Cell, Vol. 68, 799–808, February 21, 1992, Copyright © 1992 by Cell Press

Molecular Basis of Myotonic Dystrophy: Expansion of a Trinucleotide (CTG) Repeat at the 3′ End of a Transcript Encoding a Protein Kinase Family Member

J. David Brook,*† Mila E. McCurrach,* Helen G. Harley,† Alan J. Buckler,* Deanna Church,* Hiroyuki Aburatani,* Kent Hunter,* Vincent P. Stanton,* Jean-Paul Thirion,* Thomas Hudson,* Robert Sohn,* Boris Zemelman,* Russell G. Snell,† Shelley A. Rundle,† Steve Crow,† June Davies,‡ Peggy Shelbourne,‡ Jessica Buxton,‡ Clare Jones,‡ Vesa Juvonen,‡ Keith Johnson,‡ Peter S. Harper,† Duncan J. Shaw,† and David E. Housman*

*Center for Cancer Research
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139
†Institute of Medical Genetics
University of Wales College of Medicine
Heath Park
Cardiff CF4 4XN
Wales
‡Department of Anatomy
Charing Cross and Westminster Medical School
London W6 8RF
England

Summary

Using positional cloning strategies, we have identified a CTG triplet repeat that undergoes expansion in myotonic dystrophy patients. This sequence is highly variable in the normal population. PCR analysis of the interval containing this repeat indicates that unaffected individuals have between 5 and 27 copies. Myotonic dystrophy patients who are minimally affected have at least 50 repeats, while more severely affected patients have expansion of the repeat containing segment up to several kilobase pairs. The CTG repeat is transcribed and is located in the 3′ untranslated region of an mRNA that is expressed in tissues affected by myotonic dystrophy. This mRNA encodes a polypeptide that is a member of the protein kinase family.

Introduction

Myotonic dystrophy (DM) is an autosomal dominant neuromuscular disease with an estimated minimum incidence of 1 in 8000 (Harper, 1989). It is the most common form of muscular dystrophy affecting adults. The clinical picture in DM is well established but exceptionally variable (Harper, 1989). Although generally considered a disease of muscle, with myotonia, progressive weakness and wasting, DM is characterized by abnormalities to a variety of other systems. DM patients often suffer from cardiac conduction defects, smooth muscle involvement, hypersomnia, cataracts, abnormal glucose response, and, in males, premature balding and testicular atrophy (Harper, 1989). One of the striking features of this disorder is the variability of phenotype both within and between families. For clinical purposes, patients are often subdivided into three groups according to the clinical syndrome and age at onset of the disorder (Harper and Dyken, 1972). The mildest form, which is occasionally difficult to diagnose, is seen in middle or old age and is characterized by cataract with little or no muscle involvement. The classical form, showing myotonia and muscle weakness, most frequently has onset in early adult life and adolescence. The most severe form, which occurs congenitally, is associated with generalized muscular hypoplasia, mental retardation, and high neonatal mortality. Those congenitally affected offspring surviving the neonatal period invariably exhibit the classical form of the disease in late childhood or adolescence. The congenital form of DM is exclusively maternally transmitted. The phenomenon of anticipation (Howeler et al., 1989), in which the disease symptoms become more severe and age at onset earlier in successive generations, is often most strikingly manifested in a family producing a congenitally affected child.

Biochemical studies have failed to identify the defective protein in myotonic dystrophy, although several have implicated defects in membrane structure and function (Harper, 1989). Abnormalities in calcium transport (Seiler and Kuhn, 1970), membrane fluidity (Butterfield et al., 1974), sodium–potassium ATPase stoichiometry (Hull and Roses, 1976), and apamin receptor expression (Renaud et al., 1986) have all been reported for DM. Evidence of reduced phosphorylation of membrane proteins in both red blood cells (Roses and Appel, 1973) and sarcolemmal membranes from muscle biopsies of patients (Roses and Appel, 1974) has led to the suggestion by these authors that a defect in a protein kinase may be significant in DM.

In order to identify the gene responsible for this disorder, we have employed the techniques of positional cloning. Over the past few years, we and others have produced genetic (Brunner et al., 1989; Johnson et al., 1990; Harley et al., 1991; Tsilfidis et al., 1991) and physical (Korneluk et al., 1989; Smeets et al., 1990; Brook et al., 1991a, 1991b) maps of the long arm of human chromosome 19 to localize the DM gene to band 19q13.3 between DNA markers ERCC1 and D19S51. Construction and analysis of radiation-reduced hybrids (Brook et al., 1992) and YAC, cosmid, and phage libraries (Buxton et al., 1992; Jansen et al., 1992; Shutler et al., 1992) has allowed saturation mapping of the interval between these two markers. The DM region could be narrowed down further through linkage disequilibrium studies. Markers D19S63 and D19S95 are in strong linkage disequilibrium with DM (Harley et al., 1991, 1992). No disequilibrium was observed for markers flanking these loci. Screening of phage libraries derived from a radiation-reduced hybrid, 2F5 (Brook et al., 1992), produced a series of overlapping phage clones that spanned the interval between D19S63 and D19S95.

This intensive search recently resulted in the identification of DNA markers adjacent to D19S95 that detect patient specific bands on Southern blots (Harley et al., 1992; Buxton et al., 1992; Aslanidis et al., 1992). Probes pBB0.7 (Harley et al., 1992) and cDNA 25 (Buxton et al., 1992)

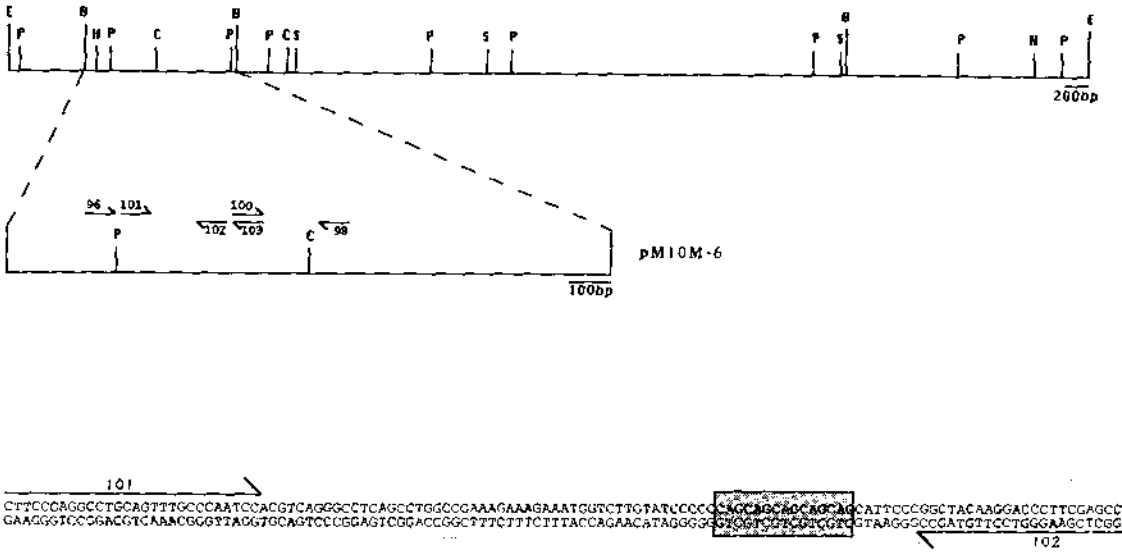


Figure 1. Restriction Map of the 10 kb EcoRI Fragment That Undergoes Expansion in Myotonic Dystrophy Patients

