Expansion of CAG Repeats in *Escherichia coli* Is Controlled by Single-Strand DNA Exonucleases of Both Polarities

Adam Jackson, Ewa A. Okely, and David R. F. Leach
Institute of Cell Biology, School of Biological Sciences, University of Edinburgh, Edinburgh, EH9 3JR, United Kingdom

ABSTRACT The expansion of CAG·CTG repeat tracts is responsible for several neurodegenerative diseases, including Huntington disease and myotonic dystrophy. Understanding the molecular mechanism of CAG·CTG repeat tract expansion is therefore important if we are to develop medical interventions limiting expansion rates. *Escherichia coli* provides a simple and tractable model system to understand the fundamental properties of these DNA sequences, with the potential to suggest pathways that might be conserved in humans or to highlight differences in behavior that could signal the existence of human-specific factors affecting repeat array processing. We have addressed the genetics of CAG·CTG repeat expansion in *E. coli* and shown that these repeat arrays expand via an orientation-independent mechanism that contrasts with the orientation dependence of CAG·CTG repeat tract contraction. The helicase Rep contributes to the orientation dependence of repeat tract contraction and limits repeat tract expansion in both orientations. However, RuvAB-dependent fork reversal, which occurs in a *rep* mutant, is not responsible for the observed increase in expansions. The frequency of repeat tract expansion is controlled by both the 5′–3′ exonuclease RecJ and the 3′–5′ exonuclease ExoI, observations that suggest the importance of both 3′and 5′ single-strand ends in the pathway of CAG·CTG repeat tract expansion. We discuss the relevance of our results to two competing models of repeat tract expansion.

EXPANDED arrays of CAG·CTG repeats are responsible for a number of debilitating human inherited diseases, including Huntington disease and myotonic dystrophy. These diseases are characterized by genetic anticipation, the increase in severity of the disease, and decrease in age of onset in subsequent generations of affected individuals. For the last two decades, the cause of anticipation has been understood to be the propensity of long arrays of trinucleotide repeats (causative of the diseases) to expand during germ line transmission. In these human diseases, there appear to be two partially separable mechanisms of germ line instability. Short arrays of CAG·CTG repeats can expand and contract via replicative mechanisms that add or remove small numbers of repeat units. This may allow a particular array to reach a threshold length whereupon large-scale expansion can occur via replication-independent mechanisms. These replication-independent mechanisms are likely to involve DNA synthesis during the repair of DNA damage. Several excellent reviews have provided a clear picture of the current understanding of trinucleotide repeat instability (Pearson *et al.* 2005; Mirkin 2006, 2007; Kovtun and McMurray 2008; Brouwer *et al.* 2009; McMurray 2010; Budworth and McMurray 2013; Kuzminov 2013).

It has also long been established that single strands of CTG or CAG repeats can form pseudohairpin structures where CG base pairs stabilize structures containing TT or AA noncanonical base pairs, and that CTG repeat pseudohairpins are thermodynamically more stable than CAG repeat pseudohairpins (Gacy *et al.* 1995; Gacy and McMurray 1998; Hartenstine *et al.* 2000). This is thought to be because a TT base pair stacks more easily in the structure than a bulky AA base pair. Greater sensitivity to modification (Mitas 1997) and initial nuclear magnetic resonance (NMR) data (Zheng *et al.* 1996) suggested that the adenine bases in CAG repeat hairpins might be extrahelical. However, a further NMR study indicated that despite the bulkiness of the AA base pair, it could be stacked in the structure (Mariappan *et al.* 1998). Investigations using 2-aminopurine in place of adenine have confirmed that the adenine residues in the stem of the pseudohairpin are

doi: 10.1534/genetics.114.168245
Manuscript received May 28, 2014; accepted for publication July 16, 2014; published Early Online July 31, 2014.
Available freely online through the author-supported open access option.
Corresponding author: Institute of Cell Biology, School of Biological Sciences, University of Edinburgh, King's Blgs., Mayfield Rd., Edinburgh, EH9 3JR, UK.
E-mail: D.Leach@ed.ac.uk

primarily stacked, while the adenines in the loop of the pseudohairpin show considerable unstacking (Degtyareva *et al.* 2009, 2011). Despite being primarily stacked, the AA base pairs in the stem are easily destabilized and a $(CAG)_8$ sequence that is paired to a shorter template strand of DNA, as would be expected to occur in a strand-slippage structure, forms an unstructured loop (Degtyareva *et al.* 2010).

The differential thermodynamic stability of CTG and CAG repeat pseudohairpins correlates with an orientation dependence of replicative instability in model systems. In *Saccharomyces cerevisiae* and *Escherichia coli* systems, it has been reported that the two orientations of the repeat array display differential instabilities (Kang *et al.* 1995; Freudenreich *et al.* 1997; Miret *et al.* 1998; Schweitzer and Livingston 1999; Zahra *et al.* 2007). In these model systems, the orientation that places the CTG repeat on the lagging-strand template is more prone to deletions than the orientation with the CAG repeat on the lagging-strand template. Conversely it has been reported that for *S. cerevisiae* chromosomal and for *E. coli* plasmid systems, expansion occurs more frequently when the CAG repeat sequence is on the lagging-strand template. The propensity for repeat array deletions and their orientation dependence have most frequently been explained by the presence of single-strand DNA on the lagging-strand template coupled to the greater stability of CTG repeat hairpins that are likely to provide a good template for replication slippage.

The greater frequency of repeat array expansion observed when the CAG repeat sequence is located on the lagging-strand template is harder to explain because the origin of a pseudohairpin on a newly synthesized DNA strand must be envisaged. Two models have been proposed for how a pseudohairpin could be generated (see Figure 1). The first model (flap processing, Figure 1A), proposes that repeat expansion occurs during the processing of DNA ends at the sites of nicks (*e.g.*, at sites of joining Okazaki fragments, which occur on the lagging strand of the replication fork). In this model, a flap that can fold back on itself may be generated at the site of a nick and form an expansion precursor. In support of this model, the flap endonuclease FEN1 limits CAG·CTG repeat expansion in several systems (Spiro *et al.* 1999; Liu *et al.* 2004, 2009; Yang and Freudenreich 2007; Goula *et al.* 2009). The second model (fork reversal, Figure 1*B*) proposes that a pseudohairpin, formed on the lagging-strand template, causes the replication fork to pause and reverse. The reversed fork may either have a single-strand region present on the tail formed by reannealing of the newly synthesized strands because the new leading strand is longer than the new lagging strand or this tail can be processed by a 5′–3′ exonuclease to generate the same single-strand region. This region can then fold into a pseudohairpin structure on the newly synthesized leading strand that can be brought back to the template DNA by returning the reversed fork to a normal fork (Mirkin 2006, 2007; Liu *et al.* 2013).

