Fidelity of Primate Cell Repair of a Double-strand Break within a (CTG)·(CAG) Tract

EFFECT OF SLIPPED DNA STRUCTURES

Received for publication, April 23, 2003, and in revised form, June 9, 2003
Published, JBC Papers in Press, June 14, 2003, DOI 10.1074/jbc.M304284200

Julien L. Marcadier‡§ and Christopher E. Pearson‡§¶

From the ‡Department of Molecular & Medical Genetics, University of Toronto, Ontario M5A 1X8, Canada and the ¶Program of Genetics & Genomic Biology, The Hospital for Sick Children, Toronto, Ontario M5G 1X8, Canada

At least 15 human diseases are caused by the instability of gene-specific (CTG)·(CAG) repeats. The precise mechanism of instability remains unknown, though bacterial and yeast models have suggested a role for aberrant repair of double-strand breaks (DSBs). Using an established primate DSB repair system, we have investigated the fidelity of repair of a DSB within a (CTG)·(CAG) repeat tract. DSB repair substrates were generated from plasmids that are stably replicated in their circular form, permitting us to highlight the effects of DSB repair on repeat stability and minimize the contribution of replication. DSBs were introduced into repeat-containing plasmids using a unique BsmI site, such that the entire repeat tract comprised one free end of the linearized plasmid. Substrates containing 17, 47, and 79 repeats, in either their linear duplex form or containing slipped structures (out-of-register interstrand mispairings at repeat sequences), were transiently transfected into primate cells. Linearized plasmids with repeats were repaired with mildly reduced efficiency, while the presence of slipped structures considerably reduced repair efficiency. The repaired products were characterized for alterations within the repeat tract and flanking sequence. DSB repair induced predominantly repeat deletions. Notably, a polarized/directional deletion effect was observed, in that the repetitive end of the DSB was preferentially removed. This phenomenon was dramatically enhanced when slipped structures were present within the repeat tract, providing the first evidence for error-prone processing of slipped-strand structures. These results suggest the existence of primate nuclease activities that are specific for (CTG)·(CAG) repeats and the structures they form.

The etiology of at least 15 human genetic diseases, including myotonic dystrophy type 1 (DM1) and Huntington’s disease, has been traced to genetic variation in the lengths of gene-specific (CTG)n·(CAG)n trinucleotide repeats (1). Tracts that are ≥34 can be genetically unstable, and expansions can be as large as 3000 repeats. In affected families there is a common tendency for expansion of the disease allele, although there have been a few DM1 families reported that preferentially transmit repeat contractions to subsequent generations (2–5). Dramatic differences in CTG length variations have been observed in several regions of the brain in DM1, HD, and DRPLA patients when compared with blood of the same patient (1). While many regions of the cerebral cortex show heterogeneous expansions larger than those observed in blood of the same patient (6), the cerebellar cortex consistently shows remarkably shorter lengths in DM1 (7, 8), DRPLA (9), HD, SBMA, SCA1, and SCA3 (1) patients, suggestive of large somatic contractions in this tissue.

The precise mechanism(s) responsible for triplet repeat instability remains to be elucidated. Expansions are dynamic, in that longer repeats tracts are more likely to incur further expansion mutations (1). The process of instability is likely to involve DNA slipped/secondary structures (out-of-register interstrand mispairings at repeat sequences). One hypothesis predicts that expansion events arise from errors at DNA replication forks (10, 11). We have shown that cis-elements such as the location of replication initiation and fork direction, relative to a repeat tract, can lead to different types and frequencies of repeat instability (10). Depending upon the location of replication initiation, (CTG)n·(CAG)n repeat tracts incurred expansions or deletions, or were stably replicated. Repeat instability may be linked to the location of the repeat tract within the single-stranded Okazaki initiation zone of the replication fork. It is possible that the replication fork may become blocked as it passes through the triplet repeat tract (12–14). This blockage may be caused by slipped structures, formed either between the nascent and template strands or between the un-replicated template strands. Elimination of the blockage and subsequent resumption of fork progression may require a double-strand break-mediated recovery mechanism (15–17), which can lead to expansions and/or deletions (18). A second hypothesis is that recombination between repeat tracts containing slipped intermediates can yield repeat expansions or deletions as a result of unequal genetic exchange (19, 20).

While these and other theories remain plausible, it is apparent that several models involve the presence of a double-strand break (DSB). Mammalian cells are constantly threatened by the occurrence of endogenously or exogenously created DSBs. In human cells, it has been estimated that DSBs arising through the DNA replication process occur at a frequency of ~10 per cell cycle, and proper repair is essential for genomic stability and cellular survival (15). Work in bacteria has shown that gene conversion can influence the expansion of CTG-CAG repeats (21). Experiments using yeast have indicated that trip-
let repeat tracts are “fragile,” such that they are susceptible to DSBs, possibly through endonucleolytic cleavage of secondary structures (22, 23). The formation of DSBs within or proximal to the repeats is length-dependent. It has been suggested that subsequent repair of the break can lead to instability of the repeat tract, in some cases yielding large deletions of the repeat and flanking sequences (22–24). The use of a donor sequence containing (CAG)₃₅₆ to repair a DSB lead to integration and expansion of the triplet repeat tract (25). However, in that study the (CTG)n (SI-DNA), equivalent to (CTG)n tract, and a (CTG)(CAG) repeat. Thus, what remains to be addressed is the influence and fidelity of DSB repair within a repeat tract upon tract stability. It is also important to determine the effects of both repeat tract length and the presence of alternative DNA structures on the fidelity of DSB repair. Furthermore, it is of interest to utilize primate cells, as yeast and mammalian cells are known to preferentially use homologous recombination and non-homologous end joining (NHEJ), respectively (26). The differences between these two DSB repair mechanisms may therefore result in distinct repair products.

