NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30, and subsequent amendments.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30, et ses amendements subséquents.

Canada

THE HIGH RESOLUTION HAPLOTYPE ANALYSIS AND ORIGIN OF THE MYOTONIC DYSTROPHY MUTATION

A thesis submitted to the School of Graduate Studies at the University of Ottawa in partial fulfillment of the requirements for the degree of Master of Science, Department of Microbiology and Immunology, Faculty of Medicine.

By Catherine E. Neville

ISBN 0-612-00489-9

Canada

UNIVERSITÉ D'OTTAWA
UNIVERSITY OF OTTAWA

ABSTRACT

Myotonic dystrophy (DM) is the most common inherited neuromuscular disease affecting adults. The mutation causing DM has been identified as an amplification of an unstable trinucleotide $(CTG)_n$ repeat located in the 3' untranslated region of a gene encoding a putative cAMP-dependent serine-threonine protein kinase. Amplification of the repeat appears to be the molecular basis for genetic anticipation since there is a broad positive correlation between the number of repeats and disease severity.

The DM mutation is in complete linkage disequilibrium with an *Alu* element polymorphism located 5 kb telomeric to the $(CTG)_n$ repeat suggesting a single common origin of the mutation. This finding was unexpected for a dominant disease which, in its severe form diminishes or abolishes (in the case of congenital DM) reproductive fitness. Such diseases are in general characterized by a high level of new mutations, which compensate for the loss of abnormal alleles due to decreased fitness. Since all cases of DM are familial and there has been no spontaneous occurrence of the disease documented, it is difficult to rationalize the relatively stable incidence of DM in the global population. Therefore, a paradox exists in DM in which the incidence of the disease has apparently remained relatively stable despite the influence of genetic anticipation and evidence to suggest it should be declining. If the incidence is not, in fact, decreasing then the possibility exists that the DM mutation may be more common than previously supposed, and may be passed down in a stable and clinically inconsequential manner for many generations. It was therefore suggested that DM could be due to recurrent mutations occurring on a predisposing allelic form of the normal gene. Such predisposition may be in the $(CTG)_n$ repeat itself or in neighbouring sequences including a putative direct effect of the 1 kb insertion/deletion *Alu* element polymorphism.

The objective of this thesis was thus to determine the origin of the DM mutation. I have used PCR-based assays of nine polymorphisms spanning a physical distance of 30 kb, within and immediately flanking the DM kinase gene, in order to examine patterns of allelic association with respect to the DM mutation. In addition, high resolution haplotype analysis employing these markers was conducted in order to predict the origin and number of DM mutations segregating in our heterogeneous study population.

Four main haplotypes (A-D) were observed in the normal population using these nine markers at the DM locus. Significantly, DM is in complete association with haplotype A, the most common haplotype in the normal population. Our data suggest the presence of two founding chromosomes: one containing a stretch of five contiguous *Alu* elements (the progenitor for haplotype A) and the other in which three of these have been deleted (the progenitor for haplotypes B, C and D). Individuals with haplotype A displayed the full spectrum of $(CTG)_n$ number, ranging from 5 to 35 repeats. The $(CTG)_5$ alleles as well as alleles with greater than 19 repeats are exclusively linked to haplotype A. In contrast, alleles in which the three *Alu* elements are deleted possess only $(CTG)_{11\text{-}14}$ repeats. Although the inference that the loss of the three *Alu* repeats may confer increased stability on the $(CTG)_n$ repeat is speculative, the narrow size range of the $(CTG)_n$ repeat on chromosomes in which the *Alu* elements have been deleted, relative to the variation seen on normal chromosomes with the DM haplotype (i.e. haplotype A), is striking. The complete allelic association of the DM mutation with all these polymorphisms may reflect a single mutational origin for DM and a haplotype predisposed to $(CTG)_n$ repeat instability.

We have proposed a model in which the DM mutation is postulated to be occurring as a multistep process. This model provides a framework in which the seemingly contradictory observations of a mutation old enough to establish a founder effect and an apparently high new mutation rate are united. The model suggests that alleles with $(CTG)_n \geq 20$ repeats arise from normal alleles and that these constitute a

reservoir from which protomutation size alleles are derived. These protomutations, in the range of 50-80 $(CTG)_n$ repeats become inherently unstable and therefore predisposed to rare amplification. It is this pool of pre-existing or predisposed carriers (individuals with protomutations) which is constantly giving rise to new clinical DM cases. Based on this model, the loss of DM alleles from the population due to the end result of genetic anticipation could be balanced by the transition of protomutations to full mutations. This may account for the relatively stable prevalence of DM in the global population.

ACKNOWLEDGMENTS

I thank Dr. Robert G. Korneluk for his guidance and support during these studies and Dr. Alexander E. MacKenzie for helpful discussions. I am grateful to the DM patients and their families for providing blood samples and to Dr. Linda Surh and the diagnostic laboratory for the use of the extracted DNA for my research. I appreciate the expert technical assistance provided by Stephen Baird on the computer and the audio-visual department at the Children's Hospital of Eastern Ontario. I especially thank my many colleagues and friends in our laboratory for their encouragement and friendship. I am grateful to Nicholas L. Bott for his love and understanding and to the System for teaching me success principles. This work was supported by operating grants to Dr. Korneluk from the Muscular Dystrophy Associations of Canada and the United States of America, the Medical Research Council of Canada and the Canadian National Centres of Excellence Genetic Disease Network.

DEDICATION

To my parents, George and Iris, my brother, Jeffrey and sister, Laura, with love.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER II
THE ORIGIN OF THE MYOTONIC DYSTROPHY MUTATION

LIST OF FIGURES

CHAPTER I
INTRODUCTION TO MYOTONIC DYSTROPHY

CHAPTER II
THE ORIGIN OF THE MYOTONIC DYSTROPHY MUTATION

LIST OF ABBREVIATIONS

A	adenine base in a nucleotide sequence
APOCII	Apolipoprotein CII gene
ASO	allele-specific oligonucleotide
bp	base pair
BRL	Bethesda Research Laboratories
°C	degrees Celsius
C	cytosine base in a nucleotide sequence
cAMP	cyclic adenosine monophosphate
cDNA	complementary DNA
CKM	muscle specific creatine kinase gene
cM	centiMorgan
C3	third component of human complement
DM	dystrophia myotonica/myotonic dystrophy
DMK	DM kinase
DNA	deoxyribonucleic acid
dNTP	deoxynucleotide triphosphate
EMG	electromyography
ERCC1	excision repair cross complementation gene 1
G	guanine base in a nucleotide sequence
HD	Huntington's disease
HMG-CoA	3-hydroxy-3-methylglutaryl coenzyme A
kb	kilobase
lod	logarithm of the odds
Mb	megabase
min.	minute

mM	millimolar
mRNA	messenger ribonucleic acid
ng	nanogram
PCR	polymerase chain reaction
RFLP	restriction fragment length polymorphism
SDS	sodium dodecyl sulphate
SSC	standard saline citrate
T	thymidine base in a nucleotide sequence
TBE	Tris-borate, EDTA buffer
Tris	tris(hydroxymethyl)aminomethane
μg	microgram
μL	microlitre
μM	micromolar
YAC	yeast artificial chromosome
Z_{max}	maximum lod score
θ	recombination fraction

CHAPTER I INTRODUCTION TO MYOTONIC DYSTROPHY

Myotonic Dystrophy or Dystrophia Myotonica (DM) is the most common form of muscular dystrophy affecting both adults and children (Harper, 1989). DM is inherited as an autosomal dominant disorder and is therefore transmitted by affected individuals of either sex to 50% of their offspring. All cases of DM are familial and there have been no reported cases of new mutation. DM has an incidence of 1 in 8000 in the Western European and North American populations (Harper, 1989) and a lower incidence of 1 in 20,000 in Japan (Davies et al., 1992). However, the real incidence is likely to be higher due to the variable penetrance of the DM gene and the variable age at onset, which can create difficulty in detecting affected individuals. Although the disease has a global distribution, French Canadians have the highest reported prevalence of DM, with frequencies of nearly 1 in 500 in the Lac St-Jean/Saguenay region of Quebec (Mathieu et al., 1990). This provides a unique opportunity for the study of DM in the Eastern Ontario and Western Quebec populations.

Clinical features of DM

There are two distinct clinical forms of DM expression: an adult-onset and a congenital form (Harper, 1989). The classical adult form of the disease is characterized by myotonia, an electrophysiological disturbance resulting in delayed muscle relaxation. Myotonia is diagnosed through physical examination and by electromyography (EMG). The repetitive myotonic potentials detected by EMG vary in frequency and amplitude before eventual decline to produce a characteristic "dive bomber" response most evident on auditory recordings. Furthermore, patients display progressive weakening and wasting of skeletal muscle, typically involving the face, jaw, neck, arms and legs. Smooth as well as skeletal muscle may be involved in the manifestation of DM. However, DM is also a heterogeneous disorder encompassing a variety of tissues and organs in addition to

muscle. Cardiac conduction defects, abnormal carbohydrate metabolism, retinal degeneration and cataracts, premature balding, infertility and mild intellectual impairment are often secondary manifestations of DM. The fact that DM is a multisystemic disease with a variable age at onset suggests that the disease gene is altered in its expression in a variety of tissues and at different times of development.

The clinical presentation of congenital DM differs significantly from adult-onset DM. Two prominent features of adult-onset DM, myotonia and cataracts, are characteristically absent in congenital DM. The affected infant presents with extreme hypotonia, muscular atrophy, feeding difficulties and neonatal respiratory distress. Congenital DM is usually fatal shortly after birth, however, if the patient survives, there is a tendency to improve during childhood. This is followed by a gradual deterioration during adolescence with eventual manifestations characteristic of the adult-onset form, however intellectual impairment is often more severe. Significantly, this form of DM is associated with a delay or an arrest of muscle maturation (Sarnat and Silbert, 1976; Soussi-Yanicostas et al., 1992).

Congenital DM is exclusively transmitted by DM mothers, conversely approximately 20% of maternally transmitted DM is the congenital form (Glanz and Fraser, 1984). Once a mother has had a congenital DM child, then the risk of having another infant with the congenital form is almost 100%, if that child has inherited the DM chromosome (Koch et al., 1991). The mechanism of influence of the transmitting parent's sex is not completely understood. The involvement of imprinting or some other epigenetic phenomenon has been suggested (Hall, 1990) while others have postulated the involvement of a maternal intrauterine toxic factor (Koch et al., 1991).

Variability in the expression of DM ranges from the severe congenital form to an asymptomatic condition associated with normal longevity. Moreover, both adult-onset and congenital forms are often seen in the same affected family. This illustrates a phenomenon characteristic of DM, genetic anticipation, which is defined as an increase in

disease severity and decrease in age at onset in successive generations. Originally, this biological phenomenon in DM pedigrees had been discounted as being simply due to ascertainment bias (Penrose et al., 1948). However, as discussed later, the unstable nature of the recently identified DM mutation may explain the molecular basis for genetic anticipation. Furthermore, anticipation in DM pedigrees typically corresponds with a decrease in reproductive fitness.

The median age of onset for the adult form of DM is 20-25 years. However, this varies widely and some individuals who carry the DM gene never develop any symptoms of the disease (Harper, 1989). The extreme cases may pose diagnostic problems since the mildly affected patients may never present with a complaint, and the severe congenital patients may die before a specific diagnosis is made. This variability in penetrance of the gene clearly complicates the genetic study of DM.

Positional cloning of the DM gene

Cloning and subsequent characterization of the gene responsible for DM has led to a better understanding of the disease. Unfortunately, there were no known chromosomal aberrations or obvious biochemical abnormalities associated with DM. Consequently, discovery of the DM gene relied upon "positional cloning", which is the isolation of the gene based on its chromosomal location. Application of this approach has been most notably successful in the identification and isolation of the gene involved in cystic fibrosis (Kerem et al., 1989; Riordan et al., 1989; Rommens et al., 1989).

Positional cloning with respect to locating the DM gene involved a three part strategy: the construction of a high resolution genetic map, the generation of a complementary long range physical map and the identification of flanking markers tightly linked to the disease gene. A genetic linkage map shows the relative location of specific DNA markers along the chromosome. The value of the genetic map is that an inherited disease can be located on the map by following the inheritance of a DNA marker present

in affected individuals, even though the molecular basis of the disease may not yet be understood nor the responsible gene identified. Physical mapping of the disease locus can only be indirectly determined by physical mapping of cloned markers showing genetic linkage to the disease locus. To obtain higher resolution genetic and physical mapping of the disease locus, the identification and mapping of flanking genetic markers which demonstrate single recombinant events are required. The crossover points of these recombinant events could therefore serve as physical landmarks defining the proximal and distal boundaries of the essential region expected to contain the disease gene.

By combining both genetic and physical linkage data, the DM locus was assigned to a specific chromosomal interval. The DM locus had been mapped to chromosome 19 through the demonstration of genetic linkage to the third component of human complement, C3 (Eiberg et al., 1983) and the previous localization of C3 to this chromosome by somatic cell hybrid studies (Whitehead et al., 1982). Subsequently, close linkage between the gene for apolipoprotein CII (ApoCII) and DM was reported (Shaw et al., 1985) and based on the physical mapping of the ApoCII gene (Hulsebos et al., 1985a, 1985b; Lusis et al., 1985), the DM locus was indirectly assigned to the chromosome 19 interval 19cen-19q13.2. Further resolution in the genetic mapping of the DM locus within this interval was achieved by extensive linkage analysis (Korneluk et al., 1989a) of a large number of DM families using seven DM linked markers. Analysis of recombinant events in this study identified flanking markers such that the DM locus was found to lie in a 10 cM interval between ApoCII/CKM and pEWRB1.4 (D19S50).

Significant progress was achieved in the genetic mapping of the DM locus once it was bracketed with flanking markers defining two key recombinant events. One recombinant occurred in a Dutch DM family (Smeets et al., 1991) localizing the DM locus distal to the DNA repair gene ERCC1 and D19S115 (identified by pE0.8) (Shutler et al., 1991). The other recombinant event was identified in a member of a French Canadian DM family (Tsilfidis et al., 1991) which maps DM proximal to D19S51

(p134c). The genetic distance between D19S51 and D19S115 was calculated to be about 1 cM or less. The actual physical distance was determined through a joint effort with Dr. Bé Wieringa of the Netherlands. We initiated a chromosome walk in a telomeric direction from pE0.8, a proximal marker tightly linked to DM, towards the DM locus (Shutler et al., 1992). Concurrently, our Dutch collaborators initiated a cosmid walk near the D19S51 locus and extended a contig in a centromeric direction towards the distal end of our genomic walk (Jansen et al., 1992a).

A number of polymorphic DNA markers have been isolated from our cosmid contig and have been used in the study of key recombinant events and in the analysis of linkage disequilibrium patterns and haplotype convergence. Linkage analysis was performed on a total of 18 RFLPs at 12 loci spanning a greater than 1 Mb region containing the myotonic dystrophy locus (Tsilfidis et al., manuscript in preparation). Two polymorphisms, identified by pKH1.1 (D19S196) and pKH1.3 (D19S194), which I analyzed and characterized were included in this study (Appendix 1). A founder effect has been previously documented in our French Canadian DM population (Korneluk et al., 1989b; MacKenzie et al., 1989). However, this analysis conducted on our entire DM kindred sample incorporating diverse linguistic groups, provided evidence for linkage disequilibrium between DM-linked markers and the disease locus extending beyond our French Canadian population.

Single copy probes derived from the contig were used for genetic and physical mapping in order to define a minimal area which contains the DM gene. Specifically, the proximal crossover point in the Dutch recombinant family was mapped between two DNA markers, pKEX0.8 (D19S118) and pKBE0.8 (D19S119) in the middle of our chromosome walk (Shutler et al., 1992). Based on this result, the closest proximal marker to DM was then established to be pKEX0.8. This represented the elimination of approximately 200 kb of sequence between pE0.8 (D19S115) and pKEX0.8 (D19S118) that would have otherwise been screened for DM candidate genes. In addition, a highly

polymorphic $(CA)_n$ repeat DNA marker, designated pX-75b (D19S112) was identified and found to cross over with DM in our French Canadian recombinant DM family (Jansen et al., 1992a). Furthermore, pX-75b was found to be located approximately 150 kb centromeric to the D19S51 locus. Therefore, the region between the D19S51 and D19S112 loci was excluded as a candidate DM area and pX-75b was established as the closest distal marker to DM. Thus, the essential region that contains the DM locus was determined to be located in the 350 kb area between D19S118 and D19S112.

A flow sorted chromosome 19 specific cosmid library (generated by the Lawrence Livermore National Laboratory in Livermore, California) was used to isolate DM region cosmids in our directed chromosome walk. Using a novel *Alu* PCR-based methodology, we isolated nearly 400 kb of contiguous human genomic DNA (Shutler et al., 1992). However, at the distal end, no additional cosmid genomic clones could be isolated. In his approach from the telomeric side, Dr. Wieringa also encountered the same unclonable stretch at the proximal end of his chromosome walk (Jansen et al., 1992a). To bridge this gap between the two cosmid contigs, we had provided Dr. Pieter de Jong (Lawrence Livermore) with PCR primers in order to detect sequence tagged sites (short, unique sequences of DNA that can be recognized by PCR assay) at the D19S118 and D19S112 loci. He used these primers to identify two overlapping yeast artificial chromosome (YAC) clones which together cover the region between D19S112 and D19S119 from which a primary cosmid library was produced (Aslanidis et al., 1992). Subsequently, we were able to bridge the gap between the distal end of our centromeric walk and the proximal end of Dr. Wieringa's telomeric walk, which corresponded to about 50 kb of genomic DNA. The collaborative effort of our lab with Dr. Wieringa's and Dr. de Jong's groups had resulted in the cloning of the complete DM gene region (approximately 700 kb), which enabled us to identify the genetic defect associated with DM. Figure 1-1 provides a summary physical and genetic map of the DM region.

Identification and characterization of the DM mutation and gene

The precise genetic and physical definition of the DM essential region greatly limited the region in which a genuine candidate gene would be contained. Genomic and cDNA probes isolated from this region were found to detect an unstable 10 kb EcoRI genomic fragment in individuals affected with DM (Aslanidis et al., 1992; Buxton et al., 1992; Harley et al., 1992). The physical map location and genetic characteristics of this unstable element were consistent with it being the cause of DM. Using Southern blot analysis, two different DNA polymorphisms mapping within the 10 kb EcoRI genomic fragment were identified (Figure 1-2A; Mahadevan et al., 1992; Aslanidis et al., 1992). The polymorphisms were detected by two genomic probes (pGB2.6 and pGP1.5) that mapped to this EcoRI fragment. In normal individuals, these probes detected an 8.5 or 9.5 HindIII or a 9.0 or 10.0 kb EcoRI insertion/deletion polymorphism. This 1 kb variation was due to the presence or absence of three contiguous *Alu* elements within a 150 bp PstI-BglI fragment (Figure 1-2A; Mahadevan et al., 1993a). The second DNA polymorphism mapped to a 1.5 kb BamHI fragment located 5.0 kb centromeric to the *Alu* element polymorphism (Figure 1-2A) and was represented by varying degrees of allelic expansion in DM affected individuals. Genetic phasing of DM families revealed that the allele with the 1 kb insertion (9.5 kb HindIII or 10.0 kb EcoRI fragment) was not only the expanding allele in DM individuals but also in absolute linkage disequilibrium with DM (Mahadevan et al., 1993a). DNA sequence analysis revealed that the unstable genomic fragment corresponded to an amplification of a $(CTG)_n$ trinucleotide repeat (Figure 1-2B; Mahadevan et al., 1992; Aslanidis et al., 1992; Buxton et al., 1992; Harley et al., 1992; Brook et al., 1992; Fu et al., 1992).

