Hairpin Formation in Friedreich's Ataxia Triplet Repeat Expansion*

Brooke L. Heidenfelder‡, Alexander M. Makhov§, and Michael D. Topal‡§¶∥**

From the §Lineberger Comprehensive Cancer Center, ¶Department of Pathology, ‡Department of Biochemistry and Biophysics, and ∥Curriculum in Genetics and Molecular Biology, University of North Carolina, Chapel Hill, North Carolina 27599-7295

Triplet repeat tracts occur throughout the human genome. Expansions of a $(GAA)_n/(TTC)_n$ repeat tract during its transmission from parent to child are tightly associated with the occurrence of Friedreich's ataxia. Evidence supports DNA slippage during DNA replication as the cause of the expansions. DNA slippage results in single-stranded expansion intermediates. Evidence has accumulated that predicts that hairpin structures protect from DNA repair the expansion intermediates of all of the disease-associated repeats except for those of Friedreich's ataxia. How the latter repeat expansions avoid repair remains a mystery because $(GAA)_n$ and $(TTC)_n$ repeats are reported not to self-anneal. To characterize the Friedreich's ataxia intermediates, we generated massive expansions of $(GAA)_n$ and $(TTC)_n$ during DNA replication *in vitro* using human polymerase β and the Klenow fragment of *Escherichia coli* polymerase I. Electron microscopy, endonuclease cleavage, and DNA sequencing of the expansion products demonstrate, for the first time, the occurrence of large and growing $(GAA)_n$ and $(TTC)_n$ hairpins during DNA synthesis. The results provide unifying evidence that predicts that hairpin formation during DNA synthesis mediates all of the disease-associated, triplet repeat expansions.

Tracts of pure repeating triplets, referred to either as trinucleotide- or triplet repeat (TR)[1] tracts, occur throughout the human genome (1). Expansions of TR tracts located in specific genes cause more than 14 hereditary neuromuscular diseases (2). TRs associated with disease include $5'$-$d(CAG)_n/5'$-$d(CTG)_n$, (abbreviated as CAG/CTG) associated with Huntington's disease and myotonic dystrophy, among others (reviewed in Ref. 2), GCC/GGC, associated with fragile X syndrome, and GAA/TTC repeats, associated with Friedreich's ataxia (FRDA). The disease expansions vary from a few triplets in a coding region of the gene (type I) to hundreds of triplets located in a noncoding region of the gene (type II). The expansions exactly mimic the sequence of the repeat tract and occur during a single transmission from parent to child.

Studies in several model systems indicate that repeat instability is dependent on the length of the repeat tract and its orientation in the genome; a TR sequence capable of forming hairpins, constructed by cloning to be oriented relative to an origin of replication so that it is the template replicated by discontinuous lagging strand replication, is predicted to suffer deletions. The same sequence, oriented so that it is the Okazaki fragment, is predicted to suffer expansions (3–5). The discontinuous nature of lagging strand replication is characterized by the presence of single-stranded regions in the template and free $5'$-ends in the growing strand (Okazaki fragment) (for reviews, see Refs. 6 and 7). The template single-stranded DNA (ssDNA) presents the opportunity for formation of secondary structures such as hairpins (8). Lagging strand replication is predicted to lead to contractions by replication across hairpins formed in the template strand and expansion by DNA slippage to give hairpin formation in the Okazaki fragment (4, 5). Factors that increase DNA slippage should favor expansion over contraction. CAG, CTG, GCC, and GGC TR tracts synthesized and studied *in vitro*, form hairpin and hairpin-related tetraplex secondary structures through self-annealing (9–12). The hairpins are predicted, by the above mechanisms, to produce contractions as well as expansions; both have been observed to occur during replication of TR tracts in bacteria and yeast (4, 5, 8, 13).

The disease-associated, single-stranded expansion products may be thousands of nucleotides in length, a great potential target for DNA repair. Secondary structure formation within the single-stranded expansion intermediate is predicted to protect the expansion from repair activities of the cell, including mismatch repair and flap endonuclease (14–20). The latter removes $5'$ flaps of Okazaki fragments displaced by DNA synthesis. Secondary structures formed from the self-annealing of synthetic single-stranded GAA and TTC TRs are much less stable (21–23) than the structures adopted by single-stranded CAG, CTG, GCC, and GGC TRs (9–12); GAA and TTC do not self-anneal under physiological conditions of temperature and salt (21–23), although $(GAA)_{15}$ may self-anneal at low temperature (23). Thus, when presented with heteroduplex DNA, generated by meiotic recombination to contain single-stranded loops with different repeat sequence tracts, yeast repairs 10-repeat GAA and TTC single-stranded repeat tracts during meiosis but is much less efficient at repairing 10-repeat CAG, CTG, GCC, and GGC single-stranded repeat tracts (19). The differences in repair presumably are caused by the different abilities of the repeat tracts to self-anneal to form hairpins.

What intermediate structures are involved in FRDA TR expansions? Mixing together GAA and TTC TR tracts gives three-stranded structures called triplexes (purine-purine-pyrimidine and pyrimidine-purine-pyrimidine) (22, 24) and duplexes that associate strongly with each other called sticky DNAs (25). The predicted formation of these structures appears to be responsible for pausing of DNA replication within the repeat tracts observed *in vitro* (24, 26–28). Blockage of DNA replication within a repeat tract induces expansion *in vitro* and is pre-

---

dicted to induce DNA slippage and thus expansion *in vivo* (29, 30). This leaves open the question of how the large single-stranded regions resulting from massive expansions characteristic of FRDA manage to avoid DNA repair.

