Published in final edited form as:

Mutat Res. 2009 February 10; 661(1-2): 71–77. doi:10.1016/j.mrfmmm.2008.11.003.

***E. coli* mismatch repair acts downstream of replication fork stalling to stabilize the expanded (GAA·TTC)_n sequence**

Rebecka L. Bourn¹, Paul M. Rindler¹, Laura M. Pollard¹, and Sanjay I. Bidichandani^{1,2}

¹Department of Biochemistry & Molecular Biology, University of Oklahoma Health Sciences Center, Oklahoma City, OK 73104, USA

²Department of Pediatrics, University of Oklahoma Health Sciences Center, Oklahoma City, OK 73104, USA

Abstract

Expanded triplet repeat sequences are known to cause at least 16 inherited neuromuscular diseases. In addition to short length changes, expanded triplet repeat tracts frequently undergo large changes, often amounting to hundreds of base-pairs. Such changes might occur when template or primer slipping creates insertion/deletion loops (IDLs), which are normally repaired by the mismatch repair system (MMR). However, in prokaryotes and eukaryotes, MMR promotes large changes in the length of (CTG·CAG)_n sequences, the motif most commonly associated with human disease. We tested the effect of MMR on instability of the expanded (GAA·TTC)_n sequence, which causes Friedreich ataxia, by comparing repeat instability in wild-type and MMR-deficient strains of *E. coli*. As expected, the prevalence of small mutations increased in the MMR-deficient strains. However, the prevalence of large contractions increased in the MMR mutants specifically when GAA was the lagging strand template, the orientation in which replication fork stalling is known to occur. After hydroxyurea-induced stalling, both orientations of replication showed significantly more large contractions in MMR mutants than in the wild-type, suggesting that fork stalling may be responsible for the large contractions. Deficiency of MMR promoted large contractions independently of RecA status, a known determinant of (GAA·TTC)_n instability. These data suggest that two independent mechanisms act in response to replication stalling to prevent instability of the (GAA·TTC)_n sequence in *E. coli*, when GAA serves as the lagging strand template: one that is dependent on RecA-mediated restart of stalled forks, and another that is dependent on MMR-mediated repair of IDLs. While MMR destabilizes the (CTG·CAG)_n sequence, it is involved in stabilization of the (GAA·TTC)_n sequence. The role of MMR in triplet repeat instability therefore depends on the repeat sequence and the orientation of replication.

Keywords

mismatch repair; triplet repeat instability; GAA triplet repeat; replication fork stalling

I. INTRODUCTION

The mismatch repair (MMR) system is a highly conserved pathway responsible for correcting base-base mismatches arising from polymerase errors and insertion/deletion loops (IDLs)

Address all correspondence and requests for reprints to: Sanjay I. Bidichandani, MBBS, PhD, 975 NE 10th St., BRC458, Oklahoma City, OK 73104, USA, Tel: (405) 271-1360, Fax: (405) 271-3910, Email: Sanjay-Bidichandani@ouhsc.edu.

CONFLICT OF INTEREST STATEMENT

The authors declare that there are no conflicts of interest.

arising from misalignment of microsatellite repeat sequences [1]. The IDLs repaired by MMR can range from a single nucleotide to 21 nucleotides or more [2,3]. Deficiency of the MMR system results in a mutator phenotype characterized by microsatellite instability and cancer [4]. The MMR process involves three steps: mismatch recognition; removal of the mispaired nucleotide(s), including several adjacent nucleotides; and resynthesis of the excised strand. In *E. coli*, this process is carried out shortly after DNA replication by the three MMR proteins, MutS, MutL, and MutH, as well as other accessory proteins [5]. A MutS homodimer recognizes and binds base-base mispairs or IDLs of up to eight nucleotides. Upon detection of a lesion, MutS forms a complex with MutL, thereby activating the endonuclease activity of MutH, which targets repair to the newly synthesized strand by preferentially nicking the unmethylated strand at hemi-methylated GATC sites. The bases from the nick to the mispaired base(s) are excised, and the strand is resynthesized, thus repairing the lesion.

In tracts of simple repeat sequences, including triplet repeats, IDLs are prone to form when DNA is either newly synthesized or denatured and reannealed, as happens during common DNA metabolic transactions, such as replication, repair, and transcription. Expanded triplet repeat tracts in particular are likely to adopt stable non-B DNA structures, which may further promote the development of mismatched templates that are substrates for MMR.

Expanded triplet repeat sequences located within, or in the vicinity of, genes are known to cause 16 different neuromuscular diseases [6]. These expanded triplet repeat tracts, involving (CTG·CAG)_n, (CGG·CCG)_n, or (GAA·TTC)_n sequences, are known to be unstable *in vivo* in somatic cells and during intergenerational transmission [7,8]. The propensity for further expansion upon germline transmission of triplet repeats in several genes results in progressive worsening of the phenotype in successive generations, a phenomenon termed genetic anticipation. Moreover, somatic instability occurring throughout the life of a patient in specific tissues is thought to be an important determinant of the progressive and tissue-specific pathology of triplet repeat diseases. For instance, in Huntington disease patients, the expanded (CTG·CAG)_n tract in the gene for huntingtin progressively undergoes further large expansions specifically in the striatum [9], the primary affected tissue in Huntington disease, and prevention of expansions delays the onset of disease in a transgenic mouse model [10]. Similarly, the expanded (GAA·TTC)_n sequence in the gene for frataxin, which causes Friedreich ataxia, undergoes further large expansions specifically in the primary site of pathology, the dorsal root ganglia, in patients and a transgenic mouse model for (GAA·TTC)_n instability [11,12].