In the flap-processing model, the preference for expansion when CAG repeats are on the lagging-stand template can be explained by the presence of the more stably folded CTG repeat sequence on the newly synthesized lagging strand that is required to fold back on itself to form the expansion precursor. This observed orientation preference for expansions is less well explained by the fork-reversal model, since the sequence that is expected to form a more stable pseudohairpin on the lagging-strand template is the CTG repeat that would be predicted to preferentially generate a CTG repeat pseudohairpin on the newly synthesized leading strand. This would predict a greater propensity to expand when the CTG repeat is the template of the lagging strand.

A clever use of zinc-finger nucleases has illuminated these transactions in human cells. Here also, a preference for deletions when CTG repeats were located on the template for the lagging strand was observed. By contrast, both expansions and deletions were observed when CAG repeats were located on the template for the lagging strand. PCR analysis was carried out following cleavage of DNA with either of the two orientations of a $(CAG{\cdot}CTG)_{102}$ repeat array by zinc-finger nucleases predicted to cleave only CTG repeat or CAG repeat pseudohairpins. The patterns of PCR fragments observed were interpreted to imply that both CTG and CAG repeat pseudohairpins were formed inside cells on both templates of the leading and lagging strands (Liu *et al.* 2010). Deletion and expansion instability of DNA containing a $(CAG{\cdot}CTG)_{45}$ repeat array was stimulated by treatment with emetine, which affects DNA synthesis of the lagging strand, leading to a conclusion that events on this strand were important (Liu *et al.* 2013). Furthermore, only oligonucleotides complementary to the lagging-strand template were able to inhibit cleavage by these pseudohairpin-directed nucleases and only these same oligonucleotides were able to inhibit emetine-induced instability (Liu *et al.* 2013). These studies revealed the importance of events occurring on the lagging strand for both deletion and expansion instability in human cells.

In summary, there is a consensus that replication-dependent instability (both deletion and expansion) is substantially affected by events occurring on the lagging strand of the replication fork. However, there are two alternative models of repeat expansion consistent with this. Either, expansion is caused by flap processing in the joining of Okazaki fragments or it is caused by lagging-strand hairpin-induced fork stalling and reversal followed by hairpin formation during the return of the reversed fork to its normal configuration. Here we confirm that the frequency of deletion of CAG·CTG repeats inserted in the chromosome of *E. coli* is orientation dependent and reveal that CAG·CTG repeat expansion is or is close to orientation independent. We show that RuvAB-dependent fork reversal is not responsible for repeat expansion in a *rep* mutant where this reaction has been characterized (Seigneur *et al.* 1998; Baharoglu *et al.* 2006). Furthermore, we demonstrate that both the single-strand exonucleases, ExoI and RecJ, participate in controlling the frequency of expansions, consistent with there being both

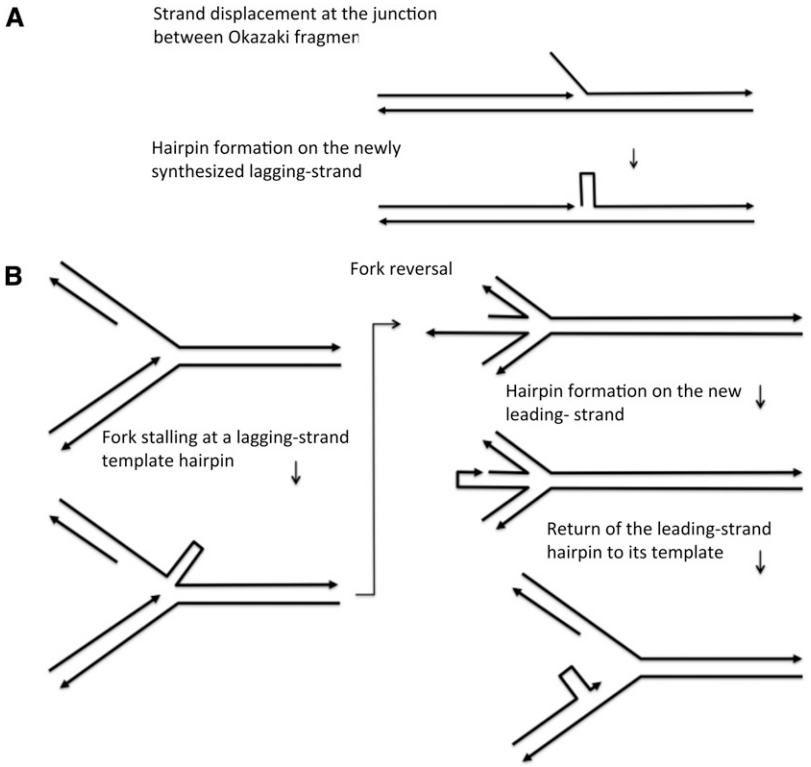

Figure 1 Models of trinucleotide repeat expansion. (A) Schematic representation depicting the flap-processing model of trinucleotide repeat expansion in which a 5′ flap is generated at the junction of Okazaki fragments. This flap then misfolds into a pseudohairpin structure that becomes incorporated into the newly synthesized strand and leads to an expansion product in the next round of DNA replication. (B) Schematic representation depicting the replication fork reversal model of trinucleotide repeat expansion in which a replication fork pauses due to the formation of a pseudohairpin structure on the template of the lagging strand. The fork then reverses and the protruding newly synthesized leading strand finds itself single stranded in the tail of the structure. This new single-stranded leading strand then folds into a pseudohairpin that remains self-annealed when the reversed fork is returned to the normal configuration and leads to an expansion product in the next round of DNA replication.

3′ and 5′ ends implicated in the formation of the expansion products. We consider that the simplest interpretation of this unexpected observation is that expansion precursors are generated during the processing of Okazaki fragments, where both 3′ and 5′ ends are present. This mechanism for replicative expansion also simply satisfies the orientation dependence observed in yeast and human cells and the involvement of FEN1 in limiting expansion observed in many studies.

Materials and Methods

Bacterial strains

E. coli strains carrying genomic CAG·CTG repeat arrays of either $(CAG)_{84}$ or $(CTG)_{95}$ integrated into the *lacZ* gene (Zahra *et al.* 2007) were used in this study (see Table 1). The trinucleotide repeat sequence present on the leading-strand template was used to designate the orientation of the array, $(CTG)_{95}$ implying 95 repeats of the sequence CTG on the leading-strand template and 95 repeats of the sequence CAG on the lagging-strand template. All mutant strains used in this study (see Table 1) were constructed using a plasmid-mediated gene replacement (PMGR) method (Merlin *et al.* 2002) or P1 transduction.