TR expansion is predicted to occur *in vivo* as the result of DNA slippage of the Okazaki fragment occurring during DNA lagging strand replication of the TR tract (31, 32) and possibly DNA slippage during DNA repair (33). *In vitro* models of DNA replication and repair have been used to generate type I and type II repeat expansion. Initial attempts at TR expansions during DNA replication used the complementary triplet repeats as template and primer (34–38). Recent studies have added unique sequences flanking the TR tract to better mimic the *in vivo* situation of the repeats (29). The latter study found that priming from within the TR tract gave much higher levels of expansion compared with priming from within the upstream unique flanking sequence. Priming from within the TR tract mimics the occurrence of an Okazaki fragment completely within a repeat tract. One of the most common lesions that arise in cellular DNA is the loss of a base to give an abasic site (39), which blocks DNA replication (40, 41). An abasic site analog, tetrahydrofuran (THF), synthesized at the 3′ terminus of the template-strand repeat tract greatly enhanced expansion during replication of the template, suggesting a possible role for DNA damage in TR expansion (29). The lesion enhanced triplet repeat expansion presumably by increasing the opportunities for DNA slippage by keeping the growing strand end within the repeat tract (29). Another recent *in vitro* model system, which uses unique sequences flanking the TR, mimics DNA slippage at a nick during DNA repair synthesis within the DNA construct (42). Using these *in vitro* replication models, type I expansions of many triplet repeats have been generated using bacterial polymerases (34–37). Human pol β (h-pol β) has been used to generate type I expansions of GCC/GGC (35) and CAG/CTG (42). Massive type II expansions of AAT, ATT, and CAG growing strand repeat tracts were achieved using the *in vitro* lagging strand model (29) with Klenow pol I *in vitro*. (Throughout this paper we refer to the sequence of the expanded growing strand. Thus, expanded TTC repeat means that the TTC growing strand grew beyond the length expected from replication of the template strand.) The DNA sequences of the type I GCC/GGC expansions (35) and the type II expansions by Klenow pol I (29, 38) were sequenced to provide evidence that the expanded reaction products completely mimic the growing strand TR sequence. Such demonstrations are important because other possible DNA expansion pathways, not based on DNA slippage, are predicted to give products during DNA replication *in vitro* that vary from the growing strand TR sequence (22, 43). Thus, *in vitro* DNA synthesis models are beginning to successfully mimic many facets of TR expansion *in vivo*.

Here we report using the lagging strand replication model (29) together with h-pol β and *Escherichia coli* Klenow pol I to demonstrate that large expansions occur when these polymerases copy GAA and TTC TRs *in vitro*. DNA sequencing and restriction enzyme cleavage demonstrate that the expansions have the correct repeat sequences. EM and nuclease cleavage experiments demonstrate that the respective expanded TTC and GAA growing strand repeats form hairpin structures and thus self-anneal.

EXPERIMENTAL PROCEDURES

GAA and TTC Expansion Reactions—Pol β replication reactions contained, in a final volume of 40 μl, a 0.28 mM concentration of either all four dNTPs or just the two dNTPs required for replication of the template repeat tract (as indicated), 0.015 mCi of ^{32}P-labeled dCTP or dATP, 5 μM annealed primer-template, and 0.025 units/μl of h-pol β (purchased from CHIMERx or purified to >95% purity as described

elsewhere (44) and shown to have similar replication kinetics and expansion behavior to h-pol β generously provided by Sam Wilson) and replication buffer (final concentrations: 50 mM Tris-HCl (pH 8.0); 10 mM MgCl$_2$, 2.5% glycerol, 20 mM NaCl, and 2 mM dithiothreitol). The reactions, which were started by the addition of dNTPs and polymerase, were incubated at 42 °C for the indicated times. Reactions were stopped by the addition of EDTA to 30 mM and cooling to 4 °C.

Klenow pol I replication reactions contained, in a final volume of 40 μl, 50 mM Tris-HCl (pH 8.0), 0.4 mM MgSO$_4$, 5 μM primer and template, 2 mM dithiothreitol, a 2 mM concentration of either all four dNTPs or the two dNTPs required to replicate the template repeat tract (as indicated), and 0.1 unit/μl of Klenow polymerase (Promega). Reactions were started by the addition of dithiothreitol, dNTPs, and polymerase, incubated for 4 h at 37 °C, and stopped by adding EDTA to 30 mM and cooling to 4 °C. Reaction products were resolved by 8% PAGE in 6 M urea.

The sequences of the primer and template were (TTC)$_3$ and 5′-dACTGTGTCTGTC(GAA)$_{10}$GCGACCTGATCC, respectively, for TTC expansions and (GAA)$_3$ and 5′-dACTGTGTCTGTC(TTC)$_{10}$GCGACCT-GATCC for GAA expansions. For study of the effects of an abasic site on expansion, the abasic site analog THF was substituted at the 5′-end of the repeat tract to give 5′-dACTGTGTCTGTC(FAA)(GAA)$_9$GCGACCT-GATCC and 5′-dACTGTGTCTGTC(FTC)(TTC)$_9$GCGACCTGATCC as described elsewhere (29). F indicates placement of THF within the repeat tract. All template strands were synthesized to contain a three-carbon spacer (Glen Research) at their 3′ terminus that cannot be removed or extended by DNA polymerase (not shown). Primers and templates were tested individually for their ability to support expansion under the above reaction conditions. Only primed templates supported TRE under our reaction conditions.

DNA Sequencing—The replication products from an 80-μl nonradio-active expansion reaction were separated by electrophoresis on a 2% agarose gel. A gel slice containing expansion products between 200 and 400 bp was excised from the gel, and the products were isolated with GenElute EtBr Minus spin columns (Sigma) and then dried. The expanded DNA was reconstituted in replication buffer (final volume 20 μl), annealed to a 12-mer primer sequence (final concentration 62 μM), and replicated for 60 min at 37 °C to give double-stranded product. The product was purified through a G-25 MicroSpin column (Amersham Biosciences). 3′-A overhangs were added for cloning using *Taq* polymerase in PCR buffer (Roche Molecular Biochemicals). The products were cloned into pCR2.1-TOPO using a TOPO TA cloning kit (Invitrogen) following the manufacturer's instructions. Plasmids were isolated from positive colonies using a QIAprep Spin miniprep kit (Qiagen). The plasmids were analyzed for large inserts via electrophoresis on 1% agarose gels. Only DNAs with inserts within the expected size range were sequenced. DNA sequencing was performed by the University of North Carolina Chapel Hill Automated DNA Sequencing Facility.

Electron Microscopy—Nonradiolabeled DNA replication using biotin-labeled oligodeoxyribonucleotides (oligonucleotides) was performed as described above. The oligonucleotides were synthesized as above to contain a biotin-labeled nucleotide at their 5′ termini. DNA replication products were prepared for EM as described elsewhere (45). TTC expansion products were cross-linked by adding 4′-hydroxymethyltrioxsalen to 0.25 μg/μl (41-μl total reaction volume) and exposing the mixture, on ice, to 366-nm UV light for 1 h. GAA expansion products were cross-linked by adding cisplatin·AgNO$_3$ to 13.3 mM (60-μl total volume) and incubating for 2 h at 37 °C.