In this study, we have used an established primate DSB repair system (27) to investigate the fidelity of repair of a DSB within a (CTG)(CAG) repeat tract. Following repair of the DSB substrates in primate cells, products were characterized for sequence alterations within the repeat tract length and the flanking sequence. The presence of a DSB within the repeat tract lead to several forms of deletion events, removing either exclusively repeats, non-repetitive flanking sequences, or both. The repeat-containing end of the DSB was preferentially deleted, relative to the non-repetitive end. The presence of intrastrand slipped structures within the repeat tract enhanced this “polared” deletion effect.

EXPERIMENTAL PROCEDURES

Construction of DSB Repair Templates—The (CTG)(CAG) DSB repair templates were made from a series of (CTG)n clones harboring either 79, 47, or 17 repeats as well as the SV40-ori (Fig. 1A). In their circular form these plasmids have been shown to be stably replicated by primate and bacterial proteins (10), thereby enabling us to highlight the contribution of DSB repair to repeat instability. These clones (pDM79ED, pDM47ED, and pDM17ED; the “ED” has been omitted throughout the text) have been described previously (10). Briefly, templates contained uninterrupted (SI-DNA), equal quantities of pDM47 EDI, and (CTG)(CAG) repeats. All clones had the repeat in the stable orientation relative to the uni-directional bacterial ColE1 origin of replication (CAG strand as lagging strand template). The SV40 viral replication origin (viral positions 128 and 5171), which contains the SV40 origin of bi-directional replication (OBR, viral position 5210/5211), is located 229 nt 5' of the CTG tract, permitting stable replication by primate cell proteins (10).

Large scale plasmid preparations were prepared from dam + Esche- richia coli cells as previously described (19). Briefly, cells were harvested and lysed with lysozyme (Invitrogen) and a detergent solution of 1% Brij 58 (Sigma) and 0.4% deoxycholate (Sigma). Plasmids were treated with RNase A and T1 (Sigma), phenol-extracted, and purified twice by cesium chloride/ethidium bromide centrifugation.

Template Linearization—The linearization of (CTG)n-containing plasmids was performed by Bsm1 digestion (New England Biolabs). The completion of linearization was verified through visualization on a 1% agarose gel, where only a single band (linearized DNA) was observed. Linearized templates were also transformed into E. coli cells and yielded no colonies, confirming the poor ability of E. coli to rescue (re-circularize) linear plasmids (28).

Heteroduplexing/Reduplexing—The induction of slipped structures in the repeats was performed as described previously (19, 29–31). For reduplexing (S-DNA), only the pDM79 plasmid was treated (5 μg), and for reduplexing (SI-DNA), equal quantities of pDM47 EDI and pDM17 EDI were mixed (2.5 μg each). Briefly, the DNAs were denatured for 5 min in denaturing solution (0.5 M NaOH, 1.5 M NaCl, pH 13), diluted 50-fold in modified 5X TE (pH 8.0), 5 μM EDTA, 0.275 mM NaCl), and renatured at 68 °C for 3 h. The relative amounts of DNA containing slipped structures were determined by densitometric anal-

yses of electrophoretic patterns of radiolabeled repeat-containing DNAs (Fig. 1C).

Cell Culture/Transfections—Cos1 cells, which express SV40 T-anti-
gen (T-Ag) were grown in DMEM supplemented with 10% (v/v) fetal bovine serum (FBS). Cells were plated to a confluency of 20% in 100-mm dishes, 1 day prior to transfection. The following day, cells at 40% confluency were incubated according to the manufacturers protocol (Stratagene) with 5 μg of linearized plasmid DNA and 50 μl of Lipotaxi transfection reagent in 4 ml of DMEM lacking FBS. The amount of DNA used per cell was such that any given cell was transfected with only one plasmid. After 4 h of incubation, 5 ml of DMEM supplemented with 20% FBS was added. The medium was replaced with fresh DMEM (10% FBS) following 24 h of incubation. Episomal DNAs were then extracted from the cells, as described (10), with Tris-HCl (pH 7.5), 1 mM EDTA, and 0.6% SDS. DNA was purified with multiple phenol/chloroform extractions, ethanol-precipitated, and resuspended in TE (10 mM Tris-HCl (pH 7.6), 1 mM EDTA).