Cell growth and GeneMapper analysis

For each strain analyzed, 120 parental colonies were taken and allowed to grow to stationary phase overnight in separate LB broth cultures at 37° with shaking. Each of these cultures was then diluted, plated onto LB agar, and grown overnight at 37° to produce single colonies. Eight sibling colonies from each plate were analyzed for repeat array instability by colony PCR. Amplification of trinucleotide repeat tracts was accomplished using Ex-Test-F (TTATGCTTCCGGCTCGTATG) and Ex-Test-R (GGCGATTAAGTTGGGTAACG) primers, the former of which was labeled with the fluorescent tag 6-fam at the 5′ end. PCR products were separated by capillary electrophoresis through a polyacrylamide medium in an ABI 3730 genetic analyzer. The array size was determined by comparison to a size standard (genescan-500 LIZ). Results were analyzed using GeneMapper software.

Instability proportions

The number of expanded, deleted, and parental length arrays detected in each sibling colony was measured and used to calculate instability proportions for each of the 120 overnight cultures (960 single colony PCR reactions per strain). The mean instability for a strain was then calculated from these results and plotted. Error bars represent the standard error of the mean for the population.

Results

An elevated frequency of expansion of a CTG repeat array in a rep mutant identifies a useful substrate for studies of expansion

CAG·CTG repeat arrays have a strong preference for deletion rather than expansion in *E. coli* cells. It has therefore been difficult to assess the genetic control of expansion in this organism. It has been particularly difficult to do this in the

Table 1

Strain	Genotype	Derivation	Source
MG1655	F⁻ lambda⁻ *ilvG rfb-50 rph-1* Δ*fnr*		Blattner *et al.* (1997)
DL2639	MG1655 *lacZ*χ⁻ *lacI*^{q} ZeoRχ⁺ *lacZ*::$(CAG)_{84}$		Zahra *et al.* (2007)
DL3692	MG1655 *lacZ*χ⁻ *lacI*^{q} ZeoRχ⁺ *lacZ*::$(CAG)_{84}$ Δ*rep*	DL2639 Δ*rep* by PMGR	This work
DL4576	MG1655 *lacZ*χ⁻ *lacI*^{q} ZeoRχ⁺ *lacZ*::$(CAG)_{84}$ Δ*recJ*	DL2639 Δ*recJ* by PMGR	Julie Blyth
DL4578	MG1655 *lacZ*χ⁻ *lacI*^{q} ZeoRχ⁺ *lacZ*::$(CAG)_{84}$ Δ*exo1*	DL2639 Δ*exoI* by PMGR	Julie Blyth
DL5003	MG1655 *lacZ*χ⁻ *lacI*^{q} ZeoRχ⁺ *lacZ*::$(CAG)_{84}$ Δ*recJ* Δ*exo1*	DL4576 Δ*exoI* by PMGR	This work
DL4804	MG1655 *lacZ*χ⁻ *lacI*^{q} ZeoRχ⁺ *lacZ*::$(CAG)_{84}$ Δ*rep* Δ*recJ*	DL4576 Δ*rep* by PMGR	This work
DL4626	MG1655 *lacZ*χ⁻ *lacI*^{q} ZeoRχ⁺ *lacZ*::$(CAG)_{84}$ *rep*::Km Δ*exo1*	DL4578 *rep*::Km by P1 transduction from JJC213 (Benedicte Michel)	This work
DL4871	MG1655 *lacZ*χ⁻ *lacI*^{q} ZeoRχ⁺ *lacZ*::$(CAG)_{84}$ Δ*rep* Δ*recJ* Δ*exo1*	DL4804 Δ*exoI* by PMGR	This work
DL4487	MG1655 *lacZ*χ⁻ *lacI*^{q} ZeoRχ⁺ *lacZ*::$(CAG)_{84}$ Δ*rep* Δ*recQ*	DL2639 Δ*rep* Δ*recQ* by PMGR	This work
DL2009	MG1655 *lacZ*χ⁻ *lacI*^{q} ZeoRχ⁺ *lacZ*::$(CTG)_{95}$		Zahra *et al.* (2007)
DL2384	MG1655 *lacZ*χ⁻ *lacI*^{q} ZeoRχ⁺ *lacZ*::$(CTG)_{95}$ Δ*rep*	DL2009 Δ*rep* by PMGR	John Blackwood
DL4730	MG1655 *lacZ*χ⁻ *lacI*^{q} ZeoRχ⁺ *LacZ*::$(CTG)_{95}$ Δ*recJ*	DL2009 Δ*recJ* by PMGR	This work
DL4579	MG1655 *lacZ*χ⁻ *lacI*^{q} ZeoRχ⁺ *lacZ*::$(CTG)_{95}$ Δ*exo1*	DL2009 Δ*exoI* by PMGR	Julie Blyth
DL5004	MG1655 *lacZ*χ⁻ *lacI*^{q} ZeoRχ⁺ *lacZ*::$(CTG)_{95}$ Δ*recJ* Δ*exo1*	DL4730 Δ*exoI* by PMGR	This work
DL4803	MG1655 *lacZ*χ⁻ *lacI*^{q} ZeoRχ⁺ *lacZ*::$(CTG)_{95}$ Δ*rep* Δ*recJ*	DL4730 Δ*rep* by PMGR	This work
DL4627	MG1655 *lacZ*χ⁻ *lacI*^{q} ZeoRχ⁺ *lacZ*::$(CTG)_{95}$ *rep*::Km Δ*exo1*	DL4579 *rep*::Km by P1 transduction from JJC213 (Benedicte Michel)	This work
DL4911	MG1655 *lacZ*χ⁻ *lacI*^{q} ZeoRχ⁺ *lacZ*::$(CTG)_{95}$ Δ*rep* Δ*recJ* Δ*exo1*	DL4803 Δ*exoI* by PMGR	This work
DL4438	MG1655 *lacZ*χ⁻ *lacI*^{q} ZeoRχ⁺ *lacZ*::$(CTG)_{95}$ Δ*rep* Δ*recQ*	DL2009 Δ*rep* Δ*recQ* by PMGR	This work
DL4845	MG1655 *lacZ*χ⁻ *lacI*^{q} ZeoRχ⁺ *lacZ*::$(CTG)_{95}$ Δ*rep* Δ*recQ* Δ*recJ*	DL4803 Δ*recQ* by PMGR	This work
DL4950	MG1655 *lacZ*χ⁻ *lacI*^{q} ZeoRχ⁺ *lacZ*::$(CTG)_{95}$ Δ*rep* Δ*ruvA*::Cm	DL2384 *ruvA*::Cm by P1 transduction from JJC3148 (Benedicte Michel)	This work

chromosome of *E. coli* since no bulk assay for expansion is available and expansion frequencies have to be obtained by individual colony PCR arrays. Despite these drawbacks, the single colony assay that we have developed, which relies on capillary electrophoresis and GeneMapper analysis of the major product amplified in a PCR reaction, is quantitative and reliable. Nevertheless, the very low frequency of expansion events observed in a wild-type host (~1%) makes statistically significant analyses of altered expansion proportions in mutant strains difficult (even when analyzing a total of 960 individual colonies per strain). The time and cost of increasing the scale of this analysis becomes prohibitive. We therefore sought to identify a mutant where the repeat arrays were somewhat destabilized and included a higher baseline of expansion events to analyze mutants with altered expansion proportions.

