National Library of Canada
Acquisitions and Bibliographic Services Branch
395 Wellington Street
Ottawa, Ontario
K1A 0N4

Bibliothèque nationale du Canada
Direction des acquisitions et des services bibliographiques
395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Your file Votre référence

Our file Notre référence

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.

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.

Canada

GENETIC AND PHYSICAL MAPPING OF THE MYOTONIC DYSTROPHY LOCUS ON HUMAN CHROMOSOME 19q13.3

A Thesis Submitted to the School of Graduate Studies University of Ottawa in partial fulfilment of the requirements for the degree of Doctor of Philosophy, Department of Microbiology and Immunology, Faculty of Medicine.

By Gary G. Shutler.

ISBN 0-315-82544-8

UNIVERSITÉ D'OTTAWA
UNIVERSITY OF OTTAWA

ABSTRACT

Myotonic dystrophy (DM) is an autosomal dominant, multisystemic disorder characterized primarily by myotonia and progressive muscle weakness. It is the most common inherited neuromuscular disease affecting both adults and children. DM shows a marked variability in expression ranging from a severe congenital form, often fatal in newborns, to an asymptomatic condition associated with normal longevity. Furthermore, an increase in the severity of the disease with successive generations has been suggested (termed genetic anticipation). The relative frequency, multisystemic involvement and variable expression are features of the disease that are poorly understood at the molecular and biochemical level. A better understanding of the disease would be facilitated by the cloning and subsequent characterization of the DM gene. Since there are no chromosomal aberrations or obvious biochemical abnormalities associated with the disease, cloning of the DM locus relied upon "positional cloning" strategies to isolate the gene based on its chromosomal location.

The cloning of the DM locus was accomplished in a three part research plan: 1) the characterization of the DNA excision repair cross complementation (ERCC1) gene region by genetic and physical mapping to determine the location of the DM gene relative to this locus, 2) the undertaking of a chromosome walk from the ERCC1 region toward the DM gene to define a minimal area that is to contain the DM locus, and 3) characterization of the essential region containing the DM locus for CpG islands and DM associated abnormalities.

Previous studies employing both linkage analysis and somatic cell hybrid mapping established the order of chromosome 19 long arm markers to be: 19qcen - pα1.4 - APOC2 - CKM - DM - D19S51 - D19S50 - 19qter and had identified the locus for the muscle creatine kinase gene (CKM) as the most closely linked marker to DM. Furthermore, the ERCC1 gene had been physically mapped distal to CKM. To map the position of ERCC1 relative to the DM locus, an informative polymorphism at or near the ERCC1 locus was essential. A polymorphic marker, pE0.8 (D19S115), was isolated for this purpose from a cosmid containing ERCC1 genomic sequences. Physical mapping of this marker placed it between CKM and ERCC1, about 15 kb centromeric from the ERCC1 locus. This marker was found to be closely linked to DM. Furthermore, a recombinant event identified in a Dutch DM family showed pE0.8 and an ERCC1 variable simple sequence marker to co-segregate with more proximal markers (D19S37, APOC2) and recombine with DM. This data clearly showed DM to be distal to ERCC1. A second key recombinant event was identified in a French Canadian DM family that mapped the DM locus proximal to D19S51. Thus the two recombinant events that flank the DM locus defined a segment of human chromosome 19q13.2-3 of less than 2 cM that was expected to contain the DM gene.

Based on the data showing DM to be distal to the markers at the ERCC1 locus and the tight linkage of D19S115 to DM it was reasoned that pE0.8 would be an appropriate start point for a chromosome walk. By using primarily ALU PCR generated probes, a cosmid contig of approximately 350 kb was constructed from a chromosome

19 cosmid library provided by the Human Genome Centre of the Lawrence Livermore National Laboratory (LLNL) in Livermore, California. The results of the chromosome walk enabled a construction of a high resolution genetic map in conjunction with a complementary long range physical map that was used to define the DM region. The proximal crossover event identified in a Dutch DM family was mapped to a 60 kb segment between two markers, pKEX0.8 (D19S118) and pKBE0.8 (D19S119), making the polymorphism at D19S118 the closest proximal marker to DM. This result also eliminated over 200 kb of sequence that would have otherwise had to be screened for DM candidate genes. At the distal end of the chromosome walk, the contig could not be extended further in a telomeric direction due to an absence of clones with adjacent sequence in the chromosome 19 library.

To address the cloning limitations of the chromosome 19 cosmid library, a collaborative chromosome walk was undertaken by a Dutch research team from a newly cloned marker, pX75b (D19S112) mapping proximal to D19S51 but distal to the first DM region cosmid contig. After cloning about 200 kb of cosmid DNA sequence no overlap with our proximal DM cosmid contig could be found again due to the same chromosome 19 cosmid library cloning deficiency for this region. To resolve the gap between the two cosmid contigs, sequence tagged sites (STSs) from the both contigs were used to screen a Yeast Artificial Chromosome (YAC) library in a collaboration with the Human Genome Centre at the LLNL in California. YAC clones bridging the two cosmid contigs were isolated and cosmid libraries constructed from these clones. Ultimately, one large cosmid

contig was established spanning a region from pE0.8 (D19S115) to just short of D19S51, a distance of nearly 700 kb.

The first step in the characterization of the essential region of the cosmid contig (the minimal region expected to contain the DM locus) was to correlate the mapping of rare cutting endonuclease sites in the cosmids with genomic data from pulsed field electrophoretic analysis. By proving the integrity of the cosmid contig, the cloning of the region of 19q13.3 which contained the DM locus was confirmed. From a broad perspective, this enabled a comparison of genetic and physical mapping data from a region in excess of 1 Mb spanning the DM locus. This data demonstrated a correlation between these two parameters only if marker intervals with high lod scores (over 22) are considered. From a more specific perspective, proof of the cosmid contig integrity permitted fine genetic and physical mapping to define the minimal region that contained the DM locus. It is from this analysis that two DNA probes were cloned, pGB2.6 and pGP1.5, which identified a heritable unstable DNA sequence unique to DM affected individuals. In normal individuals these probes identified a 9 or 10 kb EcoRI fragment (8.5 and 9.5kb HindIII fragments) but in DM patients expanded alleles from 10.5 to 16 kb were found.

Extensive linkage analysis of the heritable unstable DNA sequence with DM was conducted in our laboratory on DM families. In 70% of approximately 200 DM patients increased allele sizes ranging from 10.5 to 16 kb was revealed. Evidence from three

generation DM families showed that the increase in expansion of the allele size correlated with an increase in the severity of the disease. These observations from the analysis with pGB2.6, pGP1.5 and other DNA probes mapping to this locus (found by other members of an international DM working group) led to the proposal that the heritable unstable DNA sequence is the molecular basis of genetic anticipation in DM similar to an allelic expansion which was shown to be the molecular defect in the fragile X syndrome.

Work done in our laboratory and by others have shown the DM associated allelic expansion to be due to the amplification of a trinucleotide repeat, CTG. This repeat was mapped to the 3' untranslated region of a gene which based on sequence homology comparisons encodes a putative serine-threonine protein kinase. This is not unlike the allelic expansion found in fragile X syndrome that is due to an amplification of a CGG trinucleotide repeat mapping to the 5' untranslated region of a gene designated FMR-1. The size of the CTG repeats ascertained in 124 normal chromosomes was found to range from 5 to 30 with repeat numbers of 5 and 13 the most common. The repeat numbers in DM chromosomes was found to vary from a minimum of about 50 to over 2000. Only two out of 98 DM families did not show allelic expansion using Southern blot analysis or PCR protocols to ascertain the repeat numbers. These cases either have other mutations at or near this locus or they have another clinically similar disorder mapping elsewhere in the genome.

In summary, it is evident from work presented in this thesis that the DM locus has been cloned and that the DM mutation has been identified.

ACKNOWLEDGEMENTS

I thank Dr. Robert G. Korneluk for his guidance and support during these studies and Dr. Alexander E. MacKenzie for his expertise most notably in genetic linkage analysis. I am grateful to Dr. Peter de Jong and his laboratory for the chromosome 19 libraries and the YAC cloning work. I also thank Dr. Be' Wieringa, Dr. Bert Smeets and Gert Jansen for the exchanges of information and DNA. I especially thank my many colleagues and friends in our laboratory for their support and friendship. I thank the Royal Canadian Mounted Police for allowing me to pursue these studies.

DEDICATION

To my wife Carolyn, my children Graeme, Brett and Lindsay, and my parents George and Tena.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER III:
PHYSICAL AND GENETIC MAPPING OF pE0.8 (D19S115), A CHROMOSOME 19 ERCC1 MARKER SHOWING CLOSE LINKAGE WITH MYOTONIC DYSTROPHY

CHAPTER IV:
PHYSICAL MAPPING AND CLONING OF THE MYOTONIC DYSTROPHY GENE REGION

CHAPTER V:
IDENTIFICATION OF THE DM GENE

LIST OF FIGURES

CHAPTER II:
GENERAL MATERIALS AND METHODS

CHAPTER III:
PHYSICAL AND GENETIC MAPPING OF pE0.8 (D19S115), A CHROMOSOME 19 ERCC1 MARKER SHOWING CLOSE LINKAGE WITH MYOTONIC DYSTROPHY

CHAPTER IV:
PHYSICAL MAPPING AND CLONING THE MYOTONIC DYSTROPHY GENE REGION

CHAPTER V:
IDENTIFICATION OF THE DM MUTATION

LIST OF ABBREVIATIONS

A	Adenine base in a nucleotide sequence
APOC2	Apolipoprotein C2 gene
bp	base pair
BRL	Bethesda Research Laboratories
°C	degrees Celsius
C	Cytosine base in a nucleotide sequence
cDNA	complementary DNA
CHEF	contour clamped homogeneous electric field
Ci	Curies
CKM	muscle specific Creatine Kinase gene
cpm	counts per minute
dATP	deoxyadenosine triphosphate
dCTP	deoxycytidine triphosphate
dGTP	deoxyguanosine triphosphate
DM	dystrophia myotonica/myotonic dystrophy
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DTT	dithiothreitol
dTTP	deoxythymidine triphosphate

EDTA	ethylenediamenetetraacetic acid disodium salt
EMG	electromyography
ERCC1	excision cross complementation repair gene 1
G	guanine base in a nucleotide sequence
h	hour
Inc.	incorporated
IPTG	isopropyl-1-thio-beta-D-galactoside
kb	kilobase
l	liter
LB	Luria-Bertani medium, tryptone, yeast extract, NaCl
lod	log odds
min	minute
ml	millilitre
mM	millimolar
mmol	millimoles
mRNA	messenger ribonucleic acid
ng	nanogram
NZCYM	NZ amine, NaCl, casamino acids, yeast extract, $MgSO_4$
pg	picogram
PCR	polymerase chain reaction
PEG	polyethylene glycol
PFG	pulsed field gel

pfu	plaque forming unit
PMSF	phenylmethylsulfonyl fluoride
RNA	ribonucleic acid
s	seconds
SDS	sodium dodecyl sulphate
SSC	standard saline citrate buffer
T	thymidine base in a nucleotide sequence
TBE	Tris-borate, EDTA buffer
TE	Tris-EDTA buffer
Tris	tris(hydroxymethyl)aminomethane
uv	ultraviolet
U	uracil base in a nucleotide sequence
μCi	microcuries
μg	microgram
μmol	micromoles
YAC	yeast artificial chromosomes
Z_{max}	maximum lod score
θ	recombination fraction

Chapter I: Introduction To Myotonic Dystrophy and Molecular Genetics

Myotonic dystrophy or Dystrophia Myotonica (DM) is the most common form of muscular dystrophy affecting adults and children with an prevalence of 1 in about 8000 individuals (Harper, 1989). The disease is a genetic disorder with an autosomal dominant inheritance pattern. DM is distributed worldwide including all races. French Canadians have the highest reported prevalence of DM, with frequencies of 1 in 529 in some areas (Mathieu et al., 1990). This situation provides a unique opportunity for the study of DM in the population of Eastern Ontario and Western Quebec.

Clinical features of DM

The clinical definition of DM is very heterogenous and involves many combinations of features reflecting the multi-systemic nature of the disease (reviewed by Harper, 1989). A characteristic DM feature is myotonia, an electrophysiological disturbance resulting in delayed muscle relaxation. Furthermore, patients can display muscle weakness, typically involving the muscles of the face, jaw, neck, arms and legs. Smooth as well as voluntary muscles are involved with these symptoms. Myotonia can be elicited in almost every patient and 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. DM patients can also show the complete pattern of muscle weakness and wasting. Other reported complications of this disease include cardiac conduction disturbances, cataracts, retinal degeneration, premature balding, abnormal carbohydrate metabolism, gonadal atrophy, infertility and mental impairment. This multi-systemic nature of the disease suggests that the DM gene is expressed in a variety of tissues.

There are two distinct clinical forms of DM expression, an adult onset and a congenital form. Early studies of patients with myotonia resulted in confusion between two diseases, Thomsen's disease or myotonia congenita and Curschmann-Steinert or Dystrophia Myotonica (reviewed by Harper, 1989). Initially, it was thought that DM could not occur during infancy. Thus the congenital form of DM was thought to be a type of Thomsen's disease. Genetic data from generations of DM-affected families led to the discovery of the severe congenital DM.

The clinical presentation of congenital DM differs significantly from adult-onset DM with distinguishing features of congenital DM that are dramatic and often fatal. Affected newborns are extremely weak, have difficulty sucking and swallowing, and usually have severe respiratory problems. If the patient survives, there is a tendency to improve during childhood, followed by a gradual deterioration during adolescence with eventual manifestations characteristic of the adult-onset form. The early absence of myotonia and other features in the congenital-onset form of DM is notable in

comparison to the adult form of DM. Furthermore maternal transmission of the DM gene is a prerequisite of congenital DM. Influence of the transmitting parent's sex is not completely understood. The involvement of imprinting or some other epigenetic phenomena has been suggested (Hall, 1990) while others have postulated the presence 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. Furthermore, in the adult-onset form of DM, there appears to be an increase in the severity of the disease in successive generations. Whether this observation is due to bias in case selection (Penrose et al., 1948) or there exists a biological phenomenon of genetic anticipation (Höweler et al., 1989) has been a point of controversy.

The relative frequency, the multisystem involvement, the two distinct presentations of DM (adult-onset and congenital-onset), the influence of sex of the transmitting parent and genetic anticipation are all features of DM that are poorly understood at the molecular and biochemical level (Hall, 1990). The discovery of the DM gene would permit study of the biological process of the disease, increase accuracy of prenatal diagnosis and could fulfil a prerequisite to improved patient care. Unfortunately, there are no known chromosomal aberrations or obvious biochemical abnormalities associated with DM. Consequently, discovery of the gene relies upon "positional cloning", that is, 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

The strategy of positional cloning relative to finding the DM gene involves a number of essential resources. 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, collectively comprise these essential resources.

1. Genetic Mapping

If two polymorphic markers are consistently passed on together (cosegregate), then they are said to be genetically linked. If linkage between two markers is demonstrated then it can be stated that they are located on the same chromosome. The higher the degree of cosegregation (e.g. the greater the degree of linkage observed) between two markers, the smaller the genetic distance between them. In the case of this thesis genetic linkage between several human chromosome 19q13.3 RFLP's and the DM phenotype is presented.

It is assumed in linkage analysis that crossing over between two homologous chromosomes during meiosis occurs randomly over the length of a chromosome such

that there is a greater chance of recombination between two loci the further apart they are. The probability of a recombination occurring is called the recombinant fraction and is usually denoted by the Greek letter theta, θ. The range of θ extends from $\theta = 0$, when two loci are completely linked, to $\theta = 0.5$, when two loci are far apart and are unlinked.

The direct measure of θ, when the genotype of all family members is known, can also be obtained simply as the proportion of children with a recombinant gamete, however the maximum likelihood approach can be used to obtain an estimate of θ. The dual advantage of a statistical analysis is that it permits families in which the genotype may not be clear in all individuals to be still used to provide some information and it also provides a means to calculate a confidence interval for a given θ. A probability of observing a recombinant gamete, θ, or a nonrecombinant gamete, $1-\theta$, can be combined to give the total probability or likelihood of observing this family given by $L(\theta) = c\theta(1-\theta)$, where the factor, c, depends on the gene frequencies of the alleles. By using regression analysis, with possible values of θ as maximum likelihood estimators, the most accurate estimate of θ is attained when $L(\theta)$ is at a maximum. The larger the sample size the narrower the confidence interval for a given θ. This is ascertained by calculating the lod score. By using the odds ratio of $L(\theta)/L(1/2)$, the likelihood of observed recombination versus the likelihood of no recombination, and converting this to a logarithm, $Z(\max) = \log_{10}[L(\theta)/L(1/2)]$, a lod score can be obtained for a range of possible θ values. In this manner, the value for $L(\theta)$ does not

have to be determined to get the maximum likelihood estimation of θ. The lod scores from many families can be collectively compared with the highest pooled lod score pointing to the best estimate of θ. The difficulty of calculating the maximum likelihood estimation of θ and the lod score manually has been eliminated by the availability of computer programs for linkage analysis (Ott, 1985).

A principle concern in linkage analysis is to map the genetic distance between two loci and determine the order of loci on a chromosome. The value of θ provides an estimate of genetic distance. One centiMorgen represents a $\theta = 0.01$ or a 1% chance of recombination between two genetic markers. Although inaccurate due to variation in recombination frequency for different parts of the genome, one centimorgen is suggested to be the equivalent distance of about one million base pairs. Comparison of the value of θ between sets of different markers provides evidence for the loci order on a chromosome. Instead of two point θ comparisons, it can be more efficient to take a multipoint linkage analysis approach when there is a high density of genetic markers (Lathrop et al., 1985). Three point linkage analysis predicts the order of loci by maximizing likelihoods for recombination over portions of the genome. The strength of the predicted order is shown by the likelihood ratio test of $L(x)/L(\infty)$, the map position versus an infinite or alternative location in the genome. Linkage analysis, for the purpose of mapping genetic distance and for loci order estimation, can be accomplished with either DNA or protein polymorphisms.

The first polymorphism reported to be linked to a human disease was the Lutheran and A,B,H secretor blood group markers to DM (Mohr, 1954; Greenwalt, 1961; Renwick et al., 1971; Harper et al., 1972). Genetic distance between secretor and Lutheran was estimated at $\theta = 0.15$ ($Z_{max} = 18.4$), between secretor and DM, $\theta = 0.05$ ($Z_{max} = 2.37$), and between DM and Lutheran, $\theta = 0.10$ ($Z_{max} = 1.8$). The linkage group was further extended with C3, the third complement component (Weitkamp et al., 1974; Eiberg et al., 1983). By using a monoclonal antibody to C3, which specifically identified somatic cell hybrids expressing C3 from their human chromosome 19 component, the linkage group was assigned to chromosome 19 (Whitehead et al., 1982) thereby indirectly mapping the DM locus. Peptidase D, another blood group marker found linked to DM (Cook et al., 1972) was mapped to chromosome 19 (McAlpine, 1976). ApoE was added to this linkage group (Olaisen et al., 1982 and Gedde-Dahl et al., 1984). Thereafter, other significant genetic markers linked to DM were DNA markers.

The frequency of nucleotide variants in the human genome is sufficiently high to generate polymorphisms for a sufficient number of genetic loci to permit construction of a human gene map (reviewed by Robson, 1988). There are several approaches to detect these polymorphisms. The most familiar method involves hybridization with unique sequence probes which can detect DNA variation in the length of endonuclease restriction fragments. This class of genetic marker is referred to as a restriction fragment length polymorphism (RFLP).

The number of reported RFLP's increases substantially each year. However, the nature of these RFLP's is not as diverse as one might expect. There are only six endonucleases (TaqI, MspI, Eco RI, HindIII, PstI and BglII) accounting for 69% of the 3370 RFLP's reported at the tenth Human Gene Mapping (HGM10) conference (Kidd et al., 1989). TaqI and MspI alone detect 40% of these RFLP's. A possible explanation for the high proportion of TaqI and MspI RFLP's is that the recognition site of both enzymes include the CpG dinucleotide which may be preferentially methylated. Methylation in turn facilitates the mutation of C to T by deamination (Vogel, 1990) resulting in an RFLP. Many RFLP's have been found for linkage analysis study with DM.

The DM linkage shown by protein polymorphisms characterized in the C3 gene product was also evident for the RFLP's detected by the C3 cDNA used as a hybridization probe (Davies et al., 1983). In contrast to the mapping approach of the C3 gene, the gene for apolipoprotein C2 (APOC2) was mapped directly to chromosome 19 by DNA *in situ* hybridization (Hulsebos et al., 1985). A cluster of three apolipoprotein genes have been localized to a 55 kb region on the long arm of chromosome 19 (Myklebost and Rogne, 1988; Smit et al., 1988a). This region has yielded several RFLP's (Humphries et al., 1983; Wallis et al., 1984; Appleby et al., 1986; Korneluk et al., 1987) which have mapped the APOC2 locus to within 2-3 centiMorgans of the DM locus (θ = 0.02 - 0.03, $_{Zmax}$ > 90; Shaw et al., 1985). Linkage disequilibrium detected between APOC2 and DM for the French Canadian

population suggests the presence of a founder effect (MacKenzie et al., 1989). This observation was in agreement with an earlier report on linkage disequilibrium with DM involving the apolipoprotein E locus in this population (Laberge et al., 1985).

Linkage disequilibrium and haplotype analysis are routinely used for fine genetic mapping. The frequency of genotype combinations between two polymorphic loci is generally a function of gene frequency (linkage equilibrium). When two polymorphic loci are close together this independence between genotypes is diminished. Thus a linkage disequilibrium is evident in haplotypes for polymorphic loci that are closely linked. In the case of RFLP's linked to DM, linkage disequilibrium manifests itself by showing a deviation in the prevalence of haplotypes for markers near the DM locus. For example, when linkage disequilibrium with DM is detected at a polymorphic locus it suggests a close proximity to the DM locus. Linkage disequilibrium is calculated by using standard chi-squared formulae or variations thereof resulting in the calculation of a standardized disequilibrium constant (Hill and Robertson, 1968).

RFLP's at the locus for the creatine kinase subunit M gene (CKM), were found to be linked to DM (Brunner et al., 1989). This gene was assigned to 19q13.2-13.3 near APOC2 (Nigro et al., 1987; Schonk et al., 1989). The genetic distance between the CKM locus and the DM locus was suggested to be approximately one centimorgen (θ = 0.01, Z_{max} = 21.76, Korneluk et al., 1989b). Numerous other RFLP's from anonymous loci have also been used for linkage analysis including p17.1 or D19S8

(Shaw et al., 1986) and pJSB11 or D19S16 (Schepens et al., 1987). The D (derivative) number designation represents the locus from which the clone was derived.

RFLP's often fail to be informative in linkage analysis due to a limitation in polymorphic information content (PIC). PIC values can be calculated for a given marker (Botstein et al., 1980) and represent the probability that a given offspring of a random mating between a carrier of a rare dominant gene and a noncarrier is informative for linkage between the locus of the dominant gene and the marker. The value of PIC depends on the number and frequency of alleles detected at that locus. One class of RFLP's, that detect size changes between two endonuclease sites due to the presence of a variable number of tandem sequence repeats (VNTR's, Nakamura et al., 1987; Jefferies et al., 1985), can provide a very high PIC. These are stretches of tandem repeats (9-64 bp) that vary in size by the number of repeats present. Depending on prehybridization and hybridization conditions these RFLP's detect either a single locus or multiple loci (Shutler et al., 1988). DNA fingerprinting is using such markers under conditions of low stringency (Gill et al., 1985) while DNA typing is using such markers under conditions of high stringency (Nakamura et al., 1987). The use of VNTR's in genetic linkage is limited by a nonrandom chromosomal distribution (Royle et al., 1988; Inglehearn and Cooke, 1990). One such marker, pEWRB1 (D19S50) showed linkage to DM but was found to be some distance on the distal side, giving an estimated 9% recombination with DM (Korneluk et al., 1989). Another VNTR marker,

pEFD4.2 (D19S22; Nakamura et al., 1988a, 1988b) is situated even further distal to DM than pEWRB1 (Korneluk et al., 1989b).

Other classes of variable length markers range from the large centromere repeats (Willard et al., 1986) to those with small repeat units of one to four (Litt and Luty, 1989; Tautz, 1989; Weber and May, 1989). It is estimated that for $(GT)_n$ repeats alone over 50,000 copies per haploid genome are present (Hamada et al., 1982). Contrary to the larger VNTR repeats there is yet no evidence for nonrandom distribution (Stallings et al., 1991). The variable simple sequence motifs (VSSM's) can be detected by polymerase chain reaction (PCR) amplification of a unique sequence of DNA flanking the repeat using a radioactive end labelled primer (Saiki et al., 1985). The product is analyzed by polyacrylamide gel electrophoresis and visualized by autoradiography. One of the first applications of this class of markers to linkage analysis was for DM (Smeets et al., 1989). A highly informative $(TG)_n(CA)_n$ motif was identified in the APOC2 gene. Two key VSSM's, one in the ninth intron of the DNA excision repair gene, ERCC1 (Smeets et al., 1991) and at the locus for the anonymous marker, pX75 (Jansen et al., 1992), permitted powerful execution of linkage analysis in the DM region.

In the absence of RFLP's and variable repeat polymorphisms, variation in DNA sequence for genetic mapping can still be identified by several alternative approaches. Within a DNA region that has been well characterized, single base pair changes can

be detected by the use of allele specific oligonucleotides (ASO's) in combination with PCR amplification. Under appropriate conditions a mismatch of the primer's 3' nucleotide with the template DNA will not permit amplification to occur. Alternatively, a DNA segment can be PCR amplified and ASO's radioactively labelled for use in a dot blot assay system. ASO polymorphisms were found during characterization of the APOC2 (Smeets et al., 1988) and CKM (Bailly et al., 1991) genes. There are several other approaches that can be applied to detect single basepair variation not occurring in an endonuclease recognition site. One approach is to use denaturing gradient polyacrylamide gel electrophoresis to detect changes in DNA melting temperature caused by single base substitution (Lerman et al., 1983). Other techniques include detecting heteroduplex formation in PCR products (Surh et al., 1991) and single strand DNA conformation polymorphisms (SSCP's; Orita et al., 1989).

The use of RFLP's in linkage analysis is technologically straightforward and provides sufficient information to genetically map a particular chromosome region containing the disease locus. VSSM's and methods to detect DNA mutation not containing an endonuclease restriction site can also be used for fine resolution genetic mapping and mutation analysis of candidate genes. Complementary to construction of a genetic map, another key resource required for the positional cloning of the DM locus is a long range physical map of the DM region.

2. Physical Mapping

The oldest method for physical mapping is the localization of a cloned gene to a chromosome by *in situ* hybridization (reviewed by Buckle and Craig, 1986). The localization of APOC2 on the long arm of chromosome 19 by *in situ* hybridization (Hulsebos et al., 1985a) in combination with the close genetic linkage to DM detected at this locus (Shaw et al., 1985) indicated the potential diagnostic value of this marker. This mapping result was further complemented by hybridization to a panel of somatic cell hybrids containing various segments of chromosome 19 (Hulsebos et al., 1985b; Lusis et al., 1985; Schonk et al., 1989). Depending on the human chromosome content of a somatic cell hybrid panel, each cell line contains a particular segment of the mapping target chromosome region. Hybridization of a unique sequence DNA probe to such a somatic cell hybrid panel, on dot blots or Southern blots, physically maps that DNA sequence to a particular chromosome region. In this manner two genetic loci could thus be shown to be physically linked on the same chromosome segment.

Somatic cell hybrids are derived by fusing together human and rodent cells. Continuous passage of these cells in culture results in the loss of much of the human chromosome content. Particular cell lines can be selected to form a somatic cell hybrid panel for specific chromosomes and chromosome segments as was done in the physical mapping of the DM region (Brook et al., 1986). To obtain a higher resolution of physical mapping, somatic cell hybrids can be constructed using

cytogenetically identified translocation chromosomes (Schonk et al., 1989; Brook et al., 1991; Korneluk et al., 1989). Increased resolution can be obtained by fusing lethally irradiated somatic cell hybrids with rodent cells resulting in a reduction of the human DNA segment (Benham et al., 1989; Schonk et al., 1989).

The chromosomal breakpoint junction in reciprocal translocations can serve as important genetic and physical landmarks. The somatic cell hybrid, B-9, was generated from lymphocytes containing a chromosome 14:19 translocation associated with chronic lymphocytic leukaemia (CLL). This breakpoint involves the immunoglobin heavy chain locus at 14q32 and the q13.1 band of chromosome 19 (McKeithan et al., 1987). By karyotype analysis the B-9 hybrid was found to contain a t(14;19)(14pter-q32;19q13.1-qter) chromosome. The DNA probe pα1.4, cloned from the 14:19 breakpoint region, detected an RFLP closely linked to DM (θ = 0.03, Z_{max} = 10.44; Korneluk et al., 1989b). This marker was mapped by a somatic cell hybrid panel to 19q12-19q13.2 consistent with the karyotype analysis of B-9 (Korneluk et al., 1989a). Other genetic markers linked to DM were also found to hybridize to the B-9 hybrid cell line. This result made the pα1.4 (D19S37) translocation breakpoint locus a key physical landmark as it showed the DM locus and linked markers to be distal to pα1.4 along the long arm of chromosome 19.

The resolution of physical mapping provided by hybridization to somatic cell hybrids is complemented by restriction mapping with rare cutting endonucleases

utilizing pulsed field gel electrophoresis separation technology. Restriction enzymes that cut genomic DNA infrequently generally have six or eight basepair recognition sequences. Most of these rare cutters have one or two CpG's in their recognition sequence that are relatively rare and frequently methylated. The length of the resulting restriction fragments thus depends upon the frequency of the CpG dinucleotide and it's methylation status along the chromosome. Commonly employed methods of gel electrophoresis permit resolution of DNA restriction fragments to less than 50 kb (Sambrook et al., 1989). The introduction of the concept to use two alternating electric fields to separate large DNA fragments (50 to 2000 kb) opened the possibility of long range restriction mapping (Schwartz and Cantor, 1984). This separation technology was referred to as pulsed field gel (PFG) electrophoresis.

Early endeavours with PFG electrophoresis using nonhomogeneous electric fields featured problems with gel distortions due to velocity differences of DNA molecules with respect to gel position that caused extensive diffusion. It was thought that this type of electric field was necessary to turn the entire DNA molecule in new directions as it migrated through the gel instead of merely alternating the leading and trailing ends of the molecule with homogeneous electric fields (Southern et al., 1987). Application of the principles of electrostatics in using a hexagonal array of electrodes to generate homogeneous electric fields eliminated the nonuniform mobilities (Chu et al., 1986). The contour clamped homogeneous electric field (CHEF) system uses two alternating electric fields at a fixed reorientation angle (Chu et al., 1986) to provide

undistorted separation. The programable, autonomously controlled electrode gel electrophoresis (PACE) system is a more sophisticated version of this system with more electrodes and computer controlled settings (Clark et al., 1988). This separation technology permits accurate sizing of large DNA restriction fragments.

The size of restriction fragments resulting from a rare cutting endonuclease depends upon the methylation status and frequency of the CpG dinucleotide along the chromosome in a given region. The cleavage sites for these endonucleases often cluster at the 5' end of actively transcribed genes (Lindsay and Bird, 1986). Approximately 74% of EagI, SacII and BssHII sites were found to be associated in these CpG islands. Therefore, mapping the concentration of these endonucleases which contain CpG in their recognition sites, could lead to the identification of gene associated regions.

Rare cutting endonucleases restrict DNA into large restriction fragments and when combined with separation by PFG electrophoresis, provide a means to show physical linkage between two cloned genetic loci. In this manner the APOC1, APOC2 and APOE gene cluster was shown to be within 200 kb of pα1.4, the CLL translocation breakpoint (Shaw et al., 1989). Despite the mapping of a 1300 kb region spanning the APOC2 locus, the CKM gene could not be physically linked with the apolipoprotein gene cluster. Physical mapping by PFG electrophoresis provides approximate distance estimates between the genetic loci of linked markers. This type of analysis in

turn complements the physical linkage of genetic markers by cloning strategies such as chromosome walking. The correlation of distance between rare cutting restriction endonuclease sites used as landmarks in cloned and genomic sequence verifies the integrity of a chromosome walk contig.

A chromosome walk involves the systematic isolation of overlapping segments of contiguous DNA (reviewed by Spoerel and Kafatos, 1987). Genomic libraries are screened with a unique sequence probe isolated from the end of the starting genomic clone, leading to the isolation of clones which to varying degrees, extend into the adjacent region. These new clones are used in turn to isolate clones which extend the region further still. A genomic library is a cloned collection of endonuclease restriction fragments, normally several fold redundant, that include all the DNA sequences of a given genome or a specific portion of that genome amplified in an appropriate cloning vector. Construction of a cloned contig, that is, the cloning of adjacent sequences which span two genetic loci complements genetic and physical mapping in the positional cloning of a disease gene. This approach is very labour intensive due to the amount of cloning, restriction mapping and hybridization involved. It becomes feasible only if the genomic library from where the clones are selected contain large stable inserts (e.g. cosmids and YACs) or the distance to be cloned is relatively small (under 100 kb).

The selection of a probe for a chromosome walk step usually involves the mapping and cloning of a unique sequence restriction fragment. This process is dependent upon the distributions of restriction endonuclease sites and unique sequence DNA. Alternatively, probes can be selected by using the polymerase chain reaction with primers derived from the conserved regions of ALU repeat sequence (Karya et al., 1987) to amplify unique sequence inter-ALU regions (de Jong et al., 1987). ALU repeats consist of two monomers forming a unit of about 300 bp that is found on average every 4-7 kb randomly throughout the human genome. When two ALU repeats are in opposite orientation and a sufficient distance apart, the inter-ALU sequence will be amplified by PCR with one of the conserved ALU primers. Insert end fragments to use for probes can also be generated when ALU primers are used in conjunction with primers to vector sequence flanking the insertion site.

A unique sequence probe generated from chromosome walking or other cloning strategies can be used as a Sequence Tagged Site (STS). By determining the sequence of such a DNA probe, primers can be designed for it's PCR amplification. This segment of DNA thus becomes a landmark site identified by PCR analysis for mapping in the human genome project and can be used to screen human genomic libraries.

Another cloning approach to long range cloning as an alternative to chromosome walking, is the combined application of chromosome jumping and linking libraries. In

chromosome jumping library construction a partial digestion with a moderately frequent cutting endonuclease, such as Bam HI (Collins et al., 1987) or complete digestion with a rare cutting endonuclease, such as NotI (Poustka and Lehrach, 1987) generates large fragments. Partial digestion fragments are size selected by preparative PFG electrophoresis. These restriction fragments are ligated to a plasmid selectable marker (e.g. supF, tRNA amber suppressor gene), then completely digested with a frequent cutting endonuclease (e.g. Eco RI) and ligated into a lambda phage with a complementary selectable marker (eg. amber mutated), packaged and amplified. In the construction of linking libraries, a partial digestion with a frequent cutting endonuclease and ligation to a selectable marker is followed by complete digestion with a rare cutting endonuclease prior to the lambda phage ligation (Poustka et al., 1987). Selection of a jumping library clone provides the spliced ligation product from a several hundreds kb sized restriction fragment with only both the distal and proximal ends present. Physical linkage between the two loci at the opposite ends of the large restriction fragment is thus demonstrated in this manner. Furthermore, a jump clone from an infrequent cutting endonuclease chromosome jumping library can be used to select a linking library clone which provides the sequence adjacent to the ligated restriction fragment end that can then be used to select the next jumping library clone. Alternatively, two jumping libraries, prepared with digests from different sets of endonucleases, can be used to select overlapping clones by alternating between libraries (Spoerel and Kafatos, 1987). A general human chromosome jumping library complemented by chromosome walking libraries were used in the Cystic Fibrosis (CF)

positional cloning endeavour (Collins et al., 1987, Rommens et al., 1989). The limitations of library vector insert size are not a factor in chromosome jumping library approach, however there is a danger of cloning artifacts (e.g. a clone that is a chimeric product).

There are other potential problems with genomic libraries. Since adjacent cloned sequences are selected by hybridization, there is a great risk of selecting a clone which is not part of the contig due to cross-hybridization with related sequence elsewhere in the genome. Furthermore, failure to find the adjacent sequence may be due to one region of the genome being under-represented in the library. To minimize these risks, genomic libraries can be constructed from individual chromosomes.

Chromosome specific libraries can be generated from a hybrid cell line which contains a single human chromosome. The resultant library would however, contain a high proportion of genomic clones of rodent origin. This rodent background can be substantially reduced by flow sorting the hybrid cell line chromosomes thereby enriching for the human chromosomes in question. Such libraries have been generated by the Lawrence Livermore National Laboratory (LLNL) in Livermore, California for the Human Genome Mapping (HGM) project (Fuscoe et al., 1987; Van Dilla et al., 1986). Fluorescently stained chromosomes (using a Hoechst dye) are separated, according to chromosome type, by passing through charged deflection plates in successive runs. The degree of deflection for the human chromosome is

sufficient for separation with a 5% to 50% range of efficiency. After partial restriction enzyme digestion and size selection, the fragments are ligated into a linearized cosmid vector, packaged and amplified by standard procedures. The first chromosome specific cosmid library to be generated in this manner for the HGM project was chromosome 19 (de Jong et al., 1989). This HGM library, constructed to provide an eight fold redundancy of chromosome 19, became a valuable resource for identifying the DM locus on chromosome 19 by the positional cloning approach. An average insert size in this type of library is about 40 kb. Larger clones can be obtained by constructing a library using a yeast artificial chromosome (YAC) vector.

Early genomic libraries were limited to bacterial plasmids or phages as cloning vehicles that restricted the size of cloned insert to under 20 kb. The development of cosmids (reviewed by Evans and Wahl, 1987), the cloning of large cosmid contigs, the development of yeast vectors (Burke et al., 1987) and cloning with yeast artificial chromosomes (YACs) permitted isolation of large contiguous segments of DNA. The propagation of large DNA fragments with YAC vectors in yeast is therefore another approach to attain physical linkage. Furthermore, it has been suggested that YAC genomic libraries contain sequences poorly represented in cosmid library banks (Coulson et al., 1988). The YAC technique involves the ligation of large (50-1000 kb) genomic fragments into a vector that provides centromeric, telomeric and selective functions. The constructs are introduced into *Saccharomyces cerevisiae* and replicate in the same manner as the yeast chromosomes (reviewed by Schlessinger, 1990).

YAC clones are selected, arrayed in 96 well microtiter plates and propagated. The linking of genetic markers by YAC clones is accomplished by PCR based screening and selection strategies. Once two markers are spanned by YAC clones, a YAC derived cosmid library can be constructed for further experimentation.

To ensure that YAC clones actually represent physical linkage between two genetic loci, they must be tested for deletions and rearrangements. This can be done by mapping the distance between cloned rare cutting endonuclease sites compared to the genomic result obtained by PFG electrophoresis. A cosmid library derived from the linking YAC's permits construction of a cosmid contig. Mapping of the rare cutting endonuclease sites and identification of unique sequence probes for mapping by PFG electrophoresis can then be accomplished by characterizing the cosmid contig. If no deletions or rearrangements are evident in the physical linkage of two cloned genetic markers, then the genetic and physical distance estimates can be correlated.

