Inhibitory Effects of Expanded GAA-TTC Triplet Repeats from Intron I of the Friedreich Ataxia Gene on Transcription and Replication in Vivo*

(Received for publication, February 3, 1998, and in revised form, March 23, 1998)

Keiichi Ohshima‡‡, Laura Montermini‡‡, Robert D. Wells, and Massimo Pandolfo‡‡‡ ** ‡‡ §§
From the ‡Centre de Recherche Louis-Charles Simard, C.H.U.M., Campus Notre-Dame, 1560 rue Sherbrooke est, Montréal, Québec H2L 4M1, Canada, the §Institute of Biosciences and Technology, Center for Genome Research, Department of Biochemistry and Biophysics, Texas A&M University, Texas Medical Center, Houston, Texas 77030-3303, the **Department of Medicine, Université de Montréal, Montréal, Québec H2L 4M1, Canada, and the §§Departments of Neurology and Neurosurgery, McGill University, Montréal, Québec H3A 2B4, Canada

Friedreich ataxia (FRDA) is associated with the expansion of a GAA-TTC triplet repeat in the first intron of the frataxin gene, resulting in reduced levels of frataxin mRNA and protein. To investigate the mechanisms by which the intronic expansion produces its effect, GAA-TTC repeats of various lengths (9 to 270 triplets) were cloned in both orientations in the intron of a reporter gene. Plasmids containing these repeats were transiently transfected into COS-7 cells. A length- and orientation-dependent inhibition of reporter gene expression was observed. RNase protection and Northern blot analyses showed very low levels of mature mRNA when longer GAA repeats were transcribed, with no accumulation of primary transcript. Replication of plasmids carrying long GAA-TTC tracts (~250 triplets) was greatly inhibited in COS-7 cells compared with plasmids carrying (GAA/TTC)9 and (GAA/TTC)30. Replication inhibition was five times greater for the plasmid whose transcript contains (GAA)250 than for the plasmid whose transcript contains (UUC)270. Our in vivo investigation revealed that expanded GAA-TTC repeats from intron I of the FRDA gene inhibit transcription rather than post-transcriptional RNA processing and also interfere with replication. The molecular basis for these effects may be the formation of non-B DNA structures.

Friedreich ataxia (FRDA) is the first autosomal recessive neurodegenerative disease found to be caused by the hyperexpansion of a triplet repeat sequence (TRS) (1), a GAA-TTC repeat in the first intron of the frataxin gene. The GAA-TTC expansion accounts for about 98% of all FRDA chromosomes, with the remaining ones carrying frataxin point mutations. The recessive inheritance, nature, and intronic localization of the expanded sequence make FRDA an unique case in TRS-related diseases (2–4). However, the FRDA expanded GAA-TTC repeats show meiotic and mitotic instability as for other disease-associated TRS. In FRDA chromosomes, GAA-TTC repeat units vary from about 100 to more than 1,000 whereas less than 37 repeat units are found in normal chromosomes (1, 5, 6). FRDA patients carrying two expanded GAA-TTC repeats show very low levels of mature frataxin transcript (1, 7, 8) and of frataxin (9), indicating suppressed gene expression. Such a defect may be caused either by reduced transcription or by abnormal post-transcriptional processing (1, 9, 10). Together with the identification of frataxin point mutations resulting in a defective or truncated protein, this finding defines FRDA as a frataxin deficiency disease, in accordance with its recessive inheritance. The sizes of the GAA-TTC repeats carried by each patient correlate with the age of onset and the severity of the disease, particularly for the smaller one (5). In addition, an inverse correlation between the length of the smaller GAA-TTC repeat and the residual amount of frataxin was observed in cultured cells from FRDA patients (9).