A *rep* mutation, which slows down DNA replication (Lane and Denhardt 1974, 1975; Colasanti and Denhardt 1987), had previously been shown to destabilize dinucleotide repeat arrays (Morel *et al.* 1998). Here we show that this is also the case in our assay for CAG·CTG repeat instability. As can be seen in Figure 2, the proportion of repeat expansion was increased from ~1 to 4% in both orientations of the CAG·CTG arrays studied. It is interesting to note that the orientation dependence of repeat array contraction was substantially reduced in this mutant, through an increase in CTG leading-strand repeat array deletions. Despite this destabilization, the CTG leading-strand repeat array in a *rep* mutant can be seen to be a useful starting point from which to study repeat array expansion in *E. coli* as a ratio of ~2:1 of contraction to expansion and an elevated frequency of expansion provides a useful background to investigate mutants enhancing or reducing the frequency of expansion events.

CTG repeat array expansion in a rep mutant is not caused by RuvAB-dependent replication fork reversal

In addition to slowing down DNA replication, a *rep* mutant is characterized by RuvAB-dependent replication fork reversal (Seigneur *et al.* 1998; Baharoglu *et al.* 2006). We therefore realized that we could test whether the elevated level of expansion observed in a *rep* mutant was caused by replication fork reversal as might be anticipated according to the model of Mirkin (2006, 2007) and supported by Liu *et al.* (2013). However, the expansion proportion was not reduced in a *rep ruvA* strain, indicating that the expansion events observed in a *rep* mutant were not caused by RuvAB-dependent replication fork reversal (Figure 3). In fact both the frequencies of expansion and contraction observed in a *rep ruvA* mutant were elevated relative to the *rep* single mutant. This suggests that replication fork reversal may in fact interfere with (rather than promote) CAG·CTG instability.

CTG repeat array expansion in a rep mutant is controlled by RecJ and ExoI, two single-strand exonucleases with opposite polarities of DNA degradation

Given that RuvAB-dependent replication fork reversal was not responsible for repeat expansion in a *rep* mutant, we continued to look for other activities that might promote or inhibit repeat expansion. To our surprise, deletion of either the gene encoding the single-strand-specific DNA exonuclease RecJ or the single-strand-specific DNA exonuclease ExoI caused an increase in the frequency of expansion events

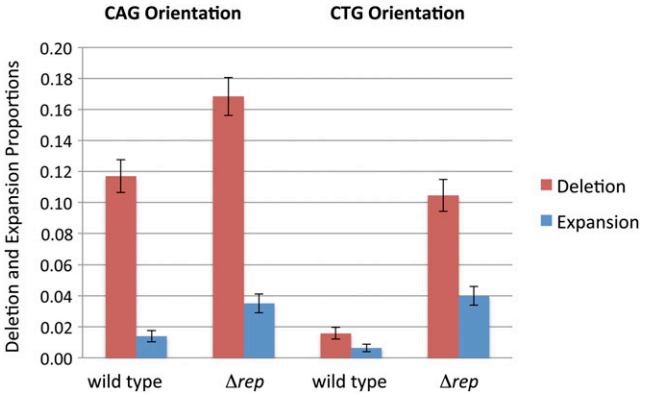

Figure 2 The behavior of a CTG leading-stand template repeat array in a *rep* mutant makes this strain a good starting point to investigate expansion events in the *E. coli* chromosome. Comparison of the expansion and deletion proportions of CAG leading-strand template and CTG leading-strand template repeat arrays in wild-type and *rep* mutant strains.

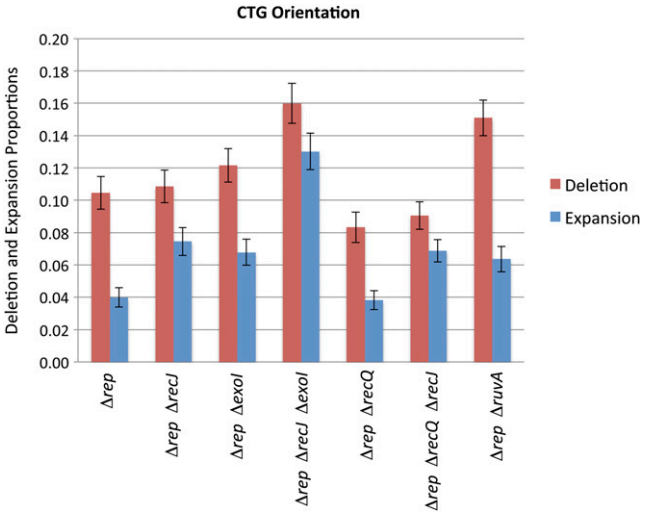

Figure 3 The effects of *recJ*, *exoI*, *recQ*, and *ruvA* mutations on instability of a CTG leading-stand template repeat array in a *rep* mutant. Comparison of the expansion and deletion proportions of CTG leading-strand template repeat arrays in *rep* mutant with *rep recJ*, *rep recQ*, *rep recQ recJ*, *rep exoI*, *rep recJ exoI*, and *rep ruvA* mutants.

(Figure 3). This was surprising because these single-strand-specific exonucleases have opposite polarities (RecJ is a 5′–3′ exonuclease and ExoI, a 3′–5′ exonuclease). These data argue that the precursor to expansion is indeed a single-strand of DNA but that it can have either a 3′ or 5′ end. Furthermore a *rep recJ exoI* triple mutant showed an enhanced expansion frequency over either a *rep recJ* or a *rep exoI* double mutant (Figure 3), confirming that both 3′ and 5′ ended substrates are available to degradation. Since the RecQ helicase is known to generate single strands that are degraded by RecJ in both recombination (see (Michel and Leach 2012) and replication (Courcelle *et al.* 2003; Courcelle and Hanawalt 1999, 2001), we sought to determine the role of RecQ in promoting or preventing CAG·CTG repeat expansion. However, it was clear that the *rep recQ* mutant behaved similarly to the *rep* mutant (Figure 3). Furthermore, the destabilizing effect of *recJ* was apparent even in the absence of Rep and RecQ in a *rep recJ recQ* triple mutant (Figure 3). Altogether, we conclude that RecJ does not inhibit expansion through concerted action with RecQ. Nor does RecQ promote expansion in the presence or absence of RecJ. Individually, the *recJ*, *exoI*, and *recQ* mutations had no significant impact on deletion frequencies, confirming that RecJ and ExoI have an impact specifically on a pathway controlling expansions. In the *rep* mutant context, the *recJ exoI* double mutant did show a modest (50%) increase in the frequency of deletions (Figure 3). However, this could be caused indirectly by the increase in frequency of expansions coupled to the known elevation of deletion frequencies in longer CAG·CTG repeat arrays (Zahra *et al.* 2007).

