The EMBO Journal (2008) 27, 2896–2906 | © 2008 European Molecular Biology Organization | All Rights Reserved 0261-4189/08
www.embojournal.org

THE EMBO JOURNAL

Chromosome fragility at GAA tracts in yeast depends on repeat orientation and requires mismatch repair

Hyun-Min Kim[1], Vidhya Narayanan[1], Piotr A Mieczkowski[2,5], Thomas D Petes[2], Maria M Krasilnikova[3], Sergei M Mirkin[4] and Kirill S Lobachev[1,*]

[1]School of Biology and Institute for Bioengineering and Bioscience, Georgia Institute of Technology, Atlanta, GA, USA, [2]Department of Molecular Genetics and Microbiology, Duke University, Durham, NC, USA, [3]Department of Biochemistry and Molecular Biology, Penn State University, University Park, PA, USA and [4]Department of Biology, Tufts University, Medford, MA, USA

Expansion of triplex-forming GAA/TTC repeats in the first intron of *FXN* gene results in Friedreich's ataxia. Besides *FXN*, there are a number of other polymorphic GAA/TTC loci in the human genome where the size variations thus far have been considered to be a neutral event. Using yeast as a model system, we demonstrate that expanded GAA/TTC repeats represent a threat to eukaryotic genome integrity by triggering double-strand breaks and gross chromosomal rearrangements. The fragility potential strongly depends on the length of the tracts and orientation of the repeats relative to the replication origin, which correlates with their propensity to adopt triplex structure and to block replication progression. We show that fragility is mediated by mismatch repair machinery and requires the MutSβ and endonuclease activity of MutLα. We suggest that the mechanism of GAA/TTC-induced chromosomal aberrations defined in yeast can also operate in human carriers with expanded tracts.

The EMBO Journal (2008) **27,** 2896–2906. doi:10.1038/emboj.2008.205; Published online 2 October 2008
Subject Categories: genome stability & dynamics; molecular biology of disease
Keywords: chromosomal fragility; chromosomal rearrangements; DNA secondary structure; mismatch repair; trinucleotide repeats

Introduction

Expansion of GAA/TTC trinucleotide repeats was recognized as a detrimental polymorphism in the human genome with the discovery of the molecular mechanisms underlying Friedreich's ataxia (FRDA) (Campuzano *et al*, 1996). FRDA is an autosomal recessive disease caused by the inheritance of two mutant alleles of the frataxin (*FXN*) gene from heterozygous parents (reviewed by De Biase *et al*, 2006; Pandolfo, 2006). In most cases (98%), inactivation of the *FXN* function in both alleles results from inhibition of gene expression by abnormal GAA repeat expansion occurring within the first intron. Although chromosomes from unaffected individuals have less than 65 triplets, disease-causing FRDA alleles contain 66–1700 GAA repeats. Premutation (34–65 triplets) and mutant (>66 triplets) alleles exhibit high levels of instability (expansions and contractions) in somatically dividing and non-dividing cells in a tissue- and an age-dependent manner (Al-Mahdawi *et al*, 2004; Clark *et al*, 2007). Premutation and disease alleles are also highly unstable during intergenerational transmission often undergoing both contractions and expansions, with hyperexpansions reaching up to a 10-fold increase in one generation (reviewed by De Biase *et al*, 2006).

Systematic analysis of the human genome revealed that the *FXN* locus is not the only location where GAA tracts can expand (Clark *et al*, 2004, 2006). Almost 1000 loci containing more than eight GAA repeats, including 29 loci with premutation size tracts, have been identified. In total, 9 out of 29 premutation alleles are highly polymorphic and prone to large expansions, which can reach up to 140 copies. To date, these expansions have not been shown to be associated with diseases.

The property of GAA repeats to inhibit *FXN* transcription and its predisposition for genetic instability are dependent on the size of the expanded tracts, which in fact reflects the ability of the repeats to adopt non-canonical DNA secondary structures (reviewed by Wells, 2008). The GAA triplet repeat is a polypurine polypyrimidine (R·Y) sequence exhibiting mirror symmetry (reviewed by Frank-Kamenetskii and Mirkin, 1995). Such RY tracts can predominantly adopt two non-B-DNA structures: triplex (or H-DNA) and sticky DNA. Triplex is formed as a result of overlaying a third strand into the major groove of the DNA double helix. The third strand pairs with the double helix through Hoogsteen or reverse Hoogsteen hydrogen bonds, thus leaving the complementary strand (either R or Y) unpaired. Sticky DNA is a more complex triplex structure formed between two remote R·Y tracts positioned in direct orientation relative to each other within the same molecule (reviewed by Wells, 2008).

Triplexes were detected *in vitro* and *in vivo* studies in model systems (reviewed by Bissler, 2007). The formation of such structures is dependent on the homogeneity of the GAA tract and is strongly favored under conditions of negative superhelicity, which *in vivo* can be provided by processes that require separation of the two strands of the duplex such as replication, transcription, recombination and repair (Frank-Kamenetskii and Mirkin, 1995). Studies *in vitro* and in model organisms show that stable secondary structures in

*Corresponding author. School of Biology and Institute for Bioengineering and Bioscience, Georgia Institute of Technology, 310 Ferst Drive, Cherry Emerson Building, Atlanta, GA 30332, USA. Tel.: +1 404 385 6197; Fax: +1 404 894 0519; E-mail: kirill.lobachev@biology.gatech.edu
[5]Present address: Department of Genetics and the Carolina Center for Genome Sciences, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA

Received: 29 July 2008; accepted: 15 September 2008; published online: 2 October 2008

 ©2008 European Molecular Biology Organization

turn can hinder transcription (Bidichandani *et al*, 1998; Ohshima *et al*, 1998; Sakamoto *et al*, 1999; Grabczyk and Usdin, 2000a, b), which can account for the GAA length-dependent inactivation of *FXN* gene function in FRDA patients (Campuzano *et al*, 1996; Cossee *et al*, 1997). It was also found that triplexes formed by GAA repeats stall the DNA synthesis *in vitro* (Ohshima *et al*, 1996; Gacy *et al*, 1998) and the progression of replication fork *in vivo* (Ohshima *et al*, 1998; Krasilnikova and Mirkin, 2004; Pollard *et al*, 2004), providing possible explanations for the GAA-associated genetic instability and the origin of expanded alleles.

Another type of expandable triplet repeat sequences, hairpin-forming CNG triplets, also cause disease by affecting gene expression, mRNA or protein function (reviewed by Mirkin, 2007). In addition, CNG repeats can compromise the integrity of the eukaryotic genomes. In humans, the expanded CCG or GCC tracts are chromosomal fragile sites that are associated with genome rearrangements (Sutherland, 2003), whereas long tracts of CCG/CGG and CTG/CAG repeats induce DSBs and chromosome instability in yeast (reviewed by Lenzmeier and Freudenreich, 2003). The CNG potential to cause chromosomal breakage and aberrations is attributed to their ability to adopt hairpin secondary structures that impede replication progression.

Although the fragility at triplex-forming GAA/TTC repeats has not been documented, a set of observations point towards their potential to cause breakage. First, in *Escherichia coli*, GAA tracts on a plasmid induce both intra- and intermolecular recombination (Napierala *et al*, 2004). Second, in yeast, GAA repeats lead to replication stalling on a plasmid in a length- and orientation-dependent manner (Krasilnikova and Mirkin, 2004). As shown from other studies, replication arrest or defects in replication machinery often lead to DSBs (reviewed by Rothstein *et al*, 2000). Third, as described above, GAA repeats can adopt triplex structures that are considered to be inducers of instability. Non-GAA mirror repeats that are prone to the formation of similar structures are also found to be hotspots for rearrangements in humans and in model organisms. These include triplex-forming sequences located in the major breakpoint cluster region at *BCL2* (Raghavan *et al*, 2005a, b), intron 21 of *PKD1* (Blaszak *et al*, 1999; Patel *et al*, 2004) and promoter region of *C-MYC* (Michelotti *et al*, 1996; Wang and Vasquez, 2004).