The GAA-TTC tract is a polypurine-polypyrimidine (Pur-Pyr) sequence, which may form an intramolecular triple helix in vitro under appropriate conditions of pH, metal ions concentrations, and supercoiling (4, 11). Increasing the length of the Pur-Pyr tract also promotes triplex formation (12). Ohshima et al. (13) demonstrated that plasmids containing 38, 58, and 103 GAA-TTC triplets, but not 16 triplets, showed supercoiled-relaxed relaxations when examined by two-dimensional-agarose gel electrophoresis, even at pH 8.3, suggesting that they may adopt a triple helical structure in vivo. Such structures inhibit gene expression by blocking the progression of RNA polymerase, as shown to occur for Pur-Pyr tracts both in vitro (14–17) and in vivo (18–20). However, GAA-containing RNA may also adopt a secondary structure interfering with post-transcriptional processing (21). Pur-Pyr sequences, including GAA-TTC tracts, can also interfere with DNA replication, since they have been shown to stall DNA polymerase in vitro, probably again as a consequence of intramolecular triplex formation (11, 13).

We used cloned GAA-TTC repeats to investigate the possible effect of this intronic sequence on gene expression in vivo. The cloned GAA-TTC tracts previously used by Ohshima et al. (13) contain interruptions and no FRDA-derived flanking sequence. Considering the effects of interruptions and of flanking sequences on the biological properties of TRS (3, 4, 6, 22–28), we constructed new recombinant plasmids containing from 9 to 500 GAA-TTC triplets along with some frataxin gene-derived

* This work was supported in part by National Institutes of Health Grants NS34192 (to M. P.) and GM52982 (to R. D. W.) and grants from the Muscular Dystrophy Association, and Medical Research Council of Canada (to M. P.), and the Robert A. Welch Foundation (to R. D. W.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

§ Recipient of a fellowship from the Human Frontier Science Program.

‡‡ Recipient of a fellowship from the Medical Research Council of Canada.

§§ To whom correspondence should be addressed. Tel.: 514-281-6000 (ext. 8928); Fax: 514-896-4762; E-mail: pandolm@magellan.umontreal.ca.

1 The abbreviations used are: FRDA, Friedreich ataxia; TRS, triplet repeat sequences; Pur-Pyr, polypurine-polypyrimidine; bp, base pair(s).
flanking sequence. These data evidence that such TRS inhibit transcription and possibly DNA replication in vivo.

**EXPERIMENTAL PROCEDURES**

**Plasmid Construction**—Genomic DNA from a patient carrying ~700 GAA/TTC repeats was amplified by polymerase chain reaction (1) using primers Not-Bam (5'-GGAGGGAGCGGGCTGCGGGAAGG-3') and 2500-F (5'-CAATCCAGGAGCTGACGTTT-3') and then digested with BamHI and PstI. The fragment containing (GAA-TTC)7, along with flanking sequences (352 bp 5' and 250 bp 3' to the TRS) was purified by a 1.2% agarose gel (13) and ligated into the BamHI site of pUC19. The ligation mixture was transformed into Escherichia coli SURE (Stratagene) by electroporation. The resultant recombinant plasmid contained 110 GAA/TTC triplets in orientation II (Fig. 1A). The SacI-HindIII digest was recloned into the SacI-HindIII site of pUC-18NotI and pUC19NotI (29) to give plasmids containing (GAA-TTC)α and (GAA-TTC)β, in two different orientations, designating pRW3804 (orientation I) and pRW3803 (orientation II), respectively. To generate longer repeats, the in vitro expansion method was performed as described previously (13, 29). Briefly, after pRW3804 was grown in E. coli DH10B (Life Technologies) or SURE, the SacI-HindIII digest was loaded on an agarose gel and regions above 70 repeats were eluted and ligated into pUC-18. The procedure was repeated several times, obtaining cloned GAA/TTC repeats ranging from 70 to 360 triplets.

To clone GAA/TTC repeats into pSPL3, inserts were excised from pUC18NotI-based plasmids by EcoRI and PstI digestion and cloned into the corresponding pSPL3 sites. The resulting constructs had the TRS and flanking sequences were excised from the orientation I pSPL3-based plasmids and recloned in the opposite orientation into the EcoRV site of pSPL3. Using the in vivo expansion-deletion method (13, 29), from 9 to 500 GAA/TTC repeats were eventually cloned in pSPL3.