Figure 1-1 Physical and genetic map of the DM region at chromosome 19q13.3. The map shows the position of various genetic loci, relative to the position of overlapping YAC clones and a proximal 350 kb (A) and distal 150 kb (B) cosmid contig. The cosmid contigs (black bars) were generated by chromosome walking, starting from the proximal ERCC1 locus and the distal D19S112 locus. YAC clones (grey bars) were used to bridge the unclonable gap between the two cosmid contigs. The essential region containing the DM locus was determined to be located in the 350 kb area between the nearest crossover markers, D19S118 and D19S112. ERCC1 and D19S51 were previously the closest flanking markers. The centromeric (CEN) and telomeric (TER) directions are indicated.

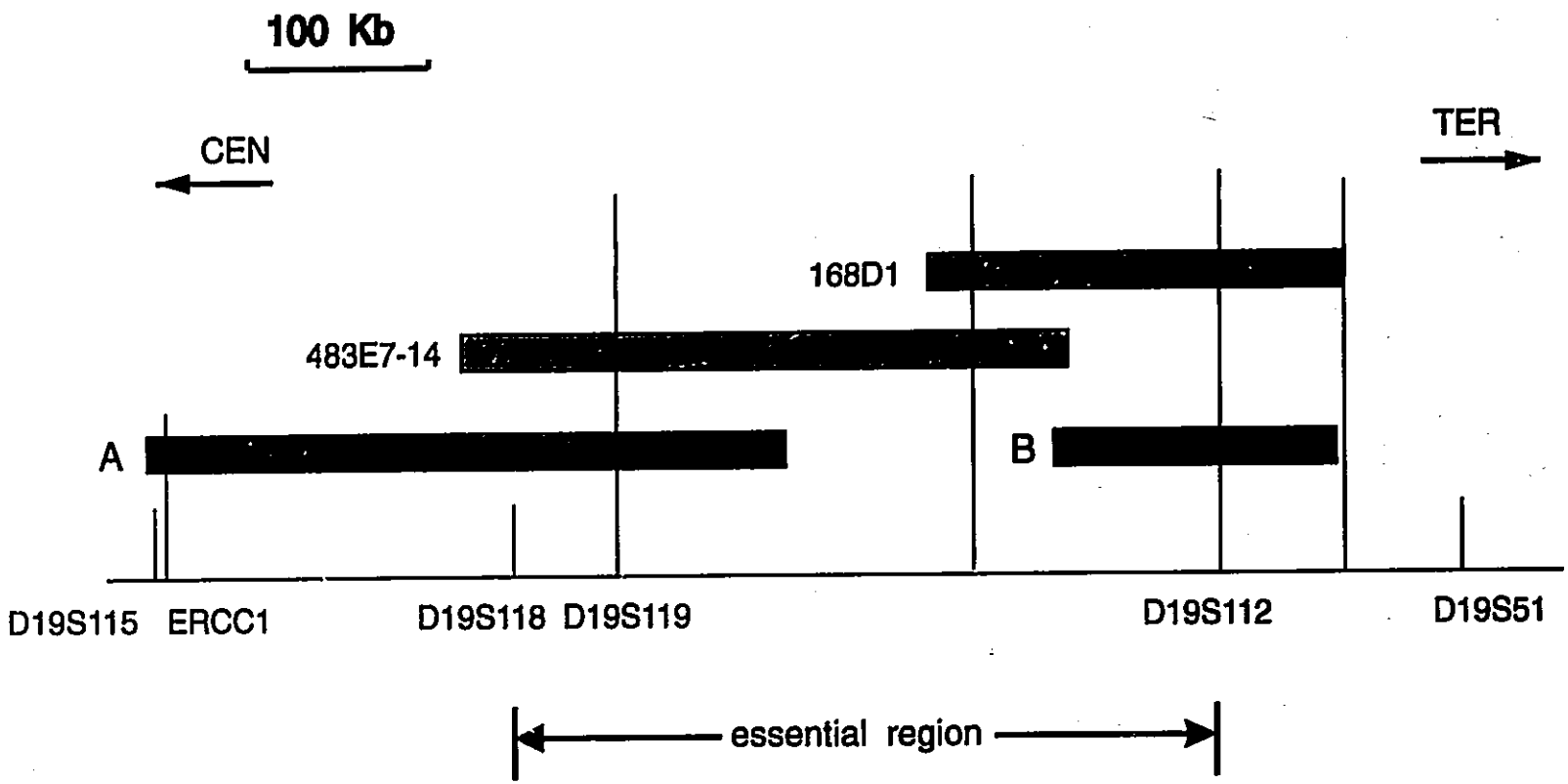
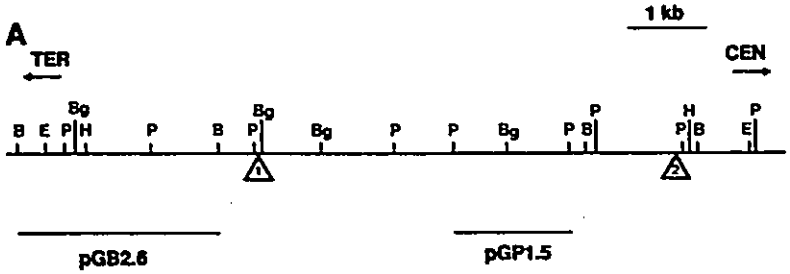

100 Kb
CEN
TER
168D1
483E7-14
A
B
D19S115
ERCC1
D19S118
D19S119
D19S112
D19S51
essential region

Figure 1-2A Restriction map of the unstable genomic segment at the DM locus. The normal insertion/deletion *Alu* element polymorphism and the allelic expansion polymorphism (represented by triangles 1 and 2, respectively) were detected by two genomic probes, pGB2.6 and pGP1.5. The orientation of the map is from telomere (TER) to centromere (CEN), and the position of relevant restriction sites are given: B, BamHI; Bg, BglI; E, EcoRI; H, HindIII and P, PstI.

Figure 1-2B Nucleotide sequence of the 3' region of a cDNA clone. The clone was isolated from a human heart cDNA library (Stratagene) using the 1.5 kb BamHI fragment containing the allelic expansion polymorphism as a probe. The sequence shown is the last 60 bp of the penultimate 3' exon (top line) followed by the entire 875 bp sequence of the 3' most exon. The nucleotide position designated by the number 1 corresponds to the first base of the last exon. A consensus polyadenylation signal sequence is underlined. Relevant restriction enzymes are shown above their recognition sequences. Primers used for sequencing and PCR amplification are underlined and numerically designated. Indicated in bold letters are the five $(CTG)_n$ repeats located between primers 409 and 410 found in this cDNA. Plasmid DNA was prepared by routine alkaline lysis methodology. Dideoxy chain termination reactions with vector primers and internal primers were performed with fluorescent dye-labeled dideoxynucleotides, according to manufacturer's specifications (Applied Biosystems), and subsequently run on an ABI 373A automated sequencer.



B

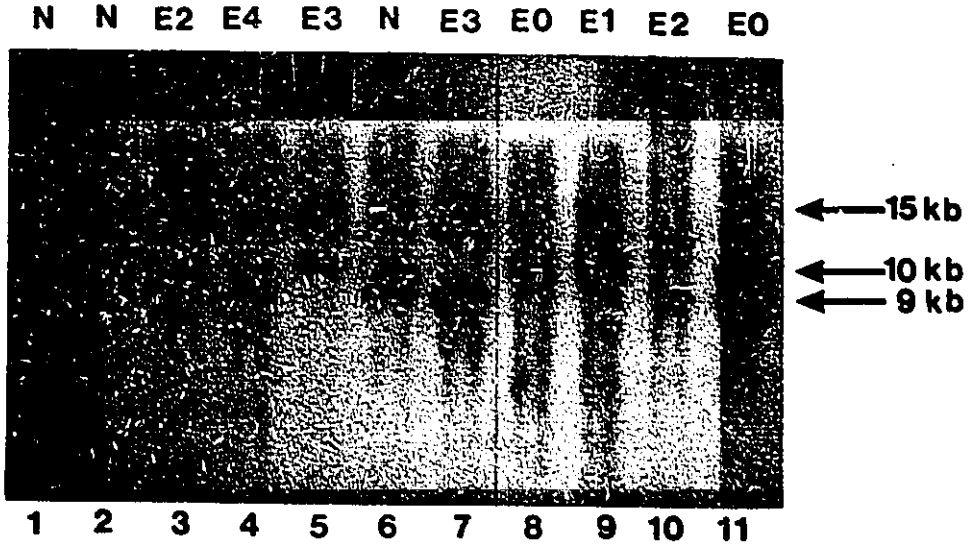
```
AGTGCCCGCTGGTGGGGCCAGGCCCCATGCACCGCCGCCACCTGCTGCTCCCTGCCAGGG
1   TCCCTAGGCCTGGCCTATCGGAGGCGCTTTCCCTGCTCCTGTTCGCCGTTGTTCTGTCTC
61  GTGCCGCCGCCCTGGGCTGCATTGGGTTGGTGGCCCACGCCGGCCAACTCACCGCAGTCT
121 GGCGCCGCCCAGGAGCCGCCCGCGCTCCCTGAACCCTAGAACTGTCTTCGACTCCGGGGC
181 CCCGTTGGAAGACTGAGTGCCCGGGGCACGGCACAGAAGCCGCGCCCACCGCCTGCCAGT
241 TCACAACCGCTCCGAGCGTGGGTCTCCGCCCAGCTCCAGTCCTGTGATCCGGGCCCGCCC
                                                    primer 409 ▷
301 CCTAGCGGCCGGGGAGGGAGGGGCCGGGTCCGCGGCCGGCGAACGGGGCTCGAAGGGTCC
                                          ◁ primer 410
361 TTGTAGCCGGGAATGCTGCTGCTGCTGCTGGGGGGGATCACAGACCATTTCTTTCTTTCGG
                                          Pst I
421 CCAGGCTGAGGCCCTGACGTGGATGGGCAAACTGCAGGCCTGGGAAGGCAGCAAGCCGGG
    ◁ primer 406                                          Hind III
481 CCGTCCGTGTTCCATCCTCCACGCACCCCCACCTATCGTTGGTTCGCAAAGTGCAAAGCT
      ◁ primer 407
541 TTCTTGTGCATGACGCCCTGCTCTGGGGAGCGTCTGGCGCGATCTCTGCCTGCTTACTCG
                                    Bam HI
601 GGAAATTTGCTTTTGCCAAACCCGCTTTTTCGGGGATCCCGCGCCCCCCTCCTCACTTGC
661 GCTGCTCTCGGAGCCCCAGCCGGCTCCGCCCGCTTCGGCGGTTTGGATATTTATTGACCT
721 CGTCCTCCGACTCGCTGACAGGCTACAGGACCCCCAACAACCCCAATCCACGTTTTGGAT
781 GCACTGAGACCCCGACATTCCTCGGTATTTATTGTCTGTCCCCACCTAGGACCCCCACCC
841 CCGACCCTCGCGAATTAAAGGCCCTCCATCTGCCC[A]n
```

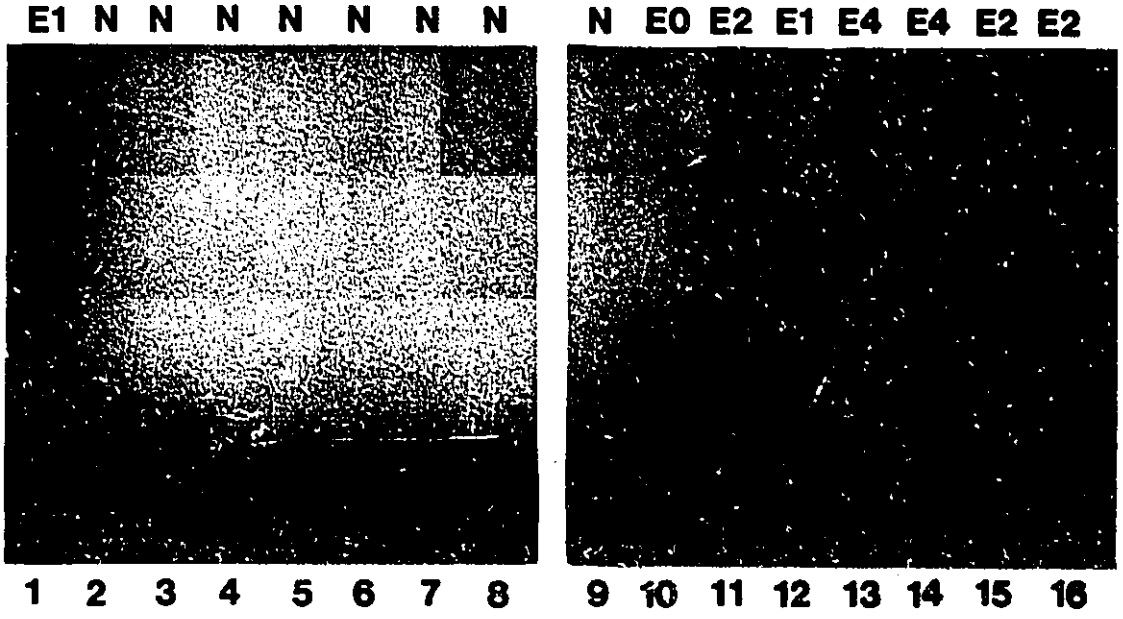
Amplification of the $(CTG)_n$ repeat is reflected in the different levels of allelic expansion observed by Southern blot analysis in DM affected individuals (Figure 1-3). However, approximately 30% of DM individuals (affected status verified by DNA typing) show no allelic expansion at the Southern blot level (Mahadevan et al., 1992). As depicted in Figure 1-3, these DM individuals could be mistakenly assumed to be normal. To circumvent this problem, an assay using Southern blot analysis of PCR-amplified genomic DNA from normal and DM-affected individuals was developed (Figure 1-4; Mahadevan et al., 1992). Southerns were probed with a synthetic $(CTG)_{10}$ oligonucleotide. PCR-amplified DNA from DM individuals showed a distinct smearing of the hybridization signal, presumably due to heterogeneity of the expanded allele. In contrast, the oligonucleotide probe hybridized only to the normal alleles in unaffected individuals. Individuals who showed no expansion at the Southern blot level showed clear expansion with the PCR-based assay. In some PCR-amplified samples from DM individuals, the hybridization smear produced by the labeled $(CTG)_{10}$ oligonucleotide was faint and could be mistaken for that of an unaffected individual. However, examination of Southern blot analysis of EcoRI- or HindIII-digested genomic DNA probed with pGB2.6 or pGP1.5 revealed the presence of a greatly expanded allele in these individuals. Presumably, a greatly increased number of $(CTG)_n$ repeats was refractory to PCR amplification under the conditions used (Mahadevan et al., 1992). Thus, appropriate molecular diagnosis of DM should include both Southern blot analysis of genomic DNA and the hybridization analysis of PCR-amplified DNA. Only a few of our DM families (2 of 98) were negative by both types of analysis. These families may represent unique mutations of the DM gene that do not involve an increase in the number of $(CTG)_n$ repeats (Mahadevan et al., 1992).

In unaffected individuals, the $(CTG)_n$ repeat is both mitotically and meiotically stable, although highly polymorphic ranging from 5 to 35 repeats. In individuals affected with DM, the $(CTG)_n$ repeat can be unstable and variation in repeat lengths may be

Figure 1-3 Southern blot analysis showing varying degrees of allelic expansion in DM-affected individuals. EcoRI allele sizes of 9 kb and 10 kb are found in the normal (N) population (lanes 1, 2 and 6). Varying degrees of expansion of the 10 kb allele are seen in DM individuals, ranging from E1 (expansion of 0 to 1.5 kb; lane 9) to E2 (expansion of 1.5 to 3.0 kb; lanes 3 and 10) to E3 (expansion of 3.0 to 4.5 kb; lanes 5 and 7) to E4 (expansion of 4.5 to 6.0 kb or more; lane 4). However, approximately 30% of DM individuals show no allelic expansion at the Southern blot level (designated E0; lanes 8 and 11) and may therefore be mistaken for normal individuals. Genomic DNA (5 μg) was digested with EcoRI, separated by electrophoresis on 0.8% agarose gels, and transferred to nylon membranes. Southerns were probed with the BamHI-EcoRI fragment of pGB2.6, washed in 0.2X SSC with 0.1% SDS at 65°C, and exposed to x-ray film for 1 to 4 days.

Figure 1-4 Southern blot analysis of PCR-amplified genomic DNA from various normal (N) and DM-affected (E) individuals. Southerns were probed with a synthetic $(CTG)_{10}$ oligonucleotide. PCR-amplified DNA from DM individuals showed a distinct smearing of the hybridization signal. In contrast, the oligonucleotide probe hybridized only to the normal alleles in unaffected individuals. Individuals who showed no expansion (E0) at the Southern blot level demonstrated clear expansion with the PCR-based assay (lane 10). In some PCR-amplified samples from DM individuals, the hybridization smear produced by the labeled $(CTG)_{10}$ oligonucleotide was faint and could be mistaken for that of an unaffected individual (lane 13). However, examination of Southern blot analysis of EcoRI- or HindIII-digested genomic DNA probed with pGB2.6 or pGP1.5 revealed the presence of a greatly expanded allele (E4) in these individuals. Genomic DNA (2 μg) was PCR-amplified with primers 406 and 409 by a standard protocol and 30 cycles of 94 °C for 1 min., 60 °C for 1 min., and 72 °C for 1.5 min.. Amplified products were separated by electrophoresis on 1% agarose gels, transferred onto nylon membranes, and probed with a labeled $(CTG)_{10}$ oligonucleotide. Membranes were washed in 6X SSC with 0.1% SDS at 40°C and exposed to x-ray film overnight.