In this study, we demonstrate that expanded GAA/TTC repeats are strong inducers of DSBs and gross chromosomal rearrangements (GCRs) in yeast. The fragility potential depends on the length of the tract and the orientation of the repeats relative to the replication origin, which correlates with their propensity to adopt triplex and to block replication fork movement. Mutants defective in the function of MutSβ and the endonuclease activity of MutLα exhibit reduced levels of GCRs and DSB formation, indicating that mismatch repair machinery (MMR) might trigger the fragility by processing the triplex structure. GCRs resulting from the GAA-mediated breaks have a specific pattern: terminal deletions coupled with non-reciprocal translocations involving expanded GAA/TTC tracts and GAA/TTC-rich regions located on non-homologous chromosomes. We propose that the mechanism of genome destabilization caused by GAA/TTC repeats defined in yeast might operate in carriers with expanded tracts at the *FXN* and other loci in the human genome.

Results

Experimental system

To assess the potential of the expanded GAA/TTC repeats to induce chromosomal fragility, we have employed two experimental assays that monitor the induction of GCRs and mitotic ectopic recombination. The GCR assay is based on the loss of *CAN1* and *ADE2* genes located on chromosome V (Figure 1). This experimental assay was used earlier to characterize the specific pattern of GCRs resulting from hairpin-capped breaks induced by inverted repeats (Narayanan *et al*, 2006). Haploid yeast strains were constructed where the left arm of chromosome V in the region of *CAN1* gene was modified. *LYS2*

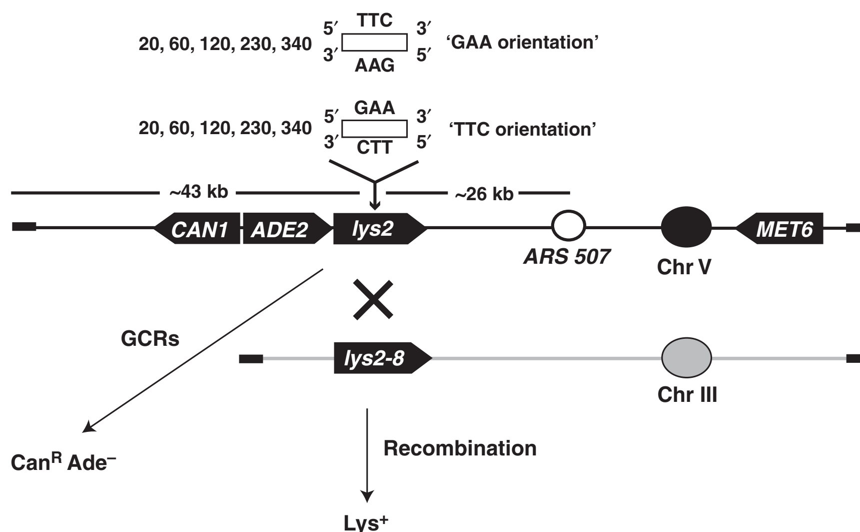

Figure 1 Experimental system to study chromosomal fragility induced by expanded tracts of GAA/TTC repeats. The breakage at the location of GAA/TTC tracts can lead to 43 kb telomere-proximal deletion resulting in Can^RAde^- clones. In a separate set of strains, the *lys2-8* allele was integrated into chromosome III, allowing us to measure the level of homologous recombination induced by GAA/TTC repeats. The 'X' denotes a recombination event generating a wild-type *LYS2* allele.

©2008 European Molecular Biology Organization

cassettes with GAA/TTC repeats were placed centromere-proximal to *CAN1*. The region between *LYS2* and the telomere does not contain essential genes and can be deleted. The *ADE2* gene was moved telomere-distal to *CAN1*. The *LYS2* cassettes contain homogeneous GAA/TTC repeats of length 20 (corresponding to normal allele size in humans), 60 (premutation size), 120, 230 or 340 (mutant sizes); all GAA/TTC insertions result in loss of *LYS2* function. Repeats were inserted into the chromosome in two different orientations with respect to the direction of replication. Replication is initiated at the *ARS507* origin and proceeds from right to left in this region (Raghuraman *et al*, 2001; Yabuki *et al*, 2002, see also replication origin database at http://www.oridb.org/). The lagging strand template contains GAA repeats in 'GAA orientation' and TTC repeats in 'TTC orientation'. A DSB in the *LYS2* region can cause deletion of the chromosome V region, including *CAN1* and *ADE2*, resulting in canavanine-resistant red colonies ($Can^R Ade^-$). Such GCR isolates can be distinguished from canavanine-resistant white colonies that are produced due to point mutations or small deletions in *CAN1*.

To determine whether the GAA/TTC tracts can stimulate mitotic ectopic recombination, we integrated a *lys2-8* allele (Lobachev *et al*, 1998) at the *LEU2* locus of chromosome III. Recombination between *lys2*::*GAA/TTC* and the *lys2-8* generates Lys^+ prototrophs, primarily through gene conversion of the insert-containing allele.

GAA repeats induce GCRs in a size- and an orientation-dependent manner

Expanded tracts of GAA/TTC repeats strongly increased the rate of chromosome V arm loss (Table I). The degree of stimulation depended on both the size and the orientation of the repetitive tracts. The strains with $(GAA)_{20}$ and $(TTC)_{20}$ tracts exhibited low rates of arm loss events, similar to that of strains containing direct *Alu* repeats, which cannot adopt secondary structures (Narayanan *et al*, 2006). There was a mild increase (~3-fold) in the levels of *CAN1* region loss for premutation size alleles (60 repeats) over normal size alleles for both orientations. However, alteration in the size of the repeat tracts from 60 to 120 led to a tremendous change in their ability to trigger chromosomal arm loss events. There were an ~200- and 650-fold increases in TTC and GAA orientation, respectively. Interestingly, further increments in the repeat tract lengths had different effects on arm loss rates for different orientations. TTC repeats of 120, 230 and 340 tract lengths had similar rates, whereas the 120, 230 and 340 GAA tracts stimulated GCRs ~2000, 7000 and 83 000 times more frequently than $(GAA)_{20}$, respectively.

Structural organization of rearranged chromosomes in $Can^R Ade^-$ isolates

To directly determine what structural changes were acquired by chromosome V as a result of GCR, we analysed the molecular karyotypes of 12 independent $Can^R Ade^-$ clones isolated from strains containing $(TTC)_{230}$ and $(GAA)_{230}$ repeats. Chromosomes from these isolates were separated using contour-clamped homogeneous electric field (CHEF) gel electrophoresis and chromosome V was examined by hybridization with a right arm-specific probe (Figure 2A). On the basis of the mobility of altered chromosomes, several different recurrent classes of rearrangements were detected for both repeat orientations. In the majority of cases, the novel chromosomes were larger than wild-type chromosome V, suggesting that the arm loss events were accompanied by the gain of genetic material. This conclusion was confirmed when genomic DNA from $Can^R Ade^-$ isolates was analysed using comparative genomic hybridization (CGH) on microarrays (Figure 2B). Among the 12 analysed $Can^R Ade^-$ isolates from $(TTC)_{230}$ strains, only two (T-2 and T-10) had a terminal deletion of V with a breakpoint near *CAN1* locus. This pattern likely reflects *de novo* telomere addition to the broken molecule following DSB induction at the repetitive tracts. The remaining 10 isolates had a deletion of the centromere-distal *CAN1* region coupled with a duplication of telomere-proximal regions of non-homologous chromosomes. One likely mechanism for generating such rearrangements is the induction of the break at the location of GAA/TTC repeats on chromosome V, followed by healing of the broken end through break-induced replication (BIR) (Malkova *et al*, 1996; Smith *et al*, 2007) involving homology or microhomology (Figure 6). Non-reciprocal translocations were confirmed by PCR analysis with primers annealing to the regions on non-homologous chromosomes that flank the breakpoints. We sequenced the PCR fragments containing the breakpoint junctions for one or several representatives of each GCR class. All such junctions were chimaeric with a GAA/TTC pure repeat region from chromosome V fused with non-homogeneous GAA/TTC-rich

Table I Length- and orientation-dependent induction of GCRs and homologous recombination by GAA/TTC repeats

Insertion in *LYS2*		Arm loss rate ($\times 10^{10}$)[a]	Recombination rate ($\times 10^{7}$)
Orientation	Tract length		
TTC	20	8 (4–17)[b]	5 (4–9)
	60	21 (19–35)	8 (6–11)
	120	4515 (3226–6149)	19 (14–23)
	230	4778 (4409–6128)	20 (17–36)
	340	4851 (4133–6659)	ND[c]
GAA	20	9 (5–15)	7 (5–9)
	60	27 (20–43)	17 (12–27)
	120	17 436 (9120–24 770)	287 (170–373)
	230	66 044 (51 054–86 480)	3808 (2410–4384)
	340	743 659 (468 938–1 085 186)	ND[c]

[a]The loss of *CAN1*- and *ADE2*-containing region was measured in strains that do not have the *lys2-8* allele.
[b]Numbers in parentheses correspond to the 95% confidence interval.
[c]ND, not determined.