This map was derived from overlapping genomic phage clones from unaffected individuals. Restriction sites for BamHI (B), EcoRI (E), HincII (C), HindIII (H), PstI (P), and SacI (S) are indicated. The subcloned 1.4 kb BamHI fragment (pM10M-6) is shown enlarged with the PstI and HincII sites, which flank the expanded region. The positions of PCR primers 96, 98, 100, 101, 102, and 103 and the sequence between primers 101 and 102 is shown.

identify the same EcoRI restriction fragment length polymorphism in the normal population, with alleles of 9 kb or 10 kb. In 43 of 53 unrelated affected individuals reported in these two studies, only one of the normal-sized alleles is present, plus an additional, larger, disease-specific band. This restriction fragment varies in length between patients, even between siblings within the same family. Furthermore, the size of the variable fragment increases in successive generations and shows a correlation between increased severity and earlier onset of the disease. The largest fragment detected is 15 kb, an increase of 5 kb over the normal size (Harley et al., 1992). In many individuals with fragments larger than 11 kb, a diffuse hybridization signal corresponding to a DNA fragment size greater than 11 kb is observed in a gel that otherwise gives tight DNA banding patterns, indicating that somatic mosaicism with respect to the precise extent of the increase in size of the DNA sequence within the EcoRI fragment has occurred. This situation is strikingly similar to that reported recently for the fragile-X mental retardation syndrome, where variation in length of a CGG repeat results in genetic instability (Dietrich et al., 1991; Fu et al., 1991; Kremer et al., 1991; Oberle et al., 1991; Verkerk et al., 1991; Yu et al., 1991).

In order to identify the mutation in DM, we have characterized genomic clones spanning the 10 kb fragment that is increased in size in this disease. We have identified a CTG repeat sequence that is highly polymorphic in the normal population and that undergoes remarkable expansion in DM patients. The modal number of repeats found in 282 normal alleles surveyed is 5 (48%), with the largest being 27. Minimally affected DM patients have at least 50 copies. The CTG repeat is transcribed and is found at a position 500 bp from the poly(A) tract of an mRNA expressed in many of the tissues affected in DM. The RNA in which the repeat resides encodes a polypeptide with strong amino acid homology to members of the protein kinase gene family.

Results

Fine Mapping of the Region Amplified in DM Patients

A series of phage clones derived from libraries of radiation-reduced hybrid 2F5 (Brook et al., 1992) have been used to span the interval between D19S63 and D19S95, the loci in linkage disequilibrium with DM. Intensive screening of this interval led to the identification of clones λM10M (detected by hybridization with cDNA 25 [Buxton et al., 1992]) and λM8L and λSM2 (which contain clone pBB0.7 [Harley et al., 1992]). These clones span the 10 kb EcoRI fragment that is increased in size in DM patients. Figure 1 shows a detailed restriction map of this interval. Sites for BamHI, HincII, HindIII, PstI, and SacI are marked. Single-copy probes mapping within this EcoRI fragment were hybridized to DNA from patient and normal individuals digested with a series of restriction enzymes. Figure 2 shows the result of one such hybridization. The normal control is heterozygous for an EcoRI RFLP. As previously reported, a patient-specific EcoRI band, larger in size than either normal allele, is observed in the DNA of the DM patient (Harley et al., 1992; Buxton et al., 1992). Digestion with BamHI, PstI, and SacI each revealed patient-specific bands. The smallest region containing the expanded sequence that could be established by hybridization was a 475 bp PstI–HincII fragment contained within the 1.4 kb BamHI fragment (pM10M-6) shown in Figure 1. Part of this fragment was sequenced to make primers for polymerase chain reaction (PCR) analysis. Amplification by PCR be-

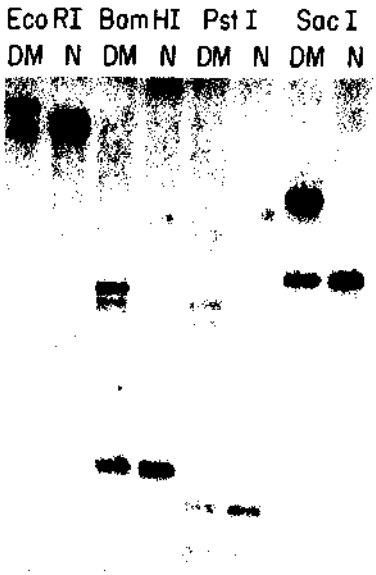


Figure 2. Southern Blot Analysis of the DNA Fragment Expanded in DM

DNA from a DM patient (DM) and an unaffected individual (N) digested with EcoRI, BamHI, PstI, and SacI and hybridized with pM10M-6 shows an increased band in the DM lane with all enzymes.



Figure 3. Variability of Repeat Length in the Normal Population

The size of labeled products derived by PCR amplification using oligos 101 and 102 was analyzed on sequencing gels. The smallest and most commonly seen allele corresponds to 5 repeats, with the majority being in the 10–16 range. Four heterozygous individuals are indicated in lanes A (13–21), J (5–21), S(5–24), and W (5–16). Lane O shows an individual homozygous for the smallest band, corresponding to 5 repeats.

tween oligos 98 and 100 produced single bands that were identical in size in patient, normal, and pM10M-6 lanes when visualized on ethidium-stained agarose gels (data not shown). In contrast, PCR using oligos 96 and 103 produced two bands in a normal human sample, a single band in DM DNA, and a single band in pM10M-6 that was smaller than any of the other bands. Analysis of the sequence derived from pM10M-6 between oligos 96 and 103 revealed tandem repeats of the trinucleotide CTG. Two other oligos, 101 and 102, which more closely flank this triplet repeat, were tested using PCR. These produced similar but more striking band size differences than with oligos 96 and 103 because of the smaller PCR product. The sequence of pM10M-6 between oligos 101 and 102 is shown in Figure 1.

Variability of the CTG Repeat in the Normal Population

In order to examine length variability of the PCR fragment produced with oligos 101 and 102 in the normal population, radiolabeled products obtained from a series of normal individuals were analyzed on sequencing gels. Figure 3 shows the results of one such experiment. The majority of individuals are heterozygous at this locus. "A" has a genotype (13, 21); "J", (5, 21); "S", (5, 24); and "W", (5, 16). Family studies show that the pattern illustrated by "O" is indicative of homozygosity for $(CTG)_5$ in normal individuals. Interestingly, shadow bands occurred in positions indicating that they differed by three bases from the major bands.

To confirm that this variability of length is due to different numbers of the triplet repeat, we cloned and sequenced the PCR products from two normal heterozygous individuals. From one individual, six clones were analyzed; three contained 12 copies of the CTG repeat, two contained 17 repeats, and one had 18 repeats (the variation between clones with 17 and 18 repeats may be an artifact of stuttering during PCR). In the second individual, two clones were sequenced and contained 5 and 11 copies of the repeat. These repeat lengths are consistent with the size of bands from these individuals determined by analysis of labeled PCR products on sequencing gels. In all clones analyzed, the sequence flanking the repeat was identical to that derived from pM10M-6, with the exception of two point mutations, which differed between clones and which may be due to errors by Taq polymerase.

