GGA-TCC-interrupted Triplets in Long GAA-TTC Repeats Inhibit the Formation of Triplex and Sticky DNA Structures, Alleviate Transcription Inhibition, and Reduce Genetic Instabilities*

Naoki Sakamoto**, Jacquelynn E. Larson, Ravi R. Iyer‡, Laura Montermini, Massimo Pandolfo§, and Robert D. Wells¶

From the Institute of Biosciences and Technology, Center for Genome Research, Texas A & M University System Health Science Center, Texas Medical Center, Houston, Texas 77030, and §Centre Hospitalier de l’Université de Montréal, Campus Notre-Dame, Montreal, Quebec H2L 4M1, Canada

Received for publication, February 28, 2001, and in revised form, April 23, 2001
Published, JBC Papers in Press, April 26, 2001, DOI 10.1074/jbc.M101852200

Large expansions of GAATTC repeats in the first intron of the frataxin (X25) gene are the principal mutation responsible for Friedreich’s ataxia (FRDA). Sticky DNA, based on R-RY triplexes, was found at the expanded GAATTC repeats from FRDA patients. The (GAA-GGA-TCC)₆₀₇₀ repeat occurs in the same frataxin locus but is nonpathogenic and does not form sticky DNA. To elucidate the behavior of sticky DNA, we introduced various extents of GGA-TCC interruptions into the long GAATTC repeat. More than 20% of GGA-TCC interruptions abolished the formation of sticky DNA. However, the GAATTC repeats with less than 11% of GGA-TCC interruptions formed triplexes and/or sticky DNA similar to the uninterrupted repeat sequence. These triplexes showed different P1 nuclease sensitivities, and the GGA-TCC interruptions were slightly more sensitive than the surrounding GAATTC repeats. Furthermore, genetic instability investigations in Escherichia coli revealed that a small number (4%) of interruptions substantially stabilized the long GAATTC tracts. Furthermore, the greater the extent of interruptions of the GAATTC repeats, the less inhibition of in vitro transcription was observed, as expected, based on the capacity of interruptions to inhibit the formation of sticky DNA. We propose that the interruptions introduce base mismatches into the R-RY triplex, which explains the observed chemical and biological properties.

The clinical features as well as the molecular pathology of Friedreich’s ataxia (FRDA)³ are summarized in the accompanying paper (1). The molecular mechanism of the biological effects of the long GAATTC repeat was proposed to be the formation of an unusual DNA structure (2–4). The GAATTC repeats have been known to form a triplex structure (5–11). Recently, the sticky DNA structure was found specifically in the long GAATTC repeats from FRDA patients (12). This structure was based on the R-RY type of triplex and was hypothesized to be formed by exchanging the pyrimidine strands between two R-RY triplexes (12). The GAATTC repeat lengths required for the formation of sticky DNA correlated well with that required for the disease phenotype (12). Furthermore, we found that sticky DNA inhibited T7 and SP6 RNA polymerase transcription effectively by sequestering RNA polymerases (1). Therefore, sticky DNA may be involved in the pathology of FRDA.

On the other hand, the GAAGGA-TCCCTTC repeat (65 units in length) was also found in the first intron of the frataxin gene but was demonstrated to be nonpathogenic (13). Unlike the GAATTC repeat, this hexamer repeat does not form a triplex and/or sticky DNA, does not inhibit transcription, and does not associate with the FRDA disease state (12, 13). This strongly suggests that triplexes and/or sticky DNA may be involved in the pathology of FRDA. However, the reason why the hexamer repeat did not form a triplex or sticky DNA was unclear.

In normal individuals with moderate lengths of GAATTC repeats, short GAGAGGA-TCCCTTC repeat interruptions were found (14, 15). This repeat appeared to also be nonpathogenic, and thus it became evident that the length of the uninterrupted GAATTC repeat was important for the pathology of FRDA. Thus, interruptions in the long GAATTC repeats may influence the ability of the repeats to form sticky DNA and to inhibit transcription.

These kinds of interruptions were shown to affect the genetic instabilities of other triplet repeat sequences (16–18). Furthermore, the GAGAGGA-TCCCTTC repeat was more stable genetically than the uninterrupted GAATTC repeat (13).

Herein, we describe the preparation and characterization of a family of direct repeat sequences of composition intermediate between uninterrupted GAATTC and GAAGGA-TCCCTTC repeats. Frequent GAATTC interruptions (more than 20%) in the long GAATTC repeat interfere with the formation of sticky DNA and triplexes, with the inhibitory effect on transcription of long GAATTC repeats, and with genetic instabilities. Mapping of P1 nuclease-cleaved sites revealed that the GAATTC repeats with less than 20% of GAATTC interruptions adopt R-RY triplexes similar to the pure GAATTC repeats but that the GAATTC interruptions may cause base mismatches in the triplexes. Hence, a high percentage of GAATTC interruptions in the GAATTC repeats inhibits the formation of sticky DNA
by introducing base-mismatches into the triplexes, which makes them unstable.