 ©2008 European Molecular Biology Organization

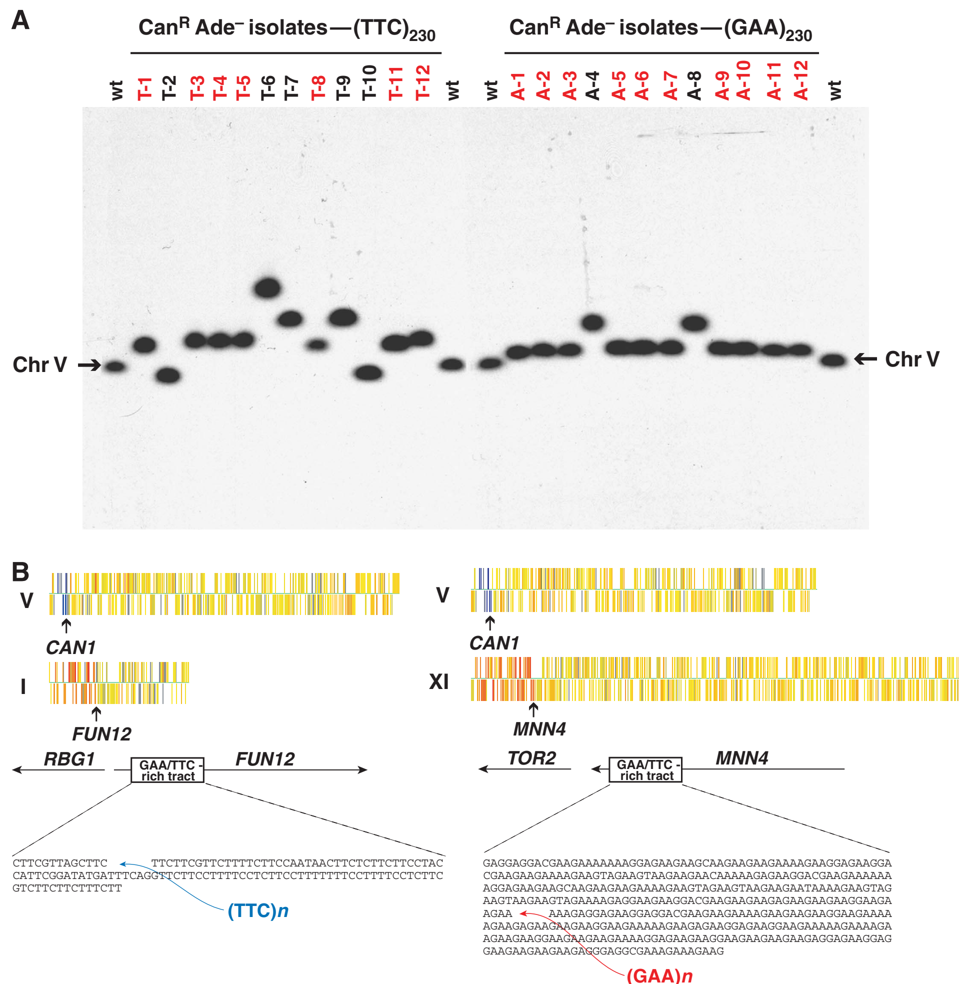

Figure 2 Structural analysis of chromosomal arm loss events stimulated by $(GAA/TTC)_{230}$ tracts. (**A**) Analysis of rearranged chromosome Vs in $Can^R Ade^-$ isolates by CHEF gels and Southern blotting. The right arm of chromosome V was highlighted using a *MET6*-specific probe in Southern analysis. Lanes labelled with 'wt' are strains containing wild-type chromosome V with $(TTC)_{230}$ and $(GAA)_{230}$ repeats. Lanes T-1 to T-12 are $Can^R Ade^-$ isolates from $(TTC)_{230}$ strains. Lanes A-1 and A-12 are $Can^R Ade^-$ isolates from strains with $(GAA)_{230}$ repeats. The primary GCR classes are labelled in red. (**B**) CGH and breakpoint analysis of the most frequent rearrangements resulting from $(TTC/GAA)_{230}$-mediated breaks. Upper panels are the microarray analysis of arm loss events. DNAs from experimental strain and control strain were labelled with different fluorescent nucleotides and hybridized in competition to DNA microarrays with yeast genes and intergenic regions. Each vertical bar corresponds to one ORF in Watson (upper bars) and in Crick (bottom bars) orientations. Colour coding is as follows: grey, repeated genomic elements; yellow, sequences present in the same dosage in the wild-type and control strains; red, sequences that were duplicated in the experimental strain relative to the control; blue, sequences that were deleted in the experimental strain relative to the control. Only those chromosomes that had a deletion or duplication are shown in this figure. Complete data for these experiments is online at GEO database (accession number GSE11425). Bottom panels depict the structure of the translocation breakpoints on chromosomes I and XI. The donor sites for BIR are shown. Blue and red arrows indicate the breakpoint junctions between GAA/TTC tracts from chromosome V and GAA/TTC-rich regions on donor chromosomes (examples are shown). The left panel is the analysis of a major class of GCRs in $(TTC)_{230}$ strains (isolates T-1, T-3, T-4, T-5, T-8, T-11 and T-12). The right panel is the analysis of a major class of GCRs in $(GAA)_{230}$ strains (isolates A-1, A-2, A-3, A-5, A-6, A-7, A-9, A-10, A-11 and A-12). The complete analysis of the breakpoints for all isolates is presented in the Supplementary Table 1.

stretches in chromosome I (T-1, T-3, T-4, T-5, T-8, T-11 and T-12), chromosome XIII (T-7 and T-9) or chromosome II (T-6) (Figure 2B; Supplementary Table I). The imperfect GAA/TTC repeats were all configured such that a BIR event initiated by a break in the pure GAA/TTC tract on chromosome V in the TTC orientation would produce a monocentric chromosome.

Similarly, in isolates derived from $(GAA)_{230}$ strains the broken end was stabilized through BIR involving chromosome XI (A-1, A-2, A-3, A-5, A-6, A-7, A-9, A-10, A-11 and A-12) or chromosome XIII (A-4 and A-8). The translocation breakpoints were mapped to 420 bp GAA/TTC-rich tract in *MNN4* gene on chromosome XI and to a 120 bp GAA/TTC-containing tract in *FPR3* gene on chromosome XIII (Figure 2B; Supplementary Table I).

In summary, these results indicate that repair of breaks triggered by GAA/TTC repeats generate specific patterns of rearrangements, wherein non-reciprocal translocations are the primary outcome of the GAA/TTC-mediated fragility.

Induction of homologous recombination between lys2 alleles depends on the orientation of repeat tracts

The disparity in the GCR potential of GAA and TTC tracts can be explained by the different propensities of the repeats to adopt secondary structures that are processed to DSBs (discussed below). Alternatively, the bias can be attributed to the efficiency of subsequent steps in the recombination process (processing of the broken ends, invasion of the broken end into a homologous template and so on). In addition, the

©2008 European Molecular Biology Organization

number of genomic templates that are potential sites for invasion by GAA/TTC tracts, their lengths, degree of sequence divergence and orientation with respect to telomere could be contributing factors to account for the differences in GCR rates. To determine whether the disparity in the GCR potential reflects DSB formation or a subsequent step in the repair, we analysed recombination between a *lys2-8* allele integrated at the *LEU2* locus of chromosome III and the *lys2* alleles containing 20, 60, 120 and 230 repeats in both orientations. The *lys2-8* allele serves as a uniform template for DSB repair and allows to measure the true repeat fragility potential in an unbiased manner. Consistent with GCR data, we found that the GAA repeats stimulate recombination more strongly than the TTC repeats. For example, the $(GAA)_{230}$ tract induces recombination between *lys2* alleles about 200 times more efficiently than the $(TTC)_{230}$ tract (Table I). These results indicate that the observed orientation dependence is primarily attributed by the differences in the propensity of the GAA and TTC repeats for breakage.

