Large-Scale Expansions of Friedreich’s Ataxia GAA Repeats in Yeast

Alexander A. Shishkin,1 Irina Voineagu,1 Robert Matera,1 Nicole Chemg,1 Brook T. Chernet,1 Maria M. Krasilnikova,2 Vidhya Narayanan,3 Kirill S. Lobachev,3 and Sergei M. Mirkin1,∗

1Department of Biology, Tufts University, Medford, MA 02155, USA
2Department of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park, PA 16802, USA
3School of Biology and Institute for Bioengineering and Bioscience, Georgia Institute of Technology, Atlanta, GA 30332
*Correspondence: sergei.mirkin@tufts.edu
DOI 10.1016/j.molcel.2009.06.017

SUMMARY

Large-scale expansions of DNA repeats are implicated in numerous hereditary disorders in humans. We describe a yeast experimental system to analyze large-scale expansions of triplet GAA repeats responsible for the human disease Friedreich’s ataxia. When GAA repeats were placed into an intron of the chimeric URA3 gene, their expansions caused gene inactivation, which was detected on the selective media. We found that the rates of expansions of GAA repeats increased exponentially with their lengths. These rates were only mildly dependent on the repeat’s orientation within the replicon, whereas the repeat-mediated replication fork stalling was exclusively orientation dependent. Expansion rates were significantly elevated upon inactivation of the replication fork stabilizers, Tof1 and Csm3, but decreased in the knockouts of postreplication DNA repair proteins, Rad6 and Rad5, and the DNA helicase Sgs1. We propose a model for large-scale repeat expansions based on template switching during replication fork progression through repetitive DNA.

INTRODUCTION

Expansions of simple DNA repeats are responsible for nearly 30 hereditary disorders in humans (for recent reviews, see Mirkin, 2007; Orr and Zoghbi, 2007). Normal and premutation alleles of the genes associated with repeat expansion diseases contain repeats that are either short or longer but carrying stabilizing interruptions. Generally, expansions begin if the length of a noninterrupted repetitive run exceeds a threshold of ~80–180 base pairs. After the threshold is overcome, further expansions ranging from dozens to thousands of repeats in a few generations become progressively more likely.

What are the mechanisms of repeat expansions? Studies conducted during the last decade suggest that the ability of expandable repeats to adopt DNA secondary structures predisposes them for instability during DNA replication, repair, or recombination. Specifically, strand misalignment promoted by the stable secondary structures during one of those processes is believed to be central for repeat instability (reviewed in McMurray, 1999; Mirkin, 2006; Pearson et al., 2005; Wells et al., 2005). Molecular mechanisms of the repeat expansion process are not yet understood in sufficient details and, in fact, differ in various experimental systems (Kovtun and McMurray, 2008).

A large group of data suggests that repeat expansions can occur during DNA replication (reviewed in Mirkin, 2006). Noncanonical DNA structures formed by expandable repeats stall DNA polymerases in vitro, resulting in misalignment of repetitive DNA strands and expansions (Gacy et al., 1998; Oshima and Wells, 1997; Usdin and Woodford, 1995). Various expandable repeats also stall the replication fork progression in vivo (Krasilnikova and Mirkin, 2004; Pelletier et al., 2003; Samadashwily et al., 1997; Voineagu et al., 2009), when the length of the repetitive run approaches the expansion threshold. The stability of expandable repeats also appears to depend on their orientation and their distance relative to the replication origin (Clery et al., 2002; Freudenreich et al., 1998; Kang et al., 1995; Miret et al., 1998; Rindler et al., 2006). Furthermore, frequencies of repeat expansions and contractions in model organisms are affected by mutations in genes that encode for replication proteins (reviewed in Mirkin, 2006).

Another group of data points to the role of DNA repair in the expansion process (reviewed in Kovtun and McMurray, 2008; Lahu and Slater, 2003). Expandable repeats appeared to be more stable in MMR-deficient E. coli strains (Jaworski et al., 1995). Furthermore, inactivation of the MSH2 or MSH3 genes markedly decreased the frequency of repeat expansions in transgenic mice models for Huntington’s disease and myotonic dystrophy (Kovtun and McMurray, 2001; Manley et al., 1999; Savouret et al., 2004; van den Broek et al., 2002). These unexpected results could be explained by the fact that the Msh2/Msh3 complex can get trapped in the mismatched hairpins formed by expandable repeats instead of repairing them (Owen et al., 2005). In dividing cells, this could lead to the stabilization of repetitive slip-outs left in the nascent DNA strands, resulting in subsequent repeat expansions. In nondividing cells, similar slip-outs could be formed during the repair of repetitive runs triggered, for example, by the removal of an oxidized guanine by 8 oxoguanine DNA glycosylase (Kovtun et al., 2007). DNA molecules carrying such repetitive slip-outs could eventually convert into a set of products with differentially expanded repeats (Panigrahi et al., 2005). Other DNA repair pathways, such as postreplication repair (Daee et al., 2007) and nucleotide excision repair (Jaworski et al., 1995), were also linked to repeat expansions.