N N E2 E4 E3 N E3 E0 E1 E2 E0
15 kb
10 kb
9 kb
1 2 3 4 5 6 7 8 9 10 11

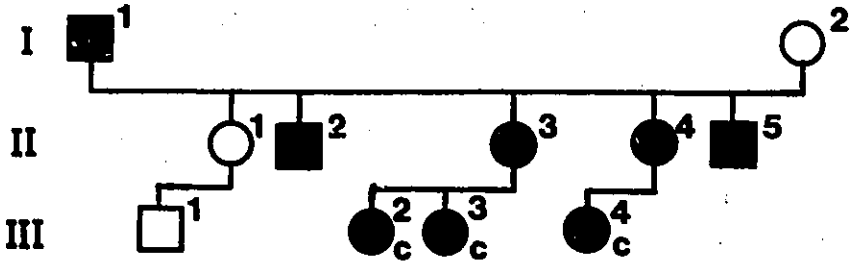

E1 N N N N N N N
N E0 E2 E1 E4 E4 E2 E2
1 2 3 4 5 6 7 8
9 10 11 12 13 14 15 16

extreme, even within a given family. Mildly affected patients possess a minimum of 50 to 80 $(CTG)_n$ repeats while severely affected individuals have up to 2000 or more copies. In fact, amplification of the repeat appears to be the molecular basis for genetic anticipation. Specifically, amplification of the $(CTG)_n$ repeat appears to correlate with increasing severity of the disease. A positive correlation has been found to exist between increased $(CTG)_n$ repeat length and increased frequency of congenital DM cases (Tsilfidis et al., 1992) and early age of onset in adult DM (Hunter et al., 1992a). Figure 1-5 represents a three generation DM family which exemplifies the phenomenon of genetic anticipation.

Additional evidence that amplification of the $(CTG)_n$ repeat is the mutation that causes DM comes from the detailed molecular analysis of the loss of the DM mutation in an individual from one of our large French Canadian pedigrees (O'Hoy et al., 1993). We observed a reduction (to within the normal range in trinucleotide $(CTG)_n$ repeats) during the transmission of a DM chromosome from father to daughter from this extensive DM kindred. Substantial evidence from the haplotype data of size polymorphic markers within the DM gene and 18 RFLPs bracketing the DM locus supported the concept that a rare gene conversion event was involved in the reduction in $(CTG)_n$ repeat copies in the daughter's DM chromosome. Moreover, the loss or reversal of the mutation was associated with the corresponding loss of the DM phenotype (Hunter et al., 1993). Specifically, the daughter was found to be clinically asymptomatic at an age at which her father was clearly affected with DM. The results of this and other genotype/phenotype correlative studies (Hunter et al., 1992; Tsilfidis et al., 1992) provide strong evidence that amplification of the unstable repeat is the basis for the variability in the clinical expression of DM.

The $(CTG)_n$ repeat was found to be located in the 3' untranslated region of a gene encoding a putative cAMP-dependent serine-threonine protein kinase (Mahadevan et al., 1993; Brook et al., 1992; Fu et al., 1992). The primary sequence characterization of the

Figure 1-5 Pedigree and Southern blot analysis of a three generation DM family. Normal individuals (white symbols) display polymorphic HindIII fragments of 8.5 and 9.5 kb. DM-affected males (black squares) and females (black circles) in generations I and II showed both clinical and electromyographic evidence of myotonia. Three affected individuals born with the congenital form of the disease (designated by 'c') are shown in generation III. The DM affected grandfather (individual I-1) displayed enlarged HindIII bands cosegregating with the DM chromosome. The affected individuals of generation II have all inherited expanded alleles, slightly larger than those seen in individual I-1. All these congenital DM patients of generation III possess even larger DM associated HindIII alleles (up to 15 kb). This demonstrated increase in disease severity and earlier age of onset in successive generations clearly illustrates the phenomenon of genetic anticipation. Genomic DNA (5μg) was digested with HindIII, separated by electrophoresis on 0.8% agarose gels, and transferred to nylon membranes. Southerns were probed with pGP1.5, washed in 0.2X SSC with 0.1 % SDS at 65 °C, and exposed to x-ray film for 1 to 4 days.


I
1
2
II
1
2
3
4
5
III
1
2
c
3
c
4
c

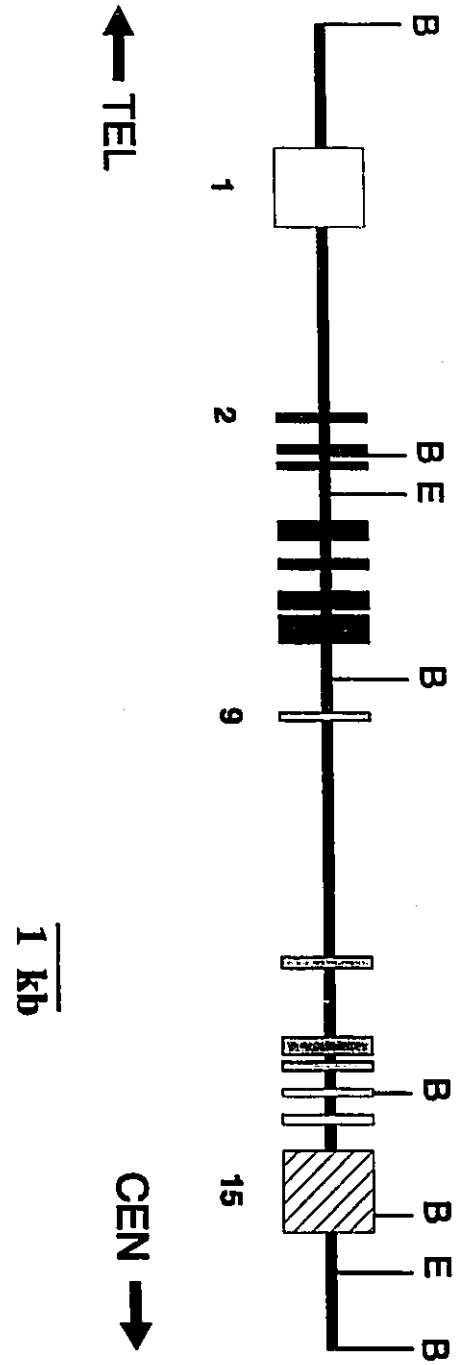

15.0
9.5
8.5

entire DM kinase gene and its corresponding mRNAs has been completed (Mahadevan et al., 1993b; Jansen et al., 1992b). Our laboratory had undertaken the analysis of the human gene, while Dr. Wieringa's group has characterized the mouse gene. In both species, the DM kinase gene consists of 15 exons with a full length mRNA size of approximately 3.4 kb. Alternate splicing of the mRNA in both species has been found (Jansen et al., 1992b) but its significance has yet to be determined. The predicted translation product of the largest human isoform is 629 amino acids with an approximate molecular weight of 69,300 Daltons (Mahadevan et al., 1993b). A schematic representation of the human DM kinase gene is shown in Figure 1-6.

A biochemical abnormality of the cell membrane in DM has been proposed as the basis for the multisystemic nature of this disorder (Harper, 1989). Since increased Ca^{2+} uptake in DM has been reported in sarcoplasmic vesicles of muscle tissue and erythrocytes, a generalized membrane abnormality of Ca^{2+} transport may be present in DM (Hockings et al., 1993). We have obtained experimental evidence that the greatly amplified repeat results in a significant increase in steady state levels of the mutant mRNA (Sabourin et al., 1993). Since such enzymes activate voltage-dependent Ca^{2+} channels by phosphorylation of a regulatory protein, the postulated Ca^{2+} channel dysfunction in DM may result directly from inherited alterations in the DM protein kinase.

In summary, DM is caused (in over 99% of all cases) by the expansion of a $(CTG)_n$ trinucleotide repeat in the 3' untranslated region of the DMK gene (Mahadevan et al., 1992; Brook et al., 1992; Fu et al., 1992). A striking, total linkage disequilibrium was observed in both Caucasian and Japanese patients, between the DM mutation and a two allele insertion/deletion *Alu* element polymorphism located 5 kb telomeric to the $(CTG)_n$ repeat, suggesting a single common origin of the mutation (Mahadevan et al., 1993a; Yamagata et al., 1992; Harley et al., 1992). This finding was unexpected for a dominant disease which, in its severe form diminishes or abolishes (in the case of

Figure 1-6 Structural organization of the human DM kinase gene. The gene maps within a 14 kb genomic region contained in five contiguous BamHI fragments. The positions of relevant restriction sites are given: B, BamHI; E, EcoRI. The orientation of the gene is from telomere (TEL) to centromere (CEN). The exons are depicted as boxes and exons 1, 2, 9 and 15 are numbered. Exons 2-8 (black boxes) encode the protein kinase domains, exons 9-12 (stippled boxes) encode the α-helical (coiled-coil) domains, and exon 15 (striped box) encodes the hydrophobic, possibly transmembranous domain and contains the $(CTG)_n$ repeat.


TEL
CEN
1
2
9
15
B
E
1 kb

congenital DM) reproductive fitness. Such diseases are in general characterized by a high level of new mutations, which compensate for the loss of abnormal alleles due to decreased fitness. Since all cases of DM are familial and there has been no reported spontaneous occurrence of the disease, it is difficult to rationalize the relatively stable incidence of DM in the global population. It was therefore suggested that DM could be due to recurrent mutations occurring on a predisposing allelic form of the normal gene (Harley et al., 1992; Barceló et al., 1993). Such predisposition may lie in the $(CTG)_n$ repeat itself, which is polymorphic in the normal population, or in neighbouring sequences including a putative direct effect of the 1 kb insertion/deletion *Alu* element polymorphism. In the case of Fragile X syndrome, which is caused by the unstable expansion of a $(CGG)_n$ repeat (Fu et al., 1991; Kremer et al., 1991; Oberlé et al., 1991), a significant but much weaker linkage disequilibrium has been observed using adjacent $(CA)_n$ repeat markers, suggesting multiple original events (Richards et al., 1992; Oudet et al., 1993).

Therefore, a paradox exists in DM in which the incidence of the disease has remained relatively stable despite the influence of genetic anticipation and evidence to suggest it should be declining. The objective of my research and the focus of this thesis was thus to determine the origin of the DM mutation. In order to address this issue, I have used PCR-based assays of nine polymorphisms, spanning a physical distance of 30 kb, within and immediately flanking the DM kinase gene, to examine patterns of allelic association with respect to the DM mutation. In addition, high resolution haplotype analysis employing these markers was conducted in order to predict the origin and number of DM mutations segregating in our heterogeneous study population.

CHAPTER II THE ORIGIN OF THE MYOTONIC DYSTROPHY MUTATION

Introduction

The mutation causing myotonic dystrophy has been identified as an expansion of an unstable trinucleotide $(CTG)_n$ repeat (Mahadevan et al., 1992; Aslanidis et al., 1992; Buxton et al., 1992; Harley et al., 1992; Brook et al., 1992; Fu et al., 1992) in over 99% of the global DM population encompassing diverse ethnic groups. The presence of strong linkage disequilibrium between DM and the three most common extended haplotypes, generated from 12 19q13.3 loci spanning a physical distance of 1.3 Mb, was observed in our study population (Tsilfidis et al., manuscript in preparation). In order to assay for the potential presence of a DM founder effect antedating that originating from our French Canadian DM population (Korneluk et al., 1989b; MacKenzie et al., 1989), all distinct extended haplotypes were counted only once regardless of the number of such haplotypes observed. This approach of collapsing the common haplotypes resulted in the loss of linkage disequilibrium between all markers and DM except for pKBE0.8 (located 160 kb centromeric to DM) and pGB2.6 (located within the DM kinase gene). The marker pGB2.6 identifies the molecular defect responsible for DM which is represented by varying degrees of allelic expansion in DM-affected individuals using Southern blot analysis. In addition, DM is in complete linkage disequilibrium with an *Alu* element polymorphism (Mahadevan et al., 1993a), identified by pGB2.6, which is located 5 kb telomeric to the $(CTG)_n$ repeat within the DM kinase gene. Linkage disequilibrium at both pKBE0.8 and pGB2.6 suggests the presence of a limited number of ancestral mutations causing DM. In addition, genealogical reconstruction of 88 DM-affected extended families in the Saguenay-Lac-Saint-Jean region resulted in a unique couple ancestor illustrating the genetic homogeneity of the DM mutation in this area of Quebec (Mathieu et al., 1990). Therefore, the presence of common extended haplotypes in our

DM population (Tsilfidis et al., manuscript in preparation) coupled with those dating from the late 1600's (Mathieu et al., 1990), supports an ancestral mutation model for DM.

Consequently, the presence of anticipation and the resulting loss in reproductive fitness in DM pedigrees would suggest that the incidence of DM should be declining. If the incidence is not, in fact, decreasing then the possibility exists that the DM mutation may be more common than previously supposed, and may be passed down in a stable and clinically inconsequential manner for many generations. Alternatively, the incidence of DM could be due to recurrent mutations occurring on a predisposing allelic form of the normal gene (Harley et al., 1992; Barceló et al., 1993). Our data suggest that either DM is a consequence of one or few ancestral mutations or alternatively, it is the result of a predisposing haplotype. However, these two models are not necessarily mutually exclusive and DM may indeed reflect the influence of both mechanisms. This hypothesis can be tested by examining patterns of linkage disequilibrium in the immediate vicinity of DM.

Recently, a study using a $(CA)_n$ repeat polymorphism (Jansen et al., 1992a) and the *Alu* element polymorphism in conjunction with the $(CTG)_n$ repeat, was reported by Imbert et al. (1993) which concluded that one or few ancestral predisposing mutations at the $(CTG)_n$ repeat occurred randomly thus becoming a predisposing haplotype for DM. The present study is an extension of preliminary results previously reported (Neville et al., 1992) using DM families from a different and diverse population. We conducted a more extensive haplotype analysis at the DM locus, which both complements and extends the work of Imbert et al.. We have used PCR-based assays of nine polymorphisms, spanning a physical distance of 30 kb, within and immediately flanking the DM kinase gene, to examine patterns of allelic association with respect to the DM mutation. Our study also included a three allele insertion/deletion polymorphism at the D19S119 (pKBE0.8) locus (Korneluk et al., 1991), a region which is located 160 kb centromeric to the DM gene. Significantly, RFLPs detected by markers mapping to this 19q13.3 region

were the first to demonstrate pronounced linkage disequilibrium with DM (Korneluk et al., 1991; Harley et al., 1991). In addition, high resolution haplotype analysis employing these markers was conducted in order to predict the origin and number of DM mutations segregating in our heterogeneous study population. Our results suggest that the DM mutation occurred on the background of a particular haplotype in which the $(CTG)_n$ repeat became inherently unstable and therefore predisposed to amplification.

Materials and Methods

Families

A total of 200 families with myotonic dystrophy were studied, of which 54 were used for extensive haplotype analysis. Approximately half of these families were French Canadian; the others were representative of the ethnic mosaicism which comprises the Canadian population. The clinical diagnosis of DM was as previously described (MacKenzie et al., 1989). A significant number of family members from a total of 54 affected families were typed to allow generation of the complete DM chromosome haplotype by chromosome phasing. Unaffected individuals from DM and non-DM families constitute the source of non-DM chromosomes.

DNA analysis

Genomic DNA was extracted from either peripheral blood samples or lymphoblast cultures by solvent extraction using a modification of Madisen et al. (1987) or by the salt precipitation method of Miller et al. (1988).

PCR-based assays of polymorphisms

DraIII polymorphism: A primer with an altered base introduced at position 21 (T→A) was synthesized in order to create a DraIII site in the presence of a polymorphic base

change (G→T). Genomic DNA (2 µg) was PCR amplified with 50 ng each of primers 839 **(5'AGG GCC CCT CAT CAA AGT CCA CGG TGT 3')** and 823 **(5'ACG GTT CTG CAG AGT GGA AGT 3')**, 1.5 mM $MgCl_2$, 1X Cetus reaction buffer and 50 µM of each dNTP in a 30 µL reaction volume. Amplification was conducted for 10 cycles at 94°C for l min., 60°C for 1.5 min., and 72°C for 1.5 min. followed by 25 cycles at 94°C for 1 min., 55°C for 1.5 min., and 72°C for 1.5 min.. Amplified products were digested with DraIII (BRL) for one hour and subsequently separated by electrophoresis, using 1X TBE buffer, in 1% agarose, 2% NuSieve gels containing ethidium bromide.

HphI polymorphism: Due to the proximity of an HphI site and the polymorphic HphI site, this polymorphism could not be resolved. Therefore, a primer with an altered base introduced at position 21 (T→G) was synthesized to create a BanI site when the polymorphic HphI site was present. Genomic DNA (2 µg) was PCR amplified with 50 ng each of primers 463 **(5'CCG TCT CCA CTC TGT CTC ACT 3')** and 600 **(5'GCT CTT GTC CCT CTT CCT AGG C3')**, 1.5 mM $MgCl_2$, 1X Cetus reaction buffer and 50 µM of each dNTP in a 30 µL reaction volume. Amplification was conducted for 10 cycles at 94°C for 1 min., 60°C for 1.5 min., and 72°C for 1.5 min. followed by 25 cycles at 94°C for 1 min., 55°C for 1.5 min., and 72°C for 1.5 min.. Amplified products were digested with BanI (BRL) for one hour and subsequently separated by electrophoresis (BRL Horizon 58), using 1X TBE buffer, in 3.5 % NuSieve gels containing ethidium bromide.

HhaI polymorphism: Genomic DNA (2 µg) was PCR amplified with 50 ng each of primers 445 **(5'GAC CTG CTG ACA CTG CTG AGC3')** and 472 **(5'GTG CCT TCC ATC CCT CAT CAG3')**, 1.5 mM $MgCl_2$, 1X Cetus reaction buffer and 50 µM of each dNTP in a 30 µL reaction volume, using 30 cycles of amplification at 94°C for 1 min., 60 °C for 1.5 min., and 72°C for 1.5 min.. Amplified products were digested with HhaI

(BRL) for one hour, and subsequently separated by electrophoresis, using 1X TBE buffer, in 1% agarose, 1% NuSieve gels containing ethidium bromide.

Insertion/deletion polymorphism: Genomic DNA (2 μg) was PCR amplified with 50 ng each of primers 405 **(5'CTG TAT ACT CAG CTA CTA GGG T3')** and 491 **(5'AAA TAG GCT GGA CCG CGG3')** and 100 ng of primer 486 **(5'GCC ACT TTA GAT AAC CCC TGA G3')**, 1.5 mM $MgCl_2$, 1X Cetus reaction buffer, 200 μM of each dNTP, in a 50 μL reaction volume using 30 cycles of amplification at 94°C for 1 min., 60°C for 1.5 min., and 72°C for 1.5 min.. Amplified products were separated by electrophoresis, using 1X TBE buffer, in 1% agarose gels containing ethidium bromide.

HinfI polymorphism: Genomic DNA (2 μg) was PCR amplified with 50 ng each of primers 458 **(5'CTG CAG AAG GTT TAG AAA GAG C3')** and 424 **(5'CAT CCT GTG GGG ACA CCG AGG 3')**, 1.5 mM $MgCl_2$, 1X Cetus reaction buffer and 50 μM of each dNTP in a 30 μL reaction volume. Amplification was conducted for 10 cycles at 94°C for 1 min., 60°C for 1.5 min., and 72°C for 1.5 min. followed by 25 cycles of amplification at 94°C for 1 min., 55°C for 1.5 min., and 72°C for 1.5 min.. Amplified products were digested with HinfI (BRL) for one hour, and subsequently separated by electrophoresis, using 1X TBE buffer, in 1% agarose, 2% NuSieve gels containing ethidium bromide.