Measuring Efficiency of DSB Repair—Intersubstrate DSB repair efficiencies (Fig. 2) were determined as follows. The linear repeat-containing repair templates pDM79, pDM79-S-DNA, and pDM79-T-DNA (linearized with Bsm1) were mixed with each of the control SV40-ori plasmid (pM1), which is void of repeats, and these mixtures were co-transfected into Cos1 cells. Following DSB repair and replication, purified products were linearized with Xho1, and unreplicated material eliminated by Dpn1 digestion. Products were electrophoresed on a 1% agarose gel to resolve the plasmid and pM1 repair products, and Southern blotting using a 32P-labeled SV40-ori probe. The relative amounts of pre-transfection starting mixtures and repaired material were determined by densitometric analysis. The efficiency of DSB repair was determined by the relative differences in the amounts of repeat-containing plasmids compared with control plasmid. The repeat-containing plasmids replicated with equal efficiency to the control pM1 plasmid.

Analysis of Stability/STRIP Assay—Repeat stability was assessed using the Stability of Trinucleotide Repeats by Individual Protein analysis (STRIP assay). This assay has been described in detail (10, 11). Briefly, DSB repair products in the Hirts lysate were digested with 10 units of Dpn1 (New England Biolabs) to eliminate unreplicated and unreplicated parental templates. Dpn1-digested material was transformed into DH5α-mcr E. coli, and only the Dpn1-resistant material gave rise to colonies. Individual bacterial colonies, each representing an individual product of primate DSB repair and replication, were picked and cultured. Primate replication-mediated instability was minimized as the starting templates are stably replicated (10). Bacterial contribution to repeat length alterations was minimized by 1) having the (CTG)(CAG) repeat in the stable orientation relative to the bacterial ColE1 replication origin, 2) the limited colony growth time (≤8 h), and 3) bacterial strain selection. Through the use of these strict specifications, 255 CTG repeats can be stably maintained in bacteria (31).

Miniprep DNA was analyzed for repeat instability through analysis of the migration pattern of BamHI/HindIII fragments on high resolution 4% polyacrylamide gels (19). The magnitude of repeat length changes were ascertained by electrophoretic sizing of the repeat-containing fragments relative to starting length material and a size marker. All expansions and samples of each deletion type were confirmed through sequence analysis.

RESULTS

Double-strand Break Assay and Substrates—To test whether the repair of DSBs within trinucleotide repeat tracts affected repeat stability in primate cells, we used a plasmid recovery system based on SV40-mediated DNA replication. This is essentially a modification of a previously described assay for the fidelity of DNA ligation and DSB repair (27). Linear SV40 templates are incapable of DNA replication, and only those templates that are re-circularized by DSB repair are capable of primate cell-mediated replication. As a result, DNA replication can be used as an indicator of DSB repair. To highlight the effects of DSB repair on repeat stability, we used as DSB repair substrates linearized plasmids (pDM79ED, pDM47ED, and pDM17ED), which we have previously shown to be stably replicated in their circular form by primate and bacterial proteins (10). As a result, the DSB-repaired (re-circularized) DNAs would be stably replicated and any repeat alterations would have been incurred predominantly through erroneous DSB repair. Therefore, in using linearized pDM substrates containing...
Slipped Structures and Fidelity of DSB Repair

**Fig. 1. DSB repair substrates.** A, the (CTG)\(\times\) (CAG) repeat is in the stable orientation relative to the uni-directional bacterial CoIE1 origin of replication. To highlight the effects of DSB repair on repeat stability, DSB repair substrates were generated by linearizing specific plasmids (pDM79, pDM47, and pDM17), which we have previously shown to be stably replicated in their circular form by primate and bacterial proteins (10). The SV40 ori is situated 229 nt 5′ of the CTG tract, which permits the stable primate replication of the repeat tracts. Thus, the DSB-repaired (re-circularized) DNAs would be stably replicated and any repeat alterations would have been incurred predominantly through erroneous DSB repair. B, digestion with BsmI causes the entire repeat tract to comprise one free end of the linearized plasmid. The two arrows shown below represent the maximum allowable deletions before compromising bacterial (left arrow) and primate (right arrow) replication. In both panels, \(n\) represents the number of repeats present in the tract, with appropriate sizes or positions displayed accordingly. Relevant restriction sites are also shown. C, fully duplexed and slipped DSB repair substrates. Linear plasmids were reduplexed (S-DNA) or heteroduplexed (SI-DNA) to induce the formation of slipped structures within the repeat tract prior to transfection into primate cells (see “Experimental Procedures”). All plasmids (pDM17, pDM47, and pDM79) were individually transfected in their fully duplexed form (column 1). Only the pDM79 linear template was reduplexed (column 2). Linear pDM47 and pDM79 templates were mixed in equal amounts for the production of heteroduplexed DNA (column 3). For the (CTG)\(\times\) (CAG)\(\times\) S-DNAs in the gel lane of column 2, the pattern was identical for radiolabeling on either strand. The heteroduplex mixture contained all molecules shown in both gel lanes of column 3. Linear, S-DNA, or SI-DNA molecules, containing the (CTG)\(\times\) strand or the (CTG)\(\times\) strand, were visualized by unique radiolabeling (*). Thus, the heteroduplex mixture contained a merger of the two lanes. The black and gray lines represent the CTG and CAG strands, respectively.

