Friedreich's ataxia: Autosomal recessive disease caused by an intronic GAA triplet repeat expansion
Campuzano, Victoria;Montermini, Laura;Molto, Maria Dolores;Pianese, Luigi;et al
Science; Mar 8, 1996; 271, 5254; ProQuest Central
pg. 1423

Friedreich's Ataxia: Autosomal Recessive Disease Caused by an Intronic GAA Triplet Repeat Expansion

Victoria Campuzano,* Laura Montermini,* Maria Dolores Moltò,† Luigi Pianese, Mireille Cossée, Francesca Cavalcanti,‡ Eugenia Monros, François Rodius, Franck Duclos,§ Antonella Monticelli, Federico Zara, Joaquin Cañizares, Hana Koutnikova, Sanjay I. Bidichandani, Cinzia Gellera, Alexis Brice, Paul Trouillas, Giuseppe De Michele, Alessandro Filla, Rosa De Frutos, Francisco Palau, Pragna I. Patel, Stefano Di Donato, Jean-Louis Mandel, Sergio Cocozza, Michel Koenig,‖ Massimo Pandolfo‖

Friedreich's ataxia (FRDA) is an autosomal recessive, degenerative disease that involves the central and peripheral nervous systems and the heart. A gene, *X25*, was identified in the critical region for the FRDA locus on chromosome 9q13. The gene encodes a 210–amino acid protein, frataxin, that has homologs in distant species such as *Caenorhabditis elegans* and yeast. A few FRDA patients were found to have point mutations in *X25*, but the majority were homozygous for an unstable GAA trinucleotide expansion in the first *X25* intron.

Friedreich's ataxia (*1*) is the most common hereditary ataxia, with an estimated prevalence of 1 in 50,000 and a deduced carrier frequency of 1 in 120 in European populations (*2*). FRDA is an autosomal recessive degenerative disease characterized by a progressive gait and limb ataxia, a lack of tendon reflexes in the legs, loss of position sense, dysarthria, and pyramidal weakness of the legs (3). Hypertrophic cardiomyopathy is found in almost all patients (*4*). Diabetes mellitus is seen in about 10% of patients, carbohydrate intolerance in an additional 20%, and a reduced insulin response to arginine stimulation in all patients (*5*). The age of onset is usually around puberty, and almost always before age 25. Most patients are confined to a wheelchair by their late 20s, and there is no treatment to slow the progression of the disease.

The first pathologic changes are thought to occur in the dorsal root ganglia with the loss of large sensory neurons, followed by deterioration of the sensory posterior columns, spinocerebellar tracts, and corticospinal motor tracts of the spinal cord, and atrophy of the large sensory fibers in peripheral nerves. Only occasional mild degenerative changes are seen in the cerebellum, pons, and medulla. Although most symptoms are a consequence of neuronal degeneration, cardiomyopathy and diabetes are thought to reflect independent sites of primary degeneration. Overall, the pathology of FRDA (*6*) is very different from that of other hereditary ataxias, particularly the dominant forms and ataxia telangiectasia, where the cerebellum is the primary site of degeneration (*7*).

The mutated gene in FRDA has been mapped to chromosome 9q13-q21.1 (*8*). We recently narrowed the FRDA candidate region to a 150-kb segment flanked by the ZO-*2* gene (distal) and the marker F8101 (proximal) (*9*). Previously proposed candidate genes are excluded: the *X104* (also referred to as *CSFA1* and ZO-*2*) gene (*10*) on the basis of the absence of deleterious mutation in patients, and the *STM7* (*11*) and *PRKACG* (*9*) genes because they lie in entirety on the centromeric side of F8101 (Fig. 1A).

Potential exons were identified in the FRDA critical region by direct complementary DNA (cDNA) selection, exon amplification, and computer prediction from random sequences (*12*). One exon, called d26, was identified independently by the last two approaches. Nested primers based on the d26 sequence, when used in a rapid amplification of cDNA 5′ end (5′-RACE) experiment on a heart cDNA template (*13*), yielded two independent but overlapping products. Sequence from these clones matched another amplified exon and an expressed sequence tag (EST) from a human liver plus spleen cDNA library (*14*). This gene, called *X25*, apparently had alternate transcripts, because the sequences at the 3′ end of the EST and RACE products were different.

The gene structure of *X25* (Fig. 1A) was resolved by obtaining intronic sequences flanking the identified exons, by inverse polymerase chain reaction (PCR), and by the direct sequencing of cosmids (*15*). Five exons (named 1 to 5a, where exon 5a corresponds to the 3′ end of the EST) were found to be spread over 40 kb. They contain an open reading frame (ORF) encoding a 210–amino acid protein (*16*), which we named frataxin (Fig. 1B). An alternative exon (5b), corresponding to d26, was localized at about 40 kb from exon 5a in the telomeric direction. Exon 5b also has an in-frame stop codon, so that the alternative transcript encodes a shorter, 171–amino acid protein, whose 11 COOH-terminal residues differ from the main isoform. Nucleotide sequences of the *X25* exons have been deposited in the GenBank database under the accession numbers U43748 to U43753. We further investigated the 3′ end of the transcript encoding the alternative form by 3′-RACE (*17*) and showed that, depending on the alternate usage of the 3′ donor splice site in exon 5b, either a transcript ending with this exon or a longer transcript including an additional noncoding exon 6 could be generated (*18*).