Orientation-dependent blockage of replication by expanded GAA tracts

To get better insights into the molecular mechanisms of GAA/TTC-associated instability, we analysed the progression of the replication fork through the chromosomal region containing GAA/TTC tracts using two-dimensional (2D) gel electrophoresis (Figure 3). Replication progression across $(GAA)_{20-230}$, $(TTC)_{230}$ and $(TTC)_{340}$ tracts was monitored in wild-type strains. Owing to the inherent instability associated with $(GAA)_{340}$ repeats (see below), wild-type strains rapidly accumulate a mix of truncated tracts on propagation, making the 2D results unclear (data not shown). Therefore, the analysis of replication forks in this strain was carried out in Δ*msh2* background that prevents large changes in the tract size (see below).

We have found that replication stalling occurs at $(GAA)_{120}$, $(GAA)_{230}$ and $(GAA)_{340}$ but not at $(GAA)_{60}$ or TTC tracts. The inhibition zone coincides with the location of repeat tracts. These results are consistent with the previous report wherein GAA repeats arrested replication of 2 μm plasmids in a length- and an orientation-dependent manner (Krasilnikova and Mirkin, 2004). It should be noted that, in this study, the blockage zone of the Y arc is shifted causing the arc interruption. This discontinuity colocalizes with the centre of the GAA tracts. This particular migration pattern of Y intermediates might be explained by the presence of secondary structures, such as H-DNA, at the arrested forks. It is also important to note that disruption of *MSH2* does not affect

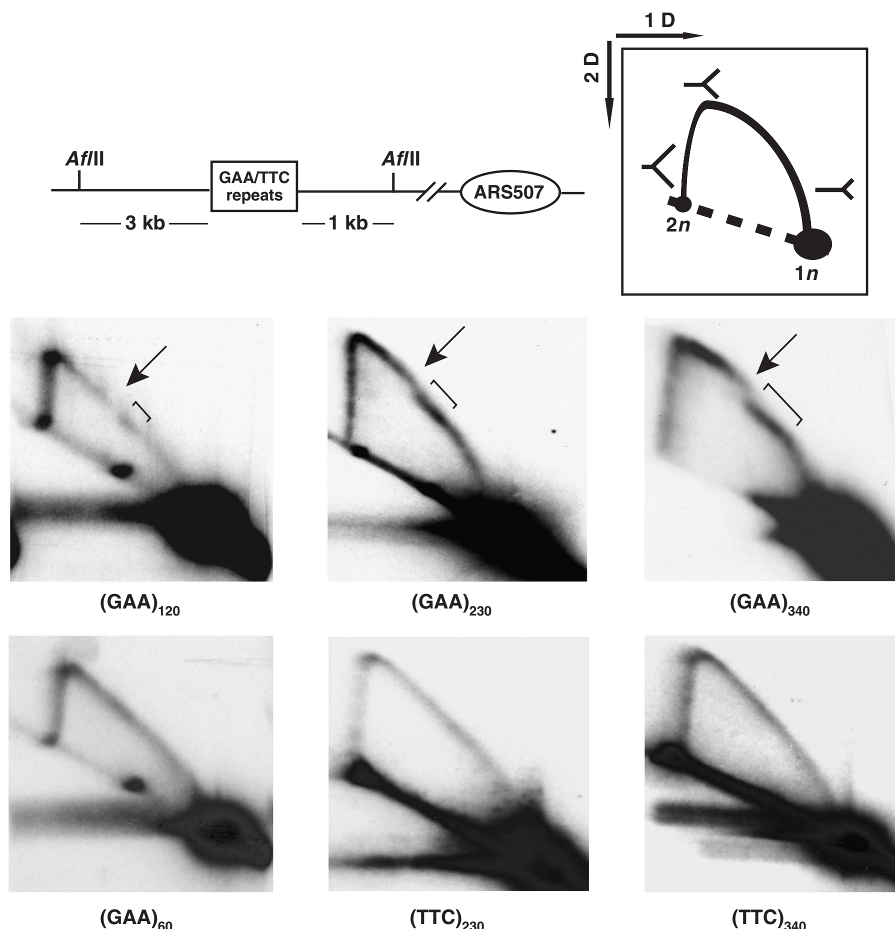

Figure 3 2D analysis of replication intermediates in strains containing GAA/TTC repeats. Neutral/neutral 2D electrophoresis was used to resolve unreplicated molecules and Y-like structures. Replication initiated at *ARS507* proceeds from right to left through the region containing the repeat tracts. Cleavage with *Afl*II positions the GAA/TTC repeats on the long shoulder of the Y-arc. The 4 kb *Afl*II-digested *LYS2* fragment was used as a probe in Southern blot hybridization. Accumulation of the replication intermediates leads to the appearance of bulges on the replication arc. Replication pausing zones are indicated by brackets. Arrows point to the Y-arc interruptions coinciding with the centre of the GAA tracts.

 ©2008 European Molecular Biology Organization

the strength of replication stalling or change the pattern of the replication intermediates in strains containing $(GAA)_{230}$ repeats (data not shown).

On the basis of these results, we suggest that the expanded GAA repeats, when present on lagging strand template, lead to the formation of triplex DNA structure that blocks the progression of the replication fork. It has been shown that the homopurine tracts are poor substrates for replication protein A (RPA) binding and for primer synthesis by the Polα–primase complex (Wold, 1997; Frick and Richardson, 2001). Hence, it is likely that GAA repeats on the template of lagging strand hinder the synthesis of Okazaki fragments, and therefore, generate long regions of single-stranded DNA, providing optimal conditions for secondary structure formation. In addition, the $R \cdot R \cdot Y$ triplex that would be formed by folding the GAA-rich strand is expected to be more stable than the $Y \cdot R \cdot Y$ conformation at the physiological pH and ionic conditions (Frank-Kamenetskii and Mirkin, 1995).

Overall, these results demonstrate that repeats in the orientation most prone for GCR and recombination also block the replication fork progression, suggesting that the induction of chromosomal instability and replication arrest are related events.

Effect of inverting the LYS2 cassette on fragility and the replication block potential of $(GAA/TTC)_{230}$ repeats

It is formally possible that the orientation-dependent bias in the fragility of the GAA repeats might be attributed to the direction of transcription in *LYS2* gene. Besides replication, transcription is another polar cellular process wherein the duplex DNA is unwound creating regions of negative superhelicity. In the GAA orientation, GAA repeats are located on the transcribed strand of *LYS2*. It has been demonstrated that the *FXN* gene expression is blocked (likely by triplex DNA) when expanded GAA tracts are on the sense strand (Campuzano *et al*, 1997; Bidichandani *et al*, 1998). In addition, it has been shown that halted transcription can attenuate replication fork progression (Krasilnikova *et al*, 1998). Hence, it is possible that replication arrest and subsequent breakage could result from defect in transcription elongation rather than from impaired lagging strand synthesis. It should be noted, in plasmid-based studies, that the GAA-induced replication block was not dependent on transcription through the repeats (Krasilnikova and Mirkin, 2004). Consistent with this study, we found that disruption of *LYS2* promoter with the *KanMX* cassette in strains containing $(GAA)_{230}$ and $(TTC)_{230}$ repeats did not affect either the GCR rates or the replication fork progression across repeat tracts (data not shown).

To directly assess whether the presence of GAA repeats on lagging strand template is responsible for the observed repeat orientation-dependent fragility, we constructed strains in which the orientation of the *LYS2* gene with $(GAA)_{230}$ was changed with respect to direction of replication from *ARS507* (Figure 4A). The flipped cassette now places the GAA repeats on the leading strand template. The arm loss events were ∼9-fold lower than in strains containing original orientation of *LYS2* cassette wherein GAA repeats are situated on the lagging strand template. It should be noted that this fold difference in arm loss events is comparable to that detected in strains with $(GAA)_{230}$ and $(TTC)_{230}$ with the *LYS2* cassette is in the original orientation. As expected, when the *LYS2* with the $(TTC)_{230}$ tract was inverted, the GCR tendency of the tract was also reversed. Moreover, the flipped GAA repeats did not compromise replication fork progression, whereas replication arrest was readily detected at the flipped TTC tracts (Figure 4B).

This observed change in GCR potential of the repeat tract on the inversion of the *LYS2* cassette demonstrates that the fragility is independent of sequences that flank the repetitive tracts, undermines the contribution of transcription and strongly implicates the role of replication.