17, 47, and 79 repeat units (10), hereafter referred to as pDM17, pDM47, and pDM79, respectively, we were able to emphasize the effects of DSB repair on repeat stability by minimizing the effects of primate and bacterial replication. The analysis of repeat stability is described in a later section, titled STRIP Assay.

**Experimental Procedures**

To determine potential differences in DSB repair efficiency, which may be caused by the presence of (CTG)\(\times\) (CAG) repeats or slipped structures, the degree of repair of repeat-containing substrates was compared with that yielded by an identical linear plasmid lacking a repeat tract (pJM1). Repeat-containing substrates and the control plasmid were mixed at equimolar concentrations (Fig. 2, “starting material”), co-transfected and repaired/replated material was isolated 48 h later. DNA from the starting plasmid mixtures was electrophoresed along with the repaired material, and relative concentrations determined through densitometric analysis (Fig. 2). It was found that the pDM79 substrate, with the repeats in their linear, Watson-Crick form, was repaired with 91% percent efficiency relative to the non-repetitive control plasmid (pJM1). The presence of homoduplex S-DNA structures decreased the DSB repair efficiency, with the pDM79-S-DNA substrate being repaired at 76% efficiency relative to pJM1, and 83% relative to the linear form of the repeats. The pDM47×79-SI-DNA substrate was repaired at 71% efficiency relative to pJM1, and 78% efficiency relative to the linear form of the repeats. The differences observed are due...
through electrophoretic comparison of the BamHI restriction digestion patterns. The occurrence of these recombinants was not considered for our results.

The STRIP assay—To address the effect of DSB repair on (CTG)/(CAG) repeat instability, we modified an assay we have developed to determine the role of primate DNA replication on the STRIP assay (10). Following enzyme-induced linearization, repeat-containing plasmids were transfected into COS1 cells, and the DSB-repaired (re-circularized) products were recovered as replicated DNAs using Hirt's lysis and DpnI digestion. To reduce contributions to DSB repair by homologous recombination, the transfections were performed such that any cell contained a single plasmid (“Experimental Procedures”). The DpnI-resistant material was transformed into bacterial cells to generate individual colonies, each derived from a single product of DSB repair. Plasmids of each colony were analyzed for repeat length alterations by restriction digestion and visualization on polyacrylamide gels.

Classification of Repeat Instabilities—As a means of determining repeat stability, the plasmids isolated from individual colonies were digested with BamHI and HindIII to release both BamHI-HindIII and the SV40-ori restriction fragments, containing the CTG repeat and the SV40-ori, respectively. An assessment of stability was then made as follows: rescued plasmids with Class II deletions displayed a BamHI-HindIII CTG-repeat fragment of greater electrophoretic mobility, and a BamHI-BamHI SV40 fragment identical to the control (Fig. 3B; see lanes labeled Class I). Also, several plasmids harbored deletions that encompassed more than just the repeat. The nature of these deletions was assessed as follows: rescued plasmids with Class II deletions displayed only the BamHI-BamHI SV40 fragment (Fig. 3B; see lanes labeled Class II), having sustained deletions of the entire repeat tract, as well as the HindIII site on their 3′-end (Fig. 3, A–ii). Another subset of rescued plasmids, termed Class III deletions, showed a single band that varied in size between plasmids, but was consistently larger than the standard BamHI-BamHI SV40 fragment (Fig. 3B; see lanes labeled Class III). These rescued plasmids had sustained deletions that removed the BamHI site nearest the BsmI site in the repeat tract, as well as the HindIII site on their 3′-end (Fig. 3, A–ii). A sample of each deletion class was confirmed through sequencing analysis.

DSB repair induced high levels of deletion products for each of the linear substrates. The pDM17 substrate incurred deletions at a frequency of 43%, of which 31, 8, and 15% were Class I, Class II, and Class III events, respectively (Fig. 4A). The pDM47 substrate incurred deletions at a frequency of 54%, of which 38, 10, and 16% were Class I, Class II, and Class III events, respectively (Fig. 4A). The pDM79 substrate incurred deletions at a frequency of 66%, of which 47, 12, and 21% were Class I, Class II, and Class III events, respectively (Fig. 4A). Class III deletions were increased only for the pDM79 substrate. It is noteworthy that while the frequency of Class II deletions was similar between the different repeat lengths, the
amount of DNA lost was greater for the longer repeat tracts. Both Class II and Class III deletion products were specific to DSB repair, as neither of these has ever been observed following bacterial or primate DNA replication (10). However, Class I deletion products, having lost only repeat units, were observed following DSB repair (herein) and bacterial or primate DNA replication.
replication (10). As stated above, the plasmids used as DSB repair substrates have been shown previously to be stably replicated in primate cells (10). This is summarized in Fig. 4B, where it can be seen that primate background deletion levels (stippled bars) are equal to bacterial background deletion levels (hatched bars), which means that the deletions observed following primate replication were originally present in the starting plasmid preparation. Likewise, a proportion of the Class I deletions observed in our DSB-repaired material can also be attributed to background length heterogeneity. Consequently, the levels of DSB repair-induced Class I deletions observed were 19% (19–0%) for pDM17, 11% (31–20%) for pDM47, and were negligible (34–36%) for pDM79 (Fig. 4B). For pDM17, Class I deletions occurred only following DSB repair, but were not detected following either bacterial or primate replication (Fig. 4B), providing further support for the enhanced deletions induced by DSB repair. The apparently reduced levels of Class I events for the longer repeats may be due to enhanced deletions into the repeat tract and the adjacent flanking regions, a phenomenon that may be stimulated by an inherent ability of the longer repeats to “fray” into alternative DNA structures (19, 29).

