Analysis of Dystrophin Point Mutations

by

Pierre Foskett

A thesis submitted for
the degree of Master of Science (Medical Science)
University of Glasgow, March 2005

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

CONTENTS

Chapter 4
Discussion and future perspectives

References

List of Figures

List of Tables

Acknowledgements

I am grateful to my supervisors, Dr Maureen Boxer and Professor Michael Connor for their encouragement and support during the course of this work.

I would also like to thank the staff of the Medical Genetics Department, Yorkhill for their assistance and co-operation during this study, in particular the members of the molecular genetics section and Dr Alison Wilcox and Dr Douglas Wilcox.

Summary

Point mutation in the dystrophin gene accounts for around 20-35% of all cases of Duchenne and Becker muscular dystrophies. Detection of the causative mutation in each case allows accurate genetic advice to be given to families, opens up the possibility of using emerging therapies and gives insight into the pathology of the disorder in each case. Novel reports of mutation/polymorphism frequency and type will add to the knowledge of the dystrophin gene and protein function.

The overall aims of this study were to search for and characterise point mutations in the dystrophin gene. This involved the development of a rapid protocol to analyse all 79 dystrophin coding exons that could be used in a clinical setting. Robotic PCR followed by automated dHPLC of a single patient sample at a time identified amplicons containing variant sequences. Sequencing was then used to characterise the changes in the nucleotide sequence in each of these fragments by comparison to a normal control. All the data was analysed and stored in a patient specific database. Restriction digests were designed to confirm each of the mutations and the inheritance of one mutation was studied within several pedigrees.

Genotype-phenotype associations in dystrophin were assessed based on reports of increased frequency of mutation in the isoform Dp71 (exons 63-79) in individuals with mental retardation. No increase in the frequency of mutation within the brain specific transcript was observed in a cohort of individuals with learning difficulties, suggesting that the distribution of cognitive impairment in affected individuals is too broad to screen this region alone in these cases. In this study the sensitivity of dHPLC for the coding region was shown to be equal to sequencing.

A study into the incidence of mutation within a single exon in a cohort of 46 individuals with DMD/BMD showed that a single mutation within exon 70 occurred three times at the same hypermutable CpG dinucleotide. Haplotype studies using microsatellite markers and SNP analysis showed that the 3 mutations most likely originated on independent backgrounds. This was confirmed by analysis of mitochondrial HV1 haplotypes showing that a greater degree of similarity existed with a random control

population than between the patients with the same mutation. This evidence suggests that certain gene regions may have an increased incidence of mutations.

To assess variance across the dystrophin gene: all 79 coding exons of the muscle transcript Dp427 were examined in a total of 22 patients with DMD and 1 with BMD. The overall sensitivity of this technique was 96% for recognised and potentially pathogenic mutations. Several novel mutations were found, nonsense mutations: 9100C>T, 5159T>A, 754G>T, 5255T>G, 1357C>T, 5089C>T the splice site mutation: 1150-1G>T and deletion insertion mutations: 10205delA, 8438delC, 8405delC, 3036-3037delinsC1021-1022insC and 3201delT. Previously reported mutations were: 10171C>T (3 times), 313A>T, 9337 C>T (all nonsense changes). Two single codon deletions were identified - a type of change reported only once before in the literature. 10099-10101delGAA has been seen once before and is associated with a less severe form of DMD. This variant is believed to affect an uncharacterised region of the C-terminal section of the protein. The second single codon deletion is novel: 482-484delCCA is expected to delete an amino acid within an internal α-helix of the actin-binding domain in the N-terminal of dystrophin. This variant is believed to affect the binding affinity of dystrophin to actin and is associated with a BMD phenotype in this study. Two uncharacterised intronic variants were also identified: 1332-9A>G is a suggested cryptic splice site mutation within intron 11 and has been reported once before; 7098+105_7098+106_insTATTTAATACTTTG is a novel intronic change shown to be absent in over 500 chromosomes from the same population. Splice site and ESE/ESS analysis of both these changes was performed and has not been reported before. Polymorphism frequency within the study population was described with the identification of the novel variants: 358-80T>C, 5448+168A>T and 8669-75C>G. One previously uncharacterised complex variant is reported as a polymorphism in this study (8729_8734delinsTGGTCG). It was also shown that mutation is associated with stretches of repeated sequences and CpG dinucleotides distributed throughout the coding sequence.

Analysis for point mutations in the dystrophin gene of patients with DMD/BMD identified several novel variants. Characterisation of these variants has contributed to the understanding of dysfunction in the dystrophin gene and identified areas for further study. PCR and dHPLC was shown to be a rapid and sensitive means of analysing genetic sequence, suitable to large genes with infrequent (but often urgent) referrals.

Publications

Point mutation detection in Dp71 in DMD/BMD
European Society for Human Genetics Conference Birmingham 2003
P Foskett, D Wilcox, A Wilcox, A Cooke, M Boxer, JM Connor

Mutation analysis of dystrophin in the Scottish population
British Society for Human Genetics Conference York 2004
P Foskett, D Wilcox, S Stenhouse, M Boxer, JM Connor

Cognitive impairment and Dp71 point mutation analysis in Duchenne and Becker muscular dystrophies.
P Foskett, D Wilcox, S Cooke, M Boxer. (in preparation)

Variants submitted to the Leiden Muscular Dystrophy Database
URL http://www.dmd.nl/

Abbreviations

A	adenine
APS	ammonium persulphate
bp	base pair
BMD	Becker muscular dystrophy
C	cytosine
cDNA	complementary DNA
DAPC	dystrophin associated protein complex
DGGE	denaturing gradient gel electophoresis
dH_2O	distilled water
dHPLC	denaturing high pressure liquid chromatography
dNTP	deoxynucleoside triphosphate
DMD	Duchenne muscular dystrophy
DNA	deoxynucleic acid
EDTA	ethylenediaminetetra-acetic acid
EtBr	ethidium bromide
ESE	exonic splicing enhancer
ESS	exonic splicing silencer
G	guanine
g	gram
HCl	hydrochloric acid
kb	kilobase pair
KCl	potassium chloride
Mb	Megabase pair
µl	microlitre
µM	micromolar
mRNA	messenger ribonucleic acid
NaAc	sodium acetate
NaCl	sodium chloride
NaOH	sodium hydroxide
ng	nanogram
OD	optical density
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PTT	protein truncation test
RNA	ribonucleic acid
SCK	serum creatine kinase
SDS	sodium dodecyl sulphate
sH_2O	sterile water
SSCP	single strand comformation polymorhism
T	thymine
TBE	tris borate EDTA
TEMED	N, N, N', N,'-tetramethylethylenediamine
UV	ultraviolet
XLDCM	x-linked dilated cardiomyopathy

CHAPTER ONE

INTRODUCTION

1.1 General introduction

Duchenne Muscular Dystrophy (DMD) is a common inherited form of muscle disease in children. It was first described in detail by an English physician Edward Meryon in 1852 - work which has since been overshadowed by that of the French physician Guillaume Duchenne de Boulogne whose name is now eponymously associated with the disease. DMD is an X-linked disorder with an incidence of 1 in 3300 male births and the disease is characterised by progressive weakness of the proximal limb musculature appearing in early childhood. Affected boys usually become wheelchair bound by the age of 12 and most die in their late teens or early twenties from respiratory or cardiac failure. In 1982 the gene for DMD (dystrophin) was located to the short arm of the X chromosome (Xp21) and it was confirmed that another X-linked condition, Becker Muscular Dystrophy (BMD) was also associated with the same gene. Characterisation of the gene led to the identification of the dystrophin protein which was shown to be absent from the muscle sarcolemma of affected patients. Subsequent work has shown that it is an inherited defect in the dystrophin gene that causes a loss or dysfunction of the dystrophin protein resulting in either DMD or BMD.

1.2 Duchenne muscular dystrophy (DMD)

1.2.1 Clinical presentation

The disorder starts in early childhood often presenting with difficulty in walking, clumsiness and a tendency to fall. Affected children can have difficulty rising from the floor and will adopt the characteristic Gower manoeuvre by climbing up their thighs and pushing down on them to extend the hip and trunk (Emery, 1993). Skeletal muscle involvement is always bilateral and appears highly selective, this effect has been elaborated upon by Bonsett (1969). The muscle weakness in DMD is progressive although there are often periods of apparent arrest. Ultimately a wheelchair becomes necessary, in most cases by the age of 12 years. Although the skeletal muscle is the most affected tissue there is evidence that the cardiac muscle and central nervous system are also affected (Blake and Kroger, 2000; Finsterer and Stollberger , 2003). The most common cause of death is pneumonia compounded by cardiac involvement and until recently most cases of DMD died in their late teens or early twenties.