Genetic mapping of a disease locus such as DM can be done directly. However, physical mapping of the disease locus can only be indirectly determined by physical mapping of cloned markers showing genetic linkage to DM. To obtain higher resolution genetic and physical mapping of the disease locus the study of single recombination events and the identification of flanking genetic markers are required. The physical mapping of flanking markers that identify single recombinant events physically defines the minimum region expected to contain the disease gene.

3. Genetic Markers and Recombinant Events Flanking the DM Locus.

By combining both genetic and physical linkage data, a disease locus such as DM can be assigned to a specific chromosome interval. The DM locus was indirectly mapped to chromosome 19 by the observed genetic linkage to a number of markers (ABH secretor/Lutheran, Renwick et al., 1971, C3, Eiberg et al., 1983) from which one, C3, was physically mapped to chromosome 19 (Whitehead et al., 1982). Subsequently, close linkage between APOC2 and DM was reported (Shaw et al., 1985). The DM locus was indirectly assigned to the chromosome 19 interval, 19cen-19q13.2, based on the physical mapping of the APOC2 gene (Hulsebos et al., 1985a, 1985b; Lusis et al., 1985). Physical mapping of numerous DM linked markers, based on an extensive somatic cell hybrid mapping panel (Schonk et al., 1989), supported the genetic linkage data assigning the DM locus to the chromosome 19cen-19q13.2 interval.

Further resolution in the genetic mapping of the DM locus within the chromosome 19, 19cen-19q13.2, interval was made possible by extensive linkage analysis (Korneluk et al., 1989b), with seven DM linked markers, of a large number of DM families (53 families comprised of 421 individuals). Analysis of recombinant events in this study identified markers flanking the DM locus. The most likely order of the chromosome 19

markers was determined by multipoint linkage analysis. This, in combination with the somatic cell hybrid panel mapping result, showed the following linkage group orientation:

19cen-p17.1-pJSB11-pα1.4-APOC2-CKM-DM-pEWRB1-pEFD4.2-19qter.

The nearest proximal marker to the DM locus by genetic linkage analysis, was estimated to be substantially closer to the DM locus (CKM; $\theta = 0.01$, $_{Zmax} = 21.76$, Korneluk et al., 1989b) than the nearest distal marker (pEWRB1; $\theta = 0.09$, $_{Zmax} = 15.45$, Korneluk et al., 1989b). The DM locus, therefore, was found to lie in the interval between APOC2/CKM and pEWRB1.

Mapping and identification of flanking markers to the DM locus localized the essential region expected to contain the DM gene to a genetic distance of 10 cM. Further progress in the positional cloning of the DM gene depended on the identification of genetic markers which mapped to this region. These genetic markers could then used to find and map at least two closely flanking recombinant events. The crossover points of these recombination events thus could serve as genetic and physical landmarks defining the proximal and distal boundaries of the essential region expected to contain the DM gene. The recent mapping of the DNA excision repair (ERCC1) gene distal to the CKM gene (Bachinski et al., 1989; Mohrenweiser et al.,

1989; Smeets et al., 1990) thus provided an appropriate locus to search for a genetic marker(s) linked to DM.

Summary of objectives

Identification of the myotonic dystrophy gene would permit study of the biological process of the disease, increase the accuracy of prenatal diagnosis and could fulfil a prerequisite to improved patient care. The major objective of this thesis was to use cloning combined with genetic and physical mapping to identify the minimal region that contained the DM gene. A three part research plan was developed as follows:

1. characterize the ERCC1 region by genetic and physical mapping to determine the location of the DM gene relative to this locus.
2. Initiate a chromosome walk from the ERCC1 region toward the DM gene enabling further study of genetic and physical linkage with the disease locus.
3. Characterize the essential region containing the DM locus for CpG islands and DM associated abnormalities.

There are no known chromosomal aberrations or obvious biochemical abnormalities associated with DM. Consequently, discovery of the gene by a planned molecular genetics approach relies upon positional cloning, that is, isolation of the gene based on its chromosomal location.

The first part of the plan was to screen cosmids known to contain the ERCC1 gene for an RFLP with a high PIC. From about 50 kb of overlapping cosmid clones, an 800 bp unique sequence was found, pE0.8 (D19S115), which identified a useful RFLP. Once identified and subcloned, this marker was then physically mapped relative to the ERCC1 and CKM genes by PFG electrophoresis. Linkage analysis of a Dutch DM family with pE0.8 identified a key recombinant event (Smeets et al., 1990b) that mapped the DM locus relative to the ERCC1 gene. This recombinant event provided a proximal boundary for the essential DM region. Furthermore, the close linkage of pE0.8 to DM confirmed the feasibility of a chromosome walk approach to cloning the DM gene region. The work is described in Chapter III.

Part two of the plan was to initiate cloning of the essential region expected to contain the DM locus by a chromosome walk approach using cosmids. This involved the preparation of cosmid pools and lifts from a chromosome 19 specific cosmid library at the LLNL in California. A novel ALU PCR approach for selection of overlapping clones was applied in the chromosome walk. The newly cloned markers from the chromosome walk generated new physical and genetic mapping information. This work is presented in Chapter IV. A second key recombinant event was identified in a French Canadian DM family (Tsilfidis et al., 1991a) resulting in the definition of the distal boundary of the essential DM region. Combining the chromosome walk of the region proximal to the DM locus (Shutler et al., 1992) with efforts from collaborators involving a chromosome walk of the region distal to the DM locus (Jansen et al., 1992)

and overlap cloning with YAC's resulted in the cloning of the entire DM region (Aslanidis et al., 1992).

Part three of the plan involved the characterization of the DM region cloned from the long range cloning endeavours. This included an RFLP search, physical mapping and the mapping of CpG islands in the essential region containing the DM locus to complement characterization of genes identified in the area. The CpG islands were also tested for DM-associated abnormal methylation. In collaboration with others, two DNA probes I cloned from the essential DM region, PGB2.6 and PGP1.4, identified an RFLP which led to the discovery of the genetic basis for DM (Aslanidis et al., 1992) and the cloning of the DM gene (Mahadevan et al., 1992).

Chapter II: General Materials and Methods

1. Method of Chromosome Walking

A focus of the Human Genome Mapping (HGM) project, at the Lawrence Livermore National Laboratory (LLNL) in California, was to generate chromosome specific genomic libraries from flow sorted human chromosome 19. Two such libraries were made and used in this study to isolate DM region clones. One consists of Lawrist 5 cosmid clones (LL19NC01, de Jong et al., 1989) and the other consists of charon 40 phage (LL19NL01, Yokobata et al., 1992). Unique sequence chromosome walk probes were used to identify DM region phage during the start of the chromosome walk to test for any major rearrangements or to determine if any cloning artifacts were encountered with the cosmids selected to that point. The phage library was again used when no further cosmids could be found at the end of the proximal chromosome walk. Except for the earliest and last chromosome walk steps, the cosmid library was exclusively employed in this study.

The chromosome 19 cosmid library consists of 10,386 clones arrayed in 108 microtitre plates. Cosmid colony filters were prepared on nylon membranes (Hybond, Amersham) by using a 96 prong stamp. The membranes were then placed on top of LB agar media to permit growth of colonies. Standard procedures were then used to prepare these membranes for hybridization as recommended by the manufacturer.

Preparations of DNA were made from each microtitre plate using cultured pools of 96 cosmids. The DNA from each plate was digested with Eco RI, run on a 1% agarose gel and transferred to nylon membranes.

A chromosome walk step, shown in figure 2-1, consists of screening by hybridization with a unique sequence probe to the cosmid pool membranes which identifies a particular microtitre plate. The individual cosmid clones are then identified by further hybridization to the indicated colony membranes. These selected cosmids can then be propagated and characterized for the next chromosome walk step.

2. PCR amplification of cosmid DNA

Unique sequence hybridization probes were generated from cosmids, as shown in figure 2-2, by the polymerase chain reaction (PCR, Saiki et al., 1988) using primers derived from conserved regions of ALU repeat sequence (Karya et al., 1987). The ALU primers used were:

PDJ33 (5'-GCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCA-3');

PDJ34 (5'-TGAGCCGAGATCGCGCCACTGCACTCCAGCCTGGG-3');

and their truncated versions,

ALU33 (5'-CTGGGATTACAGGCGTGAGCCA-3');

ALU34 (5'-CGCCACTGCACTCCAGCCTGGG-3').

Inter ALU products are generated if the repeats are within a sufficient distance and opposite orientation of each other (Fig. 2-2). End fragments were generated from

Figure 2-1: Strategy for screening the chromosome 19 cosmid library. The chromosome 19 cosmid library consists of 10,386 clones arrayed in 108 microtitre plates. Preparations of DNA were made from each plate. These pools of DNA were digested with Eco RI, electrophoretically separated and transferred to nylon membrane. Cosmid colony filters were prepared on nylon membranes by using a 96 prong stamp and placed on LB media to permit growth of colonies. The DNA on these membranes was then fixed for hybridization. A chromosome walk step consisted of screening by hybridization with a unique sequence probe to the cosmid pool membranes to identify specific microtitre plates, for example #6 and #11. The newly identified individual cosmids were selected by further hybridization to the indicated colony membranes, #6 and #11. These selected cosmids, plate #6 column E row #9 and plate #11 column B row 2, were then propagated and characterized for the next step.

SCREENING FOR COSMIDS CORRESPONDING TO PROBES

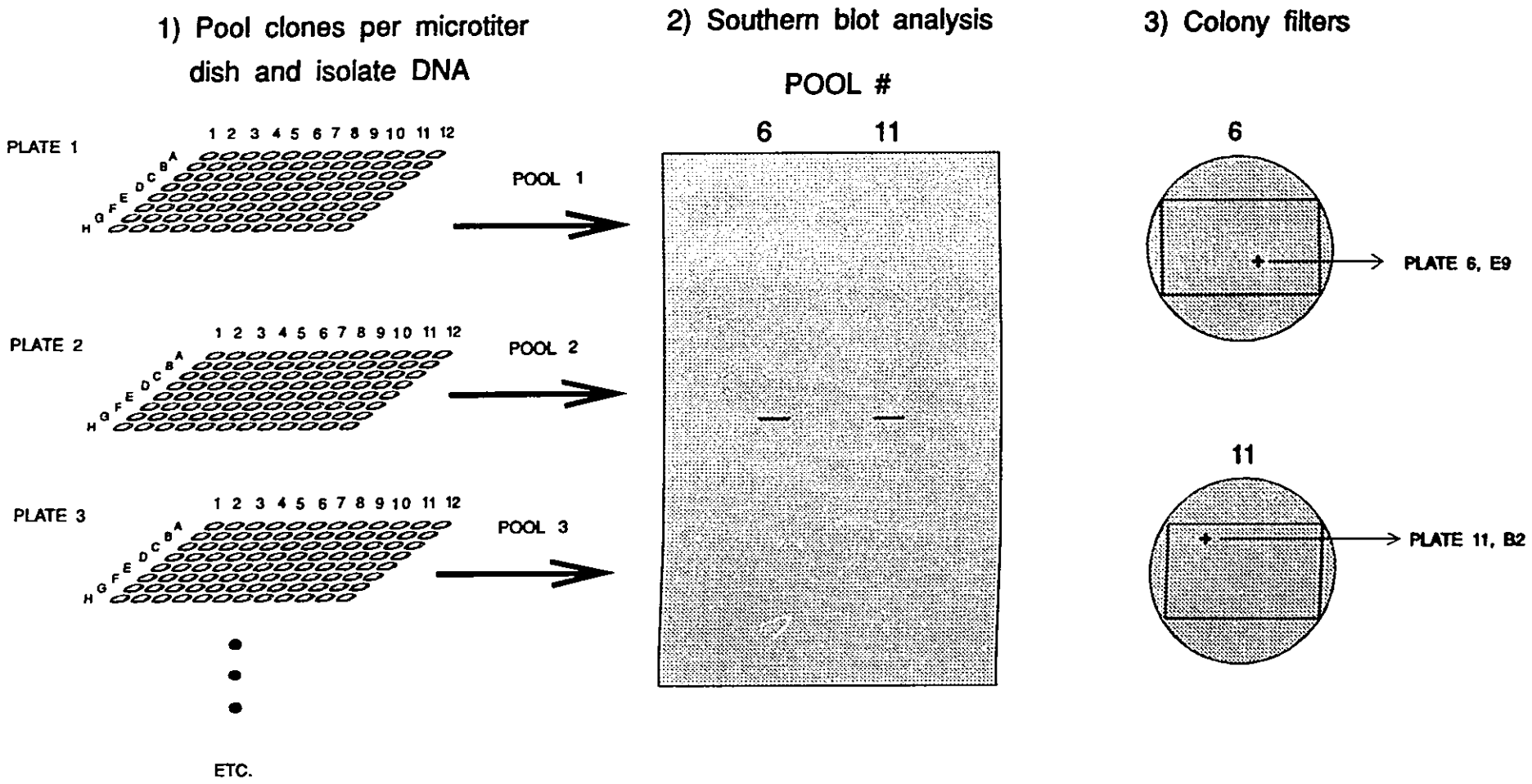


Figure 2-2: PCR amplification of unique sequence probes for chromosome walking using conserved ALU repeat sequence based primers. The production of inter ALU and ALU vector PCR products with ALU33 and ALU34 primers used alone and in combination with the vector primers Sp6 and T7 is shown.

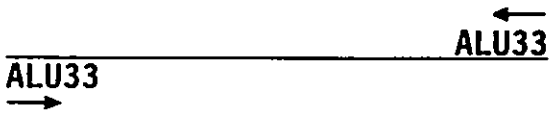
AMPLIFICATION OF UNIQUE SEQUENCE PROBES FOR CHROMOSOME WALKING

HUMAN ALU REPEAT
300 BP

LAWRIST 5 VECTOR - BAMHI INSERTION SITE FLANKED BY T7 AND SP6 PRIMER BINDING SITES.

UNIQUE SEQUENCE PCR PRODUCTS ARE EXPECTED IF:

1) TWO ALU REPEATS ARE SEPARATED BY LESS THAN 1-2 KB
2) THERE IS OPPOSITE ORIENTATION FOR THE TWO ALU REPEATS
3) NO OTHER HOMOLOGOUS REPEAT DNA SEQUENCE IS PRESENT IN BETWEEN



A) 2 ALU REPEATS IN OPPOSITE ORIENTATION: ALU33 ONLY

B) 2 ALU REPEATS IN OPPOSITE ORIENTATION: ALU34 ONLY

C) 1 ALU REPEAT NEAR THE LAWRIST 5 SP6 SITE: ALU33 OR 34 WITH SP6

D) 1 ALU REPEAT NEAR THE LAWRIST 5 T7 SITE: ALU33 OR 34 WITH T7

several cosmids when ALU-PCR was done in conjunction with the vector end primers, Sp6 (5'-GATTTAGGTGACACTATAG-3') andT7 (5'-TAATACGACTCACTATAGGG-3'). As shown in figure 2-2, if an ALU repeat is near the vector insert junction a PCR product will be generated. Cosmid DNA was amplified with *Thermus aquaticus* (Taq) DNA polymerase as recommended by the manufacturer (Cetus Perkin-Elmer) except 5% dimethyl sulfoxide (DMSO) was added to the reaction mixture. A cycle consisted of denaturation at 92°C for 1 min., annealing at 50°C for 1.5 min. and extension at 72°C for 2.5 min. Routinely 30 cycles were performed in a DNA thermal cycler (Cetus Perkin-Elmer). Amplified products were visualized by electrophoresis on ethidium bromide stained 1.0% agarose (BRL Ultrapure) gels.

3. DNA Preparation

Genomic DNA

Genomic DNA for RFLP studies and linkage analysis was extracted from either peripheral blood samples or lymphoblast cultures by using both manual (Madisen et al., 1987) and automated (Genepure, ABI, following manufacturer's instructions) phenol/chloroform procedures or by a salt precipitation method (Miller et al., 1988). The majority of the DNA extractions from patient samples was performed by technologists from the Molecular Genetics Diagnostic Laboratory at CHEO, as a routine service function.

Cosmid and Plasmid DNA Preparation

Cosmid and plasmid DNA were isolated by standard rapid alkaline preparation (Birnboim and Dolly, 1979). Cosmid clones were inoculated into 20 mls of LB broth (1% Bacto-tryptone, 0.5% yeast extract, 1% NaCl) with 20 μg/ml kanamycin and grown for a maximum of 14 h at 37°C with vigorous shaking. After the NaOH-SDS/sodium acetate and alcohol precipitation steps, the pellet was dissolved in 80 μl of sterile water to yield a final standardized concentration of about 0.2-0.3 μg/μl DNA. Cosmid clones identified as prone to deletions were grown for 8-10 h with a reduction in final yield of 50 μl standardized to a 0.2-0.3 μg/μl concentration of DNA. Plasmid clones were cultured in 5 mls of LB broth and 100 μg/ml ampicillin for at least 12 h at 37°C. This resulted in a standardized yield of 100 μl containing about 0.2-0.3 μg/μl plasmid DNA.

Phage DNA Preparation

The chromosome 19 phage library (10^7 plaque forming units, pfu) was first titred and then plated out on 20 NZCYM square plates (10% NZ amine, 5% NaCl, 1% casamino acids, 5% yeast extract, 2% $MgSO_4$, 1.5% Bacto-agar) to about 10,000 pfu/10cmX10cm plate. The host bacteria used was *Escherichia coli* strain NM538 (Stratagene). Unique sequence DNA probes, derived from the cosmid library, were used to identify phage clones from DM region. Secondary and sometimes tertiary screening was required to obtain pure picks. Phage DNA (Charon 40) was isolated by the standard plate lysate method employing polyethylene glycol precipitation and

phenol/chloroform extraction (Maniatis et al., 1982). The low titres of single plaque picks necessitated an amplification step before DNA could be prepared. In phage DNA preparation experiments, agarose was substituted for agar in both the top (0.75%) and bottom (1.5%).

4. DNA Analysis

Digestion of Genomic DNA

The identification of RFLP's involved the digestion of 5μg genomic DNA from a minimum of nine unrelated Caucasian individuals with the a variety of restriction enzymes (including; AlwNI, ApaI, ApaLI, AvaII, Bam HI, BanI, BclI, BglI, BglII, BstEII, DraI, Eco RI, Eco RV, HaeIII, HincII, HindIII, HinfI, KpnI, MspI, NcoI, NheI, NsiI, OxaNI, PstI, PvuII, RsaI, SacI, ScaI, SphI, SspI, StuI, TaqI, XbaI and XmnI).

Digestion of Cloned DNA

During chromosome walk experiments the pooled DNA of 96 cosmids from each of 108 microtitre plates, about 3-5 μg, was digested with Eco RI. The restriction mapping of individual cosmid clones required 2-3μg DNA to be routinely digested with Bam HI, Bam HI/Eco RI, Eco RI, SstI, SstII, XhoI, MluI and NotI. Cosmid subclones were restriction mapped generally using the same restriction endonucleases that have sites in the polylinker of the plasmid being characterized. All digests with restriction enzymes were performed according to the manufactures specifications (Gibco-BRL, Pharmacia, New England Biolabs and Boehringer Mannheim).

Southern Blot Analysis: RFLP's, Cosmid DNA and Cosmid Pool DNA

For RFLP studies, cosmid restriction mapping and cosmid pools, DNA was digested to completion, electrophoretically separated on 0.8% agarose gels (Gibco-BRL) in standard 1X TBE buffer (Sambrook et al., 1989) and Southern transferred to nylon membranes (Amersham Hybond-N). The DNA was fixed to the membrane by uv light using a Bioslink 312 apparatus (BIOS).

Cloned probes or total human DNA used for hybridization were labelled with α-P^{32} dCTP (>3000 Ci/mmol, Dupont) by a random priming procedure (Feinberg and Vogelstein, 1984) from commercially available kits (Amersham and Stratagene) to a specific activity of greater than 10^9 cpm/μg DNA. Blots were hybridized with radiolabelled probes in 6X SSC, 0.05% SDS, 5X Denhardt's solution and 100 μg/ml salmon sperm DNA at 65°C for at least 12 h. The final wash conditions varied depending on the DNA probe used. Generally, the final wash consisted of 0.2X SSC with 0.1% SDS at 55°C for 20 min..

The identification of cosmid end fragments was done by hybridization with Sp6 and T7 primers employed as DNA probes. Oligonucleotides Sp6 and T7 were 5' end labelled by T4 polynucleotide kinase (Pharmacia) and γ-P^{32} ATP (>3000 Ci/mmol, Dupont) using the protocol of Thein and Wallace (1986) to specific activity of greater than 10^8 cpm/μg. Blots for oligonucleotide hybridization were hybridized in 6X SSC,

0.2% SDS, 10X Denhardt's solution, 50 mmol sodium phosphate buffer pH6.8 at 42°C for at least 12 h. Final wash conditions were 6X SSC, 0.1% SDS at 37°C for 20 min.

Washed blots were exposed to film (Kodak X-OMAT AR) at -70°C for 1 to 3 h (for cloned DNA blots) or 1 to 4 days (for RFLP blots). Screen intensifiers were used for these autoradiographs (Dupont Lightning Plus and Cronex Quanta III).

Analysis of Chromosome Walk Cosmid Colony Filters

The 10,386 cosmid clones of the chromosome 19 library were inoculated on nylon membranes placed on top of LB agar media and grown overnight at 37°C. The DNA on these colony filter blots was denatured by placing the membranes, with the colonies facing up, on filter paper soaked in denaturing solution (0.5 M NaOH, 1.5 M NaCl) for 1 minute, neutralized by placing the membranes on filter paper soaked in neutralizing solution (0.25 M tris pH8.0, 1.5 M NaCl) for 5 min. and washed by placing the membrane on filter paper soaked in 2X SSC for 5 minutes. After drying, the DNA was fixed to the membranes by exposure to uv light. Prior to hybridization of individual colony filters, they were first washed in 2X SSC by immersion and rubbing off bacterial debris with gloved fingers.

Cloned probes and ALU-PCR derived probes for chromosome walk steps were labelled with α-^{32}P dCTP by random priming as described for Southern blot analysis. Hybridization conditions for cosmid colony filters were also as described for Southern

blot analysis except a minimum of a 4 h hybridization time was occasionally used. Even after the final wash, the filters were sufficiently labelled that exposure to film for only 1-3 h was required.

5. Subcloning of Cosmids

Whole Cosmid Subcloning

Subclones of Bam HI and Eco RI for entire cosmids were generated for detailed restriction mapping and to search for unique sequence DNA probes from several cosmids. Restriction fragments were cloned by standard protocol as outlined in Sambrook et al., 1989. Briefly, cosmid DNA after digestion was run on 0.6% low melting temperature agarose (Gibco-BRL) gels in the presence of ethidium bromide (Sigma), excised with a scalpel under long wave UV light (366 nm), melted at 65°C, diluted with an equal volume of TE and ligated by T4 DNA ligase (Gibco-BRL) to the appropriate linearized and dephosphorylated plasmid (pUC13 and pUC18, cut with either Bam HI or Eco RI, Pharmacia). Transformation was by heat shock (42°C, for 1 min.) of *Escherichia coli* (DH5αMCR, Gibco-BRL) made competent following the protocol described previously (Hanahan, 1985). Electroporation (Gene Pulser, Bio-Rad) was also employed for other transformations following the manufacturers instructions. Transformants appeared as white colonies on LB agar in the presence of 100 μg/ml ampicillin, and the ß-galactosidase inducer, isopropylthio-ß-galactoside (IPTG, 100 mmol, Gibco-BRL) and the enzyme's histochemical substrate, 5-bromo-4-chloro-3-indolyl-ß-D-galactoside (XGal, 2%, Gibco-BRL).

Forced and Blunt End Cloning of Various Cosmid Restriction Fragments

Subclones of specific single and double digested restriction fragments were generated for use as DNA probes using essentially the same procedure as previously described for whole cosmid subcloning. Either pSp65 (Promega) or Bluescript M13- (pBS, Stratagene) was used as the cloning vector. The plasmid was linearized by two appropriate restriction enzymes in "forced cloning" at polylinker sites, or linearized at one appropriate polylinker restriction enzyme site and dephosphorylated using calf or bacterial alkaline phosphatase (Sigma). In preparation of a plasmid for a blunt end restriction fragment ligation an alternative approach was often employed. The plasmid is first linearization by HincII, leaving blunt ends, then ligated with the target fragment and subsequently digested with SalI to reduce background. Both sites overlap on the polylinker, thus the SalI site is present only on plasmid without inserted target DNA at the HincII site.

Subcloning of PCR products

Subcloning of PCR products was accomplished with the pCR1000 plasmid (Invitrogen) following the manufacturers protocol. The vector is designed to accommodate the nontemplated addition of an extra nucleotide onto 3' ends (Clark, 1988). The addition is often a dATP which makes the PCR product compatible with the 3' dTTP added overhang of the pCR1000 vector. The ligation-transformation protocol was as previously described above except kanamycin was used for selection of transformants instead of ampicillin.

6. DNA probes

Single copy sequences from Southern blots of cosmid and cosmid subclone restriction fragments were identified by their lack of hybridization to repetitive sequences when probed with ^{32}P-dCTP labelled human DNA. After detailed restriction mapping, followed by subcloning, the unique sequence probes were tested for RFLP's and used for physical mapping and linkage analysis. These and the other DNA probes used for physical mapping and linkage analysis are detailed in subsequent chapters.

7. Linkage analysis

A total of 85 Canadian families comprising 904 individuals in total were studied. Approximately half of these families were French Canadian and the majority of the remaining families were either Scottish Canadian or Anglo Canadian. In addition, 14 Dutch families comprising 128 individuals were studied. The clinical diagnosis of DM patients in this study was in accordance with guidelines set forth by an international working group on identification of the molecular defect in myotonic dystrophy (Griggs et al., 1989). The diagnosis of DM was confirmed by clinical examination. A person was considered affected if he or she had transmitted the gene or if the family history of DM transmission was established.

Almost all of the RFLP's identified in this study were eventually incorporated into the routine protocol of the DNA Diagnostic laboratory at the Children's Hospital of

Eastern Ontario (CHEO). The data used for linkage analysis was a composite of information generated by the initial work up of the polymorphism and from data on DM families provided by the diagnostic laboratory.

Lod scores were determined using the LINKAGE package of programs (version 4.6 or 5.1, Lathrop et al., 1985) operated by Dr. Alex MacKenzie at CHEO. Briefly, the pedigree of each family is entered into the program and the phenotype for each DNA marker is entered for each individual. Compound haplotypes for loci with more than one polymorphic site (eg. APOC2 and CKM) were constructed as described by Korneluk et al., (1989b).

8. Pulsed field electrophoresis

Preparation of Human Chromosome DNA in Agarose Blocks

DNA blocks for CHEF electrophoresis were prepared from 20 lymphoblast cultures in an agarose (Clontech) matrix as described by Finney (1989). Briefly, lymphoblasts were grown to approximately 2 x 10^9 cells in tissue culture flasks (T-75, Corning). After treatment with trypsin, the cells were collected and washed in sterile physiological saline with three cycles of spinning and decanting. The lymphoblasts were then resuspended in 0.8 ml physiological saline. Block segments containing about 25 μg DNA (assuming 1 x 10^6 cells yields 10 μg DNA) were made by mixing 100 μl cell suspension with an equal volume of 1.0% molten low melting temperature agarose (Clonetech), pipetted into a mould (Bio-Rad) and cooled on ice. The samples

were lysed by Tris buffered Sarkosyl and proteinase K at 50°C for at least 12h. Prior to analysis, DNA blocks were treated with the proteinase inhibitor, phenylmethylsulfonyl fluoride (PMSF) and washed three times with buffer. DNA from blocks were cut with a scalpel to segments approximately 1/12 the size of the whole block. Each segment is expected to contain about 2 μg DNA.

Digestion of Block DNA

Prior to digestion, the agarose block segments were soaked for a minimum of 3 h at 4°C in the appropriate restriction enzyme buffer. The block DNA was then digested with 10 or 20 units of restriction enzyme (routinely NotI and MluI, Promega or Pharmacia) in a volume of 20 μl according to the manufacturer's specifications. Block segments treated as above but without restriction enzyme were used to demonstrate the integrity of the high molecular-weight DNA.

Electrophoresis Run Conditions

The digested fragments were separated by CHEF electrophoresis in 0.5% or 1.0% agarose (BRL Ultrapure) gels, stained with ethidium bromide, exposed to a uv light source, then transferred to nylon membrane (Hybond, Amersham). Optimum conditions for separating the size of restriction fragments encountered in this study were done essentially by trial and error. By keeping all other parameters constant and varying only the duration of the run, the upper (850 kb) to lower (50 kb) size range appropriate for this study could be resolved. Electrophoresis was in 0.5X TBE at 11.5

V/cm at 14°C for 38 or 42h with switching linearly ramped from 30 to 150 s. Both yeast chromosome (Bio-Rad) and concatamerized lambda phage (Pharmacia) were used as size markers.

Southern Transfer of CHEF Blots

After size separation of the DNA digested to completion, the CHEF gels were stained with ethidium bromide (1 mg/l), photographed under uv light and further exposed to uv light for up to 3 min. causing DNA fragmentation which increases the efficiency of transfer. This was followed by standard denaturation (0.5 M NaOH, 1.5 M NaCl, for 30 min.), neutralization (0.25 M Tris pH8.0, 1.5 M NaCl, for 30 minutes) and transferred to nylon membrane (Hybond, Amersham) for 48 h (500 mls 20XSSC day 1 and again for day 2). The blots were then fixed with uv light and analyzed in the same manner as a standard Southern blot.

9. DNA Sequencing

Two methods were used to determine the sequence of various cosmid subclones.

Manual DNA Sequencing

Sequence data was obtained with vectcr primers, α-^{35}S dATP (< 1000 Ci/mmol, Amersham) and Sequenase (U.S. Biochemical Corp.) in dideoxy chain termination reactions (Sanger et al., 1977) according to a sequencing protocol described previously (Korneluk et al., 1985). Briefly, 2-3 μg of BglI linearized plasmid DNA was

annealed with 1 μl (8 ng) vector sequencing primer (pUC18; M13/pUC 17-mer forward or 16-mer reverse, pBS; M13 20-mer forward or 16-mer reverse) 2 μl Sequenase buffer, to a final volume of 10μl. The primer extension labelling reaction was set up with the addition of 1 μl 0.1 M DTT, 2 μl dGTP labelling mix (diluted 5 fold), 0.5 μl of α-^{35}S DATP and 2 μl Sequenase (diluted eight fold). After 10 minutes at room temperature, 3.5 μl alliqotes were placed in four separate tubes with 2.5 μl of the appropriate A, C, G or T dideoxy termination mix for 5 minutes at 37°C prior to the addition of 4 μl stop mix. These samples were subsequently run on 6% or 8% polyacrylamide gels (Model S2 Sequencing gel electrophoresis apparatus, Gibco-BRL). The gels were then fixed, dried (Gel dryer model 583, Bio-Rad) and exposed to film (Kodak X-OMAT AR) at room temperature for 1-3 days.

Automated DNA Sequencing

Alternatively, DNA sequencing by automated methods were performed by dideoxy chain termination reactions using vector primers, Taq polymerase and fluorescent dye-labelled dideoxynucleotides, according to manufacturers specifications (Applied Biosystems). These samples were subsequently run on an ABI 373A automated sequencer. The sequence data was analyzed by ABI software on a Macintosh PC.

10. Oligonucleotide synthesis

Oligonucleotides for ALU-PCR and STSs were made on an automated DNA synthesizer (PCR-Mate, model 391, Applied Biosystems). The phosphoramidite method of oligonucleotide synthesis was followed according to manufacturers specifications. Briefly, the desired sequence is entered into the microprocessor unit and a cartridge inserted having the first 3'nucleotide linked to solid support for DNA chain growth from the 3' to the 5' direction. The process employs the addition of a phosphoramidite nucleoside monomer, blocked with a dimethoxytrityl (DMT) group, from one of four bottles fixed to the unit. There is one for each nucleotide, which binds to the chain linker at the 5' hydroxyl group in the presence of tetrazole. Chains that fail to undergo addition are then capped by acetylation of the free 5' hydroxyl group with acetic anhydride. Next follows cycles of oxidation by iodine, removal of the DMT group by trichloroacetic acid and addition of the next phosphoramidite nucleoside monomer. At the end of the synthesis, ammonium hydroxide at 55°C for 8-15 h, is used to remove the oligonucleotide from the solid support and remove other protecting groups. The DNA solution is then dried in a vacuum centrifuge (Univapo 150 H, Uniequip), reconstituted in water and quantitated by uv spectrophotometry before use in ALU-PCR and other applications.

Chapter III: Physical and Genetic Mapping of pE0.8 (D19S115), a Chromosome 19 ERCC1 Marker Showing Close Linkage With Myotonic Dystrophy.

1. Introduction

Studies employing both linkage analysis and somatic cell hybrid mapping established the order of chromosome 19 long arm markers to be: 19qcen-pα1.4-APOC2-CKM-DM-D19S50-19qter (Korneluk et al., 1989b) and had identified the gene encoding the muscle specific isoform of creatine kinase (CKM) as the marker most closely linked to DM (Korneluk et al., 1989b; Brunner et al., 1989a,b). Subsequently, the DNA excision repair gene ERCC1 was physically mapped, using somatic cell hybrids and pulsed field electrophoresis techniques, distal to CKM (Bachinski et al., 1989; Smeets et al., 1990).

To map the position of the ERCC1 gene relative to the DM locus, an informative polymorphism at or near the ERCC1 locus was essential. A polymorphic marker, pE0.8 (D19S115), was isolated for this purpose from a cosmid containing ERCC1 genomic sequences. Physical mapping of this marker placed it between ERCC1 and CKM. Results of DM linkage analysis utilizing both Canadian and Dutch DM families are presented in this chapter. Two recombinant events that corroborate the mapping of DM distal to CKM and pE0.8 were identified.

2. Materials and Methods

DNA Probes

The ERCC1 cDNA probe pLB1.14 used for the detection of cosmids containing genomic ERCC1 sequences was a subclone of plasmid pCD4A4 (Cretonne et al., 1984) and is composed of a 650 bp PstI fragment from the 3' region of the ERCC1 gene. The probes used in linkage analysis are detailed in Table 3-1 and were previously described in Korneluk et al., (1989b) with the exception of PRKCG.

DNA Analysis

Genomic DNA was extracted from either peripheral blood samples or lymphoblast cultures as stated in Chapter II, using a solvent extraction method (Madisen et al., 1987) or the salt precipitation method of Miller et al., (1988). Restriction enzyme digests and Southern blot hybridization conditions were as previously detailed in the general materials and methods (Chapter II) and described in MacKenzie et al., (1989).

Pulsed Field Electrophoresis

Lymphoblast cultures were prepared for pulsed field electrophoresis as detailed in Chapter II. DNA blocks from a minimum of three unrelated individuals were digested with 10 units of NotI or 20 units of MluI in a volume of 20 μl according to the manufacturers specifications (Promega and Pharmacia). The digested fragments were separated by CHEF electrophoresis (Bio-Rad) in 0.5% agarose (BRL Ultrapure) gels, fragmented by exposure to a uv light source, then transferred to nylon membrane

Table 3-1: Chromosome 19 Probes Used For Genetic Analysis of DM Families.

Probe (Locus)	Enzyme	Allelic Size (kb)	Allelic Frequency	PIC	Ref.
APOC2 cDNA (APOC2)	TaqI	3.8/3.5	0.52/0.48	0.38	Humphries et al., 1983
	BanI	2.5/1.6	0.76/0.24	0.30	Frossard et al., 1985
	AvaII	0.6/0.4	0.52/0.48	0.38	Korneluk et al., 1987
	NcoI	14.5/11.4	0.52/0.48	0.38	Frossard et al., 1986
3' genomic CKM (CKM)	TaqI	4.5/3.5	0.21/0.79	0.28	Perryman et al., 1988
	NcoI	3.5/2.5 + 1.0	0.44/0.56	0.37	Coewinkel-Driessen et al., 1988
PRKCG (PKC-GC)	MspI	1.6/1.4/ 1.2	0.22/0.32/ 0.46		Harley et al., 1991, Johnson et al., 1988

(Hybond, Amersham). Electrophoresis was in 0.5X TBE at 11.5 V/cm at 14°C for 38 or 42 h with switching linearly ramped from 30 to 150 s.

Linkage Analysis

Sixty-two Canadian DM families comprising 463 individuals were studied. Approximately half the families were French Canadian and the majority of the remaining families were either Scottish Canadian or Anglo Canadian. Details of the linkage analysis approach were as outlined in Chapter II.

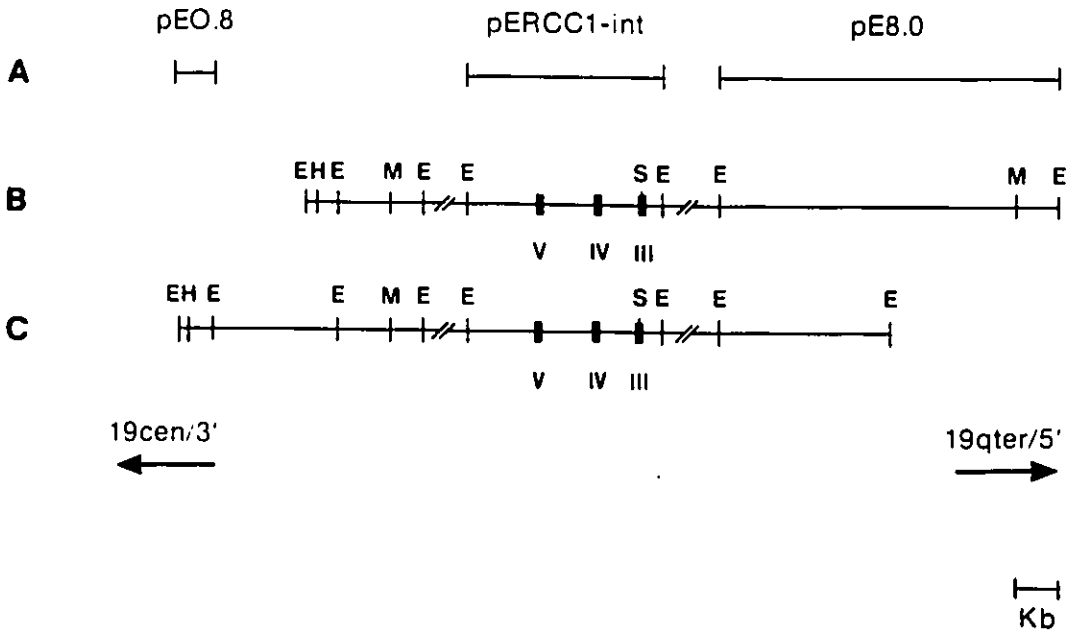
3. Results

ERCCI Cosmid Characterization

Two ERCC1 containing cosmids were selected by hybridization of the chromosome 19 flow sorted cosmid library previously described in Chapter II, with pLB1.14 which contains genomic ERCC1 sequence from the 3' end of the gene. These cosmids were given the Lawrence Livermore National laboratory identification numbers f8697 and f15123.

Overlap between the two ERCC1 cosmids was determined by restriction mapping (Fig.3-1). The mapping of rare restriction endonuclease sites was undertaken to identify them as reference landmarks of physical mapping for the comparison of genomic pulsed field data to restriction mapping from long range cloning.