Finally, repeat expansions can occur during the process of homologous recombination (reviewed in Wells et al., 2005). In
bacteria, expandable repeats increase the rate of recombination undergoing length changes in the process (Jakupciak and Wells, 1999; Napierala et al., 2002). In yeast, double-stranded breaks occur at the site of long (CAG)n-(CTG)n runs during mitotic and meiotic divisions (Freudenreich et al., 1998; Jankowski et al., 2000; Nag et al., 2004). Repair of the breaks via the synthesis-dependent strand annealing between sister chromatids could result in repeat instability (Richard et al., 2000, 2003).

Despite a wealth of data and hypotheses, the fine mechanisms of repeat expansions remain elusive. A major problem in unravelling these mechanisms is the lack of a controllable genetic experimental system to study large-scale expansions, similar to those observed in human pedigrees. One previously described yeast experimental system (Miret et al., 1998) worked with expansions of 12–25 triplet repeats—the length applicable to premutation size alleles for a subset of triplet repeat diseases. In another yeast system, larger expansions of a long CAG repeat were observed, but only during HO-induced gene conversion (Richard et al., 2000). Mouse models of repeat expansions utilize even longer repeats, but the scale of their expansions is quite modest (Koivun and McMurray, 2001; Lia et al., 1998; Manley et al., 1999) with just one exception (Gomes-Pereira et al., 2007).

To bridge this gap, we have developed an experimental system in yeast that allows selection of large-scale expansions that generate disease-size repeats. Here, we have utilized this system for studying expansions of (GAA)n repeats that are responsible for Friedreich’s ataxia, the most common hereditary ataxia in humans (reviewed in Pandolfo, 2002; Wells, 2008). In this system, 78–150 GAA repeats readily expanded to up to 450 repeat units. Unexpectedly, expansion rates did not depend on the replication fork stalling within the repeat tract. The first round of genetic analysis was performed by knocking out homologous recombination and double-strand DNA break repair proteins Rad52 and Rad50 (Krogh and Symington, 2004), the key mismatch repair protein Msh2 (Harfe and Jinks-Robertson, 2000), replication fork-stabilizing proteins Top1 and Csm3 (Calzada et al., 2005; Nedelcheva et al., 2005), RecQ (Khakhhar et al., 2003), and Pif1 (Boule and Zakian, 2006) families of DNA helicases implicated in the maintenance of genome integrity, as well as postreplication repair proteins Rad6 and Rad6 (Andersen et al., 2008). It appeared that proteins involved in homologous recombination or DNA mismatch repair had little, if any, effect on repeat expansions. The rates of GAA repeat expansions were strongly increased in the Top1 and Csm3 knockouts and decreased in the Sgs1, Rad5, and Rad6 knockouts. Thus, fork-stabilizing proteins precluded, whereas proteins involved in template switching and fork restart promoted, repeat expansions. Altogether, these data led us to propose a new model for repeat expansions based on the template switching during the replication fork progression through repetitive DNA runs.

RESULTS

A Yeast System to Select for Large-Scale Repeat Expansions

The starting idea for the development of the experimental system for large-scale repeat expansions came from the study of Yu and Gabriel, who showed that lengthening the S. cerevisiae ACT1 gene intron beyond ~1200 bp blocks RNA splicing (Yu and Gabriel, 1999). We reasoned that the large-scale expansions of a repeat positioned within an intron of the yeast reporter gene would lead to its inactivation, which can be monitored by selecting for the reporter’s loss of function. A 308 bp long ACT1 intron with varying lengths of (GAA)n-(TTC)n repeats was inserted into the URA3 gene. The 52, 78, 100, 125, and 150 homogeneous repeats were placed in such a way that the GAA runs appear in the sense strand for transcription, mimicking the situation in the human FXN gene (Campuzano et al., 1996). Even 150 noninterrupted (GAA)n repeats did not inactivate the URA3 gene expression because the lengths of introns carrying the longest repeats did not exceed 850 bp, which is below the threshold length (~1200 bp) of splicing inactivation (Yu and Gabriel, 1999). The URA3 cassettes were integrated in two orientations on chromosome III ~1 kb downstream of the active ARS306 replication origin (Figure 1A). In this system, the URA3 inactivation can result from repeat expansions, point mutations, deletions, or insertions, all of which can be monitored by selecting Ura− clones on 5-fluoro-orotic acid (5-FOA) medium (Boeke et al., 1987).