1.2.2 Pathogenesis

The most likely sequence of events in the pathogenesis of DMD is that deficiency of the dystrophin protein removes a major link between actin and the extracellular matrix in affected tissue. The cellular structure is weakened and this leaves the membrane susceptible to disruption when placed under strain (such as when the muscle is in use). The damaged membrane allows an efflux of muscle proteins (e.g. creatine kinase) and an influx of calcium. The latter leads to mitochondrial overload with ATP depletion, decreased oxidative phosporylation and the cell death seen in histochemical analysis (Emery, 2002). Fibre structure and surface-volume ratio may be important in explaining the fact that certain muscle groups are affected more than others and studies also suggested that it may not be the unique absence of dystrophin but the reduction of many dystrophin associated proteins that contributes to tissue damage (Sander *et al.*, 2000).

1.2.3 Laboratory investigations to confirm diagnosis

Confirmation of the diagnosis is essential, not only because of the serious prognostic implications but also for genetic counselling within a family. Normally the serum creatine kinase (SCK) level is up to 200 IU/l but in affected boys the levels at the beginning of the disease are elevated 50-100 times normal (Emery, 2001). As the disease progresses SCK levels gradually decline and may approach normal at later stages of the disease. Muscle biopsy specimens can be examined histologically and immunohistochemically for dystrophin. Investigations show that in the early stages of the disease muscle histology may reveal little more than an increased variation in fibre size and an increased number of eosinophilic fibres. However, by the time muscle weakness is evident fibre necrosis, invasion by mononuclear cells, phagocytosis and later replacement with fat and connective tissue are all present (Sewry, 2000). Using appropriately labelled anti-dystrophin monoclonal antibodies, in contrast to normal muscle, where dystrophin is clearly localised at the periphery of all muscle fibres, in DMD there is almost a complete absence of dystrophin apart from occasional positive revertant fibres. Revertant fibres are examples where the genetic fault is modified locally so that low level restoration of a partially functioning molecule has occurred. Mutation analysis of the dystrophin gene allows determination of carrier status as well as confirming diagnosis in the affected individual.

1.2.4 Carrier females

DMD carrier females may occasionally exhibit signs of the disease. This is attributed to skewed X-inactivation (Lyonisation) (Azofeifa *et al.*, 1995). Around 5-10% of carriers have some degree of muscle weakness that can begin at any age, symptoms can be either progressive or relatively static (Emery, 1993). In manifesting carriers SCK is invariably raised and there is some familial concordance between carrier mothers and daughters (Moser and Emery, 1974). SCK levels can be as high in asymptomatic carriers as in manifesting carriers. Cardiac involvement can occur in isolation and screening for this is recommended in all DMD carriers.

1.2.5 Management and treatment

Until a few years ago there was a general nihilism in the medical profession toward therapy in DMD. It is now acknowledged that comprehensive management requiring a multidisciplinary team approach can make substantial achievements for the patient. Promotion of ambulation, the provision of appropriate orthoses and nutritional management are all recommended. Early detection and prompt treatment of cardiac and respiratory complications is important in effective symptom control and prolongation of survival. Pharmacological therapies, in particular glucocorticoids, hold some promise in maintaining muscle strength and quality of life (Manzur, 2001). Also the possibility of gene therapy has made a resurgence in recent years. Advances have been seen in conventional gene-therapies designed to replace a shortened functional protein using vectors, plasmids and cell transplantation. Some relatively new strategies have shown promising results such as antibiotic induced read through of stop codons, utrophin upregulation and anti-sense oligonucleotide exon-skipping (van Deutekom and van Ommen, 2003).

1.3 **Becker muscular dystrophy (BMD)**

1.3.1 Introduction

This disease was named after the well-known German geneticist Prof P.E. Becker. As in DMD, others had been first to describe the features of familial muscle weakness and pseudohypertrophic calves (Kostakow and Derix, 1937), however it was Becker who recognised the disorder as a distinct entity (Becker and Keiner, 1955). Mutations in the dystrophin gene cause Becker muscular dystrophy and these mutations are associated with the production of a partially functional dystrophin protein. The disorder has a much lower prevalence than DMD, with an incidence of around 1 in 18000 live male births.

1.3.2 Clinical presentation

The condition has a wide range of onset (2-35 years) but in most cases the first symptoms are noticed between the 6^{th} and 18^{th} year of life (Emery and Skinner, 1976). Usually patients present with symptoms including falling, difficulty with climbing stairs, waddling gait and poor running, which can be ascribed to weakness of the muscles of the lower extremities. In exceptional cases the muscle weakness remains asymptomatic until very late in life (England *et al.*, 1990). Other symptoms have been reported and are covered in greater depth elsewhere (Emery, 2001). Muscle weakness is usually symmetrical and starts in the proximal muscles and lower extremities, gradually extending to the dorsal muscles and the upper extremities. Muscle hypertrophy may in fact precede muscle weakness and affects virtually every limb muscle, though it is particularly apparent in the calf muscles. In addition the heart muscle is invariably affected and this can be more pronounced than skeletal muscle disorder (Comi *et al.*, 1994).

1.3.3 Pathogenesis

In BMD the dystrophin protein is present at the cell membrane but is at reduced levels or in an incomplete or altered form. Cell structure is maintained to a greater degree and membrane damage is less severe than in DMD.

1.3.4 Laboratory investigations to confirm diagnosis

Almost all patients have increased SCK levels, usually more than five fold normal levels. Histological examination shows a similar picture to that of DMD although there are clear differences. Fibre necrosis is less marked and there is more evidence of regenerating fibres (Bradley *et al.*, 1978). Immunohistochemical staining often reveals a mosaic pattern of faintly positive disruptive dystrophin staining. As in DMD, gene mutation analysis confirms diagnosis and carrier status.

1.3.5 Disease progression, differential diagnosis and treatment

The course of the disease is not as uniform as in DMD and the clinical picture is variable even within the same family. Loss of ambulation can occur between 12 to 70 years of age, likewise age at death is highly variable (17-74 years) with a mean age of 40 years - some 25-30 years after the onset of the disease (Comi *et al.*, 1994). Because of the less severe symptoms in BMD manifesting carriers are considered rare; they show symptoms of muscle weakness and (exceptionally) heart involvement. There are two diseases that mimic BMD; limb girdle muscular dystrophy (LGMD) and spinal muscular atrophy type III. In particular LGMD can present as almost identical and in some cases only immunohistochemical analysis of the dystrophin related proteins can differentiate the two conditions (Emery, 2001). Treatment for BMD involves rehabilitation therapy for muscle weakness and regular follow-up for cardiological function. Beta-blockers may be used to reduce the risk of cardiac failure.

1.4 X-linked dilated cardiomyopathy (XLDCM)

XLDCM is a rare disorder that is associated with mutations in the dystrophin gene. Clinically it appears as similar to mild forms of BMD although there is a difference in the expression pattern of dystrophin. In XLDCM dystrophin is absent in the

myocardium but present in the skeletal muscle (Muntoni, 2003). XLDCM is associated with a small number of dystrophin mutations believed to affect tissue specific expression of the dystrophin gene.

1.5 **The dystrophin gene and proteins**

1.5.1 Full length isoforms

The dystrophin gene is the largest gene described in the human genome spanning around 2.2 million base pairs of genomic sequence. Over 99% of the dystrophin gene is intronic sequence; the coding sequence is made up of 86 exons (including seven promotors linked to unique first exons) (Muntoni, 2003) see Figure 1.1. The full-length messenger RNA is predominantly expressed in the skeletal and cardiac muscle with small amounts expressed in the brain. There are three isoforms of this full length protein each derived from three independent promotors in the brain, muscle and Purkinje cerebellar neurons. These three promotors transcribe a unique first exon which is then spliced to a common set of 78 exons (Torelli *et al.*, 1999). Expression of each promotor is specific to certain sites, for example the brain promotor drives expression primarily in the cortical neurons and the hippocampus whereas the Purkinje promotor is mainly expressed in the cerebellar Purkinje cells. The muscle promotor drives expression in the skeletal muscle and cardiomyocytes and is also expressed at low concentrations in some glial cells in the brain (Muntoni *et al.*, 1995).