EXPERIMENTAL PROCEDURES

Construction of Plasmids—The GGA-TCC interruptions were introduced into the P(RWTTT) repeat by site-directed mutagenesis (19), prRW3546, which contains the (GAA-TCC)₁₀₀ repeat in the pGEM-3Z f vector (Promega), was constructed as described by Iyer and Wells (20). The single-stranded circular prRW3546 containing the (TTTTC₉₀) sequence was prepared using the prRW3546-transfected NMS22 Escherichia coli strain and M13K07 helper phage (21). To introduce GGA-TCC interruptions into the (GAA₁₀₀) repeat by site-directed mutagenesis, both oligonucleotides were synthesized by Genesys: GAA1/24, 5′- (GAA)₉-GAA(GAA)₉; GAA1/5, 5′- (GAA)₉-GAA(GAA)₉; GAA1/12, 5′- (GAA)₉-GAA(GAA)₉; GAA1/9, 5′- (GAA)₉-GAA(GAA)₉; GAA2/9, 5′- (GAA)₉-GAA(GAA)₉; GAA3/9, 5′- (GAA)₉-GAA(GAA)₉; GAA-3′. To eliminate the nonmutated original plasmid, elimination primer, 5′-GTGCCACCTGCTGACTAA-GAACC-3′, in which the unique AaeII recognition site is mutated, was prepared. All of these primers were phosphorylated in 70 mM Tris-HCl (pH 7.6), 10 mM MgCl₂, 50 mM NaCl, in a 40-μl mixture at 25 °C for 10 min. Then PCR products generated by the CaCl₂ method (21). The entire PCR mixture was grown in 1 liter of LB broth containing 100 μg/ml XbaI-linearized pCR3.1 vector were mixed and subjected to in vitro transcription using T7 RNA polymerase.

Bacterial in vitro transcription forming the purified strand (uninterrupted or interrupted rGAA repeat) was performed in a 20-μl volume at 37 °C for 30 min in RNA polymerase buffer (40 mM Tris-HCl (pH 7.9), 6 mM MgCl₂, 2 mM spermidine, 10 mM DTT) supplemented with 500 μM each of rATP, rGTP, and rUTP, 25 μM rCTP, 0.17 μM [α-³²P]dCTP (3000 Ci/mmol, 10 μCi/ml) (Amersham Pharmacia Biotech), 40 units of Rnas-in (Promega), and 10 units of T7 RNA polymerase (New England Biolabs). For the formation of the pyrimidine strand (uninterrupted or interrupted rUUC repeat), 500 μM each of rATP, rGTP, and rUTP, 25 μM rCTP, 0.17 μM [α-³²P]dCTP (3000 Ci/mmol, 10 μCi/ml) (Amersham Pharmacia Biotech) were used. The RNA synthesized was purified by phenol/chloroform extraction, precipitated by ethanol, dissolved in formamide loading buffer (90% formamide, 20 mM EDTA, 0.05% bromphenol blue, and 0.05% xylene cyanol), and analyzed in a 4% polyacrylamide gel containing 7 M urea. The gels were exposed to x-ray film at room temperature after drying.

Recruitment Assay for the Analyses of Genetic Instabilities of Repeat Tracts—The extended repeat lengths of uninterrupted and interrupted rGAA repeat tracts were prepared as described previously (1) and transformed into E. coli AB1157 (thr-1, araC14, leuB6, Δgst-a-proA62, lacY1, tss-33, qsr-0, glvA44/AS), galK2/Oc, λ, SacO-o4(hG4/Oc), rfbD1, mgl-51, rpsS2366(Am), rpsL31(strR), hdgK51, xylA5, ntl-1, argE3(Oc), thi-1) by electroporation. The transformation mixture was plated on LB plates containing 100 μg/ml of kanamycin and incubated at 37 °C. The transformed colonies were picked and grown in 1 ml at 65 °C for 10 min in the presence of one of the following primers: K05137, 5′-CAATAGTGGTGA-3′; K07553, 5′-GTTAGCTGGATTACAGGC-3′. After neutralization by the addition of 13 μl of 3 M sodium acetate (pH 5.2), DNA was precipitated by ethanol. The DNA was dissolved in 10 μl of Sequenase buffer (40 mM Tris-HCl (pH 7.5), 20 mM MgCl₂, 50 mM NaCl) and incubated at 42 °C for 10 min to anneal the primer to the denatured template. The primer extension reaction was started by the addition of 0.1 μl of each rGAA repeat, 0.05 μl of dUTP, 0.05 μl of dCTP, and dTTP, 1 μl of [α-³²P]dATP (3000 Ci/mmol, 10 μCi/ml) (Amersham Pharmacia Biotech) and 2 μl of Sequenase version 2.0 (1.625 units/μl) (Amersham Pharmacia Biotech) and incubated at room temperature for 5 min. Then 14 μl of 80 μl dNTP mix were added and incubated at 37 °C for 5 min to complete the reaction. After purification of DNA, the extended products were analyzed in a 4% denaturing polyacrylamide gel in TBE buffer. As the marker, the sequencing reaction was performed according to the Sequenase version 2.0 (Amersham Pharmacia Biotech) sequencing protocol in parallel.

In Vitro Transcription—The pCR3.1-based plasmids were linearized by XbaI to enable the run-off of transcription. The pCR3.1 vector was also linearized by XbaI as the internal control. 100 ng of the XbaI-linearized pCR3.1-based plasmid and 100 ng of the XbaI-linearized pCR3.1 vector were mixed and subjected to in vitro transcription using T7 RNA polymerase.

Effect of GGA-TCC Interruptions on Sticky DNA

by guest on July 18, 2018http://www.jbc.org/Downloaded from
DNA Sequence and Restriction Enzyme Analyses of FRDA Patient Expanded Repeats—Expanded GAA\_\textsuperscript{z}TTC repeats from 22 Friedreich ataxia patients, including eight Acadians, were amplified from genomic DNA as described (13). The corresponding DNA fragments were separated on 1% agarose gels and purified on QIAGEN columns. To detect interruptions in GAA\_\textsuperscript{z}TTC, the purified fragments were digested with the restriction enzymes Ear\_I and Mnl\_I, whose recognition sequences are GAAGAG and GAGG, respectively. In addition, the pyrimidine strand of 11 repeats was directly sequenced on a Licor automatic sequencer using a fluorescently labeled 2500R primer (28) and the SequiTherm EXCEL\_II sequencing kit (Epicentre Technologies). Readable sequence was obtained for up to 160 triplets.

