The GAA triplet-repeat sequence in Friedreich ataxia shows a high level of somatic instability in vivo, with a significant predilection for large contractions

Rajesh Sharma¹, Saeeda Bhatti¹, Mariluz Gomez¹, Rhonda M. Clark¹, Cynthia Murray³, Tetsuo Ashizawa⁴,⁵ and Sanjay I. Bidichandani¹,²,*

¹Department of Biochemistry and Molecular Biology, and ²Department of Pediatrics, University of Oklahoma Health Sciences Center, Oklahoma City, OK, USA, ³Department of Mathematics and Statistics, University of Central Oklahoma, Edmond, OK, USA, and ⁴Department of Neurology and ⁵VA Medical Center, Baylor College of Medicine, Houston, TX, USA

Received May 14, 2002; Revised and Accepted June 28, 2002

Friedreich ataxia is commonly caused by large expansions of a GAA triplet-repeat (GAA-TR) sequence in the first intron of the FRDA gene. We used small-pool PCR to analyze somatic variability among 7190 individual FRDA molecules from peripheral blood DNA of subjects carrying 12 different expanded alleles, ranging in size from 241 to 1105 triplets. Expanded alleles showed a length-dependent increase in somatic variability, with mutation loads ranging from 47% to 78%. We noted a strong contraction bias among long alleles (>500 triplets), which showed a 4-fold higher frequency of large contractions versus expansions. Some contractions were very large; of all somatic mutations scored, 5% involved contractions of >50% of the original allele length, and 0.29% involved complete reversion to the normal/premutation length (<60 triplets). These observations contrast sharply with the strong expansion bias seen in expanded CTG triplet repeats in myotonic dystrophy. No somatic variability was detected in >6000 individual FRDA molecules analyzed from 15 normal alleles (8–25 triplets). A premutation allele with 44 uninterrupted GAA repeats was found to be unstable, ranging in size from 6 to 113 triplets, thus establishing the threshold for somatic instability between 26 and 44 GAA triplets. Analysis of an additional 7850 FRDA molecules from serially passaged lymphoblastoid cell lines carrying nine expanded alleles (132–933 triplets) showed very low mutation loads, ranging from 0% to 6.2%. Our data indicate that expanded GAA-TR alleles in Friedreich ataxia are highly mutable and have a natural tendency to contract in vivo, and that these properties depend on multiple factors, including DNA sequence, triplet-repeat length and unknown cell-type-specific factors.

INTRODUCTION

Friedreich ataxia (FRDA, OMIM 229300) is the most common inherited ataxia among Indo-Europeans (Europe, Middle East, North Africa and the Indian Subcontinent). It is an autosomal recessive, progressive and ultimately fatal disease, characterized clinically by onset before 25 years of age of limb and gait ataxia, dysarthria, areflexia, pyramidal signs, loss of position and vibration senses, secondary skeletal abnormalities, cardiomyopathy, and diabetes mellitus (1–3). The most common mutation, accounting for >95% of all mutant chromosomes, is a large expansion of a GAA triplet-repeat (GAA-TR) sequence in the first intron of the FRDA (X25) gene (4). The GAA-TR sequence is polymorphic, with normal and mutant chromosomes containing 5–60 and 100–1700 repeats, respectively (4–6). Among Indo-Europeans, 80–85% of chromosomes contain 5–11 triplets [small normal (SN) alleles], ~15% contain 12–32 triplets [long normal (LN) alleles], <1% contain 33–60 triplets (premutation alleles) and ~1% contain 100–1700 triplets (mutation/expanded alleles) (4–7). Expansion of the GAA-TR sequence beyond a critical length results in a length-dependent suppression of FRDA gene transcription mediated by an unconventional DNA structure adopted by the expanded GAA tract (8–12). Haplotype analysis has shown that LN alleles give rise to premutations, which undergo hyperexpansion to produce mutant alleles (6). At least

*To whom correspondence should be addressed at: Department of Biochemistry and Molecular Biology, University of Oklahoma Health Sciences Center, 975 N.E., 10th, BRC458, Oklahoma City, OK 73104, USA. Tel: +1 4052711360; Fax: +1 4052713910; Email: sanjay-bidichandani@ouhsc.edu
six de novo expansions of premutations have been documented by pedigree analysis (5,6,13,14, our unpublished data). Indeed, the frequency of LN alleles correlates well with the incidence of FRDA in several different populations (7).