The dystrophin gene also produces variants of each of these principal isoforms, generated through alternative splicing events. Both exon skipping and exon scrambling have been observed in various dystrophin transcripts (Sadoulet-Puccio and Kunkel, 1996). These events, which occur in a tissue specific manner, generate even further protein diversity from the one gene.

1.5.2 Full length dystrophin protein

The protein's function is to link actin in the cytoskeleton and the dystophin associated protein complex at the cell membrane which in turn binds with proteins in the extracellular matrix. This structural role for dystrophin acts to stabilise the sarcolemma and protect the muscle fibres from long-term contraction induced damage. There have

also been suggestions that the protein is involved in cellular communication by acting as a transmembrane cell signalling complex (Rando, 2001)

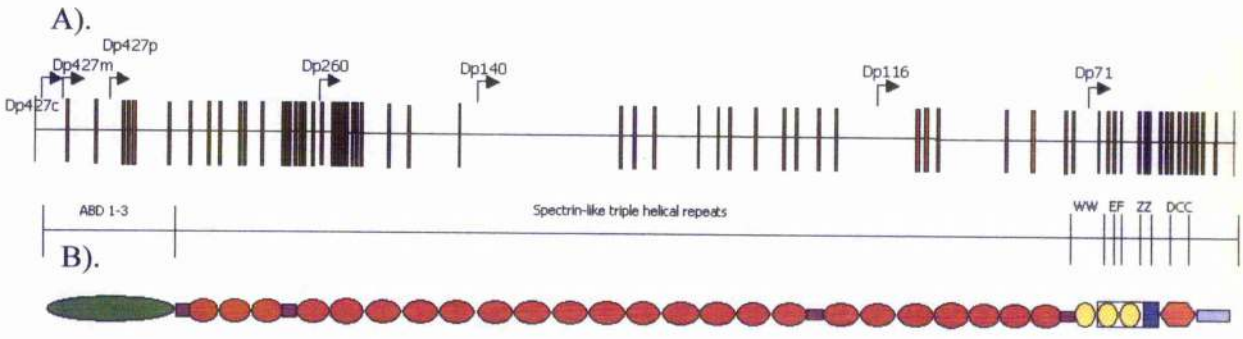


Figure 1.1 Dystrophin gene and protein organisation. A schematic representation of the dystrophin gene. A). The 2.217Mb gene showing the relative distribution of exons and the location of each of the seven promotors. The red bars signify exons. The arrows indicate promotors (transcribed from centromere to telomere). B). The position of dystrophin major structural domains in relation to the gene sequence, underneath this is a representation of the molecule with different shaped regions depicting the different structural elements of the protein, ABD = actin binding domain sites 1-3, the 24 spectrin-like repeats are shown interspersed by four hinge regions WW= WW domain, EF= EF hand calcium binding domain, ZZ= ZZ zinc finger domain, DCC = dimeric coiled coil.

Full length dystrophin is a large rod shaped protein with a molecular weight of 427 kDa. Dystrophin belongs to the spectrin family of cytoskeleton proteins and shares the common elements of this group: an amino terminal actin-binding domain, a central triple helical coiled coil domain and a carboxy terminal membrane binding domain (Figure 1.1). The actin-binding domain (ABD) has homology with α-actinin and consists of two calpain homology (CH) domains that are both α-helical globular folds. The ABD of dystrophin binds to sub-domain 1 of actin and has a conserved tertiary structure (Norwood *et al.*, 2000). The central region of dystrophin consists of 25 triple helical spectrin-like repeats and contains around 3000 residues. The amino terminal end of these repeats are also involved in actin binding (Sutherland-Smith *et al.*, 2003). The triple helical structure confers an elongated shape to the molecule that is interspersed by four hinge regions. The spectrin repeats are poorly conserved but each segment shows a common tertiary structure. The hinge regions give additional flexibility to the molecule when exposed to the contraction-relaxation forces of the muscle. Towards the carboxy-terminal there is a cysteine-rich domain of around 280 amino acids. The high level of cysteines in this region (15) show strong structural conservation. The carboxy-terminal region contains motifs for a WW, ZZ and two (calcium binding) EF hand domains

(Winder, 1997). The cysteine rich domain is involved in binding to β-dystroglycan which is the principal link to the dystrophin-associated protein complex and the extracellular matrix. Finally the terminal 325 amino acids contain binding sites for α1-syntrophin and β1-syntrophin and end in a coiled coil motif shared with dystrobrevin to which dystrophin is believed to interact with in this region (Sadoulet-Puccio *et al.*, 1997).

1.5.3 The dystrophin protein and its relationship to other proteins at the muscle cell membrane

The dystrophin protein is very similar in structure and function to utrophin on chromosome 6 and they share all the same protein domains. Utrophin is expressed in all tissues and is 'shorter' than dystrophin as it lacks some of the helical spectrin repeats in the rod region. Because of their similar function some therapeutic strategies have aimed at increasing expression of utrophin to compensate for the loss of dystrophin function (Perkins and Davies, 2002). In muscle cells dystrophin is the principal linker molecule between actin and the cell membrane where it interacts with the dystrophin-associated protein complex (DAPC) consisting of dystroglycans and sarcoglycans; utrophin acts as the principal actin-membrane link in non-muscle cells by interaction with the dystroglycans alone (Winder, 1997). The dystrophin-associated protein complex is the link between the extra-cellular matrix and the intracellular (dystrophin anchored) actin (Figure 1.2). It is a multi-functional multi-protein complex, with roles in cell signalling as well as stability.

Other forms of muscular dystrophy have been shown to arise from mutation in the laminin, sarcoglycan and caveolin DAPC protein families. Although dystroglycan mutations have not been observed, proteins involved in dystroglycan glycosylation (glysosyltransferases) have also been shown to cause muscular dystrophy (Ehmsen *et al.*, 2002).