Endonuclease Digestion—Expansion products used for P1 digestion were generated in reactions prepared to contain 0.03 μM triplet repeat primer and template and 0.03 μM h-pol β in a final volume of 20 μl. The reactions were carried out as described above. Control dsDNA was the product of a fill-in reaction using h-pol β, a "random" sequence DNA template (5′-ACTGTGTCTGTCAGGCTATCGATAGACAGTACTGCA-TACAGAGCGACCTGATCC), a primer (5′-GGATCAGGTCGC), and radiolabeled dNTPs, under replication conditions described above. Control ssDNA was made by end labeling the random DNA template using T4 polynucleotide kinase (New England Biolabs) and [^{32}P]ATP.

For P1 nuclease digestion, 0.3 pmol (total DNA) of h-pol β expansion products, control dsDNA, and control ssDNA were each added to P1 buffer containing 200 mM NaCl, 50 mM sodium acetate (pH 7.4), 1 mM ZnSO$_4$, 5% glycerol, and 0.0016 units of P1 nuclease (Sigma) in a final volume of 30 μl. Reactions were incubated at 37 °C for 3 min and stopped by adding EDTA to 50 mM and cooling to 4 °C. Excess salt was removed by passing the digestion products twice through G-25 Micro-Spin columns (Amersham Biosciences) before denaturing (7.5 M urea) 12% PAGE analysis.

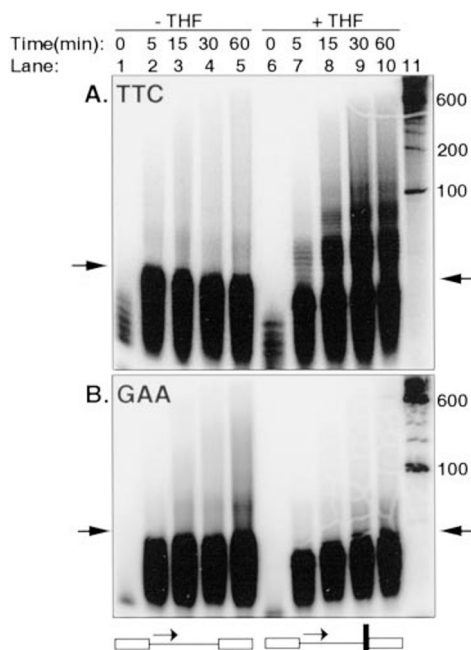*Mbo*II, *Mnl*I, and *Bbs*I (New England Biolabs) digests were per-

FIG. 1. **Expansion of TTC (*A*) and GAA (*B*) growing strand repeats during DNA replication *in vitro*.** TTC and GAA refers to the growing strand sequences that expand beyond the replication lengths (*arrow*) predicted from the lengths of the respective templates. A phosphor image of denaturing 8% polyacrylamide gel (6 M urea, ~50 °C) is shown. *Lanes 1–5* (−*THF*) and *lanes 6–10* (+*THF*), expansion times 0–60 min. *Lane 11*, 100-base pair ladder. Schematics demonstrating the template-primer construct and placement of THF are shown *under* the *gel*. The unique flanking sequences are indicated by *open rectangles*. The *solid lines* indicate the template repeat tracts. The *arrows* represent the TTC (*A*) and GAA (*B*) growing strand primers. The *solid vertical rectangle* indicates the placement of THF in the template strand at the 5′-end of the repeat tract (see Ref. 29 for detailed discussion of this template-primer construct).

formed in digestion buffer (30-μl total volume) containing 50 mM NaCl, 10 mM Tris-HCl (pH 7.9), 10 mM MgCl$_2$, and 1 mM dithiothreitol. *Mnl*I reactions also contained bovine serum albumin (100 μg/ml). For h-pol β expansion products, 27 pmol of expansion products were added to 5 units of restriction enzyme in digestion buffer and incubated at 37 °C for 30 min. Reactions were stopped by adding EDTA to 24 mM and cooling to 4 °C. Some samples were boiled in digestion buffer for 5 min and allowed to reanneal at room temperature before digestion. For *Mbo*II digestion of synthetic oligonucleotides, 99- and 51-mer GAA and TTC repeat tracts were synthesized by the LCCC DNA synthesis facility, gel-purified, and end-labeled with T4 polynucleotide kinase (New England Biolabs). End-labeled oligonucleotides were incubated in replication buffer overnight, and 0.5 pmol was digested with a total of 7.5 units of *Mbo*II (2.5 units every 1 h were added to give a final reaction volume of 31 μl) in digestion buffer at 37 °C for 3 h.

RESULTS

GAA and TTC Repeat Expansions—To study expansion of the FRDA repeats GAA and TTC, we employed an *in vitro* replication system previously shown to generate massive expansions of AAT, ATT, and CAG growing strand repeats using Klenow pol I in the presence of an abasic site analog that blocks replication (29). The analog, THF, was placed at the downstream 5′-end of the template repeat tract (Fig. 1) (29) (see "Experimental Procedures"). The abilities of GAA and TTC growing strand repeats to expand (where expansion means to a length greater than predicted from the length of the template repeat tract) varied with the polymerase used and the composition of the repeat (Fig. 1). Expansion was manifested by gel electrophoresis as resolved bands and, at higher molecular weights, an unresolved smear composed of many lengths of expansion products (Fig. 1). Klenow pol I generally gave larger expansions than h-pol β (not shown). GAA was expanded by

both polymerases without the need to incorporate a replication-blocking DNA lesion in the TTC template strand. In fact, GAA growing strand expansion was reduced by the presence of THF at the end of the TTC template repeat tract (Fig. 1). TTC also expanded in the absence of a blocking lesion but showed significantly more expansion in the presence of the replication block in the GAA template strand (Fig. 1). As found for other growing strand repeats (29), expansion was much greater when replication was primed from within the repeat tract compared with priming from the upstream 3′ unique flanking sequence of the template strand (not shown).