To assess the contribution of repeat expansions to URA3 inactivation, we analyzed the length of the (GAA)n repeats in the 5-FOA8 clones by PCR with the primers shown in Figure 1A. An example of results from a typical experiment for the expansion of the (GAA)100 repeat are shown in Figure 1B. A majority of the clones carried repeat expansions, some of them as long as 300 copies (Figure 1B) and occasionally up to 450 copies (data not shown). Besides expansions, two other types of events were observed. In some clones, the repeat length did not change. DNA sequencing revealed that these Ura− clones contained small deletions, insertions, or point substitutions in URA3 ORF outside of the intron (Figure 1C). Interestingly, these mutations occurred at significant distances from both ends of the GAA run, ranging from 270 to 513 bp. This observation is consistent with the previously reported mutagenesis at a distance associated with triplex-forming DNA sequences (Wang and Vasquez, 2004, 2006). In another group of clones, no PCR products were detected, suggesting that they carry deletions of genome sequences, including the region to which our PCR primers annealed. Using Southern blot analysis of chromosomal DNA, we have confirmed that these clones indeed contained intersitial 1–6 kb long deletions that included the URA3 gene and the flanking sequences (Figure 1D).

Rates of Expansions and Size Distributions of Expanded Repeats

As discussed above, three different mutational events were observed among the 5-FOA8 clones. In the repeat orientation wherein the homopyrimidine run is on the lagging strand DNA template, the rates for all three types of mutational events increased exponentially with the repeat length (Figures 2A and 2B). An alteration in the repeat’s length from 78 to 150 units led to an elevation in the rate of mutations by 10-fold, whereas the rate of interstitial deletions increased by 60-fold. At the same time, the rate of expansions augmented by three orders of magnitude. The exponential increase in the probability of repeat to expand in our system is quite similar to what is observed in human pedigrees (Fu et al., 1991).
The analysis of the lengths of expansions in the case of 100 and 150 repeats revealed the selection threshold of the experimental system to be 170–180 repeats (Figure 2C). Supporting this conclusion, all eight detected expansions of the (GAA)$_{78}$ were within the 170–190 repeat range (data not shown). When 100 GAA repeats expanded, we observed what looks like the right half of the normal length distribution curve, likely because our selection cuts off the left half of the curve. In contrast, for 150 copies of the GAA repeat, we detected a normal length distribution of the expanded repeats with a mean length of 220 copies, which is significantly longer than the selection threshold. We believe that the experiment with 150 repeats provided us with the unbiased view of repeat expansions, as even relatively short expansions brought yeast over the selection cutoff. On average, 70 repeats were added to the starting (GAA)$_{150}$ repeat, expanding it ~1.5 times. This result points to the existence of an expansion increment corresponding to

Potential problem, we have added either a 300 bp long nonrepetitive sequence or a 150 bp long nonrepetitive sequence to the introns of our cassettes with 50 and 100 GAA repeats, respectively, thus balancing the overall intron length between themselves and the previously described cassette carrying 150 repeats. Figure 3A compared expansion rates between the new and the previously studied constructs.

Evidently, equilibration of the intron sizes did not eliminate the dependence of the expansion rates on the repeat’s length; we still do not detect expansions for the intron-balanced (GAA)$_{50}$ repeat, and the expansion rate for the intron-balanced (GAA)$_{100}$ repeat is still ~10-fold lower than that for the (GAA)$_{150}$ repeat. At the same time, the expansion rate for the intron-balanced (GAA)$_{100}$ cassette is increased 10-fold compared to the original (GAA)$_{100}$ cassette. The latter difference is due to the change in length distribution of expanded repeats between the two cassettes: the selection cutoff is 150 repeats for the

Figure 1. Selectable System to Detect Large-Scale GAA Repeat Expansions in Yeast

(A) Selectable cassettes for repeat expansions in two orientations in chromosome III. (Gray area) Split URA3 gene. (Hatched area) ACT1 intron. (White rectangle) GAA repeat. (Arrows) PCR primers.

(B) Characteristic results of the PCR analysis of 5-FOA$^+$ clones originated from a cassette with 100 GAA repeats, showing expansions, mutations, and deletions. Black arrow designates electrophoretic mobility of the PCR product from the original (GAA)$_{100}$ repeat.

(C) Schematic representation of mutations and small-scale deletions in 5-FOA$^+$ clones. (Stars) Point mutations. (Triangle) A microinsertion. (Gapped lines) Small deletions, with sequence microhomologies shown in parentheses.

(D) Mapping of large deletions using Southern blot hybridization.

- $C_{24}$
intron-balanced cassette versus 170–180 repeats for the original cassette (Figure 3B).