A polyadenylated [poly(A)$^+$] Northern

V. Campuzano, M. Cossée, F. Rodius, F. Duclos, H. Koutnikova, J.-L. Mandel, M. Koenig, Institut de Génétique et Biologie Moléculaire et Cellulaire (IGBMC), INSERM, Université Louis Pasteur, CNRS, and Centre Hospitalier Universitaire, 1 Rue Laurent Fries, BP 163, 67404 Illkirch Cedex - Strasbourg, France.
L. Montermini, M. D. Moltò, F. Zara, S. Bidichandani, P. I. Patel, Department of Neurology, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA.
L. Pianese, A. Monticelli, S. Cocozza, Dipartimento di Biologia e Patologia Cellulare e Molecolare, University of Naples Federico II, and Centro Endocrinologia Oncologia Sperimentale Consiglio Nazionale delle Ricerche, Via Pansini 5, 80131 Naples, Italy.
F. Cavalcanti, Istituto Mediterraneo di Neuroscienze "Sanatrix," 86077 Pozzilli (Isernia), Italy.
E. Monros and F. Palau, Unidad de Genetica, Hospital Universitari La Fe, Avenida Campanar 21, 46006 Valencia, Spain.
J. Cañizares and R. De Frutos, Department de Genetica, University of Valencia, C/ Dr. Moliner 50, 46100 Burjassot (Valencia), Spain.
C. Gellera and S. Di Donato, Divisione di Biochimica e Genetica del Sistema Nervoso, Istituto Nazionale Neurologico "C. Besta," Via Celoria 11, 20133 Milan, Italy.
A. Brice, Hôpital de la Salpetriere, 47 Boulevard de l'Hôpital, 75651 Paris Cedex 13, France.
P. Trouillas, Hôpital Neurologique, Pier Wertheimer, 69003 Lyon, France.
G. De Michele and A. Filla, Dipartimento di Patologia Sistematica, Divisione di Neurologia II, University of Naples Federico II, Via Pansini 5, 80131 Naples, Italy.
M. Pandolfo, Department of Neurology, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA, and Divisione di Biochimica e Genetica del Sistema Nervoso, Istituto Nazionale Neurologico "C. Besta," Via Celoria 11, 20133 Milan, Italy.

*These authors equally contributed to this work.
†Permanent address: Departament de Genetica, University of Valencia, C/ Dr. Moliner 50, 46100 Burjassot (Valencia), Spain.
‡Present address: Dipartimento di Biologia e Patologia Cellulare e Molecolare, University of Naples Federico II, and Centro Endocrinologia Oncologia Sperimentale Consiglio Nazionale delle Ricerche, Via Pansini 5, 80131 Naples, Italy.
§Present address: Howard Hughes Medical Institute, Department of Physiology and Biophysics, University of Iowa College of Medicine, 400 Eckstein Medical Research Building, Iowa City, IA 52242, USA.
‖To whom correspondence should be addressed.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

(RNA) blot of different human tissues revealed the highest level of expression of *X25* in heart, intermediate levels in liver, skeletal muscle, and pancreas, and minimal levels in other tissues, including whole brain (Fig. 2). A 1.3-kb major transcript was identified, in agreement with the predicted size of an exon 5a–containing mRNA. Fainter bands of 1.05, 2.0, 2.8, and 7.3 kb were also detected (*19*). A Northern blot of total RNA from selected parts of the central nervous system (CNS) revealed high levels of expression of the 1.3-kb transcript in the spinal cord, with less expression in cerebellum, and very little in cerebral cortex (*20*). Overall, expression of *X25* appeared to be highest in the primary sites of degeneration in FRDA, both within and outside the CNS. To investigate the nature of the larger transcripts, we screened a fetal brain cDNA library with the EST clone (exons 2 through 5a). Among nine positive clones, four were isolated whose sequence extended beyond the limits of the previously identified *X25* mRNAs. Sequence analysis of these clones indicated that they originated from a related gene, differing from *X25* at several positions, and with stop codons in the sequence corresponding to exon 1 of *X25* (*21*). The *X25*-related gene was excluded from the critical FRDA region, and at least one intronless copy exists in the genome, as indicated by Southern (DNA) blot and PCR analysis (*22*).

A BLASTN DNA database search with the *X25* DNA sequence and a BLASTP search with the translated sequence (*23*) did not reveal any significant match. However, a TBLASTN search, in which the protein sequence was compared with the six-frame translation of the DNA databases, yielded highly significant matches with an ORF contained in a *Caenorhabditis elegans* cosmid (BLAST probability = 7.6×10^{-13}), and with a *Saccharomyces cerevisiae* EST (BLAST probability = 2.0×10^{-10}) (Fig. 1B). In both cases, the closest match involved a 27–amino acid segment of the protein (positions 141 through 167) encoded in exons 4 and 5a, showing 25 out of 27– and 22 out of 27–amino acid identity with the *C. elegans* and *S. cerevisiae* sequences, respectively, and 65% identity at the DNA level. Secondary structure predictions (*24*) for the X25-encoded protein suggested an α helical structure for the NH_2-terminal 30 amino acids and the regions between residues 90 to 110 and 185 to 195, with possible interspersed β sheet regions around residues 125 to 145 and 175 to 180. No transmembrane domain was identified. The 22 NH_2-terminal amino acids might form a cleavable signal peptide.

We amplified all six coding exons of *X25* with flanking primers and screened them for mutations in 184 FRDA patients (*25*). Three point mutations that introduce changes in the *X25* gene product were identified. The first change, in a French family with two affected sibs, consisted of a T→G transversion in exon 3 that changed a leucine codon (TTA) into a stop codon (TGA) (L106X). The second case, in a Spanish family with one affected member, was an A→G transition that disrupted the acceptor splice site at the end of the third intron, changing the invariant AG into a GG. Finally, a change from isoleucine to phenylalanine (I154F) was found in exon 4 in five patients from three Southern Italian families. This conservative change of a hydrophobic amino acid affects an invariant position within the highly conserved domain shared between human, worm, and yeast. In all three cases, affected individuals were heterozygous for the point mutation. The I154F mutation was also found in 1 out of 417 chromosomes from 210 control individuals from the same Southern Italian population, which is compatible with the possibility that this is a disease-causing mutation (*26*).