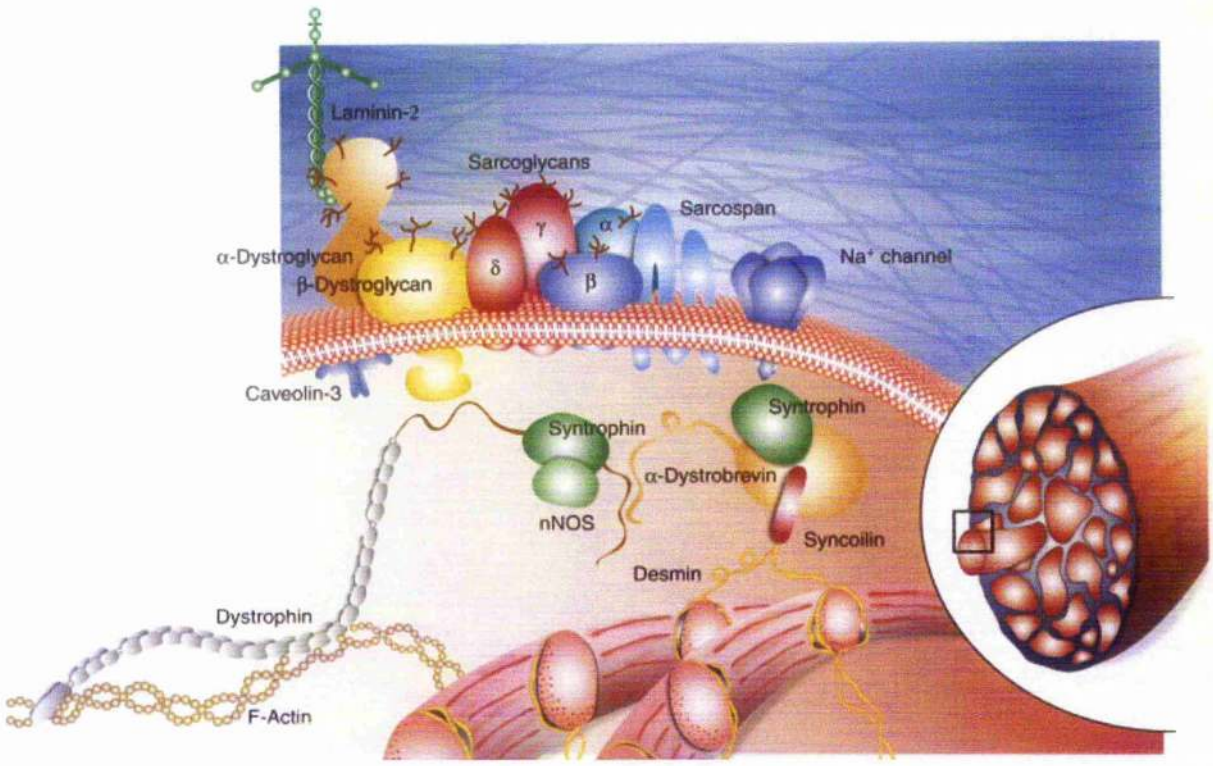


Figure 1.2 The dystrophin associated protein complex (DAPC). A diagram showing the relationships of the dystrophin protein with other proteins in the muscle cell. Image adapted from Ehmsen *et al.*, 2002. Dystrophin is principally linked to a transmembrane core of α-dystroglycan, β-dystroglycan which in turn is linked to the extracellular matrix via laminin-2. Bound to this dystroglycan core are five transmembrane proteins of the sarcoglycan family (α,β,γ, δ and ε) which also interact with dystrophin on the cytoplasmic surface (Chan *et al.*, 1998). Also associated with the sarcoglycan part of the complex is sarcospan (another transmembrane protein) which binds a cytoplasmic complex of syntrophins (α, β1, β2 and γ1 and γ2) (Ehmsen *et al.*, 2002). The syntrophins bind dystrophin near the carboxy terminal and they also bind α-dystrobrevein which in turn is associated with dytrophin at its terminal functional domain. As well as forming dystrophin-syntrophin-sarcospan interactions dystrobrevin is also involved in linking the desmin intermediate filaments to the DAPC via syncoilin (Newey *et al.*, 2000). Cell signalling is mediated by caveolin-3 which interacts with the carboxy terminus of β-dystroglycan, caveolin-3 regulates nNOS which is involved in production of nitric oxide an important regulator of local blood flow (Ehmsen *et al.*, 2002)

1.5.4 The smaller isoforms of dystrophin

The dystrophin gene has at least four internal promotors that give rise to shorter dystrophin proteins lacking successively less and less of the amino terminal of the full length protein but retaining the carboxy terminus domains. All these shorter proteins contain the binding sites for dystroglycan, dystrobrevin and syntrophin. Each isoform is expressed differently depending on tissue and their functions remain unclear. Each of the internal promotors uses a unique first exon that splices into exons of the Dp427m transcript (Figure 1.1). Dp260 is expressed in high concentrations in the retina, where it co-exists at the synapse between retinal photoreceptor and bipolar cells with the full

length brain and muscle isoforms (D'Souza *et al.*, 1995). Dp140 is expressed in brain, retina and kidney tissues during development (Lidov *et al.*, 1995). Dp116 is only expressed in adult peripheral nerves (Byers *et al.*, 1993). Dp71 is detected in most non-muscle tissues including brain, retina, kidney, liver and lung and is present in the cardiac but not skeletal muscle. Dp71 is the most abundant isoform in the early embryonic forebrain, continuing into adult life specifically in the cerebral cortex, hippocampus, pituitary and retina (Mehler, 2000). Alternative splicing of dystrophin RNA is extensive, tissue dependant and developmentally regulated. Excision of one or more exons creates different versions of each isoform lacking small regions of the protein. These findings suggest that dystrophin subserves a wide variety of cellular functions that exhibit developmental, regional and neural cell-type specificity. However, all these isoforms share similar structural domains implicating the importance of the carboxy terminal end of the protein in the role of mediating different cellular functions.

1.5.5 Dystrophin gene structure

The 79 exons of the principal Dp427m transcript range in size from 23 to 269bp and are distributed over 2.217kb of sequence. It has been predicted that the cell takes 16 hours to transcribe the complete gene (Tennyson *et al.*, 1995). Intron distances are extremely varied with some less than 1Kb and several over 100kb, the largest is intron 44 (248kb). Several clusters of exons can be identified and these are 20-28, 31-33, 38-40 and the C-terminal exons (64-79) which have only four introns over 10kb. Several very large introns are found between exons 41 and 63. The dystrophin introns have a high concentration of repetitive elements (32.1% overall) with LINE-1 elements representing the major contributor to intron size (Toffolatti *et al.*, 2002). There is a direct correlation between intron size and the number of repeat elements suggesting an incremental growth of the intron. Although dystrophin intron size appears well conserved between species the expansion of these introns has been shown to have occurred independently (Pozzoli *et al.*, 2003). The role of increased intron size is unknown but it appears to be related to the observation that out-of-frame exons are flanked by significantly larger introns than in-frame exons and that larger introns have been shown to stabilise mRNA splicing. There is evidence that the recombination events in dystrophin are mostly non-homologous (Toffolatti *et al.*, 2002). The rate of recombination in the dystrophin gene is influenced by the number of repeats involved in non-homologous recombination rather than the frequency of recombination between similar sequences. It is thought that

the repeat sequences have a propensity to form hairpin structures which are more likely to break and then subsequently rejoin with a change in sequence. Thus a recombination introducing more repeated sequence size will propagate further recombination events. This may in part explain the high recombination rate across the dystrophin gene, which occurs at a frequency of approximately 12% (Abbs *et al.*, 1990).

1.6 **Dystrophin gene mutation**

DMD and BMD are caused by certain types of genetic mutation within the dystrophin gene. The prevalence of both conditions is maintained in the population by new mutation. New mutation almost always occurs in the germline but somatic mutations have been observed in both males and females (Smith *et al.*, 1999; Helderman van den Enden *et al*, 2003). The overall ratio between female and male origin of mutation does not deviate from 1:1 but there are suggestions that point mutations arise preferentially in spermatogenesis (Grimm *et al.*, 1994). Germline mosaicism of a dystrophin mutation was first described by Bakker *et al.* in 1987 and there is a risk of 5-10% of any apparently new mutation recurring in subsequent pregnancies (van Essen *et al.*, 1997).

1.6.1 Deletions and duplications

The most common mutations in the dystrophin gene are intragenic deletions of one or more exons, which account for 60-65% of dystrophin mutations. Reports of exon duplication are less frequent and range between 5 and 15%, possibly as a result of the sensitivity of techniques used (Muntoni *et al.*, 2003). Both deletions and duplications can happen almost anywhere within the gene but most occur within two recognised hotspots. The first is the most commonly mutated region and includes exons 44-55. The second is in the 5' region of the gene and includes exons 2 to 19. These hotspots for mutation are associated with the presence of large introns, notably intron 44 which comprises around 10% of the whole gene alone. These clusters of mutations represent the basis for the use of the multiplex PCR technique, by the screening of only a set of 18-28 exons almost all exon deletion mutations can be detected (Beggs *et al.*, 1990; Abbs *et al.*, 1991) Quantitative measurement of similar multiplexes also allows duplications to be detected with a similarly high level of sensitivity (Yau *et al.*, 1996).

1.6.2 The frameshift hypothesis

Monaco *et al.* in 1988 proposed that mutations of the dystrophin gene that maintain the translational reading frame might be expected to result in an abnormal but partially functional dystrophin in muscle, this would result in the milder Becker type of dystrophy. On the other hand, mutations that shift the reading frame are expected to cause a truncated protein and would result in virtually no functional dystrophin, causing DMD. This frameshift hypothesis holds up in over 90% of cases and is of diagnostic as well as prognostic significance. Essentially the dystrophin protein can function in the absence of the central and distal rod domains. In some cases expression of an abbreviated protein has been associated with symptoms of myaglia and muscle cramps, but not muscle weakness (Angelini *et al.*, 1996). However, when the protein is prematurely truncated patients have a very similar severe phenotype of DMD. The reason for this effect is thought to be because of nonsense mediated RNA decay removing the mutant transcript before it is translated into a protein resulting in the loss of the dystrophin mediated link between the cytoskeleton and membrane proteins (Hilleren and Parker, 1999).