GAA repeat size variations and fragility are dependent on the mismatch repair system

Increases in the size of repeat tracts led to elevated levels of tract length variations with deletions observed more frequently than expansions (Supplementary Table II). The most unstable are the $(GAA)_{340}$ tracts that exhibit 80% of large contractions on propagation. We found that disruptions of *MSH2*, *MSH3*, *MLH1* and *PMS1*, but not *MSH6*, genes resulted in decreased levels of large deletions in $(GAA)_{340}$ tracts; however, there was a marked increase in the levels of small deletions (Supplementary Figure 1 and data not shown). These data are consistent with previous studies in yeast where MMR deficiency was shown to cause elevated levels of small deletions and additions in tracts of repetitive DNA (for example, Sia *et al*, 1997).

Defects in MMR also strongly reduced $(GAA)_{230}$ and $(TTC)_{230}$ repeat-induced chromosomal arm loss (Figure 5A; Supplementary Table III). Strains of the Δ*msh2* or Δ*msh3*Δ*msh6* genotypes had about 15-fold reductions in the ability of $(GAA)_{230}$ tracts to trigger GCRs. Disruption of *MSH3* led to a six-fold reduction in GCR rates, whereas Δ*msh6* strains had a modest but statistically significant decrease (1.6-fold) in the level of arm loss events. GCR rates in Δ*mlh1* and Δ*pms1* strains were comparable to those observed in Δ*msh2* mutants. We also examined the effect of *msh2*-G693A, which impairs the ATPase activity of Msh2p, but not the ability of Msh2p to form complexes with other MutS proteins (Drotschmann *et al*, 1999). In addition, we assessed the effects of the *pms1*-E707K mutation, which disrupts the newly discovered endonuclease function of MutLα but not other activities of the complex (Kadyrov *et al*, 2006, 2007). Both of these point mutations reduced the frequencies of the GAA-induced GCRs to about the same extent observed in the Δ*msh2* strain. Disruption of *MSH2* also affected GCRs in TTC strains, leading to a five-fold decrease (data not shown), indicating that $Y \cdot R \cdot Y$ triplexes are also targeted by MMR.

These results indicate that MutSβ and MutLα heterodimeric complexes are required for both GAA-mediated chromosomal fragility and tract length variations, whereas the contribution of MutSα is minor. Importantly, both the ATPase function of Msh2p and the endonuclease activity of Pms1p are necessary for the induction of GCRs.

MMR triggers chromosomal breakage at $(GAA)_{230}$ tracts

Induction of GCRs in strains containing GAA/TTC tracts is likely a consequence of the DSB formation at the location of repeats. The rate of GCRs is significantly reduced in MMR-deficient strains, indicating that MMR might be responsible for the breakage. Alternatively, the defect in MMR may negatively affect the repair of the broken molecules by

©2008 European Molecular Biology Organization

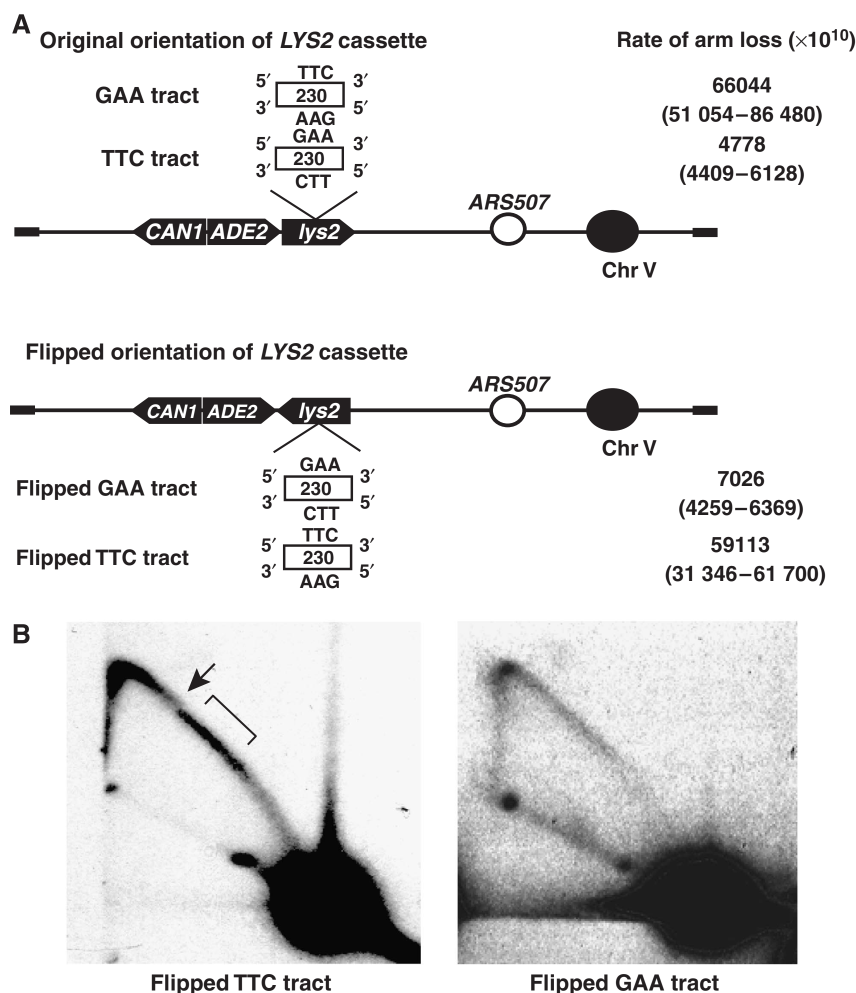

Figure 4 Induction of GCR and the ability to block fork progression by GAA/TTC repeats are affected by their orientation relative to the origin of replication. (**A**) GCR rates of original and flipped constructs. The schematic diagram of the original and the flipped *LYS2* cassette containing GAA/TTC tracts is shown on the left. The corresponding GCR rates are shown on the right. (**B**) Replication fork progression across flipped GAA/TTC tracts. 2D analysis was performed as described in Figure 3. The replication pause zone across the flipped TTC tracts is indicated by brackets.

hampering the resection of the DSB intermediates or by reducing the formation and/or extension of heteroduplex intermediates during BIR. To address this issue directly, we analysed the chromosomal DSB formation in the repeat-containing strains (Figure 5B). No DSBs were detected in strains with $(TTC)_{230}$ but were visible in strains carrying $(GAA)_{230}$ repeats (lane 3, Figure 5B) consistent with their different potential to trigger GCRs and homologous recombination (see above). The breakage in $(GAA)_{230}$ strains was compromised in Δ*msh2* and *pms1*-E707K mutants, suggesting that MMR machinery is required for efficient DSB formation and is not involved in the processing or healing of the broken ends.

Discussion

Polypurine–polypyrimidine sequences that have potential to adopt triplex secondary structure are highly polymorphic and abundant in eukaryotic genomes, ranging from yeast to humans (Cox and Mirkin, 1997). We have found that in yeast, expanded GAA/TTC tracts that belong to this class of sequence motifs strongly stimulate chromosomal fragility in a size- and orientation-dependent manner, often culminating in translocations. The MMR is a key player in the repeat-mediated breakage. This study unravels a novel function of MMR and also shows that the triplex-forming repeats can be a potent source of chromosomal aberrations similar to those observed in tumours.

Mechanism of chromosomal fragility induced by the expanded GAA/TTC tracts

In GCR and homologous recombination assays, repeats in both orientations exhibit strong breakage potential, although the fragility is more pronounced when the expanded GAA repeats are present on the lagging strand template during DNA replication (Table I; Figure 4A). Consistently, in strains with repeats in the GAA orientation, we detect a prominent replication fork arrest and accumulation of DSBs (Figures 3, 4B and 5B). One possible explanation for this orientation bias is that the purine-rich DNA template is not an ideal substrate for the proteins involved in lagging strand DNA synthesis such as RPA and Polα–primase (Wold, 1997; Frick and Richardson, 2001). Hampered Okazaki fragment synthesis would generate long single-stranded regions that could loop