The variation in repeat length was analyzed in 282 individuals, and the distribution of repeats in the normal population is shown in Figure 4. Alleles were sized against sequenced standards and sequencing ladders. Over 40% of alleles analyzed had 5 repeats. No alleles were observed with 6 or 8 triplets, and only single examples of 7 and 9 were found. The majority of repeats were in the 10–16 interval, with the highest being 27. Normal Mendelian inheritance was demonstrated for 20 meioses in two CEPH pedigrees (data not shown).

Variability of the CTG Repeat in DM Patients

During our initial attempts to characterize the variable region in patients, we observed that the DM sample tested by PCR assay produced a single band. To establish whether the single band was specific to this sample or whether it reflected something common to all patients, we analyzed by PCR 12 unrelated DM samples. PCR products were visualized on ethidium-stained gels. In all cases a single band from the normal size range (<200 bp) was present. In one sample (DMH9) a faint diffused band was observed above 360 bp on ethidium-stained gel. This patient was not severely affected and did not show signs of the disorder until age 45. In order to determine the nature of the larger fragment in patient DMH9, we cloned and sequenced the PCR product from this individual. In excess

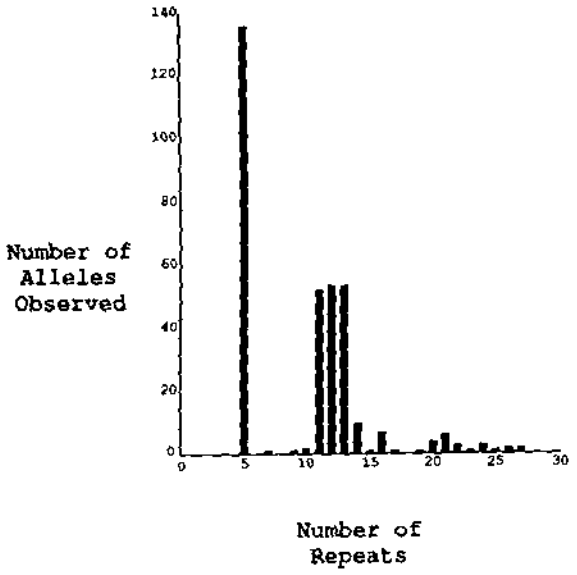


Figure 4. Distribution of Alleles within the Normal Population

The X axis represents the number of CTG repeats per chromosome, and the Y axis represents the number of alleles observed with this repeat length. The total number of chromosomes examined was 282.

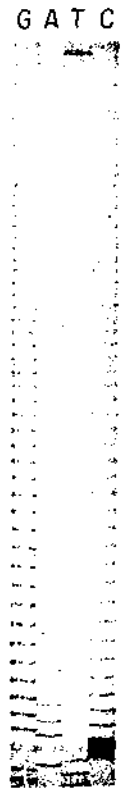


Figure 6. Amplification of the Repeat in a Mildly Affected DM Patient

The upper allele from a DM patient was cloned and sequenced. Fifty-two copies of the trinucleotide repeat are present in this clone.

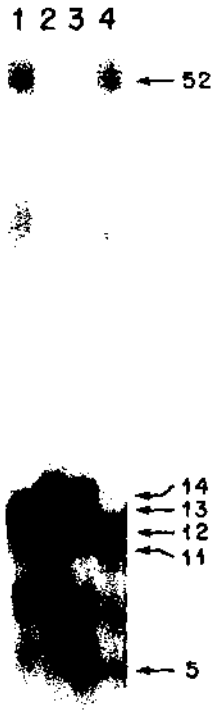


Figure 5. Visualization of a Second Allele in Two Mildly Affected Patients

Labeled PCR reactions were run on sequencing gels from two sets of grandparents. From each pair, one of the grandparents had a second band corresponding to 52 repeats. In each case the number of repeats were: lane 1, 13 and 52; lane 2, 11 and 14; lane 3, 5 and 12; lane 4, 12 and 52.

of 80 CTG repeats were present. We were unable to read beyond the CTG triplet; however, sequencing from either side of the repeat revealed that the flanking DNA was intact and had the same sequence as pM10M-6.

A single band was observed in PCR analysis of DNA from 11 of 12 DM patients, who had increases in Southern blot bands of 2 kb or more. The 12th patient, an individual who had later onset of the disorder, produced a faint upper band on PCR analysis. This suggested to us that in the other 11 patients we had been unable to amplify their second allele because of the extent of the amplified CTG repeat. In order to identify additional individuals likely to contain expanded CTG repeats that could be successfully identified by PCR, we focused on other mildly affected individuals. Three pairs of grandparents were selected from DM pedigrees in which it was difficult to decide, on clinical grounds, which was the affected individual. In each case, one of the pair had cataracts. PCR products analyzed on ethidium-stained gels showed that one individual from each pair had two alleles in the normal range, whereas the other grandparent (in each case, the one with cataracts) had one band in the normal range in addition to a second diffuse band at about 250 bp. For accurate sizing of this larger band, we analyzed labeled PCR products on sequencing gels. Figure 5 shows the size of these alleles in two pairs of grandparents. Though unrelated, these two

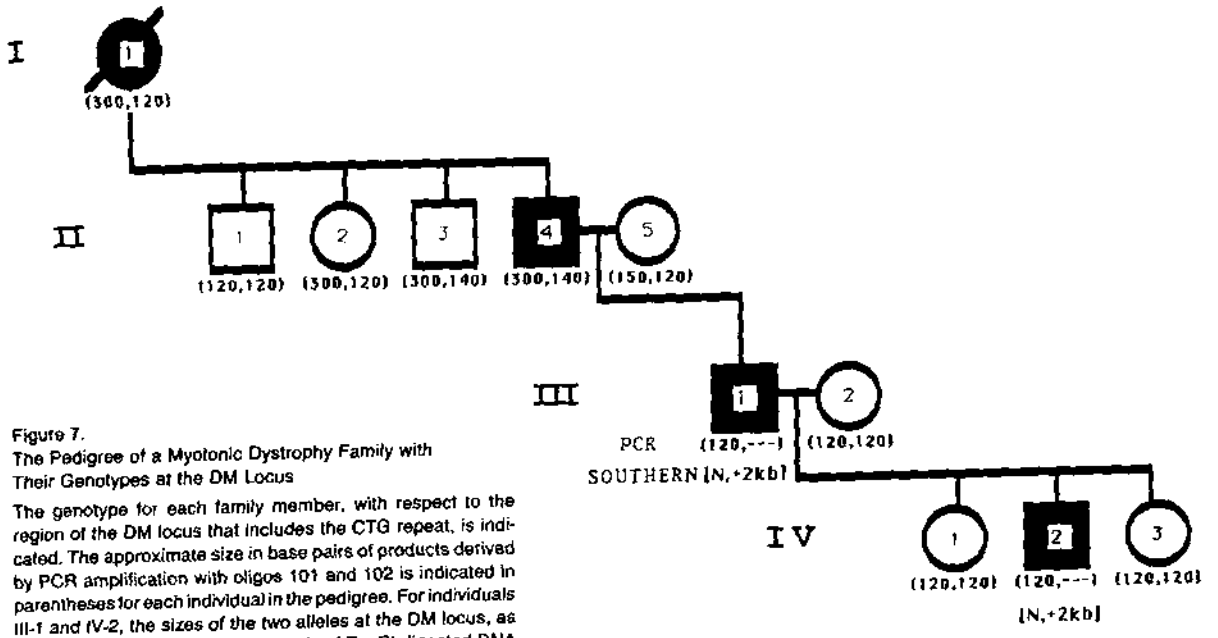


Figure 7.
The Pedigree of a Myotonic Dystrophy Family with Their Genotypes at the DM Locus