(CTG·CAG)_n sequences propagated in *E. coli* undergo both small and large changes in tract length. Whereas the *E. coli* MMR system prevents small length changes, it increases the frequency of large length changes [13-15]. Indeed, MMR proteins in mammals are required for somatic and intergenerational instability of the (CTG·CAG)_n sequence [10,16-19]. In mammals, instability of the (CTG·CAG)_n sequence seems to involve MutSβ (the heterodimer of mammalian MutS homologs that recognizes larger IDLs) binding to and stabilizing the intrastrand hairpin structure(s) adopted by CTG or CAG sequences [20]; subsequent error-prone or escaped MMR leads to instability [21]. The important roles of MutS, MutL, and MutH in mediating instability in *E. coli*, and the requirement for their homologs for instability in mammalian systems, indicate that proteins in the MMR system promote instability of the (CTG·CAG)_n sequence.

It is not clear whether the MMR system mediates the instability of triplet repeat sequences other than (CTG·CAG)_n. Specifically, MMR may have a different effect on sequences that are not as likely to form hairpins, since MMR seems to destabilize (CTG·CAG)_n sequences by binding to intrastrand hairpins [20]. For example, while there is some evidence that the GAA sequence forms hairpins [22,23], more compelling data indicate that the (GAA·TTC)_n sequence

adopts triple-helical structures (triplexes) [24-28] or sticky DNA [29,30]. To investigate whether the MMR system mediates instability of the (GAA·TTC)_n sequence we propagated plasmids containing (CTG·CAG)_n and (GAA·TTC)_n sequences in wild-type *E. coli* and isogenic MMR-deficient strains. As previously reported, the (CTG·CAG)_n sequence was significantly stabilized in the absence of MMR. In contrast, the (GAA·TTC)_n sequence was significantly destabilized in the MMR-deficient strains. Our data also show that MMR acts downstream of replication fork stalling but is independent of the RecA-mediated fork restart mechanisms, which were previously shown to be an important determinant of (GAA·TTC)_n sequence instability in *E. coli*. Our results indicate that the MMR system treats various triplet repeat motifs differently, and in addition to preventing small length changes it, also functions to prevent large contractions in the (GAA·TTC)_n sequence.

II. MATERIALS AND METHODS

Plasmid Constructs

As previously described [31], GAA-79 and TTC-79 contain a pure, patient-derived (GAA·TTC)₇₉ sequence inserted into the XbaI and PstI sites of pUC19. In GAA-79, GAA serves as the template for lagging strand synthesis (the GAA orientation), and in TTC-79, TTC serves as the template for lagging strand synthesis (the TTC orientation). pRW3246, a kind gift from Dr. Robert D. Wells (Texas A&M University), contains a pure (CTG·CAG)₉₈ sequence in the HincII site of pUC18, and CAG serves as the template for lagging strand synthesis [32].

E. coli Strains and Growth

The MMR-proficient *E. coli* strain AB1157 (*thr*-1, *ara*C14, *leu*-6, *pro*A2, *lac*Y1, *tsx*-33, *sup*E44, *gal*2, *his*G4(Oc), *rfb*D1, *mgl*-51, *rpo*S396(Am), *rps*L31(Str^R), *kdg*K51, *xyl*A5, *mtl*-1, *arg*E3(Oc), *thi*-1), and isogenic MMR deletion strains KM52 (Δ *mut*L460::*Cam*), KM55 (Δ *mut*H461::*Cam*), and KM75 (Δ *mut*S465::*Tet*) were kindly provided by Dr. Martin Marinus (University of Massachusetts) [33]. The MMR gene deletions were confirmed by PCR of *E. coli* genomic DNA followed by DNA sequencing. The *mutH* gene was amplified using primers 5'-CAGATTTGGAAAACCCGTTG-3' and 5'-CAGGCGTTAGGATCGGTTATC-3', and the *mutL* gene was amplified using primers 5'-CACGCTGCCAAACTAAGGAC-3' and 5'-GGGCAGAATCAACGCTTATC-3'. The *mutS* gene was amplified by nested PCR using primers 5'-GCTGCCTCCACCTCATTAAG-3' and 5'-GGTTGCCTTTAACCGACGTA-3' followed by 5'-TAAGCATCACGCAAAGCTG-3' and 5'-GCGATAGTGATGGGCATTG-3'. PCR products were purified and sequenced with the forward primer. The *E. coli* strain JM83 (*ara* Δ (*lac*-*proAB*) *rps*L (Str^R)[ϕ 80 *dlac* Δ (*lac*Z)M15] *thi*) and isogenic mutants DL887 (*recA*::*Cm*^R), DL902 (*mutS*::*miniTn10*), and DL1084 (*recA*::*Cm*^R, *mutS*::*miniTn10*) were kindly provided by Dr. David Leach (University of Edinburgh) [34]. All strains were grown in LB medium at 37°C. Growth medium was supplemented with ampicillin (100 μ g/ml) for all strains, plus chloramphenicol (10 μ g/ml) for KM52 and KM55, or tetracycline (10 μ g/ml) for KM75. All strains were made chemically competent and transformed with plasmids that contained the full-length repeat. For hydroxyurea treatment, 20 ml cultures were grown in LB medium supplemented with ampicillin (100 μ g/ml). At mid-log phase, cultures were aliquoted into separate 5 ml cultures that were treated with hydroxyurea (0 mM, 50mM, or 100 mM; CAS 127-07-1)[35,36]. After one hour of treatment at 37°C, cultures were washed twice in sterile water, resuspended in 5 mL LB medium, and analyzed for instability.