Expansions of the (GAA)100 repeat up to 170 copies—a selection cutoff in the original, unbalanced cassette—create a 915 bp long intron. Expansions of the same repeat to the selection cutoff of 150 repeats in the balanced cassette make a 1010 bp long intron. Finally, expansions of the original (GAA)150 repeat to the selection cutoff give rise to a 920 bp long intron. We believe, therefore, that expression of the chimeric URA3 gene in our system is shut down when its intron gets longer than ~900–1000 bp (see also next section).

Mechanisms of the Reporter Gene Inactivation upon GAA Repeat Expansions

In the original Yu and Gabriel study, an increase in the ACT1 intron length beyond ~1200 bp inhibited its splicing, thus blocking the reporter gene expression (Yu and Gabriel, 1999). In our case, expression of the chimeric URA3 gene was shut down when the repeat-containing intron became longer than 920 bp, i.e., at a significantly shorter length. These differences could be attributed to either intrinsic differences between the two independent systems or to the specific effect of the GAA repeat on gene expression. For example, several lines of evidence suggest that expanded repeats inhibit transcription elongation in vitro and in vivo (Bidichandani et al., 1998; Grabczyk and Usdin, 2000; Ohshima et al., 1998; Patel and Isaya, 2001). It was also suggested that extended GAA repeats cause aberrant splicing in HeLa cells (Baralle et al., 2008).

To distinguish between these possibilities, we measured the levels of spliced and unspliced URA3 mRNA using semiquantitative RT-PCR. To measure the amount of mature mRNA, RT-PCR was carried out with primers 1 and 2 (Figure 4A) because primer 2 can only anneal to the properly spliced mRNA. Figures 4B and 4C show that expansion of the GAA repeat leads to a gradual decrease in the amount of URA3 mRNA, dropping below 10% of the control level, when its length exceeded 180 units corresponding to an intron length exceeding 980 bp. (Note that a somewhat higher mRNA level at 200 repeats compared to 180 repeats is likely due to hard to control repeat contractions during cell growth prior to RNA

<table>
<thead>
<tr>
<th>Repeat Number</th>
<th>Rate of expansions x10^-7</th>
<th>Rate of mutations x10^-7</th>
<th>Rate of deletions x10^-7</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>N/A</td>
<td>0.45</td>
<td>0.19</td>
</tr>
<tr>
<td>52</td>
<td>N.D.</td>
<td>1.4</td>
<td>0.24</td>
</tr>
<tr>
<td>78</td>
<td>1.2</td>
<td>4.5</td>
<td>2.5</td>
</tr>
<tr>
<td>100</td>
<td>13</td>
<td>6.5</td>
<td>6.0</td>
</tr>
<tr>
<td>125</td>
<td>350</td>
<td>18</td>
<td>45</td>
</tr>
<tr>
<td>150</td>
<td>1200</td>
<td>48</td>
<td>150</td>
</tr>
</tbody>
</table>

Figure 2. Rates and Length Distributions of Expanded Repeats

(A) Rates and 95% confidence intervals of expansions, mutations, and deletions for the GAA repeats of varying lengths.

(B) Dependence of expansion rates on repeat lengths. (Red diamonds) Expansions. (Blue circles) Mutations. (Green triangles) Large deletions.

(C) Length distributions of expanded repeats among 5-FOA clones. (Black bars) Expansions of the (GAA)100 repeat. (Orange bars) Expansions of the (GAA)150 repeat.
To determine the amount of unspliced \textit{URA3} RNA, we treated RNA samples extensively with the RNase-free DNase I followed by RT-PCR with primers 3 and 4 (Figure 4A), in which primer 3 can only anneal to the intron sequence. The amount of unspliced \textit{URA3} mRNA increased upon repeat lengthening (Figures 4B and 4C). Figure 4C shows the juxtaposition of the amounts of spliced and unspliced \textit{URA3} RNA carrying GAA repeats of varying lengths. The sum total of the two values remains unchanged in the wide range of repeat lengths; thus, a decrease in the amounts of spliced \textit{URA3} mRNA is likely due to inefficient splicing. We conclude that expansions of GAA repeats in our yeast system ultimately result in blockage of RNA splicing, rather than transcription elongation. This is consistent with our previous observations showing that up to 230 GAA repeats do not block transcription elongation in yeast (Krasilnikova and Mirkin, 2004). Combining these data with the results discussed in the previous section, we conclude that the blockage of RNA splicing is largely caused by an increase in the intron’s overall length, and the length of the repeat per se contributes to this blockage to a smaller extent.

