MOLECULAR AND CELLULAR BIOLOGY, Mar. 2004, p. 2286–2295 Vol. 24, No. 6
0270-7306/04/$08.00+0 DOI: 10.1128/MCB.24.6.2286–2295.2004

Replication Stalling at Friedreich's Ataxia $(GAA)_n$ Repeats In Vivo

Maria M. Krasilnikova and Sergei M. Mirkin*

Department of Biochemistry and Molecular Genetics, University of Illinois at Chicago, Chicago, Illinois 60607

Received 21 November 2003/Returned for modification 16 December 2003/Accepted 22 December 2003

Friedreich's ataxia $(GAA)_n$ repeats of various lengths were cloned into a *Saccharymyces cerevisiae* plasmid, and their effects on DNA replication were analyzed using two-dimensional electrophoresis of replication intermediates. We found that premutation- and disease-size repeats stalled the replication fork progression in vivo, while normal-size repeats did not affect replication. Remarkably, the observed threshold repeat length for replication stalling in yeast (~40 repeats) closely matched the threshold length for repeat expansion in humans. Further, replication stalling was strikingly orientation dependent, being pronounced only when the repeat's homopurine strand served as the lagging strand template. Finally, it appeared that length polymorphism of the $(GAA)_n \cdot (TTC)_n$ repeat in both expansions and contractions drastically increases in the repeat's orientation that is responsible for the replication stalling. These data represent the first direct proof of the effects of $(GAA)_n$ repeats on DNA replication in vivo. We believe that repeat-caused replication attenuation in vivo is due to triplex formation. The apparent link between the replication stalling and length polymorphism of the repeat points to a new model for the repeat expansion.

Friedreich's ataxia, the most common form of inherited ataxias in humans, is caused by an expansion of the $(GAA)_n$ repeat within the first intron of the frataxin (X25) gene (5). This expansion apparently inhibits frataxin gene expression, leading to the disease (4). Several reports suggest that inhibition might occur at the level of transcription due to the triplex-forming potential of the $(GAA)_n$ repeat (1, 12, 27).

The mechanisms responsible for the $(GAA)_n$ repeat expansion remain to be understood. The propensity of this repeat to expand is markedly length dependent (reviewed in references 29 and 30). Normal alleles, ranging from 7 to 38 repeats, are stably inherited. The premutation alleles, carrying 38 to 65 noninterrupted repeats, might expand up to several hundred copies in a single generation. Finally, the disease alleles, containing more than 100 repeats, are highly prone to expansions. The reservoir for expansions appears to be the so-called large-normal alleles, carrying up to 55 $(GAA)_n$ repeats with some GAG interruptions (7, 20). It is believed that these interruptions prevent repeats from expanding, whereas their occasional loss converts large normal alleles into the premutation ones.

It was suggested that the length-dependent character of the $(GAA)_n$ repeat expansions could be due to its structural features. This sequence belongs to a class of homopurine-homopyrimidine mirror repeats that can adopt the triple-helical H-DNA configuration (24). Biochemical and biophysical analysis of this repetitive DNA revealed that it indeed adopts a basic triple-helical conformation in vitro (11, 22) or a more elaborate triplex structure, called "sticky DNA" (32, 38).

Triplex formation by the $(GAA)_n$ repeat impedes DNA polymerization in vitro (11, 26), and a model was put forward linking triplex-caused replication blockage with the repeat's propensity to expand (11). It was further demonstrated that cloning very long $(GAA)_n$ repeats (n of ~250) into a mammalian episomal vector severely decreased its copy number relative to an empty vector, likely due to the replication impasse of the repeat (27). Direct data on the effects of $(GAA)_n$ repeats on DNA replication in vivo were, however, absent. Here we developed an approach that allowed us to clone long noninterrupted $(GAA)_n$ repeats into an *S. cerevisiae* plasmid. Using two-dimensional analysis of replication intermediates, we found that these repeats attenuate the replication fork progression in vivo in a strikingly length- and orientation-dependent manner. The implications of our results for the mechanisms of the $(GAA)_n$ repeat expansions are discussed.

* Corresponding author. Mailing address: Department of Biochemistry and Molecular Genetics, University of Illinois at Chicago, Chicago, IL 60607. Phone: (312) 996-9610. Fax: (312) 413-0353. E-mail: mirkin@uic.edu.

MATERIALS AND METHODS

Strains. Cloning was carried out in the *Escherichia coli* XL1-Blue strain (Stratagene). Yeast replication studies were performed in *S. cerevisiae* CH1585 (*MAT***a** *leu2Δ1 trpΔ63 ura3-52 his3-200*) strain (ATCC 96098).

Plasmids. Yeast replication studies were carried out in a derivative of the pYES2 plasmid (Invitrogen). This derivative, pYES+, was obtained by inserting a 529-bp-long *Eco*RI-*Xmn*I fragment from the pTrcTACAT plasmid (17) into the *Xba*I site of pYES2, situated between the GAL1 promoter and the CYC1 terminator. This fragment corresponds to the 3′ part of the bacterial *cat* gene and does not contain prominent repeats.

Plasmid pYES-Control was obtained by inserting a 188-bp-long blunt-ended *Eco*RI-*Bam*HI fragment of YEp24 into the blunt-ended *Xho*I site of pYES+. This fragment does not contain obvious repeats.

Plasmid pYES-Bsg for cloning long $(GAA)_n \cdot (TTC)_n$ repeats was constructed in two steps. The original GTGCAG *Bsg*I site in the pYES+ plasmid site was first changed to GTGCAC. Then, another *Bsg*I site within the oligonucleotide TC GAGTGCAGACCTCAGGTTCTGCACA was introduced into the *Xho*I site of the modified plasmid.

Cloning $(GAA)_n \cdot (TTC)_n$ repeats. A starting $(GAA)_{57} \cdot (TTC)_{57}$ repeat was generated from complementary oligonucleotides $d(GAA)_{10}$ and $d(TTC)_{10}$, using the PCR strategy described in reference 35, and cloned into the *Eco*RV site of pBluescript SK(−) plasmid (Stratagene).

Plasmid pYES-TTC57 was obtained by inserting the blunt-ended *Eco*R1-*Hin*dIII fragment from pBluescript-GAA57 into the blunt-ended *Xho*I-site of pYES+.

Plasmid pYES-GAA57 was obtained by cloning the blunt-ended *Eco*R1-*Hin*dIII fragment from pBluescript-GAA57 into the blunt-ended *Eco*81I-site of the pYES-Bsg plasmid. As a result, the triplet repeat appeared to be flanked by the two adjacent *Bsg*I sites in the inverted orientation (underlined) and two distant *Eco*RI sites (italicized): *GAATTC*TGCAGATATCCATCACACTGGC

GGCCGCTCGAGTGCAGACCTCAAATTCGAT$(GAA)_{57}$GATCAAGCTTC AGGTTCTGCACATCGAGCATGCATCTA*GAATTC*.

Plasmids pYES-GAA20 and pYES-GAA40 were spontaneous deletion derivatives of the pYES-GAA57 plasmid.

To obtain the pYES-GAA114 plasmid, the $(GAA)_{57} \cdot (TTC)_{57}$ repeat was excised from the pYES-GAA58 plasmid using *Bsg*I, religated to the same plasmid underdigested with *Bsg*I, and transformed into the *E. coli* XL1-Blue strain. The length and integrity of the $(GAA)_{114} \cdot (TTC)_{114}$ repeat were verified by restriction analysis and DNA sequencing.

pYES-TTC114 was constructed by reversing the orientation of the *Eco*RI fragment carrying the $(GAA)_{114} \cdot (TTC)_{114}$ repeat within the pYES-GAA114.

To obtain the pYES-GAA228 plasmid, the $(GAA)_{114} \cdot (TTC)_{114}$ repeat from the pYES-GAA114 plasmid was excised by *Bsg*I followed by its religation to the same plasmid underdigested with *Bsg*I. The ligation mixture was transformed into the yeast CH1585 strain directly. Plasmids carrying the $(GAA)_{228} \cdot (TTC)_{228}$ repeat were identified upon restriction digestion and Southern blotting hybridization.