The genotype for each family member, with respect to the region of the DM locus that includes the CTG repeat, is indicated. The approximate size in base pairs of products derived by PCR amplification with oligos 101 and 102 is indicated in parentheses for each individual in the pedigree. For individuals III-1 and IV-2, the sizes of the two alleles at the DM locus, as determined by Southern blot analysis of EcoRI-digested DNA hybridized to probe p8B0.7, are indicated in brackets as follows: N indicates an allele equal in size to the normal population. +2kb indicates an allele approximately 2 kb greater in size than the larger allele in the normal population. The clinical diagnoses of the members of this family were as follows. Individual I-1 had few symptoms other than cataracts. II-1, II-2, and II-3 showed no clinical symptoms associated with DM. Diagnostic work on II-2 and II-3 was last carried out when they were age 61 and 64, respectively. II-4, like his mother (I-1), showed few symptoms other than cataracts. III-1 had classical adult onset. This patient shows extensive clinical symptoms, including myotonia of grip, marked facial and muscle weakness, and lens opacities. IV-2 had earlier onset than his affected father (III-1), exhibiting clear clinical signs of illness in his teenage years. II-5, III-2, IV-1, and IV-3 are clinically normal.

individuals had very similar larger bands (corresponding to 50–55 repeats). In order to determine whether the increased fragment size in these patients is due entirely to the expansion of the CTG repeat, we cloned and sequenced the larger allele from each patient following PCR amplification. Six clones were analyzed from patient DMH6 (shown in Figure 5): four contained 52 repeats, one had 54 repeats, and one had 57 repeats. Sequence from one such clone is shown in Figure 6. Analysis of five clones from patient DMH1 revealed two with 50 repeats, two with 52 repeats, and one with 61 repeats. In all clones the sequence of the DNA flanking the repeat was the same as that derived from pM10M-6. The only difference was in the length of the triplet repeat. The different CTG lengths observed in different clones from the same individual may reflect somatic mosaicism or stuttering during the PCR reaction.

A DM family that shows increased severity of disease in successive generations is shown in Figure 7. A band of approximately 300 bp is observed in the PCR product of four family members in the first two generations shown in the pedigree. This result demonstrates that a CTG repeat unit of approximately 60 repeats can be transmitted from one generation to the next without an obligatory expansion. Individuals in the first two generations had mild symptoms such as cataracts or were apparently asymptomatic. However, individuals in the third and fourth generations of the family showed severe symptoms of DM. These individuals showed only a single PCR band within the normal range. Southern blotting analysis confirmed that, for these individuals, as for most other severely affected DM patients, a dramatic expansion in allele size has occurred to increase the repeat size beyond that which can be detected by PCR analysis.

The CTG Repeat Is Transcribed

The triplet repeat sequences amplified in fragile-X syndrome and X-linked spinal and bulbar muscular atrophy (SBMA) are expressed in mRNAs. To test whether the CTG repeat amplified in DM might be within a gene, we first hybridized clone pM10M-6 to a Southern blot of DNA from different species. Figure 8 shows that DNA sequences contained within this clone are strongly conserved, suggestive of a transcribed sequence. pM10M-6 was used to screen cDNA libraries derived from several different sources including adult frontal cortex, substantia nigra, fetal muscle and fetal brain. A total of 10^6 clones were screened from these libraries and 110 positives were identified. Twenty were purified and six, from different libraries, were selected for further analysis. These clones were designated C28, C31, C34, C35, C39, and C85. The insert size of each clone and the library from which it was derived is C28 (frontal cortex), 2.5 kb; C31 (frontal cortex), 2.1 kb; C34 (substantia nigra), 1.7 kb; C35 (fetal brain), 1.7 kb; C39 (frontal cortex), 2.7 kb; C85 (fetal muscle), 2.8 kb. All six clones were hybridized to a panel of hybrid cell lines to confirm that they were from the expected region of chromosome 19. They were also mapped to filters of digested genomic phage clones λM10M and λM8L (which span the 10 kb EcoRI fragment amplified in patients) to determine how much genomic DNA they cover. cDNAs C28 and C85 each span at least 10 kb of genomic DNA.

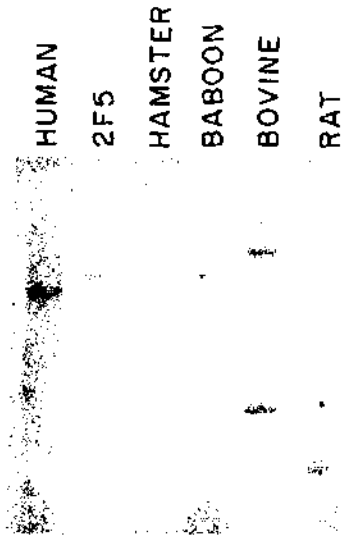


Figure 8. Clone pM10M-6, Containing the CTG Repeat, Shows Cross-Species Sequence Conservation

Southern blot analysis of HindIII-digested DNA from human, hamster, baboon, bovine, and rat hybridized with pM10M-6 reveals cross-species sequence conservation. 2F5 is a human–hamster hybrid cell line containing a small amount of human chromosome 19, which reveals a human band 1 kb higher than the band present in the human lane. This represents the upper allele of a HindIII polymorphism. The sample used in the human lane is homozygous for the lower allele.

Clone C39 was chimeric at the 5′ end, whereas the others mapped as expected. Clones C28, C34, and C35 were completely sequenced and clone C85 was partially sequenced. All clones contained the CTG repeat, and this varied in length between clones. Clones C28, C34, C35, and C85 contained 11, 5, 12, and 13 triplets, respectively.

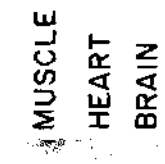




Figure 9. Northern Blot Analysis with cDNA C28

cDNA C28 hybridized to a Northern blot of baboon poly(A) RNA detects a transcript of 3.0–3.3 kb, which is expressed in heart, muscle, and, to a lesser extent, brain.

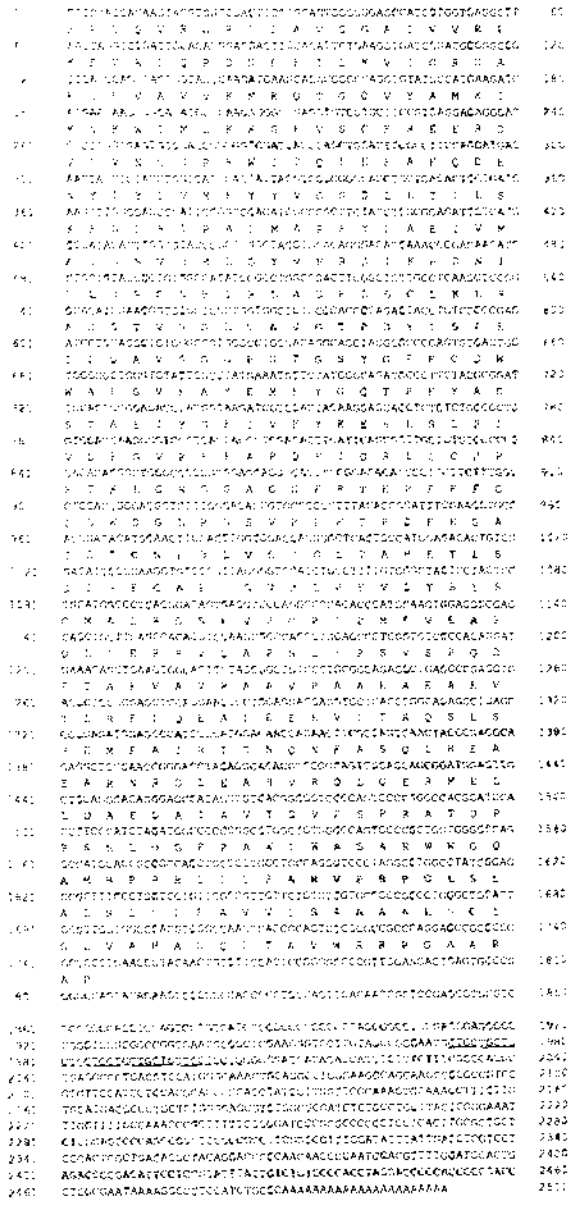


Figure 10. Nucleotide and Amino Acid Sequence of cDNA C28

The nucleotide (above) and predicted amino acid sequence (below) is shown with a presumed open reading frame of 1747 bp. The triplet repeat is underlined.