Analysis of Instability

This was performed essentially as we have previously described [31]. For GAA-79, TTC-79, and pRW3246, cultures were inoculated in triplicate from colonies that contained only the full-

length repeat, as determined by PCR. At mid-log phase, cultures were plated, and instability was assayed by colony PCR. Colony PCR for GAA-79 and TTC-79 was performed with primers GS-F and GS-R [31], or for pRW3246, DSB-F and DSB-R [37]. Repeat lengths were calculated by measuring the migration of PCR products in 2% agarose gels relative to the 1 kb Plus DNA Ladder (Invitrogen). The frequency of mutations was calculated as the number of mutated products per colony analyzed. Loss or gain of more than seven triplets was considered a large change, with changes of less than seven triplets being classified as small changes. Large changes were readily identifiable; however, repeat length variation was also confirmed by sequencing of PCR products from selected clones (with either GS-F or GS-R). Given the relative uncertainty of defining small changes by electrophoresis alone, these were additionally analyzed by PCR (using primers DSB-F and DSB-R) followed by direct sequencing of purified products (using DSB-R).

Statistical Analysis

Instability was compared with ANOVA, followed by a Bonferoni post-hoc analysis when the ANOVA yielded $P < 0.01$. To analyze the frequency of small changes, all mutants that did not meet the definition of a large change were compared by the Levene test for homogeneity of variance. All statistical tests were done with the SPSS software (version 14.0.0).

III. RESULTS

Deficiency of *E. coli* MMR increases the prevalence of large contractions of the (GAA·TTC)_n sequence specifically when GAA is the template for lagging strand synthesis

GAA-79 and TTC-79, which contain the (GAA·TTC)₇₉ sequence in either orientation with respect to the origin of replication, were propagated in the MMR deletion strains KM75 (*mutS*), KM52 (*mutL*), KM55 (*mutH*), and in AB1157, the isogenic MMR-proficient strain. The purity and length of all repeat sequences were confirmed by sequencing prior to transformation. Cultures were inoculated in triplicate from transformants containing full-length repeat tracts. Colony PCR was performed using ~50 colonies per culture (~150 colonies for triplicate analysis for each construct in each bacterial strain) to analyze length variation of the repeat tract. Mutations resulting in changes of ≤ 7 or > 7 triplets were classified as short or long mutations, respectively.

For both GAA-79 and TTC-79, significantly more short length mutations were observed in the mutant strains compared with the wild-type strain (data not shown, $P < 0.01$ for each mutant strain versus wild-type, for both GAA-79 and TTC-79). Both GAA-79 and TTC-79 were equally prone to these short mutations. Given the lower resolution of agarose gel electrophoresis for the detection of short mutations, sequencing of the repeat tract was performed for ten colonies from each of the three mutant strains and the wild-type in both the GAA and TTC orientations, which confirmed that the small changes were indeed gains or losses of ≤ 7 triplets. These data indicate that, as in the case of the (CTG·CAG)_n sequence, the MMR system also serves to prevent short mutations in the (GAA·TTC)_n sequence. Indeed, small “slippage” errors in the absence of MMR seem to be a hallmark of all simple sequence repeats, and the (GAA·TTC)_n sequence is not exceptional in this regard.

Long mutations of the (GAA·TTC)_n sequence were similarly analyzed in the wild-type and three MMR mutant strains. Only 2.7% of the large changes observed were expansions, so they were excluded from our analysis. For GAA-79, we observed an increase in the frequency of large contractions in the *mutS*, *mutL*, and *mutH* mutant strains compared with the wild-type (Figure 1A,C; $P < 0.01$). In contrast, the loss of MMR did not seem to affect large contractions of TTC-79 (Figure 1B,C). Comparing TTC-79 in wild-type and mutants, only *mutH* had a P value of less than 0.01, and this moderate increase in large contractions is unlikely to be

biologically relevant (see also Figure 2B). Whereas short mutations are prevented by MMR in both orientations, our results indicate that the MMR system stabilizes the (GAA·TTC)_n sequence by preventing large contractions specifically when GAA serves as the lagging strand template.