Figure 3-1: Partial restriction map of the ERCC1-containing cosmids. (A) Eco RI fragments derived from these cosmids that were subcloned into pUC13. (B) Cosmid human DNA insert junctions contained in cosmid f8697. (C) Cosmid human DNA insert junctions contained in cosmid f15123. The shaded boxes and Roman numerals represent exons of the ERCC1 present in the SfiI restriction site-containing subcloned region. Abbreviations: E, Eco RI; H, HindIII; M, MluI; S, SfiI.


pEO.8
pERCC1-int
pE8.0
A
B
C
EH E
M E
E
S E
E
M E
V
IV
III
EH E
E
M E
E
S E
E
E
19cen/3'
19qter/5'
Kb

SfiI digestion of the cosmids yielded a 2.5 kb fragment that has been reported to contain exon two and parts of exons one and three (van Duin et al., 1986). A relatively high density of SfiI sites, for a rare cutting endonuclease, lies in this region as yet another SfiI site was reported nearby, 32 bp from the exon 1 SfiI site (Smeets et al., 1990). The cosmid f8697 was found to contain two MluI sites at opposite ends of the clone, one within 2.0 kb of the 5' terminus and the other located at the 3' end of the cosmid (Fig. 3-1). The cosmid f15123 contained only the latter MluI site. No NotI sites were found in these cosmids.

DNA Probe Isolation

A unique sequence 500 bp PstI fragment was identified from a 4.4 kb Eco RI subclone (cloned into pUC13) of cosmid f15123. This probe, pERCC1-int, was used for physical mapping with CHEF electrophoresis blots. No polymorphisms were identified with this probe. After an exhaustive search, a unique sequence restriction fragment was identified from the cosmids that detected polymorphism. This 800 bp Eco RI fragment, unique to one end of cosmid f15123, was subcloned in pUC13. The probe identified as pE0.8 (D19S115) was used for both genetic analysis and physical mapping.

Physical Mapping

Hybridization of pCKM3', pERCC1-int and pE0.8 to NotI digested genomic DNA revealed a single common 300 kb fragment (Fig. 3-2), demonstrating physical linkage

between these loci. This result is in general agreement with the physical mapping data of Smeets et al., (1990). The pE0.8 and pCKM3' probes hybridize to a common 450 kb MluI fragment (Fig. 3-3), while pERCC1-int identifies a smaller 40 kb fragment (data not shown). This result is consistent with the mapping of pE0.8 proximal to ERCC1. Furthermore. given that pE0.8 was isolated from an area of cosmid f15123 that does not overlap with cosmid f8697 (Fig. 3-1), f15123 appears to extend centromeric to f8697. Cosmid f8697 contains two MluI sites while cosmid f15123 contains only one such site, suggesting that the additional MluI site is derived from sequences extending telomeric from cosmid f15123.

An MluI RFLP was detected in one individual with both pE0.8 and pCKM3' (lane 5, Fig. 3-3). The approximate 40 kb difference in size between the 490 and 450 kb alleles suggests that the polymorphic MluI site is one of the two sites flanking the ERCC1 locus. However, the fact that pERCC1-int does not detect this polymorphism indicates that the polymorphic MluI site is instead at the centromeric end of the 450 kb MluI fragment.

Genetic Linkage Analysis

The probe pE0.8 detects a two allele system ($p=0.62$, $q=0.38$, PIC=0.36) in DNA digested separately with at least three restriction enzymes (Bam HI, 7.0/6.5 kb; BglI, 6.5/6.0 kb; NcoI, 4.5/4.0 kb). The number of enzymes and the constant size differential between the two alleles is characteristic of an insertion polymorphism.

Figure 3-2: CHEF electrophoresis of NotI-digested genomic DNA. Southern blot analysis with pCKM3' (A), pERCC1-int (B), and pE0.8 (C) of NotI-digested DNA isolated from lymphoblasts of three unrelated individuals. Physical linkage of the loci for these three probes is demonstrated by the identification of a common 300 kb NotI fragment.

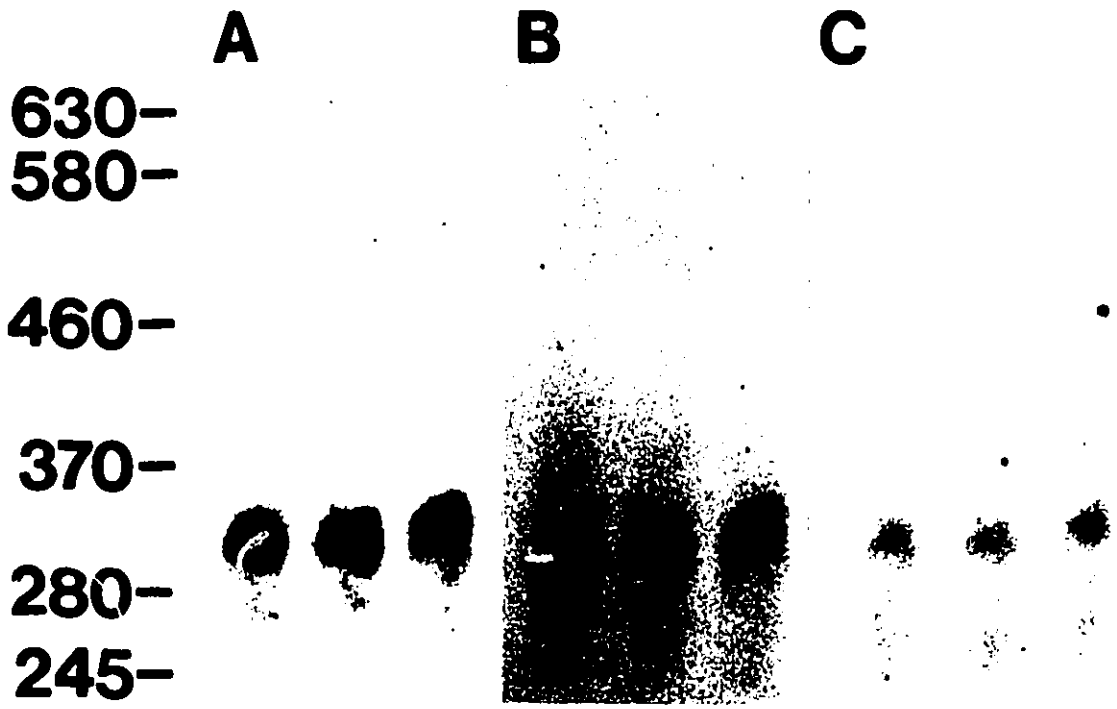
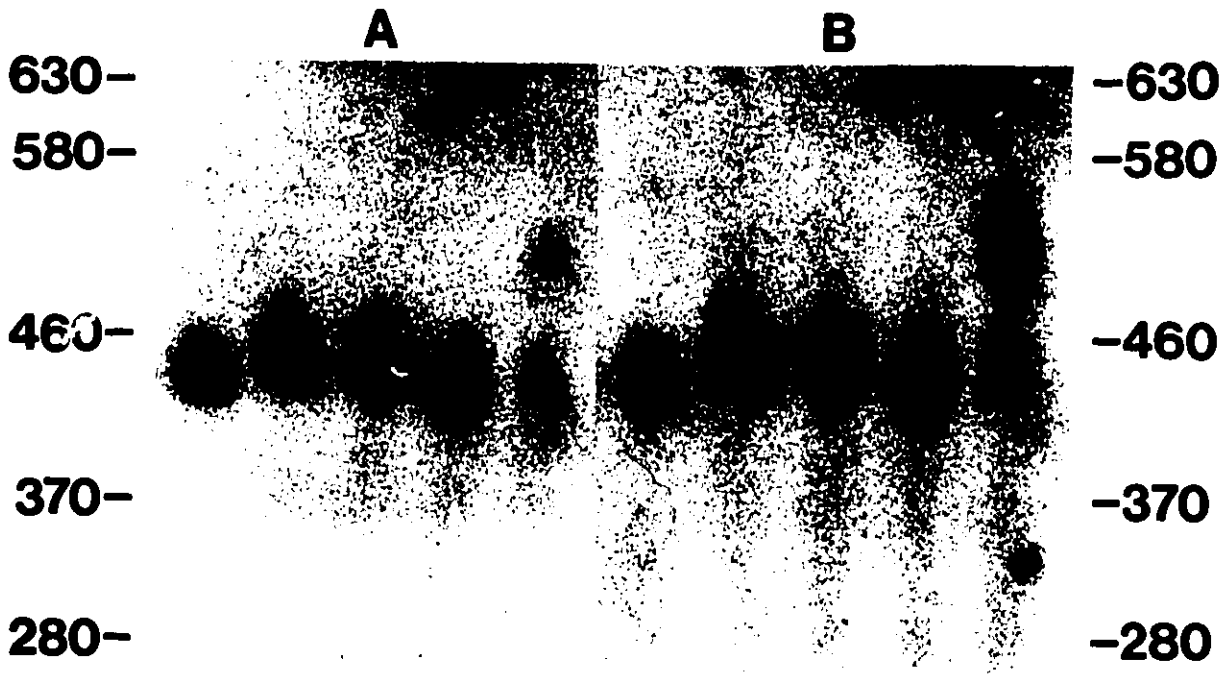

A
B
C
630-
580-
460-
370-
280-
245-

Figure 3-3: CHEF electrophoresis of MluI-digested genomic DNA. Southern blot analysis with pCKM3' (A) and pE0.8 (B) of MluI-digested DNA isolated from lymphoblasts of five unrelated individuals. Physical linkage of the loci for both of these probes is demonstrated by the identification of a common 450 kb MluI fragment.


A
B
630–
580–
460–
370–
280–
–630
–580
–460
–370
–280

The pE0.8 probe is closely linked to APOC2, CKM and DM (θ = 0.01) and shows moderate linkage to D19S50 (θ = 0.10; see Table 3-2). A recombinant event (shown in Fig. 3-4) was detected between APOC2/CKM and both pE0.8/DM and the more telomeric protein kinase C gene (PKCC; Coussens et al., 1986). Another recombinant event was identified in a Dutch DM family wherein pE0.8 cosegregates with the more proximal 19q markers (D19S37, APOC2) and recombines with DM. This recombination suggests a mapping of the disease gene locus distal to pE0.8. This family is fully detailed in Smeets et al., (1991), who document the recombination of an ERCC1 variable simple sequence marker (VSSM) with both DM and more telomeric 19q markers.

The order of chromosome 19 markers bracketing the DM locus has been shown to be 19qcen - APOC2 - CKM - DM - D19S50 - 19qter (Korneluk et al., 1989b). Consequently, the suggested placement of pE0.8 based on this data presented in this chapter is: 19qcen - APOC2 - CKM - pE0.8 (D19S115) - DM - D19S50 - 19qter. Multipoint linkage analysis of the genetic data in our Canadian families is also compatible with this ordering (Fig. 3-5).

4. Discussion

The mapping of ERCC1 distal to and within 300 kb of CKM was in agreement with the analysis of Bachinski et al., (1989), Mohrenweiser et al., (1989) and Smeets et al.,

Table 3-2: Lod Scores Characterizing the Linkage of pE0.8 to ApoC2, CKM, DM and D19S50.

θ	APOC2	CKM	DM	D19S50
0.001	18.27	26.53	15.12	-11.30
0.01	<u>28.46</u>	<u>26.93</u>	<u>19.29</u>	2.69
0.02	28.00	26.56	19.17	6.03
0.03	27.43	26.07	18.90	7.80
0.04	26.81	25.53	18.59	8.89
0.05	26.16	24.94	18.25	9.61
0.06	25.50	24.34	17.87	10.09
0.07	24.82	23.71	17.49	10.39
0.08	24.13	23.07	17.10	10.58
0.09	23.44	22.42	16.69	10.66
0.10	22.80	21.83	16.28	<u>10.68</u>
0.20	15.59	15.02	11.85	8.57
0.30	8.54	8.37	7.18	4.95
0.40	2.72	2.78	2.72	1.58

Note. Z_{max} is underlined; $\theta_m = \theta_f$

Figure 3-4: DM family showing a recombination between the proximal markers APOC2 and CKM and the distal loci ERCC1(as detected by pE0.8), DM, and PKC-GC.

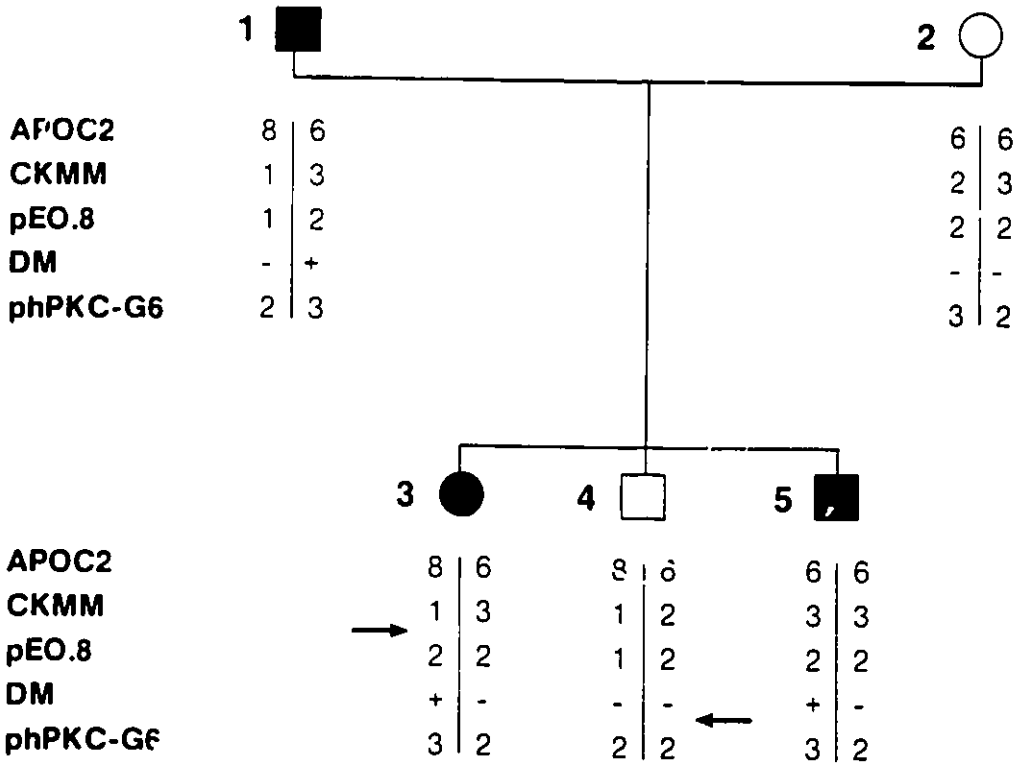
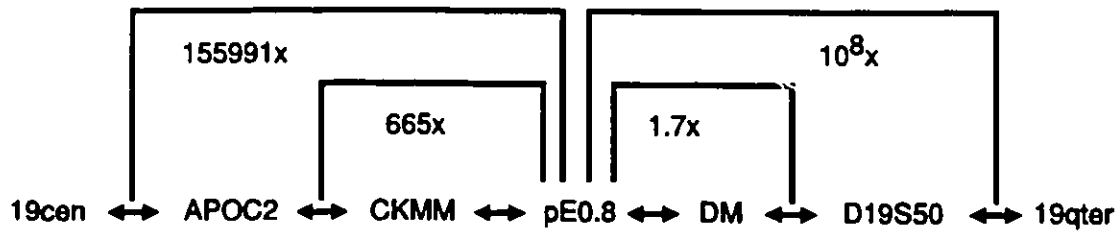

1
2
APOC2
CKMM
pEO.8
DM
phPKC-G6
8 | 6
1 | 3
1 | 2
- | +
2 | 3
6 | 6
2 | 3
2 | 2
- | -
3 | 2
3
4
5
APOC2
CKMM
pEO.8
DM
phPKC-G6
8 | 6
1 | 3
2 | 2
+ | -
3 | 2
8 | 6
1 | 2
1 | 2
- | -
2 | 2
6 | 6
3 | 3
2 | 2
+ | -
3 | 2

Figure 3-5: Multipoint linkage analysis-generated odds against placement of pE0.8 in various map intervals versus the proposed CKM-DM interval for mapping the Canadian DM families. It is 665 times less likely that pE0.8 is proximal to CKM versus the proposed distal mapping. Furthermore, it is 1.7 times less likely that pE0.8 is distal to DM versus the proposed proximal mapping.


155991x
10^8x
665x
1.7x
19cen
APOC2
CKMM
pE0.8
DM
D19S50
19qter

(1990). Despite the close genetic linkage between pE0.8 and DM, no significant linkage disequilibrium was detected between pE0.8 and DM in our French Canadian population [standardized disequilibrium constant (Hill and Robertson, 1968), ϕ = 0.15]. This result was somewhat unexpected, as there has been significant linkage disequilibrium observed with more weakly linked DM markers in the same population (MacKenzie et al., 1989; Korneluk et al., 1989a, b).

There are at least three possible explanations as to why linkage disequilibrium would be less despite being closer to the DM locus. Ongoing analysis of several DM-linked markers has revealed two major French Canadian DM haplotypes, one of which may have evolved from the other. One pE0.8 allele is on one major DM haplotype and the other pE0.8 allele is present on the second major DM haplotype. This has led to significant attenuation of the pE0.8/DM linkage disequilibrium. Other markers such as CKM have the same genotype on both DM haplotypes with no resultant diminution of linkage disequilibrium. A second possibility is that the pE0.8 insertion polymorphism is relatively unstable, resulting in a decay of pE0.8/DM linkage disequilibrium. A third factor underlies the limits of linkage disequilibrium. Genotypes that are rare in the normal population generate linkage disequilibrium which is stronger than more common alleles (genotypes). Consequently, a relatively distant locus (with a rare genotype) may demonstrate significant linkage disequilibrium with a disease gene in contrast to a locus mapping nearer to the disease gene (with a common genotype) that manifests relatively weak linkage disequilibrium.

The RFLP identified by pE0.8 in this study and the VSSM polymorphism described in Smeets et al., (1991) represented the closest proximal markers to DM reported at the time. The pE0.8 marker has proved to be of great value in the prenatal and preclinical diagnosis of DM in the CHEO diagnostic laboratory and elsewhere. In addition to close linkage between DM and the ERCC1 region markers, both this report and that of Smeets et al., (1991) mapped the disease locus distal to ERCC1. Consequently, a chromosome walk was initiated from pE0.8 in a telomeric direction toward the DM locus to map and eventually identify the genetic mutation responsible for this disease.

Chapter IV: Physical Mapping and Cloning of the Myotonic Dystrophy Gene Region.

1. Introduction

The positional cloning approach to cloning a disease gene such as DM requires a number of essential resources: 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. Once a minimal genetic and physical region has been defined, isolation and detailed characterization of the corresponding genomic sequences can be undertaken.

Two recombinant events were identified that map the DM gene to a region of about 2 cM or less on 19q13.3 (Shutler et al., 1991; Tsilfidis et al., 1991a). The order of chromosome 19 long arm markers around DM was shown to be: 19qcen - D19S37 - APOC2 - CKM - D19S115 -ERCC1 - DM - D19S51 - D19S50 - 19qter (Korneluk et al., 1989, Bachinski et al., 1989, Smeets et al., 1990a, Johnson et al., 1990 and Shutler et al., 1991a). As presented in Chapter III, the DNA probe pE0.8 (D19S115) was identified as a polymorphic marker. This marker was found to be closely linked to DM and shown to be corresponding to a genomic sequence that lies 15 kb centromeric from ERCC1 (the excision repair cross complementing 1 gene; Shutler et al., 1991a). A recombinant event identified in a Dutch DM family showed pE0.8 and an ERCC1 variable simple sequence marker (VSSM) to co-segregate with

more proximal 19q markers (D19S37, APOC2) and recombine with DM (Shutler et al., 1991a and Smeets et al., 1990b). This data clearly showed DM to be distal to ERCC1. The results of another key recombinant event in a French Canadian DM family, described by Tsilfidis et al., (1990a), mapped DM proximal to p134c (D19S51). The DNA probe p134c is an anonymous marker that was cloned from a somatic cell hybrid containing a portion of the long arm of chromosome 19 (Johnson et al., 1990).

Based on the data showing DM to be distal to markers at the ERCC1 locus and the tight linkage of D19S115 to the DM locus it was reasoned that pE0.8 would be an appropriate start point for a chromosome walk from pE0.8 towards the DM gene in a telomeric direction. In this chapter the cloning of a cosmid contig extending approximately 350 kb from pE0.8 towards 19qter is described. Single copy probes derived from the contig were used for genetic and physical mapping in order to define a minimal area that is to contain the DM gene. Furthermore, another 250 kb from the distal DM region was cloned by collaborative studies to complete the cloning of the DM region.

2. Materials and Methods

Chromosome 19 Cosmid Library

The myotonic dystrophy chromosome walk cosmids in this study were isolated from a previously described flow sorted chromosome 19 cosmid library (de Jong et al., 1989). The library consists of 10,368 single colony clones arrayed in the wells of

108 microtitre plates. Preparation of the cosmid library for this study was illustrated in figure 2-1 of Chapter II. Individual cosmids from each microtitre plate were cultured on nylon membranes in duplicate and the DNA fixed for hybridization by standard procedures. DNA was prepared from a pool of cosmids for each plate. The pools of DNA were digested with Eco RI (Gibco-BRL), size separated by agarose (Ultrapure, Gibco-BRL) gel electrophoresis and transferred to nylon membrane (Hybond, Amersham). A duplicate set of membranes was also generated. One set of cosmid pool blots consisted of three membranes with two lanes of size markers (lambda DNA digested with HindIII, Gibco-BRL) flanking 30 lanes of cosmid pool DNA for each membrane. The fourth membrane to complete the set, contained DNA from the remaining 18 pools.

Method of Chromosome Walking

A chromosome walk step consisted of hybridization of a unique sequence DNA probe to a set of cosmid pool Eco RI blots. This procedure was detailed in Chapter II. The detection of an Eco RI fragment from a particular lane of pooled DNA indicated which of the 108 microtitre plate colony filters to examine. Hybridization of the indicated colony filters with the same DNA probe then identified the individual cosmids containing unique sequence in common with the screening probe. The cosmids identified in this manner include previously identified cosmids and new cosmids representing physically linked adjacent sequence. These new cosmids were propagated and characterized for the next chromosome walk step.

DNA Probes for Chromosome Walking

Two methods of selecting DNA probes for chromosome walk steps were employed in this study. The conventional approach involved restriction mapping the cosmids and searching for unique sequence restriction fragments that map nearest to the cosmid end oriented toward the direction of the chromosome walk. Alternatively, rapid progress in chromosome walking was made when unique sequence DNA probes were identified using the polymerase chain reaction. A pair of two short (22mer) primers derived from the ALU consensus sequence separately and in combination with Sp6 and T7 vector end primers, were used to amplify unique sequence DNA from the walk cosmids as detailed in Chapter II. These PCR products were run on an ethidium bromide stained agarose gel to reveal the overlap between cosmids by fingerprinting the location of ALU repeat sequences. A non-overlapping band was sought from the most telomeric cosmid or the ALU-vector end primed PCR product for the telomeric end. These PCR product bands were used, as unique sequence probes for chromosome walking, simply by isolating the desired band after low melting temperature agarose gel electrophoresis. DNA from the gel slice was labelled by the multiprime protocol described in Chapter II.

DNA Analysis

Unique sequence DNA probes generated by both conventional and ALU-PCR means were evaluated for genetic and physical mapping purposes. The probes were screened for RFLP's as described in Chapter II. Genetic analysis of the Dutch

recombinant family detailed by Smeets et al., (1990) was performed with these newly generated markers. DNA probes employed in physical mapping were hybridized to CHEF pulsed field electrophoresis blots prepared as detailed in Chapter II. Several of these unique sequence probes from the chromosome walk were subcloned.

Sequence Tagged Sites and YAC Library Screening

A sequence tagged site (STS) consists of a short, unique DNA sequence that can be recognized by PCR assay. An STS serves as a basic landmark on the physical map of the human genome (Olson et al., 1989) and thus it's map position must be well defined. Markers generated from chromosome walk clones are amenable to STS conversion.

Three unique sequence DNA probes, pKE0.6, pKBE0.8 and an ALU34 PCR product from phage clone cw13ϕ5 were sequenced by manual and automated DNA sequencing using the protocol described in Chapter II. A primer set for pKE0.6 was devised consisting of primer #76 (5'-GCAGGGATTCAACTGCCCGCCG-3') and primer #78 (5'-TTCCTCTTCCAGCAATGGCTCT-3'). A primer set for pKBE0.8 was devised consisting of primer #73 (5'-GCCCACAACCGGAGTATAGTC-3') and primer #74 (5'-CCCGGACTGGATAATGCGGT-3'). Both these sets of primers work with the same amplification parameters. A cycle consisted of denaturation at 94°C for 1 min., reannealing at 55°C for 1.5 min. and extension 72°C for 1 min. Routinely 30 cycles were performed. Primers for PKE0.6 amplified a 580 bp product while the primers for

PKBE0.8 amplified a 750 bp product. A primer set for the ALU34 PCR product was also devised consisting of primer #269 (5'-TACCTCAGAAGGTACTTTGGAA-3') and primer #361 (5'-GAGCACCCCAGCTGAGTCTGA -3'). These primers amplified a 280 bp product. A cycle consisted of denaturation at 94°C for 1 min., annealing at 55°C for 1.5 min. and extension at 72°C for 1.5 min. A total of 30 cycles were performed. The pKE0.6 and pKBE0.8 primers along with primers from the VSSM at the pX75b locus were used in the selection of YAC clones from the DM region (Aslanidis et al., 1992).

The "St. Louis" YAC library, Burke et al., (1987), was screened with the pKE0.6 and pKBE0.8 STSs along with the STS at the pX75b locus (Jansen et al., 1992a) for DM region clones by Drs. Charalampos Aslanidis and Chris Amemiya at the Human Genome Centre of the LLNL (Livermore, California) using PCR amplification. The YAC library consists of individual clones arrayed in 96 well microtitre plates. Pooled DNAs from each microtitre plate were screened by STS PCR amplification. The YAC clones containing sequence homologous to these STSs were then identified from these pools by analyzing individual clones in the selected microtiter plate by STS PCR amplification. ALU PCR was then performed on the YAC clones to obtain probes for screening the chromosome 19 cosmid library. In addition, a cosmid library was also made from the YAC clones. This library was also screened with the ALU PCR generated probes. Cosmids selected from these libraries were used to connect the proximal and distal DM region cosmid contigs.

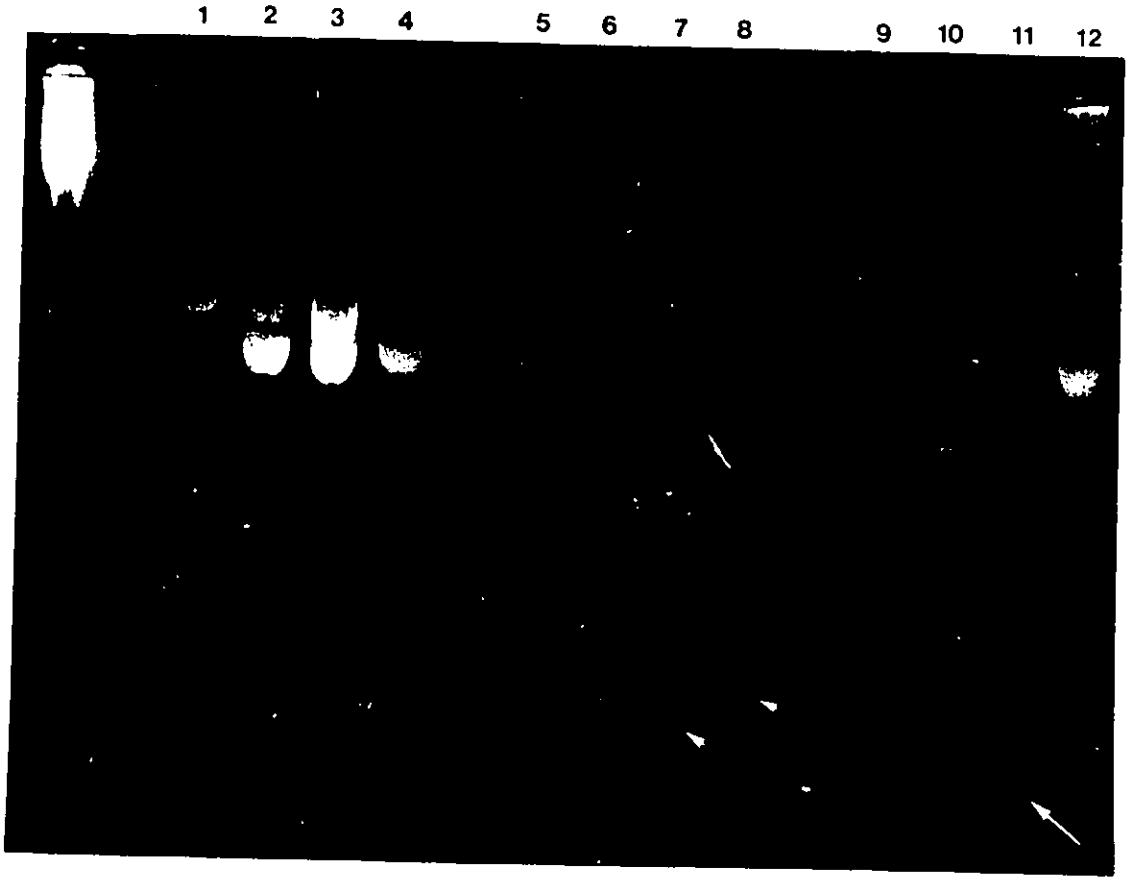
3. Results

The experimental approach in this study was designed to clone the DM region of human chromosome 19. Application of ALU-PCR to generate unique sequence DNA probes to select chromosome walk cosmids was used for this purpose.

Chromosome Walking

An example of ALU-PCR analysis for a chromosome walk step is provided in figure 4-1. Typically, amplification was performed on four or more previously selected cosmids from the last two sequential chromosome walk steps with the ALU33 and ALU34 primers alone and in combination with the vector end primers (Fig. 4-1, results from ALU34 lanes 1-4, ALU34 with Sp6 lanes 5-8 and ALU34 with T7 primer lanes 9-12 are shown). The cosmids appear in figure 4-1 ordered in a contig from the most centromeric to the most telomeric (cosmid 19; lanes 1, 5 and 9, cosmid 20; lanes 2, 6 and 10, cosmid 21; lanes 3, 7 and 11, cosmid 22; lanes 4, 8 and 12). Common bands between the cosmids are due to overlapping ALU repeat sequences. A unique band or a band appearing only in combination with a vector end primer from the most telomeric cosmid and from the telomeric end of the cosmid would be selected as a probe for the next chromosome walk step. A PCR product of ALU primer combined with a vector primer was preferable to an inter-ALU product since it could potentially be more efficient for chromosome walking than just the ALU PCR derived products. This is because the extending distance is at a maximum for cosmid end fragments. In the example provided, the ALU34 Sp6 product from cosmid 22 was selected to screen

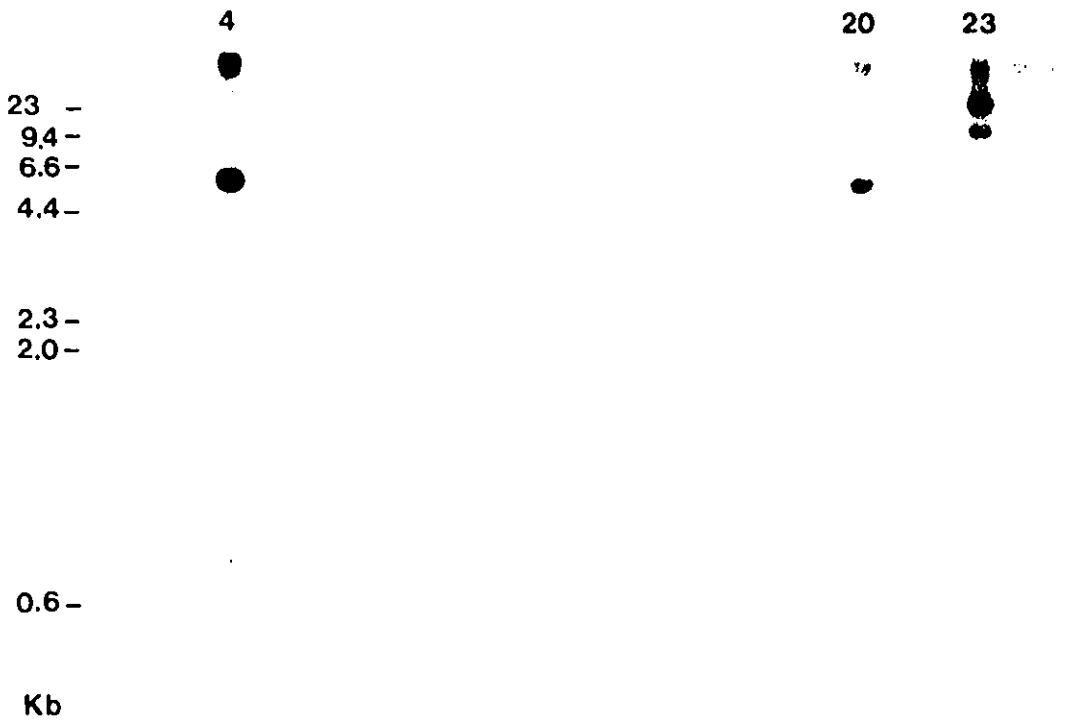
Figure 4-1: ALU PCR amplification of overlapping cosmids by ALU34 both with and without vector end primers. Amplification by PCR was performed on four overlapping cosmids previously selected from the cosmid chromosome 19 library. The ALU PCR products are visualized under uv after agarose gel electrophoresis with ethidium stain. The products are arrayed according to the originating cosmid from a centromeric to telomeric orientation (lanes 1, 5 and 9, cosmid 19; lanes 2, 6 and 10, cosmid 20; lanes 3, 7 and 11, cosmid 21; lanes 4, 8 and 12, cosmid 22) and according to the primer set used (lanes 1-4, ALU34; lanes 5-8, ALU34 and Sp6; lanes 9-12, ALU34 and T7). The size marker, first lane on left side of figure, is a 123 bp ladder. A unique band that appears only in combination with a vector end primer (indicated by arrows in lanes 7, 8 and 11) represents an insert end fragment. The arrow in lane 7 identifies a fragment from cosmid 22 that was used as a DNA probe (following the protocol previously described in Chapter II) to extend the contig by identifying adjacent cosmids from the chromosome 19 cosmid library.


1
2
3
4
5
6
7
8
9
10
11
12

the cosmid library due to it's telomeric orientation and low non-specific hybridization background.

The probes selected for the next chromosome walk step are hybridized initially to the cosmid pool membranes. One of the ALU PCR vector end probes is expected to hybridize only to pools with cosmids identified from the previous chromosome walking step while the opposite end probe identifies additional cosmid pool lanes indicating the presence of adjacent chromosome walk cosmids in these pools. Figure 4-2 shows the hybridization result for two cosmid pool blots with a unique ALU33-Sp6 primer PCR band from cosmid 14 used as a DNA probe. The identification of four cosmid pool lanes corresponding to plates 115, 166, 182 and 185 from these Eco RI digested cosmid pool DNA blots is evident (Figure 4-2 lower, lane 9; Figure 4-2 upper, lanes 4, 20 and 23 respectively). The colony filters indicated by these cosmid pool results were then hybridized with the ALU33-Sp6 cosmid 14 probe and the individual cosmids identified as illustrated in figure 4-3. The cosmid from microtitre plate 166 row H column 3 was the originating cosmid 14 clone (Lawrence Livermore identification number f20927). The cosmid from microtitre plate 182 row B column 9 (f22397) was also identified during the previous chromosome walking step, cosmid 11 (f22397), along with four others (Table 4-1). The cosmids from plates 185 and 115 (Figure 4-3 row C column 4 and row B column 6 or F15962 and F22692, respectively) represent new clones in the contig, cosmids 15 and 16 (Table 4-1). These cosmids were

Figure 4-2: Selection of cosmids from the chromosome 19 cosmid library by chromosome walking; Screening cosmid pools. The ALU33/Sp6 PCR product from cosmid 14 used as a hybridization probe to screen the cosmid pool blots prepared from the chromosome 19 cosmid library. This results in the identification of clones from four lanes, three on one blot (panel A upper - lanes 4, 20 and 23 corresponding to plates 166, 182 and 185, respectively) and one on a second blot (panel A lower - corresponding to plate 115). Two of the signals were from previously identified overlapping cosmids, plate 166 coordinants H3 or cosmid 14 (f20927) and plate 182 coordinants F4 or cosmid 13 (f22440). The other signals were from previously untested pools likely containing overlapping cosmids with an extension of adjacent sequence.


4
20
23
23
9.4
6.6
4.4
2.3
2.0
0.6
Kb

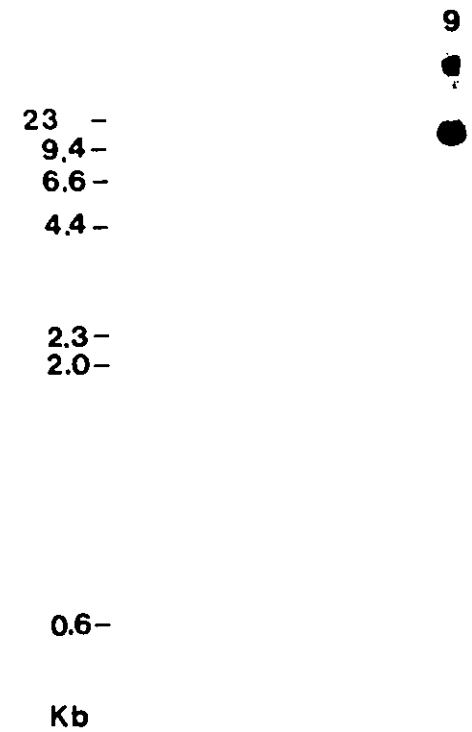
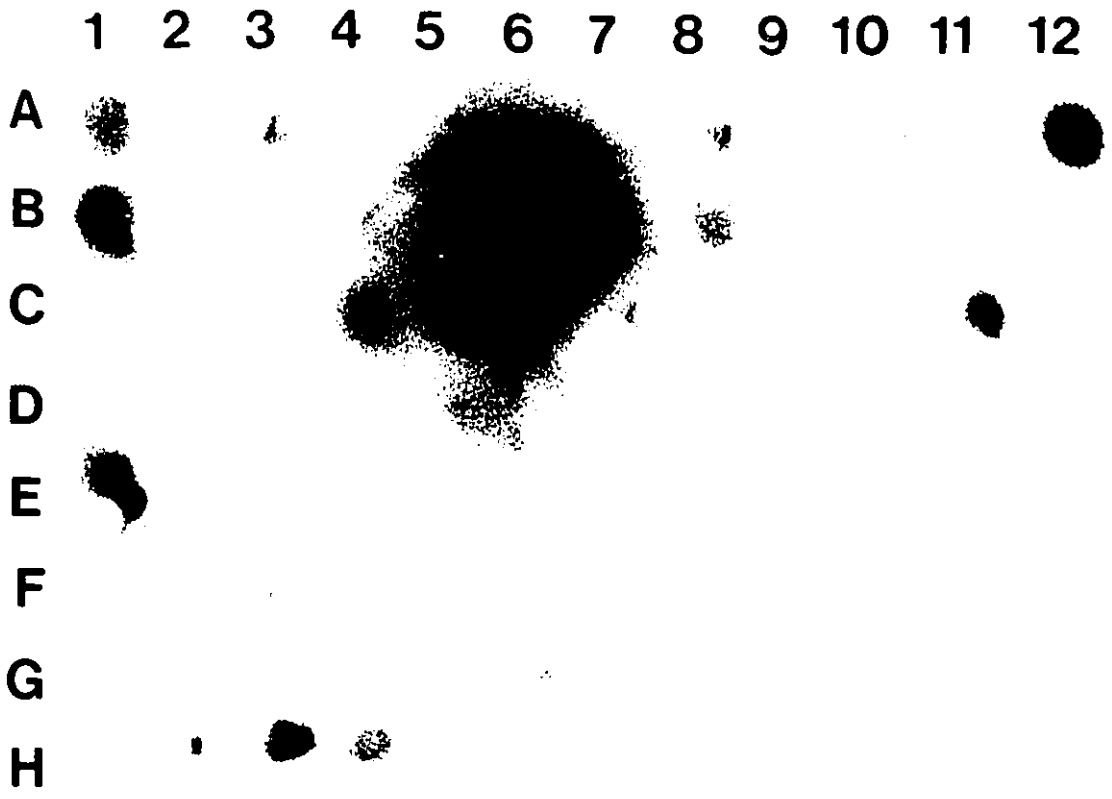

9
23
9.4
6.6
4.4
2.3
2.0
0.6
Kb

Figure 4-3: Selection of cosmids from the chromosome 19 cosmid library by chromosome walking; Identifying cosmid colonies. Individual cosmids were identified from the screened pools shown in figure 4-2 by hybridization of colony filters for both plates 115 and 185 with the ALU33/Sp6 cosmid 14 probe. Signals were obtained from coordinants B6 of plate 115 (panel B upper) and C4 of plate 185 (panel B lower). These cosmids (115/B6 or cosmid 15, given LLNL id# f15962 and 185/C4 or cosmid 16, f22692) were retrieved from their microtiter plates and propagated for further characterization.