1.6.3 Exceptions to the frameshift rule

Some patients with BMD have frameshift mutations, these mutations appear to affect certain exons, most commonly in the 5' region of the gene. In the case of deletions of exons 3-7 it is thought that an additional upstream translation start site (possibly in exon 8) may be used (Winnard *et al.*, 1995). However this phenomenon is not exclusive to mutations in the 5' region of the gene and therefore there must be some other mechanism involved in rescuing protein expression. Exon skipping events are routinely seen in patients with DMD but usually limited to a few fibres within the muscle tissue (revertant fibres). This is where additional exons are removed in RNA processing allowing the reading frame to be restored. It is believed that exon skipping is occurring at sufficient levels to produce a BMD phenotype in some frameshift mutations. Exon skipping is thought to occur in all regions of the gene but it may be limited to certain mutations (Nicholson *et al.*, 1992). Even the same mutation in different patients may show a different phenotype. One likely contributory factor could be the differences between deletion breakpoints within the introns. The splicing process particularly in

large introns is mediated by several intronic motifs, distributed along the length of the intron. Therefore different deletions might affect different exon skipping events (Muntoni *et al.*, 2003). Observations of inconsistent splicing amongst transgenic mice with the same mutation suggest that there are other non-regional factors involved (Lu *et al.*, 2000), with the added possibility of variation in the nonsense mediated decay process (Kerr *et al.*, 2001). This complexity adds uncertainty to predictions of phenotype based on mutation type alone and in some of cases confirmation of diagnosis by muscle biopsy will be required. Other rare exceptions to the frame-shift rule are in-frame deletions that cause DMD, these have been shown to remove important functional regions of the gene and are in general predictable (Nevo *et al.*, 2003). Mutations that produce a dysfunctional protein are useful in helping to characterise important domains of the dystrophin protein.

1.6.4 Point mutations

Around 20-35% of patients with DMD and BMD do not have a deletion or duplication of whole exons but instead have a change in the nucleotide sequence of the dystrophin gene. These mutations are distributed throughout the gene sequence. In accordance with the frame-shift hypothesis, almost all of the point mutations so far identified in patients with DMD are expected to lead to premature translational termination. These constitute: nonsense mutations, small frame-shift deletions/insertions and mutations that affect splicing (Roberts *et al.*, 1994). Mutations affecting splicing are varied. Point mutations in exon/intron boundaries and deep intronic regions have been shown to cause exon excision or cryptic exon insertion shifting the reading frame (Tuffery-Giraud *et al.*, 1999). Thus premature translation termination can be caused in a variety of ways. Although well conserved the dystrophin gene seems tolerant to a certain amount of non-pathogenic sequence variation. Based on this, reports of in-frame missense mutations causing DMD should be regarded with caution until the possibility of another type of mutation has been excluded (Roberts *et al.*, 1994).

The nature of BMD point mutation is more heterogeneous with around 50% shown to affect splicing resulting in an in-frame transcript and the rest being missense mutations (Roberts *et al.*, 1994). Exon skipping causing an in-frame BMD transcript has been observed with splice site mutations, cryptic splice site mutations and also with some nonsense mutations (Fajkusova *et al.*, 2001; Tuffery-Giraud *et al.*, 2003). The effects of

the splicing mutations are often unpredictable from observations of the DNA sequence alone. The missense mutations reported in BMD cluster around the N-terminal and C-terminal regions and are believed to affect the binding domains of each region (Roberts *et al.*, 1994; Norwood *et al.*, 2000). As the protein is completely retained with a missense change these alterations are expected to yield information on the structure and function of the dystrophin molecule.

1.6.5 Detection of point mutations

The detection of point mutations in dystrophin presents a considerable technical challenge, not least because of the large number of small exons, the large intervening introns and the distribution of reported mutations. Muscle RNA has been the target of choice for several laboratories with preference for the Protein Truncation Test (PTT) as a means of analysing mutations (Roest *et al.*, 1993). The drawback to these approaches is that they require muscle biopsy, an invasive procedure with a risk of complications (e.g. bleeding, infection and haematoma formation) of ~1% and one that may often be associated with psychological distress for children (Flanigan *et al.*, 2003). This technique does have several advantages particularly with respect to the effects of a DNA change on the gene transcript. Splice site mutations, intronic variants and exon skipping are all types of mutation that have been reported using RNA but would present difficulties for any DNA based analysis (Tuffery-Giraud *et al.*, 1999). However, the lack of samples has hampered the use of dystrophin mRNA analysis in many cases.

DNA samples are more readily available from patients with DMD, BMD and those where there may be a differential diagnosis (Muntoni, 2001). The technology of DNA analysis is also well developed and there are many techniques that can be used to identify variance within the DNA sequence. Single Strand Conformation Polymorphism (SSCP), heteroduplex analysis, Denaturing Gradient Gel Electrophoresis (DGGE) are commonly reported 'scanning' methods for dystrophin point mutation analysis (Prior *et al.*, 1994; Mendell *et al.*, 2001; Hofstra *et al.*, 2004). These are generally based on multiplexed PCR being analysed by differences in mobility through a polyacrylamide based matrix. High levels of variant detection have been reported in each of these methods and with automation these approaches can achieve a rapid analysis of many samples. Direct sequencing and Denaturing High Pressure Liquid Chromatography (dHPLC) analysis of the dystrophin gene has also been reported, both these techniques

are based on the analysis of each exon individually and have very high levels of sensitivity (Bennett *et al.*, 2001; Flanigan *et al.*, 2003). Although both processes are automated, direct sequencing involves a considerable amount of data analysis and dHPLC like the other scanning methods required subsequent sequencing to characterise the variants. Advances in software are improving the process of sequence data analysis and the introduction of robotics has facilitated the application of PCR to gene analysis. This will make the choice of both dHPLC and direct sequencing more equitable in terms of speed to a multiplexed approach. Given the potential for sequencing, scanning methods are still currently preferred due to the high costs of sequencing per amplicon. Combining automated PCR with dHPLC followed by semi-automated sequencing is expected to provide a system of analysis that is sensitive, rapid and economical.

1.6.6 Genotype-phenotype associations with Dp71

Based on the diversity of transcripts within the gene it has been suggested that differences in phenotype in DMD may be influenced by the location of the mutation (Moizard *et al.*, 1998). The Dp71 transcript is the most abundant dystrophin product in the brain, it comprises exons 63-79 of the Dp427m transcript with a unique 21 nucleotide first exon adjacent to its promotor. Although the protein function of Dp71 is unknown it has been shown to be involved in the binding of membrane structures and it contains the same binding domains as Dp427m (Fabbrizio *et al.*, 1994). Several reports of point mutation within the Dp71 region have been associated with a phenotype of a low IQ (<70) (Lenk *et al.*, 1993; Moizard *et al.*, 2000,) and Moizard *et al.* (1998) have also attempted to correlate IQ with region of deletion in a large group of patients. In these studies the clinical description of patients with cognitive impairment and a mutation in Dp71 ranges from moderate to severe. It has been suggested that a point mutation screen of Dp71 coding region in affected DMD/BMD patients with cognitive impairment would detect more mutations than in patients without.

1.6.7 Recurrent mutation

If there were consistent evidence for an increased incidence of either common mutations or evidence of mutational hot spots in certain exons this could be the basis of a staged approach to point mutation analysis. The relatively high prevalence of DMD is maintained by new mutations that have a short survival within families suggesting that

most mutations are unique to each pedigree (ten Kate *et al.*, 1984). Although in BMD mutations have a lower prevalence, when these are found there are likely to be more members affected due to the higher survival rate of affected individuals (Bushby and Gardner-Medwin, 1993). In almost all new cases of DMD the mutation is therefore expected to be unique whilst in BMD there will be a slightly higher chance that the mutation may have been observed before. Some point mutations have been reported multiple times in DMD, even in the same study, yet no investigations have been made to determine background to these mutations (Flanigan *et al.*, 2003; Hofstra *et al.*, 2004; Tuffery-Giraud *et al.*, 2004). Haplotype studies of the dystrophin gene region could have thrown light on the origin of these mutations. In BMD a common site of mutation could have relevance in the understanding of the protein function or in DMD there may be sites of recurrent point mutation apparent only when larger populations of affected individuals are studied. Alternatively a mutation could have originated from a common ancestor and had been passed through several generations into families unknown to each other. Thus the identity of common mutations may be important for mutation strategy within specific groups of patients.

1.6.7.1 Assessing the inheritance of a mutation

Each of the techniques available for assessing the inheritance of a mutation have their limitations and a combination of approaches may be required. The use of haplotyping using intragenic microsatellite repeats has the advantage of a high level of variance at each locus but due to high recombination rate across the dystrophin gene the determination of inheritance can be limited to immediate relatives only. Occurring at roughly one every two kilobases, SNP haplotyping offers a higher level of resolution but can be limited by the low capacity for variance at each locus. The use of mitochondrial variable sequences can also be considered in determining maternally inherited haplotypes in DMD/BMD families. The mutation rate in mitochondrial DNA (average 0.0043 per generation) is high enough to generate a large pool of haplotypes in a population yet is considered stable enough for the study of these haplotypes through tens of generations (Sigurgardottir *et al.*, 2000).