Isolation of yeast plasmid DNA minipreps. Yeast plasmid minipreps were obtained using the protocol suggested by Haber's lab. Cells were grown overnight in 4 ml of synthetic dropout media lacking uracil until saturation. Cell pellets were resuspended in 250 μl of zymolyase solution (1.2 M Sorbitol, 10 mM Tris-HCl [pH 8.0], 10 mM $CaCl_2$, 1% 2-mercaptoethanol, 3 U of zymolyase [ICN]). After 30 min of incubation at 37°C, cells were lysed by the addition of 200 μl of lysis solution (50 mM Tris-HCl [pH 8.0], 50 mM EDTA, 1.2% sodium dodecyl sulfate). Cell debris were cleared upon addition of 100 μl of 3 M KOAc, pH 5.5, and centrifugation at 13,000 × *g* for 10 min. Supernatants were precipitated by isopropanol and centrifuged at 13,000 × *g* for 5 min. Pellets were dissolved in 300 μl of Tris-EDTA (TE), incubated at 37°C with 30 U of RNase, and extracted with phenol twice, followed by chloroform extraction and isopropanol precipitation. Pellets were washed with 70% ethanol, vacuum dried, and resuspended in 30 μl of TE.

Isolation of replication intermediates from yeast cells. Cells were grown in 400 ml of complete synthetic medium lacking uracil with glucose (BIO 101) until reaching an optical density at 600 nm of 2.0. The growth was stopped by adding 4 ml of 10% NaN_3 and incubating for 2 min. Eighty milliliters of frozen 0.2 M EDTA were added, cultures were pelleted, and pellets were washed with 50 ml of ice-cold water and resuspended in 4 ml of NIB buffer (17% glycerol, 50 mM morpholinepropanesulfonic acid, 150 mM NaOAc, 2 mM $MgCl_2$, 0.5 mM spermidine, 0.15 mM spermine [pH 7.2]). An equal volume of glass beads was added, and cells were disrupted by vortexing for 7 min, chilling on ice for 30 s after every 30 s of vortexing. Beads were allowed to settle out, and the supernatant was pooled and cleared by centrifugation at 13,000 × *g* for 25 min. The pellet was resuspended in 1.5 ml of YSTE buffer (100 mM NaCl, 50 mM Tris-HCl [pH 8.0], 20 mM EDTA) and deproteinized for 1 h at 37°C by adding 0.225 ml of 10% sarcosil and 30 μl of 20-mg/ml proteinase K. Reaction mixtures were cleared by centrifugation at 12,000 × *g* for 5 min. DNA was further purified by overnight equilibrium centrifugation in a CsCl gradient with 87 μg of Hoechst 33258 trihydrochloride/ml at 50,000 rpm. The low band, corresponding to replicative intermediates, was collected, purified from the dye by butanol extraction, precipitated with ethanol, and dissolved in 30 μl of standard TE buffer.

Two-dimensional gel electrophoresis. Replication intermediates of pYES derivatives with $(GAA)_n \cdot (TTC)_n$ repeats in the *Xho*I site were digested with *Bsa*AI and *Bgl*I restriction enzymes. This was followed by phenol-chloroform extraction and ethanol precipitation. The pellets were dissolved in 15 μl of loading buffer and loaded on a 0.4% agarose gel. The lanes were then cut out of the 0.4% gel and embedded into a 1.5% agarose gel with 0.3 μg of ethidium bromide/ml for the electrophoresis in the second dimension (35). The gel was vacuum transferred onto a Nytran N membrane (Schleicher & Schuell) and hybridized with a ^{32}P-labeled probe, obtained with the Random Primers DNA labeling system (Invitrogen). A 799-bp-long *Bsa*AI-*Eco*RI fragment of pYES2 was used as a probe. Quantitative analysis was performed on a Storm 860 PhosphorImager using Imagequant software (Molecular Dynamics).

Plasmid copy number. Copy numbers for plasmids were obtained using two different assays: Southern blot and dot blot hybridizations. For Southern blot hybridization, 5 μl of plasmid minipreps were digested with *Hin*dIII, separated by agarose gel-electrophoresis, transferred onto the membrane, and hybridized with a 1,373-bp *Hin*1I-*Nhe*I fragment of pYES2, containing the URA3 gene. The plasmid copy number was determined as a ratio of the radioactivity of the plasmid-derived band (5,793-bp-long fragment) to the endogenous chromosomal band (2.0-kb-long fragment).

For the dot blot essay, 5 μl of plasmid minipreps were digested with *Eco*R1, phenol-chloroform extracted, precipitated by ethanol, and dissolved in water. The amounts of total DNA were checked by absorbance at 260 nm and/or gel electrophoresis. Equal amounts of DNA samples were denatured by boiling in 50 μl of water, followed by the addition of an equal volume of 20× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate). Dot blots were performed with a Bio-Dot microfiltration apparatus (Bio-Rad) according to the manufacturer's instructions. Hybridization was carried out with the 905-bp *Bgl*I-*Drd*I fragment of the pYES2 plasmid, containing a bacterial replication origin and *bla* gene. The relative copy numbers were obtained by comparing the intensities of the dots in various dilutions. The average of the values obtained from numerous experiments was calculated.

Analysis of the repeat length polymorphism. To study length polymorphism of $(GAA)_n \cdot (TTC)_n$ repeats depending on their orientation relative to their origins, yeast clones containing corresponding plasmids were grown on plates with synthetic dropout medium without uracil for five rounds. At each round, 10^7 cells were plated and grown for approximately 10 generations. Upon each round of cultivation, cells were scrubbed off the plate and dissolved in 100 ml of water, followed by plating of 200 μl of the resultant suspension onto a fresh selective plate. After each round of cultivation, plasmid DNA was isolated as described above. A 1-μl sample of each was digested with *Eco*RI, and EDTA was added up to a 10 mM concentration, followed by incubation at 65°C for 10 min. DNA was then phenol-chloroform extracted, precipitated by ethanol, and dissolved in 10 μl of TE. Samples were separated by 1% agarose gel electrophoresis, transferred onto the nylon membrane, and hybridized with a 439-bp-long *Eco*RI-*Eco*RI fragment of pYes-GAA115, containing the repeat.

RNA isolation and analysis. RNA was isolated from 10 ml of yeast cultures grown exponentially in the complete synthetic medium lacking uracil with galactose using the RNeasy Mini kit (Qiagen). Northern blotting hybridization was performed according to the standard protocol (36). A 112-bp-long *Pvu*II-*Xho*I fragment of pYES2, labeled using the Random Primers DNA labeling system (Invitrogen), served as a hybridization probe.

RESULTS

Cloning long noninterrupted $(GAA)_n \cdot (TTC)_n$ repeats. To study the replication fork progression through $(GAA)_n \cdot (TTC)_n$ repeats in eukaryotic cells, these repeats were cloned into an *S. cerevisiae* pYES2-derived (Invitrogen) plasmid. Short repeats were simply cloned into the pYES2 polylinker. It was critical for this study, however, to obtain long noninterrupted repeats. To this end, we used an approach previously described in reference 13 based on the ability of the restriction enzyme *Bsg*I to cut 14 of 16 nucleotides away from its recognition site. This cloning strategy is presented in Fig. 1. A $(GAA)_{57} \cdot (TTC)_{57}$ repeat was positioned between the two inverted *Bsg*I sites in such a way that the pure $(GAA)_{57} \cdot (TTC)_{57}$ fragment with GA and CT 3′ overhangs was generated upon *Bsg*I digestion. The nonpalindromic nature of these overhangs allowed self-ligation of this fragment in the head-to-tail direction only. $(GAA)_{114} \cdot (TTC)_{114}$ and $(GAA)_{228} \cdot (TTC)_{228}$ repeats were obtained by subsequent rounds of self-ligation of this fragment. These repeats were then cloned into the pYES derivative carrying a unique *Bsg*I site generating compatible overhangs upon restriction digestion. 57- and 114-mers of the $(GAA)_n \cdot (TTC)_n$ repeats appeared to be stably maintained in both *E. coli* and yeast cells. The $(GAA)_{228} \cdot (TTC)_{228}$ fragment in both orientations was cloned directly into yeast.