1 2 3 4 5 6 7 8 9 10 11 12
A
B
C
D
E
F
G
H

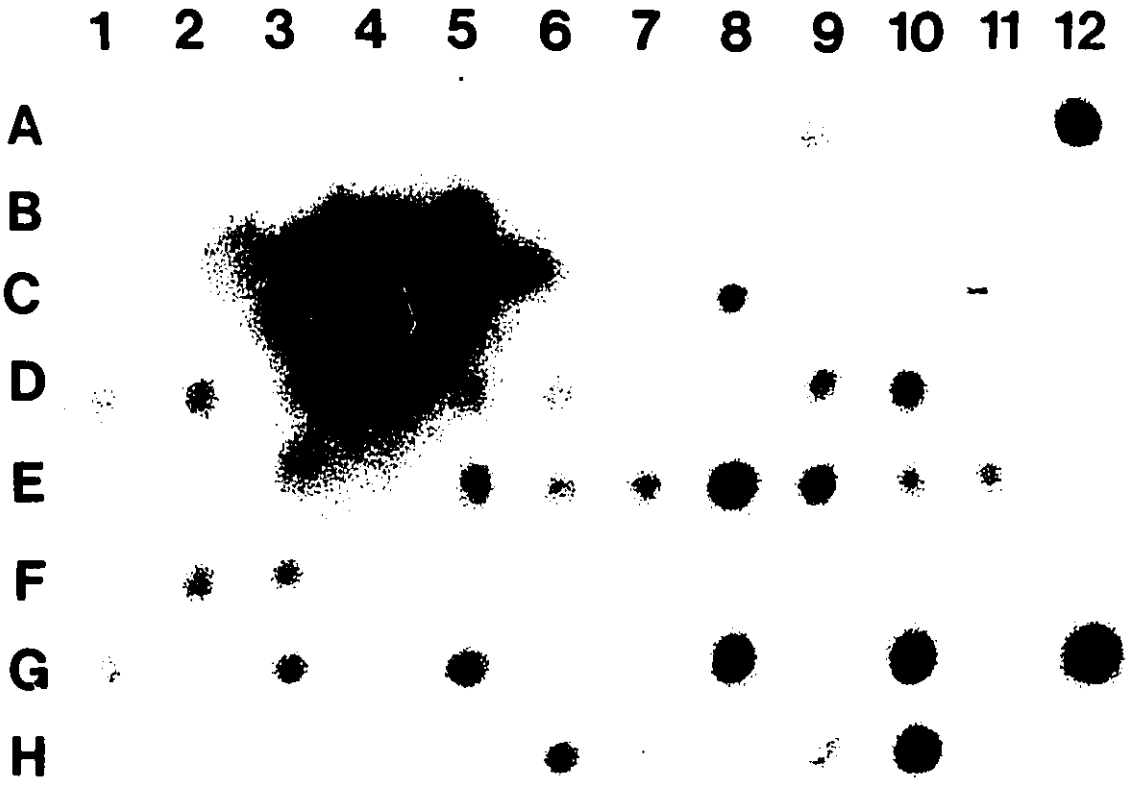

1 2 3 4 5 6 7 8 9 10 11 12
A
B
C
D
E
F
G
H

Table 4-1: Cosmids identified by a chromosome walk through the proximal segment of the myotonic dystrophy gene region.

Cosmid	LLNL ID.#	Selected by	Rare cuter site
1	f15123	pE0.8	M^1, M^2
2	f8697	pE0.8	M^2
3	f8191	pN2.1	M^2, N^1
4	f10080	pN2.1	N^1
5	f5118	pKE2.1	
6	f13544	pKE2.1	
7	f23770	pKE0.6	M^3, N^2
8	f18905	pKE0.6	M^3, N^2
9	f20720	ALU33-cosmid 8	M^3, N^2
10	f24206	ALU33-cosmid 8	M^3, N^2
11	f22397	ALU33-cosmid 8	M^3, N^2
12	f18226	ALU33-cosmid 8	M^3, N^2
13	f22440	ALU33-cosmid 8	M^3, N^2
14	f20927	ALU33-cosmid 8	M^3, N^2
15	f15962	ALU33/Sp6-cosmid 14	M^3, M^4, N^2
16	f22692	ALU33/Sp6-cosmid 14	M^4
17	f24860	ALU33/Sp6-cosmid 16	
18	f20007	ALU33/Sp6-cosmid 16	
19	f20829	ALU33/Sp6-cosmid 16	M^5
20	f16149	pKBE0.8	N^3
21	f22266	pKBE0.8	

Cosmid	LLNL ID.#	Selected by	Rare cuter site
22	f25267	ALU34/Sp6-cosmid 19	M^5, N^3
23	f20000	ALU34/Sp6-cosmid 22	N^4, N^5
24	f15243	ALU34/Sp6-cosmid 22	M^6, N^4, N^5
25	f16236	ALU34/Sp6-cosmid 24	M^6, N^4, N^5
phage	cw13ϕ5	ALU34-cosmid 25	

The LLNL ID # indicates the identification numbers of the clones in the arrayed cosmid library, as assigned at the Human Genome Centre, Lawrence Livermore National Laboratory as part of the chromosome 19 contig mapping effort.

retrieved from the appropriate microtitre plates, propagated, characterized and the procedure repeated for subsequent steps.

The alignment of cosmids and the corresponding location of the probes used in the chromosome walk are shown in figure 4-4. Nine of the fourteen probes used in the chromosome walk were ALU PCR derived. In addition, six of the nine ALU PCR derived probes were surprisingly, end fragments. This demonstrated the usefulness of the entire ALU PCR approach, particularly when combined with vector end primers. Furthermore, comparison of the inter-ALU PCR products generated with PDJ33 and PDJ34 from different cosmids was used by Dr. Peter de Jong at the Human Genome Centre of the LLNL (Livermore, California) to confirm putative overlaps. The longer primers compared to the truncated versions revealed more inter-ALU products that were preferable for fingerprinting and assigning cosmid sequence overlap. In contrast, truncated versions provided fewer and less complicated patterns that were preferable for probe selection. Figure 4-5 shows the ALU34 set of results for the cosmid contig. Many patterns appear identical which suggests a clustering of ALU repeats with the appropriate orientation in the region of overlap for those particular cosmids.

As a supplement to the cosmid library, a flow sorted chromosome 19 phage library (LL19NL01, Yokobata et al., 1992) was plated out as described in Chapter II and used in the beginning and at the end of the chromosome walk. A unique

Figure 4-4: Schematic representation of the proximal DM region cosmid contig detailing the position of rare cutter restriction sites and mapping of unique sequence DNA probes used for the chromosome walk. The positions of the cosmids derived from the walk are shown (top) in reference to the six MluI (marked and numbered M^1 - M^6) and five NotI (marked and numbered N^1 - N^5) restriction sites contained within the cosmid contig. The relative location of the probes used to identify the contig are indicated (bottom). The probes marked with a triangle were used in the selection of contiguous cosmids from the library (see Table 4-1). Most are ALU PCR products which were used as hybridization probes. For example, the ALU33/Sp6 PCR product from cosmid 16 identified cosmids 17, 18 and 19. Other probes shown were generated from the walk and used for genetic mapping studies and/or physical mapping (see Table 4-2).

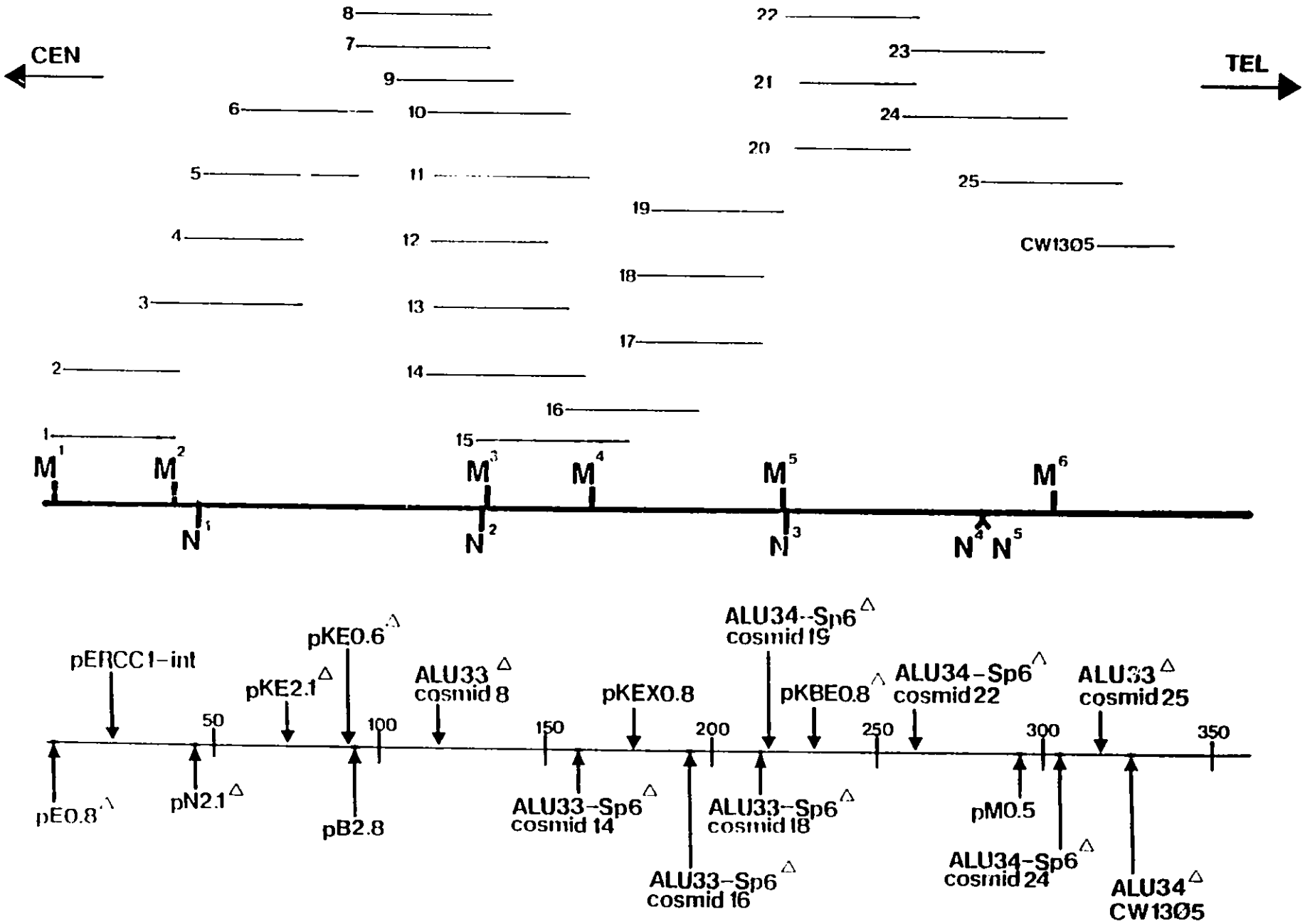
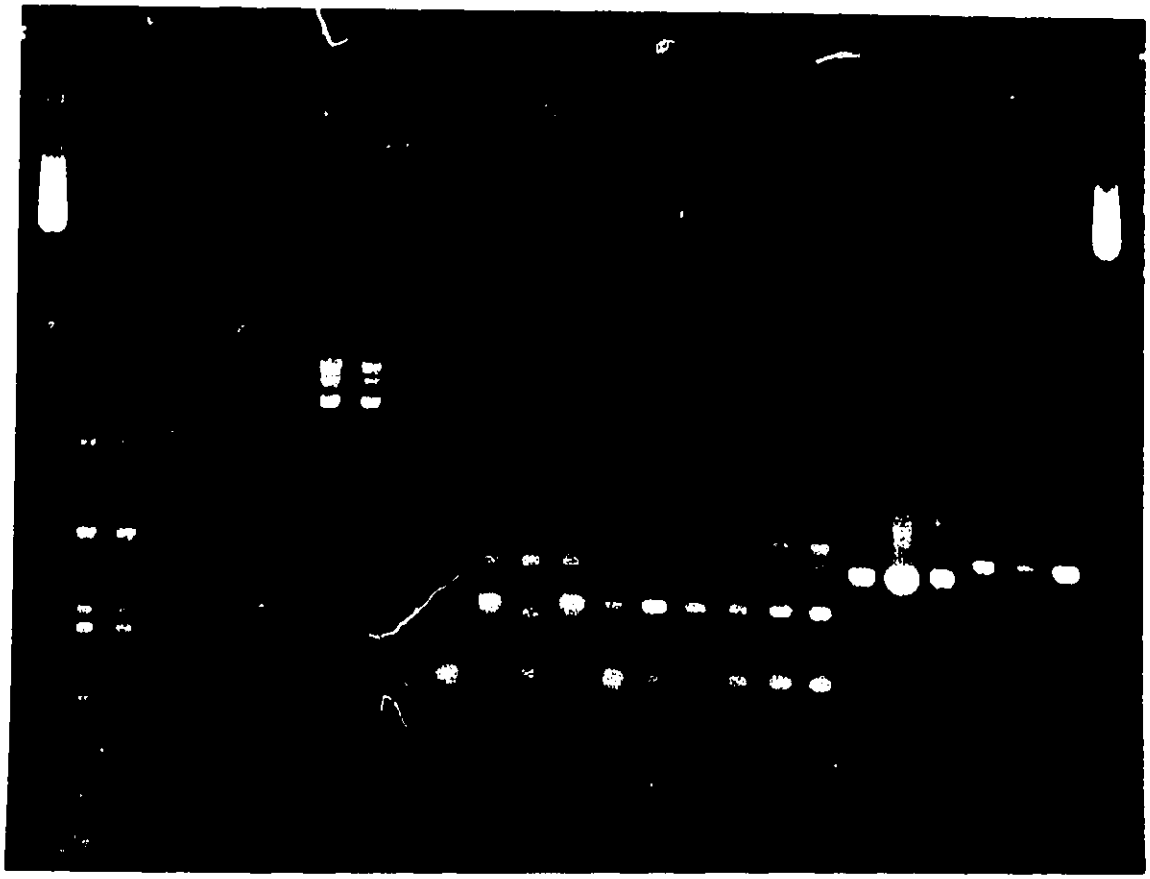

CEN
TEL
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
CW1305
M1
M2
M3
M4
M5
M6
N1
N2
N3
N4
N5
pERCC1-int
pKE0.6
pKE2.1
ALU33 cosmid 8
pKEX0.8
ALU34-Sp6 cosmid 19
pKBE0.8
ALU34-Sp6 cosmid 22
ALU33 cosmid 25
50
100
150
200
250
300
350
pE0.8
pN2.1
pB2.8
ALU33-Sp6 cosmid 14
ALU33-Sp6 cosmid 16
ALU33-Sp6 cosmid 18
pM0.5
ALU34-Sp6 cosmid 24
ALU34 CW1305

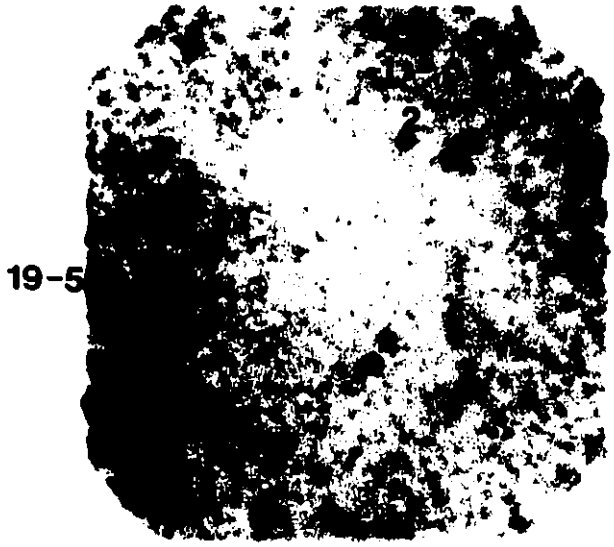
Figure 4-5: ALU34 PCR products from the proximal DM region chromosome walk cosmids. ALU-PCR reactions with the ALU34 primer from the 25 cosmid contig are shown after electrophoresis on an ethidium bromide stained 1.0% agarose gel. The 123 bp ladder was used for the two marker lanes. Overlap of cosmid clones can be seen from similar band patterns that are generated by this approach. The ALU34 primer was one of the truncated ALU primers used for the generation of inter-ALU products from cosmids. These products were used as unique sequence DNA probes in the DM region chromosome walk.



sequence probe from an 8 kb Eco RI subclone (pE8.0) of cosmid 2 (f3697) and pN2.1 containing the N^1 NotI site (cosmids 3, 4, and 5, Figure 4-4) and a 900 bp unique sequence Bam HI fragment subcloned from cosmid 3, pB900, were used for selecting phage clones in the stages of the walk. Screening of the chromosome 19 phage library during a chromosome walk step involved hybridization with a unique sequence probe and selecting at least six signals previously undetected. A typical result is shown in figure 4-6 using pN2.1 which identified numerous plaques on every phage library plate (unmarked signals were previously detected with the more centromeric pE8.0 probe). Six signals were picked from two plates. After a secondary screen, 12 single plaques were picked (Figure 4-7) and DNA prepared. The phage DNA was digested with SstI (to release the insert from the phage arms), size separated by electrophoresis with an ethidium bromide stained agarose gel, photographed under uv (Figure 4-8) and transferred to nylon membrane. Hybridization with pN2.1 revealed that six of the 12 picks had sequence homologous to pN2.1 (Figure 4-9, lanes 5-9, and 11). The unique sequence DNA probe pB900 (derived from cosmid 3 and not present in cosmids 1 or 2) identified four of the six phage picks that were extending more in the telomeric direction (Figure 4-10, lanes 5, 6, 9 and 11).

No obvious sequence discrepancies were found between the cosmids and phage in mapping experiments with other DNA probes found in both the cosmid and phage chromosome 19 libraries. The phage library was again used at the end of the

Figure 4-6: A chromosome walk step in the chromosome 19 phage library with the DNA probe pN2.1; Primary screen. The screening of the chromosome 19 phage library (LL19NL01, Yokobata et al., 1992) was done as a supplement to the chromosome 19 cosmid library (LL19NC01, de Jong et al., 1989) at the beginning and the end of the proximal DM region chromosome walk to test for any obvious sequence discrepancies. Shown here is the screening of two plates consisting of approximately 10,000 pfu/plate from the chromosome 19 phage library with pN2.1, a unique sequence DNA probe subcloned from a previously isolated phage clone. Several signals are evident on both plates. Six plugs, corresponding to the new signals, were picked (using a glass pipet end) from these plates during the initial screen (numbered 1-6, panel A). To prepare for a secondary screen with pN2.1 the phage from each plug was eluted, tittered and plated, following the protocol outlined in Chapter II.


19-5
2

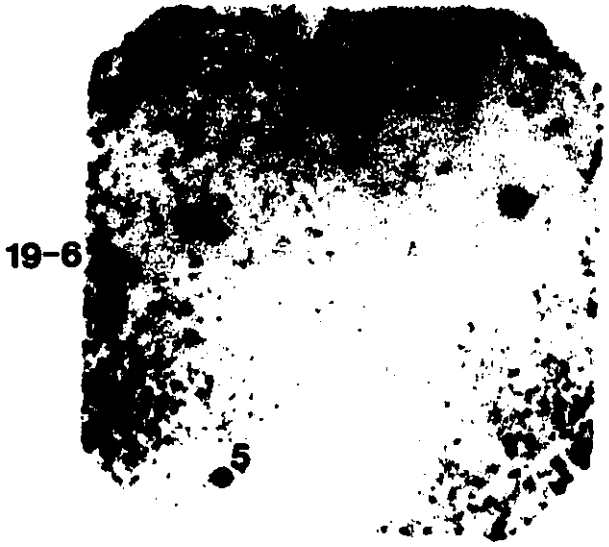

19-6
5

Figure 4-7: A chromosome walk step in the chromosome 19 phage library with the unique sequence DNA probe pN2.1; Secondary screen. The primary screen of the chromosome 19 phage library with the unique sequence DNA probe pN2.1 resulted in six newly identified signals in this chromosome walk step. These phage were titered and plated at about 200 pfu/plate (marked 1B-6B). The secondary screen with pN2.1 collaborated the primary screening result and provided 12 pure picks (single isolates numbered 1-12). DNA from these isolates was prepared for characterization to confirm and determine the overlap with previously isolated phage and cosmid clones.

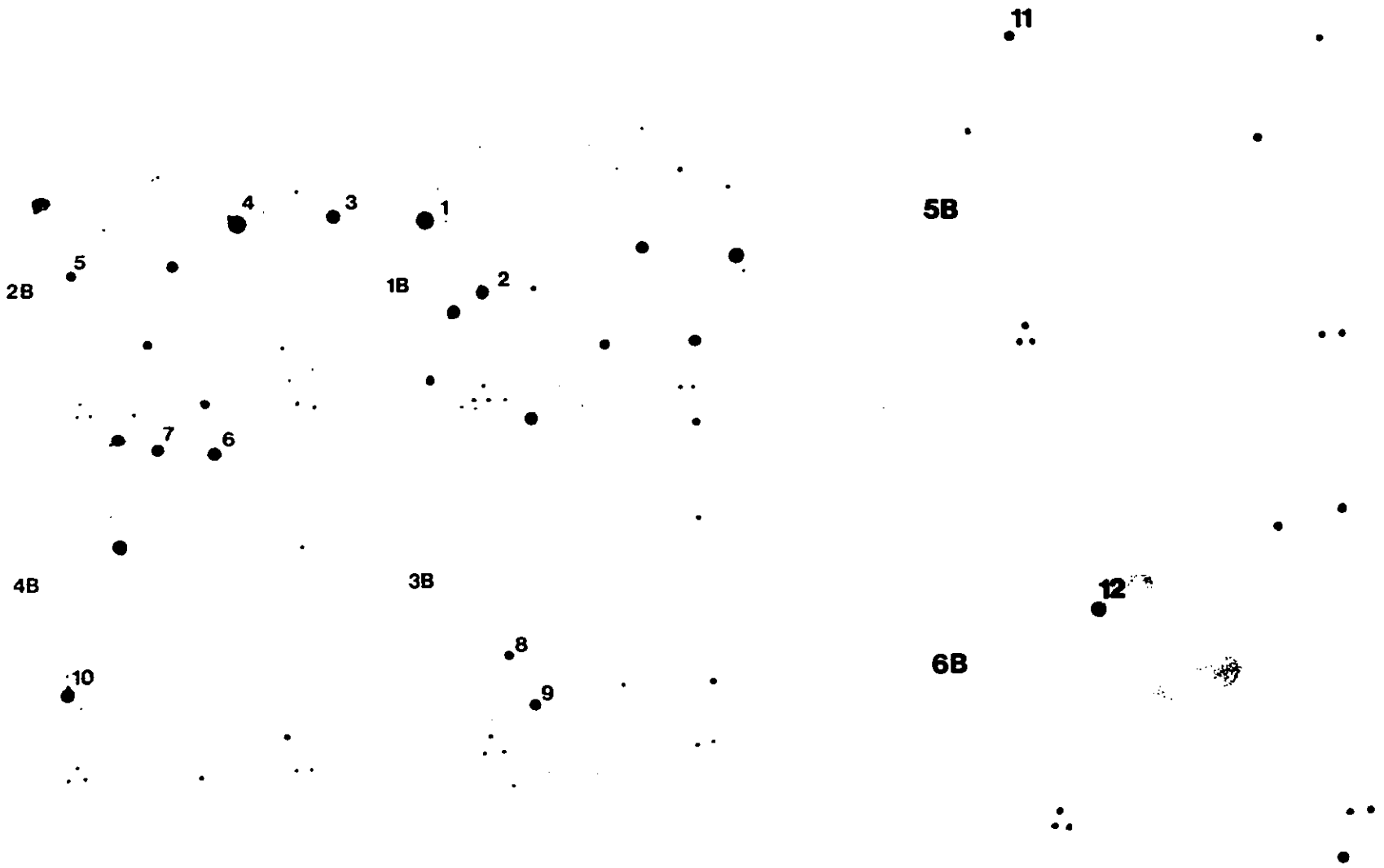

1
2
3
4
5
6
7
8
9
10
11
12
1B
2B
3B
4B
5B
6B

Figure 4-8: Characterization of pN2.1 selected chromosome 19 phage isolates; Electrophoresis of SstI digested phage DNA. Phage DNA was prepared from the 12 isolates selected by the secondary pN2.1 screen of the six initial chromosome 19 phage library primary picks. After digestion with SstI to release the vector arms the phage DNA was run on an ethidium bromide stained 1.0% agarose gel. The first marker lane is kb ladder, the second marker lane is lambda DNA digested by HindIII and the outside marker lane is 123 bp ladder. DNA from a previously selected overlapping phage isolate is in lane 1 while lanes 2-13 are DNAs from the 12 pN2.1 selected phage isolates. At least one common band was found in lanes 5, 7, 8, 9, and 11 (about 2 kb).

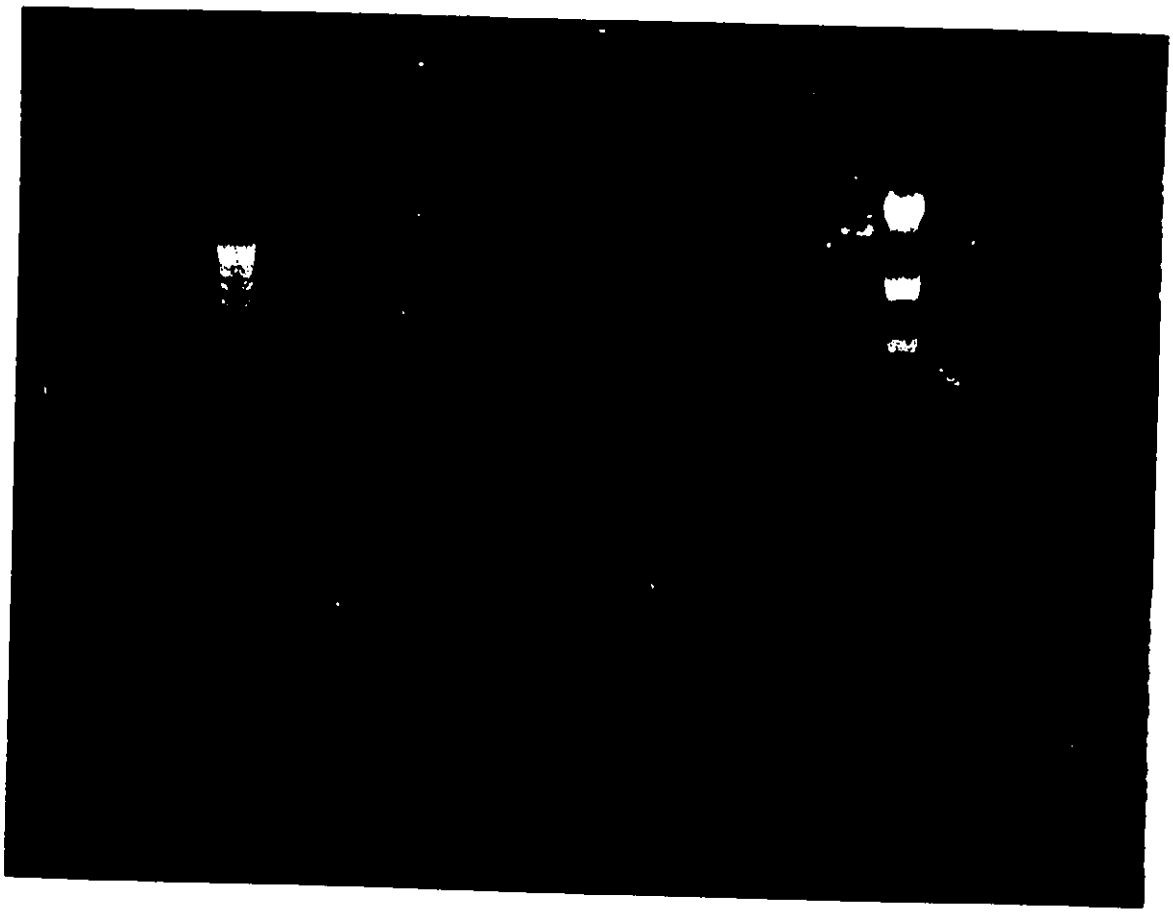


Figure 4-9: Characterization of pN2.1 selected chromosome 19 phage isolates; Hybridization with pN2.1. The gel containing the SstI digested DNA from the 12 pN2.1 selected phage shown in figure 4-8 was transferred to nylon and probed with pN2.1. A common band of about 2 kb was identified by pN2.1 hybridization for six lanes. Thus one half of the 12 picks show sequence homology to pN2.1 (lanes 5-9 and 11). This probe was also shown to be present in the chromosome 19 cosmid library (cosmids 3, 4 and 5, Figure 4-5). The phage picks that did not show homology to pN2.1 in this experiment may have initially been selected due to cross hybridization with pN2.1.

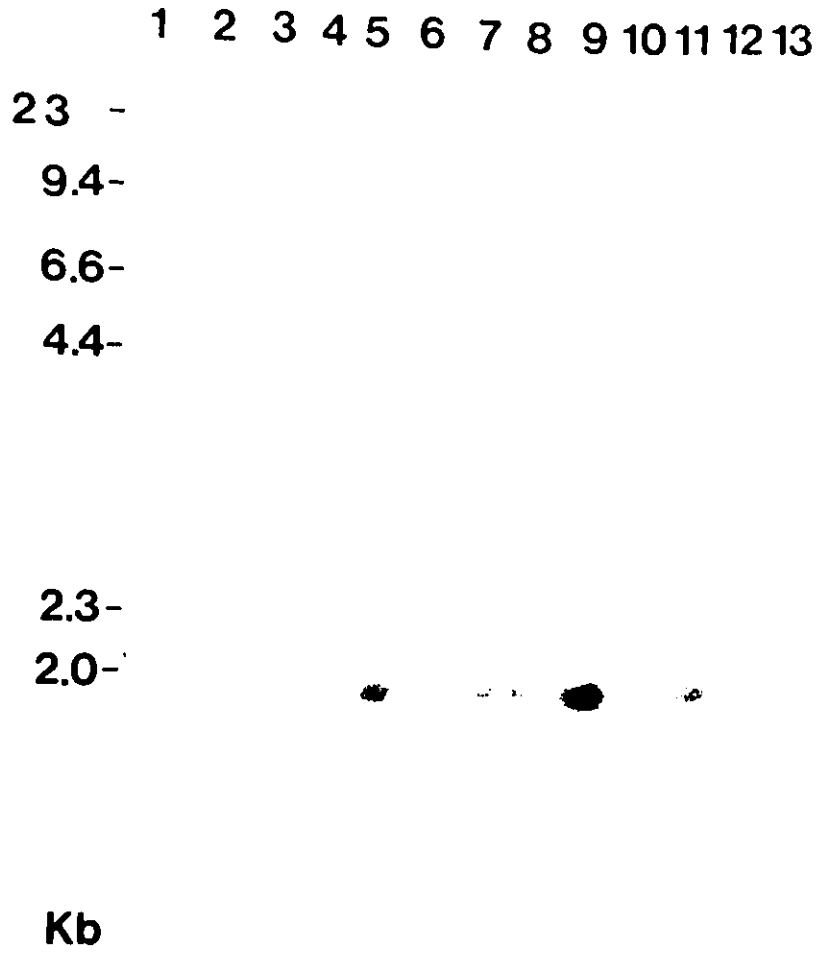
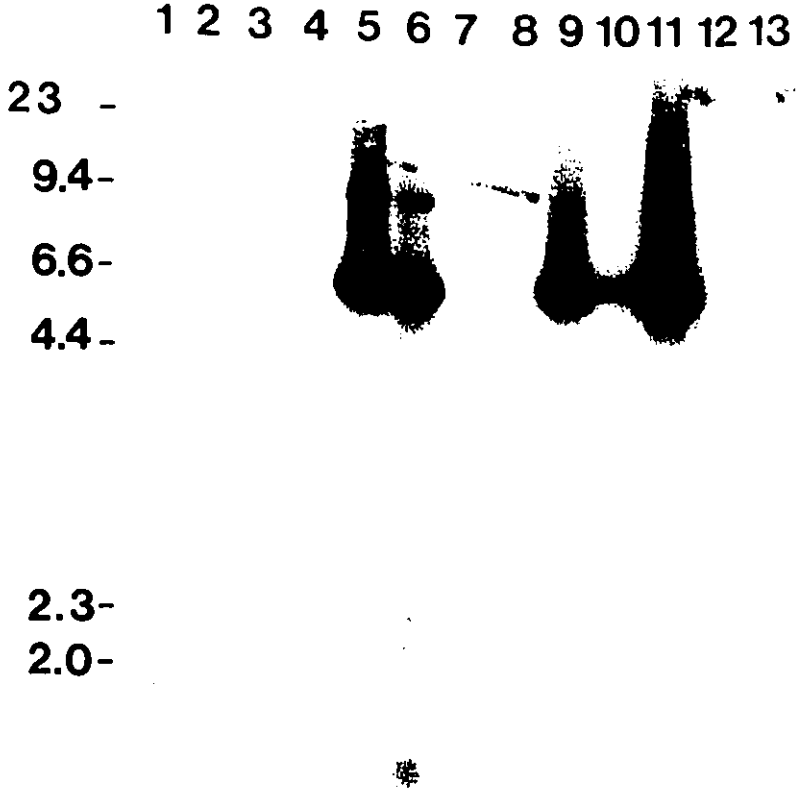

1 2 3 4 5 6 7 8 9 10 11 12 13
23
9.4
6.6
4.4
2.3
2.0
Kb

Figure 4-10: Characterization of pN2.1 selected chromosome 19 phage isolates; Hybridization with pB900. Hybridization of the blot containing SstI digested DNA from the pN2.1 selected chromosome 19 library phage with a unique sequence DNA probe, pB900 is shown. The pB900 probe was subcloned from cosmid 3. Four of the six pN2.1 homologous phage were identified by pB900 (lanes 5, 6, 9 and 11). The pN2.1 and pB900 unique sequence DNA probes are thus present in both cosmid and phage clones of this region.


1 2 3 4 5 6 7 8 9 10 11 12 13
23 -
9.4-
6.6-
4.4-
2.3-
2.0-
Kb

chromosome walk when no more cosmids could be found to extend the contig in a telomeric direction. After one phage clone was found to extend the contig, CW13ϕ5, no others were found. Thus the proximal segment of the myotonic dystrophy gene region is represented in 11 chromosome walk steps which includes 10 with cosmids and 1 with phage (Figure 4-4).

Characterization of Cosmid Clones

A total of 28 cosmids were isolated by chromosome walking. Deletions and/or rearrangements were found in propagating three of the cosmids. These unstable clones are not included in the contig as they proved to be redundant. Overlap between cosmids was evaluated by several methods including restriction mapping, ALU PCR fingerprinting and hybridization between overlapping unique sequence restriction fragments.

The integrity of the cosmid contig was confirmed by the mapping of rare endonuclease sites within the cosmids (Figure 4-4) and the subsequent correlation with CHEF data. The cosmids and the rare endonuclease sites are listed in Table 1. A total of six MluI (Figure 4-4, marked and numbered M^{1-6}) and five NotI (Figure 4-4, marked and numbered N^{1-5}) restriction sites served as landmark reference points for the chromosome walk. The presence of two MluI sites (Figure 4-4, marked M^1 and M^2) in cosmid 1 (Table 4-1) was previously reported (Shutler et al., 1991a). An additional two MluI sites (Figure 4-4, marked M^3 and M^4) were found in cosmid 15

(Table 4-1). Only one of either the proximal or distal MluI sites was present in the other overlapping cosmids. A further two MluI sites (Figure 4-4, marked M^5 and M^6) were found, one detectable in cosmids 19 and 22 and the other in cosmids 25 and 25 (Table 4-1). Cosmids 3 and 4 (Table 4-1) were found to contain a NotI site (Figure 4-4, marked N^1). A second NotI site (Figure 4-4, marked N^2) was identified in cosmids 7-15 inclusive (Table 4-1). Another NotI site (Figure 4-4, marked N^3) was found in cosmids 20 and 22 (Table 4-1). Two adjacent NotI sites (Figure 4-4, marked N^4 and N^5), within a 11 kb interval, were identified in cosmids 23-25 inclusive (Table 4-1).