RecJ and ExoI control repeat expansion in both orientations of the repeat array independently of the presence of Rep

Although the CTG leading-strand template array in a *rep* mutant provided a useful strain to investigate the genetic control of expansion, it was important to confirm that the behavior observed was not restricted to this genetic background or repeat orientation. We therefore investigated the effects of *recJ* and *exoI* mutations on the other orientation of the repeat array and in the absence of the *rep* mutation. As can be seen in Figure 4, *recJ* and *exoI* deletions in a *rep* mutant also elevated expansion frequencies in the CAG leading-stand template orientation. In wild-type cells, *recJ* and *exoI* mutations had little effect individually, but the *recJ exoI* double mutant showed a clear increase in CAG·CTG repeat expansion in both orientations of the array. The frequency of deletions was also elevated in the double mutants but to a lesser extent.

The proportions of repeat expansion are independent of the orientations of the repeat arrays and distributions of the sizes of repeat array expansions decrease exponentially in both orientations

Unexpectedly, given the results of previous studies of expansion and the marked orientation dependence of deletion frequencies (Kang *et al.* 1995; Freudenreich *et al.* 1997; Miret *et al.* 1998; Schweitzer and Livingston 1999; Zahra *et al.* 2007), the frequencies of expansion events were remarkably similar in the two orientations of the repeat array (Figure 4). The only genetic background where we observed an apparent orientation dependence of expansion was the *rep recJ exoI* triple mutant. However, the very high frequency of deletions observed in the CAG leading-strand template orientation was likely to be masking an elevated frequency of expansion (if the repeat array was deleted, it did not have the opportunity to expand).

We wondered whether any of the mutant strains investigated showed evidence of an increase in the size of the expansion products. However, analysis of individual mutants showed no significant differences in the expansion or deletion

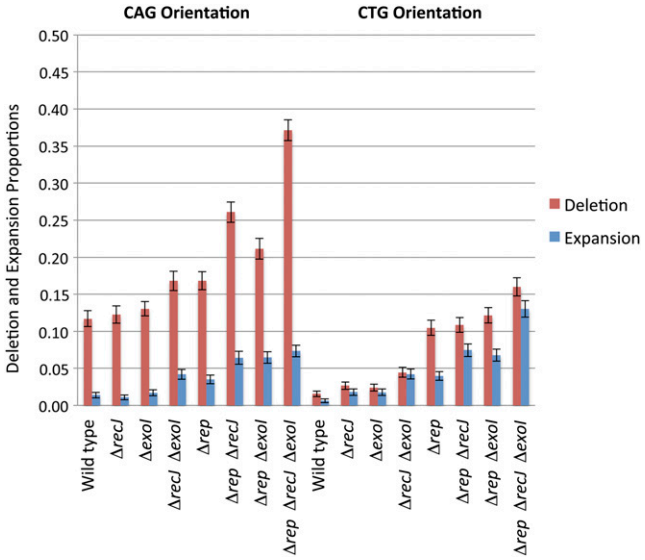

Figure 4 Control of CAG·CTG repeat expansion by *recJ* and *exoI* occurs in both orientations of the repeat array and in the presence and absence of *rep*. Comparison of the expansion and deletion proportions of CAG leading-strand template and CTG leading-strand template repeat arrays in wild-type, *recJ*, *exoI*, *recJ exoI*, and *rep* versions of these strains.

product sizes in any strain. We therefore decided to pool all the data on expansion and deletion sizes across all the mutants investigated to obtain sufficiently large datasets to clearly visualize distributions, on the explicit assumption that expansion and deletion sizes were not affected by the mutations studied. As can be seen in Figure 5, the distributions of expansion sizes and those of deletion sizes were markedly different. As previously described in several studies including for the *E. coli* chromosome (Zahra *et al.* 2007), the orientation with the CAG repeats on the leading-strand template showed a strong preference for large deletions, while the orientation with the CTG repeats on the leading-strand template showed a nearly flat distribution with no strong preference for any particular deletion length (Figure 5A). By contrast, expansions were highly skewed toward short lengths and the distribution of lengths decreased exponentially with size in both orientations of the repeat array (Figure 5B).

Discussion

Single-strand-specific exonucleases with opposite polarities of DNA degradation control the expansion of CAG·CTG repeat arrays in E. coli

We have shown here that the 5′–3′ exonuclease RecJ and the 3′–5′ exonuclease ExoI operate to reduce the frequency of CAG·CTG repeat expansion in the *E. coli* chromosome. We were surprised by this result as it implied that both 3′-ended single strands and 5′-ended single strands are precursors to expansion events in this organism. Furthermore, the RecQ helicase does not operate with RecJ in the control of expansion events, suggesting that the single-strand DNA is not generated from a double-strand precursor accessible to the RecQJ combination known to process replication forks (Courcelle *et al.* 2003; Courcelle and Hanawalt 1999, 2001) and gaps repaired by recombination (see Michel and Leach 2012). We therefore considered where 3′ and 5′ single-strand ends might be present as precursors for expansion. The 3′ ends are present at the growing tips of newly synthesized DNA strands on both the leading and lagging strands of the fork, but these would normally be complexed with the replicative polymerase (PolIII). Two reactions are expected to liberate these 3′ ends for processing by other enzymes: the termination of Okazaki fragment synthesis and replication fork reversal. Interestingly, both of these reactions also involve the presence of 5′ ends. In yeast and in several mammalian systems (though not all), repeat expansion is stimulated in *fen1* mutant strains (Spiro *et al.* 1999; Liu *et al.* 2004, 2009, 2010; Yang and Freudenreich 2007; Goula *et al.* 2009). FEN1 is a nuclease that can cleave 5′ flaps that are formed during the joining of Okazaki fragments. In *E. coli*, flap processing is thought to be carried out by DNA polymerase I encoded by the *polA* gene. Dinucleotide repeat array expansions (Morel *et al.* 1998) as well as "plus" frameshifts and duplications (Nagata *et al.* 2002) occur more frequently in *polA* mutants that are defective in the 5′–3′ exonuclease activity of the enzyme. We propose here that flap processing is responsible for CAG·CTG repeat expansion in *E. coli* and that RecJ and ExoI provide backup functions for the removal of 5′ and 3′ flaps. Further experiments are required to elucidate the interactions between the exonuclease activities of DNA polymerase I, RecJ, and ExoI in Okazaki fragment maturation and the control of CAG·CTG repeat expansion.