$(GAA)_n \cdot (TTC)_n$ repeats stall the replication fork progression in vivo. Cloned repeats were situated in the polylinker of the pYES-derived plasmid, approximately 1.4 kbp from the yeast 2μm replication *ori*. Due to the bidirectional character of the replication, this plasmid was separated into two domains replicated by different forks, where our inserts were in the left-to-right replication fork as depicted in Fig. 2A. Depending on a repeat's orientation, its individual DNA strands can be unambiguously assigned to the template for leading or lagging DNA strand synthesis for this fork. Our plasmids were named

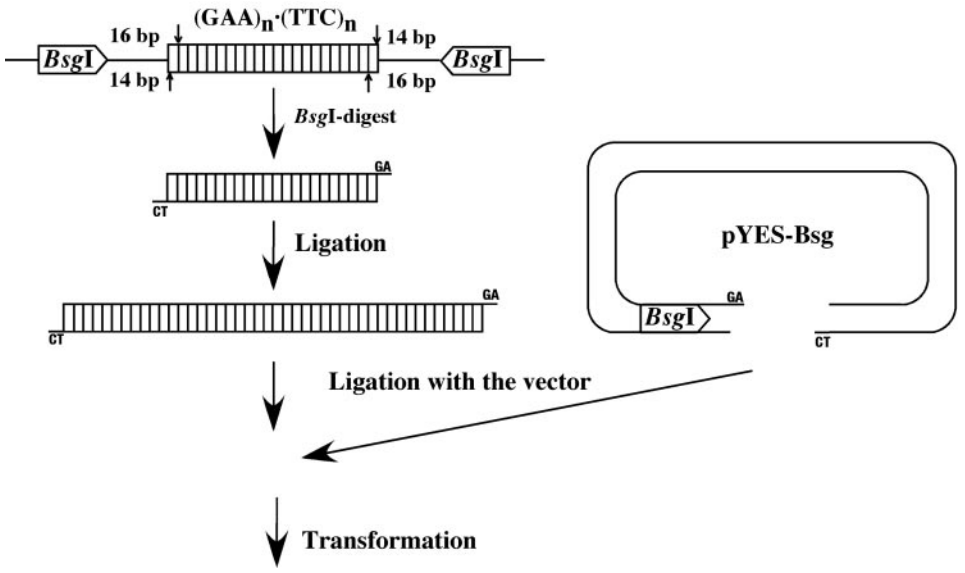

FIG. 1. Strategy for cloning long $(GAA)_n \cdot (TTC)_n$ repeats (see the text for details).

according to the repetitive sequence in the lagging strand template.

Digestion of replication intermediates by restriction enzymes *Bsa*AI and *Bgl*I leads to the appearance of Y-like structures (Fig. 2B). Those structures differ from nonreplicated DNA in their sizes and shapes, which allows one to resolve them by two-dimensional neutral-neutral agarose gel electrophoresis as a Y-arc (3) (Fig. 2B). If replication is attenuated by a triplet repeat, one should expect the appearance of a bulge on the otherwise smooth Y arc due to the preferential accumulation of replication intermediates of a specific size and shape. Our $(GAA)_n \cdot (TTC)_n$ repeats were positioned at approximately one-third of the distance from the end of a *Bsa*AI-*Bgl*I restriction fragment relative to the replication origin (Fig. 2A). We could expect, therefore, the appearance of bulges within the long shoulder of the Y-arc (Fig. 2B).

The primary data on the replication of plasmids with $(GAA)_n \cdot (TTC)_n$ inserts of various lengths and orientations are shown in Fig. 3A. The control plasmid, pYES-Control, had a relatively smooth Y-arc except for a knob at its peak. Similar knobs were also evident in all other plasmids studied, and their reproducibility depended on variables during isolation of the replication intermediates. The Y-arcs of plasmids carrying $(TTC)_n$ repeats in the lagging strand template did not differ from that of the control plasmid, even for very long repeats ($n = 228$). At the same time, one can see evident bulges (shown by arrows) in the long shoulders of the Y-arcs of plasmids with $(GAA)_n$ repeats in the lagging strand template, starting from an n value of 40. These bulges were located roughly one-third of the distance from the beginning of Y-arcs, i.e., at a position corresponding to the position of triplet repeats. We conclude, therefore, that $(GAA)_n \cdot (TTC)_n$ repeats stall DNA replication in vivo in a length- and orientation-dependent manner.

If a repeat completely blocked progression of one replication fork, the second replication fork would approach this repeat from the opposite side. This should result in the appearance of double-Y intermediates, forming a spike on the Y-arc (10). We did not see accumulation of spikes around these bulges, indicating that replication fork progression was only attenuated, rather than completely blocked, even by the longest $(GAA)_n$ repeats studied.

To quantify the strength of replication attenuation caused by $(GAA)_n \cdot (TTC)_n$ repeats, a portion of the Y-arcs (marked by arrows in Fig. 2B) for the corresponding plasmids were analyzed by phosphorimaging as previously described (31). Characteristic densitograms for repeat-containing plasmids are shown in Fig. 3B. One can see that the Y-arc of the pYES-TTC114 plasmid is quite smooth. Plasmids with long $(GAA)_n$ repeats in the lagging strand template, on the contrary, show profound peaks on their Y-arcs. The $(GAA)_{20}$ repeat does not significantly affect replication, while replication attenuation caused by the $(GAA)_{40}$ repeat is already very pronounced. Thus, there exists a clear-cut repeat threshold length responsible for the replication stalling in yeast lying between 20 and 40 repeats. In humans, the normal sizes of $(GAA)_n$ repeats in the frataxin gene range from 7 to 34 U, the permutation sizes are 34 to 80 U, and the disease-size repeats exceed 100 U. Thus, the repeat threshold length for replication blockage in yeast matches closely that for repeat expansions in humans.

Notably, the width of these peaks increased with the length of a $(GAA)_n$ run: 120 mm for n value of 40, 160 mm for n value of 57, 200 mm for n value of 114, and 260 mm for n value of 228. These width differences correlate closely with the differences in repeats lengths on a logarithmic scale, as one would expect for electrophoretic separation. At the same time, the heights of the peaks were rather similar for various repeats. Consequently, the ratio of radioactivity in the peak area to that in the corresponding area of a smooth replication arc remained roughly the same for different $(GAA)_n$ runs and corresponded to approximately 1.5. The latter ratio reflects the extent of replication stalling within a repeat. We believe, therefore, that an elementary stalling event caused by a $(GAA)_n$ run in the lagging strand template slows the replication fork down ≈1.5-fold. Multiple stalling events within long $(GAA)_n$ repeats likely

account for widening of the replication inhibition zone proportionally to the repeat's length.

Replication stalling does not depend on transcription through the repeats. Expanded $(GAA)_n \cdot (TTC)_n$ repeats disrupt frataxin gene expression, which is believed to be due to transcription elongation blockage occurring when $(GAA)_n$ runs are in the sense strand (1, 4). Further, propagation of an episomal vector carrying the $(GAA)_n \cdot (TTC)_n$ repeat in mammalian cells was predominantly inhibited when the $(GAA)_n$ strand was transcribed (27). Finally, we have previously shown that transcription stalling within a different DNA repeat could block replication (18). These considerations lead to a legitimate question: could the orientation-dependent attenuation of replication at the $(GAA)_n \cdot (TTC)_n$ repeat, described above, be due to transcription stalling at this repeat?

Our repeats were cloned within the pYES derivative in such a way that positioning of $(GAA)_n$ runs in the lagging strand template simultaneously positioned them into the sense strand for transcription. Note, however, that this area is transcribed from the *GAL1* promoter, which is induced by galactose but is repressed in the presence of glucose (40). Since yeast cultures were grown in the presence of glucose in all of the above experiments, transcription throughout the repeat should be negligible. To additionally check the possibility of transcription-replication interplay, we performed replication experiments with cells grown on either glucose or galactose. The results in Fig. 4 show that repeat-caused replication blockage was virtually identical in both sets of conditions. Transcription from the *GAL1* promoter is known to increase at least 1,000-fold upon galactose induction (40). Since replication blockage does not depend on the transcriptional status of a repeat-containing gene, it is highly unlikely to be mediated by transcription.

