMR1 trinucleotide repeats exhibit a length- and orientation-dependent instability suggestive of *in vivo* lagging strand secondary structure. *Nucleic Acids Res.*, **26**, 2353–2358.

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

44. Bacolla,A., Gellibolian,R., Shimizu,M., Amirhaeri,S., Kang,S., Ohshima,K., Larson,J.E., Harvey,S.C., Stollar,B.D. and Wells,R.D. (1997) Flexible DNA: genetically unstable CTG•CAG and CGG•CCG

45. Schulte-Frohlinde,D., Worm,K.H. and Merz,M. (1993) Double-strand breaks in plasmid DNA and the induction of deletions. *Mutat. Res.*, **299**, 233–250.

46. Chuzhanova,N., Abeysinghe,S.S., Krawczak,M. and Cooper,D.N. (2003) Translocation and gross deletion breakpoints in human inherited disease and cancer II: Potential involvement of repetitive sequence elements in secondary structure formation between DNA ends. *Hum. Mutat.*, **22**, 245–251.

47. Chuzhanova,N.A., Anassis,E.J., Ball,E.V., Krawczak,M. and Cooper,D.N. (2003) Meta-analysis of indels causing human genetic disease: mechanisms of mutagenesis and the role of local DNA sequence complexity. *Hum. Mutat.*, **21**, 28–44.

48. Chuzhanova,N.A., Krawczak,M., Thomas,N., Nemytikova,L.A., Gusev,V.D. and Cooper,D.N. (2002) The evolution of the vertebrate beta-globin gene promoter. *Evolution Int. J. Org. Evolution*, **56**, 224–232.

49. Zuker,M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.*, **31**, 3406–3415.

50. SantaLucia,J., Jr and Hicks,D. (2004) The thermodynamics of DNA structural motifs. *Annu. Rev. Biophys. Biomol. Struct.*, **33**, 415–440.

51. Freudenreich,C.H., Kantrow,S.M. and Zakian,V.A. (1998) Expansion and length-dependent fragility of CTG repeats in yeast. *Science*, **279**, 853–856.

52. Thacker,J., Chalk,J., Ganesh,A. and North,P. (1992) A mechanism for deletion formation in DNA by human cell extracts: the involvement of short sequence repeats. *Nucleic Acids Res.*, **20**, 6183–6188.

53. Sugawara,N., Ira,G. and Haber,J.E. (2000) DNA length dependence of the single-strand annealing pathway and the role of *Saccharomyces cerevisiae* RAD59 in double-strand break repair. *Mol. Cell. Biol.*, **20**, 5300–5309.

54. Liang,L., Deng,L., Chen,Y., Li,G.C., Shao,C. and Tischfield,J.A. (2005) Modulation of DNA end joining by nuclear proteins. *J. Biol. Chem.*, **280**, 31442–31449.

55. Wang,G. and Vasquez,K.M. (2004) Naturally occurring H-DNA-forming sequences are mutagenic in mammalian cells. *Proc. Natl Acad. Sci. USA*, **101**, 13448–13453.

56. Parniewski,P., Bacolla,A., Jaworski,A. and Wells,R.D. (1999) Nucleotide excision repair affects the stability of long transcribed (CTG•CAG) tracts in an orientation-dependent manner in *Escherichia coli*. *Nucleic Acids Res.*, **27**, 616–623.

57. Kang,S., Jaworski,A., Ohshima,K. and Wells,R.D. (1995) Expansion and deletion of CTG repeats from human disease genes are determined by the direction of replication in *E. coli*. *Nature Genet.*, **10**, 213–218.

58. Freudenreich,C.H., Stavenhagen,J.B. and Zakian,V.A. (1997) Stability of a CTG/CAG trinucleotide repeat in yeast is dependent on its orientation in the genome. *Mol. Cell. Biol.*, **17**, 2090–2098.

59. Cleary,J.D. and Pearson,C.E. (2003) The contribution of cis-elements to disease-associated repeat instability: clinical and experimental evidence. *Cytogenet. Genome Res.*, **100**, 25–55.