 ©2008 European Molecular Biology Organization

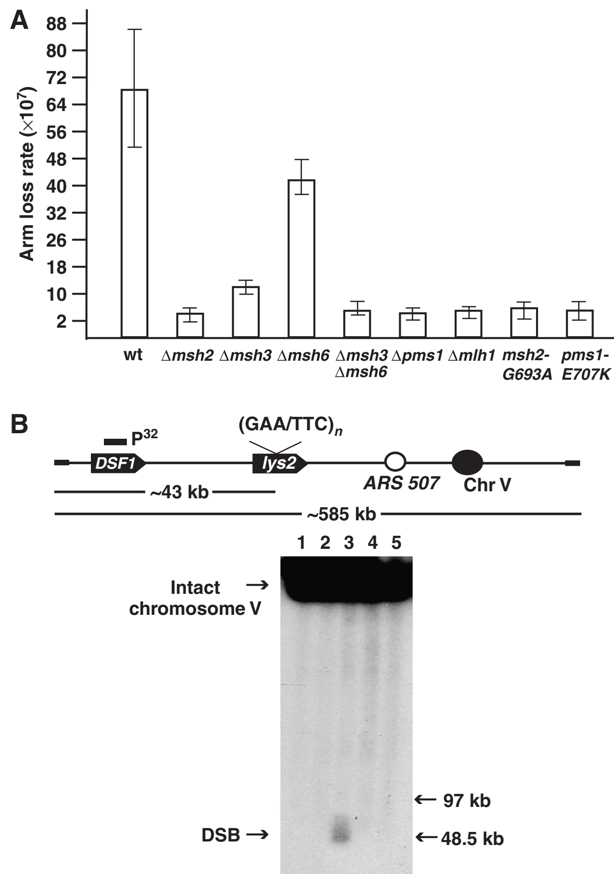

Figure 5 Chromosomal fragility at GAA/TTC tracts requires MMR. (**A**) MMR mutants strongly affect GCRs induced by $(GAA)_{230}$ tracts. Values are median rates determined in fluctuation tests using at least 14 cultures. Error bars indicate 95% confidence intervals. (**B**) Breakage of chromosome V in strains containing GAA/TTC repeats. The position of the GAA/TTC tracts on chromosome V is shown. Chromosomes were separated on the CHEF gel, transferred to a nylon membrane and hybridized with *DSF1*-specific probe to highlight the intact chromosome (∼585 kb) and broken fragment (∼43 kb). λ ladder was used as a molecular size standard shown on the right. The positions of the marker bands were determined on the ethidium bromide-stained gel prior to Southern blot hybridization. The lanes are: 1, wild-type strain with $(GAA)_{20}$; 2, wild-type strain with $(TTC)_{230}$; 3, wild-type strain with $(GAA)_{230}$; 4, Δ*msh2* strain with $(GAA)_{230}$; 5, *pms1-E707K* strain with $(GAA)_{230}$.

out and form triplexes with the double-stranded region ahead of the fork (Figure 6). Alternatively, the difference in the breakage potential could be accounted by the greater stability of the *R*·R·Y secondary structure adopted by repeats in GAA orientation versus the *Y*·R·Y triplex formed due to TTC strand folding. Hence *R*·R·Y H-DNA can be a stronger barrier for replication fork. In both of the cases, arrested fork intermediates are expected to contain the secondary structure that can explain the observed migration pattern of Y molecules in 2D gels (Figure 3).

The triplex can be recognized and targeted by MMR resulting in tract length variations and DSB formation. It should be noted that although MMR is the primary player in the fragility, Δ*msh2* does not completely eliminate the GAA/TTC-induced GCRs (Figure 5A), indicating that the triplex and/or arrested fork can lead to breakage through an alternative MMR-independent pathway. A DSB occurring within the GAA tract is expected to split chromosome V into acentric and centromere-containing fragments. There are

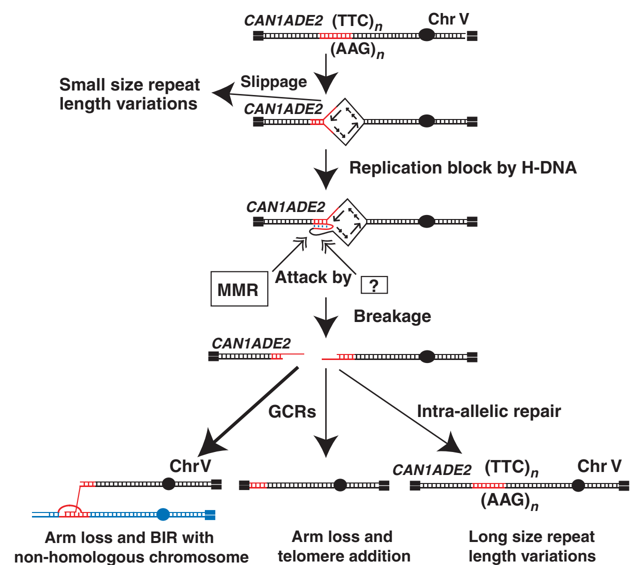

Figure 6 Model for chromosomal fragility and rearrangements mediated by triplex-forming GAA/TTC repeats. The GAA/TTC tracts are shown (not to scale) in red. Telomeres (filled rectangles) and centromeres (solid circles) are also shown. Diamonds with arrows are the bidirectional replication forks. A non-homologous chromosome is depicted in blue. GAA/TTC tracts are microsatellites that are prone to slippage during DNA synthesis. In MMR-deficient strains, this instability is manifested as small size repeat variations. We hypothesize that a triplex structure will be adopted preferentially when the GAA repeats are located on the lagging strand template. Triplex can arrest replication progression. We suggest that MMR system recognizes and processes the H-DNA leading to DSBs. DSBs can also be introduced by an alternative, MMR-independent minor pathways indicated by the boxed '?'. Following DSB induction, the broken end can be healed through intra-allelic repair (such as NHEJ or single-strand annealing) or homologous recombination with a repetitive tract on the sister chromatid, leading to large size variations. Alternatively, centromere-containing broken fragment can be repaired by BIR with GAA/TTC-rich regions on non-homologous chromosomes, resulting in arm loss and non-reciprocal translocation. Rarely, the broken end can also be capped by *de novo* telomere addition.

several pathways to repair such a break. If the broken ends are repaired by recombination with the allelic unbroken GAA tract on the sister chromatid or if the broken ends are rejoined by NHEJ or single-strand annealing, one would expect to get larger or smaller tracts without an associated translocation. Alternatively, it is possible that the acentric fragment would be lost and the centromere-containing fragment would invade GAA/TTC-rich genomic sequences located on non-homologous chromosomes. This pathway of repair would result in non-reciprocal translocations.

The function of MMR in triplex-mediated instability

Contribution of MMR to trinucleotide repeat instability has been extensively studied, in both prokaryotes and eukaryotes, for CNG tracts. In *E. coli*, defects in MMR lead to a decreased level of large deletions, but an elevated rate of small-size alterations in the hairpin-forming CTG/CAG repeats (Jaworski *et al*, 1995; Schumacher *et al*, 1998; Wells *et al*, 1998; Parniewski *et al*, 2000; Schmidt *et al*, 2000). In yeast, loss of MMR results in elevated rates of small tract alterations of CTG/CAG and CCG/CGG repeats, but has little effect on the rates of large deletions or insertions (reviewed by Lenzmeier and Freudenreich, 2003). In mice, MMR proteins are involved in regulating somatic and germline

©2008 European Molecular Biology Organization

instability of CTG/CAG repeats (promoting both expansions and contractions), MutSβ being the major player (Foiry *et al*, 2006 and references therein). *In vitro* studies show that, although Msh2p or Msh2p–Msh6p and Msh2p–Msh3p complexes efficiently bind DNA hairpins, repair does not occur (Pearson *et al*, 1997; Bowers *et al*, 2000; Owen *et al*, 2005). Owen *et al* proposed that inactive MutSβ bound to the secondary structure might prevent its processing thereby promoting tract-length changes.

We have found that similar to observations made with CTG/CAG repeats, disruptions of *MSH2*, *MSH3*, *MLH1* and *PMS1* but not *MSH6* in yeast alter the stability of $(GAA)_{340}$ repeats: decreasing the rate of large deletions and increasing the rate of small deletions (Supplementary Figure 1 and data not shown). In addition, MMR deficiency reduces the GAA/TTC-associated fragility. On the basis of these results, we suggest that besides hairpins, the triplex secondary structure might be another substrate for MMR recognition. However, unlike hairpins, H-DNA is actively processed by MutSβ and MutLα resulting in repeat size variations and DSB formation. This is strongly supported by our data that *msh2*-G693A mutants defective in the ATPase activity of Msh2 complexes or *pms1*-E707K mutants that lack the endonuclease activity of the MutLα exhibit compromised fragility and long size repeat variations.