We further looked at transcription through $(GAA)_n$ repeats in yeast. Cells carrying plasmids with various $(GAA)_n$ repeats in the sense strand for transcription were grown in the presence of glucose followed by transcription induction by galactose. RNA was isolated and analyzed using Northern blot hybridization with the probe situated immediately upstream of the repeat (Fig. 5). In our system, the full-length control transcript is expected to be approximately 1.1 kb and to become progressively longer with an increase in the repeat's length. Transcription stalling at $(GAA)_n$ runs should lead to the accumulation of truncated transcripts that are approximately 0.3 kb long. To our surprise, we didn't observe any accumulation of truncated transcripts even for the longest repeats studied.

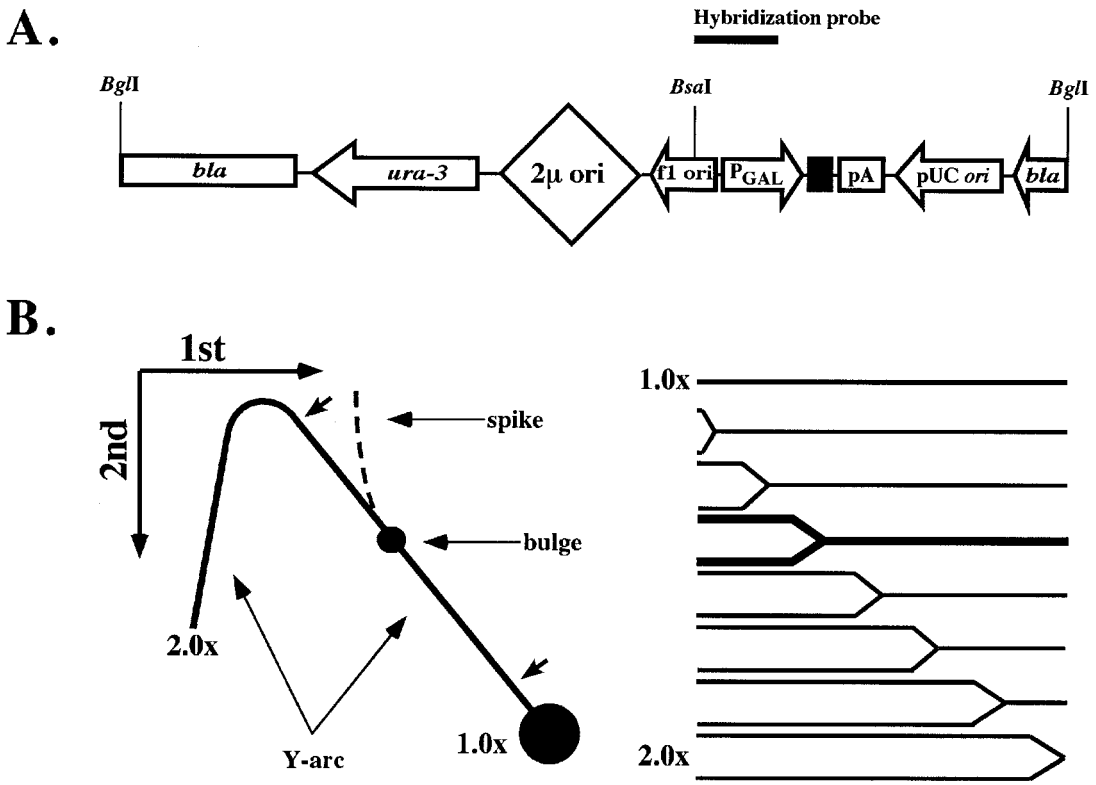

FIG. 2. Electrophoretic analysis of replication intermediates in yeast. (A) Structure of the *Bgl*I-linearized pYES derivatives containing $(GAA)_n \cdot (TTC)_n$ repeats. In this linear depiction, the 2μm replication origin is roughly in the middle of the plasmid. $(GAA)_n \cdot (TTC)_n$ repeats (black box) were cloned into the *Bsg*I site, being replicated by left-to-right replication fork. (B) Schematic representation for the two-dimensional neutral/neutral electrophoretic analysis. Cleavage of replication intermediates with restriction enzymes *Bsa*I and *Bgl*I generates Y-shaped DNA molecules, and the size of a Y increases with replication progression. Replication blockage at a repeat leads to the accumulation of replication intermediates of a given size and shape, as shown in bold (right panel). Separation by two-dimensional agarose electrophoresis reveals a Y-arc (left panel). Partial replication blockage by a repeat should result in the appearance of a bulge (black circle) on the replication arc. Complete replication blockage should lead to the appearance of a spike. Arrowheads in the right panel point to the portion of the Y-arc analyzed by phosphorimaging.