Isolation and Subcloning of DNA Probes

A total of nine unique sequence DNA probes were subcloned from cosmids. These along with four of the ALU-PCR products (Table 4-2) were used for CHEF pulsed field electrophoresis studies. One ALU34 PCR product from cw13ϕ5 was subcloned into the pCR1000 plasmid (Invitrogen) and subsequently characterized as a sequence tagged site (STS). Furthermore, five of these probes were found to detect RFLP's and were applied to subsequent linkage analysis studies (Korneluk et al., 1991; Shutler et al., 1991a, b, c; Tsilfidis et al., 1991b, c). The DNA probe, pKE2.1 (D19S116) was subcloned from a 2.1 kb Eco RI fragment from cosmid 3 (f8191) into pUC13 (Shutler et al., 1992; Tsilfidis et al., 1991b). It is homologous to chromosome 19 sequences located approximately 80 kb telomeric to pE0.8 the start point of this chromosome walk (Fig. 4-4). The DNA probe pKE0.6 (D19S117) is a 0.6 kb Eco RI fragment of cosmid 6 (f13544) subcloned in pUC13 (Shutler et al., 1991c; Shutler et

al., 1992). It is homologous to a sequence located approximately 100 kb telomeric to pE0.8 (Figure 4-4). The DNA probe pKEX0.8 is a 0.8 kb Eco RI/XbaI genomic fragment from cosmid 16 (f22692) subcloned into Bluescript (Shutler et al., 1992; Tsilfidis et al., 1991c). It is homologous to chromosome 19 sequences located approximately 180 kb telomeric to pE0.8 (Figure 4-4). The DNA probe pKBE0.8 is a 0.8 kb Bam HI/Eco RI fragment from cosmid 21 (f22266) subcloned into pSP65 (Shutler et al., 1992; Korneluk et al., 1991) and is homologous to chromosome 19 sequences approximately 240 kb telomeric to pE0.8 (Figure 4-4).

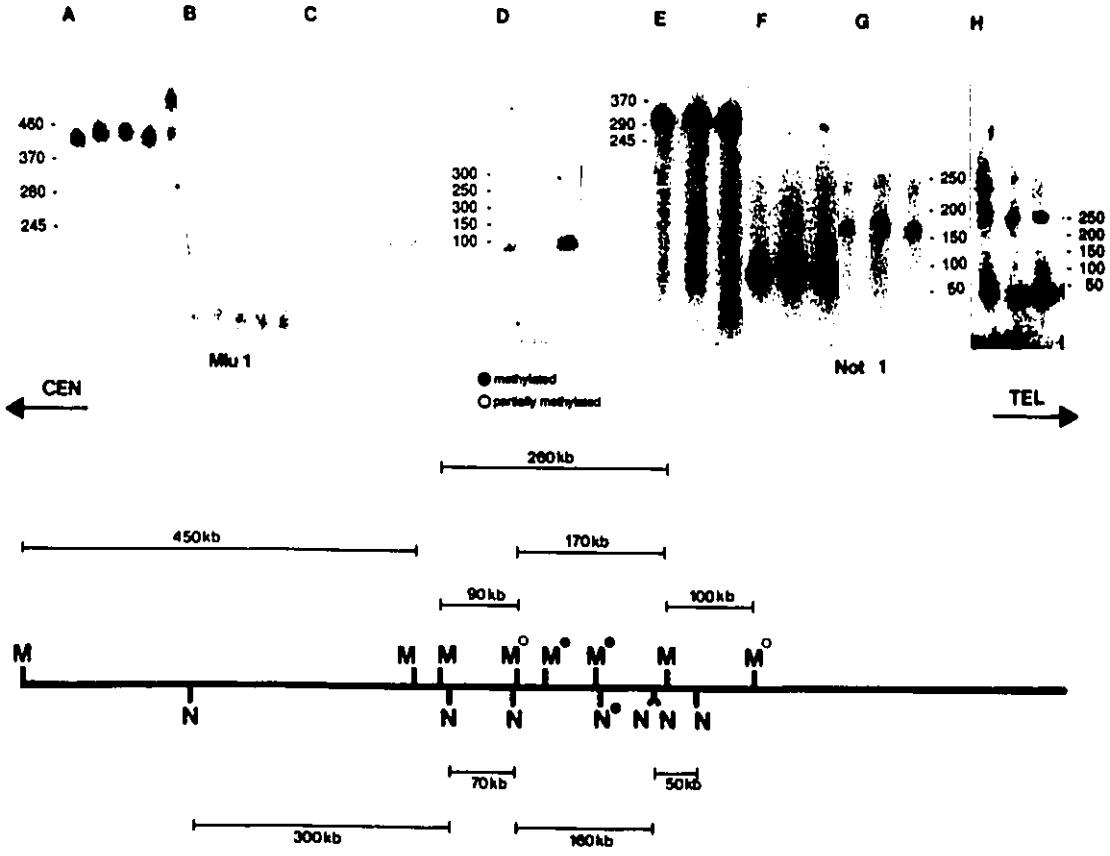
Physical Mapping

A common 450 kb MluI fragment was detected with the pE0.8 and pCKM-3' probes as previously stated (Chapter III) and reported (Shutler et al., 1991a). The next telomeric MluI fragment was found to be 40 kb in size, also as previously stated (Chapter III) and reported (Shutler et al., 1991a). These sites (Figure 4-4, marked M^1 and M^2) flank the ERCC1 gene in this 40 kb region. Hybridization with the more distal DNA probes, pN2.1, pKE2.1 and pKE0.6, revealed both a 90 kb MluI fragment and a partial 260 kb fragment (panel B, Figure 4-11). The MluI endonuclease is expected to be methylation sensitive (Nelson and McClelland, 1987). Thus a partial methylation of the next telomeric site (Figure 4-4, marked M^3) is suggested. The second MluI site in cosmids 15 and 16 (Figure 4-4, marked M^4) along with the MluI site in cosmids 19 and 22 (Figure 4-4, marked M^5) were not evident in the CHEF electrophoresis data. This

Table 4-2: DNA probes derived from the chromosome walk used for physical mapping of the DM gene region.

Probe	Cloning Reference	size of fragment (Kb) MluI	NotI
pERCC1-int	Shutler et al., 1991a	40	300
pE0.8 (D19S115)	Shutler et al., 1991a, b	450	300
pN2.1	Shutler et al., 1992	90, 260	70
pKE2.1 (D19S116)	Tsilfidis et al., 1991b Shutler et al., 1992	90, 260	70
pKE0.6 (D19S117)	Shutler et al., 1991c, 1992a	90, 260	70
pB2.8	Shutler et al., 1992	90, 260	70
ALU33/Sp6-cosmid 14	Shutler et al., 1992	170, 260	160
ALU33/Sp6-cosmid 16	Shutler et al., 1992	170, 260	160
pKEX0.8 (D19S118)	Tsilfidis et al., 1991c Shutler et al., 1992	170, 260	160
pKBE0.8 (D19S119)	Korneluk et al., 1991 Shutler et al., 1992	170, 260	160
pM0.5	Shutler et al., 1992	170, 260	50
ALU33-cosmid 25	Shutler et al., 1992	100	50
ALU34-phage cw13ϕ5	Shutler et al., 1992	100	50

Figure 4-11: Physical map of the proximal DM chromosome walk. Southern blot analysis of CHEF pulsed field gel electrophoresis are shown for MluI (panels A-D) and NotI (panels E-H) digested DNA isolated from the lymphoblasts of unrelated individuals. These top panels show the common restriction fragments identified by DNA probes from the chromosome walk (listed in table 4-2). Panel A shows the 450 kb MluI fragment identified by pE0.8. Panel B shows the common 90 and 260 kb MluI fragments identified by pN2.1, pKE2.1, pKE0.6 and pB2.8. Panel C shows the common 170 kb and 260 kb MluI fragments identified by the two ALU-PCR derived probes, ALU33/Sp6 from cosmid 14 and ALU33/Sp6 from cosmid 16, and pKEX0.8, pKBE0.8 and pM0.5. Panel D shows the 300 kb NotI fragment identified by pERCC1-int and pE0.8. Panel F shows the 70 kb fragment identified by pN2.1, pKE2.1, pKE0.6 and pB2.8. Panel G shows the 160 kb NotI fragment identified by ALU33/Sp6 from cosmid 14, ALU33/Sp6 from cosmid 16, pKEX0.8 and pKBE0.8. Panel H shows the 50 kb NotI fragment identified by pM0.5, ALU33/Sp6 from cosmid 25 and ALU34 from phage cw13ϕ5. The alignment of the MluI (M) and NotI (N) restriction sites identified within the chromosome walk relative to the pulsed field gel electrophoresis results are shown in the physical map below. Methylated restriction sites are shown with a shaded circle while partially methylated restriction sites are shown with an unshaded circle.


A
B
C
D
E
F
G
H
460
370
280
245
300
250
300
150
100
370
290
245
250
200
150
100
50
250
200
150
100
50
Mlu 1
Not 1
CEN
TEL
methylated
partially methylated
260kb
450kb
170kb
90kb
100kb
M
N
70kb
50kb
300kb
160kb

may be due to complete CpG methylation within the MluI recognition site for lymphoblast DNA. The DNA probes pKEX0.8, ALU33-SP6 PCR product from cosmids 14 and 16, and pKBE0.8 hybridize to a common 170 kb MluI fragment and the 260 kb partial MluI fragment (panel C, Figure 4-11). This is consistent with the mapping of the MluI site in cosmids 24 and 25 (Figure 4-4, marked M^6). On the distal side of the last cloned MluI site, two unique sequence DNA probes were generated by ALU-PCR, namely an ALU33 primed PCR product from cosmid 25 and an ALU34 primed PCR product from phage clone cw13ϕ5. Both of these probes detect an MluI fragment of 100 kb (panel D, Figure 4-11).

In addition to the 300 kb NotI fragment previously stated (Chapter III) and reported (Shutler et al., 1991a) to be detected by pCKM3', pERCC1-int and pE0.8, a 70 kb NotI fragment was detected by pN2.1, pKE2.1, pKE0.6 and pB2.8 (panel F, Figure 4-11). This correlated with the NotI site mapped in cosmids 7-11 (Figure 4-4, marked N^2). The NotI site found in cosmids 20 and 22 (Figure 4-4, marked N^3) was not evident in the CHEF electrophoresis data which may be the result of complete methylation in the lymphoblast DNA. NotI is also expected to be methylation sensitive (Nelson and McClelland, 1987).

The next telomeric NotI site mapped in cosmids 23-25 (Figure 4-4, marked N^4), corresponded to a 160 kb NotI fragment detected by the ALU33-SP6 PCR product from cosmid 14, pKEX0.8 and pKBE0.8. The DNA probe pM0.5 derived from cosmid

24, the ALU33 PCR product from cosmid 25 and the ALU34 PCR product from phage cw13ϕ5 all detected a 50 kb NotI fragment. These probes are distal to the tandem NotI sites mapped 11 kb apart in cosmids 23, 24 and 25 (Figure 4-4, marked M^4 and M^5).

Genetic Mapping of Key Recombinant Events

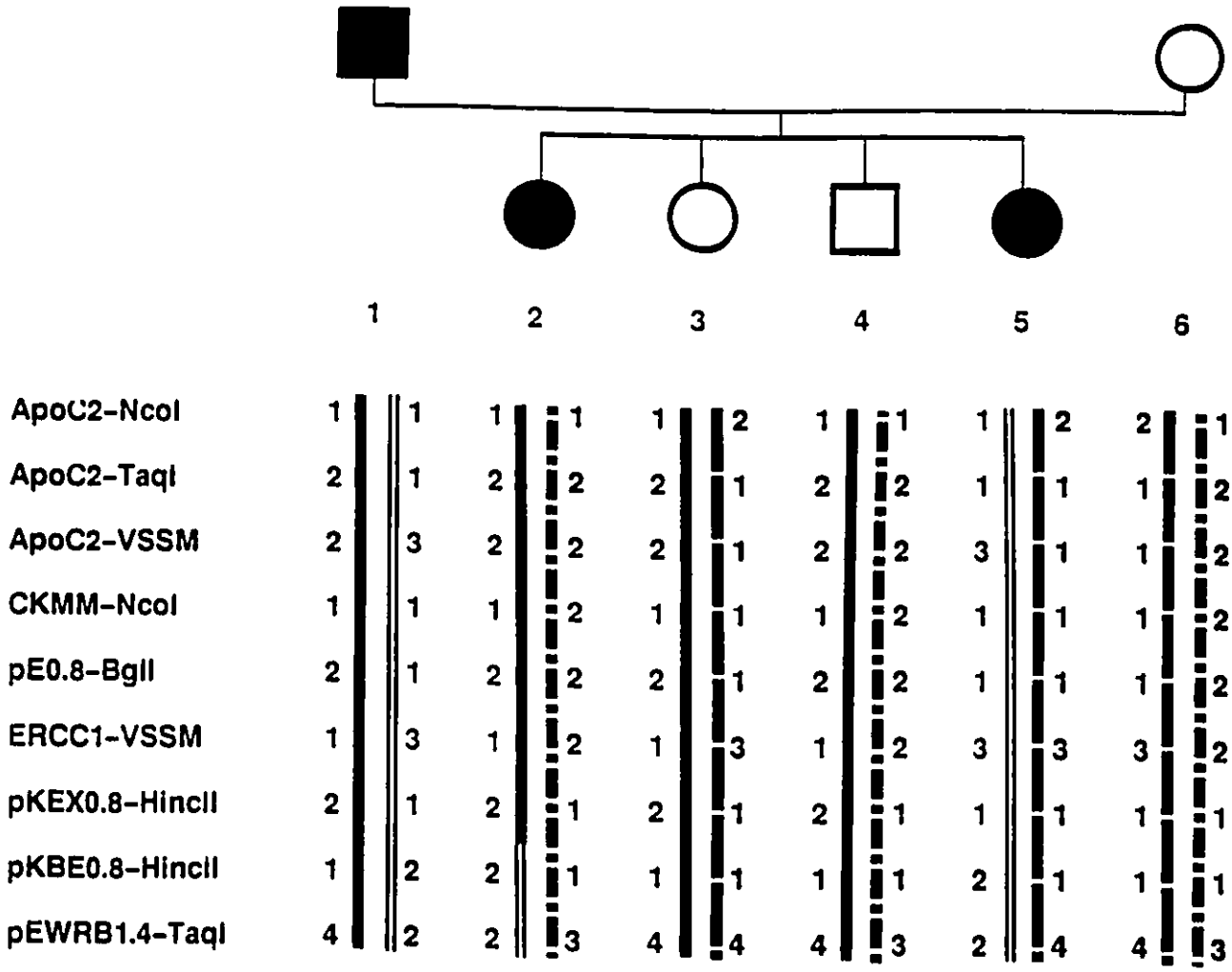
Two key recombinant events bracket the DM locus. The results of a recombinant event in a Dutch DM family, detailed by Smeets et al., (1990b) and discussed in Chapter III, maps the DM locus distal to the ERCC1 gene and pE0.8 (D19S115), the latter being the start point of the chromosome walk in this study. Figure 4-12 shows the haplotyping of this family conducted with the addition of genetic markers from this chromosome walk (listed in Table 4-3). The recombinant event (Figure 4-12, individual #2) shows pKEX0.8 (D19S118) to co-segregate with the more proximal markers pE0.8 (D19S115), ERCC1 and APOC2 and recombine with pKBE0.8 (D19S119), pEWRB1.4 (D19S50) and DM, the distal markers. Thus pKEX0.8 and pKBE0.8 flank the proximal crossover point. This recombinant event maps to within a 60 kb interval based on the physical mapping of these probes in this study (Figure 4-11).

The results of another key recombinant event in a French Canadian DM family, previously described by Tsilfidis et al., 1991a, mapped DM proximal to p134c (D19S51). This probe is closely linked to DM (θ_{max} = 0.006, Z_{max} = 35.8, Tsilfidis et

Table 4-3: Chromosome 19 probes used in genetic analysis of the recombinant event in the Dutch DM family.

Probe	Enzyme	Allelic Size (Kb)	Allelic Frequency	Reference
APOC2 (cDNA)	NcoI	14.5/11.4	0.52/0.48	Frossard et al., 1986
	TaqI	3.8/3.5	0.52/0.48	Humphries et al., 1983
APOC2 (VSSM)	PCR based	12 alleles		Smeets et al., 1989
CKM	NcoI	3.5/2.5+1.0	0.44/0.56	Coerwinkel-Driessen et al., 1988
ERCC1 (VSSM)	PCR based	10 alleles		Smeets et al., 1990a
pE0.8	BglI	6.5/6.0	0.62/0.38	Shutler et al., 1991b
pKEX0.8	HincII	15.5/11.5	0.46/0.54	Tsilfidis et al., 1991c
pKBE0.8	HincII	16/10	0.85/0.15	Korneluk et al., 1991
pEWRB1.4	TaqI	2.3-4.0, 6 alleles	0.07/0.03/0.66 0.01/0.01/0.22	Korneluk et al., 1989

Figure 4-12: Dutch DM family showing a recombination between the proximal markers APOC2, CKM, pE0.8 (D19S115), ERCC1 and pKEX0.8 (D19S118) with the distal markers pKBE0.8 (D19S119), DM and pEWRB1.4 (D19S50). This genetic data suggests that the DM gene lies telomeric to the D19S118 locus.

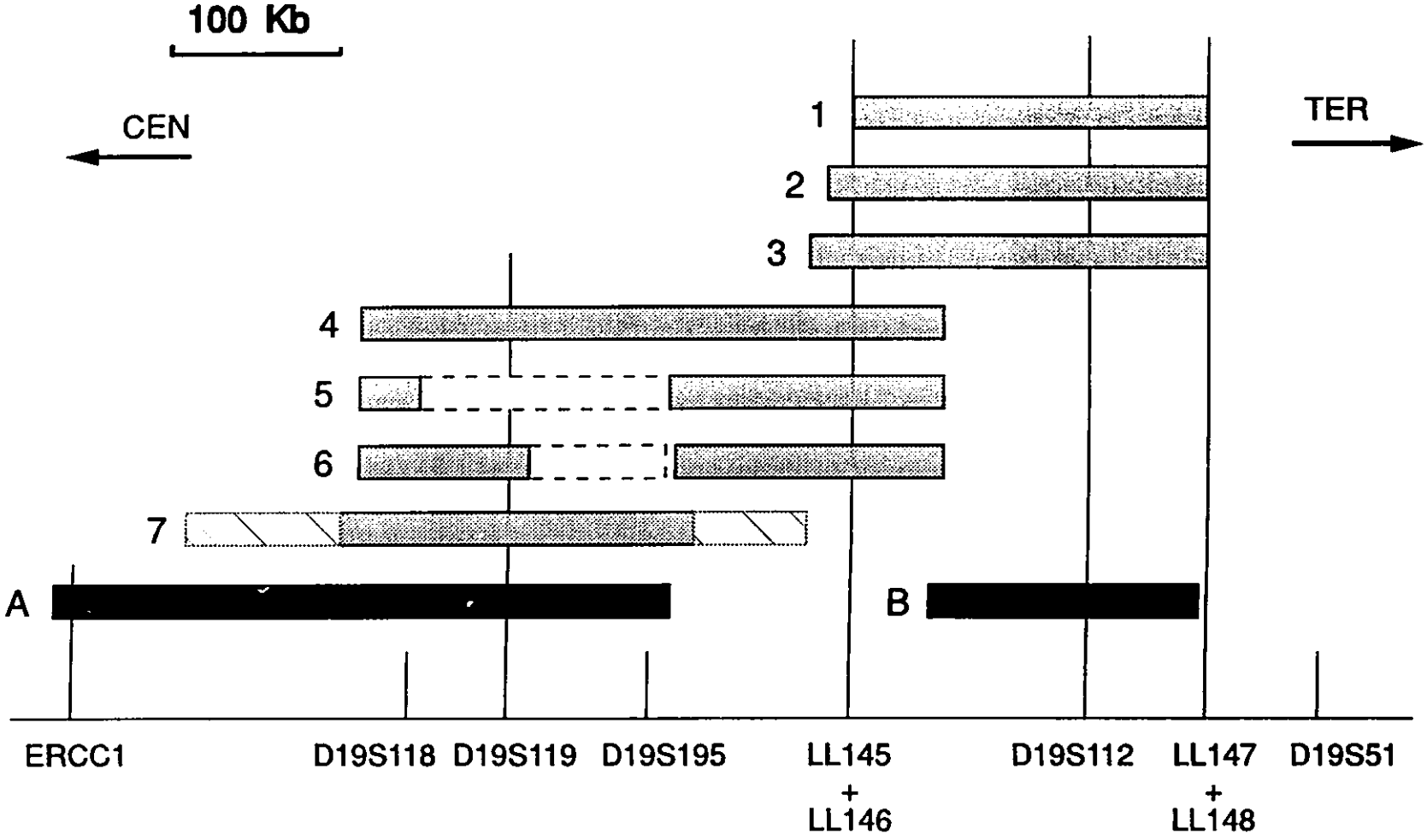

1
2
3
4
5
6
ApoC2-NcoI
ApoC2-TaqI
ApoC2-VSSM
CKMM-NcoI
pE0.8-BglI
ERCC1-VSSM
pKEX0.8-HincII
pKBE0.8-HincII
pEWRB1.4-TaqI

al., 1991a). Thus the DM locus is bracketed by D19S118 as the closest locus on the proximal side and D19S51 on the distal side.

Sequence Tagged Sites and YAC Clone Selection

Two unique sequence DNA probes, pKE0.6 and pKBE0.8 (Figure 4-4) and the ALU34 PCR product from cw13ϕ5 were sequenced and primers made for PCR amplification to generate sequenced tagged sites (STSs; short, unique sequences of DNA that can be recognized by PCR assay). One of the STSs was used to screen the chromosome 19 cosmid library. The STS from the ALU34 PCR product of cw13ϕ5 failed to identify any homogeneous cosmids. The other STS's along with an STS distal to the DM locus, at the pX75b locus (D19S112), were used to screen a YAC library by Dr. Peter de Jong's group at the Lawrence Livermore National Laboratory. The St. Louis YAC library designations were used for the YACs selected. One chimeric YAC clone, 483E7-14, selected by the pKBE0.8 STS and three YAC clones, 168D1, 215B3 and 231G8, selected by the X75 STS failed to overlap (Figure 4-13). ALU-PCR YAC vector end products were generated from YAC clone 231G8, sequenced, primers derived and the centromeric set (Figure 4-13, designated LL145/146) used to rescreen the YAC library. One of the three clones selected was a highly unstable but intact YAC clone, 483E7-14, that overlaps with the proximal walk cosmids (Figure 4-13). A cosmid library was then generated by the Human Genome Centre using a partial MboI digestion of the DM region YAC clones followed by ligation into Lawrist 13 vector. The characterization of these clones is described in Chapter V.

Figure 4-13: Physical map showing the DM region cosmid contigs and overlapping YAC clones. The region spanned by the DM proximal (A) and distal (B) cosmid contigs (shaded bars) are shown relative to the overlapping YAC clones (1-7, grey bars). YAC clones 1, 2 and 3 were selected by the D19S112 STS while YAC clone 7 was selected by the D19S119 STS. The cross hatched region shown for clone 7 indicates a chimeric segment. YAC clones 4, 5 and 6 were selected by an STS (with primers LL145 and LL146) derived from an ALU PCR YAC vector end product of clone 1. The dashed lines shown for clones 5 and 6 show deleted regions due to the instability of these clones. The St. Louis YAC library designations for these clones are as follows: clone 1, 231G8; clone 2, 215B3; clone 3, 168D1; clone 4, 483E7-14; clone 5, 483E7-C; clone 6, 483E7-11; clone 7, 101F2.


100 Kb
CEN
TER
1
2
3
4
5
6
7
A
B
ERCC1
D19S118
D19S119
D19S195
LL145
+
LL146
D19S112
LL147
+
LL148
D19S51

Discussion

An absence of associated chromosome abberations or biochemical abnormalities underscores the role of positional cloning in the isolation of the DM gene. In the search for the causative gene, the DM locus has been bracketed with flanking markers defining two recombinant events. The proximal event is found in a Dutch DM family (Smeets et al., 1990a) and the distal one in a Canadian DM family (Tsilfidis et al., 1991a). The results of the chromosome walk into the proximal segment of this defined region has enabled a high resolution genetic map to be constructed in conjunction with a complementary long range physical map that can be used to define the DM region. By using primarily ALU-PCR generated probes, a cosmid-phage contig extending approximately 350 kb telomeric from D19S115 was generated for these purposes.

The success of the chromosome walk is contingent upon the integrity of the contig and this has been validated in several ways. Restriction sites deduced from CHEF pulsed field electrophoresis are in agreement with the alignment of rare cutter sites in the cloned sequences of the contig. The new genetic markers generated from the chromosome walk all show tight linkage relationships to DM (Korneluk et al., 1991; Shutler et al., 1991b, c; Tsilfidis et al., 1991b, c) and have been used for extensive linkage analysis in Canadian DM families (Tsilfidis et al., 1992b). Furthermore, the ALU-PCR patterns generated by cosmids from successive chromosome walk steps show common centromeric components. Results of restriction mapping the cosmids

and hybridization of overlapping fragments (primarily as a result of determining the orientation of ALU-PCR primer/vector end products) also support the validity of the cosmid contig.

There are two key recombinant events that bracket the DM locus as previously addressed. One recombinant occurs in a Dutch DM family which maps DM distal to the ERCC1 gene (Smeets et al., 1990b) and D19S115 (Shutler et al., 1991a). The D19S115 locus is detected by the DNA probe pE0.8 which is the start point of the chromosome walk in this study. The other recombinant event is in a French Canadian DM family which maps DM proximal to D19S51 (Tsilfidis et al., 1991a). The genetic distance between D19S115 and D19S51 is calculated to be 2 cM or less. In this study we were able to map the proximal crossover point in the Dutch family between two probes, pKEX0.8 (D19S118) and pKBE0.8 (D19S119) which are approximately 60 kb apart (Fig. 4-4). This established pKEX0.8 as the closest proximal marker to DM in this chromosome walk. More significantly, this result eliminates approximately 200 kb of sequence between pE0.8 and pKEX0.8 that would have otherwise been screened for DM candidate genes.

At the distal end of the chromosome walk, no further telomeric cosmids could be found. A similar result was obtained with the chromosome 19 phage library. This suggests the presence of DNA sequence that is difficult to clone in bacterial hosts using cosmid and phage vectors. To approach the problem we initiated an overlap

cloning attempt from the distal side by collaboration with Gert Jansen and Dr. Bé Wieringa of the University of Nijmegen in the Netherlands, using the approach developed during the proximal DM region chromosome walk. This effort involved a chromosome walk through the chromosome 19 cosmid library from a locus defined by the anonymonous marker, pX75b (Jansen et al., 1992a).

A search for markers distal to the DM locus resulted in the cloning of the DNA probe, pX75b by the Dutch group. ALU PCR products were generated from a somatic cell hybrid (XP3542-1) containing the DM region of human chromosome 19. These products were used to screen the phage chromosome 19 library and a phage library that was made from the XP3542-1 somatic cell hybrid. Several unique sequence probes were subcloned from the phage selected and physically mapped using a hybrid cell mapping panel. Two of these probes mapped proximal to the p134c locus. One of these probes mapped at about 30 kb telomeric from the ERCC1 gene in the proximal DM region chromosome walk cosmid contig. The other probe, pX75b, was not present in this cosmid contig thus revealing its distal location. Genetic mapping of the X75b (D19S112) region was accomplished by finding a highly informative VSSM marker from this locus which was found to recombine in the French Canadian DM family previously shown to have the distal recombinant event (presented in Tsilfidis et al., 1991a). Both genetic and physical mapping data placed this locus distal to Dm1 mutation, but closer than p134c. An "essential region" that contains the DM locus is

thus flanked by the nearest crossover markers, D19S118 (pKEX0.8) and D19S112 (pX75b).

The chromosome walk initiated by our Dutch collaborators from the D19S112 locus using the chromosome 19 cosmid library established a 200 kb cosmid contig before reaching a cloning impasse where no further cosmids could be found (Jansen et al., 1992a).

To resolve the impasse in the DM cloning endeavour a YAC library (Burke et al., 1987) was screened at the Human Genome Centre at the LLNL in California with three STS systems, one for pKBE0.8 (D19S119) and one for pKE0.6 (D19S117) from the proximal chromosome walk and one for pX75b (D19S112) from the distal chromosome walk. One chimeric YAC clone was selected by the pKBE0.8 STS and three intact YAC clones were selected by the pX75b STS. These first clones failed to overlap the proximal and distal cosmid contigs. ALU-PCR products from an pX75b STS selected YAC clone, 231G8, were generated and used to probe the chromosome 19 cosmid library. Two overlapping cosmids were identified that bridged part of the gap between the proximal and distal cosmid contigs. Alu-PCR YAC vector end products were also generated from YAC clone 231G8, sequenced, primers derived and the centromeric end set (LL145/146) were used to rescreen the YAC library. A highly unstable YAC clone was found from this second screening. This clone and the YACs from the first screening formed a YAC contig that overlapped with the two cosmid contigs. A

cosmid library was then generated from these YAC clones and the gaps between the proximal and distal chromosome walk cosmid contigs were closed. A characterization of these cosmids is presented in the subsequent chapter.

Chapter V: Identification of The DM Locus

1. Introduction

An extensive region of human chromosome 19q13.3 that contains the DM locus has been cloned in this study and by a collaborative effort as addressed in Chapter IV. The endeavour consisted of: a 350 Kb chromosome walk from the pE0.8 locus (D19S115) located 15 kb centromeric to the DNA Excision Repair (ERCC1) gene, in a telomeric direction toward the DM locus (Shutler et al., 1992a); a 200 kb chromosome walk from the pX75 locus (D19S112) in a centromeric direction toward the DM locus (Jansen et al., 1992a); and linking the two segments with cosmids derived from a YAC contig spanning the region (Aslanidis et al., 1992).

The integrity of the cosmid contig from the proximal DM region chromosome walk was supported by correlating the mapping of rare endonuclease restriction sites in cosmids with results from CHEF pulsed field electrophoresis of genomic DNA as discussed in Chapter IV. Similarly, the integrity of the cosmid contig from the distal DM region chromosome walk and the cosmids derived from overlapping YAC clones was established by comparing rare restriction endonuclease sites in these cosmids with results from CHEF pulsed field electrophoresis. This was facilitated by mapping NotI and MluI sites in the distal DM region and YAC derived cosmid contig, searching for unique sequence DNA probes in this contig and analyzing CHEF Southern blots

with these DNA probes. Establishment of the cosmid contig's integrity ensures the presence of the DM locus mapped within this interval.

Both genetic and physical mapping data are essential for positional cloning strategies to find a disease gene locus. The point when tight genetic linkage indicates what cloning strategies are appropriate is often not clear. Defining a relationship between marker to marker genetic distance and physical distance estimates would assist in making such a decision. Previous studies that have correlated genetic to physical mapping data were based on large poorly characterized regions of several Mb. These studies used long range physical mapping data obtained from pulsed field gel electrophoresis with partially digested DNA (Higgins et al., 1990) or from somatic cell hybrid mapping panels (Carson and Simpson, 1991; Julier et al., 1988) and genetic data from anonymous markers. Evidence from these studies suggest a variable sex dependent relationship between marker to marker physical and genetic distance estimates.

The opportunity to compare complementary genetic and physical mapping data from a well characterized region in excess of 1 Mb spanning the DM locus was afforded during the course of mapping the DM gene by positional cloning. The cloning of large cosmid contigs, pulsed field electrophoretic analysis and somatic cell hybrid data was correlated in this study of the DM region to provide marker to marker distance measurements for six polymorphic loci over a 1.5 Mb region.

Complementary linkage analysis of these six polymorphic loci distributed through out the DM region on a total of 85 Canadian DM families comprising 904 individuals was thus undertaken to analyze the genetic distance to physical distance relationship.

In the final phase of this study the fine mapping and partial characterization of the cosmids to define the minimal area containing the DM locus is presented. This includes the cloning of unique sequence DNA probes that identify the DM mutation and the identification of CpG islands flanking the DM locus.

2. Materials and Methods

DNA Analysis

Genomic DNA was extracted from either peripheral blood samples by using a modification of Madisen et al., (1987) or by the salt precipitation method of Miller et al., (1988). Cosmid and plasmid DNA was isolated by rapid alkaline preparation (Birnboim and Dolly, 1979). Cloned probes or total human DNA used for hybridization were labelled with ^{32}P dCTP by the random priming method of Feinberg and Vogelstein (1983). Restriction enzyme digests and Southern blot hybridization conditions were as previously described (MacKenzie et al., 1989).

DNA Probes

Single copy sequences from Southern blots of cosmid restriction fragments were identified by their lack of hybridization to repetitive sequences, when probed with ^{32}P-

dCTP labelled human DNA. After subcloning, these single copy sequence probes were used for physical mapping and if they detected restriction length polymorphisms, they were also used for genetic mapping. The other DNA probes used for linkage analysis and/or physical mapping are as described in Korneluk et al., (1989a), Johnson et al., (1990), Shutler et al., (1991a, b, 1992a).

Pulsed Field Gel Electrophoresis

Contour clamped homogeneous electric field (CHEF) pulsed field electrophoresis gels were run on the CHEF DRII system (BIO-RAD). Electrophoresis run conditions and preparation of DNA from lymphoblasts were as previously described in Chapter II.

Linkage Analysis

A total of 85 Canadian DM families comprising 904 individuals were studied. Approximately half of the families were French-Canadian and the majority of the remaining families were either Scottish-Canadian or Anglo-Canadian. The clinical diagnosis of DM and linkage analysis using the LINKAGE version 5.1 package of programs (Lathrop et al., 1985) were as previously outlined (MacKenzie et al., 1989; Korneluk et al., 1989a). Compound haplotypes for loci containing more than one polymorphic site (eg. APOC2 and CKM) were constructed as previously described (Korneluk et al., 1989b).

3. Results

To construct a consensus physical map of the DM region a characterization of the cosmids derived from the distal DM region chromosome walk and overlapping YAC clones was undertaken and combined with previously characterized proximal DM region data (Chapter IV). From a broad perspective, this enables correlation of genetic and physical mapping data throughout a region in excess 1 Mb spanning the DM locus. More specifically, this also permits fine genetic and physical mapping to define the minimal region expected to contain the DM mutation.

Characterization of Cosmid Clones

In order to confirm the integrity of the cosmid contig from the distal DM region chromosome walk and Yac derived cosmids spanning the DM locus, the correlation of previously mapped landmark rare endonuclease sites in these cosmids (Jansen et al., 1992a; Aslanidis et al., 1992) with genomic CHEF pulsed field electrophoresis data derived from six DNA probes found in this region was undertaken. A total of 10 MluI sites were reported, 6 MluI sites from the proximal chromosome walk (Figure 5-1, marked and numbered M^{1-6}), Shutler et al., (1992a), and 4 MluI sites from the distal chromosome walk and YAC overlap cloning (Figure 5-1, marked and numbered M^{7-10}), Jansen et al., (1992a) and Aslanidis et al., (1992). A total of 9 NotI sites were reported, 5 NotI sites from the proximal chromosome walk (Figure 5-1, marked and numbered N^{1-5}), Shutler et al., (1992a), and 4 NotI sites from the distal chromosome walk and YAC overlap cloning (Figure 5-1, marked and numbered N^{6-9}), Jansen et al.,

(1992a) and Aslanidis et al., (1992). Other MluI sites, M^{11-14}, and NotI sites, N^{10-11}, are predicted by the CHEF pulsed field electrophoresis data.

The MluI and NotI sites mapped in the cosmids are listed in Table 5-1. The positions of cosmids derived from YACs and the distal DM chromosome walk are shown with reference to the landmark MluI and NotI sites in Figure 5-2. Cosmids 23, 24 and 25 are the three most telomeric cosmids from the proximal DM chromosome walk (Table 5-1 and Figure 5-2). The YAC derived cosmids are shown as cosmids 26 to 29 inclusive, 32 and 33 (Table 5-1 and Figure 5-2). Cosmids 30 and 31 (Table 5-1 and Figure 5-2) are two contiguous cosmids derived from the chromosome 19 cosmid library that does not overlap with other cosmids in the library. Cosmids 34 to 43 inclusive, were derived from the distal DM chromosome walk (Table 5-1 and Figure 5-2).

Isolation of DNA Probes

A total of eight unique sequence DNA probes were subcloned from cosmids in this study. Four of these (pKM1.4, pS1.8, pGB2.6 and pX1.7) along with three other chromosome 19 DNA probes were used for CHEF pulsed field electrophoresis studies (Table 5-2). Three of the probes were found to detect RFLP's (pKM1.4, pGB2.6 and pGP1.5). These new probes along with six other chromosome 19 DNA probes were

Table 5-1: Cosmids and rare endonuclease sites identified from the DM region chromosome walks and overlapping YAC clones (listed in a centromere to telomere order).

Cosmid	LLNL id#	Rare cutter site	Ref.
1	F15123	M^{1-2}	Shutler et al., 1992
2	F8697	M^{2}	*ibid*
3	F8191	M^{2}, N^{1}	*ibid*
4	F10080	N^{1}	*ibid*
5	F5118		*ibid*
6	F13544		*ibid*
7	F23770	M^{3}, N^{2}	*ibid*
8	F18905	M^{3}, N^{2}	*ibid*
9	F20720	M^{3}, N^{2}	*ibid*
10	F24206	M^{3}, N^{2}	*ibid*
11	F22397	M^{3}, N^{2}	*ibid*
12	F18226	M^{3}, N^{2}	*ibid*
13	F22440	M^{3}, N^{2}	*ibid*
14	F20927	M^{3}, N^{2}	*ibid*
15	F15962	M^{3-4}, N^{2}	*ibid*
16	F22692	M^{4}	*ibid*
17	F24860		*ibid*
18	F20007		*ibid*
19	F20829		*ibid*
20	F16149		*ibid*
21	F22266		*ibid*
22	F25267	M^{5}, N^{3}	*ibid*
23	F20000	N^{4-5}	*ibid*

Cosmid	LLNL id#	Rare cutter site	Ref.
24	F15243	M^6, N^{4-5}	*ibid*
25	F16236	M^6, N^{4-5}	*ibid*
26	Ya100075	N^6	Aslanidis et al., 1992
27	Ya100068		*ibid*
28	Ya100157	M^7, N^7	*ibid*
29	Ya100124	M^7, N^7	*ibid*
30	F18894	M^7, N^7	Jansen et al., 1992
31	F18431	M^7, N^7	*ibid*
32	Yc100259	M^7, N^7	Aslanidis et al., 1992
33	Yc100263	M^7, N^7	*ibid*
34	F16625	M^8	Jansen et al., 1992
35	F23948	M^8	*ibid*
36	F25609		*ibid*
37	F14255	M^9	*ibid*
38	F25240	M^{10}	*ibid*
39	F5864	M^{10}	*ibid*
40	F18029	M^{10}	*ibid*
41	F17058	M^{10}	*ibid*
42	F23842	N^8	*ibid*
43	F20192	N^{8-9}	*ibid*

Figure 5-1: Long range restriction map of over 1 Mb spanning the DM gene region. The approximate positions for the ApoC2, CKM, ERCCI and DM genes are indicated in brackets. Loci for the DNA probes are marked and indicated with an arrow. The landmark rare endonuclease restriction sites identified in the DM cosmid and YAC contig, MluI (M) and NotI (N), are marked and numbered in a centromere (CEN) to telomere (TER) orientation. The unshaded circles above the restriction sites indicate partial digestion and shaded circles indicate no digestion observed in genomic DNA. The unshaded square above the restriction sites indicate restriction sites subject to partial digestion predicted solely on pulsed field electrophoresis results (discussed latter in text). In the enlarged area, between MluI sites M^{7-8}, the shaded triangle marks the position of the unstable CTG repeat associated exclusively with the DM chromosome (discussed in text latter), while the unshaded triangle marks the position of the insertion polymorphism detected at the DM locus.