Southern blot analysis did not reveal any difference between FRDA patients and normal controls when DNAs digested with Msp I, Taq I, or Bst XI were hybridized with an *X25* cDNA probe, thereby excluding major rearrangements. However, hybridization of Eco RI–digested DNAs from FRDA patients revealed that the fragment con-

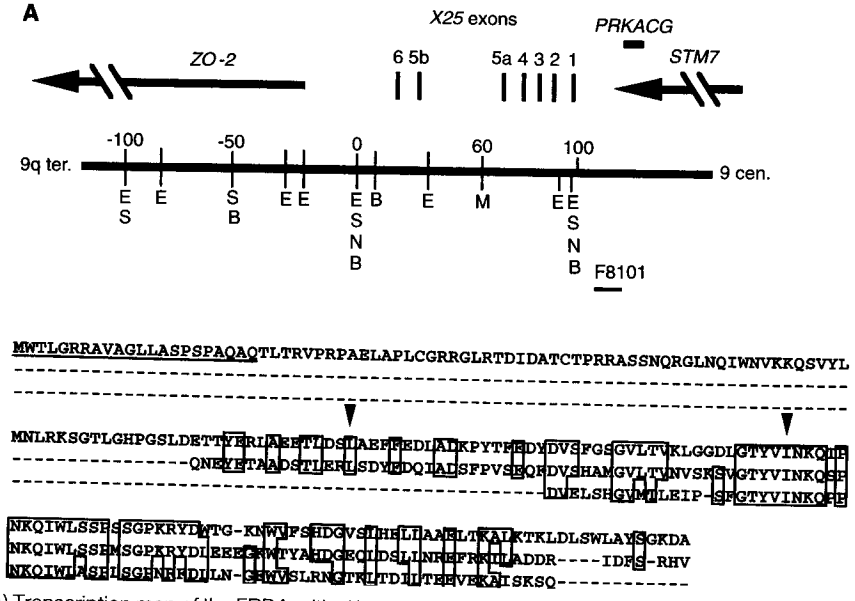

Fig. 1. (**A**) Transcription map of the FRDA critical interval. Distances are in kilobase pairs from the first Not I site upstream to the *ZO-2* gene. The critical FRDA region is between the F8101 marker and the *ZO-2* gene. B, Bss HI site; E, Eag I site; M, Mlu I site; N, Not I site; S, Sac II site; ter., telomeric; and cen., centromeric. (**B**) Alignment of the exon 5a–containing isoform of frataxin with translated ORFs contained within a *C. elegans* cosmid (CELT59G1) and a *S. cerevisiae* EST (T38910). Identical amino acids are boxed. The putative signal peptide is underlined. Amino acids involved by point mutations (L106X and I154F) are indicated by arrowheads. The exon 5b–containing isoform diverges at position 161, and its 11 COOH-terminal amino acids are RLTWLLWLFHP. Abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.

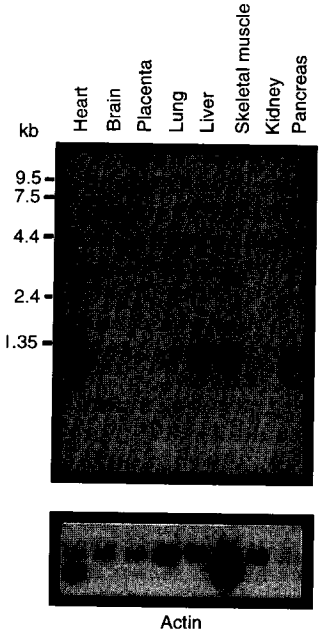

Fig. 2. Northern (RNA) blot analysis of *X25* transcripts. A ^{32}P-labeled 5′-RACE product containing exons 1 to 5b was hybridized to a multiple-tissue Northern blot (Clontech) containing 2 μg of poly(A)$^+$ RNA in each lane. The membrane was washed at 50°C with 0.1× standard saline citrate (SSC) and 0.1% SDS, then exposed to x-ray film at −70°C for 7 days. The lower panel shows a successive hybridization of the same blot with an actin probe (provided by the blot manufacturer).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

taining exon 1 was on average 2.5 kb larger than in normal control individuals, with no detectable normal band. FRDA carriers were heterozygous for an enlarged and a normal-sized fragment. The size of the enlarged fragment was clearly variable, even among FRDA carriers who were related (Fig. 3). The enlarged region was further localized to a 5.2-kb Eco RI–Not I fragment within the first intron of X25, which was subcloned from a cosmid and sequenced.

We designed oligonucleotide primers to amplify this fragment using a long-range PCR technique and confirmed its increase in size in FRDA patients (*27*). Cosmid sequence analysis revealed a $(GAA)_9$ repeat apparently derived from a poly(A) expansion of the canonical A_5TACA_6 sequence linking the two halves of an Alu repeat (Fig. 4). The $(GAA)_9$ repeat is located 1.4 kb downstream from exon 1, and restriction analysis of long-range PCR fragments from FRDA patients located the abnormal size increase within 100 bp from this triplet repeat. Digestion of the same fragments with Mbo II, whose recognition site is GAAGA, abolished size difference between patients and controls, indicating that the GAA repeat may be involved. Direct sequencing proved that the mutation consists of an almost pure GAA repeat expansion (Fig. 4). PCR primers were then designed to evaluate the presence and size of the GAA expanded repeat in FRDA patients and any variability of the repeat in normal individuals (*28*) (Fig. 5).