While not much is known about the mechanism of hyperexpansion of premutation alleles, several studies have attempted to characterize the postzygotic events leading to somatic variability/mosaicism of the expanded GAA-TR sequence. Sibs who inherit the same parental mutant chromosome often show different repeat lengths. Comparison of repeat lengths in sperm versus blood DNA from males carrying expanded alleles shows a significantly shorter repeat size in sperm, in accordance with the frequent shortening of expanded alleles observed following paternal transmission (15–17). The expanded repeat is variable among the different subregions of the brain, with developmentally related regions showing similar repeat sizes (18). The expanded allele was shown to vary slightly between DNA isolated from various tissue sources from the same patient (19,20). Lymphoblastoid cell lines, generated by Epstein–Barr virus-mediated transformation of B lymphocytes from individuals with expanded GAA-TR alleles, almost always show a shorter repeat length compared with peripheral blood DNA from the same individual (21). While all of the above are examples of relatively small changes of the ‘constitutional’ repeat length, rare examples involving somatic mosaicism for complete deletion of the expanded GAA-TR alleles have also been reported in sperm and blood DNA (16,21).

Expanded GAA-TR alleles are usually detected by PCR amplification or by genomic Southern blot analysis of the FRDA gene (2,4). The expanded ‘band’ detected by these methods is generated from analysis of 0.06–12 µg genomic DNA, or about 10^5–2 × 10^6 cells. These two assays, employed in all of the investigations described above, only estimate the repeat size of the ‘constitutional or most common’ allele (also referred to as the ‘GAA-TR’ allele in this paper). However, to investigate triplet-repeat variability among individual cells or genomes, investigators have employed single sperm analysis and small-pool PCR (SP–PCR). The CTG-CAG triplet-repeats that cause myotonic dystrophy (22–27), SCA3 (28), SCA7 (29) and Huntington disease (30), and the CGG triplet repeat that causes fragile X syndrome (31,32), have been previously analyzed by these methods. Apart from the disparate triplet sequence motif, the CTG triplet-repeat expansion in myotonic dystrophy has certain features in common with the GAA triplet-repeat expansion in FRDA: disease-causing alleles are very similar in size, they are located in non-coding sequence and the expanded alleles are unmethylated in somatic cells. Single-genome analysis of the CTG triplet repeat by SP–PCR revealed a prominent expansion bias not otherwise appreciable by conventional PCR and Southern blot analyses (22–27). Besides serving as a potential model to study CTG triplet-repeat expansion, this observation also offered a molecular explanation for the stepwise length increments of the CTG repeat with increasing age.

We have conducted a comprehensive analysis of the GAA-TR sequence in the FRDA gene using SP–PCR. We have found that disease-causing GAA-TR alleles are extremely unstable in peripheral leukocytes in vivo, with ~65% of all somatic genomes containing repeats that differ significantly in size from the constitutional GAA-TR allele detected by conventional PCR and Southern blot analyses. We show that the threshold length for the initiation of somatic variability is between 26 and 44 uninterrupted GAA triplet-repeats. Unlike the expansion bias of the CTG triplet-repeat, we show that GAA-TR alleles have a strong tendency to contract—in some cases completely reverting to the normal/premutation size range. Surprisingly, the high level of somatic variability in blood cells was not seen in lymphoblastoid cell lines containing equally large GAA-TR alleles, indicating a potential role for other, cell-type-specific factors in mediating the somatic instability. We also show that clonal variation within oligoclonal cell lines, rather than repeat-length mutations per se, accounts for most of the GAA-TR length variation previously observed in serially passaged lymphoblastoid cell lines (21).

RESULTS

Single-gene analysis of the expanded GAA-TR sequence reveals a high mutation load in peripheral blood leukocytes in vivo

Peripheral blood DNA isolated from five heterozygous carriers and four FRDA patients homozygous for expanded GAA-TR alleles were analyzed by SP–PCR (see Materials and Methods). Subjects were selected because they carried a wide spectrum of expanded alleles ranging in size from 241 to 1105 triplets (Fig. 1A). The expanded GAA-TR alleles, seen as single bands by conventional PCR, were thus resolved into multiple individual bands that differed significantly from the respective constitutional GAA-TR allele sizes (Fig. 1B, C and D). Low DNA template concentrations (6–60 pg) frequently showed either non-amplification or single allele discrimination, in which only one of two homologous alleles were amplified (Fig. 1B and C). We believe that these bands reflect the allele sizes in individual FRDA genes (haploid genomes) that are otherwise not detectable when >10^6 cells are simultaneously analyzed by conventional PCR. For the following reasons, we argue that the multiple bands detected by SP–PCR of peripheral blood DNA are not due to artifacts or contamination during PCR amplification: (i) most significantly, expanded alleles of similar size from lymphoblastoid cell line DNA did not show the multiple band pattern seen with alleles amplified from peripheral blood DNA (see below); (ii) there was a proportional increase in the number of bands using increasing quantities of peripheral blood DNA, with a linear correlation between 6 and 60 pg; (iii) the size distribution of bands obtained from similarly sized GAA-TR alleles differed from each other; (iv) roughly equal numbers of normal and expanded bands were detected (77 versus 84) in reactions that allowed single-allele discrimination in heterozygotes, which is contrary to the expectation that expanded GAA-TR alleles would generate more spurious products; and (v) we did not observe any products of amplification in 336 different ‘zero-DNA’ control reactions performed throughout the course of this study.