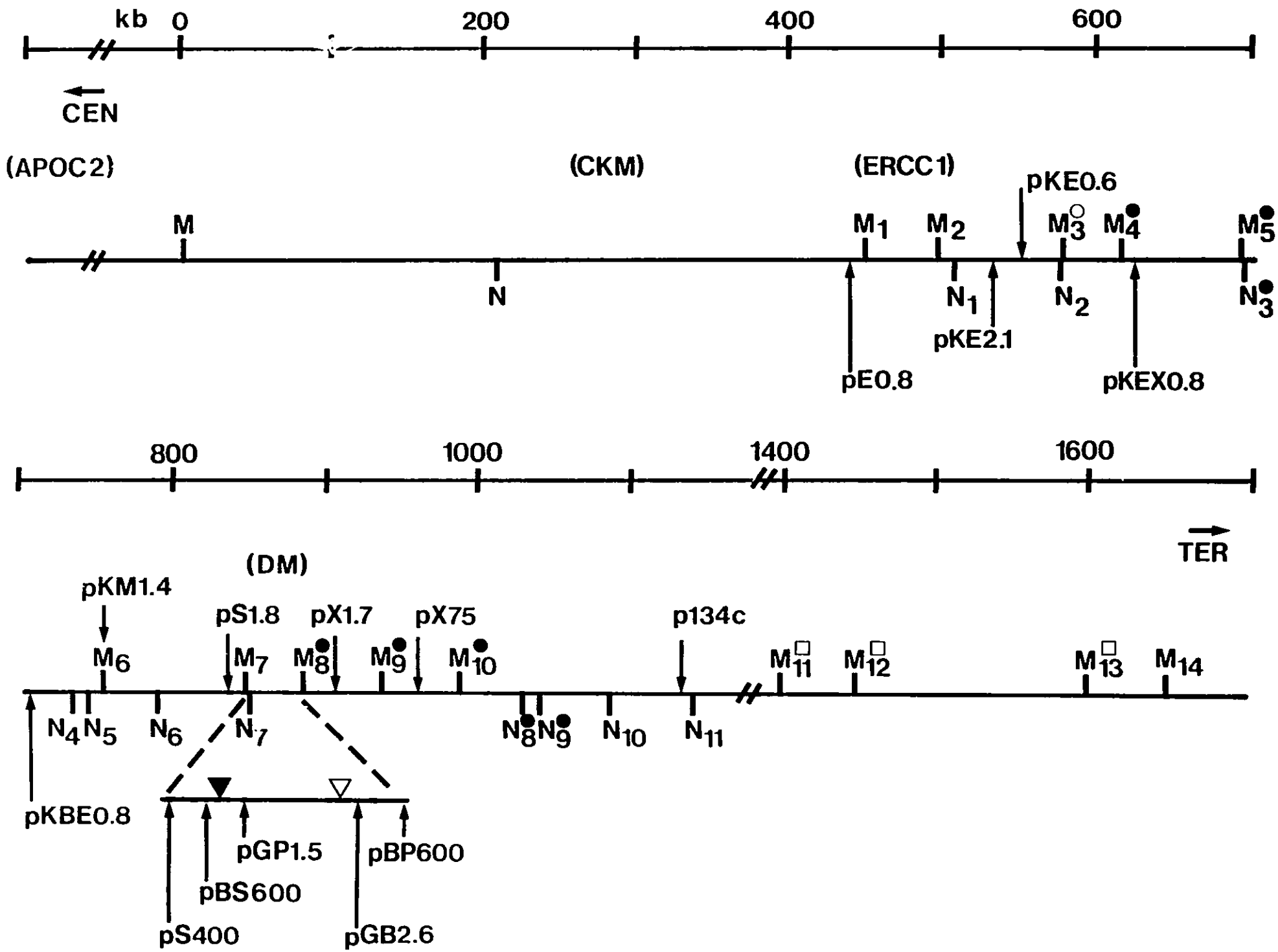

kb
0
200
400
600
CEN
(APOC2)
(CKM)
(ERCC1)
pKE0.6
M
N
M1
M2
M3
M4
M5
N1
N2
N3
pKE2.1
pE0.8
pKEX0.8
800
1000
1400
1600
TER
(DM)
pKM1.4
pS1.8
pX1.7
pX75
p134c
M6
M7
M8
M9
M10
M11
M12
M13
M14
N4
N5
N6
N7
N8
N9
N10
N11
pKBE0.8
pGP1.5
pBP600
pBS600
pS400
pGB2.6

Figure 5-2: Rare cutter restriction map of YAC derived chromosome 19 library gap and distal DM region cosmid contig. The positions are shown of the most telomeric three cosmids from the proximal DM chromosome walk (cosmids 23, 24 and 25), the YAC derived cosmids (cosmids 26 to 29, 32 and 33), a small contig of two chromosome 19 library derived cosmids (cosmids 30 and 31), and cosmids derived from the distal DM chromosome walk are shown in reference to four MluI sites (marked and numbered M^{6-10}) and seven NotI sites (marked and numbered N^{4-10}) contained in the cosmid contig. The shaded circles above the restriction sites indicate that they have not been observed in genomic DNA. The loci of three probes used for physical mapping are indicated as a reference points below, D19S195 (localized to about 100 kb from pGB2.6 and about 20 kb from the most telomeric end of cosmid 25), pGB2.6 (a unique sequence probe detecting the unstable CTG repeat polymorphism exclusively associated with DM) and D19S112 (the starting point for the distal DM chromosome walk, mapped to about 100 kb from pGB2.6).

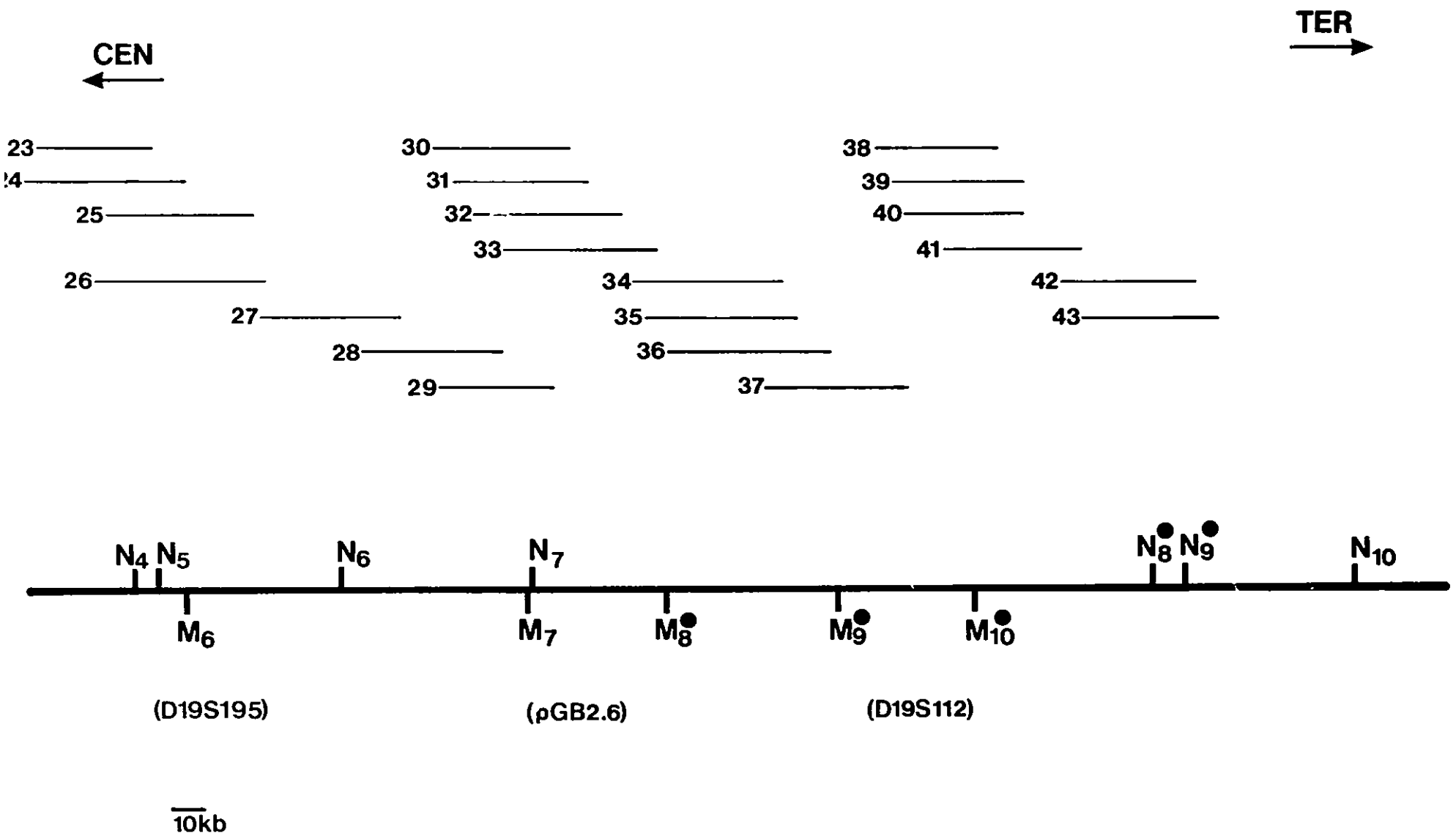

CEN
TER
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
N4 N5
N6
N7
N8 N9
N10
M6
M7
M8
M9
M10
(D19S195)
(pGB2.6)
(D19S112)
10kb

Table 5-2: Chromosome 19 probes used for extending the physical map of the DM gene region by pulsed field gel electrophoretic analysis.

Probes	size of fragment identified (kb)		Cloning Reference
	MluI	**NotI**	
pKM1.4	100	50-60	this study
pS1.8	100	50-60	this study
pGB2.6	800 (750, 600, 500)	245-260	Aslanidis et al., 1992 and this study
pX1.7	800 (750, 600, 500)	245-260	this study
pX75	800 (750, 600, 500)	245-260	Jansen et al., 1992a
p134c	800 (750, 600, 500)	50-60	Johnson et al., 1990

used in linkage analysis for this thesis and elsewhere (Table 5-3). Three of the probes, pBM800, pS400 and pBP600, were used to map CpG islands.

The pKM1.4 probe is a 1.4 kb Bam HI/HindIII fragment from cosmid 24 subcloned in pUC18. Both the pS1.8 and pS400 probes are 1.8 kb and 400 bp SstII fragments respectively from an 11 kb Eco RI subclone of cosmid 39 (f18894). The pGB2.6 and pGP1.5 probes are a 2.6 kb Bam HI fragment and a 1.5 kb PstI fragment from cosmid 33 (Yc100263) subcloned into pUC18 and pBS, respectively. The pX1.7 probe is a 1.7 kb XbaI fragment from a 10 kb Bam HI subclone of cosmid 34 (f16625). The pBP600 probe is a 600 bp Bam HI/PstI fragment from a 1.9 kb Bam HI/HindIII subclone of cosmid 24 (f16625). The pBM800 probe is a 800 bp Bam HI/MluI fragment from a 14 kb Bam HI subclone of cosmid 33 (Yc100263).

Physical Mapping

Previous studies have shown that both pE0.8 and pCKM-3' identify a common MluI fragment of 450 kb and a common NotI fragment of 300 kb (Shutler et al., 1991a). The telomeric ends of the fragments flank the ERCC1 gene (Figure 5-1, M^1 and N^1) thus mapping the CKM locus to within 250 kb of the ERCC1 locus (Shutler et al., 1991a; Smeets et al., 1990). Several studies have shown that the APOC2 locus is in close proximity, centromeric to the CKM locus (Korneluk et al., 1989a; Smeets et al., 1990; Stallings et al., 1988; Harley et al., 1991; Shaw et al., 1985; Schonk et al., 1989;

Table 5-3: Chromosome 19 probes used for linkage analysis of the DM gene region.

Probe	Enzyme	Allelic size (kb)	Allelic frequency	Reference
pα1.4 (D19S37)	BanI	3.0/2.2	0.72/0.28	Korneluk et al., 1989
APOC2 cDNA (APOC2)	NcoI	14.5/11.4	0.52/0.48	Frossard et al., 1986
	TaqI	3.8/3.5	0.52/0.48	Humphries et al., 1983
cCKM3' (CKM)	NcoI	3.5/2.5+1.0	0.44/0.56	Coerwinkel-Driessen, et al., 1988
pE0.8 (D19S115)	Bam HI BglI NcoI	7.0/6.5 6.5/6.0 4.5/4.0	0.62/0.38	Shutler et al., 1991a, b
pGB2.6 (DM gene)	Eco RI HindIII	10/9.0 9.5/8.5		Aslanidis et al., 1992
p134c (D19S51)	BglI	5.0/4.4	0.85/0.15	Johnson et al., 1990

Brooke et al., 1991). Physical linkage by pulsed field electrophoretic analysis between APOC2 and pα1.4 loci has been reported (Shaw et al., 1989), but not between the APOC2 and CKM loci due to the presence of numerous rare endonuclease sites. However, consistent coretention of APOC2 and CKM in somatic hybrid experiments suggest a close physical proximity between these loci (Stallings et al., 1988; Brooke et al., 1991). Furthermore, based on the pulsed field electrophoresis data available, APOC2 can be mapped at just over 260 kb proximal to CKM (Smeets et al., 1990).

Agreement of restriction mapping MluI sites M^1 to M^6 with CHEF pulsed field electrophoresis results has already been demonstrated in Chapter IV. Hybridization of pKM1.4 and pS1.8 (unique sequence probes subcloned from cosmids F15243 and F18894, Table 5-1 and Figure 5-1) to MluI digested genomic DNA revealed a common 100 kb fragment (Figure 5-3) which correlated to the mapping of MluI site M^7 (found in cosmids Ya100157, Ya100124, Yc100259, Yc100263, F18431 and F18894, Table 5-1 and Figure 5-2) to about 100 kb distal to M^6 (Figure 5-1).

Hybridization of four unique sequence probes from different loci (pGB2.6, pX1.7, pX75b and p134c, Table 5-2 and Figure 5-1) to MluI digested DNA, revealed a common 800 kb fragment, in the DNA tested from most individuals (Figure 5-4, lanes 1 and 2). There was however, a common complex pattern of four fragments (approximately of 800, 750, 600 and 500 kb) also observed (Figure 5-4, lanes 3 and 4).

Figure 5-3: CHEF pulsed field electrophoresis results of MluI digested genomic DNA; Presence of MluI sites M^6 and M^7. Southern blot analysis of MluI digested genomic DNA prepared from lymphoblasts and separated on a CHEF pulsed field electrophoresis gel. Hybridization of this blot with the unique sequence DNA probes pKM1.4 and pS1.8 identifies a common 100 Kb fragment. The hybridization result for pKM1.4 is shown. The identification of a 100 kb fragment with these probes is consistent with the mapping of MluI sites M^6 and M^7 about 100 kb apart in the cosmid contig.


200
150
100
50

Figure 5-4: CHEF pulsed field electrophoresis results of MluI digested genomic DNA; Absence of MluI sites M^8 and M^9. Southern blot analysis of MluI digested genomic DNA prepared from lymphoblasts and separated on a CHEF pulsed field electrophoresis gel. Hybridization of this blot with the unique sequence DNA probes pGB2.6, pX1.7, pX75 and p134c identifies two distinct band patterns (pGB2.6 is shown); a single 800 Kb band (lanes 1 and 2) or a multi-band pattern (lanes 3 and 4). MluI sites M^8, M^9 and M^{10} were mapped about 40 kb, 90 kb and 130 kb from MluI site M^7 in the cosmid contig. MluI fragments of this size range were not found in pulsed field analysis of genomic DNA in this region. The next MluI site evident in genomic DNA (M^{11}) is thus about 500 kb from MluI site M^7.

This complex pattern is difficult to resolve and had hampered previous efforts to construct a pulsed field electrophoresis based map of the region distal to MluI site M^7 (Jansen et al., 1992a; Aslanidis et al., 1992). Consequently the expected MluI sites, M^{8-10} (Figure 5-1), restriction mapped in cosmids F16625, F23948, F14255, F25240, F5864, F18029 and F17058 (Table 5-1 and Figure 5-2), are not evident in genomic lymphoblast DNA by CHEF pulsed field electrophoretic analysis. Presumably, these results can be explained by differences in the methylation status between these restriction sites. The recognition sites for both MluI and NotI include the CpG dinucleotide which when methylated, inhibits digestion by both enzymes (Nelsen and McClelland, 1987).

To substantiate the CHEF pulsed field electrophoresis evidence that MluI site M^8 is not detected in genomic DNA while MluI site M^7 is detected was provided by experiments with Bam HI/MluI double digests of leucocyte DNA. Hybridization with pBM800 (Figure 5-6), a unique sequence probe subcloned from cosmid Yc100263 (Table 5-1 and Figure 5-2), for MluI site M^7 (Figure 5-1) and pBP600 (Figure 5-5), a unique sequence probe subcloned from cosmid F16625 (Table 5-1 and Figure 5-2) for MluI site M^8 (Figure 5-1) was used to test for the presence of these MluI sites. The predicted 800 bp Bam HI/MluI double digested fragment was observed with the pBM800 hybridization result (Figure 5-6, panel A) which supports the pulsed field electrophoretic evidence for the detection of MluI site M^7 in genomic DNA.

Figure 5-5: Restriction map of the genomic segment between two landmark MluI restriction sites, $M^7$and M^8, flanking the DM locus. The map was constructed by the analysis of subclones derived from cosmids F18894, Yc100263 and F16625 covering the region from MluI site M^7 to M^8. Map orientation is from centromere to telomere. The shaded triangle marks the position of the DM variable length polymorphism while the unshaded triangle marks the position of the insertion polymorphism detected at the DM locus. The shaded diamond indicates the proposed CpG island mapped 3' to the DM gene while the unshaded diamond indicates the proposed CpG island mapped 5' to the DM gene. DNA probes used to map restriction sites of rare endonucleases inhibited by methylation of the dinucleotide CpG in their recognition site for genomic DNA are indicated below. Restriction sites that could not be identified in genomic DNA (leucocyte) are marked with a shaded circle. The position of various relevant restriction sites are given: S, SstII; M, MluI; N, NotI; B, Bam HI; E, Eco RI; P, PstI; H, HindIII; Eg, EagI; Bs, BssHII; X, XhoI.

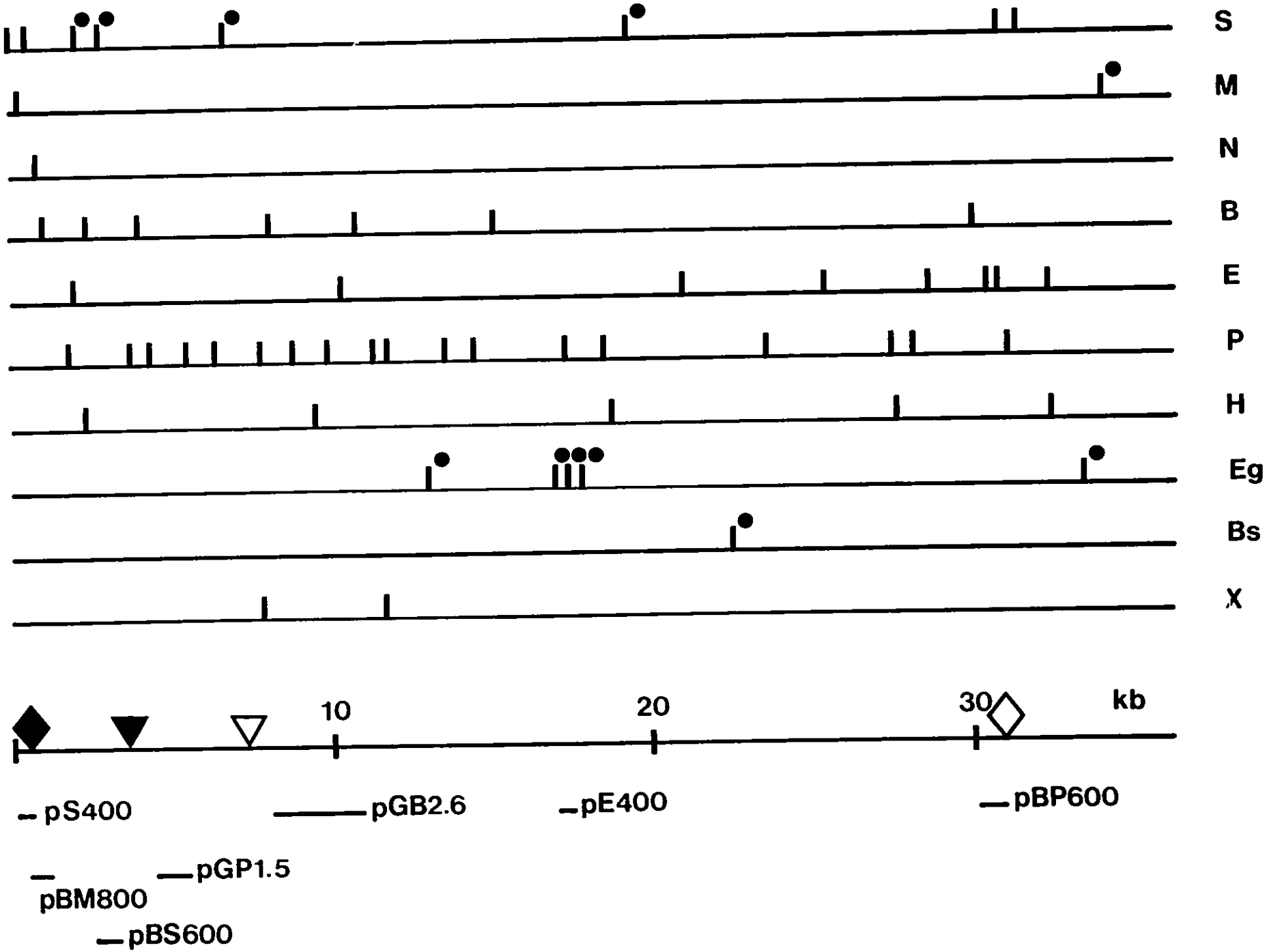
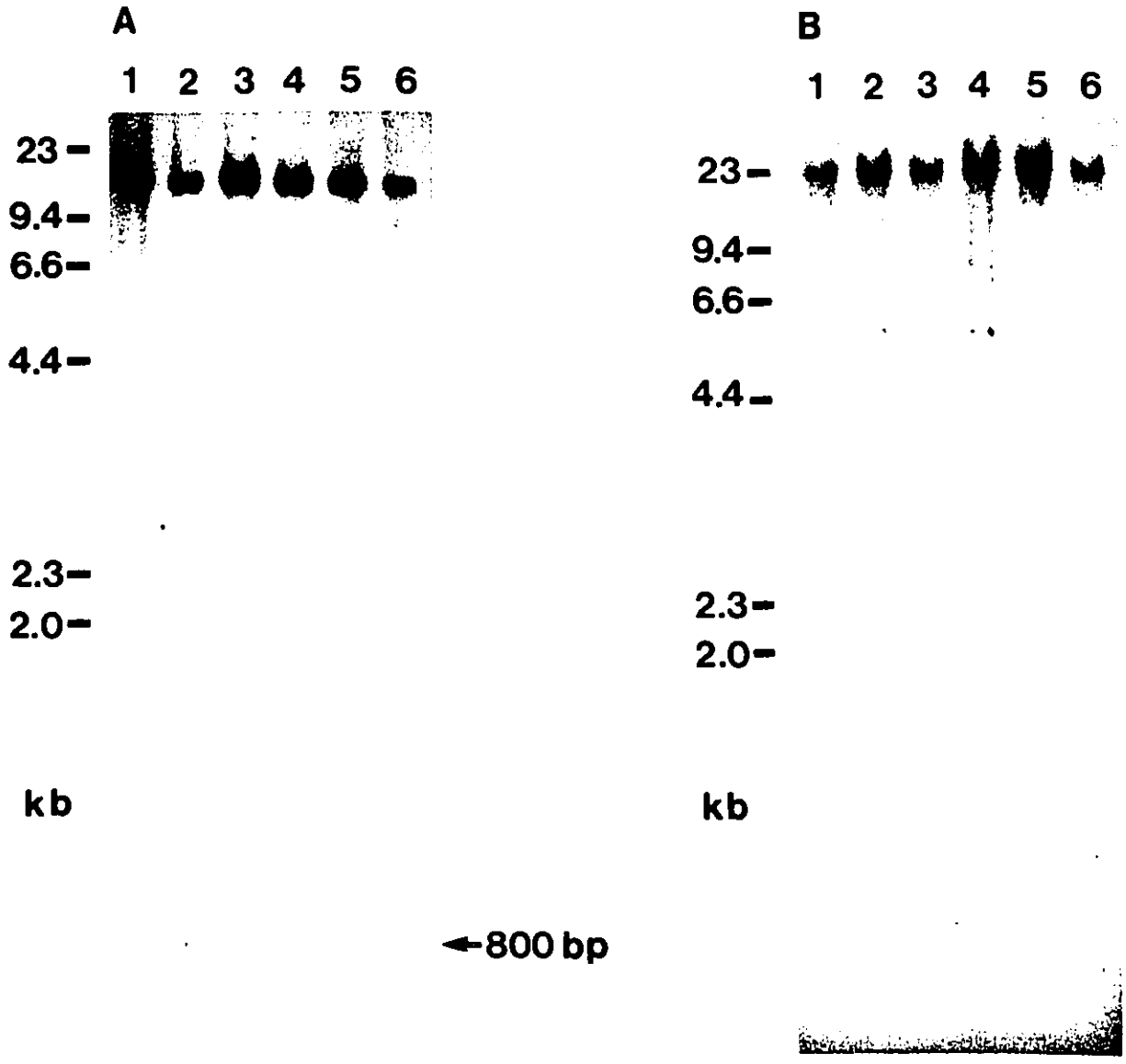

S
M
N
B
E
P
H
Eg
Bs
X
10
20
30
kb
pS400
pGB2.6
pE400
pBP600
pGP1.5
pBM800
pBS600

Figure 5-6: Southern blot analysis of MluI sites $M^7$and M^8. Normal (lanes 3 and 4) and DM patient (lanes 1, 2, 5 and 6) genomic DNA digested with Bam HI (odd numbered lanes) and double digested with Bam HI and MluI (even numbered lanes) was probed with pBM800 (A) and pBP600 (B). The double digest anticipated Bam HI to MluI site M^7 800 bp product (marked by an arrow) was detected by pBM800 (A). The expected 4 Kb product of Bam HI to MluI site M^8 was not detected by pBP600 hybridization (B).

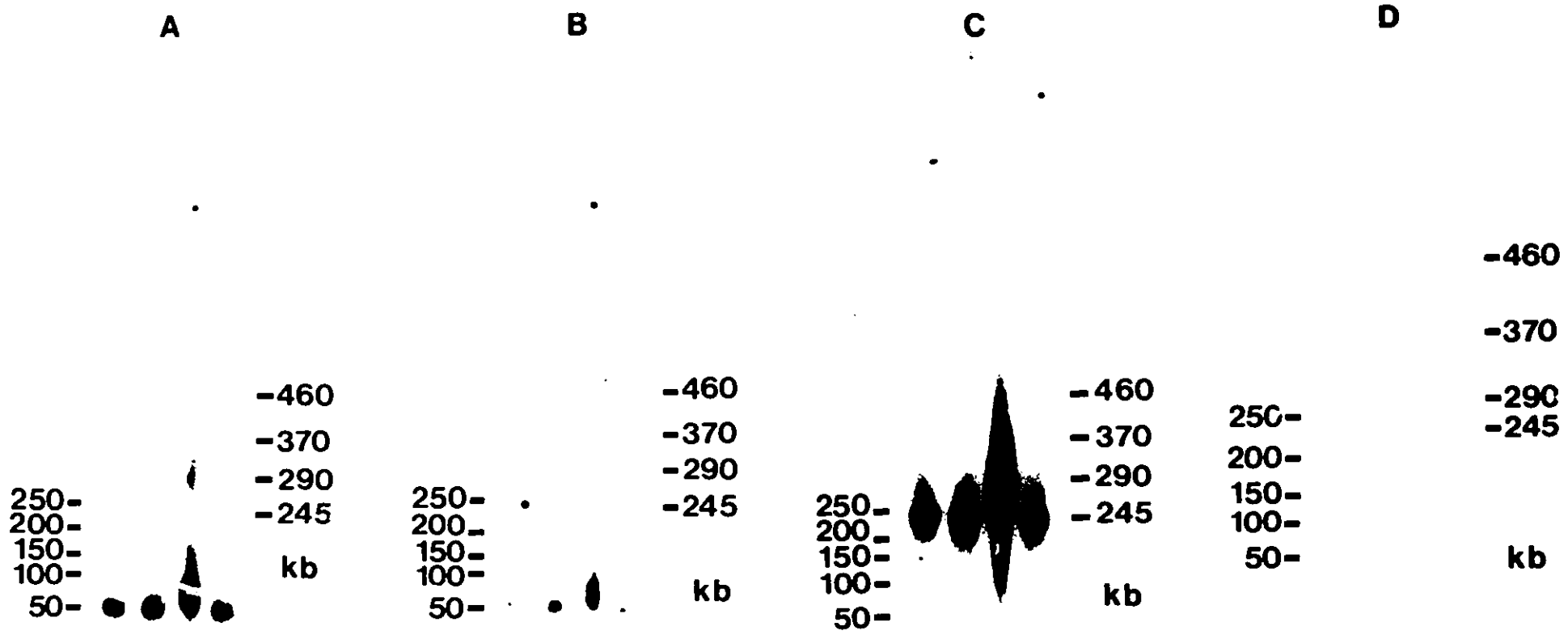

A
1 2 3 4 5 6
23
9.4
6.6
4.4
2.3
2.0
kb
B
1 2 3 4 5 6
23
9.4
6.6
4.4
2.3
2.0
kb
800 bp

Conversely, hybridization with pBP600 did not yield the expected 4 kb fragment (panel B, Figure 5-6) which supports the pulsed field electrophoretic analysis result of not detecting MluI site M^8 in genomic DNA. Both normal and DM patient leucocyte DNA were used in these experiments with no detectable differences.

The common complex pattern of MluI fragments detected by the probes with loci distal to MluI site M^7, that was evident in the lymphoblast DNA of some individuals, suggests the presence of at least four more MluI sites (Figure 5-1, marked and numbered M^{11-14}) distal to cosmid F20192 (Table 5-1 and Figure 5-2). Furthermore, partial methylation of these MluI sites, M^{11-13}, could account for the number of fragments detected in these individuals. Hybridization of these CHEF pulsed field electrophoresis blots with probes from loci proximal to MluI site M^7 did not indicate a general problem with partial digestion of these samples. This phenomenon was not observed with NotI digested DNA in the DM gene region tested proximal to the p134c locus.

Agreement of NotI sites N^{1-5} in the cosmids with genomic CHEF pulsed field electrophoresis results has already been demonstrated in Chapter IV. Hybridization of pKM1.4 to NotI digested genomic DNA revealed a 50-60 kb band (Figure 5-7, panel A) that supports the restriction mapping of NotI site N^6 (Figure 5-1) in cosmid Ya100075 (Table 5-1 and Figure 5-2). Hybridization of pS1.8 to NotI digested genomic DNA,

Figure 5-7: CHEF pulsed field electrophoresis results of NotI digested normal and DM patient genomic DNA. Southern blot analysis of NotI digested DNA prepared from lymphoblasts and separated by CHEF pulsed field electrophoresis is shown. Hybridization with pKM1.4 identifies a 50-60 Kb fragment (A), pS1.8 identifies another 50-60 Kb, fragment (B), while pGB2.6, pX1.7 and pX75 identifies a common 245-260 Kb fragment (C). Hybridization with p134c identifies a 50-60 Kb fragment (D).


A
B
C
D
-460
-370
-290
-245
250-
200-
150-
100-
50-
kb

detected a 50-60 kb fragment (Figure 5-7, panel B) which correlates with the restriction mapping of the NotI site N^7 found in cosmids Ya100157, Ya100124, F18894, F18431, Yc100259 and Yc100263 (Table 5-1 and Figure 5-2). Hybridization of pGB2.6, pX1.7 and pX75 (Table 5-2) to NotI digested genomic DNA revealed a common 245-260 kb fragment (Figure 5-7, panel C), which does not correlate with the restriction mapping of NotI sites N^{8-9} found in cosmids F23842 and F20192 (Table 5-1 and Figure 5-2). The next NotI site distal to N^7 that is evident in genomic DNA is at a distance of about 245 kb while the restriction mapping of NotI sites N^{8-9} in cosmids F23842 and F20192 indicates they are only at about 60 kb and 110 kb from N^7, respectively. The NotI site, N^{10}, is expected to occur in more distal cosmids. Hybridization of p134c (Table 5-2) with NotI digested genomic DNA shows a 50-60 kb fragment (Figure 5-7, panel D).

Correlation of Genetic and Physical Distance Estimates

Both genetic and physical mapping data are essential for locating a disease gene locus. The point when tight genetic linkage indicates that cloning strategies are appropriate is often not clear. Defining a relationship between marker to marker genetic and physical distance estimates would assist in making such a decision. Comparison of complementary genetic and physical mapping data from the over 1 Mb region spanning the DM locus was undertaken to examine this relationship.

Distance estimates between loci of the various probes used for both physical and genetic mapping were deduced by combining CHEF pulsed field electrophoresis

results and restriction mapping of MluI and NotI sites (Figure 5-8). Genotypic data was obtained from 85 families comprising 905 individuals using six DM region probes. The genetic distance in cM with confidence intervals and lod scores (Z_{max}) were calculated and are shown both in table 5-4 and figure 5-8. Lod scores from the marker to marker recombinant fractions ranged from 13.5 to 69. Genetic distance (cM) was estimated directly from the recombinant fractions for the various marker pairs. Separation of recombinant fractions based on gender showed only minor differences (Table 5-5). Thus the combined male and female recombinant fractions are used for comparison to physical distance data.

Linear regression analysis of the two parameters using the method of least squares (Figure 5-9) failed to reveal a significant linear correlation (coefficient of determination, $r^2 = 0.19$). The effect of using the most accurate linkage data was then studied. We found the lack of correlation can be eliminated if only marker pairs with the highest lod scores are considered by using marker pairs with lod scores greater than 22 to select for tighter confidence intervals. A linear correlation between the two parameters ($r^2 = 0.64$) was then observed. This correlation was within the broad confidence intervals of the eliminated marker pairs (Figure 5-9). Furthermore, the omitted marker pairs (Table 5-4 and Figure 5-9, #3, pGB2.6 to p134c; #7, pα1.4 to CKM; #9, pα1.4 to CKM; #10, CKM to p134c; #11, CKM to p134c; #13, pα1.4 to pGB2.6; #15, pα1.4 to p134c) do not originate from a common region (Figure 5-10).

TABLE 5-4: Estimates of genetic and physical distance for DM region markers.

	Interval	Mb	cM (confidence interval)	Z_{max}
1	pα1.4 to APOC2	0.25	0.001 (0.001-3.5)	32.9
2	CKM to pE0.8	0.25	0.5 (0.001-3.0)	35.9
3	pGB2.6 to p134c	0.28	0.60 (0.001-6.5)	15.0
4	APOC2 to CKM	0.26	0.3 (0.001-2.0)	69.0
5	pE0.8 to pGB2.6	0.40	0.001 (0.001-2.2)	24.4
6	APOC2 to pE0.8	0.61	0.5 (0.001-2.7)	36.4
7	pα1.4 to CKM	0.61	2.7 (0.1-6.5)	21.2
8	CKM to pGB2.6	0.65	0.4 (0.001-2.1)	50.9
9	pE0.8 to p134c	0.68	0.4 (0.001-7.5)	15.0
10	pα1.4 to pE0.8	0.76	0.4 (0.005-8.0)	13.5
11	CKM to p134c	0.93	2.8 (0.1-9.0)	18.1
12	APOC2 to pGB2.6	0.99	0.7 (0.001-2.2)	60.4
13	pα1.4 to pGB2.6	1.21	5.6 (1.5-11)	15.8
14	APOC2 to p134c	1.24	1.9 (0.001-6.0)	28.6
15	pα1.4 to p134c	1.5	0.001 (0.001-16)	11.3

Figure 5-8: Consensus physical and genetic map of the DM region. The marker to marker physical distance in kb for 6 loci from the DM region is shown on the left. Marker to marker recombinant fractions and lod scores (bracketed) are shown on the right.

Consensus Physical and Genetic Map of DM Region

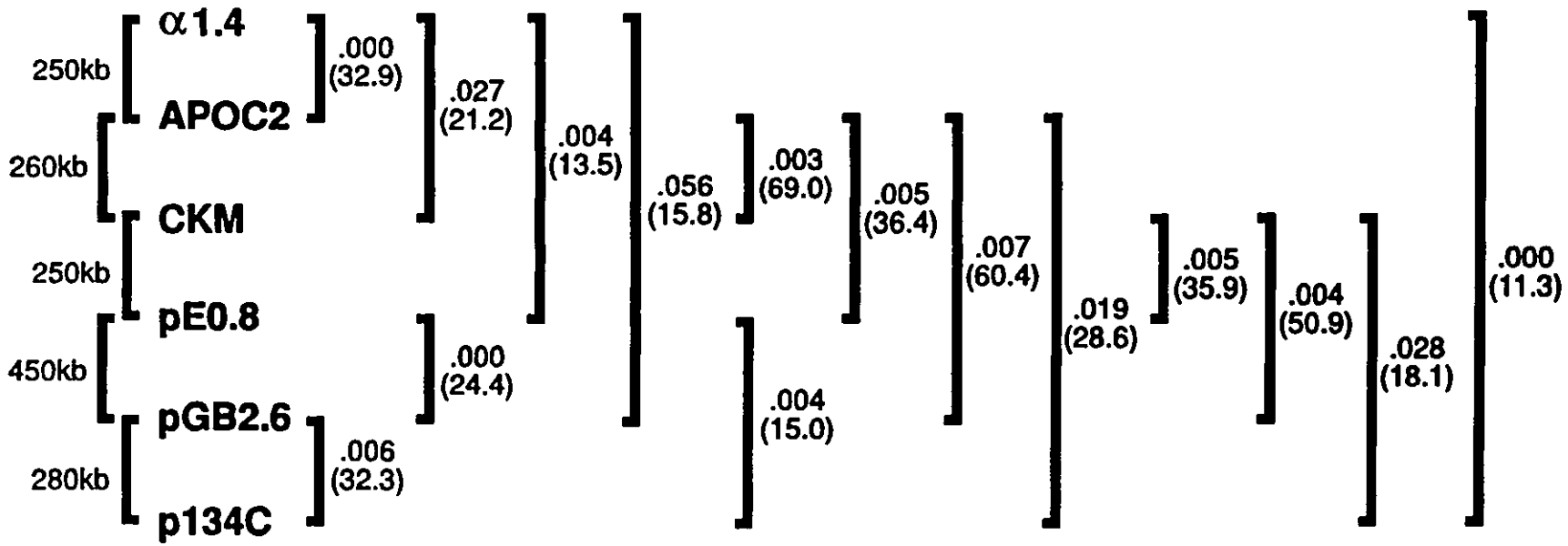


Table 5-5: Male and female recombination fractions for DM region markers.