RESULTS
Introduction of GGA\_\textsuperscript{z}TCC Interruptions into (GAA\_\textsuperscript{z}TTC)\textsubscript{150} Repeats—To create the intermediate sequences between the uninterrupted GAA-TTC repeat and the GAAGGA\_\textsuperscript{z}TCCTTC hexamer repeat, various extents of GGA-TCC interruptions were introduced into the (GAA-TTC)\textsubscript{150} repeat by a method based on site-directed mutagenesis (19). Fig. 1 shows an overview of the mutagenesis strategy. The (TTC)\textsubscript{150}-containing single-stranded circular plasmid was produced from pRW3546, which is the pGEM-3Zf(\textsuperscript{2})-based plasmid containing the (GAA\_\textsuperscript{z}TTC)\textsubscript{150} repeat, by the single-stranded phage system. The mutated repeating GAA oligonucleotides and Aat\_II elimination primers were annealed to the single-stranded (TTC)\textsubscript{150} plasmids, and the primer-directed DNA syntheses and the ligations of the newly synthesized DNA fragments were carried out by the addition of T4 DNA polymerase and T4 DNA ligase.
After transformation of the E. coli mutS strain, the plasmids were isolated and cleaved by AatII to eliminate the plasmid derived from the original template strand and then used for the transformation of E. coli SURE strain. Only the mutated and uncleaved plasmids should grow. Using this methodology, we obtained a series of plasmids containing GGA\textsubscript{z}TCC-interrupted GAA\textsubscript{z}TTC repeats; the sequences of the inserts are listed in Table I. Each mutated sequence, (GAA)\textsubscript{23}GGA, (GAA)\textsubscript{14}GGA, (GAA)\textsubscript{11}GGA, (GAA)\textsubscript{8}GGA, and (GAA)\textsubscript{3}GGA(GAA)\textsubscript{4}GGA, contains 5 (4%), 9 (7%), 10 (8%), 14 (11%), and 26 (20%) GGA\textsubscript{z}TCC interruptions, respectively, within the 130–136 repeats of the GAA\textsubscript{z}TTC sequence (Table I). All mutated sequences had a length of about 130 triplet repeats, probably because of small deletions during the process of the preparation of the single-stranded (TTC)\textsubscript{150} plasmid. However, this length is equivalent to the 65 repeats of the GAAGGA\textsubscript{z}TCCTTC hexamer sequence and thus is appropriate for the comparison between these sequences. Interestingly, the GGA\textsubscript{z}TCC interruptions were found to be aligned in a systematic order, indicating that many of the mutated oligonucleotides were closely aligned without gaps during the annealing to the single-stranded plasmid.

Sticky DNA Formation by Interrupted GAA\textsubscript{z}TTC Repeats—
We have recently reported that the uninterrupted long GAA\textsubscript{z}TTC repeat forms the sticky DNA structure, which is hypothesized to be an intermolecular bitriplex, but the GAAGGA\textsubscript{z}TCCTTC hexamer repeat of equivalent length is incapable of adopting this conformation (12). To analyze the structural properties of the intermediate sequences, the interrupted GAA\textsubscript{z}TTC repeats were recloned into the pSPL3 vector. The designations of these plasmids are listed in Table II.

When these supercoiled circular plasmids were analyzed in an agarose gel, all plasmids showed dimers and higher multimers (migrating at the top half of the gel) since they were isolated from the E. coli SURE strain, which is RecA\textsuperscript{−} strain (Fig. 2A). These plasmids contained homogeneous inserts of the expected lengths (Fig. 2B). Since all of the sequences have \textasciitilde254 and \textasciitilde354 bp of human flanking sequences on the sides, the total length of inserts was \textasciitilde1.1 kb (Fig. 2B).

Then the plasmids shown in Fig. 2A were linearized by XmnI and analyzed in an agarose gel to test for the formation of the retarded band, which represents the sticky DNA (Fig. 2C). The retarded bands were observed at the same positions on the gel for the DNAs containing 0–11% of GGA\textsubscript{z}TCC interruptions. The amount of the retarded band was about 17% of the total DNA, although a slight reduction was observed for the plasmid with 11% GGA\textsubscript{z}TCC interruptions (Fig. 2, C and D). However, when the interruptions were increased up to 20%, the retarded band disappeared completely, and the hexamer repeat (50% GGA\textsubscript{z}TCC interruptions) did not show the retarded band, as expected (Fig. 2, C and D) (12). These data show that sticky DNA, which is responsible for the retarded band, can be formed in the presence of GGA\textsubscript{z}TTC interruptions up to 11% but that 20% or more GGA\textsubscript{z}TTC interruptions abolish the formation of sticky DNA.