Single-gene analysis was accomplished by performing SP–PCR with one or two FRDA molecules per reaction (mean ± SEM: 1.61 ± 0.13) (Fig. 1C). We thus analyzed 725 expansion-bearing FRDA molecules from peripheral blood leukocyte DNA from persons carrying 12 expanded...
GAA-TR alleles that ranged from 241 to 1105 triplets. Near-complete ascertainment of all input expanded FRDA genes was obtained, since we detected 0.93 (±0.11) and 0.68 (±0.06) expanded bands per input molecule in homozygotes and heterozygotes, respectively, which is close to the theoretically expected 1 and 0.5.

Significant variability was detected in each of the expanded GAA-TR alleles analyzed. The mutation load for the 12 GAA-TR alleles, defined as the proportion of molecules analyzed that differed in length from the respective constitutional GAA-TR allele by >5%, ranged from 47.2% to 78% (mean ± SEM 64.1% ± 2.7%). This means that two-thirds of the FRDA genes...
in peripheral blood leukocytes in vivo contain repeat sizes that differ significantly from the allele sizes estimated by conventional DNA diagnostic assays. Despite the high degree of somatic variability, the modal distribution of the observed bands correlated well with the size of the respective constitutional GAA-TR alleles calculated by genomic Southern blot and long-range PCR analyses ($r = 0.98$, $P < 0.001$), indicating that the constitutional allele is the most common allele. Somatic variability increased proportionately with increasing allele length (Fig. 2). Long GAA-TR alleles (>500 triplets) were on average 5-fold more variable than short GAA-TR alleles, as judged by their variance, i.e. standard deviation from their respective constitutional GAA-TR allele sizes [see Materials and Methods for calculation of variance; median = 131.1 (range: 81.0–220.8) versus median = 24.9 (range: 20.1–48.5); Mann–Whitney test $P = 0.016$; Figs 2 and 3). A large range of variable bands was detected, with maximal expansions ranging from 3% to 54% of the GAA-TR allele length and contractions ranging from 20% to 75% (Fig. 3).

Expanded GAA-TR alleles containing >500 triplets have a significant predilection for large contractions

Analysis of somatic variation among all the expanded GAA-TR alleles showed a relative paucity of expansions compared with contractions, manifesting as a significant negative skewing of band sizes about their respective GAA-TR allele lengths (Fig. 3). The frequency of contractions correlated significantly with the length of the GAA-TR allele ($r = 0.61$, $P = 0.035$). There was a strong tendency for expanded alleles to undergo large contractions, as indicated by the 3-fold excess frequency of contractions involving >15% of the GAA-TR allele length compared with similarly sized expansions (Fig. 4A). The predisposition for large contractions was mainly dependent on the length of the GAA-TR allele, with a significant correlation evident between the degree of negative skewing and the GAA-TR allele size ($r = −0.85$, $P < 0.001$). With increasing variability, the data were more negatively skewed, indicating a strong bias towards further contraction when alleles were more mutable ($r = −0.67$, $P = 0.018$).

Comparison of the three GAA-TR alleles containing <500 GAA repeats with the nine that contained >500 triplets indicated that the predilection for large contractions was confined to the long GAA-TR alleles (Fig. 4B). Overall, the long GAA-TR alleles showed a >2-fold excess of contractions over expansions, whereas the short GAA-TR alleles actually showed a slight tendency to expand. Long GAA-TR alleles were four times more likely to undergo large contractions ($P < 0.001$). Among the short GAA-TR alleles, the frequency of large contractions and expansions was similar, whereas a >4-fold excess of large contractions over expansions was detected for the long GAA-TR alleles (Fig. 4B). Furthermore, the maximum size of contractions for long GAA-TR alleles was twice as large as for the shorter GAA-TR alleles (51% versus 25%, $P = 0.003$).

Examination of the spectrum of variability of long GAA-TR alleles revealed an upper boundary above which expansion alleles were rare, with occasional examples of expansions >25% seen in only 4 of the 12 alleles analyzed (Fig. 3). The presence of an upper boundary for somatic variability was demonstrated experimentally by analyzing an additional 935 expansion bearing molecules from five different GAA-TR alleles (with 798–1088 triplets) using 10–20 input molecules per reaction (mean ± SEM 15.4 ± 1.15) (Fig. 1D). A clear upper boundary was detected near the size of the constitutional GAA-TR allele above which bands were rarely seen. The same
negative skewing of data was reproduced, with large contractions (>15% of GAA-TR allele length) far outnumbering large expansions (41.3 ± 4.7% versus 4.6 ± 0.7%, P = 0.0015). No expansions of >25% of the GAA-TR allele size were detected, but as many as 5.2% of the changes involved contractions comprising >50% of the GAA-TR allele length.