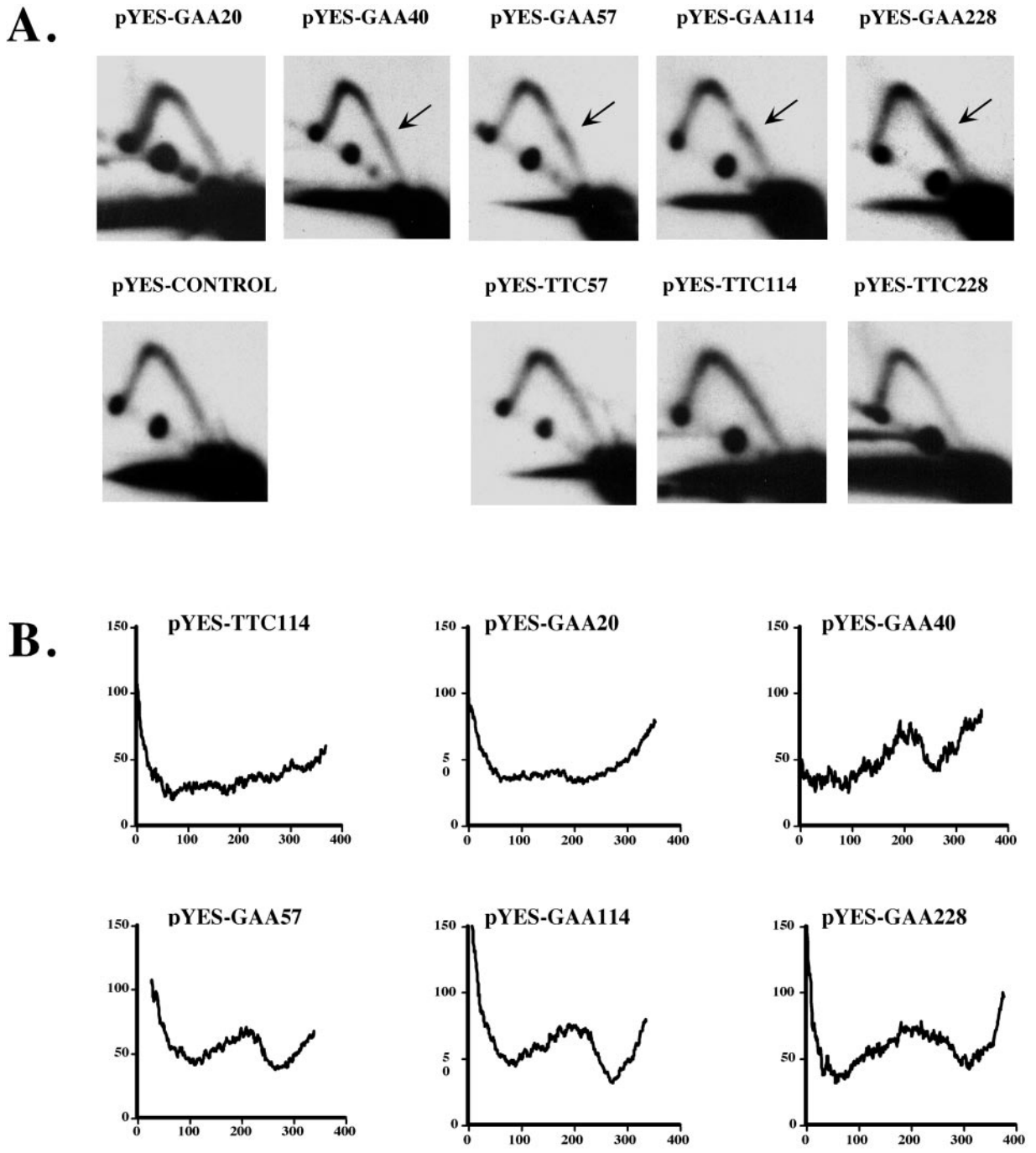

FIG. 3. Electrophoretic analysis of replication intermediates for pYES derivatives carrying various $(GAA)_n \cdot (TTC)_n$ repeats. (A) Primary electrophoretic data. Plasmids were named according to the repetitive sequence in the lagging strand template. Arrows show replication stall sites. (B) Quantitative analysis of Y-arcs for plasmids carrying $(GAA)_n \cdot (TTC)_n$ repeats. The analyzed portion of the Y-arc is marked by arrowheads in Fig. 1B. Peaks on densitograms correspond to bulges on the Y-arcs shown in Fig. 2A.

Moreover, the amount of full-length transcripts was remarkably similar for all the plasmids. We conclude, therefore, that $(GAA)_n$ repeats do not cause transcription stalling in yeast. These data differ from the previous results obtained in cultured mammalian cells that showed dramatic decrease in the amount of mRNA carrying long $(GAA)_n$ runs (1, 4), which was interpreted as the blockage of transcription elongation by this repeat. While we don't know the reasons for these differences, two explanations seem most plausible. First, they could be due to the differences between transcription apparatuses in yeast and mammals. Second, $(GAA)_n$ repeats were situated in the introns of reporter genes (possibly affecting splicing) in all mammalian studies, while they were within the exon in our case. Further studies are warranted to address this controversy.

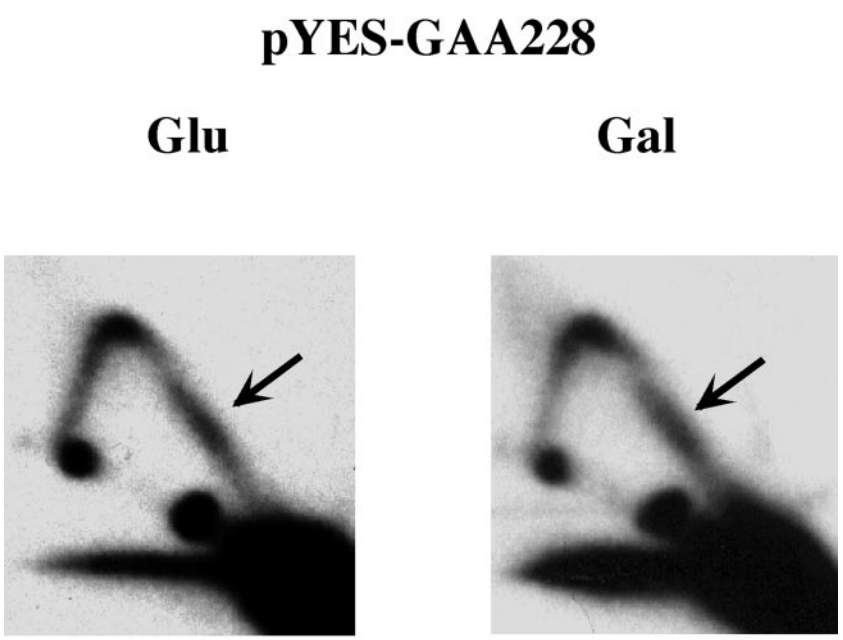

FIG. 4. Replication of the pYES-GAA228 plasmid maintained under conditions of transcriptional repression (Glu) or activation (Gal). Arrows show replication stall sites.

Biological consequences of replication stalling at $(GAA)_n \cdot (TTC)_n$ repeats. We further studied whether the orientation-dependent replication stalling at $(GAA)_n \cdot (TTC)_n$ repeats affects plasmid copy numbers and repeat length polymorphisms. To determine plasmid copy numbers, DNA samples isolated from yeast cells carrying different plasmids were analyzed by either dot blot or Southern hybridization. For dot blots, equivalent amount of various DNA samples (as determined by spectrophotometer) were digested with *Eco*RI, followed by hybridization with the plasmid-specific *amp* probe. For Southern hybridization, DNA samples were digested with *Hin*dIII, followed by their electrophoretic separation and hybridization with the *ura3* probe. The latter allowed us to normalize plasmid (5.8 kb-long fragment) to chromosomal (2.0 kb-long fragment) DNA more precisely. In fact, the two methods gave nearly identical results. Figure 6 combines the data on copy numbers for the repeat-containing plasmids relative to the control pYES2 plasmid from both type of experiments. One can see that these numbers are indistinguishable within the experimental error. We conclude, therefore, that relatively modest replication stalling caused by $(GAA)_n \cdot (TTC)_n$ repeats is insufficient to significantly affect plasmid copy number.

To study the effect of replication stalling at $(GAA)_n \cdot (TTC)_n$ repeats on their stability, we analyzed the dependence of the repeat length polymorphisms on their orientations relative to the replication origin. To this end, yeast clones with pYES-derived plasmids carrying repeats of identical lengths (n = 228) but opposite orientations have undergone several subsequent rounds of cultivation on synthetic medium plates lacking uracil, i.e., in conditions selecting for plasmid maintenance. In each round, 10^7 cells were plated and grown for approximately 10 generations. Plasmid DNAs isolated upon each round of cultivation were digested by *Eco*RI and analyzed for repeats length using Southern blot hybridization.

Figure 7A shows characteristic results for the $(GAA)_{228} \cdot (TTC)_{228}$ repeat in both orientations. We used the same plasmids as for the replication studies that were named according to the repetitive sequence on the lagging strand template. One can see that in the TTC orientation, i.e., when replication is not interrupted, the repeat remains rather stable through five cultivation cycles, as reflected by a discrete narrow band on a gel. In the GAA orientation, where replication stalling was previously observed, the repeat's band appeared to be less discrete, to begin with, and further widened upon cultivation, as reflected by a broader major band surrounded by the evident smear. Interestingly, this DNA smear migrates both underneath and above the major band, corresponding to the repeat's contractions and expansions, respectively. A bold arrow in Fig. 7A shows the length of the repeat-containing fragment in these experiments, while plain arrows correspond to reference size markers. One can see that the lower edge of the smear migrates at ~500 bp, which would result from the loss of 93 repeated units. The upper edge of the smear migrates at ~900 bp, reflecting the addition of 40 GAA repeats. Thus, both contractions and expansions observed are fairly large scale.

Figure 7B shows quantitative analysis for a typical Southern blot hybridization data using a PhosphorImager. It is apparent that there is progressive accumulation of contractions and expansions of the repeat in its GAA orientation upon cultivation. While deletions seem to accumulate faster than expansions, the latter are also quite evident. At the same time, analysis of repeat length in the TTC orientation shows few, if any, contractions or expansions. These results point to a link between replication stalling and an increase in length polymorphism for

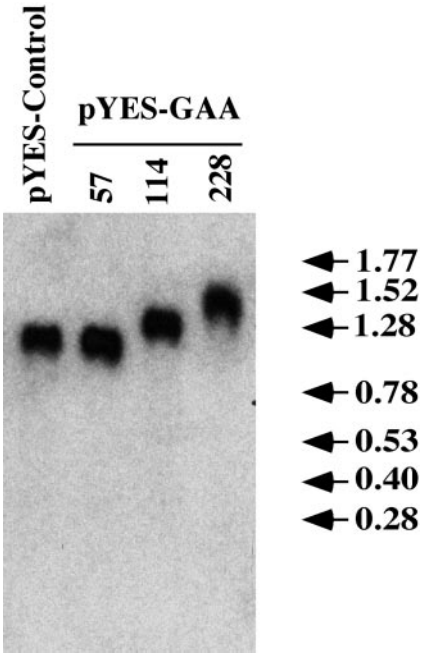

FIG. 5. $(GAA)_n$ repeats do not affect transcription in yeast. Plasmids were the same that were used in replication studies. Transcription from the GAL1 promoter was analyzed by Northern hybridization with the radiolabeled *Pvu*II-*Xho*I probe, corresponding to the +78-to-+190 part of the GAL1 transcript, immediately upstream of the repeat. A molecular weight ladder is marked by arrows.

the $(GAA)_n \cdot (TTC)_n$ repeat, since both are observed at the same orientation of the repeat relative to the *ori*.