The presence of homoduplex slipped-strand S-DNAs affected the type of DSB repair-induced deletion products, but not the overall deletion frequency. The pDM79-S-DNAs incurred overall deletion events at a frequency of 66%, which is similar to that of the fully paired linear substrate. However, the slipped substrates incurred enhanced deletions into the repeat tract, as is evidenced by the relative increase in Class II deletions and decrease in Class III deletions compared with the fully base-paired linear substrates (Fig. 4A). It therefore appears that homoduplex S-DNAs are more susceptible to deletions into the repeat tract, relative to duplex linear repeats.

DSB repair of substrates containing heteroduplex slipped-intermediate SI-DNAs also affected the types of deletion products generated. Repair products of the pDM47×79-SI-DNA substrate also contained high levels of deletions, totaling 64%, and comprising of 10% that were Class II deletions and 17% that were Class III deletions (Fig. 4A). It was not possible to assess the frequency of Class I deletions for the heteroduplexed

---

**Fig. 4. Overall stability following DSB repair.**

A, shown in the upper portion is the occurrence of DSB repair products for linear fully duplexed substrates. The middle portion shows the occurrence of DSB repair products for linear substrates containing slipped repeat structures (homoduplex S-DNAs and heteroduplex SI-DNAs). The bottom portion shows the occurrence of DSB repair products for the non-repetitive control plasmid (pJM1). Shown for each are the total number of observed products, expansions, and each deletion class (see Fig. 3A). B, shown is the background repeat length heterogeneity present in the starting plasmid preparations (cross-hatched bars), and the stable primate replication (same as background, stippled bars), the observed DSB Class I deletions (filled bars) and the DSB Class I deletions corrected for the bacterial background (hollow bars). Corrected DSB repair-induced Class I deletion frequencies were calculated as (DSB Class I deletions observed/total number of DSB repair products observed) – bacterial background deletions. This correction for background was as we have previously described (10). For each plasmid, the same preparation was used for all experiments. The primate replication results have previously been reported (10). For details please see “Results.”
Preferential Deletion of the Repeat-containing End of the DSB. A, expected frequencies of Class II and Class III deletions were determined for all repair substrates (left column), using the distance in nucleotides from the DSB to the HindIII site (H) and the proximal BsmI (B) site, respectively. For a given repair template, the ratio (distance to HindIII-distance to BsmI) is then equivalent to the expected ratio (Class III frequency/Class II frequency). The black and gray lines represent the CTG and CAG strands, respectively. Expected Class II and Class III deletion frequencies for the control plasmid (pJM1) were calculated using the distance in nucleotides from the KpnI-induced DSB to the PsiI site (Ps) and the PacI site (Pa), respectively. The white lines represent non-repetitive DNA sequence. Observed Class III:Class II frequencies (middle column) were calculated using the observed events presented in Fig. 4A. B, the expected Class III:Class II ratios for all repair templates were compared with their respective observed ratios. For the heteroduplexed SI-DNA repair template, expected values are shown for both repeat tracts containing 47 and 79 repeats.

Substrates; however, they occurred at a minimum frequency of 30% (represented by deletion products <47 repeats; see Fig. 4A).

Since the DSB was induced by BsmI digestion, it was of interest to determine if the restriction site was maintained. The status of the BsmI site in the three deletion classes varied (data not shown). For all repeat lengths, in both linear and slipped-structure conformations, Class I deletions showed a relatively even distribution of products that preserved the BsmI site and products that lost the site. The exception was for rescued pDM47×79-SI-DNA heteroduplex substrates that resulted in repeat fragments whose size was between 47 and 79 repeats, where the BsmI site was consistently preserved. Class II deletions showed a similar, seemingly random distribution of BsmI site-preservation frequencies. Finally, the BsmI site was abolished in all examples of Class III deletions, which is not surprising since Class III products have incurred loss of sequence in the non-repeating end that contains the BsmI recognition site.