**Expanded GAA-TR alleles can occasionally revert to the normal/premutation size in peripheral leukocytes**

To accurately determine the extent of very large contractions, including the detection of any complete contractions of expanded GAA-TR alleles, we analyzed DNA samples from four patients who were homozygous for GAA-TR expansions (i.e. without any normal alleles) using a primer pair that accurately estimates the size of short repeats (see Materials and Methods). The eight GAA-TR alleles thus analyzed ranged from 241 to 1105; a total of 5532 individual FRDA molecules were amplified with ~30 molecules per reaction. For convenience, we scored all large contractions as those measuring ≤1 kb in size, i.e. all bands with ≤167 triplets (Fig. 5). Of the molecules analyzed, 2.3% (126 of 5532) were scored as very large deletions, which included 16 (0.29%) cases of contraction into the normal or premutation range (<60 triplets) (Fig. 5). These very large deletions and complete reversion events were detected in each of the four homozygous individuals tested, indicating that it is a generalized phenomenon involving multiple expanded GAA-TR alleles.

**The threshold for somatic instability of the GAA-TR sequence is between 26 and 44 uninterrupted triplets**

Single-gene analysis of 226 normal FRDA molecules, from five heterozygous individuals, with normal alleles ranging from 2 to 349 triplets, showed that all alleles were stable.
8 to 17 triplets, showed that normal alleles were completely stable. Since this was carried out using primers designed for long-range PCR that amplify 1.45 kb of flanking intron 1 sequence, and may therefore not reveal small variations, we further analyzed 11 additional normal alleles using a pair of primers closer to the GAA repeat (see Materials and Methods). Of the normal alleles thus analyzed, six had either 8 or 9 uninterrupted GAA triplets (SN alleles) and the other five ranged from 17 to 25 uninterrupted triplets (LN alleles). SP–PCR analysis of 6000 normal FRDA molecules, 2950 SN and 3050 LN, showed no evidence of instability (Fig. 6A).

Additionally, we found a patient (OK99) who had a clinical presentation compatible with the diagnosis of atypical FRDA (M. Gomez et al., manuscript in preparation) and whose FRDA genes contained 44 and 696 GAA triplet repeats, respectively. Direct sequencing of the shorter allele showed 44
uninterrupted GAA triplets. SP–PCR analysis of the GAA-44 allele in peripheral blood DNA showed that it was somatically unstable (Fig. 6B). SP–PCR analysis of 2304 FRDA molecules derived from the GAA-44 allele revealed 145 variant molecules, i.e. a mutation load of 6.3%, which is 10-fold less than the 64% average mutation load seen with the larger, disease-causing expanded alleles. Moreover, analysis of the size distribution of the variant molecules clearly showed an expansion bias, with variant bands ranging from 6 to 113 repeats or 14% to 257% of the GAA-44 allele (Fig. 6C). There was an over-representation of short expansions, with variant bands containing 50–60 triplets accounting for 46% (n = 66) of all mutants. Remarkably, there was a paucity of contractions of the same magnitude, with only 2% (n = 3) of all mutants containing 28–38 triplets. It was not possible to unequivocally detect any large expansions of the GAA-44 allele, since this patient also carried an unstable, large GAA-TR allele, and therefore the actual spectrum of instability in premutation-size alleles remains to be determined. However, stability of the GAA-25 allele (on screening 1173 individual FRDA molecules) and the observed instability of the GAA-44 allele indicate that the threshold for somatic instability is likely to be between 26 and 44 uninterrupted GAA triplets.

The expanded GAA-TR is relatively stable in lymphoblastoid cell lines, with some of the observed changes being attributable to spontaneous clonal variation within oligoclonal cell lines

Initial SP–PCR analyses using one or two input molecules per reaction showed almost no variability when genomic DNA from multiple lymphoblastoid cell lines was analyzed (unpublished data). To facilitate the detection of rare changes, we therefore performed all subsequent SP–PCR assays with 6–29 FRDA molecules per reaction (mean ± SEM 13.4 ± 2.8) (Fig. 7). A total of 14 350 individual FRDA molecules were analyzed, for nine disease-causing GAA-TR alleles (range 132–933 triplets, n = 7865) and five normal alleles (range 8–9 triplets, n = 6485), which were derived from five heterozygous and two homozygous cell lines, grown in culture for 12–20 serial passages. The mutation load ranged from 0% to 6.2% (mean ± SEM 2.12 ± 0.78) for all the cell lines tested, which is at least 30-fold less than the frequency observed in peripheral leukocytes in vivo (P < 0.001) (Fig. 7A and B). However, this is likely to be an overestimation of the mutation load due to two intriguing phenomena observed upon serial passaging of lymphoblastoid cell lines. Firstly, there was a generally higher mutation load in early passages, reflecting the oligo/polyclonal nature of the cell lines. Upon subsequent passaging, however, many of the clones (bands) spontaneously disappeared, so that only one predominant clone populated the entire culture resulting in a reduction of the mutation load (Fig. 7C).