MMR proteins process multiple DNA distortions that arise during replication, DNA repair and recombination (reviewed by Jiricny, 2006). Which feature of the triplex secondary structure is recognized by MMR? It is possible that the Hoogsteen base pairs formed between the duplex and the folded strand are a good target for MutSβ binding. Alternatively, the looped-out junction at the border between the duplex and triplex can be a substrate. Biochemical characterization of the binding and the cleavage of the defined triplex substrates by purified MMR proteins, solving the structure of MutS complexes (especially MutSβ) bound to H-DNA templates, along with the assessment of *MSH2* and *MSH3* mutants that hamper mismatch recognition might help to differentiate between these scenarios.

Along with an important role in maintaining the integrity of prokaryotic and eukaryotic genomes, MMR is also implicated in the DSB generation as a consequence of 'futile cycles of repair' in cells treated with alkylating agents or antimetabolites (reviewed by Bignami *et al*, 2003). We suggest that aberrant attempt to repair the triplex structure by MMR during replication can also culminate in DSB formation. The nature of the substrate might dictate the outcome of the MMR attack. Either extensive removal of the third strand involved in the Hoogsteen interaction or nicking of the loop region of the triplex might cause DSBs as the targeted strand lacks the complementary chain.

Eukaryotic genomes contain, besides GAA/TTC tracts, other triplex-forming homopurine·homopyrimidine mirror repeats (Cox and Mirkin, 1997). In humans, several regions that contain non-GAA H-DNA adopting sequences are hotspots for rearrangements. These include the major breakpoint cluster region at the *BCL2* (Raghavan *et al*, 2005a), intron 21 of *PKD1* (Blaszak *et al*, 1999; Patel *et al*, 2004) and promoter region of *C-MYC* (Michelotti *et al*, 1996; Wang and Vasquez, 2004). The susceptibility of these regions for aberrations was attributed to the ability of the secondary structures to impede replication progression (Raghavan *et al*, 2005b) and cause DSBs (Patel *et al*, 2004). Although the Rag1/Rag2 endonuclease was implicated in promoting DSBs at the bcl2-Mbr locus (Raghavan *et al*, 2005b), our data strongly suggest that MMR might be an additional player in the breakage formation at the location of triplex structures. It is conceivable that the mechanisms governing GAA instability might be the same for other H-DNA-adopting sequences. Hence, it would be important to assess whether MMR besides GAA/TTC triplexes, can also target non-GAA H-DNA substrates.

Implications for the stability of the human genome

We find that expanded GAA/TTC repeats in yeasts are potent inducers of DSBs and chromosomal aberrations; orientation of the tracts relative to the replication origin is an important factor governing the instability. These data suggest that the human carriers of the expanded tracts such as FRDA patients might be at risk for the formation of chromosomal aberrations. Although the chromosomal rearrangements in carriers with expanded tracts have not been reported thus far, our yeast study strongly suggests that the karyotypes of these cells should be scrutinized. It is also conceivable that triplex-forming GAA/TTC tracts can function as canonical fragile sites in cytogenetic analyses either spontaneously or on induction with chemicals that stabilize triplex structures such as polycyclic compounds (Chan and Glazer, 1997).

Understanding the molecular mechanisms that govern the stability of the eukaryotic genomes is important for studying the aetiology of cancers and hereditary diseases. On the basis of this study, we propose that chromosomal regions in human carriers containing long triplex-forming repeats are predisposed for breakage and GCRs. We suggest that the length of the repetitive tracts, their location in the genome and the genetic background may be important factors that determine the susceptibility of the individuals to tumorigenic aberrations.

Materials and methods

Strains and genetic techniques

KT119 strain (*MATa*, *his7-2*, *leu2-3,112*, *trp1-Δ*, *ura3-Δ*, *lys2-Δ*, *ade2-Δ*, *bar1-Δ*, *sfa1-Δ*, *cup1-1-Δ*, *yhr054c-Δ*, *cup1-2-Δ*, *lys2::kanM-XURA3*, *ADE2*, *CUP1* and *SFA1*) is a derivative of TP strains described in Narayanan *et al* (2006). Details of constructions of strains with varying sizes of GAA/TTC repeats, including the description of genetic techniques, are given in the Supplementary data.

Molecular biology techniques

CHEF gels, 2D analysis, Southern blot hybridization and CGH were employed to characterize genome rearrangements and to detect replication and DSB intermediates. The detailed description of these techniques can be found in the Supplementary data.

Supplementary data

Supplementary data are available at *The EMBO Journal* Online (http://www.embojournal.org).

Acknowledgements

We thank Natalya Degtyareva for critical reading of the paper and helpful discussions. This study was supported by award number R01GM0825950 from NIGMS/NIH to KSL, award number R01GM52319 from NIGMS/NIH to TDP, award number R01GM60987 from NIGMS/NIH to SMM and grants from FARA and MDA foundations to MMK. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIGMS or the NIH.

©2008 European Molecular Biology Organization

References

Al-Mahdawi S, Pinto RM, Ruddle P, Carroll C, Webster Z, Pook M (2004) GAA repeat instability in Friedreich ataxia YAC transgenic mice. *Genomics* **84:** 301–310

Bidichandani SI, Ashizawa T, Patel PI (1998) The GAA triplet-repeat expansion in Friedreich ataxia interferes with transcription and may be associated with an unusual DNA structure. *Am J Hum Genet* **62:** 111–121

Bignami M, Casorelli I, Karran P (2003) Mismatch repair and response to DNA-damaging antitumour therapies. *Eur J Cancer* **39:** 2142–2149

Bissler JJ (2007) Triplex DNA and human disease. *Front Biosci* **12:** 4536–4546

Blaszak RT, Potaman V, Sinden RR, Bissler JJ (1999) DNA structural transitions within the PKD1 gene. *Nucleic Acids Res* **27:** 2610–2617

Bowers J, Tran PT, Liskay RM, Alani E (2000) Analysis of yeast MSH2–MSH6 suggests that the initiation of mismatch repair can be separated into discrete steps. *J Mol Biol* **302:** 327–338

Campuzano V, Montermini L, Lutz Y, Cova L, Hindelang C, Jiralerspong S, Trottier Y, Kish SJ, Faucheux B, Trouillas P (1997) Frataxin is reduced in Friedreich ataxia patients and is associated with mitochondrial membranes. *Hum Mol Genet* **6:** 1771–1780

Campuzano V, Montermini L, Molto MD, Pianese L, Cossee M, Cavalcanti F, Monros E, Rodius F, Duclos F, Monticelli A (1996) Friedreich's ataxia: autosomal recessive disease caused by an intronic GAA triplet repeat expansion. *Science* **271:** 1423–1427

Chan PP, Glazer PM (1997) Triplex DNA: fundamentals, advances, and potential applications for gene therapy. *J Mol Med* **75:** 267–282

Clark RM, Bhaskar SS, Miyahara M, Dalgliesh GL, Bidichandani SI (2006) Expansion of GAA trinucleotide repeats in mammals. *Genomics* **87:** 57–67

Clark RM, Dalgliesh GL, Endres D, Gomez M, Taylor J, Bidichandani SI (2004) Expansion of GAA triplet repeats in the human genome: unique origin of the FRDA mutation at the center of an Alu. *Genomics* **83:** 373–383

Clark RM, De Biase I, Malykhina AP, Al-Mahdawi S, Pook M, Bidichandani SI (2007) 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* **120:** 633–640

Cossee M, Schmitt M, Campuzano V, Reutenauer L, Moutou C, Mandel JL, Koenig M (1997) Evolution of the Friedreich's ataxia trinucleotide repeat expansion: founder effect and premutations. *Proc Natl Acad Sci USA* **94:** 7452–7457

Cox R, Mirkin SM (1997) Characteristic enrichment of DNA repeats in different genomes. *Proc Natl Acad Sci USA* **94:** 5237–5242

De Biase I, Rasmussen A, Bidichandani SI (2006) Evolution and instability of the GAA triplet repeat sequence in Friedrich's ataxia. In *Genetic Instabilities and Neurological Diseases*, Robert RD, Ashizawa T (eds), pp 305–320. New York: Academic Press

Drotschmann K, Clark AB, Tran HT, Resnick MA, Gordenin DA, Kunkel TA (1999) Mutator phenotypes of yeast strains heterozygous for mutations in the *MSH2* gene. *Proc Natl Acad Sci USA* **96:** 2970–2975

Foiry L, Dong L, Savouret C, Hubert L, te Riele H, Junien C, Gourdon G (2006) Msh3 is a limiting factor in the formation of intergenerational CTG expansions in DM1 transgenic mice. *Hum Genet* **119:** 520–526