For the construction of pMP106, a CDNA fragment was amplified by reverse transcriptase-polymerase chain reaction from total RNA isolated from COS-7 cells transfected with the Klenow fragment of E. coli DNA polymerase I (New England Biolabs) and dNTPs, and cloned into the NotI site of pSPL3. pMP129-based plasmids containing (GAA-TTC)α, were constructed by digesting the pSPL3-based plasmids with either Apal-PstI or HindIII and cloning the released inserts into the Apal-PstI or HindIII sites of pMP129, respectively.

Luciferase gene fragments (661 and 1340 bp), obtained by digesting pGL3-Control (Promega) with EcoRI or HindIII and HindIII followed by filling in the overhangs, were cloned into the EcoRV site of pSPL3 in the antisense orientation to give pMP175 and pMP177, respectively. pMP183 was constructed by inserting the HindIII digest of pMP177 into the HindIII site of pMP129.

For plasmid preparations, the recombinant plasmid DNA was transformed into E. coli SURE by electroporation and the transformant was grown in 1 liter or 100 ml of LB with 75 μg/ml ampicillin at 37 °C until the end of the logarithmic phase (OD600 = 0.9). Plasmids were isolated by the alkali lysis method (30) and purified by CsCl density gradient centrifugation or the QIAprep Plasmid Kits (Qiagen). Isolated plasmids were digested with appropriate restriction enzymes. The digest was analyzed on an agarose gel, stained with ethidium bromide, and photographed. The inserts in all of the recombinant plasmids were characterized by DNA sequencing on both strands to determine the repeat units. For plasmids containing more than 59 GAA/TTC repeats, the triplet repeat units were estimated from agarose gels and DNA sequencing to ±5 triplets.

**Cell Culture and Transient Transfection for RNA Analyses**—COS-7 cells were grown in Dulbecco’s modified Eagle’s medium (ICN Pharmaceuticals) containing 50 units/ml penicillin-streptomycin and supplemented with 2 mM glutamine and 10% fetal bovine serum (ICN Pharmaceuticals). Cultures were maintained at 37 °C in a 5% CO2 atmosphere. COS-7 cells were plated on 60-mm diameter plastic dishes at a density of 4 × 105 cells/dish. After 24 h, the cells achieved 40–70% confluence and were transfected with 1 μg of pSPL3-based plasmid (Promega) and LiAc (Life Technologies) (7 μl) according to the manufacturer’s protocol. 48 h after transfection, total RNA was isolated by the TRIzol method (Life Technologies): 800 μl of TRIzol and 200 μl of chloroform were added to the cells and the nuclear acids were subsequently recovered by precipitation with 2/3 volume of isopropl alcohol.

**RNA Protection Assays**—pMP106 and pMP125 were used as a template to make antisense RNA probes, R-MP106 and R-MP125, by in vitro transcription from the T7 and SP6 promoters, respectively, incorporating [α-32P]UTP (800 Ci/mmol, Amersham) with the MAXIscript kit (Ambion). The reaction mixtures were loaded on a 5% polyacrylamide electrophoresis gel containing 8 μM urea, and the radiolabeled RNA products were excised and eluted according to the manufacturer’s protocol. RNase protection assays were performed using the RPAII kit (Ambion): 1 μg of total RNA was hybridized to 1.2 × 106 cpm and 2.2 × 106 cpm RNA probes synthesized from pMP106 and pMP125, respectively, at 45 °C for 16 h in 20 μl of the hybridization buffer containing 80% deionized formamide, 100 mM sodium citrate (pH 6.4), 300 mM sodium acetate (pH 6.4), and 1 mM EDTA. The samples were digested with RNase A (5 μg) and RNase T1 (200 μg) of digestion buffer for 30 min at 37 °C. RNA was precipitated with ethanol and dissolved in 8 μl of gel loading buffer. Samples were fractionated on a 8% denaturing polyacrylamide gel containing 8 μM urea. After exposure of the dried gel to x-ray film at ~80 °C for 48 h, the amounts of protected products was estimated using the Alphalager version 3.0 (Alpha Innotech).