1.6.8 Whole gene analysis

Review of the dystrophin mutation database (www.dmd.nl) shows that point mutations in the dystrophin gene have been reported in virtually every exon, although some studies using limited point mutation screening have introduced a bias towards individual exons of the gene (Prior *et al.*, 1995). As results from complete studies of the gene become available a more accurate picture for mutation distribution and type will be available. Combined with quantitative techniques, whole gene point mutation analysis offers the possibility of detecting almost all dystrophin mutations. Furthermore, the characterisation of any novel base change in current studies is important for the interpretation of DNA results in future cases. Observations of polymorphism are expected to contribute to our understanding of dystrophin variance in different populations.

1.7 **Aims of the present study**

- To test the theory that cognitive impairment may be associated with a point mutation in Dp71: analyse dystrophin exons 63-79 in a cohort of patients by dHPLC.
- Compare dHPLC analysis with sequencing of dystrophin exons 63-79.
- To observe the incidence of mutation in a single exon: use and evaluate automated robotic PCR and analyse exon 70 by dHPLC in a cohort of 46 DMD/BMD patients.
- To identify whether the mutation 10171C>T found in three families is on a common genetic background.
- Establish a robotic protocol for the PCR amplification of dystrophin exons 1-79 with analysis of these products by dHPLC followed by sequencing, on a single patient basis.
- Analyse a cohort of patients with DMD/BMD for point mutation in dystrophin exons 1-79 using the protocol above.
- Characterise any potentially pathogenic variants and evaluate any evidence for the cause of each DNA mutation.
- Assess polymorphism in the dystrophin gene within the study population

CHAPTER TWO

MATERIALS AND METHODS

2.1 **Extraction of DNA from venous blood**

Genomic DNA was extracted from peripheral blood lymphocytes. 5-10ml of whole venous blood was drawn from the patient and transported in EDTA blood collection tubes (Labtek,UK). On receipt in the laboratory the blood was transferred to a 30ml sterile falcon tube and diluted to a total volume of 20ml with a solution of 155mM NH_4Cl, 10mM $KHCO_3$ and 1mM EDTA. The solution was mixed by gentle inversion and left for 30 minutes to lyse the erythrocytes.

A white cell pellet was collected by centrifugation at 1300g for 10 minutes at 10°C. The supernatant was discarded and the pellet re-suspended by vortexing for 30 seconds with a 6ml solution of 2% SDS, 25mM EDTA and left at 37°C for 30 minutes. Addition of 2ml 10M ammonium acetate and vigorous vortexing formed a protein precipitate which was pelleted by centrifugation at 1300g for 20 minutes. The clear supernatant was then removed to a sterile 20ml universal container containing 15ml isopropanol, gentle swirling precipitated the DNA forming a clump which could be extracted using a sealed sterile glass pipette. The clump was washed in a 70% ethanol solution, air dried for 1 minute and then re-suspended in 500µl of TE buffer in a 1.5ml starsted storage tube. The solution was left at 65°C for 2 hours to dissolve the DNA and then stored at -20°C until analysed.

2.2 **Determining DNA concentration**

10µl of the extracted genomic DNA sample was dissolved in 990µl of dH_2O by vortexing for 2 minutes, this was then transferred to quartz cuvette and the OD_{260} of the sample was measured in a GeneQuant II spectrophotometer against a dH_2O reference sample. The OD_{260} reading was converted to DNA concentration using the following equation:

OD_{260} X 50 X 100 = DNA concentration (ug/ml)

Where: 50 signifies ug dsDNA per 1 OD_{260} unit.
100 signifies the dilution factor.

2.3 **Agarose gel electrophoresis**

DNA fragments were fractionated by agarose gel electrophoresis using two commercial gel tanks with gels at a variety of concentrations. Volumes used were 100ml for the Electrofast® (Abgene) 96 well apparatus and 30ml for the Horizon 58® (Gibco BRL) 28 well apparatus. 1% agarose gels were made using Seachem electrophoresis grade agarose (Flowgen) and 2 to 4% agarose gels were prepared using PhoreCus™ agarose (Biogene, UK). Gels were prepared by dissolving the desired weight of agarose in 30 or 100ml of 1X TBE solution, this was achieved by boiling the agarose slurry in a microwave accompanied by regular stirring. The molten agarose was then left to cool in a 65°C waterbath. Once the apparatus was set up with the appropriate dams, EtBr was added (3ug per 100ml of agarose) and the gel poured in. The gels were left to set for a minimum of 1 hour. Once set the gel dams were removed and the gels immersed in 1X TBE buffer prior to removing the combs. PCR samples were then loaded into the wells and run at a constant 100V until the bands under study were resolved. The DNA was visualised on a UV transilluminator (UVP Inc) and a CCD image of the gel was then taken and stored as a printed image.

2.4 **PCR and oligonucleotide design**

2.4.1 Dystrophin sequence data

Sequence data for the coding region of the Dystrophin muscle transcript (Dp427m) was taken from Genbank Accession: NM_004006.1. All base changes have been described by their effect on this sequence. The nomenclature used is according to Antonarakis (1998) and den Dunnen and Antonarakis (2000). The intronic sequences were taken from the following Genbank files:

Approximate 5' locus	Accession No.
5'UTR, muscle	M32058
Exon 2	AL139401
Exon 6	AL121880
Exon 7	U60822
Exon 8	L08092
Exon 9	U06836
Exon 10	AC004468
Exon 17	AL031542
Exon 44	AC069170
Exon 47	AC021166
Exon 51	AC079864
Exon 52	AC025935
Exon 56	AC079175
Exon 57	AC079177
Exon 63	AC078958
Exon 72	AC079143
Exon 79 + 3' UTR	AC006061

Each exon nucleotide sequence and 300 hundred base pairs of flanking intronic sequence was compiled into a single MS Word document, from this an EMBL format file with all the amplicons arranged in gene order was constructed.

2.4.2 Primer design and PCR

Primer sequences amplicon size and the PCR programme used in amplification of amplicons containing exons 1 to 79 of the Dp427m transcript are listed in tables 2.1 and 2.2. These primers were designed to amplify each exon of the Dp427m transcript including at least 50bp of flanking intronic sequence. Primers were synthesised commercially (Sigma-Genosys, UK) and diluted to 5μM concentration according to the manufacturers instructions. The primers have been optimised to one of two following reaction conditions: All reaction mixes were prepared in 200 μl microtubes (Abgene):

18μl Megamix-without BSA (Microzone Ltd)
2.5μl Sense primer (5μM)
2.5μl Antisense primer (5μM)
2μl DNA (50ng/μl)

Thermocycler conditions:

File : DMD_45		File : DMD_55	
93°C	5 minutes	93°C	5 minutes
14 cycles of		14 cycles of	
94°C	30 seconds	94°C	30 seconds
55°C (-0.5°C every cycle)	30 seconds	62°C (-0.5°C every cycle)	30 seconds
72°C	45 seconds	72°C	45 seconds
20 cycles of		20 cycles of	
94°C	30 seconds	94°C	30 seconds
47°C (-0.1°C every cycle)	30 seconds	56°C (-0.1°C every cycle)	30 seconds
72°C	45seconds	72°C	45 seconds

PCR programmes were run on a GRI PTC-0200 thermocycler.

Primer sequences came from two sources. Primers taken from Bennet *et al.* 2001 were identified by the label AMER. Primers designed using Primer3 - a web based primer design programme (Rozen and Skaletsky, 2000) - have been given the label GLA

Table 2.1
Primer pairs used and resultant amplicon sizes.