RuvAB dependent replication fork reversal in a rep mutant does not promote expansion of CAG·CTG repeat arrays in E. coli

One of the best experimental systems for the study of replication fork reversal in *E. coli* utilizes the *rep* mutant (Seigneur *et al.* 1998; Michel and Leach 2012). This was the system in which replication fork reversal was first discovered (Seigneur *et al.* 1998) when it was shown that RuvAB could generate DNA double-strand ends that were processed by the DNA double-strand exonuclease RecBCD. Our observation that RuvAB did not promote CAG·CTG repeat expansion in a *rep* mutant (where it is known to be required for replication fork reversal) (Baharoglu *et al.* 2006) is not consistent with this reaction being implicated. It has also been proposed that replication forks can be reversed following UV irradiation, whereupon the reversed fork is processed by RecQ and RecJ (Courcelle *et al.* 2003). However, the lack of involvement of RecQ in CAG·CTG repeat expansion that we have observed argues against this pathway of fork reversal being responsible.

In fact, we determined that the frequencies of both CAG·CTG repeat expansions and deletions were increased modestly in a *ruvAB rep* mutant relative to a *rep* mutant. This is consistent with a role of replication fork reversal in

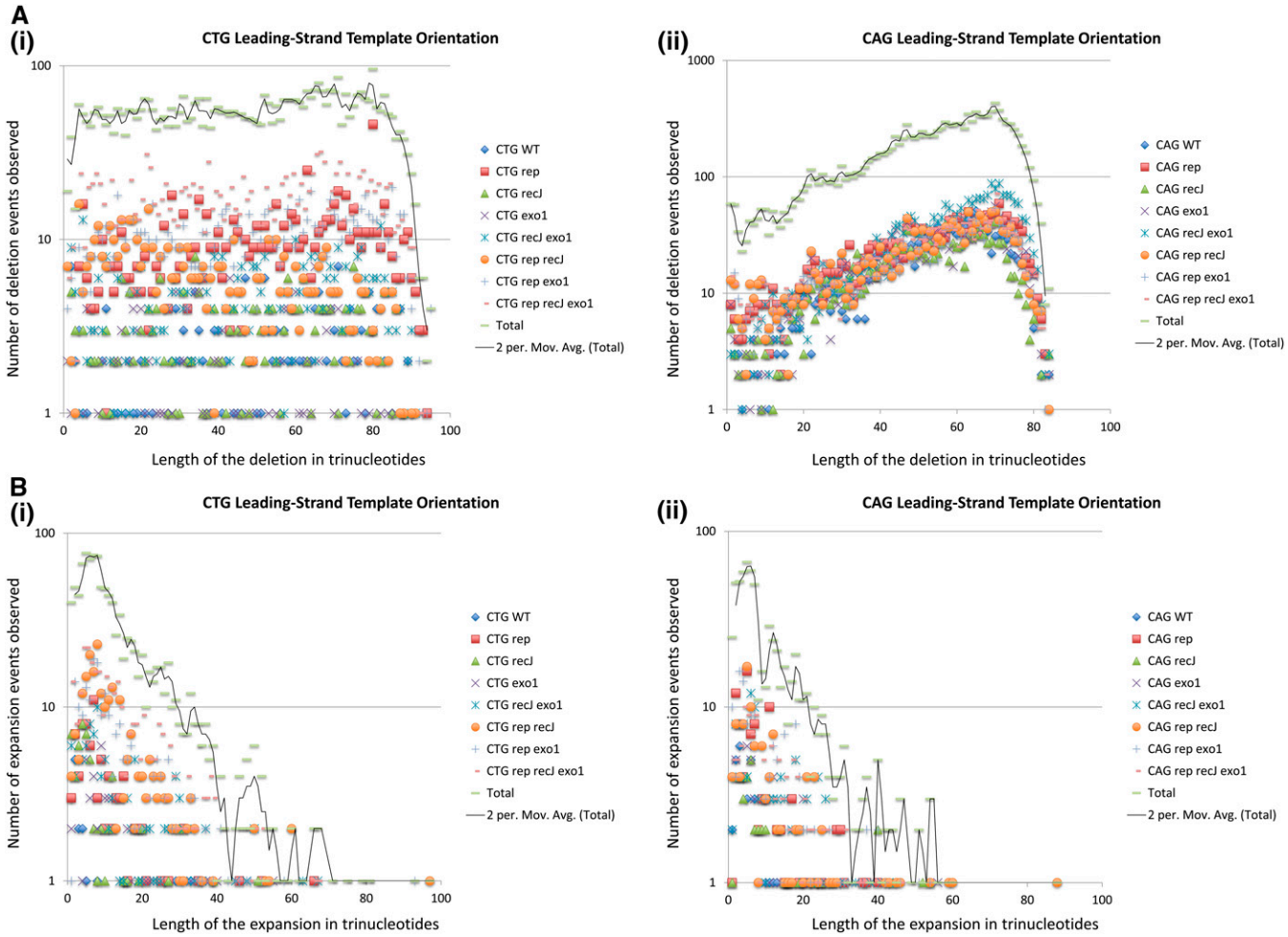

Figure 5 Distributions of deletion and expansion lengths of CTG leading-strand template and CAG leading-strand template repeat arrays. Data from all strains studied are plotted individually as are data obtained by summing across all strains. The trend lines represent moving averages with a period of 2 in the total events. (A) Distribution of deletion sizes in both orientations of the repeat array. (i) The array with the CTG repeat on the leading-strand template has a nearly flat distribution of deletion lengths, suggesting that many different deletion lengths are approximately equally probable. (ii) The array with the CAG repeat on the leading-strand template shows a very skewed distribution of deletion sizes with large deletions predominating. This is consistent with the formation of thermodynamically more stable pseudohairpins in the CTG lagging-strand template. (B) Distribution of expansion sizes in the (i) CTG leading-strand template orientation and the (ii) CAG leading-strand orientation. Here it can be seen that there is a sharp exponential decrease in the frequency of expansion products of increasing size. This is consistent with no influence of stable pseudohairpins on the size of expansion products in either orientation. The slightly larger size of the expansion events in the CTG leading-strand orientation is interesting and may reflect the nature of the structural unit of expansion.

the prevention CAG·CTG repeat instability. This may be because mismatches between the template and newly synthesized DNA stands can be removed by replication fork reversal (via degradation of the new strands and via reannealing of the parental template strands). This fits also with the conclusion that in yeast, activities that promote replication fork reversal inhibit repeat expansion (Kerrest *et al.* 2009).