BpmI polymorphism: Genomic DNA (2 μg) was PCR amplified with 50 ng each of primers 458 **(5'CTG CAG AAG GTT TAG AAA GAG C 3')** and 424 **(5'TCA TCC TGT GGG ACA CCG AGG 3')**, 1.5 mM $MgCl_2$, 1X Cetus reaction buffer and 50 μM of each dNTP in a 30 μL reaction volume. Amplification was conducted for 10 cycles at 94°C for 1 min., 60°C for 1.5 min., and 72°C for 1.5 min. followed by 25 cycles at 94°C for 1 min., 55°C for 1.5 min., and 72°C for 1.5 min.. Amplified products were digested

with BpmI (BRL) for one hour and subsequently separated by electrophoresis, using 1X TBE buffer, in 3.5% NuSieve gels containing ethidium bromide.

Fnu4HI polymorphism: Genomic DNA (2 μg) was PCR amplified with 50 ng each of primers 473 **(5'AGC CAC AGG CAG CCT TAA GC3')** and 474 **(5'TCC GGG GAA GGG GAC ACA TGA3')**, 1.5 mM $MgCl_2$, 1X Cetus reaction buffer and 50 μM of each dNTP in a 30 μL reaction volume using 30 cycles of amplification at 96°C for 1 min., 60°C for 1.5 min., and 72°C for 1.5 min.. Amplified products were digested with Fnu4HI (BRL) for one hour and subsequently separated by electrophoresis, using 1X TBE buffer, in 3.5% NuSieve gels containing ethidium bromide.

Allele-specific oligonucleotide analysis to detect G/T polymorphism (intron 14): Genomic DNA (0.5 μg) was PCR amplified with 50 ng each of primer 490 **(5'CTC CGA TCG GGT CAC CTG TC3')** and either primer 505 **(5'CGC AGC TAA GCG GGT GGC AA3')** or 506 **(5'CGC AGC TAA GCG GGT GGC AC3')**, 1 mM $MgCl_2$, 1X Cetus reaction buffer and 50 μM of each dNTP in a 30 μL reaction volume using 30 cycles of amplification at 94°C for 1 min., 65°C for 1.5 min., and 72°C for 1.5 min.. The reaction was placed at 94°C for 2 min. prior to addition of primers to eliminate non-specific priming. Amplified products were visualized by electrophoresis using 1X TBE buffer, in 1% agarose gels containing ethidium bromide.

TaqI polymorphism: Genomic DNA (2 μg) was PCR amplified with 50 ng each of primers 565 **(5'TCT AGA ACA CAT GGG AAA TAA TGT3')** and 564 **(5'ATC TAA ATG CTA AGT GGC AAG TGT3')**, 1.5 mM $MgCl_2$, 1X Cetus reaction buffer and 50 μM of each dNTP in a 30 μL reaction volume. Amplification was conducted for 10 cycles at 94°C for 1 min., 60°C for 1.5 min., and 72°C for 1.5 min. followed by 25 cycles at 94°C for 1 min., 55°C for 1.5 min., and 72°C for 1.5 min.. Amplified products were

digested with TaqI (BRL) for one hour, and subsequently separated by electrophoresis, using 1X TBE buffer, in 1% agarose, 2% NuSieve gels containing ethidium bromide.

Statistical analysis

The computer program EPISTAT was used to perform a χ^2 test of association between DM and each of the nine polymorphic markers employed in this study. The test was based on the allelic frequencies of each marker in the normal population in relation to that observed in the DM population. The raw data used in the calculations is provided in Appendix II.

Repeat sizing in normal individuals

The procedure employed was previously reported (Barceló et al., 1993), however more recent $(CTG)_n$ repeat sizing was conducted by direct visualization of PCR products in 4% Metaphor (Mandel) agarose gels containing ethidium bromide after electrophoresis in 1X TBE.

Results

In order to propose a model for the origin of DM, a high resolution genetic map of the DM locus was constructed using polymorphisms that we had discovered during the sequencing of the DM gene (Mahadevan et al., 1993b; Genbank accession no. L08835). In addition to an *Alu* deletion polymorphism (Mahadevan et al., 1993a) and three DNA sequence polymorphisms (Mahadevan et al., 1993b; O'Hoy et al., 1993) previously reported, five novel DNA sequence polymorphisms at the DM locus are presented which have been adapted for PCR assay (Figures 2-1 to 2-5). The frequency of alleles in the normal population are provided in Table 2-1. These nine polymorphisms map within a 30 kb area of 19q13.3 encompassing the entire DM kinase gene as well as flanking

genomic sequences including the 3' end of the adjacent DMR-N9 gene (Figure 2-6; Mahadevan et al., 1993b; Jansen et al., 1992a).

The *Alu* deletion polymorphism located in intron 8 of the DM kinase gene was initially detected on Southern blots of EcoRI restricted genomic DNA by the probe pGB2.6 (Mahadevan et al., 1992; Aslanidis et al., 1992). Upon sequencing this genomic region, it was apparent that the 1 kb difference between the two alleles was due to the deletion of three of five contiguous *Alu* repeats. Accordingly, primers were constructed flanking the deletion (Mahadevan et al., 1992). The TaqI polymorphism located approximately 15 kb centromeric to the DM gene was initially detected on Southern blots of TaqI restricted genomic DNA by pCN400, a 400 bp BamHI/EcoRI genomic fragment. Hybridization of the probe, pCN400, to restricted genomic clones yielded the location of the TaqI polymorphism which was subsequently adapted to PCR. The novel polymorphism at D19S463 identified by pCN400 and the previously reported polymorphism DMK (G/T) (O'Hoy et al., 1993) which flank the $(CTG)_n$ repeat are now the closest reported proximal and distal markers respectively to the DM mutation.

Significant allelic association was observed between each marker and DM with chi-square values of 33 or greater ($p<10^{-6}$). Four main haplotypes (A-D) were observed in the normal population (N=143) using these markers at the DM locus with frequencies of 49%, 27%, 16% and 8% respectively (Table 2-2). In addition, five derivative haplotypes (E-I) were observed only once (Table 2-2). These divergent haplotypes could have been generated by recombinational events at this locus. DM is in complete association with haplotype A, which is also the most common haplotype found in the normal population.

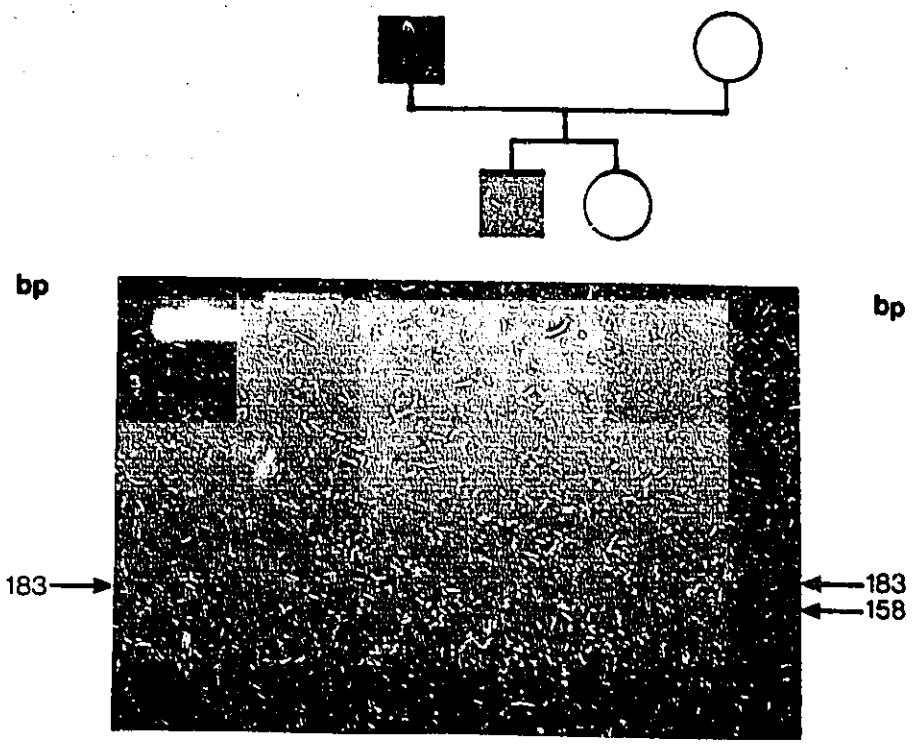
In an effort to delineate the relationship between the DM haplotype and the copy number of the $(CTG)_n$ repeat in normal individuals, PCR analysis of the $(CTG)_n$ repeat length in normal individuals with the four most common haplotypes (N=121) was

undertaken (Table 2-3). Individuals with haplotype A displayed the full spectrum of $(CTG)_n$ number, ranging from 5 to 35 repeats. Although the majority of $(CTG)_n$ alleles on haplotype A are represented by either 5 repeats (73%) or greater than 19 repeats (18%), a small number of intermediate alleles including those with 13 repeats were also observed. Haplotypes B, C and D exhibited a narrower range of trinucleotide repeats, 11 through 14, the most common of which was 13 repeats (48%).

The probe pKBE0.8 detects a three allele polymorphism at the D19S119 locus with frequencies in the normal population of 58%, 27% and 15% respectively (Korneluk et al., 1991). This frequency profile is also observed in haplotype A chromosomes from the normal population. In contrast, the pKBE0.8 allele frequency in DM individuals with a $(CTG)_n$ expansion is 50%, 1% and 49% respectively (Table 2-4A). Linkage disequilibrium has previously been reported between DM and D19S119 (Korneluk et al., 1991) but most notable is the paucity of DM on the "2" allele, identified by the probe pKBE0.8. Interestingly, examination of the distribution of pKBE0.8 alleles within haplotype A individuals with an upper normal range of $(CTG)_n$ repeats (i.e. $(CTG)_n \geq 20$) yielded a similar pattern to that observed in DM individuals (Table 2-4B). Most striking was the absence of pKBE0.8 allele "2" and the observation that the largest $(CTG)_n$ alleles (25/27/35 repeats) were all associated with pKBE0.8 allele "3".

Figure 2-1 PCR-based assays to detect sequence polymorphisms within intron 4 of the DM kinase gene and the adjacent DMR-N9 gene. Above each figure, the polymorphic base change and resultant polymorphic enzyme site are indicated. In each figure, the first lane contains the 123 bp DNA size marker, the next lane contains the uncut PCR-amplified product and the remaining lanes contain the digested PCR products from a representative DM pedigree. The black squares and circles represent DM affected males and females respectively, whereas the white symbols indicate unaffected individuals. The sizes of the uncut and digested products are indicated.

DMR-N9(G/T) DraIII



DMK (G/T) HphI

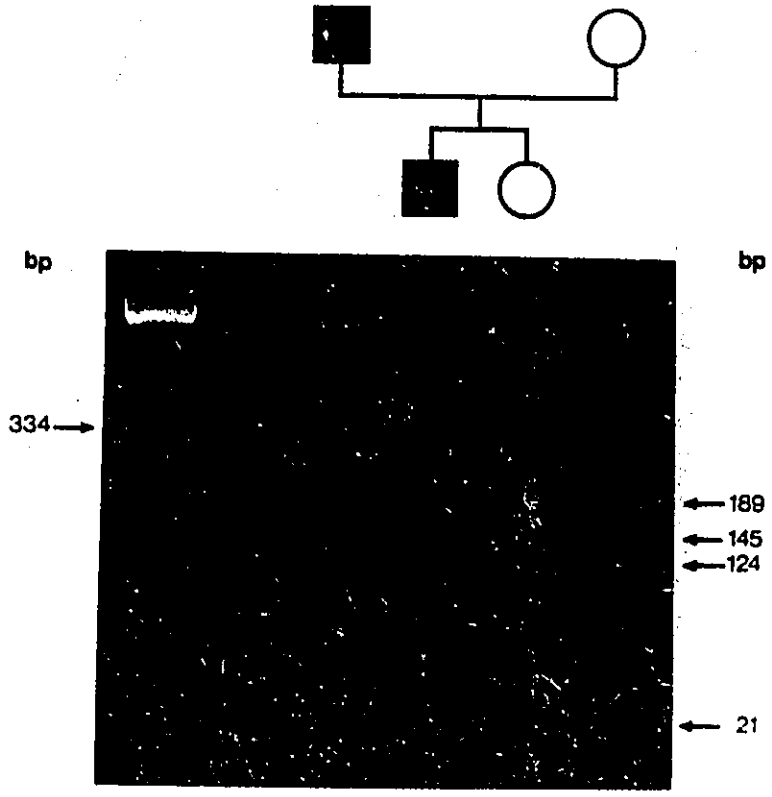


Figure 2-2 PCR-based assay to detect a sequence polymorphism within intron 5 of the DM kinase gene. Above the figure, the polymorphic base change and resultant polymorphic enzyme site are indicated. Within the figure, the first lane contains the 123 bp DNA size marker, the next lane contains the uncut PCR-amplified product and the remaining lanes contain the digested PCR products from a representative DM pedigree. The black square and circle represent a DM affected male and female respectively, whereas the white symbols indicate unaffected individuals. The sizes of the uncut and digested products are indicated.

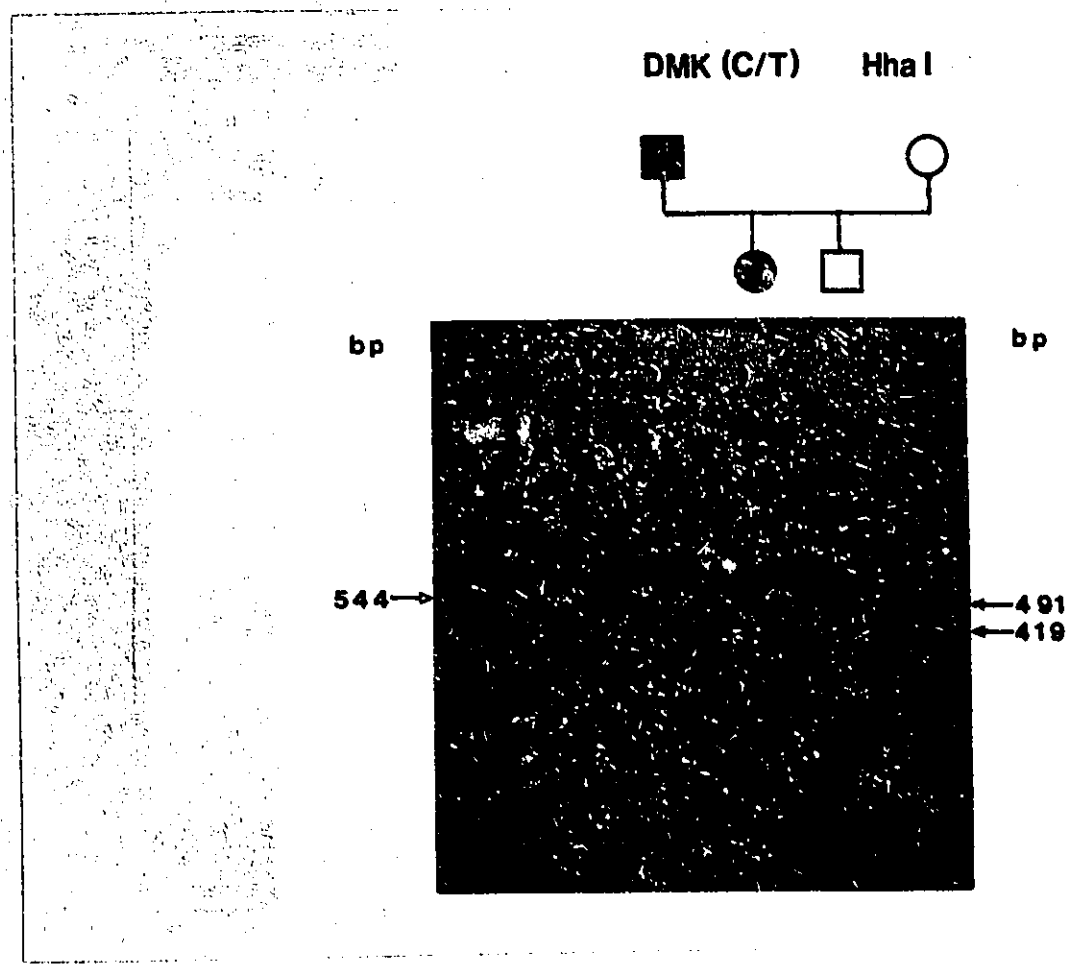
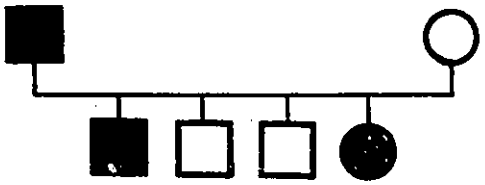

DMK (C/T)
Hha I
bp
bp
544
491
419

Figure 2-3 PCR-based assay to detect the insertion/deletion *Alu* element polymorphism within the DM kinase gene. Within the figure, the first lane contains the 123 bp DNA size marker, the next lanes contain PCR products generated from the amplification of genomic DNA across the region involved in the deletion event, using primers 405, 486 and 491. The lower band was generated from the smaller allele (representing the deletion of a 1 kb region containing three contiguous *Alu* elements) as a result of the amplification of the region between primers 491 and 486. The upper band is generated from the larger allele as a result of the amplification of the region (encompassing five contiguous *Alu* elements) between primers 491 and 486. Notably, the 1470 bp PCR product expected from the amplification of the region between primers 405 and 486 from the larger allele is absent, due to the preferential amplification of the smaller segment between primers 491 and 486. In this representative DM pedigree, as in all DM families, DM segregates with the larger allele. Black squares and circles represent DM affected males and females respectively, whereas white symbols indicate unaffected individuals.