These results contrast with those previously reported for the (CTG·CAG)_n sequence [13-15]. Despite minor differences in the data reported by different groups, there was clearly a decreased frequency of large mutations in MMR deficient strains, indicating that a functional MMR system played a causal role in the generation of mutations in the (CTG·CAG)_n sequence. Given the differences in the mutant strains and the experimental methodology used in the previous studies, we retested the instability of the (CTG·CAG)₉₈ sequence in our strains using the same experimental protocol we utilized to analyze the (GAA·TTC)₇₉ sequence. In agreement with previous reports, we found that the loss of MMR reduced the frequency of large contractions of the (CTG·CAG)₉₈ sequence by approximately one-half (Figure 1D; $P < 0.01$ in all three mutants compared with the wild-type), confirming in our experimental system that MMR produces more long mutations in the (CTG·CAG)_n sequence. Together, these data indicate that the role of the MMR system in triplet repeat instability is dependent on the sequence of the repeating unit.

Large contractions of the (GAA·TTC)_n sequence in *E. coli* MMR mutants occur in response to fork stalling

It is known that replication fork stalling occurs in the (GAA·TTC)_n sequence specifically when GAA is the lagging strand template [38]. Therefore, the orientation-specificity of endogenous replication fork stalling could explain our observation that loss of MMR increases the frequency of contractions specifically when GAA is the lagging strand template. We predicted that if stalling were induced, the absence of MMR would promote large contractions of the (GAA·TTC)_n sequence, whether GAA or TTC served as the lagging strand template. To test this hypothesis, replication fork stalling was induced with hydroxyurea. When replication fork stalling was induced with 100 mM hydroxyurea, TTC-79 showed a significant increase in the prevalence of large contractions in the treated MMR mutants, compared to the treated wild-type (Figure 2; $P < 0.01$ for each treated mutant vs. treated wild-type). Indeed, with 100 mM hydroxyurea treatment, the frequency of large contractions was similar in GAA-79 and TTC-79. Similar observations were made with 50 mM hydroxyurea treatment (data not shown). As expected, with 0 mM hydroxyurea treatment, GAA-79, but not TTC-79, showed an increased frequency of large contractions in MMR mutants compared to wild-type (Figure 2; $P < 0.01$ for GAA-79, $P = 1.0$ for TTC-79). Treatment with 100 mM hydroxyurea did not change the frequency of contractions of GAA-79 or TTC-79 in the wild-type strain ($P > 0.1$ for each untreated vs. treated wild-type), nor did it further enhance contractions of the GAA-79 construct in the MMR-deficient strains (Figure 2; $P > 0.1$ for treated vs. untreated for GAA-79 in each strain). These results suggest that MMR may stabilize the (GAA·TTC)_n sequence downstream of replication fork stalling.

MMR-mediated stabilization of the (GAA·TTC)_n sequence is independent of RecA

We previously demonstrated that efficient RecA-dependent replication fork restart mechanisms are required to prevent instability of the (GAA·TTC)_n sequence specifically when GAA is the lagging strand template. We therefore tested the hypothesis that MMR-mediated stabilization of the (GAA·TTC)_n sequence stems from the role of MMR in improving the efficacy of RecA-mediated restart of stalled replication forks. We analyzed the prevalence of large contractions of GAA-79 and TTC-79 in the wild-type *E. coli* strain JM83 and isogenic mutants lacking either RecA (DL887) or MutS (DL902), or both RecA and MutS (DL1084). Approximately 150 colonies were analyzed for each construct and strain. As expected, deficiency of either RecA or MutS resulted in approximately a two-fold increase in the

prevalence of large contractions for GAA-79, but had no effect on large contractions for TTC-79 (Figure 3; $P < 0.01$ for GAA-79 in either mutant vs. wild-type). However, the *recA mutS* double-mutant strain showed an approximately four-fold increase in large contractions for GAA-79, and a two-fold increase in large contractions for TTC-79, compared to wild-type (Figure 3; $P < 0.001$ for both orientations). For both orientations, the frequency of large contractions was approximately two-fold higher in the *recA mutS* strain than in either single mutant strain (Figure 3; $P < 0.01$ for both orientations). Given the additive effect in the double mutant strain compared with the single mutant strains, these data refute the hypothesis that MMR is acting in the same pathway as RecA, i.e., MMR does not prevent large contractions of the (GAA·TTC)_n sequence simply by ensuring fidelity of replication restart mechanisms. These data are consistent with the existence of at least two separate mechanisms that prevent large contractions of the (GAA·TTC)_n sequence in *E. coli*, both of which may act in response to replication fork stalling when GAA is the lagging strand template.

IV. DISCUSSION

The sequence of a triplet repeat and its location in the genome modify several characteristics of repeat instability, such as tissue-specificity and length distribution of changes. One probable source of instability is the propensity of all disease-associated triplet repeat sequences to form non-B-DNA structures [8,39]. Formation of distinct structures could explain, at least in part, the variations in instability observed for different repeat motifs. For example, the (CTG·CAG)_n sequence forms stable intrastrand hairpins, while the (GAA·TTC)_n sequence is more likely to form triplexes. Thus, DNA metabolic pathways may function differently on (CTG·CAG)_n and (GAA·TTC)_n sequences, leading to their unique patterns of instability.