**Northern Blot Analysis**—Antisense probes R-MP107 and R-MP125 were synthesized by in vitro transcription using T7 and SP6 RNA polymerases from pMP107 and pMP125, respectively, as described above. 1 μg of total RNA was size-fractionated on a 1.0% agarose-formaldehyde gel (Ambion) and transferred to Hybond-N nylon membrane (Amersham) using the NorthernMax kit (Ambion). After UV cross-linking, blots were hybridized to either probe R-MP107 or R-MP125 at 65 °C for 16 h. The membrane was exposed to x-ray film at ~80 °C for 24 h and 7 days for probes R-MP125 and R-MP107, respectively.

**Reporter Gene Expression Assays**—COS-7 cells were plated on 100-mm diameter plastic dishes at a density of 8 × 105 cells/dish. After 24 h, pMP129-based plasmid (3 μg) and a luciferase reporter plasmid pGL3-Control (Promega) (1 μg) were introduced into COS-7 cells by lipofection (Life Technologies). Transfected cells were harvested 48 h post-transfection and lysed by 900 μl of reporter lysis buffer (Promega). The β-galactosidase and luciferase assays were performed using the respective kits (Promega). Luciferase activity was measured using the AutoLumat LB953 (EG&G Berthold).

**Analysis of Replicated Plasmids in COS-7 Cells**—Plasmid DNA was recovered from transfected COS-7 cells as follows: 3 μg of pMP141, pMP180, pRW3823, pMP145, pMP165, pRW3827, pMP175, and pMP176 were separately introduced, along with 1 μg of pGL3-Control, into COS-7 cells in 100-mm diameter plastic dishes using lipofectamine300 (30 μg) as described above. Transfected cells were washed twice with ice-cold phosphate-buffered saline 4 h after transfection and incubated with 900 μl of lysis solution containing 0.6% SDS and 10 mM EDTA for 20 min at room temperature. The lysate was transferred into two 1.5-Ml microcentrifuge tubes, mixed with 450 μl of 2.5 M NaCl, and incubated at 4 °C for 16 h. After centrifugation at 4 °C for 4 min at 14,000 × g, the supernatant was extracted with phenol and chloroform once and precipitated with ethanol. The pellet was resuspended in 250 μl of TE buffer (10 mM Tris-Cl, 1 mM EDTA, pH 8.0) and precipitated with 750 μl of ethanol in the presence of 25 μl of 3 M sodium acetate (pH 5.2). After centrifugation, plasmid DNA was isolated and resuspended in H2O. 10 μg of plasmid DNA were digested with the appropriate restriction enzymes in the presence of RNase A and the digests were separated on agarose gels in 1 × TBE buffer (90 mTris-borate, 2 mM EDTA, pH 8.3) and repurified with 750 μl of ethanol in the presence of 25 μl of 3 M sodium acetate. After centrifugation, plasmid DNA was isolated and resuspended in H2O. 20 mM sodium phosphate, 0.1% SDS was added to the sample and the DNA was run on a 2% agarose gel in 1× TBE buffer. The DNA was visualized by ethidium bromide staining.
RESULTS

Cloning of GAA-TTC Triplet Repeat Sequences in Plasmids—Attempts were made to clone a polymerase chain reaction product from a FRDA patient containing (GAA-TTC)$_{230}$ in pUC19. E. coli SURE transformants harbored plasmids with a family of repeat lengths containing 110 GAA-TTC triplets at most. TTC triplets were in the leading strand template of all these recombinant plasmids (Fig. 1A, orientation II). This first result indicated that GAA-TTC repeats cloned into pUC19 are unstable in E. coli, and the instability may be related to the direction of replication, as previously observed for CTG-CAG (29) and CCG-CGG (31). The insert containing (GAA-TTC)$_{110}$ was then subcloned in both orientations into pUC18NolI (Fig. 1A, orientation I) and pUC19NolI (Fig. 1A, orientation II). The resulting recombinant plasmids, pRW3804 and pRW3803, contained 70 and 65 GAA-TTC triplets, respectively (Fig. 1A). These shorter GAA-TTC repeats were quite stable in both plasmids when grown in E. coli SURE, but pRW3804 was more stable than pRW3803 when grown in E. coli DE10B (data not shown). Hence, the stability of GAA-TTC repeats in E. coli is influenced by the direction of replication (see below, Fig. 1C).