DMK (Δ1kb)



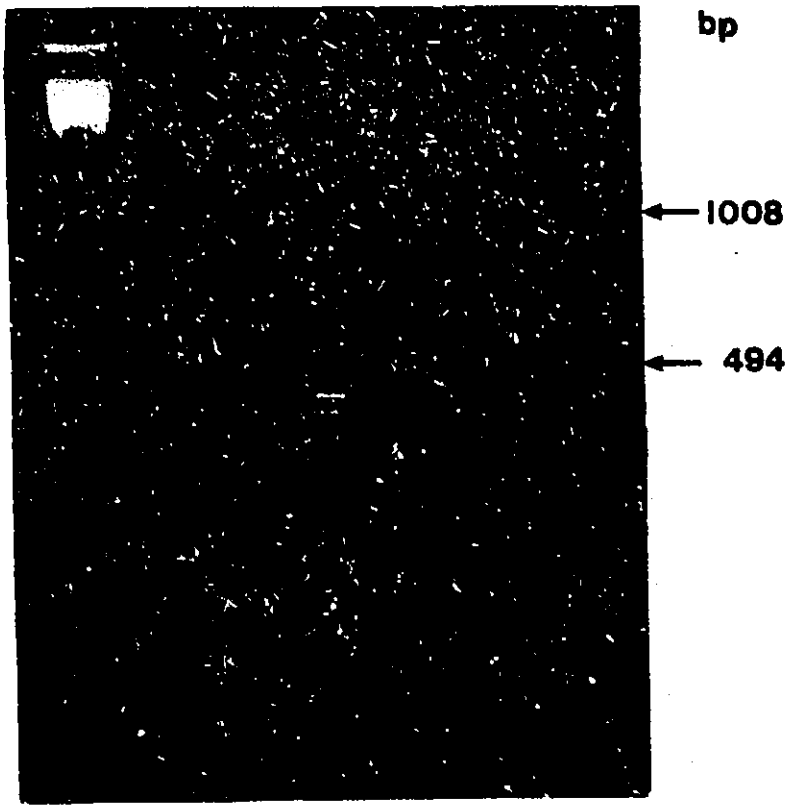
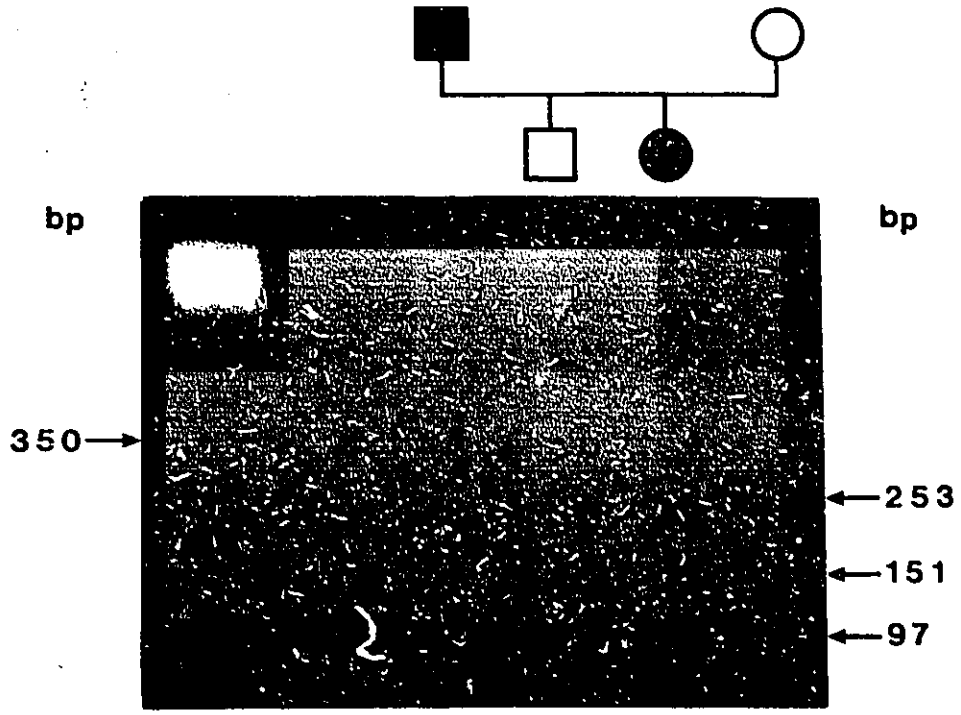


Figure 2-4 PCR-based assays to detect sequence polymorphisms within intron 9 and exon 10 of the DM kinase gene. Above each figure, the polymorphic base change and resultant polymorphic enzyme site are indicated. In each figure, the first lane contains the 123 bp DNA size marker, the next lane contains the uncut PCR-amplified product and the remaining lanes contain the digested PCR products from a representative DM pedigree. The black squares and circles represent DM affected males and females respectively, whereas the white symbols indicate unaffected individuals. The sizes of the uncut and digested products are indicated.

DMK (G/T) Hinf I



DMK (C/G) BpmI

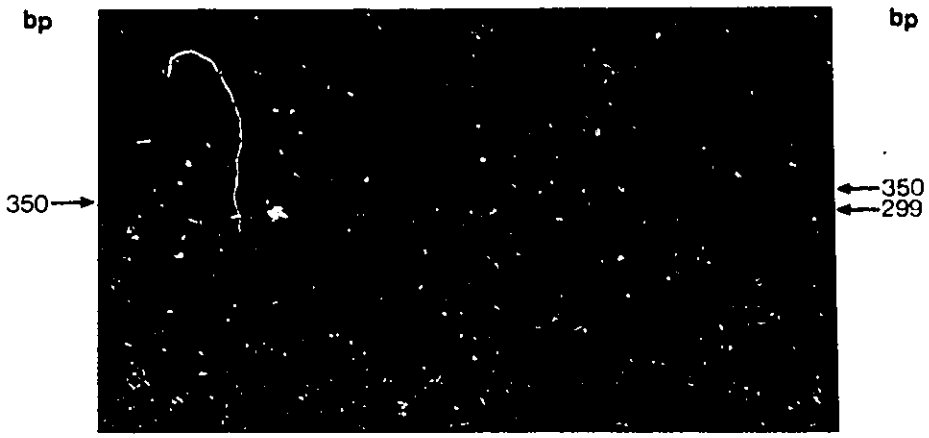
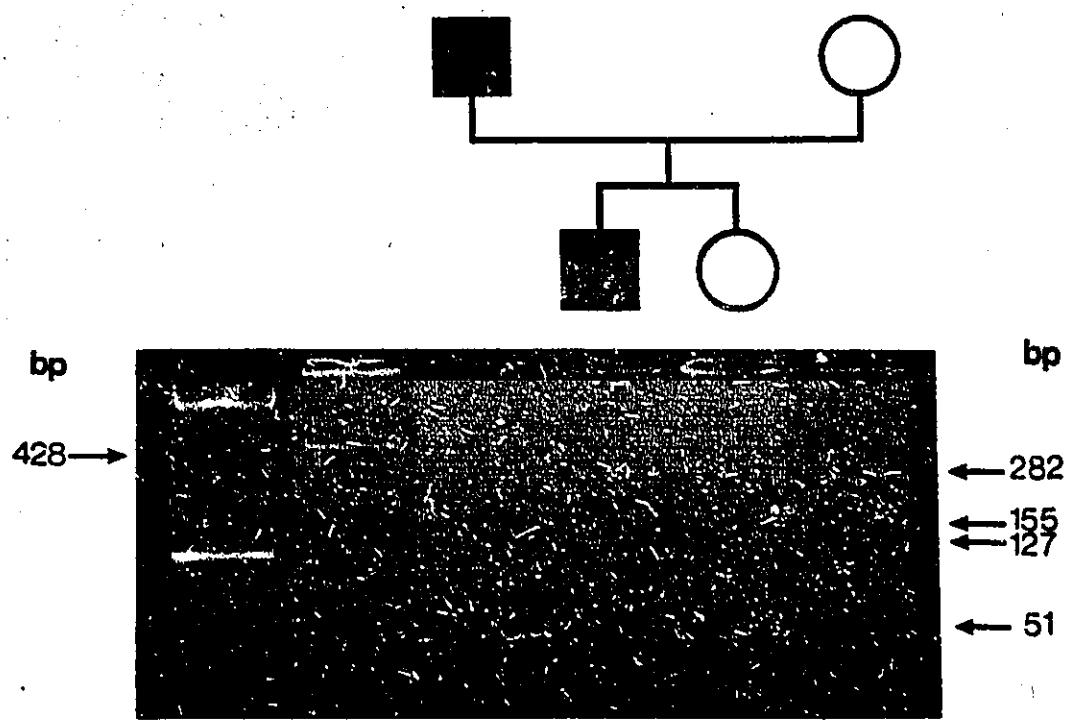


Figure 2-5 PCR-based assays to detect sequence polymorphisms within intron 11 of the DM kinase gene and a centromeric genomic segment. Above each figure, the polymorphic base change (if known) and resultant polymorphic enzyme site are indicated. In each figure, the first lane contains the 123 bp DNA size marker, the next lane contains the uncut PCR-amplified product and the remaining lanes contain the digested PCR products from a representative DM pedigree. The black squares and circles represent DM affected males and females respectively, whereas the white symbols indicate unaffected individuals. The sizes of the uncut and digested products are indicated.

DMK (G/T) Fnu4HI



pCN4OO TaqI

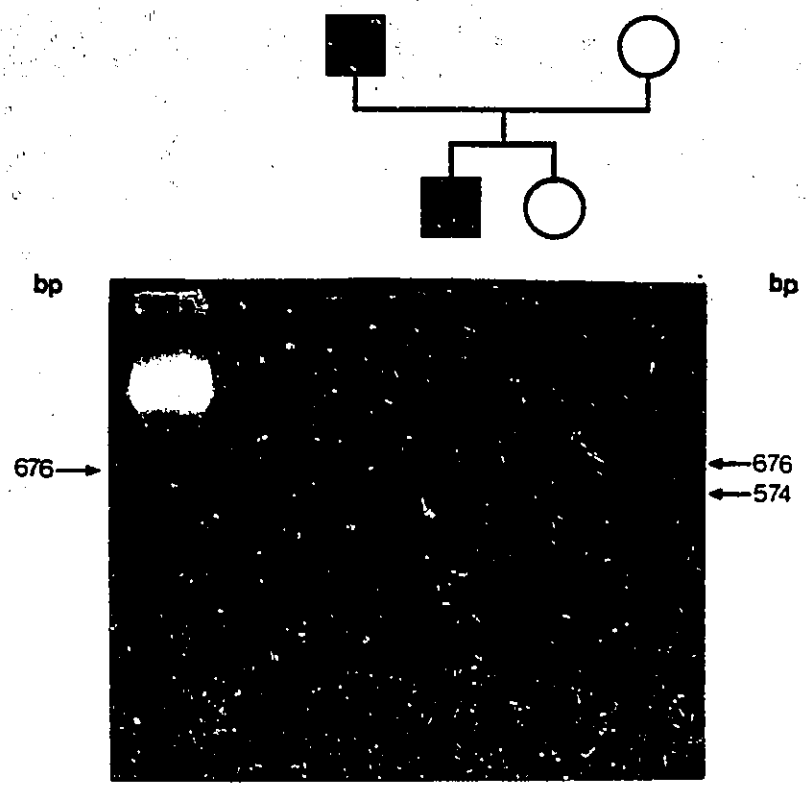


Figure 2-6 Localization of polymorphisms within the DM kinase gene and flanking genomic sequences. For each sequence polymorphism, the polymorphic base change (if known) and restriction enzyme (if applicable) are indicated. The orientation of the DM kinase gene is from telomere (TER) to centromere (CEN). The position of relevant restriction sites are given: B, BamHI; E, EcoRI. The exons are depicted as boxes and exons 1, 2, 9 and 15 are numbered. The locus D19S119 is located 145 kb centromeric to the locus D19S463, identified by the probe pCN400, which is located 15 kb centromeric to the $(CTG)_n$ repeat.

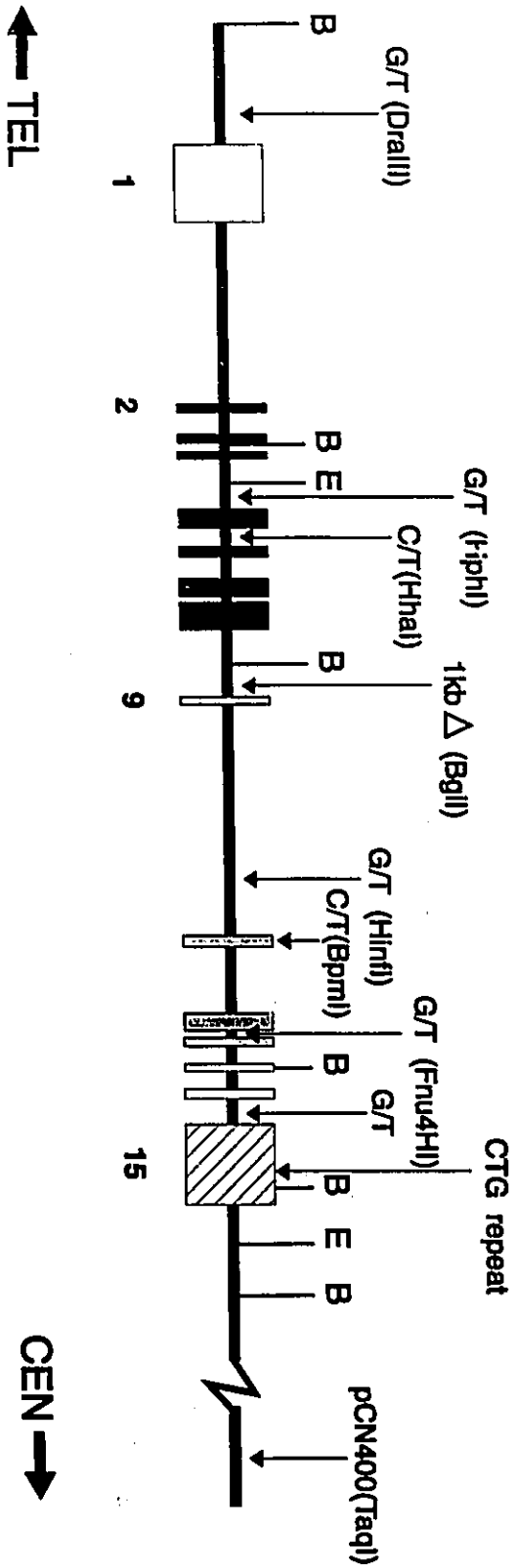

TEL
CEN
1 kb
1
2
9
15
B
G/T (DraIII)
B E
G/T (FiphI)
C/T(HhaI)
B
1kb Δ (BglI)
G/T (HinfI)
C/T(BpmI)
G/T (Fnu4HI)
B
G/T
B E B
CTG repeat
pCN400(TaqI)

Table 2-1 Sequence polymorphisms in the region of the DM kinase gene.

Sequence Polymorphism	Location in the Gene	Nucleotide Position‡	Polymorphic Enzyme Site	Frequency of Alleles in Normal Population Allele 1/Allele 2	DM Allele
DMR-N9 (G/T)	telomeric to exon 1	1069	DraIII	.64/.36	1
DMK (G/T)	intron 4	5285	HphI	.13/.87	2
DMK (C/T)	intron 5	6043	HhaI	.45/.55	1
DMK (Δ 1 kb)	intron 8	8157	N/A	.53/.47	1
DMK (G/T)	intron 9	10693	HinfI	.53/.47	2
DMK (C/G)	exon 10	10773	BpmI	.20/.80	2
DMK (G/T)	intron 11	11780	Fnu4HI	.55/.45	1
DMK (G/T)	intron 14	12581	N/A	.55/.45	T
DMK ($(CTG)_n$ Repeat)	exon 15	13230	N/A	-	-
pCN400 D19S463	15 kb centromeric	N/A	TaqI	.49/.51	2

N/A - not applicable

‡Note: The nucleotide position was obtained from the submitted sequence Genbank accession no. L08835.

Table 2-2 DM and normal haplotype frequencies.

HAPLOTYPE	SEQUENCE POLYMORPHISMS									HAPLOTYPE FREQUENCY
	DMR-N9 (G/T) DraIII	DMK(G/T) HphI	DMK(C/T) HhaI	DMK(Δ 1 kb)	DMK(G/T) HinfI	DMK(C/G) BpmI	DMK(G/T) Fnu4HI	DMK(G/T)	pCN400 TaqI D19S463	
DM	1	2	1	1	2	2	1	T	2	100%
A	1	2	1	1	2	2	1	T	2	49%
B	2	2	2	2	1	2	2	G	1	27%
C	1	1	2	2	1	1	2	G	1	16%
D	1	2	2	2	1	2	2	G	1	8%
E	2	2	2	2	1	2	2	G	2	<1%
F	1	1	2	2	1	2	2	G	1	<1%
G	1	2	1	1	2	1	1	T	2	<1%
H	2	2	2	2	2	2	2	G	2	<1%
I	1	1	2	2	2	2	1	G	1	<1%

Table 2-3 Association between haplotype and $(CTG)_n$ copy number.

HAPLOTYPE	NUMBER $(CTG)_n$
DM	≥ 50
A	5 -35
B	11 -14
C	11 -14
D	11 -14

Table 2-4A Distribution of pKBE0.8 alleles within DM and normal haplotypes.

HAPLOTYPE	NO. CHROMOSOMES		
	pKBE0.8 ALLELE		
	1	2	3
A	35	20	10
B	23	6	3
C	15	9	0
D	3	3	0
DM	22	1	20

Table 2-4B Association of pKBE0.8 alleles and $(CTG)_n \geq 20$ repeats on haplotype A chromosomes.

NO. REPEATS	NO. CHROMOSOMES	
	pKBE0.8 ALLELE 1	pKBE0.8 ALLELE 3
20	3	0
21	1	0
22	3	0
24	2	0
25	0	1
27	0	2
35	0	1
total	9	4

Discussion

To date, trinucleotide repeat expansion has been implicated in six inherited diseases including the fragile X syndrome (Kremer et al., 1991; Verkerk et al., 1991; Fu et al., 1991), spino-bulbar muscular atrophy (SBMA) (Biancalana et al., 1992), myotonic dystrophy (Mahadevan et al., 1992; Aslanidis et al., 1992; Buxton et al., 1992; Harley et al., 1992; Brook et al., 1992; Fu et al., 1992) and more recently, Huntington's disease (HD) (The Huntington's Disease Collaborative Research Group, 1993), spinocerebellar ataxia (SCA1) (Orr et al., 1993) and FRAXE mental retardation (Knight et al., 1993). The biological mechanism for the basis of the global persistence of these deleterious trinucleotide repeat mutation alleles is not known. However, the genetic analysis of the disease loci with respect to the intergenerational instability and origin of the mutation may provide insight into the mechanism of triplet repeat expansion.

In the normal population, the $(CTG)_n$ copy number at the DM locus is highly polymorphic, ranging from 5 to 35 repeats, yet it exhibits stable Mendelian inheritance patterns. In contrast, the $(CTG)_n$ repeat in DM individuals appears to display both intergenerational instability and somatic heterogeneity as well as either intergenerational amplification or reduction of the repeat number (O'Hoy et al., 1993; Brunner et al., 1993). Mildly affected DM individuals have 50 to a few hundred $(CTG)_n$ repeats, whereas those severely affected tend to have 2000 or more copies. Intergenerational amplification of the $(CTG)_n$ repeat has been proposed as the mechanism underlying genetic anticipation since there is a broad positive correlation between the number of repeats and disease severity (Hunter et al., 1992; Tsilfidis et al., 1992).

In order to account for the relatively high incidence of DM in the global population despite the negative influence of genetic anticipation on reproductive fitness, it was hypothesized that the DM mutation must be more prevalent than previously believed and therefore carried silently by a previously unascertained proportion of the population (Brook et al., 1992; Tsilfidis et al., 1992; Barceló et al., 1993). We have

suggested that the key to this enigma could be the relatively high frequency of the minimally amplified $(CTG)_n$ repeat in the range of 50-80 copies (termed the DM protomutation, Barceló et al., 1993), observed in individuals from unaffected branches of DM families. This previously undetected proportion of the population presents with a clinically inconsequential form of the disease yet has the potential to cause clinically apparent DM in subsequent generations. In keeping with this model, analysis of one of our large French Canadian pedigrees revealed the relatively stable intergenerational transmission of the protomutation through four generations (Barceló et al., 1993).