Frank-Kamenetskii MD, Mirkin SM (1995) Triplex DNA structures. *Annu Rev Biochem* **64:** 65–95

Frick DN, Richardson CC (2001) DNA primases. *Annu Rev Biochem* **70:** 39–80

Gacy AM, Goellner GM, Spiro C, Chen X, Gupta G, Bradbury EM, Dyer RB, Mikesell MJ, Yao JZ, Johnson AJ (1998) GAA instability in Friedreich's ataxia shares a common, DNA-directed and intraallelic mechanism with other trinucleotide diseases. *Mol Cell* **1:** 583–593

Grabczyk E, Usdin K (2000a) Alleviating transcript insufficiency caused by Friedreich's ataxia triplet repeats. *Nucleic Acids Res* **28:** 4930–4937

Grabczyk E, Usdin K (2000b) The GAA*TTC triplet repeat expanded in Friedreich's ataxia impedes transcription elongation by T7 RNA polymerase in a length and supercoil dependent manner. *Nucleic Acids Res* **28:** 2815–2822

Jaworski A, Rosche WA, Gellibolian R, Kang S, Shimizu M, Bowater RP, Sinden RR, Wells RD (1995) Mismatch repair in *Escherichia coli* enhances instability of CTG*n* triplet repeats from human hereditary diseases. *Proc Natl Acad Sci USA* **92:** 11019–11023

Jiricny J (2006) The multifaceted mismatch-repair system. *Nat Rev Mol Cell Biol* **7:** 335–346

Kadyrov FA, Dzantiev L, Constantin N, Modrich P (2006) Endonucleolytic function of MutLalpha in human mismatch repair. *Cell* **126:** 297–308

Kadyrov FA, Holmes SF, Arana ME, Lukianova OA, O'Donnell M, Kunkel TA, Modrich P (2007) *Saccharomyces cerevisiae* MutLalpha is a mismatch repair endonuclease. *J Biol Chem* **282:** 37181–37190

Krasilnikova MM, Mirkin SM (2004) Replication stalling at Friedreich's ataxia GAA*n* repeats *in vivo*. *Mol Cell Biol* **24:** 2286–2295

Krasilnikova MM, Samadashwily GM, Krasilnikov AS, Mirkin SM (1998) Transcription through a simple DNA repeat blocks replication elongation. *EMBO J* **17:** 5095–5102

Lenzmeier BA, Freudenreich CH (2003) Trinucleotide repeat instability: a hairpin curve at the crossroads of replication, recombination, and repair. *Cytogenet Genome Res* **100:** 7–24

Lobachev KS, Shor BM, Tran HT, Taylor W, Keen JD, Resnick MA, Gordenin DA (1998) Factors affecting inverted repeat stimulation of recombination and deletion in *Saccharomyces cerevisiae*. *Genetics* **148:** 1507–1524

Malkova A, Ivanov EL, Haber JE (1996) Double-strand break repair in the absence of RAD51 in yeast: a possible role for break-induced DNA replication. *Proc Natl Acad Sci USA* **93:** 7131–7136

Michelotti GA, Michelotti EF, Pullner A, Duncan RC, Eick D, Levens D (1996) Multiple single-stranded cis elements are associated with activated chromatin of the human *c-myc* gene *in vivo*. *Mol Cell Biol* **16:** 2656–2669

Mirkin SM (2007) Expandable DNA repeats and human disease. *Nature* **447:** 932–940

Napierala M, Dere R, Vetcher A, Wells RD (2004) Structure-dependent recombination hot spot activity of GAA.TTC sequences from intron 1 of the Friedreich's ataxia gene. *J Biol Chem* **279:** 6444–6454

Narayanan V, Mieczkowski PA, Kim HM, Petes TD, Lobachev KS (2006) The pattern of gene amplification is determined by the chromosomal location of hairpin-capped breaks. *Cell* **125:** 1283–1296

Ohshima K, Kang S, Larson JE, Wells RD (1996) Cloning, characterization, and properties of seven triplet repeat DNA sequences. *J Biol Chem* **271:** 16773–16783

Ohshima K, Montermini L, Wells RD, Pandolfo M (1998) Inhibitory effects of expanded GAA.TTC triplet repeats from intron I of the Friedreich ataxia gene on transcription and replication *in vivo*. *J Biol Chem* **273:** 14588–14595

Owen BA, Yang Z, Lai M, Gajek M, Badger II JD, Hayes JJ, Edelmann W, Kucherlapati R, Wilson TM, McMurray CT (2005) CAGn-hairpin DNA binds to Msh2–Msh3 and changes properties of mismatch recognition. *Nat Struct Mol Biol* **12:** 663–670

Pandolfo M (2006) Friedreich's ataxia. In *Genetic Instabilities and Neurological Diseases*, Robert RD, Ashizawa T (eds), pp 277–296. New York: Academic Press

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

Patel HP, Lu L, Blaszak RT, Bissler JJ (2004) PKD1 intron 21: triplex DNA formation and effect on replication. *Nucleic Acids Res* **32:** 1460–1468

Pearson CE, Ewel A, Acharya S, Fishel RA, Sinden RR (1997) Human MSH2 binds to trinucleotide repeat DNA structures associated with neurodegenerative diseases. *Hum Mol Genet* **6:** 1117–1123

Pollard LM, Sharma R, Gomez M, Shah S, Delatycki MB, Pianese L, Monticelli A, Keats BJ, Bidichandani SI (2004) Replication-mediated instability of the GAA triplet repeat mutation in Friedreich ataxia. *Nucleic Acids Res* **32:** 5962–5971

Raghavan SC, Chastain P, Lee JS, Hegde BG, Houston S, Langen R, Hsieh CL, Haworth IS, Lieber MR (2005a) Evidence for a triplex DNA conformation at the bcl-2 major breakpoint region of the t14;18 translocation. *J Biol Chem* **280:** 22749–22760

Raghavan SC, Swanson PC, Ma Y, Lieber MR (2005b) Double-strand break formation by the RAG complex at the bcl-2 major breakpoint region and at other non-B DNA structures *in vitro*. *Mol Cell Biol* **25:** 5904–5919

Raghuraman MK, Winzeler EA, Collingwood D, Hunt S, Wodicka L, Conway A, Lockhart DJ, Davis RW, Brewer BJ, Fangman WL (2001) Replication dynamics of the yeast genome. *Science* **294:** 115–121

Rothstein R, Michel B, Gangloff S (2000) Replication fork pausing and recombination or 'gimme a break'. *Genes Dev* **14:** 1–10

Sakamoto N, Chastain PD, Parniewski P, Ohshima K, Pandolfo M, Griffith JD, Wells RD (1999) Sticky DNA: self-association properties of long GAA.TTC repeats in R.R.Y triplex structures from Friedreich's ataxia. *Mol Cell* **3:** 465–475

Schmidt KH, Abbott CM, Leach DR (2000) Two opposing effects of mismatch repair on CTG repeat instability in *Escherichia coli*. *Mol Microbiol* **35:** 463–471

Schumacher S, Fuchs RP, Bichara M (1998) Expansion of CTG repeats from human disease genes is dependent upon replication mechanisms in *Escherichia coli*: the effect of long patch mismatch repair revisited. *J Mol Biol* **279:** 1101–1110

Sia EA, Kokoska RJ, Dominska M, Greenwell P, Petes TD (1997) Microsatellite instability in yeast: dependence on repeat unit size and DNA mismatch repair genes. *Mol Cell Biol* **17:** 2851–2858

Smith CE, Llorente B, Symington LS (2007) Template switching during break-induced replication. *Nature* **447:** 102–105

Sutherland GR (2003) Rare fragile sites. *Cytogenet Genome Res* **100:** 77–84

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

Wells RD (2008) DNA triplexes and Friedreich ataxia. *FASEB J* **22:** 1625–1634

Wells RD, Parniewski P, Pluciennik A, Bacolla A, Gellibolian R, Jaworski A (1998) Small slipped register genetic instabilities in *Escherichia coli* in triplet repeat sequences associated with hereditary neurological diseases. *J Biol Chem* **273:** 19532–19541

Wold MS (1997) Replication protein A: a heterotrimeric, single-stranded DNA-binding protein required for eukaryotic DNA metabolism. *Annu Rev Biochem* **66:** 61–92

Yabuki N, Terashima H, Kitada K (2002) Mapping of early firing origins on a replication profile of budding yeast. *Genes Cells* **7:** 781–789

©2008 European Molecular Biology Organization