Seventy-nine unrelated FRDA patients with typical disease, including five known to carry X25 point mutations, were tested for the GAA expansion by Southern analysis, PCR, or both. The patients previously known to carry point mutations were all heterozygous for the expansion. Segregation analysis within families indicated that the point mutation and the GAA expansion had different parental origin (Fig. 6), demonstrating that the point mutations, including the conservative missense mutation I154F, are disease causing. Homozygosity for expanded alleles was demonstrated in 71 of the 74 patients without previously detected X25 point mutations, and heterozygosity in three. Overall, according to these data, the GAA expansion accounted for about 98% of the FRDA chromosomes.

The sizes of the enlarged alleles were found to vary between 200 and more than 900 GAA units, with most alleles containing 700 to 800 repeats. Instability of the expanded repeats during parent-offspring transmission was clearly demonstrated, both directly by analysis of parent-offspring pairs, and indirectly by the detection of two distinct alleles in affected children of consanguineous parents, who are expected to be homozygous by descent at the FRDA locus. PCR products corresponding to expanded repeats appeared as slightly blurred bands, suggesting the occurrence of only a limited degree of somatic mosaicism for different size repeats due to mitotic instability, at least in lymphocyte DNA (Fig. 5B). Seventy-seven normal individuals who were tested by Southern analysis were homozygous for a normal allele. PCR analysis of an additional 98 normal controls also did not show any expansion and revealed that the GAA repeat is polymorphic, its length varying from 7 to 22 units (Fig. 5A). Smaller alleles were more prevalent.

GAA repeats, up to 30 to 40 units, are

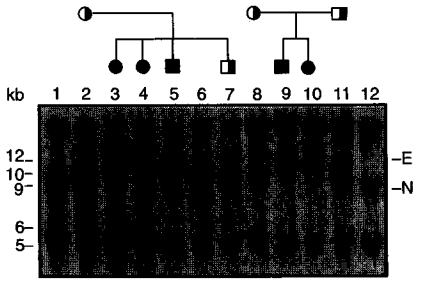

Fig. 3. Southern (DNA) blot analysis showing FRDA-associated expanded restriction fragments. Lanes 1 and 12, normal controls; lanes 2 through 7, individuals from a Saudi Arabian FRDA family; lanes 8 through 11, individuals from a Louisiana Acadian (Cajun) FRDA family. DNA from affected individuals are in lanes 3 through 5 and 9 and 10, and DNA from heterozygous carriers in lanes 2, 6 through 8, and 11. Individuals in lanes 2 and 6 are first cousins. Molecular size markers are indicated on the side. N and E indicate the normal and expanded alleles, respectively. The constant bands at the top correspond to exons 2 and 3 (15 kb), and those at the bottom to a related sequence outside of the FRDA region (5 kb). Genomic DNA (10 μg) from each individual were digested with Eco RI, run in a 0.6% agarose gel, and blotted onto a nylon membrane (Hybond +). The blot was hybridized with a ^{32}P-labeled *X25* cDNA probe. After a highest stringency wash with 0.1× SSC and 0.1% SDS for 5 min at 65°C, the blot was exposed to x-ray film at −70°C for 2 days.

Fig. 4. Cosmid DNA sequence containing the GAA repeat (bold, underlined) in the first *X25* intron. The first 29 bp and the last 30 bp correspond to the GAA-F and GAA-R primers, respectively (*28*).

```
  1  GGGATTGGTT GCCAGTGCTT AAAAGTTAGG ACTTAGAAAA TGGATTTCCT
 51  GGCAGGACGC GGTGGCTCAT GCCCATAATC TCAGCACTTT GGGAGGCCTA
101  GGAAGGTGGA TCACCTGAGG TCCGGAGTTC AAGACTAACC TGGCCAACAT
151  GGTGAAACCC AGTATCTACT AAAAAATACA AAAAAAAAAA AAAAAAAGAA
201  GAAGAAGAAG AAGAAGAAGA AGAAAATAAA GAAAAGTTAG CCGGGCGTGG
251  TGTCGCGCGC CTGTAATCCC AGCTACTCCA GAGGCTGCGG CAGGAGAATC
301  GCTTGAGCCC GGGAGGCAGA GGTTGCATTA AGCCAAGATC GCCCAATGCA
351  CTCCGGCCTG GGCGACAGAG CAAGCTCCGT CTCAAAAAAT AATAATAATA
401  AATAAAAATA AAAAATAAAA TGGATTTCCC AGCATCTCTG GAAAAATAGG
451  CAAGTGTGGC CATGATGGTC CTTAGATCAA GGACCATCAT GGCCACACTT
501  GCC
```

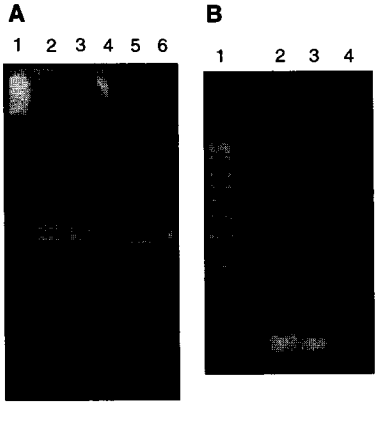

Fig. 5. (**A**) Example of PCR analysis of normal alleles of the GAA repeat. Lane 1 is the 1-kb ladder DNA size marker, and lanes 2 through 6 are alleles from normal control individuals previously identified to be heterozygous at the repeat. The GAA-F and GAA-R primers (*28*) were used for amplification. Fragments vary in size from 480 to 520 bp. (**B**) PCR amplification of the expanded GAA repeat in a FRDA carrier (lane 3) and in a patient (lane 4). Lane 1 is the 1-kb ladder DNA marker, and lane 2 is a normal control. The Bam and 2500F primers (*28*) were used for PCR. Expanded alleles have a slightly fuzzy appearance. Instability of the repeat is indicated by the presence of two distinct bands in the patient lane, although the patient is an offspring of consanguineous parents. Also, the carrier in lane 3 is the patient's mother, but the corresponding expanded allele does not exactly match in size any of the bands from her offspring.