DNA Sequencing—The GAA and TTC TR expansion products were replicated and cloned for DNA sequencing as detailed under "Experimental Procedures." For GAA, 78 colonies were isolated, and the plasmids were screened by size. Thirteen of the plasmids appeared by size to have inserts and were sequenced. Six were found by sequencing (not shown) to have inserts that varied in size from 6 to 66 repeats of perfect GAA (6, 8, 8, 11, 47, 61, and 66 repeats). For TTC, 40 colonies were isolated, and the plasmids were screened for inserts by size. Fifteen of the plasmids appeared to have inserts and were sequenced. Eight were found by sequencing (not shown) to have inserts that varied in size from 5 to 140 repeats of perfect TTC (5, 8, 10, 28, 32, 46, 62, and 140 repeats). Larger expansions were most likely not recovered because they are reported to be unstable in *Escherichia coli* (4).

Electron Microscopy—To visualize their secondary structures, the GAA and TTC growing strand expansion products from h-pol β and Klenow pol I were cross-linked, where indicated, by psoralen/UV light for TTC primer expansion products and cis-platinum for GAA expansion products. Using primer and template strands synthesized to have biotin-labeled nucleotides at their 5′-ends, we were able to recognize the 5′-ends of the expansion products by EM after the addition of streptavidin (molecules containing *visible dots* in Fig. 2). The products were visualized by EM using standard methods described elsewhere (45). Similar results were obtained without cross-linking (Fig. 2). 70 and 75% of the respective GAA and TTC expansion products (372 total molecules observed) were visually tagged by streptavidin at only one end of the expanded duplex when either the primer alone or both the primer and template strands were labeled with biotin (Fig. 2, *A–D*). The other 25 ± 7% ($n = 5$ for TTC, where n is the number of independent experiments) and 30% ± 1% ($n = 2$ for GAA) of the products were visually tagged at both ends (Fig. 2*E*). No molecules visually tagged at both ends were observed when biotin-labeled template was used with unlabeled primer for expansion. Titrating excess primer molecules with twice the concentration of template DNA reduced the occurrence of double end-tagged TTC and GAA expansion products (293 molecules observed) to 13 ± 4% ($n = 2$) and 18 ± 6% ($n = 2$), respectively.

Endonuclease Digestion—The secondary structures of the GAA and TTC growing strand expansion products were also probed using restriction enzyme *Mbo*II and single strand endonuclease P1. *Mbo*II cleaves only dsDNA and specifically recognizes the sequences GAAGA and TCTTC, whereas nuclease P1 cleaves ssDNA randomly. Nuclease P1 hydrolyzes dsDNA slowly. The expansion products were clearly susceptible to cleavage by *Mbo*II in a time-dependent manner (Fig. 3, *A* and *B*) but not by enzymes (*Bbs*I and *Mnl*I) that recognize different, but related, sequences (GAAGAC/GTCTTC and GAGG/CCTC, respectively) (not shown). The expansion products were 96% digested by *Mbo*II in 30 min (Fig. 3, *A* and *B*). Boiling the GAA and TTC expansion reaction products in replication buffer followed by incubation at room temperature gave back *Mbo*II-susceptible products (89% digested in 30 min) (Fig. 3*A*, *lane 8*).
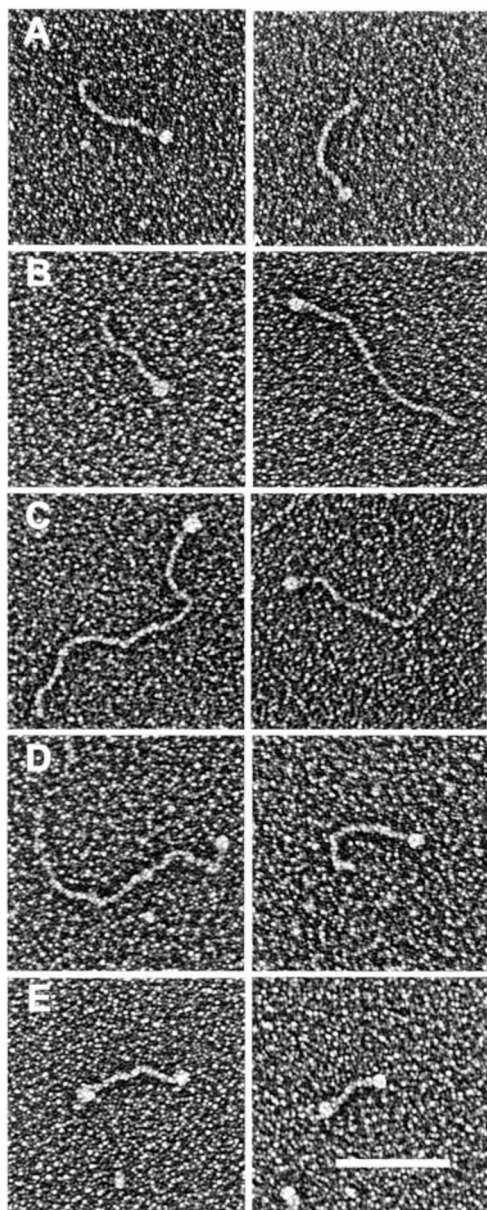
FIG. 2. **Electron microscopic visualization of expansion intermediates.** *A*, TTC growing strand expansion using Klenow pol I, template (GAA), and primer (TTC) tagged with biotin-streptavidin (*dots* at the ends of DNA molecules), cross-linked with psoralen and UV light. *B*, same as *A* but using h-pol β. *C*, GAA growing strand expansion using Klenow pol I, only primer-tagged with biotin-streptavidin (see *dots*) and no cross-linking. *D*, GAA growing strand expansion using h-pol β, primer tagged with biotin-streptavidin, cross-linked with cis-platinum. *E*, TTC growing strand expansion (*left panel*, only primer-tagged with biotin streptavidin) and GAA growing strand expansion (*right panel*, primer and template tagged with biotin-streptavidin), both using h-pol β. For visual tagging, see *dots* at both ends of DNA molecules. Both expansion products were cross-linked. The micrographs (*A–E*) are shown in reverse contrast. *Bar*, 100 nm equivalent to ~300 base pairs of B-form DNA.