60. Smith,G.K., Jie,J., Fox,G.E. and Gao,X. (1995) DNA CTG triplet repeats involved in dynamic mutations of neurologically related gene sequences form stable duplexes. *Nucleic Acids Res.*, **23**, 4303–4311.

61. Mitas,M., Yu,A., Dill,J., Kamp,T.J., Chambers,E.J. and Haworth,I.S. (1995) Hairpin properties of single-stranded DNA containing a GC-rich triplet repeat: (CTG)15. *Nucleic Acids Res.*, **23**, 1050–1059.

62. Gacy,A.M., Goellner,G., Juranic,N., Macura,S. and McMurray,C.T. (1995) Trinucleotide repeats that expand in human disease form hairpin structures *in vitro*. *Cell*, **81**, 533–540.

63. Chen,X., Mariappan,S.S.V., Catasti,P., Ratliff,R., Moyzis,R.K., Laayoun,A., Smith,S.S., Bradbury,E.M. and Gupta,G. (1995) Hairpins are formed by the single DNA strands of the fragile X triplet repeats: structure and biological implications. *Proc. Natl Acad. Sci. USA*, **92**, 5199–5203.

64. Bohenzky,R.A., LeFebvre,R.B. and Berns,K.I. (1988) Sequence and symmetry requirements within the internal palindromic sequences of the adeno-associated virus terminal repeat. *Virology*, **166**, 316–327.

65. Lin,C.T., Lin,W.H., Lyu,Y.L. and Whang-Peng,J. (2001) Inverted repeats as genetic elements for promoting DNA inverted duplication: implications in gene amplification. *Nucleic Acids Res.*, **29**, 3529–3538.

66. Meservy,J.L., Sargent,R.G., Iyer,R.R., Chan,F., McKenzie,G.J., Wells,R.D. and Wilson,J.H. (2003) Long CTG tracts from the myotonic

dystrophy gene induce deletions and rearrangements during recombination at the APRT locus in CHO cells. *Mol. Cell. Biol.*, **23**, 3152–3162.

67. Harland,R.M., Weintraub,H. and McKnight,S.L. (1983) Transcription of DNA injected into *Xenopus* oocytes is influenced by template topology. *Nature*, **302**, 38–43.

68. Pruitt,S.C. and Reeder,R.H. (1984) Effect of topological constraint on transcription of ribosomal DNA in *Xenopus* oocytes. Comparison of plasmid and endogenous genes. *J. Mol. Biol.*, **174**, 121–139.

69. Weintraub,H., Cheng,P.F. and Conrad,K. (1986) Expression of transfected DNA depends on DNA topology. *Cell*, **46**, 115–122.

70. Krasilnikova,M.M., Samadashwily,G.M., Krasilnikov,A. and Mirkin,S.M. (1998) Transcription through a simple DNA repeat blocks replication elongation. *EMBO J.*, **17**, 5095–5102.

71. Schumacher,S., Pinet,I. and Bichara,M. (2001) Modulation of transcription reveals a new mechanism of triplet repeat instability in *Escherichia coli*. *J. Mol. Biol.*, **307**, 39–49.

72. Pearson,C.E. and Sinden,R.R. (1996) Alternative structures in duplex DNA formed within the trinucleotide repeats of the myotonic dystrophy and fragile X loci. *Biochemistry*, **35**, 5041–5053.

73. Courcelle,J. and Hanawalt,P.C. (2003) RecA-dependent recovery of arrested DNA replication forks. *Annu. Rev. Genet.*, **37**, 611–646.

74. Sargentini,N.J. and Smith,K.C. (1986) Quantitation of the involvement of the recA, recB, recC, recF, recJ, recN, lexA, radA, radB, uvrD, and umuC genes in the repair of X-ray-induced DNA double-strand breaks in *Escherichia coli*. *Radiat. Res.*, **107**, 58–72.

75. Michel,B., Ehrlich,S.D. and Uzest,M. (1997) DNA double-strand breaks caused by replication arrest. *EMBO J.*, **16**, 430–438.