MMR proteins promote large mutations in (CTG·CAG)_n repeat tracts in bacteria [13-15] and mammals [10,16,17,18,19]. In mammals, expansion or contraction of (CTG·CAG)_n repeats requires binding of the mammalian MutS homolog to CTG / CAG hairpins, followed by error-prone or escaped MMR [20,21]. Based on these data, it has been suggested that the MMR system destabilizes all triplet repeat sequences [1]. We tested this hypothesis by analyzing the prevalence of long mutations in (CTG·CAG)₉₈ and (GAA·TTC)₇₉ sequences propagated in isogenic MMR-proficient and -deficient strains of *E. coli*. Consistent with previous observations, we detected a significant decrease in large contractions of (CTG·CAG)₉₈ in the absence of MMR. In contrast, the frequency of large contractions of (GAA·TTC)₇₉ was significantly increased by the loss of MutS, MutL, or MutH. This protective effect of MMR was only apparent when GAA served as the template for lagging strand synthesis. Therefore, our data indicate that the effect of MMR on triplet repeat sequences depends on multiple factors, including the repeat motif and its orientation with respect to the origin of replication.

Taken together, our results suggest that MMR may prevent large contractions of the (GAA·TTC)_n sequence, acting downstream of replication stalling. First, the loss of MMR only increased the frequency of large contractions in the presence of stalled replication forks, suggesting that the effect of MMR on (GAA·TTC)_n sequences is related to replication fork stalling. In the GAA orientation, when replication fork stalling is known to occur [38], the loss of MMR increased the frequency of large contractions. Conversely, in the TTC orientation, when replication stalling is not known to occur, the loss of MMR had no effect on large contractions. Furthermore, when the stalling of replication was promoted in the TTC orientation, either by hydroxyurea-induced stalling, or by preventing the resolution of stalled forks in a RecA-deficient background, the frequency of large contractions was increased in MMR mutants. Finally, chemically-induced replication fork stalling had no effect in the presence of MMR, indicating that MMR may be acting downstream of replication stalling. Interestingly, in the GAA orientation, hydroxyurea treatment did not further enhance the frequency of large contractions. This could mean that MMR acts after replication fork stalling,

but exposure to hydroxyurea did not substantially increase the frequency of stalling above the level normally seen when GAA serves as the lagging strand template. However, it should be noted that these data do not exclude the possibility that the effect of MMR in the GAA orientation may be independent of replication fork stalling.

We have shown that efficient RecA-dependent restart of stalled replication forks is required to prevent large contractions of the (GAA·TTC)_n sequence specifically in the GAA orientation [40]. In addition, MMR is known to be important for fidelity during recombinational repair of stalled replication forks [41], and misalignments may result in large mutations. However, our data indicate that the stabilizing effect of MMR is likely to be independent of RecA-mediated pathways. These data are consistent with the existence of two separate mechanisms for stabilizing the (GAA·TTC)_n sequence: RecA-dependent replication restart and MMR, each acting in independent pathways to stabilize the (GAA·TTC)_n sequence (Figure 4). Replication fork stalling when GAA serves as the lagging strand template requires proficient RecA-dependent replication restart mechanisms to maintain the length of the repeat tract [40]. Error-prone replication restart could directly lead to changes in the length of the repeat tract, while inefficient replication restart would also allow persistently stalled replication forks to produce double-strand breaks (DSB) [42-44]. Repair of DSBs within the (GAA·TTC)_n sequence is almost always error-prone, leading to large contractions regardless of RecA status [40]. Another possible outcome of replication stalling, or other DNA metabolic processes, is the formation of IDLs. In MMR-proficient strains, IDLs would be accurately repaired to maintain the length of the repeat tract. However, in MMR-deficient strains, failure to repair IDLs would lead to changes in the length of the repeat tracts of daughter molecules following replication. In *recA mutS* double mutants, large contractions could therefore occur due to error-prone replication restart, failure to repair IDLs, and increased frequency of error-prone DSB repair.

The expanded (GAA·TTC)_n sequence displays considerable somatic instability in tissues derived from Friedreich ataxia patients, and the tissue-specific pathology seen in Friedreich ataxia may result from the distinct bias toward large expansions in the dorsal root ganglia. In all other human tissues, somatic instability mainly involves large contractions [11]. Indeed, in actively proliferating tissues, such as peripheral blood cells, the expanded (GAA·TTC)_n sequence shows a marked propensity for large contractions. It is therefore tempting to speculate that the mechanism(s) of somatic instability, at least in actively dividing cells *in vivo*, may share some similarities with the large contractions seen in our simple replication model system.

In conclusion, we have demonstrated that *E. coli* MMR has sequence- and orientation-dependent effects on the ability of triplet repeat tracts to undergo large mutations. In contrast with its destabilizing effect on (CTG·CAG)_n repeats, the *E. coli* MR system tends to prevent large contractions of expanded (GAA·TTC)_n sequences when GAA is the template for lagging strand synthesis. Furthermore, these data are consistent with the notion that secondary structures are a common source of triplet repeat instability. The ability of MMR and other repair systems to distinguish between different structures may explain the unique features of instability observed for different repeat motifs.