DISCUSSION

Our data show, for the first time, that Friedreich's ataxia $(GAA)_n$ repeats substantially attenuate replication fork progression in vivo. This effect is remarkably orientation dependent, being pronounced only when the homopurine strand of the repeat is in the lagging strand template for replication. There exists an excellent correlation between threshold lengths responsible for the repeat-caused replication blockage in yeast and its expansions in humans. We believe, therefore, that there is a link between the replication blockage and the repeat's expansion. What could be the mechanisms of replication blockage and expansions of the repeat?

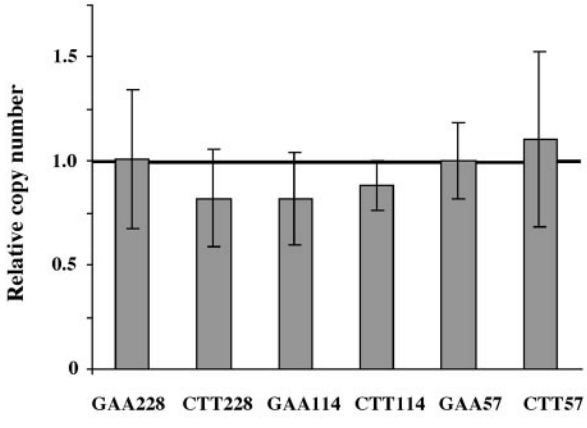

FIG. 6. Relative copy numbers of plasmids containing various $(GAA)_n \cdot (TTC)_n$ repeats as determined by the combination of dot blot and Southern hybridization assays. Plasmids are named according to sequences of the lagging strand template. The horizontal line reflects the copy number of a control plasmid, carrying a nonrepetitive sequence, used for normalization.

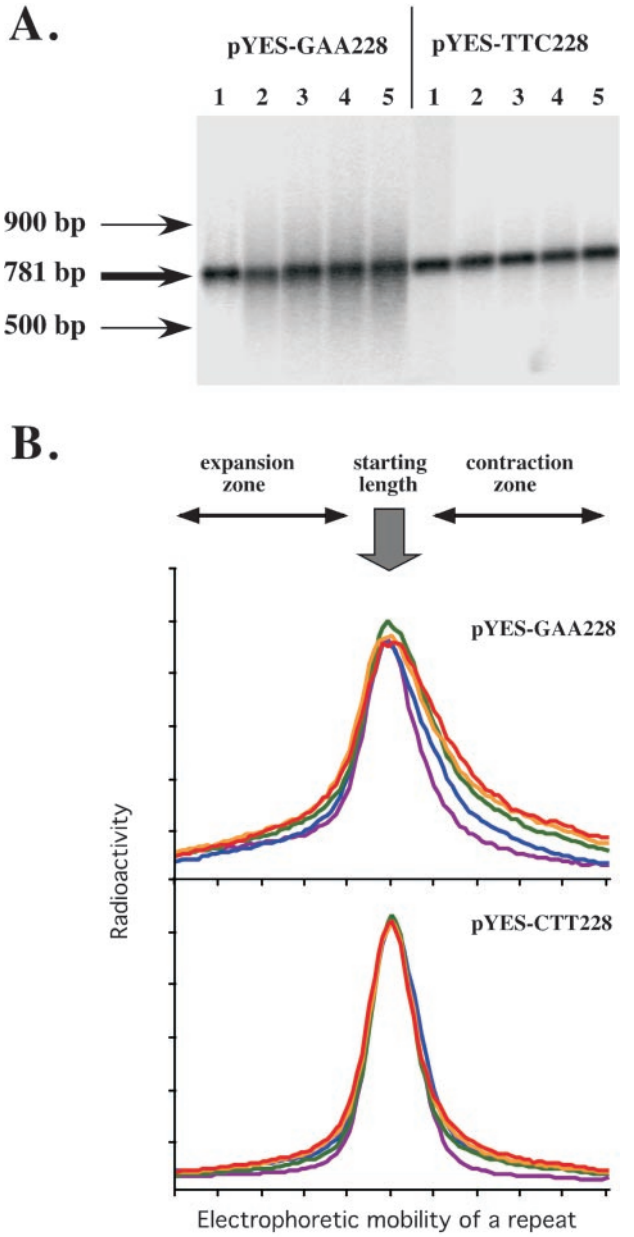

FIG. 7. Length polymorphism of the $(GAA)_{228} \cdot (TTC)_{228}$ repeat in the two orientations relative to the origin. (A) Experimental Southern hybridization data. Plasmids are named according to sequences of the lagging strand template. Numbers 1 to 5 correspond to five consecutive rounds of cultivation. (B) Quantitative analysis of the Southern hybridization data using a PhosphorImager. Purple, blue, green, yellow, and red lines correspond to cultivation rounds one through five, respectively.