Sense Primer	Sequence	Antisense Primer	Sequence	Size of PCR product
DMD_DHPLC_GLA_Ex01s	TCCTGGCATCAGTTACTGTGTT	DMD_DHPLC_GLA_Ex01as	CACTCCATTTCTGTTGAAAAGTT	476
DMD_DHPLC_GLA_Ex2s	ATGGAAAGTTACTTTGGTTG	DMD_DHPLC_GLA_Ex2as	CAGGTACATAGTCCATTTTG	212
DMD_DHPLC_GLA_Ex3s	TCAAAAGGGGATAATCGTGAA	DMD_DHPLC_GLA_Ex3as	GGTATTGCTGTTTCAATCAGTACCT	291
DMD_DHPLC_AMER_Ex4s	CGTACCAGGCCAAGGACAATTAGC	DMD_DHPLC_AMER_Ex4as	CCAAAGCCCTCACTCAAACATGAA	322
DMD_DHPLC_GLA_Ex5s	TTGCAACTAGGCATTTGGTC	DMD_DHPLC_GLA_Ex5as	TTGTTTCACACGTCAAGGGTA	228
DMD_DHPLC_AMER_Ex6s	GCCCTAAAATTTCTATTTATCACT	DMD_DHPLC_AMER_Ex6as	CTGGGGAAAAATATGTCATCA	356
DMD_DHPLC_GLA_Ex7s	AGGACTATGGGCATTGGTTG	DMD_DHPLC_GLA_Ex7as	CAGATGAAAACATTAAACTCTACCAT	293
DMD_DHPLC_GLA_Ex8s	GAAACCAAAAATTGATGTGTAGTG	DMD_DHPLC_GLA_Ex8as	TGCATATAAAACAGAAAACATCTTG	274
DMD_DHPLC_GLA_Ex9s	TATGGTTTTTCCCCCTCCTC	DMD_DHPLC_GLA_Ex9as	GGCACTGAAAAATTCAAGCA	272
DMD_DHPLC_GLA_Ex10s	GCAAAGACATTAATTGTGTAACACC	DMD_DHPLC_GLA_Ex10as	GGATGACTTGCGATTATAAC	295
DMD_DHPLC_GLA_Ex11s	CCACACCGATTTACCTAGAGTTC	DMD_DHPLC_GLA_Ex11as	CAAAATAATCACAAGCTTCCAAAA	300
DMD_DHPLC_GLA_Ex12s	CATTGTGATGTTCAGTAATAAGTTGC	DMD_DHPLC_GLA_Ex12as	CCATGTCATCTGTGTTACTGTGT	300
DMD_DHPLC_AMER_Ex13s	GCAGAAAATTGGCTTGGAATGGTT	DMD_DHPLC_AMER_Ex13as	CCCATCCGCAGTTAGTTACTATTG	477
DMD_DHPLC_AMER_Ex15as	TTTGAATAATCTATGATCCAAGCA	DMD_DHPLC_AMER_Ex14s	GTGTCCCTTCCAACTTCTAGC	569
DMD_DHPLC_AMER_Ex16s	GCAAATGAGCAAATACACGCAAAA	DMD_DHPLC_AMER_Ex16as	CACAGGGCAAAAACTAATCTGGTT	388
DMD_DHPLC_GLA_Ex17s	AAGGGGCGCTAAATTGACTT	DMD_DHPLC_GLA_Ex17as	TTGCTGAAGTGAAAACACCA	498
DMD_DHPLC_AMER_Ex18s	GTCAGGCAGGAGTCTCAGATTGA	DMD_DHPLC_AMER_Ex18as	GAAAGGCATCCCTAGTCAGTCACA	456
DMD_DHPLC_AMER_Ex19s	GATGGCAAAAGTGTTGAGAAAAAGTC	DMD_DHPLC_AMER_Ex19as	TTCTACCACATCCCATTTTCTTCCA	459
DMD_DHPLC_AMER_Ex20s	GGATGTGTTGGCTTTCAGATCATTTC	DMD_DHPLC_AMER_Ex20as	TGGAATGCCAAGAAATACCTATTGA	377
DMD_DHPLC_GLA_Ex21s	ATGTATGCAAAGTAAACGTG	DMD_DHPLC_GLA_Ex21as	TGGAAAATGTCAAGTTAGCC	270
DMD_DHPLC_GLA_Ex22s	GGAAAACATGGCAAAGTGTG	DMD_DHPLC_GLA_Ex22as	GCTCAATGGGCAAACTACCA	332
DMD_DHPLC_GLA_Ex23s	ACAGTTAATTTCACTAAAACTCATC	DMD_DHPLC_GLA_Ex23as	AAGATGCTGAAGGTCAAATGC	394
DMD_DHPLC_GLA_Ex24s	TTGTGTTAAAAGTAATCAGCACACC	DMD_DHPLC_GLA_Ex24as	CCCCAGCTGTAAAACACTGA	250
DMD_DHPLC_GLA_Ex25s	TCCATATGCAATGCCATCAG	DMD_DHPLC_GLA_Ex25as	GGAACAAAGCCTTAACCAAAA	362
DMD_DHPLC_GLA_Ex26s	TTCATCACTGTCAATAATCGTGTTT	DMD_DHPLC_GLA_Ex26as	CAACTTCAAGCATTGTTGCAT	297
DMD_DHPLC_AMER_Ex27s	TCTCATTCTAACTGGGATGTTGTG	DMD_DHPLC_AMER_Ex27as	CCATGTTCTTAACCACTATGC	370
DMD_DHPLC_AMER_Ex28s	GCTGCATTTTGAATTACCTGCTACA	DMD_DHPLC_AMER_Ex28as	TGGGTTGTTTTCTTTGGATTTGTC	374
DMD_DHPLC_GLA_Ex29s	TGAGCATTTGCTGATAATCCA	DMD_DHPLC_GLA_Ex29as	TCTGAGAGGCCTGTATCTGCT	292
DMD_DHPLC_AMER_Ex30s	CATTTATTGTTCAGCAGGATTAC	DMD_DHPLC_AMER_Ex30as	CCCATGGAAAACTAGTTGAATAA	344
DMD_DHPLC_GLA_Ex31s	CATGGTAGAGGTGGTTGAGGA	DMD_DHPLC_GLA_Ex31as	TGCCAATAATCAAGTTGTCCA	295
DMD_DHPLC_GLA_Ex32s	CCAGTTATTGTTTGAAAGGCAAA	DMD_DHPLC_GLA_Ex32as	GCCACAATACATGTGCCAAT	387
DMD_DHPLC_GLA_Ex33s	ACCGCTGCAAAATGCTACTC	DMD_DHPLC_GLA_Ex33as	TGCCCGTTGCTTTACAATTT	360
DMD_DHPLC_GLA_Ex34s	AAGTGGTTATAACGAAATTTGA	DMD_DHPLC_GLA_Ex34as	TTATGTGTTTTCACGTATGTT	288
DMD_DHPLC_GLA_Ex35s	CCATACAGAAAGCCGTTTCA	DMD_DHPLC_GLA_Ex35as	GCCAGCATATACGTAGAATTGAGA	375
DMD_DHPLC_AMER_Ex36s	CGCAATATTCTATATGAAAATACCAC	DMD_DHPLC_AMER_Ex36as	AGGAAGGAAAAAGGGAAAGATAAAG	415
DMD_DHPLC_AMER_Ex37s	TCGCTCACTTGCTCACTCGCTCTG	DMD_DHPLC_AMER_Ex37as	CAGAGTACTGCGCAACCTTC	338
DMD_DHPLC_AMER_Ex38as	CCGTAAGTGCTCCTATATTACCATACC	DMD_DHPLC_AMER_Ex38s	TGATTAGTTTAGCAACAGGAGGTTGA	428
DMD_DHPLC_AMER_Ex39s	AGAAAGGCTATGAGCACAGTATC	DMD_DHPLC_AMER_Ex39as	CTGAAGCAGATTTTATTAATGCAC	386
DMD_DHPLC_GLA_Ex40s	AAAAAGATGAGGGACGCAAA	DMD_DHPLC_GLA_Ex40as	CAAGGAAATGCATCAAATCAAA	379
DMD_DHPLC_GLA_Ex41s	CTTGCAAGTCGGTTGATGTG	DMD_DHPLC_GLA_Ex41as	CCAAAGTGAGGGAAACCACT	443
DMD_DHPLC_AMER_Ex42s	CAATTGTCAGCTGTAGAATGAGACC	DMD_DHPLC_AMER_Ex42as	TGAAGCCAACCACACTATCAAGTA	445
DMD_DHPLC_AMER_Ex43s	AGAATTGCAACACCATTTGCTACC	DMD_DHPLC_AMER_Ex43as	TGAGAGTGATACTTCTTTTTCCCTGTC	403
DMD_DHPLC_AMER_Ex44s	AATGTTGTGTGTACATGCTAGGTGTG	DMD_DHPLC_AMER_Ex44as	CATCACCCTTCAGAACCTGATCTTT	427
DMD_DHPLC_AMER_Ex45s	TGCCAGTACAACTGCATGTGGTAG	DMD_DHPLC_AMER_Ex45as	GCTTATAATCTCTCATGAAATATTC	362
DMD_DHPLC_AMER_Ex46s	GTTTGTGTCCCAGTTTGCATTAAC	DMD_DHPLC_AMER_Ex46as	GGCAGAAAACCAATGATTGAATTA	318
DMD_DHPLC_AMER_Ex47s	GGGGTGAGTGTTTCAGTCAATC	DMD_DHPLC_AMER_Ex47as	CATATAGCCAAAGCAAACGGTC	488
DMD_DHPLC_AMER_Ex48s	TAAACATTTTGGCTTATGCCTTGA	DMD_DHPLC_AMER_Ex48as	TGGTGCCTGTGCCTATTGTGTTAT	408
DMD_DHPLC_AMER_Ex49s	GCTACATTTTTGTGCCCTTA	DMD_DHPLC_AMER_Ex49as	GCAATGACTCGTTAATAGCCTTA	450
DMD_DHPLC_AMER_Ex50s	CACCAAATGGATTAAGATGTTCAT	DMD_DHPLC_AMER_Ex50as	CTCACCCAGTCATCACTTCATAGT	266
DMD_DHPLC_AMER_Ex51s	GGCTCTTTAGCTTGTGTTTC	DMD_DHPLC_AMER_Ex51as	GTGGGAAATGGTCTAGGAG	398
DMD_DHPLC_GLA_Ex52s	CAACGCTGAAGAACCCTGAT	DMD_DHPLC_GLA_Ex52as	TTGTGTGTCCCATGCTTGTT	193
DMD_DHPLC_GLA_Ex53s	AGAATTCCTCCAGACTAGCATTT	DMD_DHPLC_GLA_Ex53as	TTTCAGCTTTAACGTGATTTTCTG	392
DMD_DHPLC_AMER_Ex54s	GACACTCCAACTAGAGATTTC	DMD_DHPLC_AMER_Ex54as	CCCATTATTACAGCCAAC	268
DMD_DHPLC_AMER_Ex55as	TTGTCCCTGGCTTGTCAGTT	DMD_DHPLC_AMER_Ex55s	CCATCTTTCTCTTTTTATGGAGTT	361
DMD_DHPLC_AMER_Ex56s	TACGCCAAGAAAAGGGATTTGAGA	DMD_DHPLC_AMER_Ex56as	CCAGTTACTTGTGCTAAGACAATGAGG	423
DMD_DHPLC_AMER_Ex57s	ACACTTCTAGATATTCTGACATGG	DMD_DHPLC_AMER_Ex57as	GTCACTGGATTACTATGTGCTTAAC	267
DMD_DHPLC_AMER_Ex58s	GCACCCAGGATTAATTTTGAGAAGA	DMD_DHPLC_AMER_Ex58as	CCAGACCCTGGCAGCAAGAAT	277
DMD_DHPLC_AMER_Ex59s	CAGTAGGTTACCCTCTTGTTCAAC	DMD_DHPLC_AMER_Ex59as	GGGAAGATAACACTGCACTCAAGT	466
DMD_DHPLC_AMER_Ex60s	CCCTAAAGAGAATAAGCCCAGGTA	DMD_DHPLC_AMER_Ex60as	TCCTATCCTCACAAATATTACCATGAA	361
DMD_DHPLC_AMER_Ex61s	GTTGCTTTAGTGTTCTCAGTCTTGGA	DMD_DHPLC_AMER_Ex61as	GGATGATTTATGCTTCTACTGCATCTG	332
DMD_DHPLC_AMER_Ex62s	CCTGTTTGCGATGAATTTGACCTC	DMD_DHPLC_AMER_Ex62as	ACAGGTTAGTCACAATAAATGCTCTT	170
DMD_DHPLC_AMER_Ex63as	CAAGTAACTTTCACACTGCAAACT	DMD_DHPLC_AMER_Ex63s	GCAAAAATCATGTTGTTGTTATTG	243
DMD_DHPLC_GLA_Ex64s	AAATCACTGGGCGTCGG	DMD_DHPLC_GLA_Ex64as	TCTAAGCAAAGACATAGTATCAAGATC	206
DMD_DHPLC_GLA_Ex65s	AAGGAAGGTTTTACTCTTTGAGTC	DMD_DHPLC_GLA_Ex65as	CTAAGCCTCCTGTGACAGAGC	302
DMD_DHPLC_GLA_Ex66s	AAGGGTCTAGTAATTGTTTTCTGC	DMD_DHPLC_GLA_Ex66as	TCTAGAACTAGGGTAATTAGCCAAC	191
DMD_DHPLC_GLA_Ex67s	TTGGATGTCAGGTTCTGCTG	DMD_DHPLC_AMER_Ex67as	CCCATACCTACTGCCTACTGAAG	346
DMD_DHPLC_AMER_Ex68s	GATATACACCTCCTTTGCCATC	DMD_DHPLC_AMER_Ex68as	TGGCACAGGAGATAAAAGATCAAG	323
DMD_DHPLC_GLA_Ex69s	TCGAAGAAATACATACGTGTTTG	DMD_DHPLC_GLA_Ex69as	AGGTGAACTAACTCTCACGTCAG	192
DMD_DHPLC_GLA_Ex70s	TTAGTTTTGAAATCATCCTGTCC	DMD_DHPLC_GLA_Ex70as	ATACATCAAACAAGAGTGTGTTCTG	233
DMD_DHPLC_GLA_Ex71s	CTGAGAAAGCGTGTGTCTCC	DMD_DHPLC_GLA_Ex71as	AGGAAGGGGAATTAATATGTCC	218
DMD_DHPLC_GLA_Ex72s	GTGTGGTGGGTTTTTTCTCC	DMD_DHPLC_GLA_Ex72as	TATCATAGGTTAGCTTTCCTTGG	201
DMD_DHPLC_GLA_Ex73s	GACGTCACATAAGTTTTAATGAGC	DMD_DHPLC_GLA_Ex73as	CTACCTCTAAATCCCTCAAAGC	176
DMD_DHPLC_GLA_Ex74s	GAGTCCCTAACCCCCAAAG	DMD_DHPLC_GLA_Ex74as	TGCACTCTGCATACCAATGAC	267
DMD_DHPLC_AMER_Ex75 s	AGTGCCTCTTTTGCTTGCTGTTCT	DMD_DHPLC_AMER_Ex75 as	TAAGAATTGATGCTTGTTGCACCT	498
DMD_DHPLC_GLA_Ex76s	GTTTATTATGAAAAGTAATTCTGTTTTC	DMD_DHPLC_GLA_Ex76as	CGACTCTACCTTTCTTCAGACAAC	219
DMD_DHPLC_AMER_Ex77s	GTAATCATGGCCCTTTAATATC	DMD_DHPLC_AMER_Ex77as	CCCAAATTGCCATTCTGATACTGC	207
DMD_DHPLC_GLA_Ex78s	TCCCTTTCTGATATCTCTGCC	DMD_DHPLC_GLA_Ex78as	ATGATGACACAATGTGTAATACACAC	199
DMD_DHPLC_GLA_Ex79s	GAGTGATGCTATCTATCTATCTGCACC	DMD_DHPLC_GLA_Ex79as	TGTCTAATCCTCTTTGTTGTATGAA	224