Recently, the molecular defect responsible for Huntington's disease (HD) was linked to a trinucleotide $(CAG)_n$ repeat in the coding region of the IT-15 gene (The Huntington's Disease Collaborative Research Group, 1993). Chromosomes with the most common HD haplotype at this locus and carrying a $(CAG)_n$ repeat length in the HD gene in the upper end of the normal range were found to be segregating in two families with putative new mutations for HD. Since clinical appearance of HD in these two cases was associated with expansion of the trinucleotide repeat, it was suggested that a particular subset of normal chromosomes may have the capacity to undergo expansion of the repeat into the HD range (The Huntington's Disease Collaborative Research Group, 1993). This same phenomenon has been proposed to occur with DM (Imbert et al., 1993) although transition from a normal size allele to protomutation size has not yet been observed. This could be attributed to ascertainment bias as families are studied upon presentation of a patient with the full mutation. Therefore the initial mutation, having occurred several generations earlier, would be inaccessible in those families. Ultimately, until such a transition from a $(CTG)_n$ repeat allele in the upper normal range to the protomutation size has been documented, the existence of a predisposing haplotype responsible for DM, while highly probable, is not conclusively proven.

Our study reveals the complete association of DM with one haplotype constructed from nine polymorphisms. The complete allelic association of the DM mutation with all

these polymorphisms may reflect a single mutational origin for DM and/or a haplotype predisposed to $(CTG)_n$ repeat instability. Out of 2^9 haplotypes theoretically possible for these nine RFLPs, it is perhaps surprising that only nine were observed in our normal population. However, only a limited number of chromosomes were analysed. Furthermore, the strong marker to marker allelic association observed between the nine markers in the normal population is likely a reflection of their relative proximity to each other and may also be due to their localization within a quiescent area of recombination. Interestingly, nearly two thirds of the haplotypes were represented by one of two complementary haplotypes (A and C). Although the mechanism underlying this phenomenon is unknown, we have previously observed similar complementary haplotypes at the ApoCII locus (MacKenzie et al., 1991), located approximately 1 Mb centromeric to the DM mutation.

The identification of a single DM haplotype, whose normal equivalent haplotype A (Table 2-2), is associated with the widest spectrum of $(CTG)_n$ repeats (Table 2-3) suggests one of three possible models of DM origin: i) an ancestral mutation which occurred once or rarely, ii) a particular haplotype in which the $(CTG)_n$ repeat is inherently unstable and therefore predisposed to mutation or iii) a haplotype of such ancient origin that the $(CTG)_n$ repeat length has had time to evolve from a single size to the distribution observed today. The corollary of model (iii) is that those haplotypes in which only a limited range of $(CTG)_n$ repeats are seen may be of more recent origin, suggesting that with time they would also become prone to sustaining the DM mutation. This possibility seems unlikely given that these haplotypes (B-D; Table 2-2) occur with a frequency of 51% in the general population and that the DM mutation is not found on these haplotypes.

Previously, the marker pKBE0.8, which detects a three allele insertion/deletion polymorphism, was found to demonstrate marked linkage disequilibrium with DM (Korneluk et al., 1991). Significantly, linkage disequilibrium was also observed in the

European DM population at the D19S63 locus with pD10 (Harley et al., 1991), a probe which identifies the same polymorphism detected by pKBE0.8. Our analysis of this polymorphism revealed a similar allelic distribution for haplotype A individuals with an upper normal range of $(CTG)_n$ repeats (i.e. $(CTG)_n \geq 20$) and DM individuals. This finding parallels those reported by Imbert et al. (1993) using a $(CA)_n$ repeat (at D19S112), located approximately 90 kb telomeric to the $(CTG)_n$ repeat. Thus, our results extend the previously reported allelic association (Imbert et al., 1993), 160 kb in a centromeric direction. This further strengthens the possibility that these $(CTG)_n \geq 20$ repeat alleles may represent the pool from which disease causing mutations are formed. The presence of larger $(CTG)_n$ alleles associated with pKBE0.8 allele "3" may represent two independent ancestral mutations occurring on the background of pKBE0.8 alleles "1" and "3". Alternatively, a single ancestral mutation with a single recombination event onto the "3" allele could also be responsible for the observed results, as the polymorphism is located 160 kb centromeric to the $(CTG)_n$ repeat.

Our data suggest the presence of two founding chromosomes: one containing a stretch of five contiguous *Alu* elements (the progenitor for haplotype A) and the other in which three of these have been deleted (the progenitor for haplotypes B, C and D). Since the $(CTG)_5$ allele is the most common repeat allele associated with haplotype A and this number of repeats is conserved in the mouse equivalent DM kinase gene (Jansen et al., 1992b), this may represent the primary ancestral chromosome from which all others were derived. The $(CTG)_5$ alleles as well as alleles with greater than 19 repeats are exclusively linked to haplotype A. This association is consistent with the work by Imbert et al. (1993) and is extended unchanged by additional polymorphisms derived from our high resolution haplotype analysis.

In contrast, alleles in which the three *Alu* elements are deleted possess only $(CTG)_{11-14}$ repeats. This distribution is quite restricted and distinct from that observed for haplotype A. We have previously suggested that the second of five *Alu* elements,

which has a direct repeat at either end, was introduced by retrotransposition followed by a deletion of three contiguous *Alu* repeats in intron 8 of the DM kinase gene (Mahadevan et al., 1993a). As the $(CTG)_{13}$ allele is the most common allele associated with chromosomes carrying the *Alu* deletion, the deletion event may have initially occurred on a haplotype A chromosome with 13 repeats. This is consistent with haplotypes B, C and D being associated with the most recent chromosomes. However, the equal frequency of both founding chromosomes in the global population suggests that they are both relatively ancient. Although the inference that the loss of the three *Alu* repeats may confer increased stability on the $(CTG)_n$ repeat is speculative, the narrow size range of the $(CTG)_n$ repeat on chromosomes in which the *Alu* elements have been deleted, relative to the variation seen on normal chromosomes with the DM haplotype (i.e. haplotype A), is striking.

The minority (approximately 10%) of normal alleles with 20-30 repeats are likely derivatives of the $(CTG)_5$ allele by very few primary events. In keeping with previous work (Imbert et al., 1993), we propose that these alleles would be at risk for low frequency recurrent multistep amplification of the $(CTG)_n$ repeat which may lead, over many generations, to a disease causing mutation. In fact, these $(CTG)_{20-30}$ alleles may act as a reservoir for new mutations such as duplication events which would generate protomutation size alleles. The protomutation can be transmitted in a relatively stable manner through many successive generations (Barceló et al., 1993), accounting for the prevalence of the DM mutation whose frequency may only be accurately estimated by an extensive population survey. As a consequence, the real prevalence, as well as the inherited stability of the protomutation is not known. Yet, on the basis of this high resolution genetic analysis the existence of an unstable 19q13.3 haplotype which encompasses both the models of (i) an ancestral mutation and (ii) a particular haplotype in which the $(CTG)_n$ repeat is inherently unstable, appears possible. Based on this model, the loss of DM alleles from the population due to the end result of genetic

anticipation could be balanced by the introduction of new mutations onto this haplotype. This may account for the relatively stable prevalence of DM in the global population.

CHAPTER III GENERAL DISCUSSION

We have proposed a model in which the DM mutation is postulated to occur as a multistep process. This model provides a framework in which the seemingly contradictory observations of a mutation old enough to establish a founder effect and an apparently high new mutation rate are united. The model suggests that alleles with $(CTG)_n \geq 20$ repeats arise from alleles with fewer repeats and that these constitute a reservoir from which protomutation size alleles are derived. Individuals with protomutations (Barceló et al., 1993), in the range of 50-80 $(CTG)_n$ repeats, are clinically normal. Moreover, the protomutation can be passed down in a stable manner for many generations. This stability would allow the required time for the establishment of an ancestral haplotype, resulting in the observed linkage disequilibrium. Linkage disequilibrium mapping is based upon the expectation that a chromosomal region lying in close proximity to the mutant gene may have descended from a common ancestral mutation and therefore will show a common haplotype reflecting that of the original ancestral chromosome. Accordingly, my results suggest that the DM mutation occurred on the background of a particular haplotype in which the $(CTG)_n$ repeat became inherently unstable and therefore predisposed to amplification. It is this pool of pre-existing, or predisposed carriers (individuals with protomutations) which is constantly giving rise to new clinical DM cases. Based on this model, the loss of DM alleles from the population due to the end result of genetic anticipation could be balanced by the transition of protomutations to full mutations. This may account for the relatively stable prevalence of DM in the global population.

However, the nature of predisposing alleles and the event which triggers the dramatic expansion of the $(CTG)_n$ repeat is currently unknown. One hypothesis is that the $(CTG)_n$ repeat may be inherently unstable. Evidence that other simple repeat elements in the human genome have frequencies of mutation approaching 1% (Hästbacka

et al., 1992) suggests that this phenomenon may not be a rare occurrence. Alternatively, it may be possible that the expansion of the $(CTG)_n$ element is independent of the repeat itself, but is controlled by a flanking DNA element. Alterations in the length of microsatellite loci are known to be controlled by flanking modulating elements (Jeffreys et al., 1990).

Repeated DNA sequences interspersed throughout the human genome have become recognized as an important source of DNA polymorphism in molecular genetic research (Jeffreys et al., 1985; Nakamura et al., 1987). Among this group are the microsatellites or simple tandem repeated sequences which are evenly distributed in the genome. The rate of mutation of microsatellites is related to their copy number (Weber, 1990). Since the mutability of the product of a change in copy number is different from that of its predecessor, this mechanism has been termed dynamic mutation (Richards and Sutherland, 1992a). The unique properties of dynamic mutation account for the unusual genetic characteristics of myotonic dystrophy. However, the mechanics of dynamic mutation are as yet unexplained. Meiotic recombination as a cause of instability has been ruled out by analysis of informative flanking markers (Jeffreys et al., 1990) and by the fact that a decrease in linkage disequilibrium is not observed. Experiments *in vitro* demonstrate that slippage during replication is not dependent upon length (Schlötterer and Tautz, 1992). Since there is a relationship between change in copy number and length, slippage in itself does not appear to account for dynamic mutation.

The recent finding that fragile X syndrome, like myotonic dystrophy, has founder chromosomes raises a fundamental question about the molecular events involved in dynamic mutation. Two models for such founder mutations are possible. In one model, an increase in copy number beyond a certain length may occur rarely, but once beyond this length, subsequent mutation is relatively frequent. Simple tandem repeated sequences are often interrupted by an imperfection in the repeated motif that appears to limit their mutability (Richards and Sutherland, 1992b). Interruptions involving a G to T

substitution have been observed in the fragile X syndrome $(CCG)_n$ repeat in normal individuals (Oberlé et al., 1991; Kremer et al., 1991; Verkerk et al., 1991). Therefore, the degree of polymorphism in simple tandem repeats is proportional to perfect repeat length (Weber, 1990). Thus, the founder mutation could be either a rare increase in copy number of only perfectly repeated sequence or a mutation of an imperfect interruption back to the normal motif, which establishes a perfect repeat of unstable length (Richards and Sutherland, 1992a).

Analysis of normal individuals with the high resolution haplotype common to DM individuals (haplotype A) revealed a subset of chromosomes (18%) carrying $(CTG)_n$ alleles with greater than 19 repeats. This same phenomenon has been observed among normal individuals with one of the haplotypes common to fragile X individuals in which a high incidence of high (normal) copy number ($(CTG)_n > 39$) was revealed (Richards and Sutherland, 1992b). This suggests that there may be a pool of individuals whose descendants are at risk of subsequent mutation due to their high copy number and relatively high mutability.

While it is clearly established that a repeated sequence capable of amplification is a necessary condition for dynamic mutation disease, it is not clear whether this sequence alone is sufficient. The second model for founder mutations suggests that mutation of another sequence unrelated to the trinucleotide repeat in some way predisposes the repeat to greater instability. For example, the acquisition of a DNA binding site for a protein involved in DNA mechanics may confer a greater rate of subsequent copy number variation than is normally accorded these sequences. More importantly, the relative proximity of a contiguous stretch of *Alu* repeats in the DM kinase gene to the highly polymorphic $(CTG)_n$ repeat may increase the susceptibility of this trinucleotide repeat to slipped strand mispairing, resulting in an increased number of repeats (Mahadevan et al., 1993a). The *Alu* repeat contig may confer a conformational change on its adjacent sequences thereby increasing the likelihood of slipped strand mispairing.

Given the deleterious consequences of including potentially unstable trinucleotide repeats within a gene, it seems logical to conclude that the presence of unstable repeats in genes must have functional significance. However, specific functions for known tandem repetitive sequences so far discovered remain obscure. Nevertheless, a subset of minisatellites share a common "core" sequence which, based on sequence similarity to the generalized recombination signal of Escherichia coli, Chi (χ), has led to speculation that the core sequence may serve as a recombination signal to promote unequal crossing over at minisatellites (Jeffreys et al., 1985). The presence of core-like sequences near a meiotic recombination hotspot (Steinmetz et al., 1986) and the preferential localization of minisatellites near the ends of human chromosomes in regions involved in chromosome synapsis and recombination (Royle et al., 1988) support this hypothesis.

Repeats in coding sequences may be necessary if the function of the encoded protein requires a particular homopolymer. In this case, there would seem to be pressure to select for wobble mutations, so that the DNA sequence would lose its repetitive nature yet maintain the homopolymer coding capacity. The high mobility group protein of trout appears to have such wobble mutations: its $(ACG)_n$ repeat frequently mutates to $(ATG)_n$ yet conserves the coding capacity of its mRNA (Boylan et al., 1990). However, the finding of repeated sequences in untranslated regions suggests that a fundamental constraint other than coding capacity maintains the presence of the repeats. The trinucleotide repeat $(CCG)_n$ in the 5' untranslated region of the breakpoint cluster region (BCR) mRNA is the binding site for a nuclear protein factor (Zhu et al., 1990). Changing the copy number of the repeat could change the number of sites for the factor and therefore affect transcriptional activity of the gene.

Fragile X syndrome, Huntington's disease and myotonic dystrophy are three distinct genetic diseases which are caused by dynamic mutation (Fu et al., 1991; Kremer et al., 1991; The Huntington's Disease Collaborative Research Group, 1993; Mahadevan et al., 1992; Fu et al., 1992; Brook et al., 1992). A trinucleotide repeat is implicated in

each disorder, however the location of the repeated element within the respective gene and therefore the mechanics of disease in each case is unique. This type of defect has been proven to be a more general mechanism of mutation which likely underlies many more genetic disorders in which the phenomenon of anticipation occurs.

Fragile X syndrome, associated with intellectual impairment and minor dysmorphic features, is characterized by a cytogenetic fragile site at Xq27.3. An amplification of a trinucleotide $(CGG)_n$ repeat has been identified within the 5' untranslated region of the FMR-1 gene on the X chromosome. In affected individuals, the repeat expands dramatically ($(CCG)_n \geq 200$) and as a consequence the FMR-1 gene becomes highly methylated and gene expression is lost (Hirst et al., 1993). As is seen in DM, an increasing copy number of the unstable sequence is paralleled by an increased penetrance of the disorder in subsequent generations. This genetic property, anticipation, is referred to in the fragile X syndrome as the Sherman paradox (Sherman et al., 1984, 1985).

While this disorder is clearly X linked, phenotypically normal males identified as obligate carriers by their position in the pedigree are common. These transmitting males have 60-200 copies of the repeat (termed "premutation", Pembrey et al., 1985) whereas normal individuals have 6-60 copies (Richards and Sutherland, 1992b). The DM protomutation is unique in that the range of repeat sizes does not overlap with that of normal individuals and it may be associated with mild clinical symptoms. The premutation of fragile X, similar to the DM protomutation, represents a pre-existing population of mutation carriers which accounts for the high frequency of the syndrome despite the negative influence of genetic anticipation on reproductive fitness. As is the case with DM, the generation of fragile X syndrome is a multistep process. Larger $(CCG)_n$ normal alleles convert to highly unstable premutation alleles which generate new mutations. Amplification mostly occurs when the sequence is transmitted by females to the extent that transmission through a female is essential to generate a copy number that

will manifest the disorder (Richards and Sutherland, 1992a). In DM, this is true only for the congenital form of the disorder.

Using haplotype analysis of microsatellite markers flanking the fragile X unstable element, a few founder mutations were determined to be responsible for the majority of fragile X cases (Richards and Sutherland, 1992b). Nevertheless, 44% of fragile X chromosomes carry other haplotypes, suggesting that mutations may have arisen independently on other genetic backgrounds (Hirst et al., 1993). However, diversity of haplotypes may also be the result of mutations in the $(CA)_n$ dinucleotide markers themselves (Hirst et al., 1993). This is unlike DM in which the mutation occurs on the background of one particular haplotype.

Huntington's disease (HD) is an inherited neurodegenerative disorder of mid-life onset transmitted in an autosomal dominant manner and characterized by chorea, dementia and personality disorder (Myers et al., 1993). An amplification of a trinucleotide $(CAG)_n$ repeat has been identified within the coding region of the IT-15 gene at 4p16.3 (The Huntington's Disease Collaborative Research Group, 1993). The $(CAG)_n$ sequence is likely transcribed and translated as a polyglutamine tract which interferes with normal protein function.

A newly recognized intermediate allele (IA) in the parental generation has been demonstrated to contain 30-38 $(CAG)_n$ repeats in the HD gene which is greater than usually seen in the general population ($(CAG)_{11-34}$) but below the range seen in patients with HD ($(CAG)_n \geq 37$). These IAs or premutations are meiotically unstable and in the sporadic cases, expand to the full mutation associated with the phenotype of HD (Goldberg et al., 1993). This expansion has been shown to occur only during transmission through the male germline, in contrast to fragile X where expansion of the premutation originates in the female germline (Goldberg et al., 1993). In DM, the protomutation is also meiotically unstable, but in contrast to both fragile X and HD, it may be transmitted by either parent (Barceló et al., 1993). In DM and fragile X, the

length of the repeat has been demonstrated to be the major source of recurrent DNA mutations once the repeat has reached an intermediate range (Fu et al., 1991; Kremer et al., 1991; Mahadevan et al., 1992; Fu et al., 1992; Brook et al., 1992). In HD, the $(CAG)_n$ repeat length is only one of the factors contributing to instability once the number of repeats reaches a threshold level ($\geq$ 30 repeats) (Goldberg et al., 1993). Mutations causing HD may be more likely to occur on the paternally derived allele due to a higher mutation rate in the male germline. Errors in DNA replication during germ cell division may be more likely in males as the number of germ cell divisions per generation is much greater than that seen in females. In addition, advanced paternal age in some undetermined way increases the susceptibility of the IA to full mutation in HD (Goldberg et al., 1993).