ACKNOWLEDGEMENTS

This research was made possible by grants from the National Institutes of Health (NIH / NINDS), Muscular Dystrophy Association, and Friedreich Ataxia Research Alliance to S.I.B. We thank Dr. Robert D. Wells (Texas A&M University) for kindly providing us with the repeat-containing plasmid, pRW3246, Dr. Martin Marinus (University of Massachusetts) for providing us with the MMR-deficient bacterial strains, and Dr. David Leach (University of Edinburgh) for providing us with the RecA- and MutS-deficient bacterial strains.

REFERENCES

- [1]. Jiricny J. The multifaceted mismatch-repair system. *Nat. Rev. Mol. Cell Biol* 2006;7:335–346. [PubMed: 16612326]
- [2]. Dohet C, Dzidic S, Wagner R, Radman M. Large non-homology in heteroduplex DNA is processed differently than single base pair mismatches. *Mol. Gen. Genet* 1987;206:181–184. [PubMed: 3472034]
- [3]. Fang WH, Wang BJ, Wang CH, Lee SJ, Chang YT, Chuang YK, Lee CN. DNA loop repair by *Escherichia coli* cell extracts. *J. Biol. Chem* 2003;278:22446–22452. [PubMed: 12692132]
- [4]. Peltomaki P. Role of DNA mismatch repair defects in the pathogenesis of human cancer. *J. Clin. Oncol* 2003;21:1174–1179. [PubMed: 12637487]
- [5]. Lahue RS, Au KG, Modrich P. DNA mismatch correction in a defined system. *Science* 1989;245:160–164. [PubMed: 2665076]
- [6]. Orr HT, Zoghbi HY. Trinucleotide repeat disorders. *Annu. Rev. Neurosci* 2007;30:575–621. [PubMed: 17417937]
- [7]. Pearson CE, Edamura KN, Cleary JD. Repeat instability: Mechanisms of dynamic mutations. *Nat. Rev. Genet* 2005;6:729–742. [PubMed: 16205713]
- [8]. Wells RD, Dere R, Hebert ML, Napierala M, Son LS. Advances in mechanisms of genetic instability related to hereditary neurological diseases. *Nucleic Acids Res* 2005;33:3785–3798. [PubMed: 16006624]
- [9]. Shelbourne PF, Keller-McGandy C, Bi WL, Yoon SR, Dubeau L, Veitch NJ, Vonsattel JP, Wexler NS, US-Venezuela Collaborative Research Group. Arnheim N, Augood SJ. Triplet repeat mutation length gains correlate with cell-type specific vulnerability in Huntington disease brain. *Hum. Mol. Genet* 2007;16:1133–1142. [PubMed: 17409200]
- [10]. Wheeler VC, Lebel LA, Vrbanc V, Teed A, te Riele H, MacDonald ME. Mismatch repair gene Msh2 modifies the timing of early disease in hdh(Q111) striatum. *Hum. Mol. Genet* 2003;12:273–281. [PubMed: 12554681]
- [11]. De Biase I, Rasmussen A, Endres D, Al-Mahdawi S, Monticelli A, Cocozza S, Pook M, Bidichandani SI. Progressive GAA expansions in dorsal root ganglia of Friedreich's ataxia patients. *Ann. Neurol* 2007;61:55–60. [PubMed: 17262846]
- [12]. Clark RM, De Biase I, Malykhina AP, Al-Mahdawi S, Pook M, Bidichandani SI. The GAA triplet-repeat is unstable in the context of the human FXN locus and displays age-dependent expansions in cerebellum and DRG in a transgenic mouse model. *Hum. Genet* 2007;120:633–640. [PubMed: 17024371]
- [13]. Jaworski A, Rosche WA, Gellibolian R, Kang S, Shimizu M, Bowater RP, Sinden RR, Wells RD. Mismatch repair in *Escherichia coli* enhances instability of (CTG)_n triplet repeats from human hereditary diseases. *Proc. Natl. Acad. Sci. U. S. A* 1995;92:11019–11023. [PubMed: 7479928]
- [14]. Schmidt KH, Abbott CM, Leach DR. Two opposing effects of mismatch repair on CTG repeat instability in *Escherichia coli*. *Mol. Microbiol* 2000;35:463–471. [PubMed: 10652107]
- [15]. Parniewski P, Jaworski A, Wells RD, Bowater RP. Length of CTG-CAG repeats determines the influence of mismatch repair on genetic instability. *J. Mol. Biol* 2000;299:865–874. [PubMed: 10843843]
- [16]. Manley K, Shirley TL, Flaherty L, Messer A. Msh2 deficiency prevents in vivo somatic instability of the CAG repeat in Huntington disease transgenic mice. *Nat. Genet* 1999;23:471–473. [PubMed: 10581038]
- [17]. van den Broek WJ, Nelen MR, Wansink DG, Coerwinkel MM, te Riele H, Groenen PJ, Wieringa B. Somatic expansion behaviour of the (CTG)_n repeat in myotonic dystrophy knock-in mice is differentially affected by Msh3 and Msh6 mismatch-repair proteins. *Hum. Mol. Genet* 2002;11:191–198. [PubMed: 11809728]
- [18]. Foirey L, Dong L, Savouret C, Hubert L, te Riele H, Junien C, Gourdon G. Msh3 is a limiting factor in the formation of intergenerational CTG expansions in DM1 transgenic mice. *Hum. Genet* 2006;119:520–526. [PubMed: 16552576]