**Dependence of Expansion Rates on the Repeat’s Orientation Relative to the Replication Origin**

We have previously shown that carrier- and disease-size (GAA)$_n$ repeats, when either in a yeast plasmid or on chromosome V, stalled the replication fork progression in an orientation-dependent way: when the homopurine run was on the lagging strand DNA template (Kim et al., 2008; Krasilnikova and Mirkin, 2004). To determine whether there is a link between the replication stalling and the propensity of a repeat to expand in our selectable system, we assessed the replication progression across the (GAA)$_{100}$ repeats and rates of their expansions when those repeats were placed in two orientations relative to the replication origin. In the direct orientation of our cassette, the GAA run was on the lagging strand template, whereas it was on the leading strand template in the inverted orientation of the cassette. Using two-dimensional gel electrophoresis of chromosomal replication intermediates, we observed the same phenomenon that we previously described for the plasmid template: the GAA/TTC repeat stalled the replication fork progression in a strictly orientation-dependent manner only when the homopurine run was on the lagging strand template (Figure 5A). At the same time, the rate of expansions was barely (1.5-fold) higher in the repeat’s orientation associated with the replication fork stalling than in the opposite orientation (Figure 5B). A similar tendency was also observed for 150 GAA repeats (data not shown). Even this minor elevation in the expansion rate for the directly oriented cassette could, in fact, be due to the somewhat lower level of the \textit{URA3} mRNA in this orientation compared to the opposite orientation (data not shown), lowering the selection pressure.

In a separate study, we have found that long GAA repeats caused chromosomal fragility, induced gross chromosomal rearrangements, or underwent frequent contractions in the orientation responsible for the replication stalling in yeast (Kim et al., 2008). The likelihood of all of these events depended on the activity of the DNA mismatch repair system. It was foreseeable, therefore, that the actual rate of repeat expansions in the orientation of our cassette associated with the replication stalling could be masked by the high contraction rate. To address this concern, we compared (GAA)$_{100}$ expansion rates in this orientation in the wild-type strain with \textit{Dmsh2} mutant, in which the rate of repeat contractions was drastically decreased (Kim et al., 2008). It appeared that the absence of the Msh2 protein only marginally (1.5-fold) affected the expansion rate in this orientation of the GAA repeat tract (data not shown). We believe, therefore, that large-scale expansions of the GAA repeat are not directly linked to the replication fork stalling.

<table>
<thead>
<tr>
<th>Starting Repeat</th>
<th>$\mu_{\text{expansion}} \times 10^{-7}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(GAA)$_{20}$</td>
<td>N.D.</td>
</tr>
<tr>
<td>(GAA)$_{20} + 300$ bp</td>
<td>N.D.</td>
</tr>
<tr>
<td>(GAA)$_{100}$</td>
<td>13**</td>
</tr>
<tr>
<td>(GAA)$_{100} + 150$ bp</td>
<td>11 – 16</td>
</tr>
<tr>
<td>(GAA)$_{200}$</td>
<td>137**</td>
</tr>
<tr>
<td></td>
<td>900 – 1,600</td>
</tr>
</tbody>
</table>

*** Differences with $p < 0.001$ in Mann-Whitney U-test.
To get an insight into the mechanisms of GAA repeat expansions, we conducted a first-round screening of yeast mutants to establish the impact of various DNA transactions in this process. We compared the expansion rates for the GAA100 repeats in our wild-type strain with those in the individual knockouts for homologous recombination and double-strand DNA break repair proteins (Krogh and Symington, 2004), the mismatch repair proteins (Harfe and Jinks-Robertson, 2000), replication fork-stabilizing proteins (Calzada et al., 2005; Nedelcheva et al., 2005), RecQ and Pif1 families of DNA helicases (Boule and Zakian, 2006; Khakhar et al., 2003), and the postreplication repair proteins Rad5 and Rad6 (Andersen et al., 2008).

Because we did not see much difference between the two orientations of the repeat in the expansion rates, these analyses were carried out for just the inverted orientation of our cassette (Figure 1A).

The data on the rates of expansions, mutations, and deletions in these mutants are presented in Table 1. Due to the intrinsic noise in our system, we took a conservative approach to consider only those differences that exceeded 2-fold and had the p value of less than 0.001.

The expansion rates were elevated in the Tof1 and Csm3 knockouts: 6- and 4-fold, respectively. Expansion rates were decreased ~3.5-fold in the Sgs1, Rad6, and Rad5 knockouts. We have further confirmed the effects of these mutations on the expansions of 125 GAA repeats (data not shown). At the same time, deletion of either the RAD52 or RAD50 genes did not affect expansions, strongly suggesting that homologous recombination and/or recombinational fork restart are not involved in this process. The Msh2 inactivation led to a small, if any, increase in the expansion rates, arguing against the significant role of the mismatch repair in the process. Finally, Pif1-like DNA helicases Pif1 and Rrm3, as well as RecQ DNA helicase Srs2, had little, if any, effect on expansions.