Disequilibrium studies have revealed that one haplotype occurs on about one third of HD chromosomes (MacDonald et al., 1992). A wide variety of other haplotypes are represented among the rest of the HD chromosomes suggesting either that many different independent mutations have given rise to HD or that the defect can transfer to other haplotypes by a mechanism other than simple recombination (MacDonald et al., 1992). Gene conversion and double recombination are mechanisms which may account for the HD haplotype diversity consistent with an original mutant chromosome model for HD. Moreover, it is possible that there is a very ancient original premutation for HD predating the occurrence of DNA polymorphisms in the HD region which at different times in various populations has undergone expansion (MacDonald et al., 1992). With either model, a multistep mutational process is consistent with the development of HD, similar to myotonic dystrophy and the fragile X syndrome.

The human genome contains many interspersed repetitive DNA sequences that are polymorphic in length because of variability in the number of repeat units in a tandemly repeated sequence. A newly identified class of tandemly repetitive polymorphic sequences which are located adjacent to *Alu* repeats has recently been described as "*Alu*

sequence-related polymorphisms" (Zuliani and Hobbs, 1990). The basis of this polymorphism was determined to be a trinucleotide $(TTA)_n$ repeat. Tandem repeats of this trimer were identified in or near the genes for HMG-CoA reductase, β-tubulin, enkephalin, factor IX, α-globin, interleukin-1α and fibrinogen. Four of these $(TTA)_n$ repeats were associated with *Alu* sequences and three of the four were found to be polymorphic in length. The $(TTA)_n$ repeat in the HMG-CoA reductase gene is located immediately 3' of an *Alu* sequence and in the other two (β-tubulin and interleukin-1α) genes the repeats are part of, or are immediately adjacent to, one of the direct repeats flanking an *Alu* sequence. No size polymorphism was identified in the repeats not associated with *Alu* sequences. The presence or absence of polymorphism can not be attributed to differences in the number of repeat units since the same number of $(TTA)_n$ repeats was found in all of these genes . Therefore, there appears to be an association between the frequency of length polymorphism and the physical proximity to *Alu* sequences.

Alu sequences (which constitute approximately 10% of the genome) have been shown to be concentrated in the R (reverse) bands by *in situ* hybridization studies of human chromosomes (Kornberg and Rykowski, 1988). These regions are very GC rich and have been found to be more transcriptionally active than adjacent sequences. Both the close physical association of *Alu* sequences with transcribed genes and the possibility that some *Alu* sequences are transcriptionally active may make sequences adjacent to *Alu* repeats more prone to slipped-strand mispairing events (Zuliani and Hobbs, 1990). Recombination seems to occur more frequently in the R bands, and this may be related to the concentration of *Alu* sequences within these regions. *Alu* sequences have been implicated as sites of recombinational events responsible for the duplication and evolution of new genes in the human genome (Barsh et al., 1983; Kudo and Fukudo, 1989). Therefore, sequences in close proximity to *Alu* repeats may be more susceptible to

slipped-strand mispairing events, owing to a higher frequency of recombinational events associated with *Alu* sequences.

Direct repeats are found immediately 3' of the poly A tail and at a variable distance 5' of most *Alu* repeats. The sequences of the direct repeats are not part of the *Alu* consensus sequence and are thought to be generated at the site of *Alu* integration into the genome (Bains, 1986). *Alu* sequences seem to preferentially insert into A-rich regions of the genome, which may account for the AT rich composition of these polymorphic repeats (Zuliani and Hobbs, 1990). Interestingly, at both the α-globin and interleukin-1α loci, only one of the two $(TTA)_n$ repeats flanking the *Alu* sequence was found to be polymorphic in length, suggesting that the size polymorphism was generated subsequent to the insertion of the *Alu* sequence. Similarly, we suggest that the $(CTG)_n$ repeat instability at the DM locus may result from a contiguous sequence of five *Alu* repeats located 5 kb telomeric to the repeat. Normal alleles in which the three *Alu* elements at the DM locus are deleted possess only $(CTG)_{11-14}$ repeats. In contrast, alleles in which the five *Alu* elements at the DM locus are intact possess a wider range in $(CTG)_n$ copy number (5 to 35 repeats) in the normal population. Although the inference that the loss of the three *Alu* repeats may confer increased stability on the $(CTG)_n$ repeat is speculative, the narrow size range of the $(CTG)_n$ repeat on chromosomes in which the *Alu* elements have been deleted, relative to the variation seen on normal chromosomes with the DM haplotype (i.e. haplotype A), is striking. We have previously suggested that the second of five *Alu* elements, which has a direct repeat at either end, was introduced by retrotransposition followed by a deletion of three contiguous *Alu* repeats in intron 8 of the DM kinase gene (Mahadevan et al., 1993a). We suggest that the $(CTG)_n$ length polymorphism was generated subsequent to the insertion of the contiguous *Alu* sequence at the DM locus. The insertion of the *Alu* elements may have been the initial ancestral mutation which predisposed haplotype A to $(CTG)_n$ repeat instability. Protomutation size alleles, derived from a pool of upper normal range repeat sizes (ie. $(CTG)_{20-35}$) on

haplotype A, are responsible for the continual introduction of new DM mutations into the human gene pools.

In summary, the objective of this thesis was to study the origin of the DM mutation. The mutation causing DM has been identified as an amplification of an unstable trinucleotide $(CTG)_n$ repeat in over 99% of the global DM population (Mahadevan et al., 1992; Brook et al., 1992; Fu et al., 1992). It is in complete linkage disequilibrium with an *Alu* element polymorphism within the DM kinase gene, suggesting that DM is a consequence of one or few ancestral mutations (Mahadevan et al., 1993a; Yamagata et al., 1992; Harley et al., 1992). However, the low reproductive fitness of individuals with congenital DM (the endpoint of genetic anticipation in DM) suggests a significant rate of new mutations (Harley et al., 1992; Barceló et al., 1993). I have presented a high resolution haplotype analysis of the DM locus using PCR-based assays of nine polymorphisms, spanning a physical distance of 30 kb, within and immediately flanking the DM kinase gene at chromosome 19q13.3. The complete allelic association of the DM mutation consistently observed with these polymorphisms suggests that the DM mutation occurred on the background of a particular haplotype (haplotype A) in which the $(CTG)_n$ repeat is slightly, but unequivocally, unstable and therefore predisposed to rare amplification from the normal to the protomutation range of trinucleotide repeat size. Specifically, one or few ancestral expansion mutations likely occurred via the mechanism of slipped-strand mispairing due to the proximity of the *Alu* insertion polymorphism to the $(CTG)_n$ repeat. This created a minority class (10%) of alleles with $(CTG)_{20-30}$ repeats which, serving as a reservoir for low frequency multistep amplification, resulted in the formation of the DM protomutation. The protomutation size alleles (ie. $(CTG)_{50-80}$), capable of clinically inconsequential intergenerational transmission yet committed to further amplification into the DM range (ie. $(CTG)_n \geq 100$), could be responsible for the observed linkage disequilibrium associated with this ancestral DM mutation. Based on this model, the loss of DM alleles from the population

due to the end result of genetic anticipation could be balanced by the transition of protomutations to full mutations. This may account for the relatively high prevalence of DM in the global population.

REFERENCES

Aslanidis, C., Jansen, G., Amemiya, C., Shutler, G., Tsilfidis, C., Mahadevan, M., Chen, C., Allemen, 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: 548-550.

Bains, W. (1986) The multiple origins of human *Alu* sequences. J. Mol. Evol. 23: 189-199.

Barceló, J., Mahadevan, M.S., Tsilfidis, C., MacKenzie, A.E. and Korneluk, R.G. (1993) Intergenerational stability of the myotonic dystrophy protomutation. Hum. Mol. Gent. 2: 705-709.

Barsh, G.S., Seeburg, P.H. and Gelinas, R.E. (1983) The human growth hormone gene family: structure and evolution of the chromosomal locus. Nucl. Acids Res. 11: 3939-3958.

Biancalana, V., Serville, F., Pommier, J., Julien, J., Hanauer, A. and Mandel, J.L. (1992) Moderate instability of the trinucleotide repeat in spino bulbar muscular atrophy. Hum. Molec. Genet. 1: 255-258.

Boylan, K.B., Ayres, T.M., Popko, B., Takahashi, N., Hood, L.E. and Prusiner, S.B. (1990) Repetitive DNA $(TGGA)_n$ 5' to the human myelin basic protein gene: a new form of oligonucleotide repetitive sequence showing length polymorphism. Genomics 6: 16-22.

Brook, J.D., McCurrach, M.E., Harley, H.G., Buckler, A.J., Church, D., Aburatani, H., Hunter, K., Stanton, V.P., Thirion, J-P., Hudson, T., Sohn, R., Zemelman, B., Snell, R.G., Rundle, S.A., Crow, S., Davies, J., Shelbourne, P., Buxton, J., Jones, C., Juvonen, V., Johnson, K., Harper, P.S. and Shaw, D.J. (1992) Molecular basis of myotonic dystrophy: expansion of a trinucleotide (CTG) repeat at the 3' end of a protein kinase family member. Cell 68: 799-808.

Brunner, H.G., Jansen, G., Nillesen, W., Nelen, M.R., de Die, C.E.M., Howeler, C.J., van Oost, B.A., Wieringa, B., Ropers, H-H. and Smeets, H.J.M. (1993) Brief report: reverse mutation in myotonic dystrophy. New Engl. J. Med. 238: 476-480.

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.

Davies, J., Yamagata, H., Shelbourne, P., Buxton, J., Ogihara, T., Nokelainen, P., Nakagawa, M., Williamson, R., Johnson, K. and Miki, T. (1992) Comparison of the myotonic dystrophy associated CTG repeat in European and Japanese populations. J. Med. Genet. 29: 766-769.

Eiberg, H., Mohr, J., Nielson, L. and Simonson, N. (1983) Genetics and linkage relationships of the C3 polymorphism: discovery of C3-Se linkage and assignment of LES-C3-DM-Se-PEPD-Lu synteny to chromosome 19. Clin. Genet. 24: 159-170.

Fu, Y-H., Pizzuti, A., Fenwick, R.G., King, J., Rajnarayan, S., Dunne, P.W., Dubel, J., Nasser, G.A., Ashizawa, T., de Jong, P., Wieringa, B., Korneluk, R.G., Perryman, M.B., Epstein, H.F. and Caskey, C.T. (1992) An unstable triplet repeat in a gene related to myotonic dystrophy. Science 255: 1256-1258.

Fu, Y-H., Kuhl, D.P.A., Pizzuti, A., Pieretti, M., Sutcliffe, J.S., Richards, S., Verkerk, A.J.M.H., Holden, J.J.A., Fenwick Jr., R.G., 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.

Glanz, A. and Fraser, F.C. (1984) Risk estimates for neonatal myotonic dystrophy. J. Med. Genet. 21: 186-188.

Goldberg, Y.P., Kremer, B., Andrew, S.E., Theilmann, J., Graham, R.K., Squitieri, F., Telenius, H., Adam, S., Sajoo, A., Starr, E., Heiberg, A., Wolff, G. and Hayden, M.R. (1993) Molecular analysis of new mutations for Huntington's disease: intermediate alleles and sex of origin effects. Nat. Genet. 5: 174-179.

Hall, J.G. (1990) Genomic imprinting: review and relevance to human diseases. Am. J. Hum. Genet. 46: 857-873.

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.

Harley, H.G., Brook, D.J., Floyd, J., Rundle, S.A., Crow, S., Walsh, K.V., Thibault, M-C., Harper, P.S. and Shaw, D.J. (1991) Detection of linkage disequilibrium between the myotonic dystrophy locus and a new polymorphic DNA marker. Am. J. Hum. Genet. 49: 68-75.

Harper, P.S. (1989) *Myotonic Dystrophy*, second edition, W.B. Saunders, London and Philedelphia.

Hästbacka, J., de la Chapelle, A., Kaitila, I., Sistonen, P., Weaver, A. and Lander, E. (1992) Linkage disequilibrium mapping in isolated founder populations: diastrophic dysplasia in Finland. Nat. Genet. 2: 204-211.

Hirst, M.C., Knight, S.J.L., Christodoulou, Z., Grewal, P.K., Fryns, J.P. and Davies, K.E. (1993) Origins of the fragile X syndrome mutation. J. Med. Genet. 30: 647-650.

Hockings, G.I., Grice, J.E., Crosbie, G.V., Walters, M.M. and Jackson, R.V. (1993) Altered hypothalamic-pituitary-adrenal axis responsiveness in myotonic dystrophy: *in vivo* evidence for abnormal dihydropyridine-insensitive calcium transport. J. Clin. Endo. and Metab. 76: 1433-1438.

Hulsebos, T., Brunner, H., Wieringa, B., Friedrich, U., Smeets, D., Oei, T., Hustinx, T., Scheres, J., Wienker, T., Humphries, S., Myklebost, O., Junien, C., Haarter, B. and Ropers, H.H. (1985a) Regional assignment of C3, Gp1, ApoC2 and β-HCG and their linkage relationship with DM and 19 cen. Cytogenet. Cell Genet. 40: 658.

Hulsebos, T., Wieringa, B., Hochstenbach, R., Smeets, D., Schepens, J., Oerlemans, R., Zimmer, J. and Ropers, H.H. (1985b) Toward early diagnosis of myotonic dystrophy: construction and characterization of a somatic cell hybrid with a single human der (19) chromosome. Cytogenet. Cell Genet. 43: 47-56.

Hunter, A., Tsilfidis, C., Mettler, G., Jacob, P., Mahadevan, M., Suhr, L. and Korneluk, R.G. (1992) The correlation of age of onset with CTG trinucleotide repeat amplification in myotonic dystrophy. J. Med. Genet. 29: 774-779.

Hunter, A.G.W., Jacob, P., O'Hoy, K., MacDonald, I., Mettler, G., Tsilfidis, C. and Korneluk, R.G. (1993) Decrease in the size of the myotonic dystrophy CTG repeat during transmission from parent to child: implications for genetic counselling and genetic anticipation. Am. J. Med. Genet. 45: 401-407.

Imbert, G., Kretz, C., Johnson, K. and Mandel, J-L. (1993) Origin of the expansion mutation in myotonic dystrophy. Nat. Genet. 4: 72-76.

Jansen, G., de Jong, P.J., Amemiya, C., Aslanidis, C., Shaw, D.J., Harley, H.G., Brook, D.J., Fenwick, R., Korneluk, R.G., Tsilfidis, C., Shutler, G.G., Herman, R., Wormskamp, N.G.M., Smeets, H.J.M. and Wieringa, B. (1992a) Physical and genetic characterization of the distal segment of the myotonic dystrophy area on 19q. Genomics 13: 509-517.

Jansen, G., Mahadevan, M., Amemiya, C., Wormskamp, N., Segers, B., Hendriks, W., O'Hoy, K., Baird, S., Sabourin, L., Lennon, G., Jap, P.L., Iles, D., Coerwinkel, M., Hofker, M., Carrano, A.V., de Jong, P.J., Korneluk, R.G. and Wieringa, B. (1992b) Characterization of the myotonic dystrophy region predicts multiple protein isoform-encoding mRNAs. Nat. Genet. 1: 261-266.

Jeffreys, A.J., Neumann, R. and Wilson, V. (1990) Repeat unit sequence variation in minisatellites: a novel source of DNA polymorphism for studying variation and mutation by single molecule analysis. Cell 60: 473-485.

Jeffreys, A.J., Wilson, V. and Thein, S.L. (1985) Hypervariable 'minisatellite' regions in human DNA. Nature 314: 67-73.

Kerem, B-S., Rommens, J.M., Buchanon, J.A., Markiewiez, D., Cox, T.K., Chakravarti, A., Buchwald, M. and Tsui, L-C. (1989) Identification of the cystic fibrosis gene: genetic analysis. Science 245: 1073-1080.

Knight, S.J.L., Flannery, A.V., Hirst, M.C., Campbell, L., Christodoulou, Z., Phelps, S.R., Pointon, J., Middleton-Price, H.R., Barnicoat, A., Pembrey, M.E., Holland, J., Oostra, B.A., Bobrow, M. and Davies, K.E. (1993) Trinucleotide repeat amplification and hypermethylation of a CpG island in FRAXE mental retardation. Cell 74: 127-134.

Koch, M.C., Grim, T., Harley, H.G. and Harper, P.S. (1991) Genetic risks for children with myotonic dystrophy. Am. J. Hum. Genet. 48: 1084-1091.

Kornberg, J.R. and Rykowski, M.C. (1988) Human genome organization: *Alu*, lines and the molecular structure of metaphase chromosome bands. Cell 53: 391-400.

Korneluk, R.G., MacKenzie, A.E., Nakamura, Y., Dube, I., Jacob, P. and Hunter, A.G.W. (1989a) A reordering of human chromosome 19 long arm markers and identification of markers flanking the myotonic dystrophy locus. Genomics 5: 596-604.

Korneluk, R.G., MacLeod, H.L., McKeithan, T.W., Brook, D.J. and MacKenzie A.E. (1989b) A chromosome 19 clone from a translocation breakpoint shows close linkage and linkage disequilibrium with myotonic dystrophy. Genomics 4: 146-151.

Korneluk, R.G., Tsilfidis, C., Shutler, G., Mahadevan, M., Bailly, J. and Surh, L.C. (1991) A three allele insertion polymorphism is identified by the human chromosome 19q13.3 probe pKBE0.8 (D19S119). Nucl. Acids Res. 19: 1157.

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.

Kudo, S. and Fukudo, M. (1989) Structural organization of glycophorin A and B genes: glycophorin B gene evolved by homologous recombination at *Alu* repeat sequences. Proc. Natl. Acad. Sci. USA 86: 4619-4623.

Lusis, A.J., Heinzmann, C., Sparkes, R.S., Gellar, R., Sparkes, M.C. and Mohandas, T. (1985) Regional mapping on chromosome 19: apolipoprotein E, apolipoprotein CII, low density lipoprotein (LDL) receptor, peptidase D, glucose phosphate isomerase. Cytogenet. Cell Genet. 40: 683.

MacDonald, M.E., Novelletto, A., Lin, C., Tagle, D., Barnes, G., Bates, G., Taylor, S., Allitto, B., Altherr, M., Myers, R., Lehrach, H., Collins, F.S., Wasmuth, J.J., Frontali, M. and Gusella, J.F. (1992) The Huntington's disease candidate region exhibits many different haplotypes. Nat. Genet. 1: 99-103.

MacKenzie, A.E., MacLeod, H.L., Hunter, A.G.W. and Korneluk, R.G. (1989) Linkage analysis of the apolipoprotein C2 gene and myotonic dystrophy on human chromosome 19 reveals linkage disequilibrium in a French Canadian population. Am. J. Hum. Genet. 44: 140-147.

MacKenzie, A.E., MacLeod, H.L., Leblond, S.C., Monteith, N., Lahey, D. and Korneluk, R.G. (1991) DNA polymorphism and linkage disequilibrium within the apolipoprotein CII locus on human chromosome 19. Hum. Hered. 41: 188-194.