- [19]. Gomes-Pereira M, Fortune MT, Ingram L, McAbney JP, Monckton DG. Pms2 is a genetic enhancer of trinucleotide CAG·CTG repeat somatic mosaicism: Implications for the mechanism of triplet repeat expansion. *Hum. Mol. Genet* 2004;13:1815–1825. [PubMed: 15198993]
- [20]. Owen BA, Yang Z, Lai M, Gajek M, Badger JD 2nd, Hayes JJ, Edelmann W, Kucherlapati R, Wilson TM, McMurray CT. (CAG)_(n)-hairpin DNA binds to Msh2-Msh3 and changes properties of mismatch recognition. *Nat. Struct. Mol. Biol* 2005;12:663–670. [PubMed: 16025128]
- [21]. Panigrahi GB, Lau R, Montgomery SE, Leonard MR, Pearson CE. Slipped (CTG)·(CAG) repeats can be correctly repaired, escape repair or undergo error-prone repair. *Nat. Struct. Mol. Biol* 2005;12:654–662. [PubMed: 16025129]
- [22]. Heidenfelder BL, Makhov AM, Topal MD. Hairpin formation in Friedreich's ataxia triplet repeat expansion. *J. Biol. Chem* 2003;278:2425–2431. [PubMed: 12441336]
- [23]. Suen IS, Rhodes JN, Christy M, McEwen B, Gray DM, Mitas M. Structural properties of Friedreich's ataxia d(GAA) repeats. *Biochim. Biophys. Acta* 1999;1444:14–24. [PubMed: 9931411]
- [24]. Tiner WJ S, Potaman VN, Sinden RR, Lyubchenko YL. The structure of intramolecular triplex DNA: Atomic force microscopy study. *J. Mol. Biol* 2001;314:353–357. [PubMed: 11846549]
- [25]. Potaman VN, Oussatcheva EA, Lyubchenko YL, Shlyakhtenko LS, Bidichandani SI, Ashizawa T, Sinden RR. Length-dependent structure formation in Friedreich ataxia (GAA)_n·(TTC)_n repeats at neutral pH. *Nucleic Acids Res* 2004;32:1224–1231. [PubMed: 14978261]
- [26]. Mariappan SV, Catasti P, Silks LA 3rd, Bradbury EM, Gupta G. The high-resolution structure of the triplex formed by the GAA/TTC triplet repeat associated with Friedreich's ataxia. *J. Mol. Biol* 1999;285:2035–2052. [PubMed: 9925783]
- [27]. Wells RD. DNA triplexes and Friedreich ataxia. *FASEB J* 2008;22:1625–1634. [PubMed: 18211957]
- [28]. Shimizu M, Hanvey JC, Wells RD. Multiple non-B-DNA conformations of polypurine-polypyrimidine sequences in plasmids. *Biochemistry* 1990;29:4704–4713. [PubMed: 2372553]
- [29]. Sakamoto N, Chastain PD, Parniewski P, Ohshima K, Pandolfo M, Griffith JD, Wells RD. Sticky DNA: Self-association properties of long GAA·TTC repeats in R·R·Y triplex structures from Friedreich's ataxia. *Mol. Cell* 1999;3:465–475. [PubMed: 10230399]
- [30]. Vetcher AA, Napierala M, Wells RD. Sticky DNA: Effect of the polypurine-polypyrimidine sequence. *J. Biol. Chem* 2002;277:39228–39234. [PubMed: 12161438]
- [31]. Pollard LM, Sharma R, Gomez M, Shah S, Delatycki MB, Pianese L, Monticelli A, Keats BJ, Bidichandani SI. Replication-mediated instability of the GAA triplet repeat mutation in Friedreich ataxia. *Nucleic Acids Res* 2004;32:5962–5971. [PubMed: 15534367]
- [32]. Bowater RP, Rosche WA, Jaworski A, Sinden RR, Wells RD. Relationship between *Escherichia coli* growth and deletions of CTG·CAG triplet repeats in plasmids. *J. Mol. Biol* 1996;264:82–96. [PubMed: 8950269]
- [33]. Loh T, Murphy KC, Marinus MG. Mutational analysis of the MthH protein from *Escherichia coli*. *J. Biol. Chem* 2001;276:12113–12119. [PubMed: 11124943]
- [34]. Pan X, Leach DR. The roles of mutS, sbcCD and recA in the propagation of TGG repeats in *Escherichia coli*. *Nucleic Acids Res* 2000;28:3178–3184. [PubMed: 10931934]
- [35]. Morganroth PA, Hanawalt PC. Role of DNA replication and repair in thymineless death in *Escherichia coli*. *J. Bacteriol* 2006;188:5286–5288. [PubMed: 16816201]
- [36]. Godoy VG, Jarosz DF, Walker FL, Simmons LA, Walker GC. Y-family DNA polymerases respond to DNA damage-independent inhibition of replication fork progression. *EMBO J* 2006;25:868–879. [PubMed: 16482223]
- [37]. Pollard LM, Bourn RL, Bidichandani SI. Repair of DNA double-strand breaks within the (GAA·TTC)_n sequence results in frequent deletion of the triplet-repeat sequence. *Nucleic Acids Res* 2008;36:489–500. [PubMed: 18045804]
- [38]. Krasilnikova MM, Mirkin SM. Replication stalling at Friedreich's ataxia (GAA)_n repeats in vivo. *Mol. Cell. Biol* 2004;24:2286–2295. [PubMed: 14993268]
- [39]. Mirkin SM. Expandable DNA repeats and human disease. *Nature* 2007;447:932–940. [PubMed: 17581576]