Intriguingly, the rates of interstitial deletion formation were not significantly changed in mutants that affected repeat expansions. These deletions were completely absent, however, in the Rad52 or Msh2 knockouts. They were also decreased 4-fold in the Srs2 knockout. As for mutations, their rates were strongly elevated only in the Δmsh2 (40-fold) and the Δrad52 (10-fold) strains, in accordance with their known mutator phenotypes (Tran et al., 1997; Von Borstel et al., 1971).

**DISCUSSION**

We have developed a convenient experimental system to analyze large-scale repeat expansions in yeast. Differently from previously described assays, it allows one to monitor expansions of the premutation-size (78–150 copies) repeats well into the disease range (200–450 copies), providing a unique opportunity to study characteristics and genetic controls of large-scale expansions. This study investigates the mechanisms and consequences of expansions of (GAA)n repeats, which are responsible for Friedreich’s ataxia, the most common hereditary ataxia in humans (Pandolfo, 2002). In the other experimental yeast system that monitors expansions of up to 25 copies of triplet repeats...
increase in repeat length (Figure 2). Specifically, doubling the size of the (GAA)\textsubscript{n} repeat from 78 to 150 copies led to a threefold increase in the propensity of a repeat to expand. This discrepancy could be due to the fact that our earlier observations were made primarily for small repeats, whereas our current system monitors large expansions of shorter GAA repeats on a yeast chromosome.

The propensity to expand increases exponentially with an increase in repeat length (Figure 2). Specifically, doubling the size of the (GAA)\textsubscript{n}, repeat from 78 to 150 copies led to a threefold increase in the propensity of a repeat to expand. This dramatic difference can be explained by: (1) a length-dependent increase in the propensity of a repeat to expand, (2) a decrease in the number of expansion steps required for longer repeats to reach the selection threshold, or (3) the combination of both factors. Analysis of expansions of 150 GAA repeats hinted to the existence of an incremental step of the expansions, corresponding to roughly 1.5 times the repeat size. If it is applicable to repeats of other lengths, short GAA repeats would need two or more steps to reach the selection cutoff. We suspect that this is the likely explanation of their much lower expansion rates compared to the WT strain.

One of the unexpected observations in our study was the lack of orientation dependence for the GAA repeat expansions. Similar to what was reported previously (Krasilnikova and Mirkin, 2004), the replication fork stalled at the repeat in one orientation only, when the GAA run was a part of the lagging strand template (Figure 5A). However, the rates of large-scale expansion in both orientations were quite similar (Figure 5B). These results are different from our previous data that expansions of long GAA repeats depended on their orientation within a plasmid (Krasilnikova and Mirkin, 2004). This discrepancy could be due to the fact that our earlier observations were made primarily for small expansions of a much longer GAA repeat (280 copies) positioned within a multicopy plasmid, whereas our current system monitors large expansions of shorter GAA repeats on a yeast chromosome.

The first-round genetic screening gave us important clues to the mechanisms of repeat expansions. There were no differences in the expansion rates between the wild-type and Rad52 knockout strains, whereas the rate of expansions in hyperrecombinogenic \textit{Jsgs1} mutant (Watt et al., 1996) decreased. Therefore, the involvement of genetic recombination in the large-scale expansions of GAA repeats can be ruled out, contrary to what was described in a bacterial system (Napierska et al., 2004). We also did not see much effect of the Msh2 protein on the GAA repeat expansions both in this and previous (Kim et al., 2008) studies, making the mismatch repair system an unlikely player in the repeat expansion process.

Disruption of the \textit{TOF1} or \textit{CSM3} genes led to a strong elevation of the expansion rates. These genes encode components of the so-called replication-pausing complex (Tof1-Mrc1-Csm3) (Katou et al., 2003) that prevents the replication fork stalling caused by hydroxyurea treatment (Calzada et al., 2005; Nedelcheva et al., 2005). DNA damage (Foss, 2001), or unusual DNA structures (Voineagu et al., 2008) by averting uncoupling of the replicative DNA helicases from stalled forks (Nedelcheva et al., 2005). In addition, Tof1p facilitates the replication fork pausing at protein-mediated barriers (Mohanty et al., 2006).
In contrast, disruption of the SGS1, RAD6, or RAD5 genes inhibited repeat expansions. The SGS1 gene encodes a 3’-to-5’ DNA helicase homologous to the human Bloom’s syndrome DNA helicase (Gangloff et al., 1994). SGS1 mutants are hypersensitive to UV light and hydroxyurea (Chakraverty et al., 2001) and display hyperrecombination phenotype (Watt et al., 1996). The Sgs1 protein was implicated in the restart of stalled replication forks (Torres et al., 2004) and in the repair of defects accumulated during the lagging strand synthesis (Li and Brill, 2005). Rad6 is a ubiquitin-conjugating enzyme that acts in a complex with Rad18 ubiquitin ligase to regulate DNA damage tolerance pathway in yeast (reviewed in Lawrence, 1994). The latter includes a Rad5-dependent template-switching branch and the translesion DNA synthesis branch (Andersen et al., 2008). Rad5, a member of the SWI/SNF family, has ATPase and E3 ubiquitin ligase activities (Klein, 2007). Its ATPase activity is stimulated by the presence of branched DNA structures, which triggers a helicase-like reaction of template switching and/or fork regression (Blastyak et al., 2007).