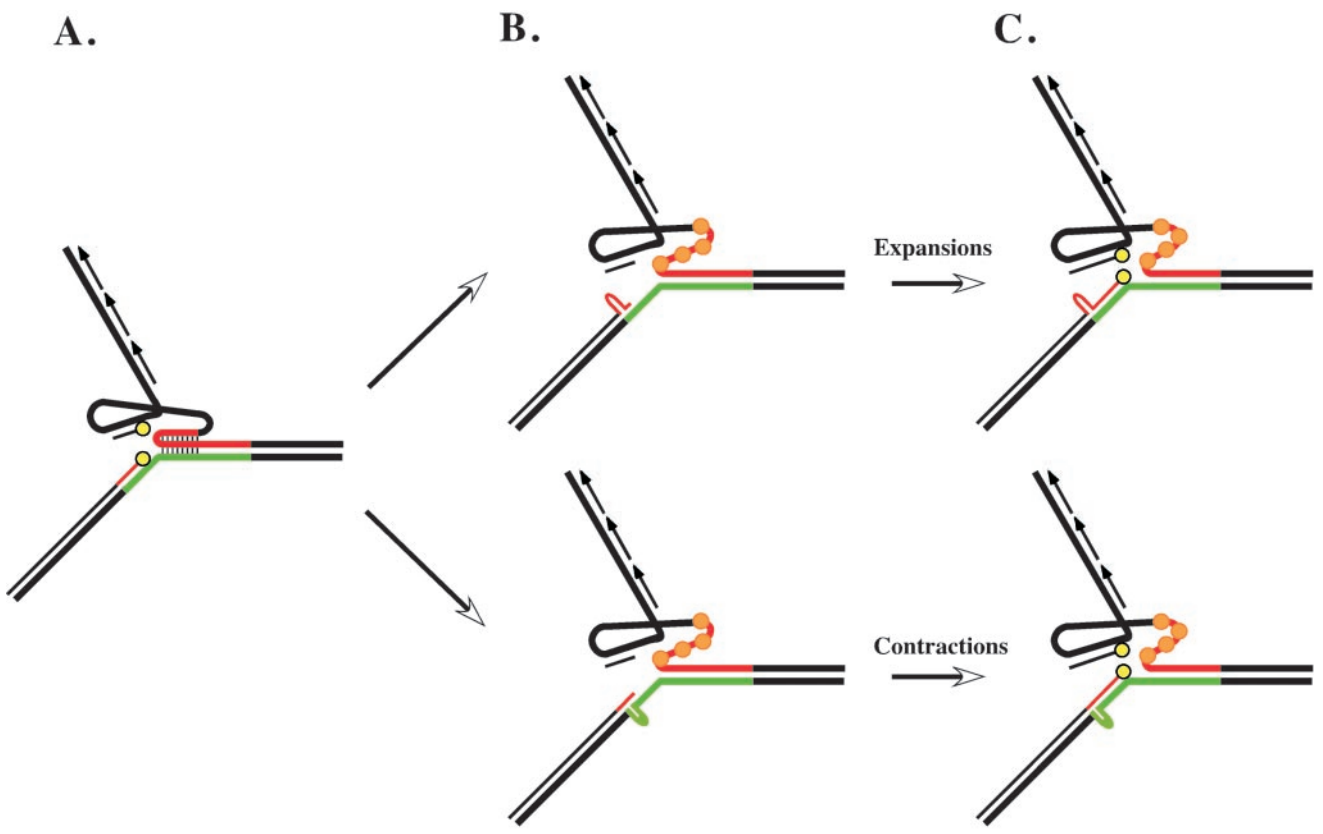

FIG. 8. A model for replication blockage by $(GAA)_n \cdot (TTC)_n$ repeats leading to their expansions. (A) A portion of the lagging strand template with a $(GAA)_n$ run folds back to form a stable triplex that stalls the leading DNA polymerase. (B) Slow unraveling of this triplex could be accompanied by polymerase dissociation and misalignment of the newly synthesized and template DNA strands. (C) Resumption of DNA synthesis upon triplex dismantling would then lead to repeat expansions or contractions. Red line, homopurine strand; green line, homopyrimidine strand; black line, flanking DNA. Arrows depict Okazaki fragments. Yellow circles represent leading and lagging DNA polymerases. Brown circles symbolize genome guardians, such as DNA helicases, SSB proteins, etc.

Our data exclude aberrant transcription of the repeat. Replication arrest could also be caused by a concerted binding of multiple protein molecules to a repetitive run as described by us for $(GA)_n \cdot (TC)_n$ repeats (19). In the latter case, however, replication inhibition was equally pronounced in both orientations of the repeat relative to the *ori*, as one would expect for protein binding. This model does not seem to apply for the $(GAA)_n \cdot (TTC)_n$ repeat, since it fails to explain the extreme orientation dependence of its inhibitory effect on replication.

We believe that the mechanisms of replication blockage by the $(GAA)_n \cdot (TTC)_n$ repeat could be grounded in its structural features. This repeat can adopt triple-helical H-DNA conformation in vitro (11, 22, 32), preventing efficient DNA polymerization through it (11, 26). Generally, the efficiency of DNA polymerization through various H-forming repeats depends on their orientation (33). Displacement of the homopurine strand during DNA synthesis leads to stable triplex formation and subsequent polymerase halt, while displacement of the homopyrimidine strand is relatively harmless. This phenomenon could explain our data on orientation-dependent blockage of DNA replication in vivo by $(GAA)_n \cdot (TTC)_n$ repeats.

During DNA replication, a portion of the lagging strand template (of about Okazaki fragment size) remains single stranded (8). Note that the size of an Okazaki fragment in eukaryotes is approximately 150 nucleotides, i.e., equivalent to roughly 50 triplet repeats. When this single-stranded portion contains a $(GAA)_n$ run, it can fold back to form a stable triplex with the remaining double-stranded part of the repeat causing DNA polymerase to stop and possibly even dissociate (Fig. 8A). It did not escape our attention that the formation of triplexes in this case could be additionally facilitated by two factors. First, the pure homopurine run should be considered the worst possible template for the DNA primase, since replication primers usually start from purines (reviewed in reference 9). Second, binding of the replication protein A to homopurine templates is severely ($\approx$50-fold) compromised compared to that with the homopyrimidine counterparts (39). The combination of these two factors would increase the likelihood for $(GAA)_n$ runs on the lagging template to remain

single stranded, thus increasing their propensity to adopt the triple-helical configuration.

Our experimental data show that length polymorphism of the $(GAA)_n \cdot (TTC)_n$ repeats drastically increases in repeats associated with replication stalling. Notably, the levels of both repeat contractions and expansions are significantly elevated when the repeat's replication is impaired. These data point to a clear-cut link between the repeat's replication blockage and its propensity to expand and contract. How could this be explained in the frame of our triplex model? We believe that to continue replication upon repeat-mediated triplex formation, genome guardians, such as replication protein A, DNA helicases, and others, dismantle the triplex (6, 16, 21, 34). While this dismantling is under way, newly synthesized and template DNA strands in the stalled replication fork might dissociate and misalign at a repetitive sequence (28). This misalignment can clearly occur in two ways, such that the repetitive segment on either the template or the newly synthesized DNA strand is looped out (Fig. 8B). It can also be additionally stabilized by the formation of hairpin-like structures by $(GAA)_n$ runs (14). Resumption of DNA synthesis would then lead to the repeat's contractions or expansions, respectively (Fig. 8C).

Importantly, for a $(GAA)_n \cdot (TTC)_n$ run that exceeds the size of an Okazaki fragment, i.e., when n is more than 50, this scenario can recur during its replication. It is tempting, therefore, to explain widening of replication stall areas with an increase in the repeat's length (Fig. 3B) as being the consequence of multiple replication stalling events within long repeats. Obviously, both triplex formation and DNA strand misalignment are fortuitous events. Thus, the longer the repeat, the more probable replication slowing and expansions should become. This could explain an increase in probability of expansions for longer repeats. Finally, our model assumes that expansions occur during the leading strand synthesis. While most current models for triplet repeat expansions suggest that they occur during lagging strand synthesis (2, 23, 37), expansions in the course of leading strand synthesis have been detected as well (15).

Our hypothesis could also give some clues to the origin of Friedreich's ataxia. Familial analysis has demonstrated that the reservoir for expansions here are the so-called large-normal alleles, carrying up to 55 GAA repeats. Further, most, if not all, of the large-normal alleles were derived from a single founder chromosome (7, 20). What could predispose $(GAA)_n$ repeats in the founder chromosome for expansions? We believe that expansions follow replication stalling caused by $(GAA)_n$ runs in the lagging strand template. It is tempting to speculate, therefore, that the first event triggering expansions in the founder chromosome could be the inactivation of the regular replication origin situated on one side of the triplet repeat and concurrent activation of a cryptic origin on its other side, positioning the $(GAA)_n$ runs in the lagging strand template. This model, called ORI-SWITCH (25), can apply to other triplet repeat diseases, as well.

ACKNOWLEDGMENTS

We thank George Samadashwily, Gordana Raca, and Labib Rouhana for their help in plasmid construction, Randal Cox and Andrey Krasilnikov for valuable comments, and Gerald Buldak for editorial help.

This work was supported by the grant GM60987 from NIH to S.M.M.