To clone longer GAA-TTC repeats, we used the in vivo expansion method (13, 29, 32). pRW3804, in which (GAA-TTC)$_{110}$ is more stable, was chosen as the starting material. This repetitive procedure successfully generated repeats containing between 70 and 360 GAA-TTC triplets (Fig. 1A). The instability of the longest repeats was evident during their propagation (Fig. 1B). While shorter repeats as (GAA-TTC)$_{110}$ (lane 1) and (GAA-TTC)$_{150}$ (lane 3) were quite stable, (GAA-TTC)$_{230}$ (lane 5) and (GAA-TTC)$_{290}$ (lane 7) generated multiple deletion products, visible as distinct bands on a smeary background, as previously seen with CTG-CAG (29).

GAA-TTC Triplet Repeats Were More Stable in pSPL3 Than in pUC Vectors—To investigate the effects of GAA-TTC triplet repeats on transcription and splicing (see below), we subcloned several such repeats from pUC vectors into the multicloning site of pSPL3 in both orientations (Fig. 2A). The subsequent in vivo expansion-deletion procedure generated GAA-TTC repeats ranging from 9 to 500 triplets (Fig. 1A). Surprisingly, GAA-TTC repeats were much more stable in pSPL3 than in pUC vectors (Fig. 1B). pSPL3-based plasmids containing up to 270 repeats were completely stable (Fig. 1B, lanes 2, 4, and 6), and pRW3824 containing 360 repeats showed only slight instability (lane 8). As expected, constructs containing 470 and 500 repeats showed an increasing frequency of deletions (data not shown). When comparing the stability of (GAA-TTC)$_{290}$ in different orientations, the insert in orientation II (Fig. 1C, lane 9) was less stable than the one in orientation I (lane 8), as observed for pUC-based plasmids. To avoid contamination by deleted products, we used only pSPL3-based plasmids in which the cloned GAA-TTC repeat was completely stable for further experiments. These contained up to 230 triplets in orientation II, and up to 270 triplets in orientation I (Fig. 1C).

Length and Orientation-dependent Reduction of Gene Expression by GAA-TTC in Transfected COS-7 Cells—pSPL3 (Fig. 2A) harbors a reporter gene, derived from the HIV gag 120 gene, composed by two exons (exons 1 and 2) separated by an intron derived from the HIV tat gene (33). This intron contains a multiple cloning site where (GAA-TTC)$_n$ repeats were inserted. Transcription is controlled by the SV40 early promoter. (GAA-TTC)$_n$-repeat-containing pSPL3-based plasmids were transfected into COS-7 cells and the reporter gene transcripts were detected by RNase protection analysis (Fig. 2B). We observed a reduction in the amount of the mature transcript containing both exons 1 and 2 (indicated as I in Fig. 2B) as the length of GAA-TTC repeats increased. Such reduction was much greater when (TTC)$_n$ was in the template strand, i.e. when GAA-containing RNA was synthesized (Fig. 2A, lower inset), than when (GAA)$_n$ was in the template strand, i.e. when UUC-containing RNA was formed (Fig. 2A, upper inset). Specifically, the mature transcript derived from (GAA)$_{230}$ containing RNA was 6 times less abundant than the one derived from the (UUC)$_{270}$-containing RNA (3% (Fig. 2B, lane 7) versus 17% (lane 4) of pSPL3 (lane 8)). The protected fragment indicated as II in Fig. 2B, whose size corresponds to exon 2 only, was also reduced as the repeat length increased, suggesting that accumulation of unspliced RNA was not occurring. The protected fragment indicated as III in Fig. 2B was also reduced in amount as the repeat length increased and is of uncertain nature, possibly resulting from an alternate splicing of exon 1 with part of exon 2. In the single figure, IV indicates a protected fragment corresponding to exon 1 only. Interestingly, its abundance remained stable in all samples, suggesting that the initiation of transcription was not affected by the length of the TRS.