Expansion frequencies are independent of repeat array orientation and the size distribution of expansion products decays exponentially

We have confirmed a strong orientation dependence in the frequency of CAG·CTG repeat deletion events and in the sizes of the deletion products. The CAG leading-strand template orientation of the repeat array (where the CTG repeat array is present on the template for the lagging strand) showed an elevated frequency of deletion events and the distribution of deletion sizes was strongly skewed toward large deletions. Surprisingly, however, we observed that the frequencies of CAG·CTG repeat expansion were remarkably similar to each other in the two orientations of the repeat array. This differs from what has been observed in *E. coli* plasmids and in the chromosomes of yeast and human cells (Kang *et al.* 1995; Freudenreich *et al.* 1997; Miret *et al.* 1998; Schweitzer and Livingston 1999; Zahra *et al.* 2007; Liu *et al.* 2009, 2010, 2012, 2013). The lack of orientation dependence seen here for the frequency of CAG·CTG repeat expansion in the *E. coli* chromosome suggests that secondary structure formation in

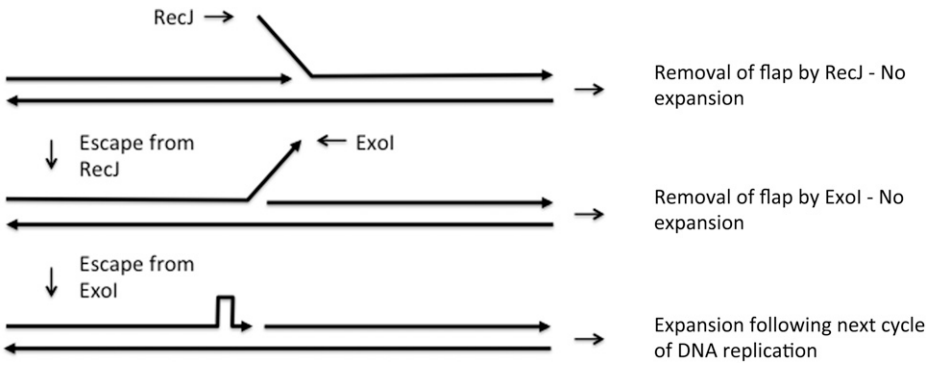

Figure 6 Extension of the flap-processing model to allow both 3′- and 5′-end processing. Since we have observed the control of expansion frequencies by exonucleases of opposite polarities, we propose that the precursor for expansion may be allowed to interchange between a 5′ overhang and a 3′ overhang at the site of maturation of Okazaki fragments. In the presence of RecJ, the 5′ overhang can be digested and in the presence of ExoI, the 3′ overhang can be removed. When both of these DNA exonucleases are absent, a pseudohairpin can form on the Okazaki fragment, which will lead to a repeat array expansion after the next DNA replication cycle.

the pseudohairpin expansion precursors is not rate limiting for these events. This is consistent with our observation that the sizes of the CAG·CTG repeat expansion products were small and their frequency decreased exponentially with size in both orientations of the repeat array. The slopes of the exponential decrease in the size distribution of expansion products were approximately the same in both orientations of the repeat array, suggesting that there was no preferential stabilization of large hairpins in one orientation of the DNA sequence. The slightly larger size of the expansion events in the CTG leading-strand orientation that we have detected is interesting and may reflect the nature of the structural unit of expansion.

Conclusions

We have used a nonselective system to investigate CAG·CTG repeat expansion in the *E. coli* chromosome. We have shown that single-strand-specific exonucleases with opposite polarities of DNA degradation, RecJ and ExoI, limit the frequency of expansion events. This suggests the operation of an expansion pathway that involves both 5′ and 3′ DNA single-strand substrates. Furthermore, the control of expansion by RecJ is not mediated by the coordinated action of RecJ with RecQ, and RuvAB does not promote expansion in a *rep* mutant (where it is known to mediate replication fork reversal). Finally, the small size of the expansion loops and the orientation independence of expansion frequencies argue against the importance of differential stabilities of pseudohairpins being important for expansion in an *E. coli* chromosomal context. These data are consistent with expansion being promoted by flap processing during Okazaki fragment completion as predicted by the involvement of FEN1 in the control of expansion in yeast and mammalian cells (Spiro *et al.* 1999; Liu *et al.* 2004, 2009, 2010; Yang and Freudenreich 2007; Goula *et al.* 2009) and the strong effects of treatments affecting lagging-strand synthesis in human cells (Liu *et al.* 2010, 2013). A simple extension of the model of flap processing to allow the interconversion of 5′ and 3′ flaps would enable exonucleases of opposite polarities to control expansion frequency as shown in Figure 6. In rep^+ cells, inactivation of both RecJ and ExoI is required to observe an elevated proportion of expansion events arguing that either of these nucleases can digest unprocessed flaps. However, in a *rep* mutant our data argue that some flaps can escape processing by one or the other exonuclease, explaining the elevation of expansion in the presence of both nucleases, and in single as well as double mutants. The frequency of expansion events in *E. coli* is relatively low compared to eukaryotic cells and we wonder whether this may be because of the presence of more active exonucleases in this organism that contribute to a nonrepetitive and streamlined genome. Our work suggests that studies should be carried out to determine whether expression of bacterial DNA single-strand exonucleases in mammalian cells could be used to limit the frequencies of somatic expansion events.

Acknowledgments

We thank Elise Darmon for critical reading of the manuscript. A.J. was supported by a studentship from the Medical Research Council (MRC) and the research was supported by an MRC programme grant to D.L.

Literature Cited

Baharoglu, Z., M. Petranovic, M. J. Flores, and B. Michel, 2006 RuvAB is essential for replication forks reversal in certain replication mutants. EMBO J. 25: 596–604.

Blattner, F. R., G. Plunkett, 3rd, C. A. Bloch, N. T. Perna, V. Burland *et al.*, 1997 The complete genome sequence of Escherichia coli K-12. Science 277: 1453–1462.

Brouwer, J. R., R. Willemsen and B. A. Oostra, 2009 Microsatellite repeat instability and neurological disease. BioEssays 31: 71–83.

Budworth, H., and C. T. McMurray, 2013 A brief history of triplet repeat diseases. Methods Mol. Biol. 1010: 3–17.

Colasanti, J., and D. T. Denhardt, 1987 The Escherichia coli rep mutation. X. Consequences of increased and decreased Rep protein levels. Mol. Gen. Genet. 209: 382–390.

Courcelle, J., and P. C. Hanawalt, 1999 RecQ and RecJ process blocked replication forks prior to the resumption of replication in UV-irradiated Escherichia coli. Mol. Gen. Genet. 262: 543–551.

Courcelle, J., and P. C. Hanawalt, 2001 Participation of recombination proteins in rescue of arrested replication forks in UV-irradiated Escherichia coli need not involve recombination. Proc. Natl. Acad. Sci. USA 98: 8196–8202.

Courcelle, J., J. R. Donaldson, K. H. Chow, and C. T. Courcelle, 2003 DNA damage-induced replication fork regression and processing in Escherichia coli. Science 299: 1064–1067.

Degtyareva, N. N., M. J. Reddish, B. Sengupta, and J. T. Petty, 2009 Structural studies of a trinucleotide repeat sequence using 2-aminopurine. Biochemistry 48: 2340–2346.