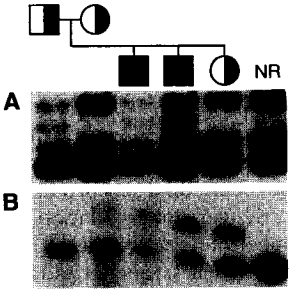

Fig. 6. Segregation of the L106X mutation and of the GAA expansion in a FRDA family. (**A**) The single-strand conformation polymorphism pattern indicates the paternal origin of the point mutation, and (**B**) Southern blot analysis indicates the maternal origin of the expansion. NR, unrelated normal control.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

common in many organisms and are sometimes polymorphic, as in the 3′ untranslated region of the rat polymeric immunoglobulin (Ig) receptor (*29*); they have not previously been associated with disease. A recently proposed theoretical model suggested that the ability to form a hairpin structure is crucial for the susceptibility of trinucleotide repeats to give rise to large expansions (*30*). According to this model, CAG/CTG or CGG/CCG repeats were predicted to be expansion prone, whereas the GAA/CCT repeat had the lowest propensity to expand, making the FRDA expansion an unexpected finding. A striking linkage disequilibrium between FRDA and a polymorphism in a newly identified exon of the *ZO-2* gene (about 120 kb telomeric to the expanded triplet repeat) in French and Spanish families (*31*) suggests a single origin for the FRDA expansion, but it is also compatible with a multistep or recurrent expansion on an allele at risk (*32*).

The fact that FRDA is autosomal recessive makes the natural history of the mutation at the population level strikingly different from any other known disease due to trinucleotide expansions. In fragile X and myotonic dystrophy, where expansions of comparable size occur in noncoding sequences, carriers have severe early-onset disease and a strong reproductive disadvantage. Large expansions in these diseases are newly formed from unstable alleles of intermediate sizes, resulting in the phenomenon of anticipation (*33*). In FRDA large expanded alleles are transmitted by asymptomatic carriers, and new expansion events in heterozygotes would go undetected at the phenotypic level. The absence of negative selection against heterozygotes plays the key role in maintaining the frequency of large FRDA expanded alleles as high as 1 per 250 chromosomes, at least one order of magnitude higher than any other characterized trinucleotide expansion. Conversely, deletions of CTG repeats in myotonic dystrophy with reversion to normal-sized alleles have been observed (*34*). In our sample of FRDA families, large expanded alleles were present in all tested asymptomatic carriers, and despite their size instability, neither new expansions deriving from an intermediate allele nor reversions to normality were detected. Although we cannot exclude the occasional occurrence of such events in the general population, given the large number of heterozygous individuals, it appears that their frequency is low enough not to introduce detectable distortions in the pattern of FRDA inheritance, particularly inconsistencies in linkage results.

FRDA patients appear to have shown either undetectable or extremely low mRNA amounts when compared with carriers and unrelated controls (*35*). This observation suggests that either an abnormality in RNA processing or an interference with the transcription machinery occurs as a consequence of the intronic GAA expansion. Patients with deleterious point mutations affecting *X25* clearly demonstrate that no other gene in the region, which could in principle be affected by a GAA expansion, is involved in the causation of FRDA. The restricted expression of *X25* in the sites of degeneration or malfunction distinguishes FRDA from the dominant ataxias and from ataxia telangiectasia, where expression of the causative gene is ubiquitous (*36*). A severely reduced *X25* mature mRNA is expected to result in a similarly low level of frataxin. Reduced frataxin in spinal cord, heart, and pancreas is probably the primary cause of neuronal degeneration, cardiomyopathy, and increased risk of diabetes. Functional studies on frataxin are likely to be facilitated by its evolutionary conservation and will hopefully generate new therapeutic possibilities.