	Interval	θ (male)	θ (female)	Z_{max}
1	pα1.4 to APOC2	0.000	0.000	33
2	CKM to pE0.8	0.011	0.000	36
3	p134c to pGB2.6	0.000	0.013	33
4	APOC2 to CKM	0.006	0.000	69
5	pE0.8 to pGB2.6	0.000	0.000	24
6	APOC2 to CKM	0.009	0.000	37
7	pα1.4 to CKM	0.000	0.044	22
8	CKM to pGB2.6	0.008	0.000	51
9	pE0.8 to p134c	0.000	0.000	15
10	pα1.4 to pE0.8	0.000	0.008	13
11	CKM to p134c	0.033	0.025	18
12	APOC2 to pGB2.6	0.007	0.008	60
13	pα1.4 to pGB2.6	0.054	0.057	16
14	APOC2 to p134c	0.017	0.022	29
15	pα1.4 to p134c	0.000	0.000	11

Figure 5-9: Plot of the genetic distance versus the physical distance between markers from the DM region. A comparison of genetic distance in cM (including confidence intervals) to physical distance in Mb for 15 marker to marker intervals in the DM region is shown. Linear regression analysis using the method of least squares revealed a correlation between the two parameters only if intervals with high lod scores are considered. The coefficient of determination for this correlation was determined to be 0.64, the coefficient of correlation was found to be 0.80 while the standard error of estimate was shown to be 0.40. The shaded squares represent genetic to physical distance ratios for intervals with high lod scores ($Z_{max} > 22$) the unshaded squares indicate genetic to physical distance ratios for intervals with lower lod scores ($Z_{max} < 22$).

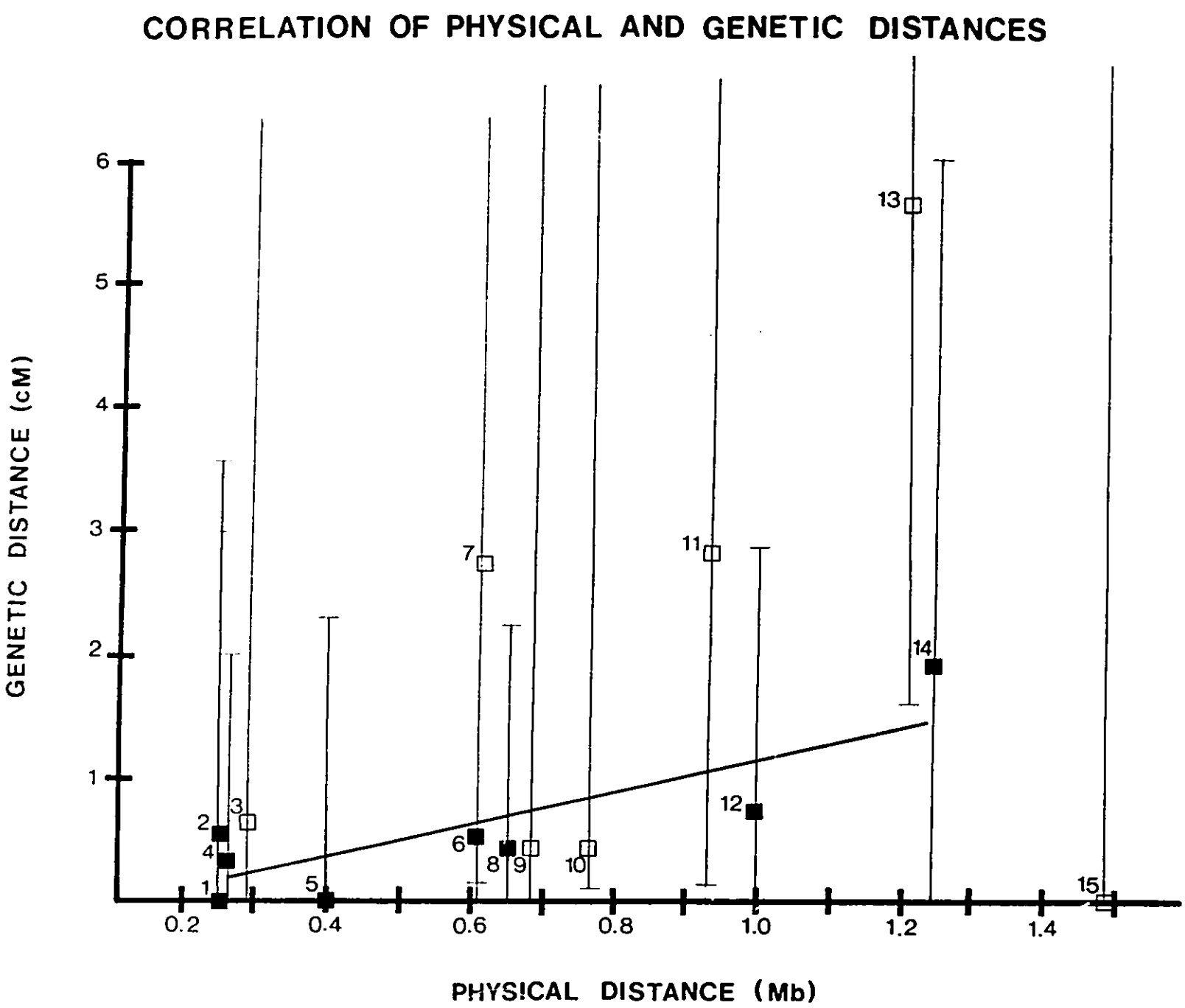
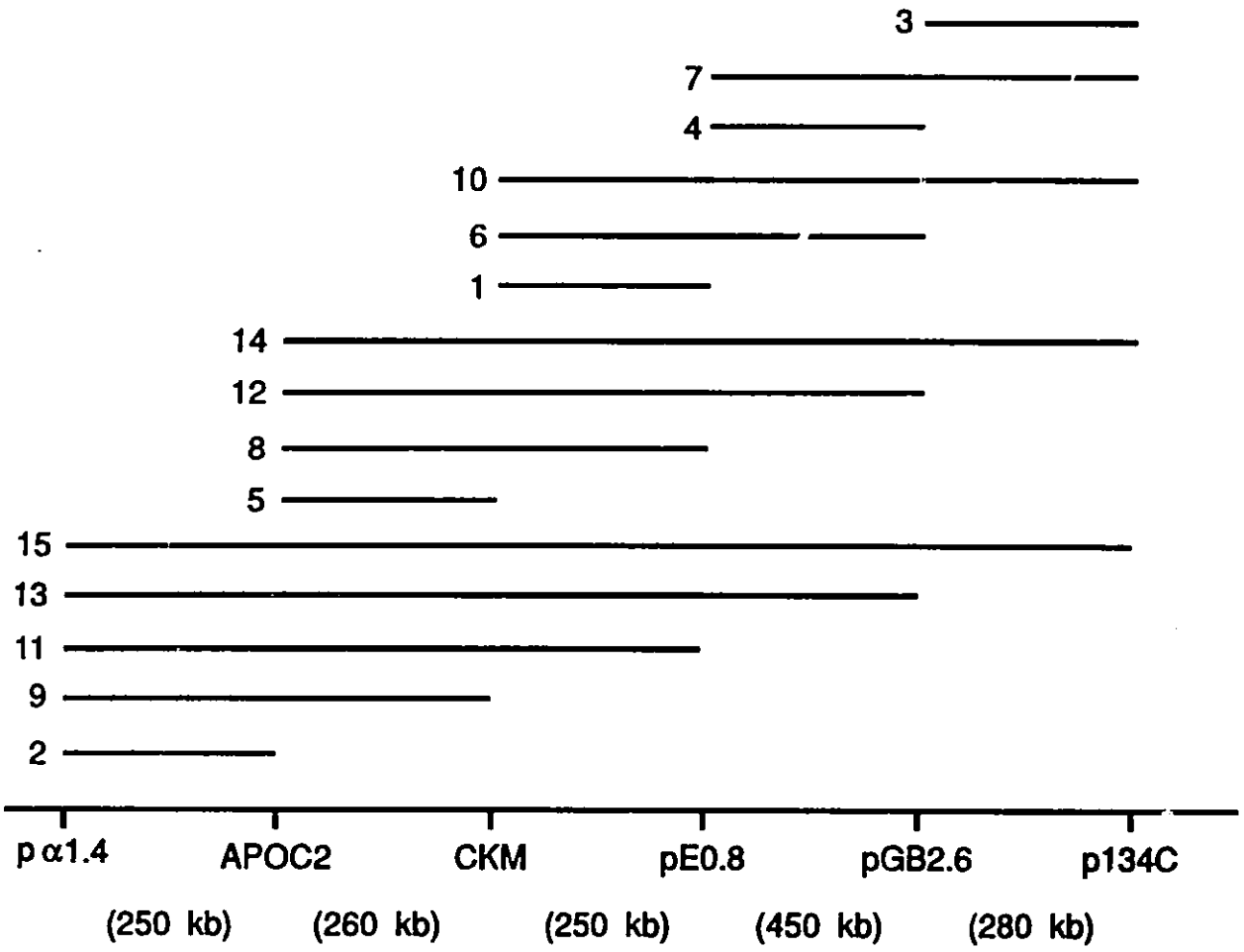

CORRELATION OF PHYSICAL AND GENETIC DISTANCES
GENETIC DISTANCE (cM)
6
5
4
3
2
1
0.2
0.4
0.6
0.8
1.0
1.2
1.4
PHYSICAL DISTANCE (Mb)
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15

Figure 5-10: Location of DM region marker intervals used for comparison of genetic to physical distances. The locations of the marker intervals from the DM region used for genetic distance to physical distance comparison (from Figure 5-9) are shown. There are seven points with lod scores below 22 that were omitted from the linear regression analysis shown in figure 5-9 (interval #3, pGB2.6 to p134c; interval #7, pα1.4 to CKM; interval #9, pα1.4 to CKM; interval #10, CKM to p134c; interval #11, CKM to p134c; interval #13, pα1.4 to pGB2.6; interval #15, pα1.4 to p134c). It can be seen here from the location of those intervals that they do not originate from a common region. The physical distance between the loci for the six DM region markers is shown separately in brackets.

Location of DM marker intervals used for comparison of genetic to physical distance

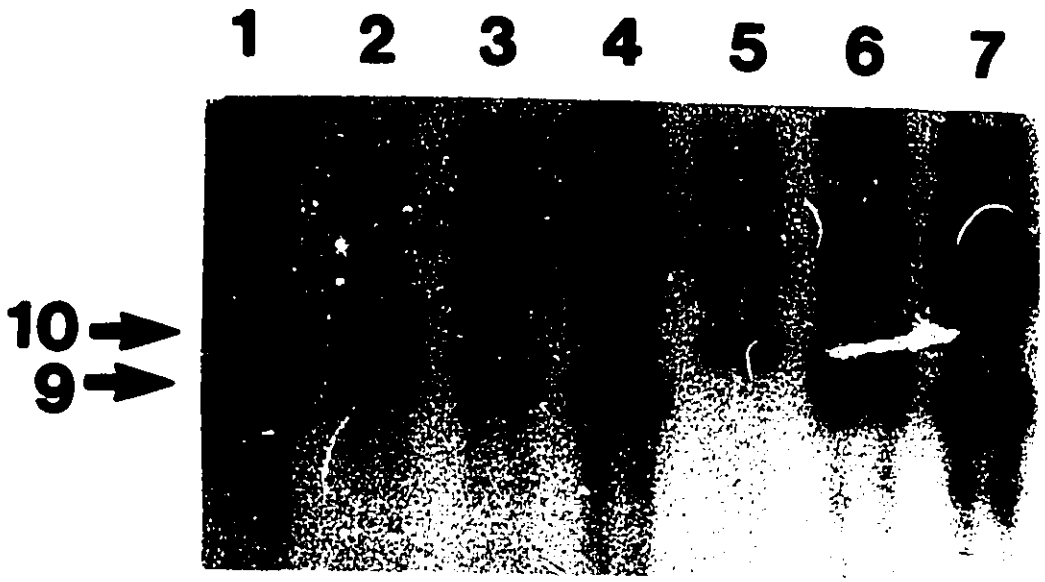


Identification of the DM Mutation with Genomic Probes

The DNA probes pGB2.6 and pGP1.5 are a 2.6 kb Bam HI fragment and a 1.5 kb PstI fragment subcloned from cosmid Yc100263 into pUC18 and Bluescript plasmids, respectively. Both of these probes map to a 9 kb Eco RI fragment present in the YAC derived cosmid Yc100263 and a 10 kb Eco RI fragment in the chromosome 19 cosmid library derived cosmid f18894 (Fig. 5-5). These two Eco RI fragments represent the alleles of an insertion polymorphism evident in genomic DNA from normal individuals (lanes 1, 2 and 6, Figure 5-11). Both of these probes also detect a variable length polymorphism with Eco RI or HindIII digested genomic DNA from DM affected individuals (lanes 3, 4, 5 and 7, Figure 5-11). An extensive linkage analysis of this variable length polymorphism with DM, primarily using pGB2.6, was conducted in our laboratory on DM families (Aslanidis et al., 1992). In nearly 70% of approximately 200 DM patients, increased allele sizes ranging from 10.5 to 15 kb was revealed. This mutation was subsequently identified in our laboratory as an amplification of a CTG trinucleotide repeat found to be present in the 3' untranslated region of a gene for a serine/threonine protein kinase (Mahadevan et al., 1992).

In the analysis of DM pedigrees it was discovered that the expansion on the DM chromosome observed in successive generations correlated with the increased severity of the disease (Aslanidis et al., 1992; Mahadevan et al., 1992; Buxton et al., 1992; Harley et al., 1992; Brook et al., 1992; Fu et al., 1992; Tsilfidis et al., 1992). The amplification of the CTG trinucleotide repeat thus forms the molecular basis for genetic

Figure 5-11: Southern blot analysis showing the various degrees of allelic expansion in DM-affected individuals. Genomic and cDNA probes that map to the same 10 kb Eco RI from human chromosome 19q13.3 detect an unstable CTG repeat unique to individuals with DM. The two Eco RI alleles, 9 kb and 10 kb, identified by these probes (pGB2.6 is shown) are found in the normal population (lanes 1, 2 and 6). The 10 kb allele contains the trinucleotide repeat that is polymorphic but stably inherited in the normal population. The trinucleotide repeats range in size from 5 to 30 with repeat numbers of 5 and 13 the most common. In DM patients there are increases in the size of the 10 kb allele that are due to an increased number of CTG repeats ranging from 50 repeats to several kb. The CTG trinucleotide repeat maps to the 3' untranslated region of a putative serine-threonine protein kinase gene. Nearly all cases of DM displayed expansion of the CTG repeat region (discussed in text and latter in Chapter VI). An increase in the severity of the disease in successive generations (genetic anticipation) is accompanied by an increase in the number of CTG trinucleotide repeats. Expansion of DM alleles seen by Southern blot analysis was classified in the following manner: E0, no expansion; E1, expansion of 0 to 1.5 kb; E2, expansion of 1.5 to 3.0 kb (lane 3); E3, expansion of 3.0 to 4.5 kb (lanes 5 and 7); E4, expansion of 4.5 kb or greater (lane 4).


1
2
3
4
5
6
7
10
9

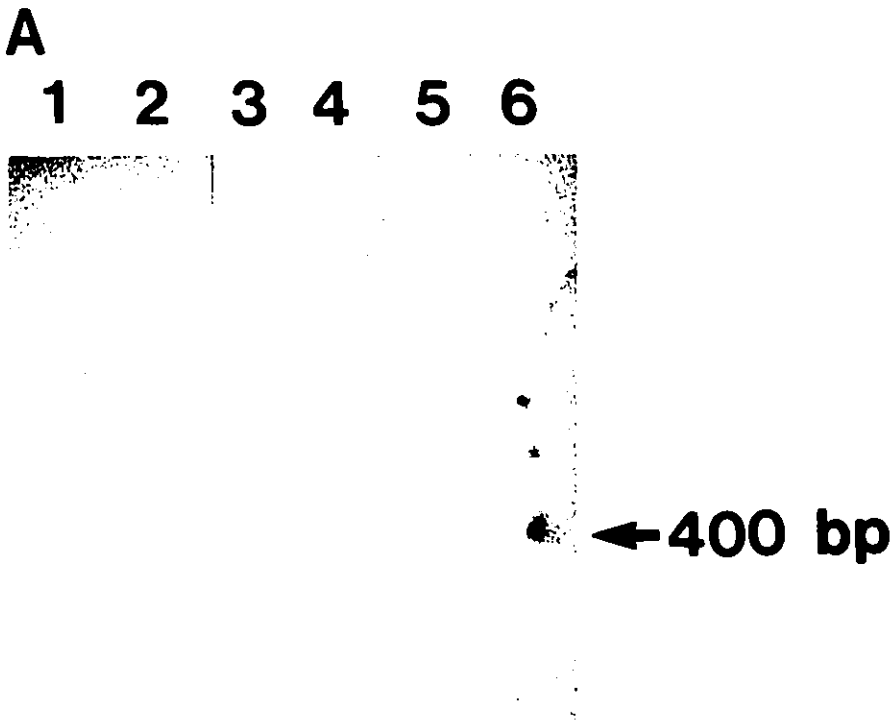
anticipation observed in DM patients. These observations led to the conclusion that the expansion likely constitutes the DM mutation. Results of collaborative studies support these conclusions and are further examined in the next chapter.

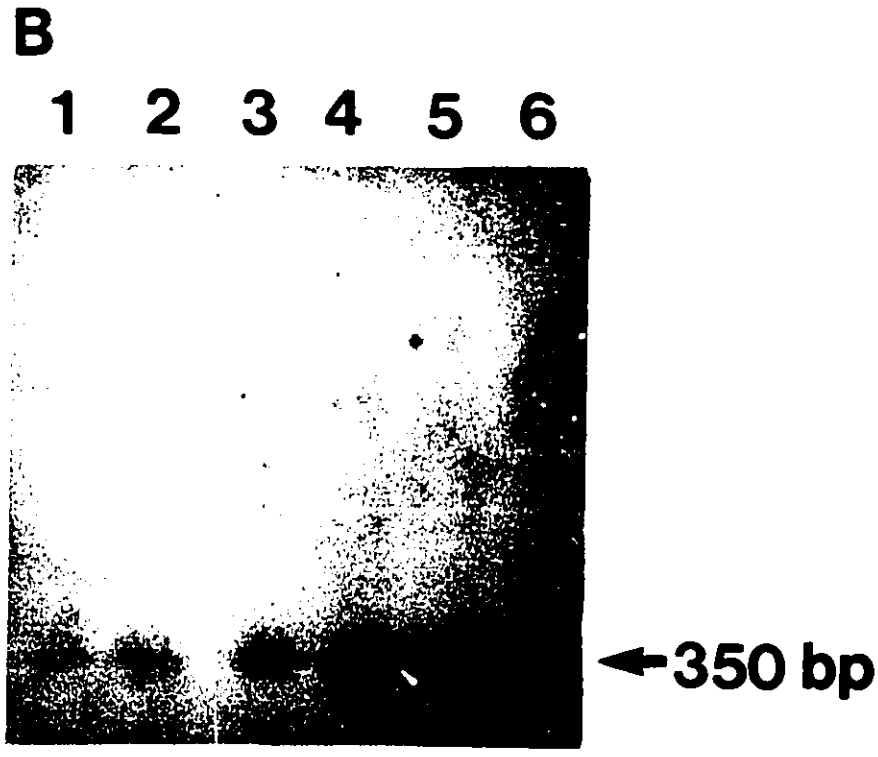
The DM Mutation and Mapping CpG Islands that Flank the DM Gene

In the characterization of the DM locus a search for associated CpG islands was conducted. CpG islands are CpG rich areas that are hypomethylated and near the 5' end of "housekeeping" genes and many tissue specific genes (Bird, 1987; Razin and Ceder, 1991). Rare cutting endonucleases such as SstII, EagI, BssHII and MluI have the CpG dinucleotide in their recognition sequences and are often associated with CpG rich areas (Lindsay and Bird, 1987). The 10 kb Eco RI fragment from cosmid f18894 and the homologous 9 kb Eco RI fragment from cosmid Yc100263 that contains the DM mutation maps between MluI sites M^7 and M^8 (Figure 5-1). The MluI site M^7 maps to a 400bp SstII fragment (Figure 5-5) which suggests it to be a CpG rich area. This 400bp SstII fragment was found to be unique sequence DNA and was used as a probe to determine whether it represents a CpG island. Hybridization of pS400 to SstII digested genomic DNA revealed a 400bp fragment both for normal and DM patient leucocyte DNA (Figure 5-12, panel A). The two SstII sites around MluI site M^7 likely reside in a CpG island (Figure 5-5, marked with a shaded diamond).

To map the adjacent CpG island, distal to the MluI M^7 site and proximal to the MluI M^8 site, Bam HI and Eco RI subclones of cosmids f18894, Yc100263 and f16625 were

Figure 5-12: Southern blot analysis of CpG islands flanking the DM locus. Normal (lanes 1, 3, 4 and 6) and DM patient (lanes 2 and 5) genomic DNA digested with SstII was probed with pS400 (A) and pBP600 (B). The expected restriction fragment size was identified with both probes, 400 bp by pS400 (A, marked by arrow) and 350 bp by pS350 (B, marked by arrow). The small size of the fragments detected by these probes is presumed to be the cause of the faintness in the signal.


A
1 2 3 4 5 6
400 bp

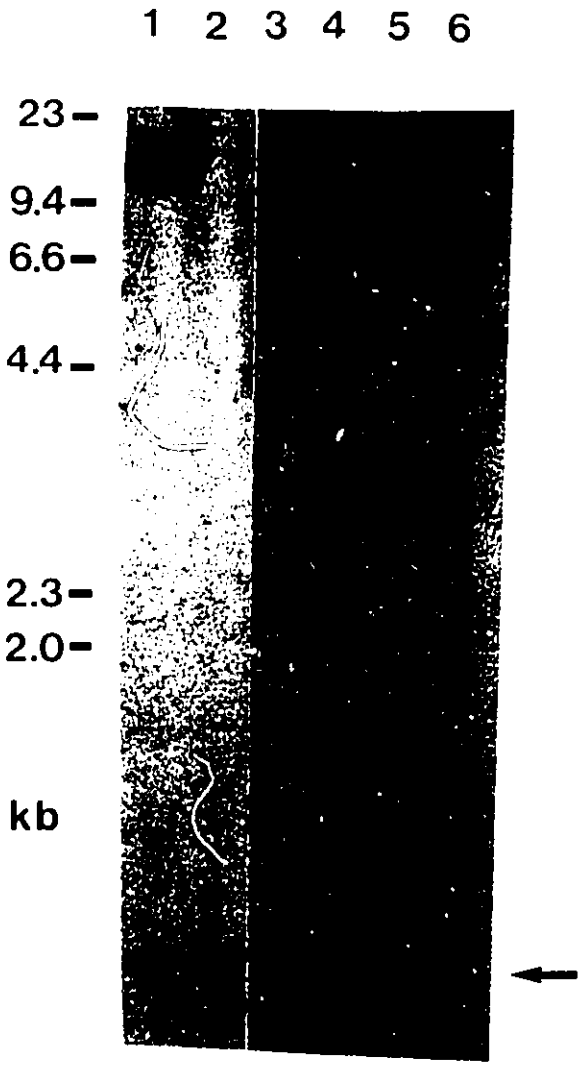

B
1 2 3 4 5 6
350 bp

surveyed for SstII, EagI and BssHII restriction sites (Figure 5-5). After restriction mapping these sites, unique sequence probes were identified that would test for these sites in genomic DNA (Figure 5-5, pBS600, pE400 and pBP600). The five EagI, one BssHII and four of the six SstII sites mapped in these cosmids were not evident in leucocyte DNA from normal and DM patients (Figure 5-5, marked with a shaded circle). Thus no evidence of a CpG island was found with these restriction sites. However, two SstII sites about 350 bp apart found less than 30 kb distal to MluI site M^7 and at least 3 kb proximal to MluI site M^8 were detected by hybridization with pBP600 in both normal and DM patient leucocyte DNA (Figure 5-12, panel B). This region likely represents the next CpG island mapping distal to the CpG island at MluI site M^7 (Figure 5-5, marked with an unshaded diamond).

Recently, a gene expressed in brain and testes was mapped between the putative DM protein kinase gene and the CpG island distal to MluI site M^8 (Jansen et al., 1992b). There is no evidence of another CpG island between these two genes. Thus the closest CpG island to the DM mutation is around MluI site M^7, approximately 3 kb centromeric to the CTG repeat.

Additional evidence that the area around MluI site M^7 is a CpG island was provided by the MspI and HpaII restriction fragments identified by the DNA probe pS400 in the genomic DNA from both normal individuals and DM patients (Figure 5-

Figure 5-13: Southern blot analysis of HpaII and MspI digested genomic DNA with pS400. Normal (lanes 1 and 2) and DM patient (lanes 3-6) genomic DNA was digested with either HpaII (lanes 1, 3 and 5) or MspI (lanes 2, 4 and 6) was probed with pS400. Equivalent sized fragments of about 300 bp were evident for both digests in normal and DM individuals.


1 2 3 4 5 6
23
9.4
6.6
4.4
2.3
2.0
kb

13). A common HpaI and MspI fragment was identified by hybridization with pS400 in DNA from both normal and DM individuals. Since both restriction endonucleases have a common recognition sequence (CCGG) and HpaII is sensitive to CpG methylation while MspI is not, they are often used to confirm methylation patterns. Furthermore, no evidence of an abnormal methylation pattern was apparent in DNA tested from DM patients at the closest CpG island to the DM mutation detectable in leucocyte DNA (Figure 5-13).

5. Discussion

Critical to the mapping of the DM gene in the overlap cloning endeavour is the integrity of the cosmid contig spanning over 500 kb from the pE0.8 locus to the pX75b locus. Previous correlation of rare cutting restriction endonuclease sites in cosmids with CHEF pulsed field analysis of genomic DNA supported the integrity of the 350 kb cosmid contig from the proximal DM region chromosome walk. Similarly, the integrity the cosmid contig from the distal chromosome walk and the cosmids derived from overlapping YAC clones was established by comparing rare cutting restriction endonuclease sites in these cosmids and the CHEF pulsed field data from genomic DNA analysis in this chapter. Consequently, the presence of the DM locus is ensured within this cosmid contig.

Comparison of the genetic and physical maps

The genetic and physical mapping data generated in the cloning of the DM region permitted comparison of genetic distance to physical distance in over a 1.5 Mb region of 19q13.3 spanning the DM locus. The cloning of large cosmid contigs, pulsed field electrophoretic analysis and somatic cell hybrid analysis has furnished marker to marker distance estimation to provide a consensus physical map between loci of the DNA probes used in linkage analysis (Shutler et al., 1992a; Jansen et al., 1992a; Aslanidis et al., 1992; Shutler et al., 1991; Smeets et al., 1990; Korneluk et al., 1989; Stallings et al., 1988; Harley et al., 1991; Shaw et al., 1985; Schonk et al., 1989; Brook et al., 1991; Shaw et al., 1989). Genetic typing data was obtained from a total of 85 families comprising 904 individuals. Linkage analysis was then performed to calculate marker to marker recombination fractions that were converted directly to genetic distance (cM). The two sets of data equated about 1.5 Mb of physical distance to a previously estimated less than 2 cM genetic distance on human chromosome 19q13.3 (Tsilfidis et al., 1991a).

Earlier studies that compared genetic maps to physical maps over large regions of several Mb suggested a variable relationship between marker to marker genetic distance and physical distance. Carson and Simpson (1991) compared the genetic and physical maps of chromosome 10 by converting physical localizations to centimorgans (cM) based on estimates for the whole chromosome as opposed to the

whole genome based assumption of 1cM equivalent to 1 Mb (Ott, 1985). They consequently identified a region of high recombination frequency for chromosome 10.

This approach could also be tested for to chromosome 19. It has been estimated that chromosome 19 is about 62 Mb (Mendelsohn et al., 1973; Southern et al., 1982; Darnel et al., 1986) and the genetic linkage map has been estimated to cover a genetic distance of 137 cM in males and 189 cM in females (Nakamura et al., 1988). The marker to marker genetic distance estimates can then be compared to the physical distance estimates using the alternative assumption of 1 cM is the equivalent to 0. 45 Mb in males and 0.32 Mb in females or 0.39 Mb combined for chromosome 19.

The initial analysis of the DM region data did not reveal a statistically significant correlation between physical and genetic distance. Linear regression analysis of the two parameters showed a correlation only if marker intervals with higher lod scores (with corresponding narrow confidence intervals) were considered. Consequently, seven marker pairs were eliminated. Furthermore, those intervals defined by the seven marker pairs do not originate from a common region. Thus, there is no indication of recombination suppression or the presence of a recombination "hot spot" in the DM region to explain the initial lack of correlation observed. Rather, the related inaccuracy of the broad confidence intervals for the seven marker pairs (with lod scores > 22)

resulted in the initial lack of correlation. Calculation of the 1 cM equivalent in Mb using only marker intervals with high lod scores gave a mean of 1.2 Mb with a standard deviation of 0.4 Mb. Although a genetic/physical distance correlation was shown, predictions of physical distance on the basis of genetic distance should be limited to intervals with high lod scores. Two previous studies that compared genetic to physical maps and found poor correlation did not mention the possible influence of lod scores (Higgins et al., 1990, for the retinoblastoma chromosome 13q14 region and Fulton et al., 1989, for the cystic fibrosis chromosome 7q31 region).

There is evidence to suggest that uneven distribution of recombination frequency reported for the chromosome 10p13 to 10q23 region is gender specific (Carson and Simpson, 1991). In contrast to the chromosome 10 study, no sex difference was found in the genetic map of chromosome 22 (Julier et al., 1988). Estimates of gender specific recombination rates for the DM region either revealed roughly equivalent rates or so few recombinant events (less than or equal to 3) for a given interval that it is impossible to derive any significance to the gender differences observed. A possible exception is the pα1.4-CKM interval for which no male recombinants were seen but four female recombinants were seen (Table 5-5). The genetic map by Nakamura et al., (1988) for chromosome 19 reflects however a strong gender difference, 137 cM in males and 189 cM in females.

DM mutation

Two unique sequence DNA probes, pGB2.6 and pGP1.5 subcloned from cosmid Ya100263, along with a mouse brain cDNA probe cloned by collaborators in the Netherlands, pDMR B15, were found to identify two polymorphisms. One of the polymorphisms is an insertion polymorphism of about 1 kb that is evident as an 8.5 or 9.5 kb HindIII fragment and a 9 kb or 10 kb Eco RI fragment in the genomic DNA of normal individuals. The other polymorphism is an expanded allele found exclusively in DM individuals. The degree of allele expansion from one generation to the next correlates with the increase in severity of the disease, genetic anticipation (Aslanidis et al., 1992). In a survey by our laboratory of over 100 normal individuals and over 200 DM individuals the expanded allele was found only in DM individuals. Increased allele sizes in the range of 10.5 to 15 kb were found in nearly 70% of the DM individuals.

Our laboratory and others have shown the allele expansion found exclusively in DM individuals is due to an amplification of a CTG trinucleotide repeat (Mahadevan et al., 1992; Fu et al., 1992; Brooke et al., 1992). The amplification of the CTG repeat gives a molecular basis for genetic anticipation. The unstable CTG repeat sequence in DM chromosomes is found only in the large Eco RI fragment. This sequence was further mapped and shown to be present within a 1.5 kb Bam HI fragment (Figure 5-5, shaded triangle) approximately 3 kb distal to MluI site M^7. The mutation maps to a locus for the 3' untranslated region of a gene that encodes a putative serine-threonine protein kinase (Mahadevan et al., 1992; Fu et al., 1992; Brooke et al., 1992). The

unstable CTG repeat amplification is not unique to DM as Fragile X syndrome (Verkerk et al., 1991; Kremer et al., 1991; Fu et al., 1991) and Kennedy disease (La Spada et al., 1991) also have trinucleotide repeat mutation sites.

CpG islands and methylation

DM and fragile X syndrome seemingly share a common molecular basis, a hereditable unstable trinucleotide repeat (Oberle et al., 1991; Kremer et al., 1991; Fu et al., 1991). Abnormal methylation of a CpG island led to the mapping of the fragile X locus and may be associated with the trinucleotide repeat amplification. This prompted a preliminary search for DM associated abnormal methylation. The DM mutation maps to a 10 kb Eco RI fragment (Aslanidis et al., 1992; Buxton et al., 1992; Harley et al., 1992) situated between MluI sites M^7 and M^8. CHEF pulsed field electrophoresis data indicates that MluI site M^7 is evident in lymphoblast genomic DNA but MluI site M^8 is not in both normal and DM individuals. Double digest experiments with leucocyte DNA confirmed these results.

The MluI site M^7 is about 3 kb distal to the unstable CTG repeat DM mutation. This restriction endonuclease site maps to a 400 bp SstII fragment (Figure 5-5). The SstII fragment is readily detectable in digests of genomic DNA from both DM and normal individuals with the DNA probe pS400 (panel A, Figure 5-12). Both MluI and SstII have the dinucleotide CpG in their recognition sites which reveals two significant features for this 400 bp area which are: 1) identification of the area as CpG rich, 2)

indication of the methylation status of the area because the CpG dinucleotide when methylated inhibits digestion. Areas that are CpG rich and hypomethylated generally occur at the 5' end of housekeeping genes and many tissue specific genes and are called CpG islands (Bird et al., 1987; Razin et al., 1991). Since the DM mutation maps to the 3' untranslated region of a serine-threonine protein kinase encoded at this locus, the island identified by pS400 is most likely associated with an adjacent gene proximal to the DM locus.

A search for a CpG island on the 5' side of the DM gene led to the identification of a 350 bp SstII fragment detected by the DNA probe pBP600, present in both normal and DM individuals, that maps approximately 30 kb distal to the 3' CpG island and about 3 kb proximal to the MluI site M^8. There is another gene at this locus in addition to the protein kinase gene between these two sites (Jansen et al., 1992a). There is no evidence of another CpG island between these two genes. Thus there is either no CpG island for the DM gene or the CpG island is due to the presence of both genes or solely the testes and brain gene. Furthermore, no evidence was found to suggest any DM associated aberrant methylation at these CpG islands for lymphoblast and leucocyte DNA. However a complete study on a possible role for methylation in DM would require that other tissues, in which the expression of the gene is likely most significant such as heart, skeletal muscle and brain, be analyzed. In addition, the fate of the DM region during development could provide vital information.

Chapter 6: General Discussion

Myotonic dystrophy (DM) is the most common inherited neuromuscular disease affecting both adults and children with a global incidence of 1 in 8000 individuals (Harper, 1989). The disease is an autosomal dominant disorder characterized primarily by myotonia and progressive muscle weakness, although central nervous system, cardiovascular and ocular manifestations also frequently occur. These features reflect the multi-systemic nature of this disease. DM shows a marked variability in expression, ranging from a severe congenital form, often fatal in newborn, to an asymptomatic condition associated with normal longevity. Furthermore, the involvement of genetic anticipation, an increase in the severity of the disease with successive generations has been suggested for DM families (Höwler et al., 1990). The relative frequency, multisystem involvement and variable expression are features of DM that are poorly understood at the molecular and biochemical level (Hall, 1990; Koch et al., 1991). The opportunity to study these features would be facilitated by the cloning and subsequent characterization of the DM locus. Since there are no known chromosomal aberrations or obvious biochemical abnormalities associated with the disease, cloning of the DM locus relies upon "positional cloning" strategies, that is, the isolation of the gene based on its chromosomal location.

The objective of this thesis was to use long range cloning methods along with genetic and physical mapping to define the minimal region expected to contain the DM

locus. The strategy shown in Chapter III addressed the initial concern of localizing the DM gene relative to the ERCC1 locus. A polymorphic marker, pE0.8 (D19S115) was found in the ERCC1 region and used for physical and genetic mapping. CHEF pulsed field electrophoretic analysis of human genomic DNA with three probes, pCKM3' (CKM), pERCC1-int (ERCC1) and pE0.8 (D19S115) supported the mapping of ERCC1 to within 300 kb distal to CKM. The mapping of a cosmid containing ERCC1 sequence positioned pE0.8 about 15 kb centromeric from the ERCC1 locus. Genetic mapping revealed close linkage of pE0.8 to DM (θ = 0.01, Z_{max} = 19.29) Furthermore, a key recombinant event was identified in a Dutch DM family wherein pE0.8 cosegregates with the more proximal markers (D19S37, APOC2) and recombines with DM. This recombination maps the disease gene locus distal to pE0.8.

A second key recombinant event was identified in a French Canadian DM family, previously described by Tsilfidis et al., (1991a) that maps the DM locus proximal to p134c (D19S51). This probe was isolated from a cosmid library of a somatic cell hybrid containing a small portion of chromosome 19 including the DM region (Johnson et al., 1990). These two recombinant events define a less than 2 cM region expected to contain the DM locus from pE0.8 to p134c (Tsilfidis et al., 1991a). Based on the tight linkage of pE0.8 (D19S115) to DM and the recombinant event which maps the DM locus distal to the ERCC1 locus it was reasoned that pE0.8 would be an

appropriate start point for a chromosome walk from D19S115 towards the DM locus in a telomeric direction.

By using primarily ALU PCR generated probes, a cosmid contig of approximately 350 kb was constructed from a chromosome 19 cosmid library as detailed in Chapter IV. The results of the chromosome walk has enabled construction of a high resolution genetic map in conjunction with a complementary long range physical map that was used to define the DM region. The proximal crossover event identified in a Dutch DM family (Chapter III) was mapped to a 60 kb region between two probes centrally located in the chromosome walk, pKEX0.8 (D19S118) and pKBE0.8 (D19S119), establishing pKEX0.8 as the closest proximal marker to DM. This result also eliminates over 200 kb of sequence that would have otherwise been screened for DM candidate genes. At the distal end of the chromosome walk the contig could not be extended further in a telomeric direction due to an absence of clones with adjacent sequence in the chromosome 19 libraries.

To resolve the cloning limitations of the chromosome 19 cosmid library, a chromosome walk was initiated from a newly cloned marker, pX75b (D19S112), that mapped between DM and p134c (D19S51). The work was done by collaboration with Gert Jansen and Dr. Be' Wieringa from the University of Nijmegen in the Netherlands. Utilizing the method of chromosome walking by ALU PCR developed during the proximal chromosome walk endeavour, a bidirectional chromosome walk established a

200 kb contig before no more cosmids could be found at the centromeric end. Overlap of the proximal and distal cosmid contigs was accomplished by using two STS's, one from the proximal side at pKBE0.8 (D19S119) and one from the distal side at pX75b (D19S112), to screen a YAC library for YAC clones spanning the two cosmid contigs. This work was done by collaboration with Dr. Peter de Jong at the Human Genome Centre in the Lawrence Livermore National Laboratory in Livermore, California.

A third STS from an ALU PCR generated vector end product of a pX75b STS selected YAC clone was used to screen the "St. Louis" YAC library (Burke et al., 1987) in order to find intact YAC sequence that spanned the proximal and distal DM cosmid contigs. Cosmid libraries were made from the YAC clones for this region. Cosmid clones that connected the proximal and distal DM region cosmid contigs were then selected.