Degtyareva, N. N., C. A. Barber, B. Sengupta, and J. T. Petty, 2010 Context dependence of trinucleotide repeat structures. Biochemistry 49: 3024–3030.

Degtyareva, N. N., C. A. Barber, M. J. Reddish, and J. T. Petty, 2011 Sequence length dictates repeated CAG folding in three-way junctions. Biochemistry 50: 458–465.

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

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

Gacy, A. M., and C. T. McMurray, 1998 Influence of hairpins on template reannealing at trinucleotide repeat duplexes: a model for slipped DNA. Biochemistry 37: 9426–9434.

Goula, A. V., B. R. Berquist, D. M. Wilson, 3rd, V. C. Wheeler, Y. Trottier *et al.*, 2009 Stoichiometry of base excision repair proteins correlates with increased somatic CAG instability in striatum over cerebellum in Huntington's disease transgenic mice. PLoS Genet. 5: e1000749.

Hartenstine, M. J., M. F. Goodman, and J. Petruska, 2000 Base stacking and even/odd behavior of hairpin loops in DNA triplet repeat slippage and expansion with DNA polymerase. J. Biol. Chem. 275: 18382–18390.

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

Kerrest, A., R. P. Anand, R. Sundararajan, R. Bermejo, G. Liberi *et al.*, 2009 SRS2 and SGS1 prevent chromosomal breaks and stabilize triplet repeats by restraining recombination. Nat. Struct. Mol. Biol. 16: 159–167.

Kovtun, I. V., and C. T. McMurray, 2008 Features of trinucleotide repeat instability in vivo. Cell Res. 18: 198–213.

Kuzminov, A., 2013 Inhibition of DNA synthesis facilitates expansion of low-complexity repeats: Is strand slippage stimulated by transient local depletion of specific dNTPs? BioEssays 35: 306–313.

Lane, H. E., and D. T. Denhardt, 1974 The rep mutation. III. Altered structure of the replicating Escherichia coli chromosome. J. Bacteriol. 120: 805–814.

Lane, H. E., and D. T. Denhardt, 1975 The rep mutation. IV. Slower movement of replication forks in Escherichia coli rep strains. J. Mol. Biol. 97: 99–112.

Liu, G., and M. Leffak, 2012 Instability of (CTG)n*(CAG)n trinucleotide repeats and DNA synthesis. Cell Biosci. 2: 7.

Liu, G., X. Chen, J. J. Bissler, R. R. Sinden, and M. Leffak, 2010 Replication-dependent instability at (CTG) x (CAG) repeat hairpins in human cells. Nat. Chem. Biol. 6: 652–659.

Liu, G., X. Chen, and M. Leffak, 2013 Oligodeoxynucleotide binding to (CTG). (CAG) microsatellite repeats inhibits replication fork stalling, hairpin formation, and genome instability. Mol. Cell. Biol. 33: 571–581.

Liu, Y., H. Zhang, J. Veeraraghavan, R. A. Bambara, and C. H. Freudenreich, 2004 Saccharomyces cerevisiae flap endonuclease 1 uses flap equilibration to maintain triplet repeat stability. Mol. Cell. Biol. 24: 4049–4064.

Liu, Y., R. Prasad, W. A. Beard, E. W. Hou, J. K. Horton *et al.*, 2009 Coordination between polymerase beta and FEN1 can modulate CAG repeat expansion. J. Biol. Chem. 284: 28352–28366.

Mariappan, S. V., L. A. Silks, 3rd, X. Chen, P. A. Springer, R. Wu *et al.*, 1998 Solution structures of the Huntington's disease DNA triplets, (CAG)n. J. Biomol. Struct. Dyn. 15: 723–744.

McMurray, C. T., 2010 Mechanisms of trinucleotide repeat instability during human development. Nat. Rev. Genet. 11: 786–799.

Merlin, C., S. McAteer, and M. Masters, 2002 Tools for characterization of Escherichia coli genes of unknown function. J. Bacteriol. 184: 4573–4581.

Michel, B., and D. R. Leach, 2012 Homologous recombination: enzymes and pathways, pp. 1–46 in *EcoSal Plus*, edited by S. T. Lovett and A. Kuzminov. ASM Press, Washington, DC.

Miret, J. J., L. Pessoa-Brandao, and R. S. Lahue, 1998 Orientation-dependent and sequence-specific expansions of CTG/CAG trinucleotide repeats in Saccharomyces cerevisiae. Proc. Natl. Acad. Sci. USA 95: 12438–12443.

Mirkin, S. M., 2006 DNA structures, repeat expansions and human hereditary disorders. Curr. Opin. Struct. Biol. 16: 351–358.

Mirkin, S. M., 2007 Expandable DNA repeats and human disease. Nature 447: 932–940.

Mitas, M., 1997 Trinucleotide repeats associated with human disease. Nucleic Acids Res. 25: 2245–2254.

Morel, P., C. Reverdy, B. Michel, S. D. Ehrlich, and E. Cassuto, 1998 The role of SOS and flap processing in microsatellite instability in Escherichia coli. Proc. Natl. Acad. Sci. USA 95: 10003–10008.

Nagata, Y., K. Mashimo, M. Kawata, and K. Yamamoto, 2002 The roles of Klenow processing and flap processing activities of DNA polymerase I in chromosome instability in Escherichia coli K12 strains. Genetics 160: 13–23.

Pearson, C. E., K. Nichol Edamura, and J. D. Cleary, 2005 Repeat instability: mechanisms of dynamic mutations. Nat. Rev. Genet. 6: 729–742.

Schweitzer, J. K., and D. M. Livingston, 1999 The effect of DNA replication mutations on CAG tract stability in yeast. Genetics 152: 953–963.

Seigneur, M., V. Bidnenko, S. D. Ehrlich, and B. Michel, 1998 RuvAB acts at arrested replication forks. Cell 95: 419–430.

Spiro, C., R. Pelletier, M. L. Rolfsmeier, M. J. Dixon, R. S. Lahue *et al.*, 1999 Inhibition of FEN-1 processing by DNA secondary structure at trinucleotide repeats. Mol. Cell 4: 1079–1085.

Yang, J., and C. H. Freudenreich, 2007 Haploinsufficiency of yeast FEN1 causes instability of expanded CAG/CTG tracts in a length-dependent manner. Gene 393: 110–115.

Zahra, R., J. K. Blackwood, J. Sales, and D. R. Leach, 2007 Proofreading and secondary structure processing determine the orientation dependence of CAG × CTG trinucleotide repeat instability in Escherichia coli. Genetics 176: 27–41.

Zheng, M., X. Huang, G. K. Smith, X. Yang, and X. Gao, 1996 Genetically unstable CXG repeats are structurally dynamic and have a high propensity for folding. An NMR and UV spectroscopic study. J. Mol. Biol. 264: 323–336.

Communicating editor: S. Sandler