REFERENCES AND NOTES

1. N. Friedreich, *Virchows Arch. Pathol. Anat.* **26**, 391 (1863).
2. H. Skre, *Clin. Genet.* **7**, 287 (1975); G. Romeo *et al.*, *Am. J. Hum. Genet.* **35**, 523 (1983).
3. G. Geoffroy *et al.*, *Can. J. Neurol. Sci.* **3**, 279 (1976); A. E. Harding, *Brain* **104**, 589 (1981); *The Hereditary Ataxias and Related Disorders* (Churchill Livingstone, Edinburgh, 1984).
4. A. E. Harding and R. L. Hewer, *Q. J. Med.* **208**, 489 (1983); B. Pentland and K. A. A. Fox, *J. Neurol. Neurosurg. Psychiatry* **46**, 1138 (1983).
5. G. Finocchiaro, G. Baio, P. Micossi, G. Pozza, S. Di Donato, *Neurology* **38**, 1292 (1988).
6. J. T. Hughes, B. Brownell, R. L. Hewer, *Brain* **91**, 803 (1968); R. L. Hewer, *Br. Med. J.* **3**, 649 (1968); J. B. Lamarche, B. Lemieux, H. B. Liu, *Can. J. Neurol. Sci.* **11**, 592 (1984); G. Said, M.-H. Marion, J. Selva, C. Jamet, *Neurology* **36**, 1292 (1986); S. Jitpimolmard *et al.*, *Acta Neuropathol.* **86**, 29 (1993).
7. T. Klockgether *et al.*, *J. Neurol. Neurosurg. Psychiatry* **53**, 297 (1990); T. A. Waldmann *et al.*, *Ann. Intern. Med.* **99**, 367 (1983).
8. S. Chamberlain *et al.*, *Nature* **334**, 248 (1988); R. Fujita *et al.*, *Genomics* **4**, 110 (1989); A. Hanauer *et al.*, *Am. J. Hum. Genet.* **46**, 133 (1990).
9. L. Montermini *et al.*, *Am. J. Hum. Genet.* **57**, 1061 (1995).
10. F. Duclos, F. Rodius, K. Wrogemann, J.-L. Mandel, M. Koenig, *Hum. Mol. Genet.* **3**, 909 (1994); M. Pandolfo *et al.*, *Biochem. Med. Metab. Biol.* **52**, 115 (1994).
11. J. Carvajal *et al.*, *Hum. Mol. Genet.* **4**, 1411 (1995).
12. Twelve cosmids spanning 120 kb of the critical FRDA interval plus 80 kb immediately proximal were individually subcloned as Bam HI–Bgl II fragments into pSPL1 and pSPL3 exon-trapping vectors and transfected into COS-7 (A6) cells for splicing of potential exons [D. M. Church *et al.*, *Nature Genet.* **6**, 98 (1994)]. The same cosmids were used for hybridization-selection from uncloned cDNAs synthesized from human cerebellum and spinal cord polyadenylated [poly(A)$^+$] RNA [J. G. Morgan *et al.*, *Nucleic Acids Res.* **20**, 5173 (1992)]. Finally, seven of the cosmids were subcloned as Sau 3AI, Apo I, and Hae III fragments, and ~1500 random single-pass sequences were generated, which were then analyzed with the GRAIL1a and GRAIL2 [E. C. Uberbacher and R. J. Mural, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 11261 (1991)] and FEXH [V. V. Solovyev, A. A. Salamov, C. B. Lawrence, *Nucleic Acids Res.* **22**, 5156 (1994)] programs. These three approaches yielded 19, 5, and 17 potential coding sequences, respectively, including two that matched known genes, namely the protein kinase A γ catalytic subunit gene (obtained by cDNA selection and random sequencing) and a mitochondrial adenylate kinase 3 pseudogene (obtained by random sequencing).
13. The 5′-RACE experiment was performed with the Clontech RACE-Ready cDNA kit, according to the manufacturer's instruction.
14. The EST (GenBank accession number R06470) represent 5′ sequence of cDNA clone 126314 determined by the Washington University–Merck EST Project.
15. The EST clone contained four exons, and the longer RACE product contained one additional 5′ exon. This exon mapped within the CpG island at position 100 on the genomic map. A transcription start site was predicted 388 bp upstream of the exon 1 donor splice site, and a TATA box could be found 28 bp further upstream by the TSSG program [V. V. Solovyev, A. A. Salamov, C. B. Lawrence, in preparation].
16. An in-frame ATG preceded 108 bp upstream by an in-frame stop codon was found in exon 1, and it was assumed to be the translation start site.
17. The 3′-RACE experiment was carried out as described [E. M. Frohman and J. B. Martin, *Technique* **1**, 165 (1989)], with total RNA from HeLa cells (2 mg) and nested primers in exon 5b.
18. This longer 3′-RACE product ended with the poly(A) tail of a downstream Alu sequence. The genomic sequence of exon 6 showed that it contains three Alu sequences in tandem, followed by a poly(A) signal 1050 bp away from the acceptor splice site. Exon 6 was mapped 13 kb telomeric to exon 5b (Fig. 1A). Splice sites of all seven exons (1 to 4, 5a, 5b, and 6) conform to the canonical consensus.
19. Further hybridizations of the Northern blot with exon 5a– and 5b–specific probes revealed that the 1.05- and 2.0-kb bands contained exon 5b, whereas sequences matching exon 5a were found in the 2.8- and 7.3-kb bands in addition to the major 1.3-kb band.
20. V. Campuzano *et al.*, data not shown.
21. Three of the cDNAs, which are identical in the portion that has been sequenced so far, extend respectively for 0.5, 1, and 2 kb upstream of exon 1. Their sequence presents numerous divergences from *X25* in the part corresponding to exon 1, mostly CpG dinucleotides changed to TG or CA, then the sequences are almost identical in the part corresponding to exons 2 to 4. An additional 1.6-kb cDNA begins with a sequence closely matching exon 5a, even in its untranslated region, with only occasional single-base changes and short insertions or deletions [L. Montermini *et al.*, in preparation].
22. Southern blot analysis with a cDNA probe of *X25* exons 1 to 5a revealed a prominent 5-kb Eco RI band in genomic DNA that did not correspond to any exon and was absent in yeast artificial chromosome and cosmid DNA from the critical FRDA region. Several additional bands, also absent from cloned DNA from the FRDA region, appeared when blots were washed at lower stringency (1× SSC at room temperature). The primers nF2 (5′-TCCCGCGGCCGGCAGAGTT-3′) and E2R (5′-CCAAAGTTCCAGATTTCCTCA-3′), which can amplify a 173-bp fragment spanning exons 1 and 2 of the *X25* cDNA, generated a PCR product of corresponding size from genomic DNA, but not from cloned DNA from the FRDA region, indicating the presence of sequences with high similarity to a processed *X25* transcript elsewhere in the genome.
23. S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, *J. Mol. Biol.* **215**, 403 (1990).
24. Secondary structure prediction was performed with the SSP and NNSSP programs, which are designed to locate secondary structure elements [V. V. Solovyev and A. A. Salamov, *CABIOS* **10**, 661 (1994)]. The TMpred program was used to predict putative transmembrane domains [K. Hoffmann and W. Stoffel, *Biol. Chem. Hoppe-Seyler* **374**, 166 (1993)]. PSORT was used to predict possible protein sorting signals [K. Nakai and M. Kanehisa, *Proteins Struct. Funct. Genet.* **11**, 95 (1991)].
25. The following intronic primers were used to amplify the *X25* exons: exon 1 (240 bp), F: 5′-AGCAC-