Madisen, L., Hoar, D.I., Holroyd, C., Crisp, M. and Hodes, M.E. (1987) DNA banking: the effects of storage of blood and isolated DNA on the integrity of DNA. Am. J. Med. Genet. 27: 379-390.

Mahadevan, M., Tsilfidis, C., Sabourin, L., Shutler, G.G., Amemiya, C., Jansen, G., Neville, C., Narang, M., Barceló, J., O'Hoy, K., Earle-MacDonald, J., de Jong, P., Wieringa, B. and Korneluk, R.G. (1992) Myotonic dystrophy mutation: an unstable CTG repeat in the 3' untranslated region of the gene. Science 255: 1253-1255.

Mahadevan, M.S., Foitzik, M.A., Surh, L.C. and Korneluk, R.G. (1993a) Characterization and polymerase chain reaction (PCR) detection of an *Alu* deletion polymorphism in total linkage disequilibrium with myotonic dystrophy. Genomics 15: 446-448.

Mahadevan, M.S., Amemiya, C., Jansen, G., Sabourin, L., Baird, S., Neville, C.E., Wormskamp, N., Segers, B., Batzer, M., Lamerdin, J., de Jong, P., Wieringa, B. and Korneluk, R.G. (1993b) Structure and genomic sequence of the myotonic dystrophy (DM kinase) gene. Hum. Mol. Genet. 2: 299-304.

Mathieu, J., De Braekeleer, M. and Prévost, C. (1990) Genealogical reconstruction of myotonic dystrophy in the Saguenay-Lac-Saint-Jean area (Quebec, Canada). Neurol. 40: 839-842.

Miller, S.A., Dykes, D.D. and Polesky, H.F. (1988) A simple salting out procedure for extracting DNA from human nucleated cells. Nucl. Acids Res. 16: 1215.

Myers, R.H., MacDonald, M.E., Koroshetz, W.J., Duyao, M.P., Ambrose, C.M., Taylor, S.A.M., Barnes, G., Srinidhi, J., Lin, C.S., Whaley, W.L., Lazzarini, A.M., Schwarz, M., Wolff, G., Bird, E.D., Vonsattel, J-P.G. and Gusella, J.F. (1993) *De novo* expansion of a $(CAG)_n$ repeat in spordic Huntington's disease. Nat. Genet. 5: 168-173.

Nakamura, Y., Leppert, M., O'Connell, P., Wolff, R., Holm, T., Culver, M., Martin, C., Fujimoto, E., Hoff, M., Kumlin, E. and White, R. (1987) Variable number of tandem repeats (VNTR) markers for human gene mapping. Science 235: 1616-1622.

Neville, C.E., Mahadevan, M., Tsilfidis, C., MacKenzie, A.E. and Korneluk, R.G. (1992) High resolution genetic mapping of the DM locus suggests few ancestral mutations. Am. J. Hum. Genet. 51: A610.

O'Hoy, K.L., Tsilfidis, C., Mahadevan, M.S., Neville, C.E., Barceló, J., Hunter, A.G.W. and Korneluk, R.G. (1993) Reduction in size of the myotonic dystrophy trinucleotide repeat mutation during transmission. Science 259: 809-812.

Oberlé, I., Rousseau, F., Heitz, D., Kretz, C., Devys, D., Hanauer, A., Boué, 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.

Orr, H.T., Chung, M-Y., Banfi, S., Kwiatkowski Jr., T.J., Servadio, A., Beaudet, A.L., McCall, A.E., Duvick, L.A., Ranum, L.P.W. and Zoghbi, H.Y. (1993) Expansion of an unstable trinucleotide CAG repeat in spinocerebellar ataxia type 1. Nat. Genet. 4: 221-226.

Oudet, C., von Koskull, H., Nordström, A.M., Peippo, M. and Mandel, J.L. (1993) Striking founder effect for the fragile X syndrome in Finland. Eur. J. Hum. Genet. 1: 181-189.

Pembrey, M.E., Winter, R.M. and Davies, K.E. (1985) A premutation that generates a defect at crossing over explains the inheritance of fragile X mental retardation. Am. J. Med. Genet. 21: 709-717.

Penrose, L.S. (1948) The problem of anticipation in pedigrees of dystrophia myotonica. Ann. Eugen. (London) 14: 125-232.

Richards, R.I., Holman, K., Friend, K., Kremer, E., Hiller, D., Staples, A., Brown, W.T., Goonewardena, P., Tarleton, J., Schwartz, C. and Sutherland, G.R. (1992) Evidence of founder chromosomes in fragile X syndrome. Nat. Genet. 1: 257-260.

Richards, R.I. and Sutherland, G.R. (1992a) Dynamic mutations: a new class of mutations causing human disease. Cell 70: 709-712.

Richards, R.I. and Sutherland, G.R. (1992b) Heritable unstable DNA sequences. Nat. Genet. 1: 7-9.

Riordan, J.R., Rommens, J.M., Kerem, B-S., Alon, N., Rozmahel, R., Grzelczak, Z., Zielenski, J., Lok, S., Plavsic, N., Chou, J-L., Drumm, M.L., Iannuzzi, M.C., Collins, F.S. and Tsui, L.C. (1989) Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. Science 245: 1066-1073.

Rommens, J.M., Iannuzzi, M.C., Kerem, B-S., Drumm, M.L., Melmer, G., Dean, M., Rozmahel, R., Cole, J.L., Kennedy, D., Hidaka, N., Zsiga, M., Buchwald, M., Riordan, J.R., Tsui, L-C. and Collins, F.S. (1989) Identification of the cystic fibrosis gene : chromosome walking and jumping. Science 245: 1059-1065.

Royle, N.J., Clarkson, R.E., Wong, Z. and Jeffreys, A.J. (1988) Clustering of hypervariable minisatellites in the proterminal region of human autosomes. Genomics 3: 352-360.

Sabourin, L.A., Mahadevan, M.S., Narang, M., Lee, D.S.C., Surh, L.C. and Korneluk, R.G. (1993) Effect of the myotonic dystrophy (DM) mutation on mRNA levels of the DM gene. Nat. Genet. 4: 233-238.

Sarnat, H.B. and Silbert, S.W. (1976) Maturational arrest of fetal muscle in neonatal myotonic dystrophy: a pathologic study of four cases. Arch. Neurol. 33: 466-474.

Schlötterer, C. and Tautz, D. (1992) Slippage synthesis of simple sequence DNA. Nuc. Acids Res. 20: 211-215.

Shaw, D.J., Meredith, A.L., Sarfarazi, M., Huson, S.M., Brook, J.D., Myklebost, O. and Harper, P.S. (1985) The apolipoprotein C2 gene: subchromosomal localization and linkage to the myotonic dystrophy locus. Hum. Genet. 70: 271-273.

Sherman, S.L., Morton, N.E., Jacobs, P.A. and Turner, G. (1984) The marker (X) syndrome: a cytogenetic and genetic analysis. Ann. Hum. Genet. 48: 21-37.

Sherman, S.L., Jacobs, P.A., Morton, N.E., Froster-Iskenius, U., Howard-Peebles, P.N., Nielsen, K.B., Partington, M.W., Sutherland, G.R., Turner, G. and Watson, M. (1985) Further segregation analysis of the fragile X syndrome with special reference to transmitting males. Hum. Genet. 69: 289-299.

Shutler, G., Leblond, S., Bailly, J., MacKenzie, A.E., Tsilfidis, C. and Korneluk, R.G. (1991) An insertion polymorphism identified by the probe pE0.8 (D19S115) at 19q13.3. Nuc. Acids Res. 19: 1159.

Shutler, G.G., Korneluk, R.G., Tsilfidis, C., Mahadevan, M., Bailly J., Smeets, H., Jansen, G., 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 13: 518-525.

Smeets, H.J., Hermens, L.R., Brunner, H.G. and Wieringa, B. (1991) Identification of variable simple sequence motifs (VSSMs) in 19q13.2-qter: markers for the myotonic dystrophy locus. Genomics 9: 257-263.

Soussi-Yanicostas, N., Chevallay, M., Laurent-Winter, C., Tome, F.M., Fardeau, M. and Butler-Browne, G.S. (1991) Distinct contractile protein profile in congenital myotonic dystrophy and X-linked myotubular myopathy. Neuromusc. Dis. 1: 103-111.

Steinmetz, M., Stephan, D. and Lindahl, K.F. (1986) Gene organization and recombinational hotspots in the murine major histocompatibility complex. Cell 44: 895-904.

The Huntington's Disease Collaborative Research Group (1993) A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. Cell 72: 971-983.

Tsilfidis, C., MacKenzie, A.E., Shutler, G., Leblond, S., Bailly J., Johnson, K., Shelbourne, P., Williamson, R., Siegel-Bartelt, J. and Korneluk, R.G. (1991) A phase known recombination confirms the location of D19S51 distal to the myotonic dystrophy locus on 19q. Am. J. Hum. Genet. 49: 961-965.

Tsilfidis, C., MacKenzie, A.E., Mettler, G., Barceló, J. and Korneluk, R.G. (1992) CTG trinucleotide repeat amplification patterns in the myotonic dystrophy (DM) gene: correlation between repeat length and frequency of severe congenital myotonic dystrophy. Nature Genet. 1: 192-195.

Tsilfidis, C., MacKenzie, A., Shutler, G., Neville, C.E., Mahadevan, M., Wood, S., Andermann, E. and Korneluk, R.G. Haplotype analysis and linkage disequilibrium patterns in myotonic dystrophy (DM). (manuscript in preparation)

Verkerk, A.J.M.H., Pierreti, 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.

Weber, J.L. (1990) Informativeness of human $(dC\text{-}dA)_n \cdot (dG\text{-}dT)_n$ polymorphisms. Genomics 7: 524-530.

Whitehead, A.S., Solomon, E., Chambers, S.P., Povey, S. and Bodmes, W.F. (1982) Assignment of the structural gene for the third complement component of human complement to chromosome 19. Proc. Nat. Acad. Sci. USA 79: 5021-5025.

Yamagata, H., Miki, T., Ogihara, T., Nakagawa, M., Higuchi, I., Osame, M., Shelbourne, P., Davies, J. and Johnson, K. (1992) Expansion of unstable DNA region in Japanese myotonic dystrophy patients. Lancet 339: 692.

Zhu, Q-S., Heisterkamp, N. and Groffen, J. (1990) Unique organization of the human BCR gene promotor. Nucl. Acids Res. 18: 7119-7125.

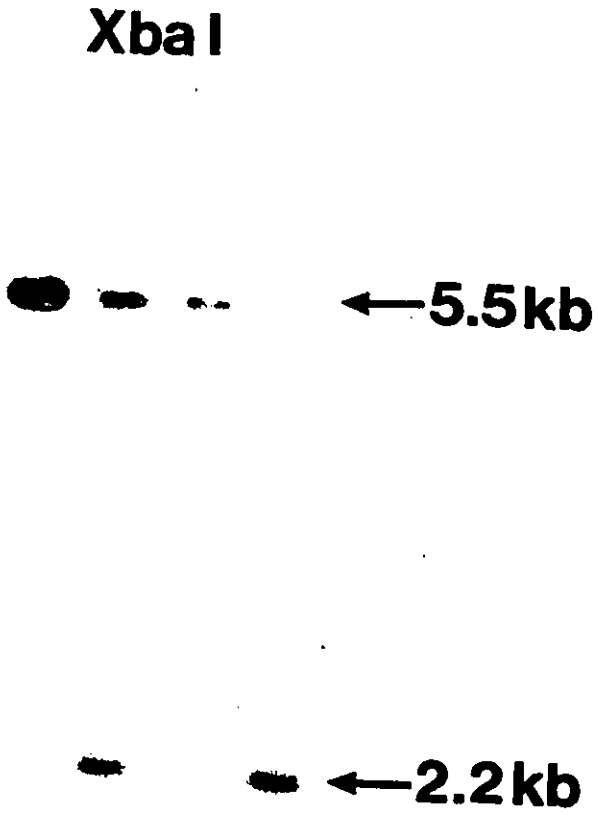
Zuliani, G. and Hobbs, H.H. (1990) A high frequency of length polymorphisms in repeated sequences adjacent to *Alu* sequences. Am. J. Hum. Genet. 46: 963-969.

APPENDIX I

An insertion polymorphism is identified by the probe pKH1.1 at the locus D19S196 at human chromosome 19q13.3. A two allele insertion polymorphism is detected by TaqI, BanI (3.8 kb, 3.5 kb) or NcoI (4.4 kb, 4.1 kb) restriction endonucleases. Allele sizes generated from TaqI restricted genomic DNA are provided in the figure. The allele frequency in the normal population (395 unrelated individuals) was 0.69 and 0.31 for the larger and smaller fragments respectively. A codominant segregation pattern was observed in over 100 DM families. Close linkage is observed between pKH1.1 and the DM locus ($Z_{max} > 20.00$, $\theta_{max} = 0.00$). The probe pKH1.1 is a 1.1 kb HincII genomic fragment isolated from a cosmid containing human chromosome 19 sequences which mapped to a genomic region approximately 205 kb telomeric to the locus identified by the probe pE0.8 (Shutler et al., 1991). The physical linkage of pKH1.1 with pE0.8 is based upon the isolation of contiguous DNA sequences derived from a chromosome walk in a human genomic cosmid library (Shutler et al., 1992). Genomic DNA (5μg) was digested with TaqI, separated by electrophoresis on 0.8% agarose gels and transferred to nylon membranes. Southerns were probed with pKH1.1, washed in 0.2X SSC with 0.1% SDS at 55°C and exposed to x-ray film for 1 to 4 days.

Taq I

An XbaI RFLP identified by the probe pKH1.3 maps to the locus D19S194 at human chromosome 19q13.3. The restriction endonuclease XbaI identifies a two allele polymorphism with fragment sizes of 5.5 kb and 2.2 kb, as indicated on the figure. The allele frequencies in the normal population (233 unrelated individuals) are 0.66 and 0.34 for the larger and smaller fragments respectively. A codominant segregation pattern was observed in over 100 DM families. Close linkage is observed between pKH1.3 and the DM locus ($Z_{max} > 20.00$, $\theta_{max} = 0.00$). The probe pKH1.3 is a 1.3 kb HincII genomic fragment isolated from a human chromosome 19 cosmid library as a result of a chromosome walk initiated from pE0.8 (Shutler et al., 1992). The probe pKH1.3 is homologous to a genomic sequence which is located approximately 210 kb telomeric to pE0.8 (Shutler et al., 1991). Genomic DNA (5 µg) was digested with XbaI, separated by electrophoresis on 0.8% agarose gels and transferred to nylon membranes. Southerns were probed with pKH1.3, washed in 0.2X SSC with 0.1% SDS at 50 °C and exposed to x-ray film for 1 to 4 days.


XbaI
5.5kb
2.2kb

APPENDIX II

RAW DATA FOR χ^2 TEST OF ALLELIC ASSOCIATION BETWEEN DM AND NINE PCR-BASED POLYMORPHISMS

Sequence Polymorphism	No. Chromosomes Normal Population		Total No. Normal Chromosomes	No. Chromosomes DM Population		Total No. DM Chromosomes
	Allele 1	Allele 2		Allele 1	Allele 2	
DMR-N9 (G/T) DraIII	101	57	158	52	1	53
DMK (G/T) HphI	17	113	130	1	53	54
DMK (C/T) HhaI	74	91	165	52	1	53
DMK (Δ 1 kb)	148	132	280	253	5	258
DMK (G/T) HinfI	103	92	195	2	51	53
DMK (C/G) BpmI	34	138	172	1	53	54
DMK (G/T) Fnu4HI	78	65	143	49	1	50
DMK (G/T) ASO* assay	(G)56	(T)46	102	(G)0	(T)36	36
pCN400 (TaqI) D19S463	98	101	199	1	44	45

*allele-specific oligonucleotide

APPENDIX III

PUBLICATIONS

Refereed Papers

Mahadevan, M.S., Tsilfidis, C., Sabourin, L., Shutler, G., Amemiya, A., Jansen, G., **Neville, C.E.,** Narang, M., Barceló, J., O'Hoy, K., Leblond, S., Earle-MacDonald, J., de Jong, P., Wieringa, B. and Korneluk, R.G. (1992) Myotonic dystrophy mutation: an unstable CTG repeat in the 3' untranslated region of the gene. Science 255: 1253-1255.

Mahadevan, M.S., Amemiya, C., Jansen, G., Sabourin, L., Baird, S., **Neville, C.E.,** Wormskamp, N., Segers, B., Batzer, M., Lamerdin, J., de Jong, P., Wieringa, B. and Korneluk, R.G. (1993) Structure and genomic sequence of the myotonic dystrophy (DM kinase) gene. Hum. Molec. Genet. 2: 299-304.

O'Hoy, K., Tsilfidis, C., Mahadevan, M., **Neville, C.E.,** Barceló, J., Hunter, A.G.W. and Korneluk, R.G. (1993) Reduction in size of the myotonic dystrophy trinucleotide repeat mutation during transmission. Science 259: 809-812.

Neville, C.E., Mahadevan, M.S., Barceló, J.M. and Korneluk, R.G. (1994) High resolution genetic analysis suggests one ancestral predisposing haplotype for the origin of the myotonic dystrophy mutation. Hum. Molec. Genet. 3: 45-51.

Abstracts

Korneluk, R.G., Bailly, J., Shutler, G., Tsilfidis, C., Mahadevan, M., **Neville, C.E.,** Narang, M., Foitzik, M. and MacKenzie, A.E. (1991) Progress in cloning the myotonic dystrophy gene. Am. J. Hum. Genet. 49: A404.

Tsilfidis, C., **Neville, C.E.,** MacKenzie, A.E., Shutler, G., Mahadevan, M. and Korneluk, R.G. (1991) Haplotype analysis of a French Canadian myotonic dystrophy (DM) population suggests the DM gene is contained within a 350 kb region of chromosome 19q13.3. Am. J. Hum. Genet. 49: A2036.

Shutler, G., Tsilfidis, C., **Neville, C.E.** and Korneluk, R.G. (1991) Characterization of a CpG-rich island from a region of human chromosome 19q13.3 which contains the myotonic dystrophy (DM) gene. Am. J. Hum. Genet. 49: A2370.

O'Hoy, K., Tsilfidis, C., Mahadevan, M., **Neville, C.E.,** Barceló, J., Hunter, A.G.W. and Korneluk, R.G. (1992) Myotonic dystrophy: a gene conversion event results in loss of the mutation and disease. Am. J. Hum. Genet. 51: A150.

Shutler, G., MacKenzie, A.E., **Neville, C.E.,** Mahadevan, M., Tsilfidis, C., Jansen, G., de Jong, P., Wieringa, B. and Korneluk, R.G. (1992) Correlation of the consensus genetic and physical maps of the myotonic dystrophy (DM) region. Am. J. Hum. Genet. 51: A1578.

Neville, C.E., Mahadevan, M., Tsilfidis, C., MacKenzie, A.E. and Korneluk, R.G. (1992) High resolution genetic mapping of the DM locus suggests few ancestral mutations. Am. J. Hum. Genet. 51: A610.