Nuclease P1 was titrated against random sequence 54-base pair (bp) dsDNA and the random sequence template strand to find the concentration of nuclease P1 that best differentiated between dsDNA and ssDNA by cleavage efficiency (not shown). Conditions were found that cleaved 98 ± 2% ($n = 4$) of the ssDNA, as measured by loss of the 54-mer band, but only 37 ± 7% ($n = 4$) of the dsDNA (Fig. 4, *lanes 3–6*). Under these conditions, 68 ± 10% ($n = 3$) of the total expanded GAA growing strand DNA (Fig. 4) and 46 ± 8% ($n = 3$) of the total expanded TTC growing strand DNA (not shown) were digested
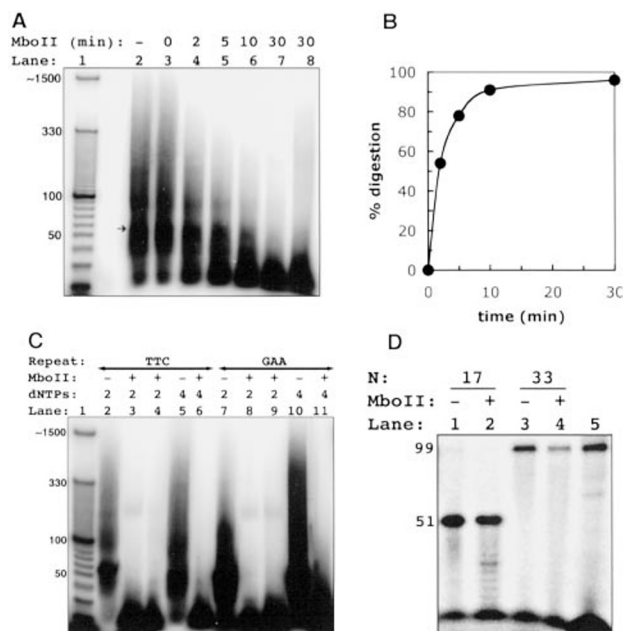

FIG. 3. **Digestion of triplet repeat expansion products by *Mbo*II restriction endonuclease.** *A*, digestion of TTC growing strand expansion products. *Lane 1*, molecular weight markers. *Lane 2*, no enzyme. *Lanes 3–7*, 5 units of *Mbo*II. *Lane 8*, expansion products were boiled, reannealed, and treated with *Mbo*II. Native 6% PAGE resolved the digestion products. Gels were analyzed using a PhosphorImager. The *arrow* indicates finished replication of template molecule. *B*, quantitation of the relative amounts of *Mbo*II digestion in *A*. The areas *above* the *arrow* in *lanes 3–7* were measured by a PhosphorImager and plotted as percentage decrease (increase in digestion) relative to *lane 3*. *C*, digestion of TTC and GAA growing strand expansion products from replication with two *versus* four dNTPs. *Lane 1*, molecular weight markers. *Lanes 2–6*, TTC growing strand expansion. *Lanes 7–11*, GAA growing strand expansion. h-pol β expansions in *lanes 5*, *6*, *10*, and *11* used all four dNTPs (dATP, dGTP, dCTP, and dTTP), those in *lanes 2–4* contained only dTTP and dCTP, and those in *lanes 7–9* contained only dGTP and dATP. *Lanes 2*, *5*, *7*, and *10*, no *Mbo*II. *Lanes 3*, *4*, *6*, *8*, *9*, and *11* were digested with 5 units of *Mbo*II. *Lanes 4* and *9*, reaction products were boiled and reannealed before digestion with *Mbo*II. The presence of two *versus* four dNTPs is indicated by the numbers *2* and *4 above* the *gel* in the *dNTPs row*. *D*, *Mbo*II digestion of (TTC)$_{17}$ and (TTC)$_{33}$. *Lane 1*, (TTC)$_{17}$ and no enzyme. *Lane 2*, digestion of (TTC)$_{17}$ with *Mbo*II. *Lane 3*, (TTC)$_{33}$ and no enzyme. *Lane 4*, digestion of (TTC)$_{33}$ with *Mbo*II. *Lane 5*, 100-bp marker. Radiolabeled reaction products were resolved by denaturing 8% PAGE containing 6 M urea and analyzed by a PhosphorImager. Repeat tracts (99-mer and 51-mer) were synthesized by the LCCC DNA synthesis facility and gel-purified. End-labeled (using T4 polynucleotide kinase and [^{32}P]ATP) 99- and 51-mer (0.5 pmol) were incubated in expansion buffer for 3 h and digested with *Mbo*II (7.5 units total; 2.5 units added at the beginning of each hour for 3 h).

by nuclease P1. Total expanded DNA refers to all products with higher mobility than the fully replicated template (indicated by the *arrow* in Fig. 4 at 41-base length; since priming is within the repeat tract, the longest labeled replicated product expected on a denaturing gel is 41-mer). Digestion of ssDNA by nuclease P1 was characterized by almost complete loss of the ssDNA band and production of a wide range of faster mobility products dispersed below the starting material (Fig. 4, *lane 6*). Nuclease P1 favors production of 5′ mononucleotides. The labeled mononucleotide products were mostly removed by gel filtration to remove salt for gel analysis. The effects of nuclease P1 on the expanded products did not follow this pattern. As seen from the digestion of the GAA growing strand expansion products by nuclease P1 (Fig. 4, *lanes 1* and *2*), the expanded intermediate bands at ~90 and 120 bp are each apparently reduced in size by less than 10 bp. Quantitating the relative densities of the apparent 90-bp region (Fig. 4, *lane 1*) and 80-bp
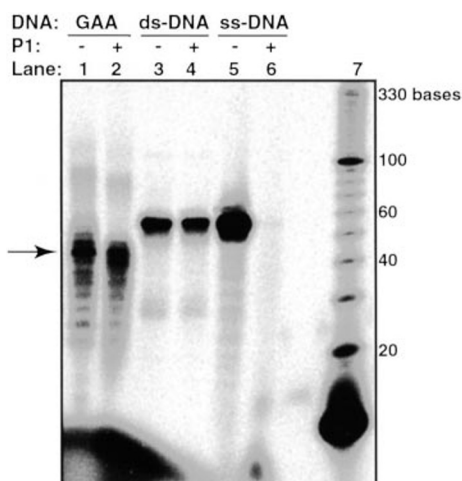
FIG. 4. **Digestion of GAA growing strand expansion products by P1 nuclease.** A phosphor image is shown of 12% denaturing polyacrylamide gel with 7.5 M urea. *Lanes 1* and *2*, GAA growing strand expansion. *Lanes 3* and *4*, dsDNA control. *Lanes 5* and *6*, ssDNA control. *Lane 7*, molecular weight markers. *Lanes 1*, *3*, and *5*, no P1. *Lanes 2*, *4*, and *6*, 0.0016 units of P1 added. An *arrow* indicates finished replication of template molecule.
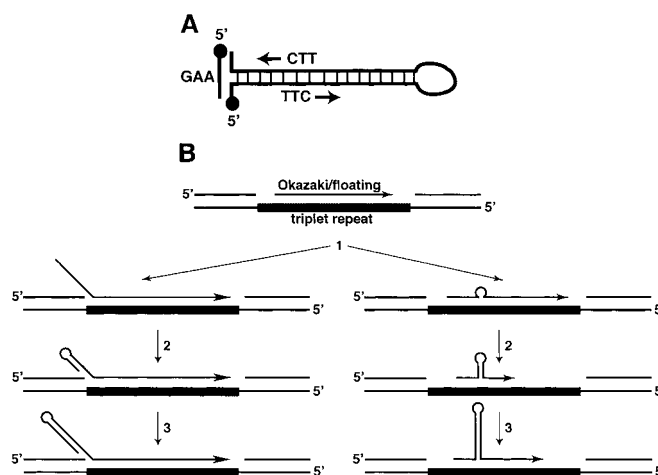