Comparison of the cDNAs with the genomic clones indicates that the gene is transcribed in the orientation telomere-to-centromere and that the CTG triplet is on the coding strand.

Northern blot analysis with C28 is shown in Figure 9. The full-length transcript is between 3.0 and 3.3 kb in length and is highly expressed in heart and to a lesser extent in muscle. Prolonged autoradiographic exposure indicates that this transcript is also expressed in brain, consistent with the identification of cDNAs from this tissue.

The sequence of C28 is shown in Figure 10. This clone contains a complete 3′ terminus, which includes the poly-

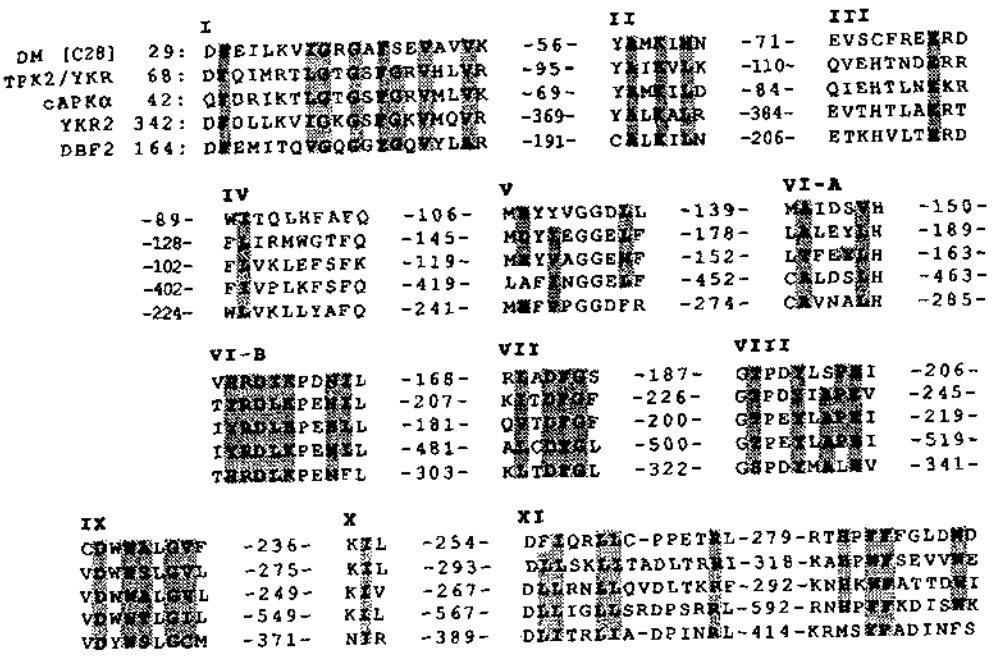


Figure 11. Amino Acid Alignment to Protein Kinase Family Members

The predicted amino acid sequence of part of cDNA C28 (listed as DM) is shown along with four other protein kinase family members (Hunter, 1991). The roman numerals indicate the major domains of the protein kinase family. The shadowed regions indicate the key amino acids in each domain. The numbers between domains refer to the position of the next amino acid shown.

adenylation addition signal AAUAAA. The largest predicted open reading frame extends from the beginning of the sequence to position 1747 with a coding capacity of 582 amino acids. The C28 sequence was compared to the nonredundant sequence database, which combines all available protein databases. This sequence search revealed homology to the cyclic AMP–dependent protein kinases. The highest score was to the protein kinase TKR-YKR from Saccharomyces cerevisiae, with a smallest poisson probability score of 7.2×10^{-34}. The 11 protein kinase domains are found within the first 300 residues of the deduced amino acid sequence, and are illustrated in Figure 11. Beyond the kinase domain, some slight homology to the chicken myosin heavy chain (smallest poisson probability score 1.1×10^{-8}) was observed.

Discussion

Myotonic dystrophy shows a clear autosomal dominant pattern of inheritance. However, there are several aspects of this disorder that are particularly challenging to explain from a molecular genetic perspective. First is the considerable variability of phenotype between affected individuals, even within the same family. Second, there is an association of DM with specific haplotypes in the population (Harley et al., 1991) indicating that most cases have resulted from a small number of genetic events. Third is the multisystemic nature of the phenotype. Fourth, there is an apparent increase in severity of symptoms and reduction in age at onset that is observed during transmission of the gene within families.

Several molecular genetic features of DM appear to be directly comparable to fragile-X syndrome. In fragile-X syndrome, increasing allele size at the FMR-1 locus, measured by Southern hybridization, is due to expansion of a CGG repeat at the 5′ end of the FMR-1 gene. Increased allele size correlates with the severity of disease (Fu et al., 1991). The extent of fragment size increase in DM also shows a clear correlation with increased severity and age at onset of the disease (Harley et al., 1992; Buxton et al., 1992). The identification of a CTG repeat, which is highly variable in the normal population and which is greatly expanded in DM patients, extends the parallels between DM and fragile-X.

Analysis of cloned PCR products reveals that the increase in size of the PCR products observed in mildly and minimally affected DM patients is due entirely to increased number of the CTG repeats. The DNA flanking the repeat is intact in all clones examined. Two minimally affected patients, DMH1 and DMH6, had repeat lengths of 50 and 52 CTGs, respectively, whereas a slightly more severely affected patient, DMH9, had in excess of 80 repeats. PCR analysis of the most severely affected patients reveals only a single band, which is in the normal size range. In these individuals the expanded allele cannot easily be visualized by PCR assay. Thus, as with fragile-X, phenotypic severity correlates with the number of repeats.

Other similarities and differences between DM and fragile-X are noteworthy. In fragile-X, individuals are categorized as normal, premutation, or full-mutation on the basis of CGG repeat number (Fu et al., 1991). Transmissions of 46 repeats and below are within the normal range

and are stable. The transition from stability to instability occurs within the 46–52 repeat range. Premutations showing no phenotypic effect range from 52 to 200 repeats and are meiotically unstable. For myotonic dystrophy there is, as yet, no clear classification of premutation. However, it seems likely that minimally affected individuals could go undiagnosed, were it not for the appearance of a more severely affected individual in a subsequent generation. This point is further illustrated in Figure 7, in which two siblings appeared normal on clinical examination at ages 64 and 61, yet showed the same upper allele as their mother, who lived to age 89 and who did not show myotonia but did have cataracts. These minimally affected individuals may be comparable to the premutational state seen in fragile-X.