- [40]. Pollard LM, Chutake YK, Rindler PM, Bidichandani SI. Deficiency of RecA-dependent RecFOR and RecBCD pathways causes increased instability of the (GAA·TTC)_n sequence when GAA is the lagging strand template. *Nucleic Acids Res* 2007;35:6884–6894. [PubMed: 17932052]
- [41]. Evans E, Alani E. Roles for mismatch repair factors in regulating genetic recombination. *Mol. Cell. Biol* 2000;20:7839–7844. [PubMed: 11027255]
- [42]. Michel B, Ehrlich SD, Uzest M. DNA double-strand breaks caused by replication arrest. *EMBO J* 1997;16:430–438. [PubMed: 9029161]
- [43]. Seigneur M, Bidnenko V, Ehrlich SD, Michel B. RuvAB acts at arrested replication forks. *Cell* 1998;95:419–430. [PubMed: 9814711]
- [44]. Marians KJ. Mechanisms of replication fork restart in *Escherichia coli*. *Philos. Trans. R. Soc. Lond. B. Biol. Sci* 2004;359:71–77. [PubMed: 15065658]

Figure 1.

E. coli MMR prevents large contractions (>7 triplets) of the (GAA·TTC)_n sequence specifically when GAA is the lagging strand template. (A,B) Representative agarose gels showing PCR products generated from colonies containing the (GAA·TTC)₇₉ repeat tract in the GAA (A) and TTC (B) orientations in the MMR-proficient (wild-type) strain and three isogenic MMR-deficient mutants, *mutS*, *mutL*, and *mutH*. Short changes in length (<7 triplets) were more common in the mutants than in the wild-type for both GAA-79 and TTC-79 ($P < 0.01$). Arrowheads indicate the position of the full-length repeat. Horizontal lines along the right edge of the gel pictures represent the position of 100-bp, 200-bp, 300-bp, 400-bp and 500-bp, based on the 1 Kb Plus DNA Ladder (Invitrogen). (C) The frequency of large contractions, calculated

as the number of large contractions per successful PCR, was significantly enhanced for GAA-79 when propagated in the three MMR-deficient strains relative to the isogenic MMR-proficient strain ($P < 0.01$ for each comparison of mutant vs. wild-type). No such increase in the frequency of large contractions was noted for TTC-79. The moderate increase in large contractions observed for *mutH*, while statistically significant ($P < 0.01$), is unlikely to be biologically relevant. **(D)** In contrast, the frequency of large contractions was significantly reduced for CTG-98 when propagated in the three MMR-deficient strains compared to the isogenic MMR-proficient strain ($P < 0.01$ for each mutant). Error bars depict $\pm 2\text{SEM}$.

Figure 2.

Replication stalling promotes large contractions of the (GAA·TTC)_n sequence in *E. coli* MMR mutants. Cultures were either treated with 100 mM hydroxyurea to induce replication fork stalling, or with 0 mM hydroxyurea as a control. As expected, in the absence of hydroxyurea, GAA-79 (**A**) showed a significant increase in the frequency of large contractions in the MMR-deficient strains *mutS*, *mutL*, and *mutH* compared to the isogenic wild-type, but TTC-79 (**B**) showed no such increase ($P < 0.01$ for GAA-79; $P = 1.0$ for TTC-79). After the induction of replication fork stalling with hydroxyurea, the loss of MMR led to an increased frequency of large contractions for TTC-79 ($P < 0.01$); however, hydroxyurea treatment did not further

increase large contractions of GAA-79 ($P > 0.1$ for treated vs. untreated in each strain). Error bars depict ± 2 SEM.

Figure 3.

Loss of MMR promotes large contractions of the (GAA·TTC)_n sequence independently of RecA status. The individual mutation of *recA* or *mutS* significantly increased the frequency of large contractions for GAA-79 ($P < 0.01$), but had no effect on TTC-79. The *recA mutS* double mutant showed a higher frequency of large contractions than either single mutant for both GAA-79 and TTC-79 ($P < 0.001$ in both orientations). Error bars depict ± 2 SEM.

Figure 4.

Independent functions of RecA and MMR in maintaining stability of the (GAA·TTC)_n sequence in *E. coli*. The replication fork is known to stall when GAA serves as the lagging strand template. Efficient RecA-dependent restart of stalled replication forks is required to maintain the length of the repeat tract. In a RecA-deficient state, error-prone replication restart promotes large contractions. Additionally, failure to repair stalled forks may lead to DSB formation, and repair of DSBs within the (GAA·TTC)_n sequence is known to cause large contractions. The role of MMR in correcting IDLs that may form at the stalled fork is likely to be critical for preventing large contractions. In the absence of MMR, large contractions of the (GAA·TTC)_n sequence may result from bypass of unresolved IDLs, resulting in a shorter nascent strand during replication.