FIG. 5. **Schematics of secondary structure consistent with EM visualized expansion products labeled at one end (*A*) and generation of triplet repeat expansion using floating primer (*B*).** *A*, the template contains a GAA repeat tract. The TTC expanded during replication to give a TTC expanded region that formed a hairpin. The 5′-ends of the primer and template are labeled with biotin-streptavidin. The template region is too small, compared with the expanded region, to resolve by EM. *B*, proposed mechanism to generate type II expansions during DNA replication *in vitro*. Two mechanisms are shown. *Left panel* (*1*), slippage and 5′-tail formation; *right panel* (*1*), one or both ends slip to form an internal loop. Following the initial slippage, step 2 is formation of the initial hairpin, and step 3 shows growing hairpin formation caused by continued replication and slippage within the TR tract.

region (*lane 2*), assuming that all of the density of the apparent 80-bp band came from the 90-bp band, showed that nuclease P1 digested $23 \pm 4\%$ ($n = 3$) of the 80-bp region. This was significantly less than the overall amount of digestion of expanded product (see above). Therefore, the ability of nuclease P1 to cleave different size expansion products was assessed by measuring identical corresponding areas of *lanes 1* and *2* that encompassed 50–65- and 80–95-base lengths, as estimated from size markers in *lane 7*. For GAA, $85 \pm 2\%$ ($n = 3$) of the 50–65-base length products was digested, whereas $64 \pm 7\%$ ($n = 3$) of the 80–95-length products was digested by nuclease P1.

Two Versus Four dNTPs—Occasional interruptions in the expanded products caused by misinsertions of nucleotides complementary to the repeat could contribute to stabilization of secondary structure. To eliminate this possibility, the above expansion reactions for GAA and TTC were repeated using only the two required dNTPs in each reaction. Results of expansion reactions, of DNA sequencing of the expansions, and of *Mbo*II cleavage of the expansion products (Fig. 3*C*) were similar to the experimental results when all four dNTPs were present. This is additional evidence against the occurrence of GAA sequences within the expanded TTC products (and *vice versa*) being responsible for hairpin formation.

Self-annealing of Synthetic Oligonucleotides—Small synthetic GAA and TTC single-stranded DNAs do not easily self-anneal (19, 21–23). We clearly observed by EM and nuclease cleavage experiments, however, that expanded GAA and TTC growing strand DNAs form structures consistent with hairpins. The GAA and TTC expansion products that formed hairpins were significantly longer than the small synthetic GAA and TTC repeats that provided evidence against hairpin formation. We tested synthetic oligonucleotides synthesized to give GAA and TTC repeats 99 and 51 nucleotides long (TTC results shown in Fig. 4*D*). *Mbo*II cleaved both the 99-mers (GAA cleaved $17 \pm 4\%$, $n = 6$; TTC cleaved $44 \pm 16\%$, $n = 7$) and the 51-mers (GAA cleaved $12 \pm 6\%$, $n = 6$; TTC cleaved $48 \pm 10\%$, $n = 4$).

DISCUSSION

To provide evidence for the secondary structures of the intermediates involved in expansion of the repeats associated with FRDA, we generated massive expansions of GAA and TTC repeats *in vitro* using h-pol β and Klenow pol I. A lagging strand replication model was used to generate the expansions (29). The model uses a template strand with unique sequences flanking the repeat tract and a covalent block to template expansion synthesized at the 3′ terminus. A replication-blocking lesion, found necessary with this model to generate massive expansions of AAT, ATT, and CAG growing strands (29), was unnecessary for, and in fact interfered with, GAA growing strand expansion. It significantly aided TTC growing strand expansion, but TTC showed expansion without DNA damage. This is consistent with the proclivity of GAA/TTC to form intramolecular triplexes that block DNA replication primed farther upstream (26–28). The presence of the damage in the template strand may interfere with formation of the triplex when GAA is the template repeat.

Given the different possible ways to generate apparent expanded molecules (22, 43), we characterized the intermediate expansion products using DNA sequencing, EM, and endonuclease digestion. In addition, we repeated the experiments using only the two nucleotides required to replicate the template TR tract so that contaminating complementary nucleotides could not be incorporated into the expansion products at a low level. Finally, we synthesized 99- and 51-mer GAA and TTC oligonucleotides, and we determined their abilities to reanneal using endonuclease digestion experiments.

DNA sequencing of cloned reaction products from the GAA and TTC expansion reactions demonstrated that the expansion products mimicked the growing strand repeat tract. The sizes of the sequenced products are consistent with the instability of repeat tracts in *E. coli* (4). We cannot rule out the possibility that sequences that varied from the GAA/TTC repeats were preferentially deleted during cloning in *E. coli*. That seems unlikely, however, since interrupted sequences are relatively stable in *E. coli* (46), and the repeated isolation of pure GAA/TTC sequences argues that the *in vitro* expansion products mimicked the growing strand repeat sequences.