How do our results compare with the other studies on repeat expansions? In the best-characterized yeast system that dealt with smaller-scale expansions of CAG repeats, Tof1 inactivation led to a 7-fold increase in the expansion rate (Razidlo and Lahue, 2008), which is quantitatively similar to our observations. We conclude, therefore, that the activity of Tof1p universally opposes small- and large-scale expansions of various triplet repeats. At the same time, our data with Sgs1, Srs2, and Rad6 knockouts contrast previous observations: disruption of the Sgs1 helicase did not affect CAG expansions (Bhattacharyya and Lahue, 2004), whereas inactivation of the Srs2 DNA helicase (Bhattacharyya and Lahue, 2004) or Rad6 repair pathway (Daee et al., 2007) resulted in a dramatic increase in CAG expansions. These profound differences could either reflect separate mechanisms for the large- and small-scale expansions or could be due to different structural features of CAG and GAA repeats. Future studies of CAG repeat expansions in our system could distinguish between these scenarios.

Altogether, our data left us with the following paradox. On one hand, all genes affecting GAA expansions in our system, including Tof1, Csm3, Rad6, Rad5, and Sgs1, are implicated in fork stabilization, reversal, or restart. On the other hand, we do not see the link between the repeat-mediated replication fork stalling and their propensity to expand. We hypothesize that the model, loosely based on the template-switching mechanism proposed in Goldfless et al. (2006), could resolve this paradox. We propose that, during replication of a repetitive DNA run (Figure 6A), a leading strand DNA polymerase can accidentally (~10^-3 per replication) switch its template to continue DNA synthesis along the nascent lagging strand (Figure 6B). Notably, in a long repetitive run, each sequence in the nascent lagging strand sequence is repeated multiple times in the leading strand template. This could make the template switch more feasible, compared to the unique DNA sequences, as an unwound portion of the repetitive leading strand can pair with multiple points along the repetitive lagging strand. After reaching the end of the Okazaki fragment (Figure 6C), the polymerase should switch back to its primary leading strand template in order for replication to continue. This results in an expanded repetitive run within the leading DNA strand (Figure 6D). These reactions would likely depend on the activities of the 3’-to-5’ DNA helicases, such as Sgs1, and the template switcher Rad5. The Tof1/Csm3/Mrc1 fork-stabilizing complex, in contrast, would be expected to block template switching. Our data are in agreement with these predictions.

In general, this model assumes that the maximum size of the one-step repeat expansion should be less than or equal to an Okazaki fragment. The biochemical measurement of the sizes of Okazaki fragments in eukaryotes showed that they vary significantly: from 40–290 nt (Anderson and DePamphilis, 1979; Raschle et al., 2008). Our average expansions (50–70 triplet repeats) are within these ranges. How can we explain bigger-size expansions also observed in our experiments? We believe that an expansion cycle presented in our model could recur more than once, particularly when the repeat’s size exceeds that of an Okazaki fragment. This could account for the largest-scale expansions observed in our hands, as well as for the so-called catastrophic expansions observed for GAA repeats in human pedigrees (Montermini et al., 1997).

Importantly, this rare sequence of events should not be linked to the much more frequent fork stalling at the repeat. The latter is due to the formation of the triplex DNA structure, when the homopurine run is situated on the lagging strand template during DNA replication (Krasilnikova and Mirkin, 2004). We have previously found that such stalling results in repeat contractions, chromosomal fragility, and chromosomal rearrangements mediated by the mismatch repair machinery (Kim et al., 2008). In the current study, we observed that the formation of interstitial deletions, but not expansions, depended on the presence of functional Msh2 and Rad52 proteins. Overall, we believe that interstitial deletions result from the double-stranded breaks at GAA repeats mediated by the MMR proteins followed by their repair via single-strand annealing (SSA).
We are fully aware that the proposed model for repeat expansions is by no means final, and future studies including full-genome screening are needed to gain a better insight in the mechanisms of this process. The immediate conceptual advance of our model is that it is applicable for a variety of DNA repeats, notwithstanding of their specific secondary structures. This could nicely explain how similar expansion principles apply to such structurally different repeats, including quadruplex-forming CGG repeats, hairpin-forming CTG repeats, triplex-forming GAA repeats, or DNA-unwinding ATTCT repeats. At the same time, not every repetitive sequence is known to expand (McMurray, 1999). This could be explained if a propensity for the template switching also depends on certain biophysical properties of repetitive DNAs, such as the “stickiness” characteristic of long (GAA)n repeats (Sakamoto et al., 2001).