The molecular mechanisms that determine the stability during inheritance of DM alleles is unclear at present. There is considerable variability at this locus in the normal population: over 75% of normal individuals are heterozygous. The largest allele observed was 27 repeats. The most common allele in the population, 5 repeats, is found at a frequency of 0.48. Analysis of repeat length in 141 individuals revealed no alleles with less than 5 copies of the repeat. Furthermore, only two alleles were found in the 6–9 repeat range. Replication of the 5 repeat allele appears to be stable. Duplication or triplication with subsequent slippage during DNA replication (Jeffreys et al., 1988) may account for the generation of other alleles. Unequal crossing over would seem to be unlikely as the mechanism generating allelic diversity, in view of the shortage of alleles between 5 and 10 repeats. Clearly, further study of the mechanisms that underlie variation in repeat length at the DM locus will be of great interest. In addition, it will be important to determine the extent of tissue mosaicism in the expansion of the CTG repeat. Such mosaicism could have a significant impact on clinical phenotype.

One of the families documented by Fu et al. (1991) appears to represent a fragile-X premutation segregating in the normal population, which has yet to undergo expansion to the full mutation.

It is unclear in fragile-X whether new mutation is a frequent event. It seems quite possible, however, that transition from a large allele in the normal population to premutation allele in fragile-X could occur by the same mechanism that generates diversity amongst alleles in the normal size range (Fu et al., 1991). Thus, large alleles in the normal population and fragile-X premutations would represent a continuum across a stability threshold, with the fragile-X phenotype generated by multiple independent events. At the DM locus, however, two observations suggest that expansion of the CTG repeat to a clinically significant level is likely to occur only in a specific population subgroup.

First, there is clear evidence in heterogeneous populations of linkage disequilibrium between DM and polymorphisms at "nearby" loci (Harley et al., 1991, 1992). This implies either that there are few mutations, possibly a single ancestral event, or that specific nearby polymorphisms predispose to the generation of DM mutations. It is difficult to envisage a mechanism by which multiple polymorphisms at distances of up to 70 kb (in the case of D19S63) from the CTG repeat could predispose to DM. On the other hand, if there are very few, if any, new mutations, some mechanism must maintain the disease allele in the population, particularly in view of the genetic endpoint represented by severely affected individuals. Thus, there may be a large unrecognized pool of individuals in the population who carry and transmit the DM premutation with little if any phenotypic effect. Extensive studies of normal populations (possibly focusing on individuals with cataracts) will be necessary to test this possibility.

Second, the distribution of CTG repeat alleles is quite distinct in the normal population and in DM patients. In DM the smallest number of CTG repeats observed is 50, almost double the largest number of repeats seen in the normal population. It is possible that a doubling or tripling in repeat number is the ancestral event that predisposes an allele at the DM locus to further expansion into an allele associated with the complete disease phenotype.

Unraveling the connections between expansion of the CTG repeat and the symptoms of DM is an important challenge for future investigation. While it is certainly possible that amplification of the CTG affects the expression of several transcription units in the immediate vicinity of the repeat, it seems very likely that the transcription unit we have identified in this study plays an important role in the pathophysiology of DM. The mechanism through which expansion of the CTG repeat affects the function of the DM gene remains to be elucidated. Since DM is a dominantly inherited disorder, mutant alleles must exert an effect in the presence of a normal allele. There are a number of possible ways in which an amplified sequence in the 3′ UTR of a gene could exert an effect on the function of that gene. One possibility is that, analogous to fragile-X, the expansion of the CTG repeat causes a loss of expression of the allele carrying the expanded repeat. If this is the case, then the DM gene must indeed be extremely sensitive to gene dosage, since the range of gene expression levels in the presence of a normal allele can range only between 50% and 100% of normal. Genomic imprinting cannot be invoked to increase this range too much further, since DM can be inherited from either the father or the mother with quite severe symptoms. The issue of congenital DM, which shows exclusively maternal transmission and may well be affected by imprinting, is not addressed in this manuscript. Alternatively, the expansion of the CTG repeat may lead to a gain of function in the DM gene. Gain-of-function mutations in the 3′ UTR of the *fem-3* and *lin-14* genes of Caenorhabditis elegans have recently been demonstrated (Ahringer and Kimble, 1991; Wightman et al., 1991). In both cases, deletion or inactivation of a binding site for a negative control element is thought to result in unregulated activity of these genes. Amplification of the CTG repeat in DM may be producing a similar effect.

The similarity of the DM gene to members of the protein kinase family, in particular cAMP-dependent protein kinase (cAPK), opens a broad range of physiological questions that should be directly tested. cAPKs (Hunter, 1991) are known to modulate the activity of excitable cells by phosphorylation of ion channels, exert control of glycogen

and lipid metabolism through cascades of enzyme phosphorylation, and modify gene expression (Yamamoto et al., 1988; Foulkes et al., 1991). Abnormalities in function or regulation of such a molecule fit well with the diverse phenotypic effects exhibited by DM patients. It is intriguing that biochemical data produced nearly 20 years ago by Roses and Appel (1973, 1974) indicated that a defect in a protein kinase may be significant in DM. A resolution of the underlying physiological basis of DM should prove to be an important step both in developing appropriate therapeutic intervention and in gaining further insight into the multiple organ systems that exhibit physiological disruption in DM.

Experimental Procedures

PCR Analysis

The PCR analysis was performed as follows. Reactions (10 μl) were set up using standard PCR conditions (50 mM KCl, 1.5 mM $MgCl_2$, 10 mM Tris (pH 8.3], 200 μM dNTPs, 1 μM of each primer, and 20 ng template DNA). For the radiolabeled experiments, oligo 101 was incubated for 30 min at 37°C with T4 polynucleotide kinase (3μ) in 20 μl reaction with 50 mM Tris-HCl (pH 7.5), 10 mM $MgCl_2$, 5 mM dithiothreitol, 0.1 mM spermidine, and 1 μl of [γ-^{32}P]ATP (3000 ci/mmol). The PCR reaction was spiked with the labeled primer in a ratio of 30:1 unlabeled to labeled. PCR reactions were carried out on a Stratagene 96 well thermocycler. Cycling conditions were as follows: 1 × (3 min, 94°C), 35 × (10 s, 94°C, 30 s, 62°C; 30 s, 72°C), 1 × (5 min, 72°C), 15°C soak. Samples then were either loaded on agarose gels or were diluted 1:1 with sequencing buffer, denatured, and loaded on 8% sequencing gels.

The sequences of Oligos used in PCR reactions was as follows:

96, GGT GCG TGG AGG ATG GAA CAC GGA C
98, GCG TGC GAG TGG ACT AAC AAC AGC TG
100, CAC GCT CGG AGC GGT TGT GAA CTG G
101, CTT CCC AGG CCT GCA GTT TGC CCA TC
102, GAA CGG GGC TCG AAG GGT CCT TGT AGC
103, CCA GTT CAC AAC CGC TCC GAG CGT G

Genomic Digests and Southern Blots

Genomic DNAs were digested with restriction endonucleases PstI, SacI, EcoRI, or BamHI (New England Biolabs) in 30 μl or 40 μl reactions with NEBuffer 10 × according to the manufacturer's instructions. Digested DNAs were run on 0.8% agarose gels (FMC), denatured in 0.5 M NaOH, 1.5 M NaCl, neutralized in 1 M Tris (pH 7.0), 1.5 M NaCl, and transferred to Zetabind (AMF) membranes in 10 × SSC.