Despite the overall lower frequency of mutations, the same length-dependent contraction bias seen in leukocytes was also detected among the changes seen in lymphoblastoid cell lines. The three short GAA-TR alleles (<500 triplets) were extremely stable, with mutation loads ranging from 0% to 0.043% (mean ± SEM 0.15 ± 0.14), compared with 0.41%–6.2% (mean ± SEM 2.75 ± 0.89) for GAA-TR alleles containing >500 triplets (P = 0.03). As in the case of leukocytes, the frequency of contractions (mean ± SEM 0.82 ± 0.11) in lymphoblastoid cell lines was significantly greater than that of expansions (mean ± SEM 0.066 ± 0.02, P < 0.001).

DISCUSSION

We have shown that the expanded GAA-TR sequence is highly mutable in peripheral leukocytes in vivo and that on average almost two-thirds of the genes contain repeat lengths that differ significantly from the constitutional GAA-TR allele. Implicit in our analyses is the assumption that the constitutional GAA-TR allele is the initial allele size from which all variable bands originate. While this assumption is not entirely accurate, since mutant alleles themselves are also likely to serve as templates for additional mutations, it nevertheless serves as a useful way to perform a semiquantitative analysis of the observed mutational spectrum. Our data revealed a length dependence in the somatic mutational spectra of GAA triplet repeats. Alleles with <25 triplets are completely stable, and somatic instability (at least in leukocytes) initiates above a threshold length between 26 and 44 triplets. Expanded GAA-TR alleles show a length-dependent increase in mutability. Alleles with
<500 triplets predominantly undergo small length changes with a slight expansion bias, while expanded alleles with >500 triplets, the most frequent disease-causing alleles, show a strong bias towards large contractions.

Our data are consistent with the presence of two types of changes: (i) small changes of <5% of the GAA-TR allele length and (ii) large changes involving >15% of the GAA-TR allele length. Among the small changes, expansions and contractions were equally frequent, and no difference was noted when comparing the mutability of long versus short GAA-TR alleles. However, among the large changes, contractions outnumbered expansions, and this bias was predominantly seen in GAA-TR alleles with >500 GAA triplets. Short length variations in repeat tracts are thought to occur as a result of slippage and mispairing during DNA replication (33–37). It is likely that this is the mechanism by which the equally frequent short contractions and expansions (<5%) of the GAA-TR alleles are generated. However, we propose that the strong tendency of long GAA-TR alleles to undergo large contractions in vivo arises from errors caused during lagging-strand synthesis (38). Most expanded GAA-TR alleles associated with FRDA contain 600–1200 triplets (2,4,20,39), which is sufficiently long to be spanned by several eukaryotic Okazaki fragments. Therefore, multiple Okazaki fragments would have to be initiated, extended, processed and ligated within the replicating repeat tract. Stable or metastable secondary structure(s) adopted by the single-stranded GAA-TR template (40–42) would result in the bypassing of a variable number of repeats in the nascent strand, thus effectively producing contractions. We hypothesize that the single-stranded GAA-TR is likely to have more freedom to adopt such structures during lagging-strand replication as compared with the complementary TTC sequence due to the 50-fold reduced binding affinity of human replication protein A for polypurine sequences (43,44). Longer repeats would provide multiple opportunities for such extrusion events and result in larger contractions, consistent with our experimental observations. It should be noted that our data do not rule out any