The thicknesses and secondary structures of the DNA products determined by EM (Fig. 2) are consistent with the EM of duplex DNA (45). The spreading technique used to prepare the

DNA for EM collapses ssDNA (45). The 5′ ends of the DNA molecules were tagged with biotin-streptavidin that is easily visualized by EM. Most of the GAA and TTC expanded DNA molecules seen by EM were tagged at only one end. This is consistent with hairpin formation (Fig. 5*A*). The expanded growing strand becomes double-stranded by folding back to give a hairpin. The large size of the expanded DNA relative to the replicated template means that the molecule will appear tagged at one end independent of whether the primer, template, or both are biotin-labeled (Fig. 5*A*). The duplex DNA molecules observed increased in size with increasing reaction time. A percentage of the GAA and TTC growing strand expansion products observed by EM were tagged at both ends only when the primer was biotin-labeled. One possibility is that biotin-labeled primer molecules interacted with single-stranded loops formed by DNA fold back to give the hairpin structures (Fig. 5*A*). To test this possibility, the concentration of unlabeled template DNA in the expansion reaction was doubled to compete for unannealed primer molecules. The percentage of double-stranded structures observed by EM to be tagged at both ends was reduced by almost half by the 2-fold increase in template concentration. The reduction supports the notion that a small amount of biotin-labeled primer annealed to the expanded end of some of the double-stranded structures observed by EM (Fig. 5*A*). The annealing may be because of the presence of a single-stranded loop. The EM results eliminate the possibility that the observed expansion products could come from any kind of fill-in reactions involving linear arrays of multiple overlapping molecules (22, 43). Such reactions would have produced visual tags between the two ends of the expanded molecules when the primer and template were biotin-labeled. The products of such reactions are also predicted to have DNA sequences that vary from true TR expansions.

Endonuclease digestion gave independent confirmation of the double-stranded nature of the expanded primer strand. *Mbo*II recognizes the double-stranded sequences GAAGA and TCTTC and cleaves downstream of both sequences (47). *Mbo*II cleaved both the GAA and TTC expansion products in a time-dependent manner (Fig. 3*A*). On the other hand, the expanded products were relatively resistant to nuclease P1 under conditions that cleaved 98% of the control ssDNA but only 37% of the control dsDNA. Moreover, the results of nuclease P1 digestions of GGA repeat DNA provided evidence that longer dsDNA expansion products (80–95 bp) are more resistant to cleavage with nuclease P1 than shorter products (50–65 bp). These results suggest that GAA hairpin formation is dependent on DNA length and may help explain why earlier studies (21–23), which used shorter DNA lengths, failed to demonstrate GAA and TTC hairpin formation under physiological conditions. To test this hypothesis, longer GAA and TTC TRs (17 and 33 repeats) were synthesized and were shown to be partially susceptible to *Mbo*II cleavage under physiological conditions of temperature and salt concentration. Thus, the endonuclease cleavage experiments, like the EM experiments, support hairpin formation within the expanded strands. GAA and TTC hairpins are most likely stabilized by GA and TC base pairs, respectively. Alternating d(GA)$_n$ and d(TC)$_n$ repeats self-anneal (48–50), and possible base-pairing schemes have been determined from crystallographic and NMR studies (50–53).

Hairpin formation occurred during DNA replication *in vitro* using a primer complementary to the repeat tract. Priming within the repeat tract gave dramatic amounts of type II triplet repeat expansion. No type II expansion was detected when replication was primed from the unique upstream flanking sequence. The difference in expansion abilities argues that free 3′- and 5′-ends within the repeat tract are important for expansion. Free 3′- and 5′-ends could arise from the occurrence of an Okazaki fragment completely within the boundaries of a TR tract (31, 32) (Fig. 5*B*). The association with free 5′-ends is consistent with the preferred occurrence of TR expansion during DNA lagging strand replication (3–5). The requirement for a free 5′-end suggests at least two different possible pathways for DNA slippage. The first pathway predicts that the growing strand slips back along the template repeat tract, resulting in a 5′ flap caused by displacement of the 5′-end when it tries to overlap the unique upstream flanking sequence. This flap should be a substrate for flap endonuclease-1 (14–18, 20), unless either hairpin formation occurs to protect the 5′-end from flap endonuclease-1 (20) or flap endonuclease-1 activity is absent. The second pathway involves an accordion-like slippage that predicts movement of both ends of the growing strand to form a loop within the growing strand. The loop should be a substrate for mismatch repair (19), unless either the loop grows large enough to form a stable hairpin structure or mismatch repair is defective (19). The requirement for longer repeat tracts for GAA and TTC hairpin formation suggests that *in vivo* FRDA-associated premutations may be longer than those for the other triplet repeat diseases; the longer premutagenic TR tract may enable a single replication slippage event to generate an initial expansion long enough to form a hairpin predicted to protect the expanded single-stranded region from repair.

CONCLUSION

The results demonstrate massive expansions of GAA and TTC trinucleotide repeats implicated in Friedreich's ataxia during DNA replication by h-pol β and Klenow pol I. The results further demonstrate that both expansions form hairpins *in vitro* and thus self-anneal. The results predict that all of the triplet repeat diseases are mediated by hairpin formation.