Preferential Deletion of the Repeat-containing End—Remarkably, it was found that the repeat-containing end of the DSB was preferentially degraded during DSB repair. Moreover, the presence of slipped structures within the repeat tract enhanced this directional deletion effect. DNA ends at a dou-

ble-strand break are susceptible to exonucleolytic degradation, and subsequent ligation of these ends can result in deletions. Assuming that the rate of exonucleolytic digestion of DNA ends is insensitive to DNA sequence or structure, one would expect both the repeat end and the non-repeat end to be degraded at similar rates. Intuitively, Class II deletion frequencies should decrease as repeat tract length increases, simply due to the greater number of nucleotides between the BsmI DSB and the HindIII site (Fig. 5A). However, Class III deletion frequencies should remain relatively constant, since the number of nucleotides between the BsmI DSB and the proximal BsmI site is the same for all repair substrates. In this manner, the proportion of Class III deletions should increase relative to Class II deletions for substrates with longer repeat tracts, and the Class III/Class II ratio for a given substrate would be directly related to the differential lengths from each end to their respective diagnostic sites.

Taking these assumptions into account, we have estimated the expected proportion of Class III to Class II frequencies based on the nucleotide lengths that would be lost for each event. More specifically, the length ratios are the distances from the DSB site to the Class II and Class III diagnostic restriction sites, namely from BsmI to HindIII and BsmI to BsmI, respectively (Fig. 5A, left column). Comparing these expected III/II ratios to the experimentally observed III/II ratios revealed considerably lower values in the latter (Fig. 5B), indicating a marked bias for the deletion of the repeat-containing end of the DSB.

The difference between observed and expected values for the slipped structure-containing pDM79-S-DNA substrate was statistically significant ($\chi^2$-test, $p = 0.0025$). Additionally, as both the S-DNA and SI-DNA mixes did not consist entirely of molecules containing slipped structures (70 and 50%, respectively), the effect of slipped structures on the observed deletion bias may have been underestimated due to the presence of repeat tracts in their linear Watson-Crick form. Thus, the presence of slipped structures within the repeat tract enhanced the preferential deletion of repeats.

This method of analysis was verified by performing an identical DSB repair experiment using a plasmid (pJM1) similar to the pDM substrates, lacking only the repeat tract (Fig. 5A). Repair products of the linearized control plasmid produced an observed Class III/Class II ratio similar to that which was expected (Fig. 5B). This further illuminates the effect that both repeat sequence and slipped repeat structures have on the outcome of DSB repair.

To address the possibility that repeat tracts of different lengths were deleted at different rates, we compared the expected over observed Class III/Class II ratios between different substrates (Fig. 5A, right column). As predicted, there was little difference observed for the control pJM1 plasmid, which lacked repeats (0.85), and a minimal difference for the pDM17 substrate that contained only 17 repeats (1.31). For both the pDM47 and pDM79 substrates, which contained longer repeat tracts, the difference was considerable (2.06 and 2.16, respectively). The pDM79-S-DNA substrate, which contained slipped structures, showed the greatest difference (4.71). Thus, the rate of polarized deletion was ~2-fold greater for tracts that contained greater than 17 repeats, while the presence of slipped structures led to a nearly 5-fold enhanced deletion rate. Thus, not only are repeat-containing ends more susceptible to deletions than non-repeat ends, but longer repeat tracts are deleted more frequently or rapidly than shorter tracts. To our knowledge, this is the first observation of a polarized deletion bias for any sequence or DNA structure.
DISCUSSION

Through the use of primate cell repair machinery, we have shown that the repair of a double-strand break located within a (CTG)(CAG) trinucleotide repeat tract resulted in alterations of repeat tract length. Overall, DSB-induced deletions were more frequent than what was observed for primate replication-induced instability (10). This is evidenced by the 43% deletions generated by DSB repair of the linearized (CTG)_{17} substrate, which contrasts with the consistently stable replication of the same template in its circular form (10). Furthermore, the presence of slipped structures within the repeat-containing end enhanced the preferential deletion of repeats, providing the first line of evidence that processing of slipped structures can lead to repeat instability. Our results are consistent with studies in bacterial (21) and yeast (22–25) models that propose a role for recombination and DSB repair in trinucleotide repeat instability.

The improper repair of a DSB located within a triplet repeat tract resulted most commonly in deletions. As tract length increased above the unstable threshold length of 34, the occurrence of deletions into the repeat end increased. This is the first time this phenomenon has been observed for a DSB located within a repeat tract in primate cells. Similar double-strand break repair assays that have not concerned repeating sequences have also found deletions to be the most common result of improper DSB repair (27, 32). It has been proposed that this bias toward deletions is likely a result of fraying/unwinding at the two free ends of the double-strand break, thereby becoming suitable substrates for repair/recombination enzymes (33). For the repeat-containing substrates used herein, the preferential deletion of the repeat end in duplex molecules may be due to the increased ability of this end to fray and form substrates for nucleases that would enhance deletions. Free ends with different base compositions are known to vary in their degree of end fraying (breathing) (34), which may affect their ability to form alternative structures, as in the terminal repeats of the linear adenovirus (35). It is possible that CTG-CAG repeats at free ends are inherently prone to breathing and structure formation, with subsequent enzymatic processing possibly leading to deletions (Fig. 6). This model is supported by the enhanced deletion of DSB substrates containing slipped structures.