Thermostability of Sticky DNA Formed by Interrupted GAA\textsubscript{z}TTC Repeats—To evaluate the thermostability of the sticky DNA structures formed by the interrupted GAA\textsubscript{z}TTC repeats, prW3822 containing uninterrupted (GAA\textsubscript{z}TTC)\textsubscript{150} and prW2114 containing 8% of GGA\textsubscript{z}TTC interruptions were linearized by XmnI, and the thermostabilities of the sticky DNA

---

**Table I**

GGA\textsubscript{z}TCC-interrupted GAA\textsubscript{z}TTC sequences created by site directed mutagenesis in this study.

| Abbreviated sequence of repeating unit | Schematic representation of the location of the GGA interruptions within the GAA repeat tracts | GGA/Total triplet repeats (fraction %) |
|---------------------------------------|-----------------------------------------------------------------|-------------------------------------|
| (GAA)\textsubscript{150}              | ![Schematic Diagram](image)                                       | 0/150 (0)                           |
| (GAA)\textsubscript{23}GGA            | ![Schematic Diagram](image)                                       | 5/131 (4)                           |
| (GAA)\textsubscript{14}GGA            | ![Schematic Diagram](image)                                       | 9/130 (7)                           |
| (GAA)\textsubscript{11}GGA            | ![Schematic Diagram](image)                                       | 10/130 (8)                          |
| (GAA)\textsubscript{8}GGA             | ![Schematic Diagram](image)                                       | 15/136 (11)                         |
| (GAA)\textsubscript{3}GGA(GAA)\textsubscript{4}GGA | ![Schematic Diagram](image)                                       | 26/131 (20)                         |
| GAAGGA                                | ![Schematic Diagram](image)                                       | 65/130 (50)                         |
structures were tested in the presence or absence of EDTA. The sticky DNA formed by pRW3822 was extremely thermostable in the absence of EDTA (Fig. 3, open circles), whereas the retarded band of pRW3822 completely disappeared at 60 °C in the presence of EDTA as described before (Fig. 3, filled circles) (12). Similarly, the sticky DNA structure adopted by pRW2114

**TABLE II**

List of plasmids used in this study

| Abbreviated sequence | pSPL3-based | pCR3.1-based (formation of rGAA RNA) | pCR3.1-based (formation of rUUC RNA) |
|----------------------|-------------|--------------------------------------|--------------------------------------|
| GAA                  | pRW3822     | pRW4245                              | pRW4250                              |
| (GAA)$_3$GGA         | pRW2116     | pRW2145                              | pRW2150                              |
| (GAA)$_5$GGA         | pRW2115     | pRW2120                              | pRW2148                              |
| (GAA)$_7$GGA         | pRW2114     | pRW2119                              | pRW2148                              |
| (GAA)$_9$GGA         | pRW4222     | pRW2118                              | pRW2147                              |
| (GAA)$_{11}$GGA(GAA)$_3$GGA | pRW2113     | pRW2117                              | pRW2146                              |
| GAAGGA               | pMP193      | pRW4246                              | pRW4251                              |

**FIG. 2.** Effect of GGA-TCC interruptions on the formation of sticky DNA. A, supercoiled plasmids isolated from E. coli SURE cells. Plasmids harboring various percentages of GGA-TCC interruptions (Table I) were analyzed in a 0.7% agarose gel. B, EcoRI–PstI-digested plasmids. The plasmids listed in Table I were digested with EcoRI and PstI, and the length of the insert was analyzed in a 1.5% agarose gel. C, XmnI-linearized plasmids. The plasmids listed in Table I were linearized with XmnI, and the formation of the retarded band was analyzed in a 0.7% agarose gel. The percentages of GGA-TCC interruptions are shown below the gel. M, 1-kb DNA ladder (Life Technologies). The sizes (in kb) of the 1-kb DNA ladder are shown to the left of the gels. The positions of the linear band and the retarded band are shown on the right. Black triangle, position of the retarded band; white triangle, position of the linear band. D, quantitation of the effect of GGA-TCC interruptions on the formation of sticky DNA. The amount of the retarded band in C was quantitated by densitometric analysis. The y axis shows the percentage of the retarded band, and the total amount of DNA in each lane is defined as 100%. The x axis shows the percentage of GGA-TCC interruptions.

**FIG. 3.** Thermostabilities of sticky DNA formed by interrupted GAA-TTC repeats. XmnI-linearized pRW3822 and pRW2114 were incubated at different temperatures for 10 min in TE buffer (10 mM Tris-HCl (pH 8.0) and 1 mM EDTA). The amount of the retarded bands was analyzed by 0.7% agarose gel electrophoresis.

**FIG. 4.** P1 nuclease sensitivities on each strand of interrupted GAA-TTC repeats. A, map of plasmid used in this analysis. The positions of restriction enzyme recognition sites used in this assay are shown. Shaded box, GAA-TTC repeat or interrupted GAA-TTC repeats; open boxes, human flanking sequences. The orientation of each strand is indicated by an arrow. B, analysis of P1 nuclease sensitivities by “weak” reaction. Supercoiled plasmids shown in Fig. 2A were cleaved by 1 unit of P1 nuclease at 37 °C for 1 min, and P1 nuclease sensitivities on each strand were analyzed as described (22). The percentages of GGA-TCC interruptions are shown at the top of the gel. M, 1-kb DNA ladder. The positions of the pyrimidine and purine tracts are shown on the left and right, respectively. Vertical lines, flanking sequences; asterisks, positions of $^32$P end labeling.
first, we performed a “weak” digestion by P1 nuclease (1 unit for 1 min) so that a single cleavage occurred on each plasmid molecule. The cleaved sites were mapped on each strand as described previously (12). For analysis of the pyrimidine strand, we observed smeared signals within the repeat for plasmids containing 0% (uninterrupted) to 11% GGA-TTC interruptions (Fig. 4B). On the other hand, the purine tract did not show such intense signals (Fig. 4B). These results suggest that the GGA-TTC interruptions up to 11% in the long GAA-TTC repeat do not interfere with the ability of the long GAA-TTC repeat to form triplex structures and that all of these interrupted sequences form similar R-R-Y type of triplexes. This is consistent with the ability of these plasmids to form the sticky DNA structure (Fig. 2).

The GAAGGA-TCCCTTC hexamer repeat (50% GGA interruptions) showed an intense signal at the 5′-end of purine strand where a poly(A-T) tract exists (Fig. 4B). This indicates that the poly(A-T) sequence adjacent to the GAAGGA-TCCCTTC repeat on its 5′ side is unpaired when 50% of GAA-TTC triplets are substituted alternately by GGA-TTC triplets.