REFERENCES

1. Tóth, G., Gáspári, Z., and Jurka, J. (2000) *Genome Res.* **10,** 967–981
2. Cummings, C. J., and Zoghbi, H. Y. (2000) *Hum. Mol. Genet.* **9,** 909–916
3. Kunst, C. B., and Warren, S. T. (1994) *Cell* **77,** 853–861
4. Kang, S., Jaworski, A., Ohshima, K., and Wells, R. D. (1995) *Nat. Genet.* **10,** 213–218
5. Freudenreich, C. H., Stavenhagen, J. B., and Zakian, V. A. (1997) *Mol. Cell. Biol.* **17,** 2090–2098
6. Bambara, R. A., Murante, R. S., and Henricksen, L. A. (1997) *J. Biol. Chem.* **272,** 4647–4650
7. Waga, S., and Stillman, B. (1998) *Annu. Rev. Biochem.* **67,** 721–751
8. Trinh, T. Q., and Sinden, R. R. (1991) *Nature* **352,** 544–547
9. Mitas, M., Yu, A., Dill, J., Kamp, T. J., Chambers, E. J., and Haworth, I. S. (1995) *Nucleic Acids Res.* **23,** 1050–1059
10. Chen, X., Mariappan, S. V., Catasti, P., Ratliff, R., Moyzis, R. K., Laayoun, A., Smith, S. S., Bradbury, E. M., and Gupta, G. (1995) *Proc. Natl. Acad. Sci. U. S. A.* **92,** 5199–5203
11. Nadel, Y., Weisman-Shomer, P., and Fry, M. (1995) *J. Biol. Chem.* **270,** 28970–28977
12. Usdin, K. (1998) *Nucleic Acids Res.* **26,** 4078–4085
13. Veaute, X., and Fuchs, R. P. P. (1993) *Science* **261,** 598–600
14. Tishkoff, D. X., Filosi, N., Gaida, G. M., and Kolodner, R. D. (1997) *Cell* **88,** 253–263
15. Gordenin, D. A., Kunkel, T. A., and Resnick, M. A. (1997) *Nat. Genet.* **16,** 116–118
16. Schweitzer, J. K., and Livingston, D. M. (1998) *Human Mol. Genet.* **7,** 69–74
17. Spiro, C., Pelletier, R., Rolfsmeier, M. L., Dixon, M. J., Lahue, R. S., Gupta, G., Park, M. S., Chen, X., Mariappan, S. V. S., and McMurray, C. T. (1999) *Mol. Cell* **4,** 1079–1085
18. White, P. J., Borts, R. H., and Hirst, M. C. (1999) *Mol. Cell. Biol.* **19,** 5675–5684
19. Moore, H., Greenwell, P. W., Liu, C.-P., Arnheim, N., and Petes, T. D. (1999) *Proc. Natl. Acad. Sci. U. S. A.* **96,** 1504–1509
20. Henricksen, L. A., Tom, S., Liu, Y., and Bambara, R. A. (2000) *J. Biol. Chem.* **257,** 16420–16427
21. Gacy, A. M., Goellner, G. M., Juranic, N., Macura, S., and McMurray, C. T. (1995) *Cell* **81,** 533–540
22. LeProust, E. M., Pearson, C. E., Sinden, R. R., and Gao, X. (2000) *J. Mol. Biol.* **302,** 1063–1080
23. Suen, I.-S., Rhodes, J. N., Christy, M., McEwen, B., Gray, D. M., and Mitas, M. (1999) *Biochim. Biophys. Acta* **1444,** 14–24
24. Gacy, A. M., Goellner, G. M., Spiro, C., Chen, X., Gupta, G., Bradbury, E. M., Dyer, R. B., Mikesell, M. J., Yao, J. Z., Johnson, A. J., Richter, A., Melancon, S. B., and McMurray, C. T. (1998) *Mol. Cell* **1,** 583–593
25. Sakamoto, N., Chastain, P., Parniewski, P., Ohshima, K., Pandolfo, M.,

Griffith, J. D., and Wells, R. D. (1999) *Mol. Cell* **3,** 465–475

26. Wells, R. D., Collier, D. A., Hanvey, J. C., Shimizu, M., and Wohlrab, F. (1988) *FASEB J.* **2,** 2939–2949
27. Frank-Kamenetskii, M. D., and Mirkin, S. M. (1995) *Annu. Rev. Biochem.* **64,** 65–95
28. Ohshima, K., Montermini, L., Wells, R. D., and Pandolfo, M. (1998) *J. Biol. Chem.* **273,** 14588–14595
29. Lyons-Darden, T., and Topal, M. D. (1999) *J. Biol. Chem.* **274,** 25975–25978
30. Viguera, E., Canceill, D., and Ehrlich, S. D. (2001) *EMBO J.* **20,** 2587–2595
31. Eichler, E. E., Holden, J. J. A., Popovich, B. W., Reiss, A. L., Snow, K., Thibodeau, S. N., Richards, C. S., Ward, P. A., and Nelson, D. L. (1994) *Nat. Genet.* **8,** 88–94
32. Richards, R. I., and Sutherland, G. R. (1994) *Nat. Genet.* **6,** 114–116
33. Sinden, R. R. (2001) *Nature* **411,** 757–758
34. Schlötterer, C., and Tautz, D. (1992) *Nucleic Acids Res.* **20,** 211–215
35. Ji, J., Clegg, N. J., Peterson, K. R., Jackson, A. L., Laird, C. D., and Loeb, L. A. (1996) *Nucleic Acids Res.* **24,** 2835–2840
36. Petruska, J., Hartenstine, M. J., and Goodman, M. F. (1998) *J. Biol. Chem.* **273,** 5204–5210
37. Nakayabu, M., Miwa, S., Suzuki, M., Izuta, S., Sobue, G., and Yoshida, S. (1998) *Nucleic Acids Res.* **26,** 1980–1984
38. Lyons-Darden, T, and Topal, M. D. (1999) *Nucleic Acids Res.* **27,** 2235–2240
39. Lindahl, T. (1992) *Nature* **362,** 709–715
40. Paz-Elizur, T., Takeshita, M., and Livneh, Z. (1997) *Biochemistry* **36,** 1766–1773
41. Miller, H., and Grollman, A. P. (1997) *Biochemistry* **36,** 15336–15342
42. Hartenstine, M. J., Goodman, M. F., and Petruska, J. (2002) *J. Biol. Chem.* **277,** 41379–41389
43. Tuntiwechapikul, W., and Salazar, M. (2002) *Biochemistry* **41,** 854–860
44. Beard, W. A., and Wilson, S. H. (1995) *Methods Enzymol.* **262,** 98–107
45. Griffith, J. D., and Christiansen, G. (1978) *Annu. Rev. Biophys. Bioeng.* **7,** 19–35
46. Kang, S., Ohshima, K., Jaworski, A., and Wells, R. D. (1996) *J. Mol. Biol.* **258,** 543–547
47. Roberts, R. J., and Macelis, D. (2000) *Nucleic Acids Res.* **28,** 306–307
48. Huertas, D., Bellsolell, L., Casasnovas, J. M., Coll, M., and Azorin, F. (1993) *EMBO J.* **12,** 4029–4038
49. Shiber, M. C., Braswell, E. H., Klump, H., and Fresco, J. R. (1996) *Nucleic Acids Res.* **24,** 5004–5012
50. Jaishree, T. N., and Wang, A. H.-J. (1994) *FEBS Lett.* **337,** 139–144
51. Hunter, W. N., Brown, T., and Kennard, O. (1986) *J. Biomol. Struct. Dyn.* **4,** 173–191
52. Privé, G. G., Heinemann, U., Chandrasegaran, S., Kan, L. S., Kopka, M. L., and Dickerson, R. E. (1987) *Science* **238,** 498–504
53. Gao, Y. G., Robinson, H., Sanishvili, R., Joachimiak, A., and Wang, A. H. (1999) *Biochemistry* **38,** 16452–16460