Our results propose a role for sequence- or structure-specific protein(s) which preferentially remove triplet repeat-induced slipped structures located at a DSB as an early step in the repair process (Fig. 6). We have recently demonstrated that both S-DNA and SI-DNA structures can be specifically recognized and cleaved by both junction- and single strand-specific nucleases (36). Mammalian cells are believed to use non-homologous end joining (NHEJ) as a primary means of DSB repair (26), making proteins such as Ku70, Ku80, DNA-PKcs, DNA ligase IV, XRCC4, and Artemis likely candidates. Similarly, the human Mre11-Rad50-NBS1 complex, which is a part of the homologous recombination process, has been directly implicated in the scission and/or unwinding of DNA hairpins (37). Other junction- or structure-specific endonucleases (38) may act upon slipped structures. The structure-specific WRN DNA helicase may participate in the repeat instability observed (39). Many of these proteins are involved in the maintenance of telomeric repeats through telomere-specific DNA structures (40). Thus, triplet repeats situated at a DSB end may not be protected from end degradation, as are telomeric repeats at chromosome ends (41). While it is not known which specific proteins participate in the processing of slipped repeat structures, we have clearly shown that when present at a DSB, their processing results in deletions.

Repeat tract length changes may result from a combination of replication fork blockage and DSB repair (12–14, 16, 18). The re-establishment of stalled fork progression in mammalian cells may involve the induction of DSBs and subsequent break repair (15, 17, 18). Previously, we reported template-specific instability of (CTG)(CAG) tracts following episomal replication in primate cells (10). Depending upon replication direction and the location of the repeat tract relative to replication initiation, we observed either a bias for deletions, a bias for expansions or stable replication. Replication fork blockage at the repeats and DSB-induced re-establishment of fork progression may contribute to the type of mutations incurred by specific template configurations. The results presented here suggest that DSBs within triplet repeats increase the frequency of repeat dele-
tions. This would suggest that the circular replication templates pDM79AP, pDM79HD, and pDM79E that show a predominance of deletions (10) may be more commonly subject to fork blockage within the repeat tract, and that DSB induction and aberrant repair lead to preferential repeat contractions.

The DSB in our substrates may share some, but not all, structural features present in DSBs associated with stalled replication forks, and as such may not be subject to the same repair/recombination processes. Such differences may explain the paucity of expansion events observed in our DSB repair assay. Conversely, the DSB present in our substrates may more closely reflect breaks that occur spontaneously throughout the life of a cell (22). The deletions incurred by genome maintenance-associated repair may reflect the age-dependent length heterogeneity observed in non-proliferative cells of mice and patients harboring expanded trinucleotide repeats (42–46). Our results suggest that the natural or environmental induction of DSB within a (CTG)\(\text{CAG}\) tract might lead to repeat deletions.

Large repeat deletions have been detected in the cerebellar cortex of DM1 (7, 8), DRPLA (9), HD, SBMA, SCA1 and SCA3 (1) patients. Furthermore, a few DM1 families have been shown to preferentially pass on repeat contractions of the expanded repeat to subsequent generations (2–5). It is unknown how and when these deletions arise. Perhaps these somatic or intergenerational contractions are the result of faulty DSB repair specific to these families/tissues. Further studies establishing the natural causes of DSBs in human triplet repeat tracts, and their repair by human proteins, will further our understanding of the manifestation and treatment of trinucleotide repeat-associated diseases.

Acknowledgments—We thank members of the Pearson laboratory, in particular John Cleary for technical assistance, Jayne Danska and Susanna M. Lewis for comments on the manuscript, and Andrew Paterson for statistical analysis. We thank Erin Montgomery for the electrophoretic analysis of slipped DNAs (Fig. 1C).