Next, we carried out a “strong” P1 nuclease digestion (10 units for 10 min) of these mutated plasmids. After P1 nuclease digestion, plasmids were cleaved by NheI and analyzed by an agarose gel electrophoresis. If a double-stranded cleavage occurred within the repeats, 1.3–1.75-kb fragments should be observed (Fig. 5A). As seen in the “weak” reaction (Fig. 4), P1 nuclease cleavages were observed within the repeat for plasmids with interrupted GAA-TTC repeats (Fig. 4A). In the analysis of the pyrimidine strand, we observed smeared signals within the repeat for plasmids containing 0% (uninterrupted) to 11% GGA-TTC interruptions (Fig. 4B). On the other hand, the purine tract did not show such intense signals (Fig. 4B). These results suggest that the GGA-TTC interruptions up to 11% in the long GAA-TTC repeat do not interfere with the ability of the long GAA-TTC repeat to form triplex structures and that all of these interrupted sequences form similar R-R-Y type of triplexes. This is consistent with the ability of these plasmids to form the sticky DNA structure (Fig. 2).
mids containing from 0% (uninterrupted) up to 11% GGA-TCC interruptions (Fig. 5B).

Interestingly, when these cleaved plasmids were analyzed in a long agarose gel, each cleavage pattern revealed different distinct bands (Fig. 5C). The uninterrupted GAA-TTC repeat showed a smeared pattern, indicating that P1 nuclease cleaved the repeat randomly. However, the GAA-TTC repeats interrupted by 4, 7, 8, and 11% of GGA-TCC triplets showed about 4, 7, 8, and 10 separate bands, respectively, although there were smeared backgrounds in each lane. This suggests that there are favored P1 nuclease cleavage sites in the interrupted GAA-TTC repeats. The distance between the adjacent bands decreased as the extent of the GGA-TCC interruptions increased. For the DNAs with 4% GGA-TCC interruptions, the distance between each band was ~80 bp, similar to the interval between each of the GGA-TCC interruptions, 69 bp. For the DNAs with 7 and 8% GGA-TCC interruptions, the distance between each band was estimated to be ~51 and ~41 bp, respectively. These values are also close to the expected distance between the GGA-TCC interruptions in each sequence, 42 and 33 bp. However, it is not clear if the GGA-TCC interruptions or the short GAA-TTC repeats were cleaved by P1 nuclease.

Finally, we attempted to precisely map the P1 nuclease cleaved sites using a primer extension method. After digestion of pRW2116 by 10 units of P1 nuclease, the cleaved sites were analyzed on each strand by the primer extension method. When the interrupted GAA strand was analyzed, we observed a smeared pattern, indicating the absence of specific cleaved sites on the GAA strand (data not shown). However, the analysis on the interrupted TTC strand showed slightly intense signals at the positions of the TTC interruptions, although there was a smeared background signal (Fig. 6). These results suggest that the GGA-TTC interrupted sites are less fully base-paired than the more perfect neighboring GAA-TTC regions.

**GGA-TCC Interruptions Genetically Stabilize Long GAA-TTC Repeat Tracts—**Prior studies (3, 4) showed that the presence of CTA or AGG triplets interspersed within CTG or CCG repeats drastically reduced the extent of expansions and deletions of these repeats. Therefore, we hypothesized that the introduction of GGA-TCC triplets into long homogenous GAA-TTC repeat tracts would prevent the rampant expansions and deletions observed for these sequences. To test this hypothesis, we analyzed the genetic instabilities of the pure and interrupted GAA-TTC repeat tracts in pRW2450, pRW2150, pRW2148, and pRW4251 (Table I) in E. coli AB1157 (see “Experimental Procedures”). The plasmids were incorporated into the cells and propagated in log phase growth by successive cultivations. The cultures were harvested after each recultivation, and the plasmid DNA was isolated and analyzed by restriction digestion and polyacrylamide gel electrophoresis. The genetic instabilities were measured by determining the amount of full-length repeat tract as a relative percentage of the total amount of repeat tract in the sample. Fig. 7A shows the 7% polyacrylamide gel analysis of the excised StuI–XmnI inserts from the recultivation assays. The quantitative analyses of these gels is shown in Fig. 7B. The data clearly show that the uninterrupted (GAA-TTC)150 repeat is the most unstable of all the repeats investigated. After just one recultivation, as little as 8% of the initial full-length triplet repeat remained (Fig. 7B). However, the presence of just five GGA-TCC interruptions (4%) within 131 GAA-TTCs served to dramatically increase the stability of these sequences as observed for pRW2150. As much as 85% of the full-length repeat remained after three recultivations (Fig. 7B). When the number of GGA-TCC interruptions was increased to 10 (8%), the genetic instability of the repeat tract was almost negligible (82% of full-length repeat tract remaining after three recultivations) (Fig. 7B). Unexpectedly, the (GAA-TTC)150 hexamer repeat tract in pRW4251 was substantially more unstable than the GAA2,3-GGA and the GAA1,2-GGA repeat tracts. After three recultivations, 63% of the full-length repeat remained. However, this repeat was still dramatically more stable than the pure (GAA-TTC)150 tract. This result was also verified in other E. coli strains. We postulate that the homogeneity of the hexamer repeat tract results in the unexpected levels of expansion and deletion (see “Discussion”).

Thus, our data show that the presence of even a small number of interruptions significantly stabilizes the (GAA-TTC)α repeats. However, this effect is reversed when the number of GGA-TCC interruptions is so high as to abolish the sequence heterogeneity; the relative homogeneity of the hexamer repeat probably allows the formation of slipped structures, which in turn mediate expansions and deletions of these repeats.