Northern blot analysis of total RNA from transfected cells (Fig. 2C) confirmed that GAA-TTC triplet repeats caused a length- and orientation-dependent reduction in the abundance of mature reporter gene RNA (arrow) (lanes 1–7). The most marked decrease was observed for the mature RNAs derived from (GAA)$_{170}$ and (GAA)$_{230}$ containing primary transcripts (lanes 6 and 7). Small amounts of unspliced transcripts (bracket) were detected for all templates, including the control plasmid (pSPL3 without insert, lane 8), possibly because of the high level of transcription generated by the SV40 promoter. The (GAA)$_{170}$ and (GAA)$_{230}$ containing primary transcripts (lanes 6 and 7) were reduced as much as the corresponding mature transcripts, and no partially spliced intermediates were revealed. This lack of accumulation of unspliced or partially spliced RNAs strongly suggests that long GAA-TTC triplet repeats interfere with transcription rather than splicing.

We then investigated the effects of (GAA-TTC)$_n$ repeats on the expression of the protein product of the reporter gene. The lacZ gene open reading frame was inserted into exon 2 of the pSPL3-based plasmids containing (GAA-TTC)$_n$ (Fig. 3). β-Galactosidase activity, resulting from lacZ expression, decreased in transfected COS-7 cells as the length of GAA-TTC repeats increased. However, this decrease significantly exceeded that observed as a general consequence of inserting longer sequences upstream of the lacZ open reading frame only when GAA-containing RNA was synthesized. β-Galactosidase activity in cells transfected with pMP163, whose primary transcript contains (GAA)$_{230}$, was significantly lower than in cells transfected with pMP183, in which a random sequence of similar length was cloned instead of a GAA-TTC repeat, while lacZ expression in cells transfected with pMP153, which synthesize RNA containing (UUC)$_{270}$, was not significantly lower than in pMP183-transfected cells.

Analysis of Replicated Plasmids Containing GAA-TTC Repeats in COS-7 Cells—pSPL3-based plasmids containing between 9 and 270 GAA-TTC triplets in both orientations were transfected into COS-7 cells to study their replication properties. After 48 h, plasmid DNAs were isolated, digested with restriction enzymes to produce fragments containing (GAA-TTC)$_n$, and analyzed by Southern blot (Fig. 4). No evidence of instability was found for GAA-TTC repeats containing up to 270 triplets in orientation A (lane 3) and 230 triplets in orientation B (lane 6). Additional analysis using different restriction enzymes of KpnI and PstI, which produced shorter fragments to give better separation between the shortest and longest repeats on a 1.0% agarose gel, also showed no apparent instability (data not shown). However, the hybridization sig-
The cloning of (GAA)$_n$ and (GAA)$_{TTC}$ repeats in plasmids and analysis of their replication stability in E. coli.

Cloning of (GAA)$_n$ and (GAA)$_{TTC}$ repeats in plasmids and analysis of their replication stability in E. coli. Interestingly, we recovered five times less pRW3827, which generates a (GAA)$_{230}$ containing transcript (Fig. 2A, orientation B), than pRW3823, which generates a (UUC)$_{270}$ containing transcript (Fig. 2A, orientation A). This finding suggests that the replication of a long GAA-TTC repeat may be influenced by transcriptional activity and possibly by interference of RNA molecules containing (GAA)$_n$ (see “Discussion”).

**DISCUSSION**

The level of frataxin mRNA and protein is very low in tissue samples and cultured cells from FRDA individuals carrying intronic GAA-TTC expansions in both homologs of the frataxin gene (1, 7–9). By using a simple in vivo model, we showed that an expanded intronic GAA-TTC TRS can suppress gene expression in a length- and orientation-dependent manner, consistent with the observations in FRDA. Inhibition of transcription seems to be the most likely mechanism. In addition, we obtained new information about factors affecting GAA-TTC repeat stability, along with evidence suggesting that this sequence may interfere with DNA replication.