Figure 6. The threshold for somatic instability of the GAA-TR sequence is between 26 and 44 uninterrupted triplets. (A and B) SP–PCR analysis of normal and premutation-sized GAA-TR alleles showing that alleles with ≤25 uninterrupted triplets are completely stable (A) whereas an allele with 44 uninterrupted triplets is somatically unstable (B). (C) Of the 2304 individual FRDA molecules analyzed from the GAA-44 allele, 145 (6.3%) variant molecules (range 6–113 triplets) were identified. The variant molecules are shown on a scatter plot with the repeat sizes indicated on the vertical axis. The vast majority of bands contained 44 triplets, and a line at the appropriate level represents these. Observations above and below the line indicate expansions and contractions, respectively. The mutational spectrum indicates that the GAA-44 allele has a distinct tendency to expand, with almost half of all variant molecules ranging between 50 and 60 triplets (see text).
Figure 7. The GAA-TR sequence is relatively stable in lymphoblastoid cell lines. (A and B) SP-PCR analysis showing a remarkable degree of stability of expanded GAA-TR alleles (GAA-110 and GAA-801 are shown here) in lymphoblastoid cell lines. Asterisks indicate rare variants. (C) Spontaneous clonal purification in a serially passaged lymphoblastoid cell line. Multiple bands are seen in an early passage (passage 1), reflecting the polyclonal nature of the cell line; however, upon serial passaging, a single major clone containing 734 triplets is seen (passage 12). Note the presence of some variant bands (indicated by the asterisk), but it is not possible to determine if these are de novo mutation events or examples of clonal repopulation. (D) Spontaneous clonal variation in a serially passaged lymphoblastoid cell line. Bands are seen in an early passage (passage 1), reflecting the oligoclonal nature of the cell line; however, upon serial passaging, the GAA-734 clone is completely replaced by a GAA-995 clone (passage 12). Note that the GAA-995 repeat is seen in passage 1, and, likewise, the GAA-734 clone is also detected in passage 12.
potential role that triplex formation by GAA-TR sequences (8–12) may have in mediating the observed somatic instability. It is unclear what determines the threshold for somatic instability of GAA-TR alleles between 26 and 44 triplets. While it is possible that this length represents the threshold for a slippage-mediated mechanism, it is, however, intriguing that this repeat length is also approximately the size of a eukaryotic Okazaki fragment. Given the strong expansion bias of the GAA-44 premutation allele in peripheral leukocytes, and that premutation alleles involved in germline hyperexpansions at the FRDA locus range from 34 to ~60 triplets (5,6,13,14, our unpublished data), it is likely that the factors determining the threshold for somatic and germline instability may be similar. Similar-sized alleles of other triplet-repeat disorders have also been analyzed by SP–PCR of leukocyte DNA, and they seem to have a very low mutation frequency compared with the 6% mutation load that we observed with the GAA-44 allele. The CGG-55 allele at the fragile X locus showed a somatic mutation frequency of \(3.6 \times 10^{-4}\) (31). Another study found that an uninterrupted CGG-28 allele had a mutation frequency of \(7.3 \times 10^{-5}\) (32). A CTG-42 allele at the myotonic dystrophy locus was found to be completely stable by SP–PCR of leukocyte DNA (27). The variation in mutation frequencies may stem from differences in the primary triplet-repeat sequence, secondary structure, epigenetic modification, nucleosomal organization or other locus-specific factors.

We previously reported a patient (21) who had inherited two expanded GAA-TR alleles, one from either parent, but in addition showed somatic mosaicism with a heterogeneous mix of GAA-TR alleles in the normal range (9–29 triplets). A similar example of complete reversion of an expanded GAA-TR allele to the normal range has been reported in sperm DNA (16). Here we have shown that the rate of reversion of expanded alleles to the normal/premutation size range (≤60 triplets) is 0.29%. While this frequency is likely to be too low to account for any phenotypic variation, it nevertheless demonstrates that complete contractions are mechanistically possible in somatic cells in vivo. We believe that these reversions do not represent PCR contamination or artifacts because we did not observe any products of amplification in 336 different ‘zero-DNA’ reactions performed throughout the course of this study, because the sizes of the observed bands differed significantly from each other (range 5–60 triplets) and because no such contraction events were seen using DNA from lymphoblastoid cells.

Given the observed mean mutation load of ~65% in peripheral leukocytes, we employed a simple model to estimate the somatic mutation frequency of expanded GAA-TR alleles. Notwithstanding the different developmental lineages of the various classes of leukocytes, for the sake of simplicity if we assume that an adult human produces \(10^{11}\) peripheral blood leukocytes per day from \(10^7\) hematopoietic stem cells (see Materials and Methods), a postmitotic leukocyte would have undergone approximately 13 binary divisions prior to entering the peripheral circulation. We therefore estimate that the mutation frequency, i.e., the proportion of mitotic divisions that result in mutations of the expanded GAA-TR sequence, is ~10% per division, since that would give rise to a 65% mutation load in 13 binary divisions (a 1% mutation frequency, with one mutant daughter cell per defective mitosis, would result in a cumulative mutation load of 0.005 per division). This is clearly a rough estimate of the actual mutation frequency, given our assumptions that all mutations are caused by single events, that all stem cells contain the same repeat length (the constitutional GAA-TR allele length) and that mitotic divisions during the development of all types of peripheral leukocytes have similar mutational frequencies and spectra. This mutation frequency, although very high when compared with common microsatellite repeat polymorphisms \((10^{-3}–10^{-6})\) (45–47), is similar to what is observed with expanded CTG alleles containing >200 triplet repeats at the myotonic dystrophy locus (22,24).