**GGA-TCC Interruptions Alleviate Transcription Inhibition by Long GAA-TTC Repeat Tracts—**Long GAA-TTC repeat tracts inhibit transcription from an upstream promoter (3, 23). We showed that this behavior is due to the formation of sticky DNA by these repeats; when the sticky DNA structure was disrupted by heat or EDTA, there was no appreciable inhibition of transcription (1). Since the formation of sticky DNA is dramatically reduced when the number of GGA-TCC interruptions is increased, we postulated that these interruptions would also alleviate the inhibition of transcription. To test this hypothesis, we performed in vitro transcription studies with the T7 RNA polymerase using the pCR3.1-derived plasmids, which contained pure and interrupted GAA-TTC repeats (Table II) as described under “Experimental Procedures.” The plasmids were linearized with XbaI and incubated with T7 RNA polymerase and rNTPs. An equimolar amount of pCR3.1 (which did not contain a triplet repeat insert) was included in each reaction as an internal control. The transcripts generated in the reaction were analyzed on 4% denaturing acrylamide gels. The amount of transcript produced from the insert-containing plasmid was compared with that from the control plasmid. The measured ratios revealed that the relative amount of full-length transcript was the least for the (GAA-TTC)150-containing plasmid. However, increasing amounts of GGA-TCC interruptions resulted in the generation of significantly higher amounts of full-length transcript (data not shown). This effect was observed for both orientations of the triplet repeat inserts, i.e. when the transcript was purine-rich (rGAA) or pyrimidine-rich (rUUC).

**Interruptions in FRDA Chromosomes—**The potential role of interruptions in expanded GAA-TTC repeat sequences in the human FRDA chromosomes that may affect phenotypic variability is uncertain. To evaluate this question, we performed DNA sequence analyses and restriction mapping experiments in order to determine the frequency of interruptions in patient materials (Table III). The pyrimidine strands of 11 expanded alleles were sequenced for as long as technically possible, usually up to 200 triplets. Interruptions were found and were clustered at the 3′-ends of the expanded repeats, affecting the last 10–15 triplets. These interruptions usually consisted of stretches of three or four A/T pairs and occasional GGA-TCC triplets. The remainder of the repeats contained only GAA-TTC triplets within the sequencing limitations; these data included eight alleles, which were sequenced in their entirety. These types of interruptions were not significantly more common in chromosomes from Acadian patients, which are known to have milder phenotypes compared with non-Acadian patients carrying repeats of comparable lengths.

In order to further probe the presence of interruptions, re-

---

\(^2\) R. R. Iyer, A. Pluciennik, M. L. Hebert, A. Vetcher, J. E. Larson, and R. D. Wells, manuscript in preparation.
FIG. 7. Effect of GGA\textsuperscript{z}TCC interruptions on the genetic instabilities of GAA\textsuperscript{z}TTC repeats. A, plasmids pRW4250, pRW2150, pRW2148, and pRW4251 were propagated in *E. coli* AB1157 in log phase by stepwise recultivation as described under “Experimental Procedures.” The triplet repeat containing *Stu*I–*Xma*I inserts were labeled and electrophoresed through 6% native polyacrylamide gels. The gels were exposed to x-ray film as well as a PhosphorImager screen for quantitation. Lanes 1–3 contain DNAs isolated after 1–3 recultivations, respectively. The arrow indicates the band that contains the full-length triplet repeat. The boxed regions designate the expansion and deletion products. B, the genetic instabilities of plasmids pRW4250 (○), pRW2150 (▼), pRW2148 (■), and pRW4251 (▲) were quantitatively determined by scanning the exposed PhosphorImager screens with a Molecular Dynamics PhosphorImager. The signal intensity of the full-length triplet repeat-containing band was measured as a proportion of the total signal intensity in the lane above and below the band but not including the vector bands. The average percentage of full-length insert from three independent experiments was plotted on the *y* axis against the number of recultivations on the *x* axis. The error bars represent the S.D.

TABLE III

Summary of studies to identify interruptions in expanded GAA\textsuperscript{z}TTC repeats

| FRDA patient | GAA1 (no. of triplets) | GAA2 (no. of triplets) | Interruption detected by restriction enzyme digestion | Sequence |
|--------------|------------------------|------------------------|-----------------------------------------------------|----------|
| Individuals not identifying themselves as Acadian | | | | |
| M316 | 200 | 200 | Yes | (CTT)$_{160}$… |
| M326 | 490 | 750 | No | (TTC)$_{130}$… |
| M374 | 620 | 620 | No | (TTC)$_{140}$… |
| M524 | 590 | 790 | No | |
| M527 | 500 | 700 | No | |
| M621 | 590 | 920 | Yes | |
| M622 | 720 | 920 | No | |
| M623 | 720 | 1090 | No | |
| M642 | 920 | 1150 | No | |
| M649 | 620 | 1120 | No | |
| M662 | 490 | 950 | No | |
| M663 | 720 | 820 | No | |
| M665 | 750 | 850 | No | |
| M669 | 750 | 750 | No | |
| FRDA-Acadian | | | | |
| H39 | 800 | 800 | Yes | (TTC)$_{155}$… |
| M237 | 90 | 620 | Yes | (TTC)$_{130}$… |
| M258 | 806 | 806 | Yes | |
| M339 | 750 | 750 | No | (TTC)$_{136}$… |
| R1043 | 650 | 800 | No | (TTC)$_{136}$… |
| R1102 | 580 | 750 | No | (TTC)$_{132}$… |
| R1126 | 580 | 580 | No | (TTC)$_{130}$… |
| R1151 | 550 | 550 | No | (TTC)$_{130}$… |
striction enzyme analyses were conducted to attempt to identify GGA\textsubscript{TCC} triplets that were buried more deeply into the repeats (Table III); these analyses were positive for only two non-Acadian and two Acadian patients. Thus, neither the DNA sequence analyses nor restriction enzyme determinations enabled an exhaustive scan for interruptions; the former determination suffers from the weakness that it is not possible to sequence an entire expanded repeat tract of the average size found in Friedreich's ataxia patients (700–900 triplets). The restriction enzyme analyses suffer from the lack of specificity, since only a few enzymes are available that recognize the sequence configurations found by exchanging G\textsubscript{A}C for A\textsubscript{C}T base pairs in the GAA\textsubscript{TTC} triplet repeat sequence (i.e. the interruptions may, in fact, be present in the patient samples but might not have been detected with these probes).