A consensus physical map of the DM cosmid contig consisting of the proximal DM region chromosome walk, the distal DM region chromosome walk and overlapping YAC derived cosmids was presented in Chapter V. The YAC derived and distal chromosome walk cosmids were screened for single copy sequence. CHEF pulsed field electrophoretic analysis of human genomic DNA with the newly generated probes from the distal DM region was undertaken to correlate rare cutting endonuclease sites in the Yac derived and distal cosmid contig to confirm the integrity of the long range

cloning effort. Establishment of the cosmid contig integrity ensured the presence of the DM locus within this cloned expanse of DNA.

The region of human chromosome 19q13.3 is well characterized due to the positional cloning strategy to identify the DM gene. This provided the opportunity for comparison of genetic distance to physical distance estimates between markers on a consensus physical map spanning 1.5 Mb. Linear regression analysis of marker to marker genetic distances to physical distances showed a linear correlation between these two parameters only when intervals with high lod scores (over 22) were considered. There was no evidence for recombination suppression or for a recombination "hot spot" in the DM region. On average 1 cM was found to be the equivalent to 1.2 Mb +/- 0.4 Mb for this region.

Several unique sequence DNA probes were cloned and used for CHEF pulsed field electrophoretic analysis of the region. This was done for three main purposes. The first purpose was to confirm the integrity of the YAC derived cosmid contig through the chromosome 19 cosmid library cloning gap and the distal DM region cosmid contig. The second purpose was to extend the physical map from the proximal DM region cosmid contig as previously discussed. A third purpose was to search for polymorphic markers to use in the high resolution genetic mapping of the region expected to contain the DM gene. Two of the newly cloned probes, pGB2.6 and pGP1.5, found to be within 10 kb of MluI site M[7], also mapped to a 9 kb Eco RI

fragment in YAC derived cosmid Yc100263 and a 10 kb Eco RI fragment in the chromosome 19 derived cosmid f18894. These two Eco RI fragments were identified as the two alleles of an insertion polymorphism evident in the DNA from normal individuals. More significantly, these two probes identified an expansion of the 10 kb Eco RI allele in genomic DNA exclusively from DM affected individuals.

Extensive linkage analysis of the expanded allele with DM was conducted in our laboratory on DM families. In 70% of approximately 200 DM patients increased allele sizes ranging from 10.5 to 15 kb was revealed (Aslanidis et al., 1992). Evidence from three generation DM families show that the increase in expansion of the allele size correlated with an increase in the severity of the disease (genetic anticipation).

The expansion in the unstable alleles was found to be due to amplification of a CTG trinucleotide repeat. The smallest subclone from cosmid Yc100263 containing the repeat was a 1.5 kb Bam HI fragment cloned into pUC18. Primers flanking the CTG repeat were designed from the sequence of this subclone. PCR amplification of genomic DNA using this primer set was done with one of the primers radioactively labelled with γ-[^{32}P] adenosine triphosphate. The PCR products along with a sequencing ladder were run on an 8% polyacrylamide sequencing gel. The dried gels were autoradiographed to reveal the size of the repeat. The size of repeats in 124 normal chromosomes analyzed ranged from 5 to 30 with repeat numbers of 5 and 13 the most common. The repeat numbers found in DM chromosomes were greater

than 50, smeared and often too large to resolve. Southern blot analysis of PCR amplified genomic DNA from DM individuals with a radioactively labelled $(CTG)_{10}$ oligonucleotide probe showed a distinct smearing of the hybridization signal above the normal alleles likely due to the heterogeneity of the expanded allele. In contrast to the DM unstable CTG repeat, the CTG repeat in normal individuals was found to be both meiotically and mitotically stable despite being very polymorphic.

In all populations studied thus far, the amplification of the CTG repeat seems to be the sole mutation responsible for DM (Mahadevan et al., 1992; Brook et al., 1992; Fu et al., 1992; Jansen et al., 1992b). The only exception was for two out of 98 DM families that did not show allelic expansion using both methods of detection (Mahadevan et al., 1992). Either the two negative families have other mutations at or near this locus or they have another clinically similar disorder. Amplification of the trinucleotide repeat is exclusive to DM patients.

The CTG trinucleotide repeat amplification has been proposed as the molecular basis for genetic anticipation in DM families. There is a correlation between the clinical severity of DM with the increase in amplification of the CTG repeat (Aslanidis et al., 1992; Mahadevan et al., 1992; Buxton et al., 1992; Harley et al., 1992; Brook et al., 1992; Fu et al., 1992). Furthermore, a study in our laboratory of the CTG repeat status in 272 DM patients revealed that infants with severe congenital DM, as well as their mothers, have on average greater amplification of the trinucleotide repeat than

observed in the non-congenital population (Tsilfidis et al., 1992a). Thus, the trinucleotide amplification in DM is likely the molecular basis for genetic anticipation.

Heritable unstable DNA sequences are not unique to DM. Trinucleotide amplifications have also been shown to be the molecular defect in the fragile X syndrome (Verkerk et al., 1991; Kremer et al., 1991; Fu et al., 1991) and Kennedy disease which is also known as spinal and bulbar muscular atrophy (La Spada et al., 1991). In fragile X, allelic expansion is due to an amplification of a trinucleotide repeat, CGG, mapping to the 5' untranslated region of a gene identified as FMR-1. Amplification of a CAG trinucleotide repeat that encodes a polyglutamine in the androgen receptor gene was found to be responsible for Kennedy disease. Therefore, there are three examples to date of hereditable unstable DNA sequences which result in genetic disease. However, unstable DNA sequence manifesting in a heritable disease could be more wide spread.

The CTG trinucleotide repeat maps to the 3' untranslated region of a gene which based on sequence homology comparisons encodes a putative serine-threonine cAMP dependent protein kinase (Mahadevan et al., 1992; Fu et al., 1992; Brook et al., 1992). By Northern blot analysis the gene was found to have the highest expression in heart and to a lesser extent in skeletal muscle. There is also some expression evident in brain as detected by Northern analysis and lymphocytes as seen by reverse transcriptase PCR. The full length transcript is between 3 to 3.3 kb with evidence of

tissue specific alternative splicing (Jansen et al., 1992b). The gene is oriented in a 5' to 3' direction from telomere to centromere. Another gene located just upstream is also oriented in the same transcriptional direction. This gene was found to be highly expressed in brain and testes (Jansen et al., 1992b). The coding region for both of these genes is flanked by MluI sites M^{7-8} (Chapter V, Figure 5-1).

The MluI site M^7 is about 3 kb distal to the CTG repeat. Evidence was presented in Chapter V that this CpG rich area constituted a CpG island. Since the DM associated VLP has been mapped to the 3' untranslated region of the putative serine-threonine protein kinase and CpG islands are generally found at the 5' end of all housekeeping genes and many tissue specific genes (Stein et al., 1983; Bird et al., 1986; Razin and Cedar, 1991), this CpG island is most likely associated with an adjacent gene proximal to this locus. The only candidate CpG island distal to the CTG repeat was found to be about 3 kb proximal to MluI site M^8 and is likely associated with the 5' end of the gene expressed in brain and testes upstream from the protein kinase. There are several interpretations for this CpG island mapping. One explanation is that there simply is no CpG island associated with the protein kinase gene. A second possible explanation is that the island is associated with both genes. Alternatively, another CpG rich region, possibly around the three EagI sites 15 kb distal to the CTG repeat, is hypomethylated but only in tissues that show high expression of the protein kinase such as heart and muscle. Therefore an extensive study using DNA from these different tissues should be conducted.

The similarity between allele expansion of DM and fragile X syndrome along with the involvement of abnormal methylation at the CpG island associated with fragile X syndrome prompted a search for abnormal methylation associated with DM. A comparison in lymphoblasts between normal and DM individuals of methylation status for CpG islands located 5' and 3' to the protein kinase using methylation sensitive endonucleases did not reveal any differences in methylation patterns between affected and unaffected individuals. However a complete study on a possible role for methylation in DM would require that other tissues, in which the expression of the gene is likely most significant (eg. heart and skeletal muscle) be analyzed.

Concluding Remarks

By using the basic strategy of positional cloning the DM locus has been identified. Over 350 kb from the proximal DM region was cloned in this study and another 250 kb from the distal DM region was cloned by collaborative studies. The resultant consensus physical map of the DM region spanned 1.5 Mb of human chromosome 19q13.3. Unique sequence probes subcloned in this study and their use in genetic linkage analysis of the DM families served by the Childrens Hospital of Eastern Ontario resulted in the identification of a heritable unstable CTG repeat sequence found exclusively in DM patients.

The increasing of the unstable CTG trinucleotide repeat amplification correlated with an increasing severity of the disease found in DM families (genetic anticipation) is

possibly the strongest argument that this is the molecular defect causing DM. The finding of allelic expansion based on the amplification of a trinucleotide repeat in fragile X syndrome and spinal and bulbar muscular atrophy makes this argument even more compelling. The mapping of the CTG repeat to the 3' untranslated region of a putative serine-threonine protein kinase expressed in a variety of tissues, including heart and skeletal muscle, may account for at least in part of the multisystemic nature of the disease.

While it is evident that the DM locus has been found there are many questions yet to be answered. These include the reason for the instability of the CTG repeat in DM patients, the consequences of the trinucleotide repeat amplification to the protein kinase and other genes in the region and the biological process of the disease. These and other pertinent questions may now by studied as a result of the discovery of the genetic defect responsible for myotonic muscular dystrophy.

References

Appleby, V.L., Coleman, R.T. and Frossard, P.M. (1986) Linkage disequilibrium at the human apopolipoprotein CII gene locus. Am. J. Hum. Genet. 39 (suppl): A145.

Aslanidis, C., Jansen, G., Amemiya, C., Shutler, G., Tsilfidis, C., Mahadavan, 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.

Bachinski, L.L., Liu, P. and Siciliano, M.J. (1989) A gene order for 22 markers on human chromosome 19. Cytogen. Cell Genet. 51: 954.

Baily, J., MacKenzie, A.E., Leblond, S. and Korneluk, R.G. (1991) Assessment of a creatine kinase isoform M defect as a cause of myotonic dystrophy and the characterization of two novel CKMM polymorphisms. Hum. Genet. 86: 457-462.

Benham, F., Hart, K., Crolla, J., Bobrow, M., Francavilla, M. and Goodfellow, P. (1989) A method for generating hybrids containing nonselected fragments of human chromosomes. Genomics 4: 509-517.

Bird, A.P. (1986) CpG rich islands and the function of DNA methylation. Nature 321: 503-504.

Bird, A.P. (1987) CpG islands as gene markers in the vertebrate nucleus. Trends in Genet. 3: 342-347.

Birnboim, H.C. and Dolly, J. (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. 7: 1513-1523.

Botstein, D., White, R.L., Skolnick, M. and Davis, R.W. (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am. J. Hum. Genet. 32: 314-331.

Brook, J.D., Shaw, D.J., Thomas, N.S.T., Meredith, A.L., Cowell, J. and Harper, P.S. (1986) Mapping genetic markers on human chromosome 19 using sub-chromosomal fragments in somatic cell hybrids. Cytogenet. Cell Genet. 41: 30-7.

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

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., Korneluk, R.G., Coerwinkel-Driessen, M., MacKenzie, A., Smeets, H., Lambermon, H.M.M., van Oost, B.A., Wieringa, B. and Ropers, H.H. (1989a) Myotonic dystrophy is closely linked to the gene for muscle-type creatine kinase (CKMM). Hum.Genet. 81: 308-310.

Brunner, J.D., Smeets, H., Lambermon, H.M.M., Coerwinkel-Driessen, M., van Oost, B.A., Wieringa, B. and Ropers, H.H. (1989b) A multipoint linkage map around the locus for myotonic dystrophy on chromosome 19. Genomics 5: 589-595.

Buckle, V.J. and Craig, I.W. (1986) in Human genetic diseases a practical approach, Davies, K.E. (ed) IRL Press, Oxford, Washington D.C., 85-100.

Burke, D.T, Carle, F.G. and Olson, M.V. (1987) Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. Science 236: 806-812.

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.

Carson, N. and Simpson, N. (1991) A physical map of human chromosome 10 and a comparison with an existing genetic map. Genomics 11: 379-388.

Chu, G., Vollrath, D. and Davis, R.W. (1986) Separation of large DNA molecules by contour-clamped homogeneous electric fields. Science 234: 1582-1585.

Clark, J.L. (1988) Novel non-templated nucleotide addition reactions catalyzed by prokaryotic and eucaryotic DNA polymerases. Nucleic Acids Res. 16. 9677-9686.

Clark, S.M., Lai, E., Birren, B.W. and Hood, L. (1988) A novel instrument for separating large DNA molecules with pulsed homogeneous electric fields. Science 241: 1203-1205.

Coewinkel-Driessen, M., Schepens, J., van Zanvaart, P., van Oost, B., Mariman, E. and Wieringa, B. (1988) NcoI RFLP at the creatine kinase-muscle type gene locus (CKMM, chromosome 19). Nucleic Acids Res. 16: 8743.

Collins, F.S., Drumm, M.L. Cole, J.L., Lockwood, W.K., Vande Wonde, G.F. and Innuzzi, M.C. (1987) Construction of a general human chromosome jumping library, with application to Cystic Fibrosis. Science 235: 1046-1049.

Cook, P.J.L., Povey, S. and Robson, E.B. (1972) Linkage studies on peptidases A, B, C, and D in man. Ann. Hum. Genet. 36: 89-97.

Coulson, A, Waterston, R, Kiff, J., Sulston, J. and Kohara, Y. (1988) Genome linking with yeast artificial chromosomes. Nature 335: 184-186.

Coussens, L., Parker, P.J., Rhee, L., Yang-Feng, T.L., Chen, E., Waterfield, M.D., Franke, U. and Ullrich, A. (1986) Multiple, distinct forms of bovine and human protein kinase c suggest diversity in cellular signalling pathways. Science 233: 859-866.

Coerwinkel-Driessen, M., Schepens, J., Van Zandvoort, P., Van Oost, B., Mariman, E. and Wieringa, B. (1988) NcoI RFLP at the creatine kinase-muscle type gene locus (CKMM, chromosome 19). Nucleic Acids Res. 16: 8743.

Cretonne, A. Hoeijmakers, J.H.L., van Duin, M., de Wit, D., Odijk, H., Pastink, A., Wood, R.D. and Bootsma, D. (1984) Molecular cloning of a human DNA repair gene. Nature 310: 425-429.

Davies, K.E., Jackson, J., Williamson, R., Harper, P.S., Ball, S., Sarfarazi, M., Meredith, L. and Fey, G. (1983) Linkage analysis of myotonic dystrophy and sequences on chromosome 19 using a cloned complement 3 gene probe. J. Med. Genet. 21: 259-263.

de Jong, P.J., Tokobata, K., Chen, C., Lohman, F., Pederson, L., McNinch, J. and Van Dilla, M. (1989) Human chromosome-specific partial digest libraries in lambda and cosmid vectors. Cytogenet. Cell Genet. 51: 985.

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.

Evans, G.A. and Wahl, G.M. (1987) Cosmid vectors for genomic walking and rapid restriction mapping. In Methods in Enzymology, Academic Press 152: 604-615.

Feinberg, A.P. and Vogelstein, B. (1984) Addendum to "A technique for radiolabeling DNA restriction endonuclease fragments to a high specific activity." Anal. Biochem. 137: 266-267.

Finney, M. (1989) In. "Current protocols in Molecular Biology" Suppl. 4. Edited By: F.M. Ausubel, R. Brent, R.E. Kingston, D.D. Moore, J.G. Seidman, J.A. Smith and K. Struhl. Wiley, New York. 2.5.9.-2.5.15.

Frossard, P.M., Coleman, R.T. and Assman, G. (1986) NcoI RFLP at the human apolipoprotein CII locus. Nucleic Acids Res. 14: 5120.

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.

Fulton, T.R., Bowcock, A.M., Smith, D.R., Daneshvar, L., Green, P., Cavalli-Sforza, L.L. and Donis-Keller, H. (1989) A 12 megabase restriction map at the cystic fibrosis locus. Nucleic Acids Res. 17, 271-284.

Fuscoe, J.C., Clark, L.M., and Van Dilla (1986) Construction of fifteen chromosome-specific libraries from flow-purified chromosomes. Cytogenet. Cell Genet. 43; 79-86.

Gedde-Dahi, T. Jr., Olaisen, B., Teisberg, P. Wihelmy, M.C., Mevag, B. and Helland, R. (1984) The locus for apolipoprotein E (apoe) is close to the lutheran (Lu) blood group locus on chromosome 19. Hum. Genet. 67: 178-182.

Gill, P., Jeffreys, A.J. and Werrett, D.J. (1985) Forensic applications of DNA fingerprints. Nature 318: 577-579.

Greenwalt, T.J. (1961) Conformation of linkage between the Lutheran and Secretor genes. Am. J. Hum. Genet. 13: 69-88.

Griggs, R.C., Wood, D.S., the working group on the molecular defect in myotonic dystrophy. Neurology **39**: 420-421.

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

Hamada, H., Petrino, M.G. and Kakunaga (1982) A novel repeated element with Z DNA forming potential is widely found in evolutionary diverse eukaryotic genomes. Proc. Natl. Acad. Sci. USA 79: 6465-6469.

Hanahan, D. (1985) Technique for transforming E-coli. In DNA Cloning. Vol. 1. 109-135.

Harley, H.G., Walsh, K.V., Rundle, S., Brook, J.D., Sarfarazi, M., Koch, M.C., Floyd, J.L., Harper, P.S. and Shaw, D.J. (1991) Localization of the myotonic dystrophy locus to 19q13.2-19q13.3 and its relationship to twelve polymorphic loci on 19q. Hum.Gent. 87: 73-80.

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

Harper, P.S., Rivas, M.L., Bias, W.B., Hutchinson, J.R., Dyken, P.R. and McKusick, V.A. (1972) Genetic linkage confirmed between the locus for myotonic dystrophy and the ABH-secretion and Lutheran blood group loci. Am. J. Hum. Genet. 24: 310-316.

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

Higgins, M.J., Turmel, C., Noolandi, Neumann, P.E. and Laland, M. (1990) Construction of the physical map for three loci in chromosome band 13q14: Comparison to the genetic map. Proc. Nat. Acad. Sci. USA. 87: 3415-3419.

Hill, W.G. and Robertson, A. (1968) Linkage disequilibrium in finite populations. Theor. Appl. Genet. 38: 226-231.

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

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

Hulsebos, T., Wieringa, B., Hochstenbach, R., Smeets, D., Schepens, J., Oerlemans, F., 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.

Humphries, S.E., Jowett, N.I., Williams, L., Rees, A., Vella, M., Kessling, A., Myklebost, O., Lyndon, A., Seed, M., Galton, D.J. and Williamson, R. (1983) A DNA polymorphism adjacent to the apolipoprotein CII gene. Mol. Biol. Med. 1: 463-471.

Inglehearn, C.F. and Cooke, H.J. (1990) A VNTR immediately adjacent to the human pseudoautosomal telomere. Nucleic Acids Res. 18: 471-476.

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., Hermens, 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., Iles, D., Coerwinkel, M., Hofker, M., Carrano, A., de Jong, P., Korneluk, R.G. and Wieringa, B. (1992) Characterization of the myotonic dystrophy (DM) region predicts multiple protein kinase isoform-encoding mRNAs that differ at their 3' ends. Nature Genetics 1: 261-266.

Johnson, K., Shelbourne, P., Davies, J., Buxton, J., Nimmo, E., Siciliano, M.J., Bachinski, L.L., Anvert, M., Harley, H., Rundle, S., Miki, T., Brunner, H. and Williamson (1990) A new polymorphic probe which defines the region of chromosome 19 containing the myotonic dystrophy locus. Am. J. Hum. Genet. 46: 1073-1081.

Julier, C., Lathrop, G.M., Abderrezak, R., Szanjnert, M.-F., Lalouel, J.-M. and Kaplan, J.-C. (1988) A linkage and physical map of chromosome 22 and some applications to gene mapping. Am. J. Hum. Genet. 42: 297-308.

Karya, Y., Kato, K., Hayashizaka, Y., Himeno, S., Tarui, S. and Matsubura, K. (19870 revision of consensus sequence of human Alu repeats- A review. Gene 53: 1-10.

Kerem, B-S., Rommens, J.M., Buchanan, J.A., Markiewicz, 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.

Kidd, K.K., Bowcock, A.M., Schmidtke, J., Track, R.K., Ricciuti, F., Hutschings, G., Bale, A., Pearson, P. and Willard, H.F. (1989) Report of the DNA committee and catalogs of cloned and mapped genes and DNA polymorphisms. Cytogenet. Cell Genet. 51: 622-947.

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.

Korneluk, R.G., Quan, F. and Gravel, R.A. (1985) Rapid and reliable sequencing of double-stranded DNA. Gene 40: 317-323.

Korneluk, R.G., MacLeod, H.L., Leblond, S.C., Monteith, N.L., Boralle, F.E. and Hunter, A.G.W. (1987) Ava II RFLP at the human apolipoprotein CII (APOCII) gene locus. Nucleic Acids Res. 15: 6769.

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

Korneluk, R.G., MacKenzie, A.E., Nakamura, Y., Dube, I., Jacob, P. and Hunter, A.G.W. (1989b) 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., 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). Nucleic Acids Res. 19: 1157.

Kremer, E.J., Pritchars, 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.

Laberge, C., Gaudet, A., Morrissette, J., Moorjani, S. and Thiault, M-J. (1985) Linkage of myotonic dystrophy and APOE in a French Canadian isolate. Cytogenet. Cell Genet. 40: 675.

La Spada, A., Wilson, E., Lubahn, Harding, A. and Fischbeck, K. (1991) Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy. Nature 352: 77-79.

Lathrop, G.M., Lalond, J.M., Julier, C. and Ott, J. (1985) Multilocus analysis in humans: Detection of linkage and estimation of recombination. Am. J. Hum. Genet. 37: 428-498.

Lerman, L.S., Fischer, S.G. and Lumelsky, N. (1983) Inferring variations in genotype from properties of DNA. In Recombinant DNA and medical genetics. Edited by A. Messer and I. Porter. New York. pp. 157-187.

Lindsay, S. and Bird, A.P. (1987) Use of restriction enzymes to detect potential gene sequences in mammalian DNA. Nature 37: 336-338.

Litt, M. and Lutty, J.A. (1989) A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. Am. J. Hum. Genet. 44: 397-401.

Lusis, A.J., Heinzmann, C., Sparkes, R.S., Geller, 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.

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

Madisen, L., Hoar, D., 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. Amer. J. Med. Genet. 27: 379-380.

Mahadevan, M., Tsilfidis, C., Sabourin, L., Shutler, G.G., Ameniya, C., Jansen, G., Neville, C., Narang, M., Barcelo, 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.

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

Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) In: Molecular cloning: A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, USA.

McAlpine, P.J., Mohandas, T., Ray, M., Wang, H. and Hammerton, J.L. (1976) Assignment of the peptidase D gene locus (PEPD) to chromosome in man. Cytogen. Cell Genet. 16: 204-205.

McKeithan, T.W., Rowley, J.D., Shows, T.B. and Diaz, M.O. (1987) Cloning of the chromosome translocation breakpoint of the t (14, 19) in chronic lymphocytic leukaemia. Proc. Nat. Acad. Sci. USA 84: 9257-9260.

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

Mohr, J. (1954) A study of linkage in man. Copenhagen: Munksgaard.

Mohrenweiser, H.W., Carrano, A.V., Fertitta, A., Perry, B., Thompson, L.H., Tucker, J.D. and Weber, C.A. (1989) Refined mapping of the three DNA repair genes, ERCC1, ERCC2, XRCC1, on human chromosome 19. Cytogenet. Cell Genet. 52: 11-14.

Myklebost, O. and Rogne, S. (1988) A physical map of the apolipoprotein gene cluster on human chromosome 19. Hum. Genet. 78: 244-247.

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

Nelson, M. and McClelland, M. (1987) The effect of site-specific methylation on restriction-modification enzymes. Nucleic Acids Res. 15: r219-r230.

Nigro, J.M., Schweinfest, C.W., Rajkovic, A., Pavlovic, J., Jamal, S., Dottin, R.P., Hart, J.T., Kamarck, M.E., Rae, P.M.M., Carty, M.D. and Martin-DeLeon, P. (1987) cDNA cloning and mapping of the human creatine kinase M gene to 19q13. Am. J. Hum. Genet. 40: 115-125.

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

Olaisen, B., Teisberg, P. and gedde-Dahl, T. Jr. (1982) The locus for apolopoprotein E (apoE) is linked to the complement component C3 (C3) locus on chromosome 19 in man. Hum. Genet. 26: 233-236.

Olson, M., Hood, L., Cantor, C. and Botstein, D. (1989) A common language for physical mapping of the human genome. Science 245: 1434-1435.

Orita, M., Iwahana, H., Kanazawa, H., Hayashi and Sekiya, T. (1989) Detection of polymorphisims of human DNA by gel electrophoresis as single-strand conformation polymorphisms. Proc. Nat. Acad. Sci. Usa 86: 2766-2770.

Ott, J. (1985) The analysis of human genetic linkage. Baltimore, MD. The Johns Hopkins University Press.

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

Poustka, A., Pohl, T.M., Barlow, D.P., Frischauf, A-M. and Lerach, H. (1987) Construction and use of human chromosome jumping libraries from NotI digested DNA. Nature 325: 353-355.

Razin, A. and Cedar, H. (1991) DNA methylation and gene expression. Microbiol. Rev. 55: 451-458.

Renwick, J.H., Bundey, S.E., Fergeson,-Smith, M.A. and Izatt, M.M. (1971) Confirmation of the linkage for loci for myotonic dystrophy and A,B,H secretion. J. Med. Genet. 8: 407-416.

Riordan, J.R., Rommens, J.M., Kerem, B-S, Alon, N., Rozmahel, R., Grezelczak, Zielenski, J., Lok, S., Plausic, 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.

Robson, E.B. (1988) The human gene map. Phil. Trans. R. Soc. Lond. B. 319: 229-237.

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 regions of human autosomes. Genomics 3: 352-360.

Saiki, R.K., Bugawan, T.L., Horn, G.T., Mullis, K.B. and Erlich, H.A. (1986) Analysis of enzymatically amplified ß-globin and HLA-DQα DNA with allele-specific oligonucleotide probes. Nature 324: 163-166.

Saki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B. and Erlich, H.A. (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239: 487-491.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) In: Molecular cloning, A laboratory manual. second edition. Cold Spring Harbour Laboratory Press, N.Y.

Sanger, F., Nicklen, S. and Coulson, A.R. (1977) DNA sequencing with chain terminating inhibitors. Proc. Nat. Acad. Sci. USA 74: 5463-6467.

Schlessinger, D. (1990) Yeast artifical chromosomes: tools for mapping and analysis of complex genomes. Trends Genet. 6: 253-258.

Schonk, D., Coerwinkel-Dreiessen, M., Dalan van, I., Oerlemans, F., Smeets, B., Schepens, J., Hulsebos, T., Cockburn, D., Boyd, Y., Davis, M., Rettig, W., Shaw, D., Roses, A., Ropers, H. and Wieringa, B. (1989) Definition of subchromosomal intervals around the myotonic dystrophy region at 19q. Genomics 4: 384-396.

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.

Shaw, D.J., Harley, H.G., Brook, J.D. and McKeithan, T.W. (1989) Long-range restriction map of a region of human chromosome 19 containing the apolipoprotein genes, a CLL-associated translocation breakpoint, and two polymorphic MluI sites. Hum. Genet. 83: 71-74.

Shutler, G.G., Monteith, N. and Korneluk, R.G. (1988) Influences on RFLP's generated by alteration of prehybridization and hybridization conditions. Genome 30 (Suppl. 1): 236.

Shutler, G.G., MacKenzie, A.E., Brunner, H., Wieringa, B., de Jong, P., Lohman, F.P., Leblond, S., Bailly, J. and Korneluk, R.G. (1991a) Physical and genetic mapping of a novel chromosome 19 ERCC1 marker showing close linkage with myotonic dystrophy. Genomics 9: 500-504.

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

Shutler, G., Tsilfidis, C., Leblond, S. and Korneluk, R.G. (1991c) RFLP identified by the probe pKE0.6 (D19S117) at human chromosome 19q13.3. Nucleic Acids Res. 19: 1158.

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. (1992a) Physical mapping and cloning of the proximal segment of the myotonic dystrophy gene region. Genomics 13: 518-525.

Smeets, B., Poddighe, J., Brunner, H., Ropers, H.H. and Wieringa, B. (1988) Tight linkage between myotonic dystrophy and apolipoprotein E genes revealed with allele-specific oligonucleotides. Hum. Genet. 80: 49-52.

Smeets, H.J.M., Brunner, H.G., Ropers, H.H. and Wieringa, B. (1989) Use of variable sequence motifs as genetic markers: application to study of myotonic dystrophy. Hum. Genet. 83: 245-251.

Smeets, H., Bachinski, L., Coerwinkel, M., Schepens, J., Hoeijmakers, J., van Duin, M., Grzeschik, K.H., Weber, C.A., de Jong, P., Siciliano, M.J. and Wieringa, B. (1990) A long range map of the human chromosome 19q13 region: close physical linkage between CKMM and the ERCC1 and ERCC2 genes. Am. J. Hum. Genet. 46: 492-501.

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

Smit, M., Kooij-Meijs van der E., Frants, R.R., Havekes, L. and Klasen, E. (1988) Apolipoprotein gene cluster on chromosome 19. Definite localization of the Apoc2 gene and the polymorphic HpaI site associated with type II hyperlipoproteinemia. Hum. Genet. 78: 90-93.

Southern, E.M., Anand, R. and Fletcher, D.S. (1987) A model for the separation of large DNA molecules by crossed field gel electrophoresis. Nucleic Acids Res. 15: 5925-5943.

Spoerel, N.A. and Kafatos, F.C. (1987) Isolation of full-length genes: Walking the chromosome. In Methods in Enzymology, Academic Press 152: 598-603.

Stallings, R.L., Olson, E., Strauss, A.W., Thompson, L.H. and Siciliano, M.J. (1988) Human creatine kinase genes on chromosome 15 and 19, and the proximity of the gene for the muscle form to the genes for apolipoprotein C2 and excision repair. Am. J. Hum. Genet. 43: 144-151.

Stallings, R.L., Ford, A.F., Nelson, D., Torney, D.C., Hildebrand, C.E. and Moyzis, R.K. (1991) Evolution and distribution of (GT)n repetitive sequences in mammalian genomes. Genomics 10: 807-815.

Stein, R., Sciaky-Gallili, N., Razin, A. and Cedar, H. (1983) Pattern of methylation of two genes coding for housekeeping functions. Proc. Nat. Acad. Sci. USA 79: 61-65.

Swartz, D.C. and Cantor, C.R. (1984) Separation of yeast chromosome size DNAs by pulsed field gel electrophoresis. Cell 37: 67-75.

Tautz, D. (1989) Hypervariability of simple sequences as a general source of polymorphic DNA markers. Nucleic Acids Res. 17: 6463-6471.

Thein, S.L. and Wallace, R.B. (1986) In: Human genetic diseases a practical approach, Edited by K.E. Davies, IRL Press, Oxford and Washington D.C.

Tsilfidis, C., MacKenzie, A.E., Shutler, G., Leblond, S., Bailly, J., Johnson, K., Shelborn, P., Williamson, R., Siegel-Bartelt, J. and Korneluk, R.G. (1991a) 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., Shutler, G., Leblond, S. and Korneluk, R.G. (1991b) An SstI RFLP detected by the probe pKE2.1 (D19S116) localized to human chromosome 19q13.3. Nucleic acids Res. 9: 1158.

Tsilfidis, C., Shutler, G., Mahadevan, M. and Korneluk, R.G. (1991c) A frequent HincII polymorphism identified by human chromosome 19 probe pKEX0.8 (D19S118). Nucleic Acids Res. 9: 1157.

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

Tsilfidis, C., MacKenzie, A., Shutler, G., Neville, C., Mahadevan, M., Wood, S., Andermann, E., Korneluk, R. (1992b). Haplotype analysis and linkage disequilibrium patterns in Myotonic Dystrophy (DM). Am. J. Hum. Gen. (submitted).

Van Dilla, M.A., Deavan, L.L., Albright, K.L., Allen, N.A., Aubuchon, M.R., Barthholdi, M.F., Bromne, N.C., Campbell, E.W., Carrano, A.V., Clark, L.M., Cram, L.S., Crawford, B.D., Fuscoe, J.C., Gray, J.W., Hildebrand, C.E., Jackson, P.J.,Jett, J.H., Longmire, J.L., Lozes, C.R., Lendemann, M.L., Martin, J.C., McNinch, J.S., Meincke, L.J., Mendelsohn, M.L., Meyne, J., Moyzis, R.K., Perlman, J., Peters, D.C., Silva, A.J. and Trusk, B.J. (1986) Human chromosome-specific DNA libraries: construction and availability. Biotechnology 4: 537-552.

Van Duin, M., de Wit, J., Odijk, H., Cretonne, A., yasui, A., Koken, M.H.M., Hoeijmakers, J.H.J. and Bootsma, D. (1986) Molecular characterization of the human DNA repair gene ERCC1: cDNA cloning and amino acid homology with the yeast DNA repair gene rad 10. Cell **44**:913-923.

Vogel, F. (1990) Mutation in man. In: Principles and Practice of Medical Genetics, second edition, Edited by A.E. Emery and D.L. Rinein, Churchill Livingstone, Edinburgh, London, Melbourne, New York.

Wallis, S.C., Rogne, S., Gill, L., Markhyem, A., Edge, M., Woods, D., Williamson, R. and Humphries, S. (1983) The isolation of cDNA clones for human apolipoprotein E and the detection of apoE RNA in hepatic and extra-hepatic tissues. EMBO J. 12: 2369-2373.

Weber, J.L. and May, P.E. (1989) Abundant class of DNA polymorphisms which can be typed using the polymerase chain reaction. Am. J. Hum. Genet. 44: 388-396.

Weitkamp, L.R., Johnston, E. and Guttormsen (1974) Probable genetic linkage between the loci for the lewis blood group and the complement C3. Cytogenet. Cell Genet. 13: 183-184.

Whitehead, A.S., Solomon, E., Chambers, S.P., Povey, S. and Bodmer, 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.

Willard, H.F. (1990) Centromeres of mammalian chromosomes. Trends Genet. 6: 410-416.

Yokobata, K.E., McNinch, J.S., Van Dilla, M.A. and de Jong, P.J. (1992) construction and characterization of a human chromosome 19-specific larger-insert lambda library. Cytogen. Cell Genet. (in press).

APPENDIX 1

PUBLICATIONS

Shutler, G., MacKenzie, A.E., Brunner, H., Wieringa, B., de Jong, P., Lohman, F.P., Leblond, S., Bailly, J. and Korneluk, R.G. (1991) Physical and genetic mapping of a novel chromosome 19 ERCC1 marker showing close linkage with myotonic dystrophy. Genomics 9: 500-504.

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

Surh, L.C., Shutler, G. and Korneluk (1991) Simple and rapid detection of PCR heteroduplexes in DNA mutations and polymorphisms. Clin. Chem. 37: 2142.

Shutler, 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.

Aslanidis, C., Jansen, G., Amemiya, C., Shutler, G., Tsilfidis, C., Mahadevan, M., Chen, C., Alleman, J., Wormskamp, N.G.M., Vooijs, M., Buxton, J., Johnson, K., Smeets, H.J.M., Lennon, G.G., Carrano, A.V., Korneluk, R.G., Wieringa, B. and de Jong P.J. (1992) Cloning of the essential myotonic dystrophy region and mapping of the putative defect. Nature 355: 548-550.

Mahadevan, M., Tsilfidis, C., Sabourin, L., Shutler, G., Amemiya, C., Jansen, G., Neville, C., Narang, M., Barcelo, 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.

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

Tsilfidis, C., MacKenzie, A., Shutler, G., Neville, C., Mahadevan, M., Wood, S., Andermann, E. and R. Korneluk (1992b) Haplotype and linkage disequilibrium patterns in myotonic dystrophy (DM). Am. J. Hum. Genet. (submitted).

REFEREED SHORT REPORTS

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

Shutler, G., Tsilfidis, C., Leblond, S. and Korneluk, R.G. (1991c) RFLP identified by the probe pKE0.6 (D19S117) at human chromosome 19q13.3. Nucleic Acids Res. 19: 1158.

Tsilfidis, C., Shutler, G., Leblond, S. and Korneluk, R.G. (1991b) An SstI RFLP detected by the probe pKE2.1 (D19S116) localized to human chromosome 19q13.3. Nucleic acids Res. 9: 1158.

Tsilfidis, C., Shutler, G., Mahadevan, M. and Korneluk, R.G. (1991c) A frequent HincII polymorphism identified by human chromosome 19 probe pKEX0.8 (D19S118). Nucleic Acids Res. 9: 1157.

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). Nucleic Acids Res. 19: 1157.

ABSTRACTS

Shutler, G., Leblond, S., Earle-MacDonald, J., de Jong, P. and R.G. Korneluk (1990) Rapid chromosome walking toward the myotonic dystrophy gene by an ALU based PCR technique. Am. J. Hum. Gen. 47: A262.

Korneluk, R.G., Leblond, S., Shutler, G., Bailly, J., Tsilfidis, C. and MacKenzie, A.E. (1990) Linkage disequilibrium between myotonic dystrophy (DM) and an extended chromosome 19 haplotype in French-Canadians. J. Neurol. Sci. 98: 38.

Korneluk, R.G., Tsilfidis, C., Shutler, G., Leblond, S., Bailly, J., MacLeod, H., Lahey, D. and MacKenzie, A.E. (1990) Two distinct chromosome 19 haplotypes show linkage disequilibrium with myotonic dystrophy (DM) in a French-Canadian population. Am. J. Hum. Genet. 47: A186.

Tsilfidis, C., MacKenzie, A.E., Shutler, G., Leblond, S., Bailly, J., Johnson, K., Davies, J., Buxton, J. and Korneluk, R.G. (1990) A recombinant event maps the DNA probe p134c telomeric to the myotonic dystrophy (DM) locus on 19q. Am. J. Hum. Gen. 47: A202.

Jansen, G., de Jong, P., Voovs, M., Aslanidis, C., Korneluk, R.G., Shutler, G., Wormskamp, N., Smeets, B., Ropers, H.H. and Weiringa, B. (1991) Genetic and physical definition of the myotonic dystrophy gene region. Am. J. Hum. Genet. 49: A405.

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

Shutler, G., Tsilfidis, C., Neville, C. 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.

Tsifidis, 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 in a 350 kb region of chromosome 19q13.3. Am. J. Hum. Genet. 49: A2036.

Sabourin, L.A., Mahadevan, M., Narang, M., Shutler, G.G. and Korneluk, R.G. (1992) Myotonic Dystrophy: Transcriptional repression of the DM kinase in affected individuals is not an immediate effect of CTG trinucleotide expansion. Am. Soc. of Hum. Genet. (in press)

Shutler, G., MacKenzie, A.E., Neville, C., Mahadevan, M., Tsifidis, C., Jansen, G., de Jong, P., Weiringa, B. and Korneluk, R.G. (1992) Correlation of consensus genetic and physical maps of the myotonic dystrophy (DM) region. Am. Soc. of Hum. Genet. (in press).