REFERENCES

1. Cleary, J. D., and Pearson, C. E. (2003) Cytogeten. Genome Res., in press
2. Ashizawa, T., Ansret, M., Baiget, M., Barcelo, J. M., Brunner, H., Cobo, A. M., Dallapiccola, B., Fenwick, R. G., Jr., Grandell, U., Harley, H., et al. (1994) Am. J. Hum. Genet. 54, 414–423
3. Brunner, H. G., Jansen, G., Nillesen, W., Nelen, M. R., de Die, C. E., Howeler, C. J., van Oost, B. A., Wieringa, B., Rogers, H. H., and Smeets, H. J. (1993) N. Engl. J. Med. 328, 476–480
4. O’Hoy, K. L., Tettidin, C., Mahadevan, M. S., Neville, C. E., Barcelo, J., Hunter, A. G., and Kornzelki, R. G. (1993) Science 259, 801–812
5. Shelbourne, P., Winquist, R., Kunert, E., Davies, J., Leisti, J., Thiele, H., Bachmann, H., Buxton, J., Williamson, B., and Johnson, K. (1992) Hum. Mol. Genet. 1, 467–473
6. Sergeant, N., Sablonniere, B., Schraen-Maschke, S., Ghestem, A., Maurage, C. A., Wattez, A., Vermeersch, P., and Delacourte, A. (2001) Hum. Mol. Genet. 10, 2143–2155
7. Thornton, C. A., Johnson, K., and Mosley, R. T., III (1994) Ann. Neurol. 35, 104–107
8. Ishii, S., Nishio, T., Sunohara, N., Yoshihara, T., Takemura, K., Hikiji, K., Tsuchino, S., and Sakuragawa, N. (1996) Hum. Genet. 98, 138–140
9. Aoki, M., Abe, K., Tohda, M., Kameya, T., Watanabe, M., and Itoyama, Y. (1996) Clin. Genet. 50, 199–201
10. Cleary, J. D., Nichol, K., Wang, Y. H., and Pearson, C. E. (2002) Nat. Genet. 31, 37–46
11. Panagrahi, G. B., Cleary, J. D., and Pearson, C. E. (2002) J. Biol. Chem. 277, 13993–13994
12. Samadashwily, G. M., Raca, G., and Mirkin, S. M. (1997) Nat. Genet. 17, 298–304
13. Pelletier, R., Krasilnikova, M. M., Samadashwily, G. M., Lahue, R., and Mirkin, S. M. (2003) Mol. Cell. Biol. 23, 1349–1357
14. Freudenreich, C. H., Stavenhagen, J. B., and Zakian, V. A. (1997) Trends Biochem. Sci. 22, 271–275
15. Michel, B. (2000) Trends Biochem. Sci. 25, 173–178
16. Saintigny, Y., Delacote, F., Vares, G., Petitot, F., Lambert, S., Averbeck, D., and Lopez, B. S. (2001) EMBO J. 20, 3861–3870
17. Flores, M. J., Bienere, H., Ehrlich, S. D., and Michel, B. (2001) EMBO J. 20, 619–629
18. Pearson, C. E., and Sinden, R. R. (1996) Biochemistry 35, 5041–5053
19. Richard, G. F., and Pasques, F. (2000) EMBO Rep. 1, 122–126
20. Napiarla, M., Parniewski, P., Pluciennik, A., and Wells, R. D. (2002) J. Biol. Chem. 277, 34097–34100
21. Freudenreich, C. H., Kastraw, S. M., and Zakian, V. A. (1998) Science 279, 853–856
22. Jankowski, C., Nasaar, F., and Nag, D. K. (2000) Proc. Natl. Acad. Sci. U. S. A. 97, 2134–2139
23. Richard, G. F., Dujon, B., and Haber, J. E. (1999) Mol. Gen. Genet. 261, 871–882
24. Richard, G. F., Geiellner, G. M., McMurray, C. T., and Haber, J. E. (2000) EMBO J. 19, 2381–2390
25. Honeikers, J. H. (2001) Nature 411, 366–374
26. Runger, T. M., and Kremer, R. H. (1989) EMBO J. 8, 1419–1425
27. Conley, E. C., Saunders, V. A., Jackson, V., and Saunders, J. R. (1986) Nucleic Acids Res. 14, 8919–8932
28. Pearson, C. E., and Sinden, R. R. (1997) Hum. Mol. Genet. 6, 1117–1123
29. Pearson, C. E., and Sinden, R. R. (1998) Curr. Opin. Struct. Biol. 8, 321–330
30. Conley, E. C., and Saunders, J. R. (1984) Mol. Biol. Genet. 194, 211–218
31. Roychoudhury, R., Tsu, C. F., and Wu, R. (1979) Nucleic Acids Res. 6, 1223–1233
32. Ganesh, A., North, P., and Thacker, J. (1993) Mutat. Res. 299, 251–259
33. Conley, E. C., and Saunders, J. R. (1984) Mol. Gen. Genet. 194, 211–218
34. Kamath-Loeb, A. S., Loeb, L. A., Johansson, E., Burgers, P. M., and Fry, M. (2001) J. Biol. Chem. 276, 16439–16446
35. Griffith, J. D., Comeau, L., Rosenfield, S., Stansel, R. M., Bianchi, A., Moss, H., and de Lange, T. (1999) Cell 97, 503–514
36. Cervantes, R. B., and Lundblad, V. (2002) Curr. Opin. Cell Biol. 14, 351–356
37. Chong, S. S., McCall, A. E., Cota, J., Subramony, S. H., Orr, H. T., Hughes, M. R., and Zoghbi, H. Y. (1995) Nat. Genet. 9, 344–350
38. Hashida, H., Goto, J., Kurisaki, H., Mizusawa, H., and Kanazawa, I. (1997) Ann. Neurol. 41, 505–511
39. Takano, H., Onodera, O., Takahashi, H., Igarashi, S., Yamada, M., Oyake, M., Ikeuchi, T., Kinde, R., Tanaka, H., Iwabuchi, K., and Tsuji, S. (1996) Am. J. Hum. Genet. 58, 1212–1222
40. Tanaka, F., Sobue, G., Doya, M., Itsu, Y., Yamamoto, M., Shimada, N., Yamamoto, K., Eku, S., Ihezume, Y., and Mitusa, T. (1996) J. Neurol. Sci. 135, 43–50
41. Telesius, H., Kremer, B., Goldberg, Y. P., Theilmann, J., Andrew, S. E., Zeisler, J., Adam, S., Greenberg, C., Ives, E. J., Clarke, L. A., et al. (1994) Nat. Genet. 4, 409–414