Patients with FRDA show a wide range of clinical presentations. The length of the expanded GAA-TR in peripheral blood DNA from FRDA patients correlates significantly with various parameters of disease severity, including age of onset, rate of disease progression, loss of deep tendon reflexes, and the presence and severity of cardiomyopathy and secondary skeletal abnormalities (2,17,20,39). However, these correlations are quite variable, and it is often not possible to predict the clinical severity in individual patients based on the size of their GAA triplet repeats. For example, 33–73% of the variability in the age of onset has been attributed to the length of the shorter of the two expanded GAA-TR alleles (2,17,39,48). Sibs who inherit the same parental expanded GAA-TR alleles can have widely varying clinical presentations (19,49–51). All of the clinical correlation studies are based on the size of the constitutional GAA-TR alleles, which we have now shown to vary considerably represent only 22–53% of somatic genes in vivo. We believe that this interindividual variability in the levels of somatic instability is a potential phenotypic determinant, and may account for some of the discrepancies observed during genotype–phenotype correlation.

The CTG triplet-repeat expansion in myotonic dystrophy shows an exact opposite pattern of somatic mutation. The expanded CTG triplet repeat shows a prominent expansion bias during germline transmission (27,52), in somatic cells in vivo, especially with increasing age of patients (22,24,53), and upon serial passaging of patient-derived cells in vitro (23). The CTG repeat in myotonic dystrophy has some features in common with the GAA-TR sequence in FRDA. Disease-causing alleles of both disorders are similar in size, are located in non-coding sequences (unlike the polyglutamine-encoding CAG repeats) and, in contrast to the CGG triplet repeat in fragile X syndrome, both are unmethylated in somatic cells. The opposite orientation of mutations detected in our experiments is possibly due to differences mediated by the primary triplet sequence, such as secondary DNA structure(s), nucleosomal organization and differential interactions with trans-acting factors.

It is not clear why the mutation load in serial passages of lymphoblastoid cell lines was so low even though approximately 30–50 cell divisions had occurred over the course of our experiments. Despite the significantly lower mutation load, it is noteworthy that the preponderance of contractions in leukocytes was also observed in this cell type. While the mutational spectrum in lymphoblastoid cell lines is consistent with our model explaining the predilection for large contractions in long GAA-TR alleles, the lower overall frequency of mutations points to a role for other cell-type-specific factors, including, for example, the rate of cellular proliferation, state(s) of...
differentiation and possible effects of transformation on DNA metabolism.

We have previously shown that expanded GAA-TR alleles can vary considerably in size during serial passaging of lymphoblastoid cell lines in vitro (21). Those experiments, performed using genomic Southern analysis, represented changes in the constitutional GAA-TR allele length. However, by SP–PCR analysis of one such change, we detected a small proportion of cells in previous passages that contained the ‘new’ GAA-TR allele, indicating that the apparent change in GAA-TR length was likely due to spontaneous variation in clonal subpopulations rather than by de novo mutagenesis. It should be noted that our data do not rule out the possibility that these changes are de novo mutations in triplet repeats. Similar clonal variations were also detected in cell lines that did not show a complete change in the constitutional GAA-TR allele length. Another intriguing finding involved spontaneous clonal purification where multiple bands in early passages converted to one predominant band over a few passages. These findings indicate that variations among clonal subpopulations occur spontaneously and continuously during the growth of lymphoblastoid cell lines in vitro. This is an important and poorly appreciated property of lymphoblastoid cell lines that may have been responsible for some previously reported variation in continuous cultures (54–58), and needs to be considered when interpreting data obtained from such experiments. We therefore believe that the mutation load observed in lymphoblastoid cell lines is not an accurate estimate of the actual mutation frequency in this cell type.

In summary, our data demonstrate that the expanded GAA-TR alleles in FRDA are highly mutable and result in very high mutation loads in somatic cells in vivo. Disease-causing, expanded GAA-TR alleles have a strong natural tendency to contract and can spontaneously revert to the normal/pemutation size. Understanding the molecular mechanism(s) of these contractions may provide unique opportunities to effect or accelerate such changes as a potential therapeutic strategy.