EXPERIMENTAL PROCEDURES

Making of Selectable Cassettes and Yeast Strains

Cloning of the (GAA)228 repeat has been previously described in Krasilnikova and Mirkin, 2004. We first inserted a 308 bp long ACT1 intron into the StuI site of the yeast URA3 gene within the pYES2 plasmid. The (GAA)228(TTC)228 repeat with the MCS flanks was excised from the pYES-GAA228 plasmid (Krasilnikova and Mirkin, 2004) using EcoRI followed by cloning into the unique MunI site of the intron, such that the GAA run was in the sense strand for transcription.

To create a collection of GAA repeats of varying lengths in the intron-containing URA3 gene, we transformed the plasmid pYES-Int-GAA228 into the S. cerevisiae CH1585 (MATa leu21 trp63 ura3-52 his3-200) strain using Ura+ selection. Because GAA repeats readily contract under selective pressure, plasmids with 52, 78, and 100 GAA repeats were isolated from yeast Ura+ clones and retransformed into the E. coli DH5alpha strain (Invitrogen). The resultant cassettes with varying lengths of repeats were integrated, using PCR-based integration with flanks of homology, in two orientations—rightwards from the ARS306 replication origin in chromosome III of the CH1585 strain via Ura+ selection, replacing the region corresponding to positions 75,423–75,715 on the third chromosome. Primers 7 and 8 (Table S2) were used (Table S2). Correct integration was confirmed by PCR with primers 13 and 15 (for direct orientation) or primers 14 and 15 (for invert orientation) (Table S2).

Rates of Mutational Events

Rates of mutational events were determined by the method of mutant accumulation (Drake, 1991). Large colonies (~106 cells) grown for 75 hr on complete media (YPD) plates were dissolved in 1 ml of sterile H2O. The cell suspension was serially diluted and plated on 5-FOA media or YPD media and incubated for 72 hr at 30°C. Approximately 50 5-FOA-resistant clones from each experiment were analyzed by PCR (see Supplemental Data) to determine the relative ratio of expansions, mutations, and deletions. These ratios were then used to approximate frequencies for each event in individual experiments. The median rate and 95% confident interval (CI) of each event was then calculated using the formula \( f = \ln(N) / \ln(n) \), in which N was the number of cells in an individual YPD-grown colony and f was the mutation frequency in this colony (Drake, 1991) as described in Lobachev et al. (2002).

Electrophoretic Analysis Replication Intermediates

Replication intermediates were analyzed as follows. Early logarithmic culture (500 ml) was synchronized with a factor (Zymo Research) to a final concentration of 0.1 μM (all strains used for isolation of replication intermediates were Jbart). After 2 hr, cells were released into fresh YPD medium. Cell growth was stopped at 20 min postrelease by adding NaN₃ to a final concentration of 0.1% w/v. From this point, the protocol in Wu and Gilbert, 1995 was followed, using a G/100 QIAGEN column. Isolated DNA (10 μg) was digested with 300 U of Dral. The electrophoresis conditions are as described previously (Voineagu et al., 2008). The agarose gel was transferred to a HybondXL membrane (Amersham) and hybridized with a 32P-labeled probe. The probe was PCR amplified from yeast genomic DNA using primers 7 and 8 (Table S2).

RNA Analysis

RNA analysis was performed using RT-PCR using Superscript III Reverse Transcriptase (Invitrogen) as described in the Supplemental Data.

SUPPLEMENTAL DATA

Supplemental Data include Supplemental Experimental Procedures and two tables and can be found with this article online at http://www.cell.com/molecular-cell/supplemental/S1097-2765(09)00430-4.

ACKNOWLEDGMENTS

We thank Catherine Freudereich, Hanna Klein, Robert Lahue, Steven Brill, Mitch McVey, Susan Lovett, and members of Tufts RRR seminar for many useful comments, suggestions, and discussions. This study was supported by the NIH grant GM60987 and generous contribution of the White family to S.M.M. and by the GM0825950 from NIGMS/NIH to K.S.L.

Received: August 22, 2009
Revised: January 7, 2009
Accepted: June 18, 2009
Published: July 9, 2009

REFERENCES