Hybridizations

Hybridizations to both Northern and Southern blots were performed at 42°C in 50% formamide with 5 × SSC, 1 × Denhardt's solution, 0.02 M $NaPO_4$, 100 μg/ml single-stranded DNA, 10% dextran sulphate.

DNA probes were labeled by random priming (Feinberg and Vogelstein, 1983).

Northern blots

Total RNA was extracted from baboon tissue using the method of Auffray and Rougeon (1980) with modifications from Buckler et al. (1991). Poly(A) RNA was isolated with oligo(dT)-cellulose, and gels and Northern blots were set up as described in Buckler et al. (1991).

DNA Sequencing

Three parallel sequencing strategies were adopted. Much of the sequencing was carried out using a U.S. Biochemical sequence kit according to manufacturer's instructions on cDNA constructs. In general, vector oligonucleotides were used as primers, and in some cases, specific oligonucleotides were synthesized from deduced sequence. We also used cycle sequencing with a Promega Kit and some automated sequencing on the ABI model system 373A.

cDNA Libraries

Four cDNA libraries were screened. Three, constructed in λZAP from frontal cortex, substantia nigra, and fetal brain, were kindly supplied by Dr. Marcy MacDonald. A fetal muscle library in λGT10 was kindly supplied by Dr. L. Kunkel.

DNA Database Searching

DNA databases were searched on a digital VAX computer using the GCC software package (Altschul et al., 1990) and the BLAST network service from the National Center for Biotechnology Information.

Acknowledgments

We wish to acknowledge the contribution of a series of invaluable clinical research fellows: William Reardon, Andrew Norman, Peter Lunt, and Susan Huson. We thank Dr. Marcy MacDonald and Dr. Lou Kunkel for cDNA libraries; Dr. André Bernards for helpful advice about protein kinases; Dr. Bob Horvitz for discussions on gain-of-function mutations in C. elegans; the Massachusetts Institute of Technology Cancer Center Biopolymers Laboratory, especially Sandy Schultz and Richard Cook, for excellent service; and Kevin Porter for artwork and photographic prints. We also thank friends and colleagues, especially Dr. Dave Munroe and Dr. Jerry Pelletier for many useful discussions, help, advice, and various other distractions. We thank Sara Dell for much help in the preparation of this manuscript. This work was funded by the Muscular Dystrophy Association and the Piton Foundation, the Muscular Dystrophy Group of Great Britain, the Wellcome Trust, the Hereditary Disease Foundation, Program of Excellence Grant HL-41484 of the National Heart, Lung and Blood Institution, and Human Genome Initiative Grant HG 00299 from the National Institutes of Health (to D. E. H.).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 USC Section 1734 solely to indicate this fact.

Received February 5, 1992; revised February 7, 1992.

References

Ahringer, J., and Kimble, J. (1991). Control of the sperm–oocyte switch in *Caenorhabditis elegans* hermaphrodites by the *fem-3* 3′ untranslated region. Nature *349*, 346–348.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). A basic local alignment search tool. J. Mol. Biol. *215*, 403–410.

Aslanidis, C., Jansen, G., Amemiya, C., Shutler, G., Tsilfidis, C., Mahadevan, M., Chen, C., Alleman, J., Wormskamp, N. G. M., Vooijs, M., Buxton, J., Johnson, K., Smeets, H. J. M., Lennon, G. G., Carrano, A. V., Korneluk, R. G., Wieringa, B., and de Jong, P. J. (1992). Cloning of the essential myotonic dystrophy region: mapping of the putative defect. Nature *355*, 548–551.

Auffray, C., and Rougeon, F. (1980). Purification of mouse immunoglobulin heavy-chain messenger RNAs from total myeloma tumor RNA. Eur. J. Biochem. *107*, 303–314.

Brook, J. D., Knight, S. L., Roberts, S. H., Harley, H. G., Walsh, K. V., Rundle, S. A., Freyne, K., Koch, M. C., Epstein, N. D., Wieringa, B., Smeets, H., Haddingham, K., Siciliano, M. J., Palmer, D. H., Miles, J. S., Wolf, C. R., Fonatsch, C., and Shaw, D. J. (1991a). The physical map of chromosome arm 19q: some new assignments, confirmations and re-assessments. Hum. Genet. *87*, 65–72.

Brook, J. D., Harley, H. G., Walsh, K. V., Rundle, S. A., Siciliano, M. J., Harper, P. S., and Shaw, D. J. (1991b). Identification of new DNA markers close to the myotonic dystrophy locus. J. Med. Genet. *28*, 84–88.

Brook, J. D., Zemelman, B. V., Haddingham, K., Siciliano, M. J., Crow, S., Harley, H. G., Rundle, S. A., Buxton, J., Johnson, K., Almond, J. W., Housman, D. E., and Shaw, D. J. (1992). Radiation reduced hybrids for the myotonic dystrophy locus. Genomics, in press.

Brunner, H. G., Smeets, H., Lamberman, H. M. M., Coerwinkel-Driessen, M., Van Oost, B. A., Wieringa, B., and Roper, H.-H. (1989). A multipoint linkage map around the locus for myotonic dystrophy on chromosome 19. Genomics *5*, 589–595.

Buckler, A. J., Pelletier, J., Haber, D. A., Glaser, T., and Housman, D. E. (1991). Isolation, characterization, and expression of the murine Wilms' tumor gene (WT1) during kidney development. Mol. Cell. Biol. *3*, 1707–1712.

Butterfield, D. A., Roses, A. D., and Copper, M. L. (1974). A comparative electron spin resonance study of the erythrocyte membrane in myotonic muscular dystrophy. Biochemistry *13*, 5078–5082.

Buxton, J., Shelbourne, P., Davies, J., Jones, C., Van Tongeren, T., Aslanidis, C., de Jong, P., Jansen, G., Anvret, M., Williamson, R., and Johnson, K. (1992). Detection of an unstable fragment of DNA specific to individuals with myotonic dystrophy. Nature *355*, 547–548.

Dietrich, A., Kioschis, P., Monaco, A. P., Gross, B., Williams, S. V., Sheer, D., Heitz, D., Oberle, I., Toniolo, D., Warren, S. T., Lehrach, H., and Poustka, A. (1991). Molecular cloning and analysis of the fragile-X region in man. Nucl. Acids Res. *19*, 2567–2572.

Dyken, P. R., and Harper, P. S. (1973). Congenital dystrophia myotonica. Neurology *23*, 465–473.

Edwards, A., Hammond, H. A., Jin, L., Caskey, C. T., and Charkraborty, R. (1992). Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. Genomics *12*, 241–253.

Feinberg, A. P., and Vogelstein, B. (1983). A technique for radiolabeling DNA restriction fragments to high specific activity. Biochem. Biophys. Res. Commun. *111*, 47–54.

Foulkes, N. S., Borrelli, E., and Sassone-Corsi, P. (1991). CREM gene: use of alternative DNA-binding domains generates multiple antagonists of cAMP-induced transcription. Cell *64*, 739–749.

Fu, Y.-H., Kuhl, D. P. A., Pizzuti, A., Pieretti, M., Sutcliffe, J., Richards, S., Verkerk, A. J. M. H., Holden, J. J. A., Fenwick, R. G, Jr., Warren, S. T., Oostra, B. A., Nelson, D. L., and Caskey, C. T. (1991). Variation of the CGG repeat at the fragile-X site results in genetic instability: resolution of the Sherman paradox. Cell *67*, 1047–1058.