To perform this study, we utilized cloned GAA-TTC repeats originally amplified from the first intron of the frataxin gene along with some flanking sequence. Previously cloned GAA-TTC TRS had been obtained using synthetic oligonucleotides (13). The longest repeat contained 103 triplets and was interrupted by AA and AGG in the lagging strand. n indicates the total number of GAA-TTC repeats. Inserts containing less than 70 triplets were entirely sequenced. pMP145, pMP141, pMP146, and pMP142 contain perfect GAA-TTC repeat triplets. pMP145 and pRW3832 contain (GAA)$_n$ in the lagging and leading strand templates, respectively. pRW3835 and pRW3821 contain (GAA)$_n$, A/GAA/GAG in the lagging and leading strand templates, respectively. pRW3830 and pRW3821 contain (GAA)$_n$, A/GAA/GAG in the lagging and leading strand templates, respectively. pRW3830 and pRW3821 contain (GAA)$_n$, A/GAA/GAG in the lagging and leading strand templates. B, differential stability of (GAA-TTC)$_n$ in pUC versus pSPL3 vectors grown in E. coli SURE. pUC19NotI (U) and pSPL3 (S) containing the number of the repeats indicated on the top were digested with EcoRI and PstI and run on a 1.2% agarose gel. M1 and M2 indicate λ DNA-BstEI digest and 100-bp ladder size markers, respectively, with sizes in base pairs (bp) shown on the left. The positions of the EcoRI-PstI fragments containing (GAA-TTC)$_n$ are shown by arrows on the right. C, stability of pSPL3-based plasmids containing (GAA-TTC)$_n$ grown in E. coli SURE. pSPL3-based plasmids containing (GAA-TTC)$_n$ in orientations I and II, as defined in A, were digested with EcoRI and NdeI and run on a 1.2% agarose gel. M1 and M2 indicate λ DNA-BstEI digest and 100-bp ladder size markers, respectively, with sizes in base pairs (bp) shown on the left.

**Fig. 1.** Cloning of (GAA-TTC)$_n$ in plasmids and analysis of their stability in E. coli. A, cloned (GAA-TTC)$_n$ in plasmids. Insets include flanking sequence from the first intron of the frataxin gene, extending 352 bp 5’ and 250 bp 3’ of the GAA-TTC repeat. Plasmids with a
and recent studies also suggested a role of flanking sequences (22–28). Therefore, we chose to work with repeats as close as possible to the naturally occurring sequence in the FRDA gene.

Under certain conditions, TRS are unstable during replication in recombinant plasmids in E. coli (29, 31, 34–37) and Saccharomyces cerevisiae (38–40). Contractions are much more common than expansions. The initial cloning of GAA-TTC repeats from polymerase chain reaction products also indicated that this TRS shows intrinsic instability in E. coli, which was
were cloned in the intron of pMP129, a pSPL3 derivative containing the lacZ gene within exon 2. Plasmids whose transcripts contain (GAA), indicated with closed bars, those transcribing (UUC), with hatched bar. β-Galactosidase activity values were normalized to luciferase activity from co-transfected pGL3-Control, and are relative to pMP129 as 100%. pMP183 contains a luciferase gene fragment (UUC)270 (pMP153) and (GAA)230 (pMP163). All experiments were done at least in triplicate. * indicates the SV40 early promoter. E1 and E2 designate exon 1 and 2, respectively. MCS designates the multicloning site. Results of t tests between pairs of samples are indicated as follows; NS, not significant; *, p < 0.05; **, p < 0.01; ***, p < 0.001.

Several factors are known to influence the stability of TRS, including length, presence of interruptions, characteristics of the vector, orientation relative to the unidirectional replication origin, and genetic backgrounds of host cells (29, 31, 35–37). The instability of the GAA repeats was clearly dependent on the direction of replication, as previously observed for CTG-CAG (29) and CCG-CGG (31). Single-stranded CTG, CAG, CCG, and CGG repeats are known to form hairpin structures of variable stability (4, 41, 42). The molecular basis of expansions versus deletions of CTG-CAG (29) and CCG-CGG (31) was explained on the basis of preferential stabilization of transient loop structures during replication (29). The current study shows that the instability of GAA-TTC repeats was greater when GAA was the lagging strand template than when it was the leading strand template. According to the previously proposed model for CTG-CAG (29), this differential instability may be due to the ability of single-stranded GAA to adopt a more stable DNA secondary structure during replication than single-stranded TTC. The nature of such structure for the FRDA sequence remains undetermined and may differ from a hairpin, since the structures of both single-stranded GAA and TTC can adopt were reported to be much less stable than those formed by single-stranded CTG, CAG, CCG, and CGG (41).