**DISCUSSION**

To clarify the role of interruptions in GAA\textsubscript{TTC} repeats in their capacity to form triplexes and sticky DNA as well as to inhibit transcription and their involvement in genetic instabilities, we conducted the studies described herein. Considering the role of direct repeat sequences in the molecular pathology of the FRDA, two extreme cases exist. One is the long, uninterrupted GAA\textsubscript{TTC} repeats, which form the sticky DNA structure, inhibit transcription, and are associated with the disease (12). The other is the GAAGGA\textsubscript{TCC}TCCTTC hexamer repeat, which does not form sticky DNA, does not inhibit transcription, and does not cause the disease (12). However, the molecular reason for these phenomena was not clear. Several sequences of intermediate composition of these repeats were synthesized; up to 11% of GGA\textsubscript{TCC} interruptions in the long GAA\textsubscript{TTC} repeat did not influence the formation of sticky DNA. Furthermore, we showed that these GAA\textsubscript{TTC} repeats interrupted up to 11% form similar R\textsubscript{R}Y triplexes. However, these structures were slightly different from the structure of pure GAA\textsubscript{TTC}, because the GGA\textsubscript{TCC} interruptions were more sensitive to P1 nuclease compared with the surrounding GAA\textsubscript{TTC} repeats, indicating that the GGA\textsubscript{TCC} interruptions may be, at least partially, unpaired.

Thus, we hypothesize that the GAAGGA\textsubscript{TCC}TCCTTC repeats cannot form a triplex and/or sticky DNA structure due to their deleterious effect on base-pairing capabilities. The R\text{R}Y type
triplexes are based on TA*G and CG*G triads, which consist of normal Watson-Crick base pairs and reverse Hoogsteen pairs (Fig. 8B). In this case, the reverse Hoogsteen base pairs contain two hydrogen bonds (Fig. 8B). When we consider all three possible reading frames of Hoogsteen base pairings, the uninterrupted GAATTCC repeat can form a stable R-R-Y triplex using only TA*G and CG*G triads in one frame (Fig. 8A, left column). However, in the case of the GAAGGATTCTTC hexamer repeat, TA*G and CG*G triads must be included in all six possible frames (Fig. 8A, right column). As shown in Fig. 8C, both TA*G and CG*G triads contain only one hydrogen bond in each reverse Hoogsteen pair. Although the CG*G triad can have two hydrogen bonds in the reverse Hoogsteen base pair if the adenine is protonated at N-1 position (Fig. 8C, right column), this scheme is unlikely because protonation usually requires an acidic pH (24, 25). In the case of the hexamer repeat, at least 33% of bases have CG*G or TA*G mismatched triads (Fig. 8C, left column), and these mismatches probably destabilize the triplex and/or sticky DNA structures. We calculate that 11% of GAATTCC interruptions make ~7.4% of mismatched base triads in the minimum case. Hence, this may be the upper limit of mismatched base triads to maintain the triplex structure.

There were great differences between 11 and 20% of GAATTCC interruptions on the formation of sticky DNA and triplexes. We tried to prepare sequences with intermediate content to better understand the effect of the GAATTCC interruptions. However, our attempts to make (GAA)$_2$GGAA were unsuccessful, although the introduction of both more and fewer GAATTCC interruptions were successful. Three different experiments using different oligonucleotides with GGA interruptions at different positions were conducted. However, all cases resulted in the creation of the (GAA)$_2$GGAA sequence (data not shown). The reason for the failure to prepare the (GAA)$_2$GGAA is unclear but may be related to a DNA secondary structure, which is unfavorable for the (GAA)$_2$GGAA sequence.

We also evaluated the role of the GAATTCC interruptions on the genetic instabilities of long GAA/TTT repeats in the Escherichia coli system (2, 26, 27). A small number of interruptions (4%) served to substantially stabilize the GAA(TTT)$_n$ sequence. The hexamer repeat has a substantially interrupted sequence. The hexamer repeat has a substantially interrupted sequence, which is unfavorable for the (GAA)$_2$GGAA sequence.

We attempted to evaluate the role of interruptions in expanded GAATTCC repeat sequences in FRDA chromosomes in terms of their effect on phenotypic variability. Our current data do not permit the conclusion that interruptions in the GAATTCC expanded repeat sequence are a common cause of the variability in FRDA; the level of interruptions required to cause an effect on sticky DNA formation and in vitro transcription might have been detected in our analyses of patient materials. Alternatively, natural sequence variants, in addition to occurring in nonpathogenic alleles (13), may also play a role in rare patients with atypical phenotypes that have short interrupted repeats (29). One example is offered by two sibs in the series of atypical cases reported by Moseley et al. (30) who had a very slowly evolving disease and carried a (GAA)$_{10}$GGAA-(GAA)$_2$GGAA backbone (GAGGA)$_2$(GAA)$_{12}$ repeat along with a large expansion. Thus, the potential role of interruptions in FRDA phenotypic variability remains to be clarified.