Hanks, S. K., and Quinn, A. M. (1991). Protein kinase catalytic domain sequence database: identification of conserved features of primary structure and classification of family members. Meth. Enzymol. *200*, 38–62.

Hanks, S. K., Quinn, A. M., and Hunter, T. (1988). The protein kinase family: conserved features and deduced phylogeny of the catalytic domains. Science *241*, 42–52.

Harley, H. G., Walsh, K. V., Rundle, S., Brook, J. D., Sarfarazi, M., Koch, M. C., Floyd, J. L., Harper, P. S., and Shaw, D. J. (1991). Localisation of the myotonic dystrophy locus to 19q 13.2–19q 13.3 and its relationship to twelve polymorphic loci on 19q. Hum. Genet. *87*, 73–80.

Harley, H. G., Brook, J. D., Rundle, S. A., Crow, S., Reardon, W., Buckler, A. J., Harper, P. S., Housman, D. E., and Shaw, D. J. (1992). Expansion of an unstable DNA region and phenotypic variation in myotonic dystrophy. Nature *355*, 545–546.

Harper, P. S. (1989). Myotonic Dystrophy, 2nd ed. (London: W. B. Saunders Co.).

Harper, P. S., and Dyken, P. R. (1972). Early onset dystrophia myotonica—evidence supporting a maternal environmental factor. Lancet *2*, 53–55.

Howeler, C. J., Busch, H. F. M., Geraedts, J. P. M., Niermeijer, M. F., and Staal, A. (1989). Anticipation in myotonic dystrophy: fact or fiction. Brain *112*, 779–797.

Hull, K. L., Jr., and Roses, A. D. (1976). Stoichiometry of sodium and potassium transport in erythrocytes from patients with myotonic muscular dystrophy. J. Physiol. *254*, 169–181.

Hunter, T. (1991). Protein kinase classification. Meth. Enzymol. *200*, 3–37.

Jansen, G., de Jong, P. J., Amemiya, C., Aslanidis, C., Shaw, D. J., Harley, H. G., Brook, J. D., Fenwick, R., Korneluk, R. G., Tsilfidis, C., Shutler, G., Hermens, R., Wormskamp, N. G. M., Smeets, H. J. M., and Wieringa, B. (1992). Physical and genetic characterization of the distal segment of the myotonic dystrophy area on 19q. Genomics, in press.

Jeffreys, A. J., Royle, N. J., Wilson, V., and Wong, Z. (1988). Spontaneous mutation rates to new length alleles at tandem-repetitive hypervariable loci in human DNA. Nature *332*, 278–281.

Johnson, K., Shelbourne, P., Davies, J., Buxton, J., Nimmo, E., Siciliano, M. J., Bachinski, L. L., Anvret, M., Harley, H., Rundle, S., Miki, T., Brunner, H., and Williamson, R. (1990). A new polymorphic probe which defines the region of chromosome 19 containing the myotonic dystrophy locus. Am. J. Hum. Genet. *46*, 1073–1082.

Korneluk, R. G., MacKenzie, A. E., Nakamura, L., Dube, I., Jacob, P., and Hunter, A. G. W. (1989). A re-ordering of human chromosome 19 long-arm DNA markers and identification of markers flanking the myotonic dystrophy locus. Genomics *5*, 596–604.

Kremer, E. J., Pritchard, M., Lynch, M., Yu, S., Holman, K., Baker, E., Warren, S. T., Schlessinger, D., Sutherland, G. R., and Richards, R. I. (1991). Mapping of DNA instability at the fragile X to a trinucleotide repeat sequence p(CCG)n. Science *252*, 1711–1714.

La Spada, A. R., Wilson, E. M., Lubahn, D. B., Harding, A. E., and Fischbeck, K. H. (1991). Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy. Nature *352*, 77–79.

Oberle, I., Rousseau, P., Heitz, D., Kretz, C., Devys, D., Hanauer, A., Bour, J., Bertheas, M. F., and Mandel, J. L. (1991). Instability of a 550 base pair DNA segment and abnormal methylation in fragile X syndrome. Science *252*, 1097–1102.

Renaud, J. F., Desnuelle, C., Schmid-Antomarchi, M., Hugues, M., Serratrice, G., and Lazunski, M. (1986). Expression of apamin receptor in muscles of patients with myotonic muscular dystrophy. Nature *319*, 678–680.

Roses, A. D., and Appel, S. H. (1973). Protein kinase activity in erythrocyte ghosts of patients with myotonic muscular dystrophy. Proc. Natl. Acad. Sci. USA *70*, 1855–1859.

Roses, A. D., and Appel, S. H. (1974). Muscle membrane protein kinase in myotonic muscular dystrophy. Nature *250*, 245–247.

Seiler, D., and Kuhn, E. (1970). Kalzium Transport der isolierten Vesikel des Sarkoplasmatischen Retikulums von Patienten mit Myotonia congenita und Myotonia dystrophica. Schweiz Med. Wochenschr. *100*, 1374–1376.

Shutler, G., Korneluk, R. G., Tsilfidis, C., Manadevan, M., Bailly, J., Smeets, H., Wieringa, B., Lohman, F., Aslanidis, C., and de Jong, P. J. (1992). Physical mapping and cloning of the proximal segment of the myotonic dystrophy gene region. Genomics, in press.

Smeets, H., Bachinski, L., Coerwinkel, M., Schepens, J., Hoelimakers, J. Duin, M., van Grzeschik, K. H., Weber, C. A., De Jan, P., Siciliano, M. J., and Wieringa, B. (1990). A long range restriction map of the human chromosome 19q13 region: close physical linkage between cumm and the ERCC1 and ERCC2 genes. Am. J. Hum. Genet. *46*, 492–502.

Tsilfidis, C., Mackenzie, A. E., Shutler, G., Leblond, S., Bailly, J., Johnson, K., Williamson, R., Siegel-Bartelt, J., and Korneluk, R. G. (1991). D19S51 is closely linked with and maps distal to the myotonic dystrophy locus on 19q. Am. J. Hum. Genet. *49*, 961–965.

Verkerk, A. J. M. H., Peiretti, M., Sutcliffe, J. S., Fu, Y.-H., Kuhl, D. P. A., Pizzuti, A., Reiner, O., Richards, S., Victoria, M. F., Zhang, F., Eussen, B. E., van Ommen, G. J. B., Blonden, L. A. J., Riggins, G. J., Chastain, J. L., Kunst, C. B., Galjaard, H., Caskey, C. T., Nelson, D. L., Oostra, B. A., and Warren, S. T. (1991). Identification of a gene (*FMR-1*) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. Cell *65*, 905–914.

Wightman, B., Bürglin, T., Gatto, J., Arasu, P., and Ruvkun, G. (1991). Negative regulatory sequences in the *lin-14* 3′-untranslated region are necessary to generate a temporal switch during *Caenorhabditis elegans* development. Genes Dev. *5*, 1813–1824.

Yamamoto, K. K., Gonzalez, G. A., Biggs, W. H., III, and Montminy, M. R. (1988). Phosphorylation-induced binding and transcriptional efficacy of nuclear factor CREB. Nature *334*, 494–498.

Yu, S., Pritchard, M., Kremer, E., Lynch, M., Nancarrow, J., Baker, E., Holman, K., Mulley, J. C., Warren, S. T., Schlessinger, D., Sutherland, G. R., and Richards, R. I. (1991). Fragile X genotype characterized by an unstable region of DNA. Science *252*, 1179–1181.