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

CCAGCGCTGGAGG-3′, R: 5′-CCGCGGCTGTTC-CCGG-3′; exon 2 (168 bp), F: 5′-AGTAACGTACT-TCTTAACTTTGGC-3′, R: 5′-AGAGGAAGATACC-TATCACGTG-3′; exon 3 (227 bp), F: 5′-AAAATG-GAAGCATTTGGTAATCA-3′, R: 5′-AGTGAACTAA-AATTCTTAGAGGG-3′; exon 4 (250 bp), F: 5′-AAG-CAATGATGACAAAGTGCTAAC-3′, R: 5′-TGGTC-CACAATGTCACATTTCGG-3′; exon 5a (223 bp), F: 5′-CTGAAGGGCTGTGCTGTGGA-3′, R: 5′-TGTC-CTTACAAACGGGGCT-3′; and exon 5b (224 bp), F: 5′-CCCATGCTCAAGACATACTCC-3′, R: 5′-ACA-GTAAGGAAAAAACAAACAGCC-3′. Amplifications for exons 2, 3, 4, 5a, and 5b consisted of 30 cycles with the following parameters: 1 min at 94°C, 1 min at 55°C, and 1 min at 72°C. To amplify the highly GC-rich exon 1, we raised the annealing temperature to 68°C and 10% dimethyl sulfoxide was added to the reaction. The search for mutations was conducted with single-strand conformation polymorphism analysis [M. Orita, Y. Suzuki, T. Sekiya, K. Hayashi, *Genomics* **5**, 874 (1989)] in 168 FRDA patients, and chemical cleavage [J. A. Saleeba, S. J. Ramus, R. J. H. Cotton, *Hum. Mutat.* **1**, 63 (1992)] in 16 patients.

26. Assuming a FRDA carrier frequency in Italy of 1 out of 120 individuals (*2*) and a frequency of I154F of 1 out of 40 FRDA chromosomes in Southern Italians, one individual in 3300 in that population is expected to be a carrier of I154F. Finding such a person in a random sample of 210 individuals can occur with >6% probability.
27. The Perkin-Elmer XL long-PCR reagent kit was used to set up the reactions, and standard conditions were used as suggested by the manufacturer with primers 5200Eco (5′-GGGCTGGCAGATTCCTCCA-G-3′) and 5200Not (5′-GTAAGTATCCGCGCCGG-GAAC-3′). Amplifications were performed in a Perkin-Elmer 9600 machine and consisted of 20 cycles of the following steps: 94°C for 20 s, 68°C for 8 min, followed by 17 cycles in which the length of the 68°C-step was increased by 15 s per cycle. The generated amplification product is 5 kb from normal chromosomes, and about 7.5 kb from FRDA chromosomes.
28. The primers GAA-F (5′-GGGATTGGTTGCCAGT-GCTTAAAAGTTAG-3′) and GAA-R (5′-GATCTAAG-GACCATCATGGCCACACTTGCC-3′) flank the GAA repeat and generate a PCR product of 457 + 3*n* bp (*n* = number of GAA triplets). With these primers, efficient amplification of normal alleles could be obtained by using the traditional PCR procedure with Taq polymerase after 30 cycles consisting of the following steps: 94°C for 45 s, 68°C for 30 s, and 72°C for 2 min. Enlarged alleles were much less efficiently amplified, particularly when present together with a normal allele; therefore, use of these primers is not indicated for FRDA carrier detection. A more efficient amplification of expanded alleles, also in FRDA carriers, could be obtained by using the primers Bam (5′-GGAGGGATC-CGTCTGGGCAAAGG-3′) and 2500F (5′-CAATCCAG-GACAGTCAGGGCTTT-3′). These primers generated a 1.5-kb normal fragment. Amplification was conducted with the long PCR protocol, in 20 cycles composed of the following steps: 94°C for 20 s, 68°C for 2 min and 30 s, followed by 17 cycles in which the length of the 68°C step was increased by 15 s per cycle.
29. K. S. Koch *et al.*, *Nucleic Acids Res.* **23**, 1098 (1995).
30. A. M. Gacy, G. Goellner, N. Juranic, S. Macura, C. T. McMurray, *Cell* **81**, 533 (1995); X. Chen *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 5199 (1995).
31. E. Monros *et al.*, in preparation. The polymorphism has two alleles, one of which is present in 84% of FRDA chromosomes and in 23% of normal chromosomes.
32. See G. Imbert *et al.*, *Nature Genet.* **4**, 72 (1993), where the absolute linkage disequilibrium in myotonic dystrophy is explained by recurrent mutations on such an at-risk allele.
33. G. Bates and H. Lehrach, *Bioassays* **16**, 277 (1994); J.-L. Mandel, *Nature Genet.* **7**, 453 (1994); E. R. Fearon *et al.*, *Science* **247**, 49 (1990).
34. P. Shelbourne *et al.*, *Hum. Mol. Genet.* **1**, 467 (1992); K. L. O'Hoy *et al.*, *Science* **259**, 809 (1993).
35. S. I. Bidichandani *et al.*, in preparation.
36. S. Banfi *et al.*, *Nature Genet.* **7**, 513 (1994); Y. Kawaguchi *et al.*, *ibid.* **8**, 221 (1994); Y. Trottier *et al.*, *Nature* **378**, 403 (1995); K. Savitsky *et al.*, *Science* **268**, 1749 (1995).