REFERENCES

1. Sakamoto, N., Ohshima, K., Montermini, L., Pandolfo, M., and Wells, R. D. (2000) J. Biol. Chem. 275, 27171–27177
2. Ohshima, K., Kang, S., Larson, J. E., and Wells, R. D. (1996) J. Biol. Chem. 271, 16773–16783
3. Bidichandani, S. I., Ashizawa, T., and Patel, P. I. (1988) Am. J. Hum. Genet. 42, 111–121
4. Ohshima, K., Montermini, L., Wells, R. D., and Pandolfo, M. (1998) J. Biol. Chem. 273, 14558–14565
5. Wells, R. D., Collier, D. A., Hanvey, J. C., Shimizu, M., and Wohlrab, P. (1988) FASEB J. 2, 2939–2949
6. Frank-Kamenetski, M. D., and Mirkin, S. M. (1995) Annu. Rev. Biochem. 64, 65–95
7. Grabczyk, E., and Usdin, K. (2000) J. Biol. Chem. 275, 4704–4713
8. Sakamoto, N., Chastain, P. D., Parniewski, P., Ohshima, K., Pandolfo, M., Griffith, J. D., and Wells, R. D. (1999) Mol. Cell 3, 465–475
9. Ohshima, K., Campuzano, V., Reutenauer, L., Moutou, C., Mandel, J.-L., and Vescovi, A. L. (2000) J. Biol. Chem. 276, 5944–5959
10. Shimizu, M., Hanvey, J. C., and Wells, R. D. (1990) Biochemistry 29, 4704–4713
11. Cossee, M., Schmitt, M., Campuzano, V., Reutenauer, L., Moutou, C., Mandel, J.-L., and Koenig, M. (1997) Proc. Natl. Acad. Sci. U. S. A. 94, 7452–7457
12. Eichler, E. E., Holden, J. J., 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
13. Nancevarrow, J. K., Holman, K., Mangel, M., Hori, T., Denton, M., Sutherland, G. R., and Richards, R. I. (1995) Hum. Mol. Genet. 4, 367–372
14. Parniewski, P., Jaworski, A., Wells, R. D., and Bowater, R. P. (2000) J. Mol. Biol. 299, 865–874
15. Deng, W. P., and Nickoloff, J. A. (1992) J. Biol. Chem. 267, 23438–23446
16. Cossee, M., Schmitt, M., Campuzano, V., Reutenauer, L., Moutou, C., Mandel, J.-L., and Koenig, M. (1997) Proc. Natl. Acad. Sci. U. S. A. 94, 7452–7457
17. Eichler, E. E., Holden, J. J., 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
18. Nancevarrow, J. K., Holman, K., Mangel, M., Hori, T., Denton, M., Sutherland, G. R., and Richards, R. I. (1995) Hum. Mol. Genet. 4, 367–372
19. Parniewski, P., Jaworski, A., Wells, R. D., and Bowater, R. P. (2000) J. Mol. Biol. 299, 865–874
20. Deng, W. P., and Nickoloff, J. A. (1992) J. Biol. Chem. 267, 23438–23446
21. Cossee, M., Schmitt, M., Campuzano, V., Reutenauer, L., Moutou, C., Mandel, J.-L., and Koenig, M. (1997) Proc. Natl. Acad. Sci. U. S. A. 94, 7452–7457
22. Eichler, E. E., Holden, J. J., 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
23. Nancevarrow, J. K., Holman, K., Mangel, M., Hori, T., Denton, M., Sutherland, G. R., and Richards, R. I. (1995) Hum. Mol. Genet. 4, 367–372
24. Parniewski, P., Jaworski, A., Wells, R. D., and Bowater, R. P. (2000) J. Mol. Biol. 299, 865–874
25. Deng, W. P., and Nickoloff, J. A. (1992) J. Biol. Chem. 267, 23438–23446
26. Cossee, M., Schmitt, M., Campuzano, V., Reutenauer, L., Moutou, C., Mandel, J.-L., and Koenig, M. (1997) Proc. Natl. Acad. Sci. U. S. A. 94, 7452–7457
27. Eichler, E. E., Holden, J. J., 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
28. Nancevarrow, J. K., Holman, K., Mangel, M., Hori, T., Denton, M., Sutherland, G. R., and Richards, R. I. (1995) Hum. Mol. Genet. 4, 367–372
29. Parniewski, P., Jaworski, A., Wells, R. D., and Bowater, R. P. (2000) J. Mol. Biol. 299, 865–874
30. Deng, W. P., and Nickoloff, J. A. (1992) J. Biol. Chem. 267, 23438–23446
31. Cossee, M., Schmitt, M., Campuzano, V., Reutenauer, L., Moutou, C., Mandel, J.-L., and Koenig, M. (1997) Proc. Natl. Acad. Sci. U. S. A. 94, 7452–7457
GGA·TCC-interrupted Triplets in Long GAA·TTC Repeats Inhibit the Formation of Triplex and Sticky DNA Structures, Alleviate Transcription Inhibition, and Reduce Genetic Instabilities

Naoaki Sakamoto, Jacquelynn E. Larson, Ravi R. Iyer, Laura Montermini, Massimo Pandolfo and Robert D. Wells

J. Biol. Chem. 2001, 276:27178-27187. doi: 10.1074/jbc.M101852200 originally published online April 26, 2001

Access the most updated version of this article at doi: 10.1074/jbc.M101852200

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 28 references, 9 of which can be accessed free at http://www.jbc.org/content/276/29/27178.full.html#ref-list-1