REFERENCES

1. **Bidichandani, S. I., T. Ashizawa, and P. I. Patel.** 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.
2. **Bowater, R. P., and R. D. Wells.** 2001. The intrinsically unstable life of DNA triplet repeats associated with human hereditary disorders. Prog. Nucleic Acid Res. Mol. Biol. **66:**159–202.
3. **Brewer, B. J., and W. L. Fangman.** 1987. The localization of replication origins on ARS plasmids in S. cerevisiae. Cell **51:**463–471.
4. **Campuzano, V., L. Montermini, Y. Lutz, L. Cova, C. Hindelag, S. Jiralerspong, Y. Trottier, S. J. Kish, B. Faucheux, P. Trouillas, F. J. Authier, A. Durr, J.-L. Mandel, A. L. Vescovi, M. Pandolfo, and M. Koenig.** 1997. Frataxin is reduced in Friedreich ataxia patients and is associated with mitochondrial membranes. Hum. Mol. Genet. **6:**1771–1780.
5. **Campuzano, V., L. Montermini, M. D. Molto, L. Pianese, M. Cossee, F. Cavalcanti, E. Monros, F. Rodius, F. Duclos, A. Monticelli, F. Zara, J. Canizares, H. Koutnikova, S. I. Bidichandari, C. Gellera, A. Brice, P. Trouillas, G. De Michele, A. Filla, R. De Frutos, F. Palau, P. I. Patel, S. Di Donato, J.-L. Mandel, S. Cocozza, M. Koenig, and M. Pandolfo.** 1996. Friedreich's ataxia: autosomal recessive disease caused by an intronic GAA triplet repeat expansion. Science **271:**1423–1427.
6. **Cheung, I., M. Schertzer, A. Rose, and P. M. Lansdorp.** 2002. Disruption of dog-1 in Caenorhabditis elegans triggers deletions upstream of guanine-rich DNA. Nat. Genet. **4:**405–409.
7. **Cossee, M., M. Schmitt, V. Campuzano, L. Reutenauer, C. Moutou, J. L. Mandel, and M. Koenig.** 1997. Evolution of the Friedreich's ataxia trinucleotide repeat expansion: founder effect and premutations. Proc. Natl. Acad. Sci. USA **94:**7452–7457.
8. **DePamphilis, M. L., and P. M. Wassarman.** 1980. Replication of eukaryotic chromosomes: a close-up of the replication fork. Annu. Rev. Biochem. **49:**627–666.
9. **Frick, D. N., and C. C. Richardson.** 2001. DNA primases. Annu. Rev. Biochem. **70:**39–80.
10. **Friedman, K. L., and B. J. Brewer.** 1995. Analysis of replication intermediates by two-dimensional agarose gel electrophoresis. Methods Enzymol. **262:**613–627.
11. **Gacy, A. M., G. M. Goellner, C. Spiro, X. Chen, G. Gupta, E. M. Bradbury, R. B. Dyer, M. J. Mikesell, J. Z. Yao, A. J. Johnson, A. Richter, S. B. Melancon, and C. T. McMurray.** 1998. GAA instability in Friedreich's ataxia shares a common, DNA-directed and intraallelic mechanism with other trinucleotide diseases. Mol. Cell **1:**583–593.
12. **Grabczyk, E., and K. Usdin.** 2000. 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.
13. **Grabczyk, E., and K. Usdin.** 1999. Generation of microgram quantities of trinucleotide repeat tracts of defined length, interspersion pattern, and orientation. Anal. Biochem. **267:**241–243.
14. **Heidenfelder, B. L., A. M. Makhov, and M. D. Topal.** 2003. Hairpin formation in Friedreich's ataxia triplet repeat expansion. J. Biol. Chem. **278:**2425–2431.
15. **Iyer, R. R., and R. D. Wells.** 1999. Expansion and deletion of triplet repeat sequences in Escherichia coli occur on the leading strand of DNA replication. J. Biol. Chem. **274:**3865–3877.
16. **Kopel, V., A. Pozner, N. Baran, and H. Manor.** 1996. Unwinding of the third strand of a DNA triple helix, a novel activity of the SV40 large T-antigen helicase. Nucleic Acids Res. **24:**330–335.
17. **Krasilnikov, A. S., A. Podtelezhnikov, A. Vologodskii, and S. M. Mirkin.** 1999. Large-scale effects of transcriptional DNA supercoiling in vivo. J. Mol. Biol. **292:**1149–1160.
18. **Krasilnikova, M. M., G. M. Samadashwily, A. S. Krasilnikov, and S. M. Mirkin.** 1998. Transcription through a simple DNA repeat blocks replication elongation. EMBO J. **17:**5095–5102.
19. **Krasilnikova, M. M., E. V. Smirnova, A. S. Krasilnikov, and S. M. Mirkin.** 2001. A new trick for an old dog: TraY binding to a homopurine-homopyrimidine run attenuates DNA replication. J. Mol. Biol. **313:**271–282.
20. **Labuda, M., D. Labuda, C. Miranda, J. Poirier, B. W. Soong, N. E. Barucha, and M. Pandolfo.** 2000. Unique origin and specific ethnic distribution of the Friedreich ataxia GAA expansion. Neurology **54:**2322–2324.
21. **Maine, I. P., and T. Kodadek.** 1994. Efficient unwinding of triplex DNA by a DNA helicase. Biochem. Biophys. Res. Commun. **204:**1119–1124.
22. **Mariappan, S. V. S., P. Catasti, L. A. Silks, E. M. Bradbury, and G. Gupta.** 1999. The high-resolution structure of the triplex formed by the GAA/TTC triplet repeat associated with Friedreich's ataxia. J. Mol. Biol. **285:**2035–2052.
23. **McMurray, C. T.** 1995. Mechanisms of DNA expansion. Chromosoma **104:**2–13.

24. **Mirkin, S. M., and M. D. Frank-Kamenetskii.** 1994. H-DNA and related structures. Annu. Rev. Biophys. Biomol. Struct. **23:**541–576.
25. **Mirkin, S. M., and E. V. Smirnova.** 2002. Positioned to expand. Nat. Genet. **31:**5–6.
26. **Ohshima, K., S. Kang, J. E. Larson, and R. D. Wells.** 1996. Cloning, characterization, and properties of seven triplet repeat DNA sequences. J. Biol. Chem. **271:**16773–16783.
27. **Ohshima, K., L. Montermini, R. D. Wells, and M. Pandolfo.** 1998. Inhibitory effects of expanded GAA.TTC triplet repeats from intron I of the Friedreich's ataxia gene on transcription and replication in vivo. J. Biol. Chem. **273:**14588–14595.
28. **Ohshima, K., and R. D. Wells.** 1997. Hairpin formation during DNA synthesis primer realignment in vitro in triplet repeat sequences from human hereditary disease genes. J. Biol. Chem. **272:**16798–16806.
29. **Pandolfo, M.** 1999. Molecular pathogenesis of Friedreich ataxia. Arch. Neurol. **56:**1201–1208.
30. **Patel, P. I., and G. Isaya.** 2001. Friedreich ataxia: from GAA triplet-repeat expansion to frataxin deficiency. Am. J. Hum. Genet. **69:**15–24.
31. **Pelletier, R., M. M. Krasilnikova, G. M. Samadashwily, R. S. Lahue, and S. M. Mirkin.** 2003. Replication and expansion of trinucleotide repeats in yeast. Mol. Cell. Biol. **23:**1349–1357.
32. **Sakamoto, N., P. D. Chastain, P. Parniewski, K. Oshima, M. Pandolfo, J. D. Griffith, and R. D. Wells.** 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.
33. **Samadashwily, G. M., A. Dayn, and S. M. Mirkin.** 1993. Suicidal nucleotide sequences for DNA polymerization. EMBO J. **12:**4975–4983.
34. **Samadashwily, G. M., and S. M. Mirkin.** 1994. Trapping DNA polymerases using triplex-forming oligodeoxyribonucleotides. Gene **149:**127–136.
35. **Samadashwily, G. M., G. Raca, and S. M. Mirkin.** 1997. Trinucleotide repeats affect DNA replication in vivo. Nat. Genet. **17:**298–304.
36. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual—2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
37. **Usdin, K., and E. Grabczyk.** 2000. DNA repeat expansions and human disease. Cell. Mol. Life Sci. **57:**914–931.
38. **Vetcher, A. A., M. Napierala, R. R. Iyer, P. D. Chastain, J. D. Griffith, and R. D. Wells.** 2002. Sticky DNA, a long GAA · GAA · TTC triplex that is formed intramolecularly, in the sequence of intron 1 of the frataxin gene. J. Biol. Chem. **277:**39217–39227.
39. **Wold, M. S.** 1997. Replication protein A: a heterotrimeric, single-stranded DNA-binding protein required for eukaryotic DNA metabolism. Annu. Rev. Biochem. **66:**61–92.
40. **Youkum, R. R., S. Hanley, R. West, and M. Ptashne.** 1984. Use of *lacZ* fusions to delimit regulatory elements of the inducible divergent GAL1-GAL10 promoter in *Saccharomyces cerevisiae*. Mol. Cell. Biol. **4:**1985–1999.