37. This work is dedicated to the memory of Dr. A. E. Harding. Supported by grants from NIH (NS34192, M.P. and P.I.P.); from the Muscular Dystrophy Association (M.P.); from the Association Francaise contre les Myopathies, the Groupement de Recherche et d'Etudes sur les Genomes, CNRS, and INSERM (M.K.); from the Italian Telethon (S.C.); from the Comision Interministerial de Ciencia y Tecnologia; and from the Generalitat Valenciana (F.P.). V.C. is a recipient of a fellowship from Ministerio de Educacion y Ciencia, Spain, M.D.M. is a recipient of a fellowship from the Generalitat Valenciana and from the University of Valencia, L.M. is a recipient of a fellowship from the Italian Telethon, M.C. is a recipient of a fellowship from the Fondation pour la Recherche Médicale, F.D. is a recipient of a fellowship from the Association Française contre l'Ataxie de Friedreich (AFAF), E.M. and J.C. are recipients of fellowships from the Generalitat Valenciana, and S.B. is a recipient of a fellowship from the Muscular Dystrophy Association. We wish to thank S. Vicaire (IGBMC) and P. Milasseau and C. Crueau (Généthon, Evry) for large-scale sequencing. We thank the Baylor College of Medicine Human Genome Center for access to DNA analysis resources and databases. We are grateful to the many clinicians who referred their patients to us, and to all patients and their families for their willingness to participate in this research and for their continuous encouragement.

18 December 1995; accepted 28 January 1996

Egr-1–Induced Endothelial Gene Expression: A Common Theme in Vascular Injury

Levon M. Khachigian,* Volkhard Lindner,† Amy J. Williams, Tucker Collins‡

A number of pathophysiologically relevant genes, including platelet-derived growth factor B-chain (PDGF-B), are induced in the vasculature after acute mechanical injury. In rat aorta, the activated expression of these genes was preceded by a marked increase in the amount of the early-growth-response gene product Egr-1 at the endothelial wound edge. Egr-1 interacts with a novel element in the proximal PDGF-B promoter, as well as with consensus elements in the promoters of other genes induced by endothelial injury. This interaction is crucial for injury-induced PDGF-B promoter–dependent expression. Sp1, whose binding site in the PDGF-B promoter overlaps that of Egr-1, occupies this element in unstimulated cells and is displaced by increasing amounts of Egr-1. These findings implicate Egr-1 in the up-regulated expression of PDGF-B and other potent mediators in mechanically injured arterial endothelial cells.

Vascular endothelium constitutes a nonthrombogenic surface of normally quiescent cells that line blood vessels and regulate molecular and cellular movement across the vessel wall. In response to denuding injury, endothelial cells at the wound edge spread and migrate into the vacant area, undergo proliferation, and secrete factors that stimulate endothelial and smooth muscle cell growth. These responses provide an important homeostatic mechanism for maintaining normal vascular function. PDGF has been implicated in the regenerative events that follow vascular injury. The induction of PDGF expression in vascular endothelium may have profound chemotactic and mitogenic effects on the underlying smooth muscle cells and may contribute to the structural remodeling that typically occurs in experimental arterial repair, in restenosis, and in the pathogenesis of atherosclerotic vascular disease (*1*). Despite a wealth of descriptive studies that correlate the formation of vascular occlusive lesions with the inappropriate expression of PDGF and other growth-regulatory molecules (*2*, *3*), a direct link between a transcription factor and the induced expression of a pathophysiologically relevant gene has not yet been demonstrated in the context of arterial injury.

In a survey of immediate-early genes that could be induced by acute vascular injury in rat aorta (*4*, *5*), we examined the expression of Egr-1, a serum-inducible zinc-finger nuclear phosphoprotein and member of a family of related transcription factors (*6*). In situ hybridization techniques that visualize the endothelium of the vessel wall en face revealed that Egr-1 expression was markedly induced exclusively at the endothelial wound edge within 30 min of partial denudation (Fig. 1B). Egr-1 expression was undetectable in endothelium from unmanipulated arteries (Fig. 1A). Induced Egr-1 mRNA remained apparent after 2 hours (Fig. 1D), and the time-dependent decrease in the specific hybridization signal (Fig. 1, B to D) demonstrates the transient induction of endothelial Egr-1 expression by injury. In contrast, the

L. M. Khachigian, A. J. Williams, T. Collins, Vascular Research Division, Department of Pathology, Brigham and Women's Hospital, 221 Longwood Avenue, Boston, MA 02115, USA.
V. Lindner, Department of Pathology, University of Washington, Seattle, WA 98195, USA.

*Present address: Centre for Thrombosis and Vascular Research, School of Pathology, University of New South Wales, Sydney, NSW 2052, Australia.
†Present address: Department of Surgery, Maine Medical Center Research Institute, South Portland, ME 04106, USA.
‡To whom correspondence should be addressed.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.