**MATERIALS AND METHODS**

**SP–PCR analysis at the FRDA locus**

SP–PCR was performed using a modification of a previously published protocol (22). Serial dilutions of genomic DNA, ranging from 6 to 1200 pg, were prepared in siliconized microfuge tubes from 12 ng/ml stock solutions made using a fluorometer. PCR was performed using either of two primer pairs: (i) T3F (5'-GGC CGC AAC AAT TAA CCC TCA CTA AAG GGA ACA GGA GGG ATC CGT CTG GGC AAA GG-3') and T7R (5'-CCG CGC AAT TAA TAG GAC TCA CTA TAG GGC GAC AAT CCA GCA CAG TCA GGG TTT T3'), which amplify 1.45 kb of flanking, non-repeat intron 1 sequence using the GeneAmp XL PCR kit (Applied Biosystems) and the long-range PCR cycle described previously (4), and enable efficient amplification of expanded GAA-TR sequences; (ii) GAA-104F and GAA-629R (16), which amplify 499 bp of flanking intron 1 sequence, and allow accurate sizing of normal alleles and short expansions (<500 triplets). All disease-causing GAA-TR allele sizes were measured using T3F/T7R and 0.1–0.5 µg peripheral blood DNA. Normal and premutation alleles were directly sequenced to determine the exact length and purity of the GAA-TR sequence. SP–PCR was assembled in a dedicated PCR chamber (PLAS Labs, Michigan) following a 20 min exposure to short-wave ultraviolet irradiation to inactivate any pre-PCR contamination. ‘Zero-DNA’ blanks were made for every master mix. Strict separation of pre- and post-PCR zones was maintained at all times. PCR products were resolved by electrophoresis on 0.8–2% agarose gels, depending on the size of the bands being analyzed. The products were analyzed by Southern blotting using an end-labeled TTC-9 oligonucleotide probe that specifically hybridizes to the GAA-TR sequence.

**Data analysis and statistical calculations**

Band sizes were estimated from exponential curves generated by two DNA size standards run on every gel. The concentration of template DNA that resulted in multiple failed reactions was used to calculate the average number of DNA molecules per reaction by Poisson analysis (59) using the formula $-\ln(m/n)$, where $m$ is the number of lanes where no signal was detected and $n$ is the total number of reactions. The ‘gene equivalent’, i.e. the amount of genomic DNA required to detect one amplifiable FRDA molecule for the various DNA preparations, was calculated from multiple such reactions. A 6.5-fold excess of input DNA was required to amplify one FRDA molecule in heterozygotes (mean±SEM 41.7±7.6 pg) versus homozygotes (mean±SEM 6.39±0.75 pg). This may reflect the reduced efficiency of PCR during simultaneous amplification of alleles of widely varying size, since the gene equivalent for normal individuals (homozygous for GAA-TR alleles with <25 repeats) was similar to that seen with DNA from patients who were homozygous for large GAA-TR expansions. Multiple reactions were performed using two different template concentrations; 1–2 gene equivalents per reaction to determine ‘frequent events’ of somatic variation, and 10–30 gene equivalents per reaction to detect ‘rare events’.

Mutation load was estimated as the proportion of molecules that differed in length from the respective constitutional GAA-TR allele by >5%. Variance was measured as the standard deviation of variant molecule sizes from the respective constitutional GAA-TR allele length, and was calculated as

$$\sqrt{\frac{(O_L - G_L)^2}{n - 1}}$$

where $O_L$ is the observed variant band length, $G_L$ is the constitutional GAA-TR allele length and $n$ is the number of observations. Skewness is the measure of lack of symmetry of distribution of the observed variable bands about the respective constitutional GAA-TR allele, and was calculated as the third moment about the GAA-TR allele: $m_3 = m_3^* - \frac{3}{2}m_2^*$, where

$$m_3 = \frac{\sum (O_L - G_L)^3}{n - 1}$$

$$m_2 = \frac{\sum (O_L - G_L)^2}{n - 1}$$

$\text{where}$
All statistical tests were performed using SAS v6.12 (Cary, NC) and Analysis ToolPak in Microsoft Excel 2000.

The following data were used to calculate the numbers of peripheral leukocytes and hematopoietic stem cells in an adult human weighing 70 kg. (i) Adult humans produce $0.85 \times 10^9$ neutrophils/kg/day, i.e. $6 \times 10^{10}$ cells/day (60), and since neutrophils account for $\sim 60\%$ of the peripheral leukocyte population, we assumed the total leukocyte production to be $10^{11}$ per day. (ii) Stem cells are thought to represent $10^{-4}$ cells of the total nucleated bone marrow population. The total nucleated cell population in human bone marrow is estimated to be $10^9$ to $10^{10}$ g of body weight (61), i.e. $7 \times 10^{11}$ in a 70 kg adult, and therefore the number of stem cells is $7 \times 10^7$. We assumed the total stem cell population to be $\sim 10^7$, since a fraction of them (approximately one-third) are estimated to be actively dividing (61,62).

**ACKNOWLEDGEMENTS**

We are grateful to the patients and their families for participating in this study and to Drs Astrid Rasmussen and Alexander Cooke for providing us with samples. We thank Drs Darren Monckton, Richard Sinden and Gillian Dalgliesh for critically reviewing the manuscript and for their many useful suggestions. This work was supported in part by grants from the American Heart Association, American Diabetes Association, OCAST (Oklahoma Center for Advancement of Sciences and Technology) and FARA (Friedreich Ataxia Research Alliance) to S.I.B.

**REFERENCES**