The stability of long GAA-TTC sequences in E. coli was also strongly influenced by the cloning vector. Long GAA-TTC repeats were much more stable when cloned into pSPL3 rather than into pUC vectors. Transcriptional activity may be involved, because CTG-CAG repeats have been shown to be more unstable when transcribed, both in E. coli (34) and in transgenic mice (22). In pSPL3-based plasmids, the GAA-TTC repeats are located in the intronic region and were transcribed under the control of the SV40 early promoter. No transcription occurs when these plasmids are propagated in E. coli. Conversely, the GAA-TTC repeats are localized within a transcription unit in pUC vectors, possibly accounting for the observed greater instability.

We analyzed the effect of intronic GAA-TTC repeats on gene expression by transfecting COS-7 cells with constructs harboring GAA-TTC repeats of different lengths and orientations in an intron of a reporter gene. When (GAA), was in the transcripts, as is the case in the frataxin gene, transcription and expression of the reporter gene were reduced proportionally to the repeat length. Repeats containing more than 33 triplets, close to the upper limit for normal alleles of the frataxin TRS (1, 5, 6), started to inhibit gene expression. No increase in unspliced or partially spliced transcript was observed, suggesting that a defect in RNA splicing caused by the expanded GAA-TTC repeat, proposed as a cause of reduced frataxin gene expression in FRDA (10), is unlikely. Along with the observation that transcription initiation is probably not affected, as suggested by RNase protection experiments, the occurrence of a transcriptional block at the repeat seems to be the most likely explanation for reduced gene expression. According to our observations, such a block is orientation-dependent, occurring only with transcription of GAA-containing RNA. Such purine-specific inhibition is in agreement with previous in vitro studies of Pur-Pyr sequences (14–17), which indicated that under physiological conditions Pur-Pur-Pyr triplex structures are preferentially formed and in vitro transcription of purine-rich RNA is specifically reduced. Griffin et al. (14–16) suggested that the underlying molecular mechanism is the formation of an intermolecular RNA-DNA hybrid triplex structure (Fig. 5B). Grabczcyk and Fishman (17) proposed instead that purine-rich RNA may bind to the single pyrimidine-rich DNA strand generated by the formation of an intramolecular DNA triplex, resulting in its stabilization. According to this model, a wave of negative supercoiling following transcription (Fig. 5C) would trigger intramolecular DNA triplex formation. We propose that the GAA-rich transcript may bind the duplex DNA template, as in the previously proposed models (Fig. 5, B and C), interfering with RNA elongation and preventing further transcription.

Because TRS may be unstable when propagated in eukaryotic cells, including yeast (38–40) and transgenic mice (22,
which was sequentially hybridized with different probes. First, the digests were applied on a 1.2% agarose gel, run at 70 V for 4.5 h at 25 °C.

Relative amounts of replicated pSPL3-based plasmids, normalized to the co-transfected pGL3-Control and relative to the value of pMP145 as 100%.

Variation in transcriptional activity was not correlated with the amount of DNA template. This might be due to saturation of transcription machinery in COS-7 cells even with the lower level of DNA template, as previously suggested (51).

Our analysis of the effect of intronic GAA·TTC repeats on transcription and replication provides an initial understanding of the molecular mechanisms underlying the loss of function in FRDA. Some of these mechanisms may be common to other long tracts of intronic Pur·Pyr sequences, including GAAGAGA·CTC·CTC repeats in the human tumor necrosis factor receptor p75 gene (52), and GAA·TTC and GAG·CTC triplets in the cardiac α-myosin heavy chain (MYH6) gene (53), whose biological roles are currently unknown.

Acknowledgments—We thank Drs. Albino Bacolla, Pawel Parniewski, and Margaret Labuda, and Sarn Jiralerspong for helpful suggestions. We also thank Dr. Jacques Drouin and Gino Poulin for use and assistance of the luminometer, and Dr. Pierre Chartrand for use of the electroporator.