TABLE 2.2 A list of amplicons with dHPLC temperatures and elution time shifts using Navigator Software.

Amplicon	PCR Program	dHPLC Temp	Time Shift	dHPLC Temp	Time Shift
DMD_DHPLC_GLA_5'-3-Exon_1	DMD_45	DMD Ex01@55	0	DMD Ex01@56	0
DMD_DHPLC_GLA_Ex2	DMD_45	DMD Ex02@54	0	DMD Ex02@55	0.5
DMD_DHPLC_GLA_Ex3	DMD_55	DMD Ex03@58	1	DMD Ex03@61	1.5
DMD_DHPLC_AMER_Ex4	DMD_55	DMD Ex04@56	0	DMD Ex04@59	0.8
DMD_DHPLC_GLA_Ex5	DMD_55	DMD Ex05@56	0	DMD Ex05@57	0.5
DMD_DHPLC_AMER_Ex6	DMD_45	DMD Ex06@56	0	DMD Ex06@59	1.5
DMD_DHPLC_GLA_Ex7	DMD_55	DMD Ex07@57	0	DMD Ex07@60	1.5
DMD_DHPLC_GLA_Ex8	DMD_55	DMD Ex08@56	0	DMD Ex08@59	1
DMD_DHPLC_GLA_Ex9	DMD_55	DMD Ex09@60	0.5	DMD Ex09@61	0.5
DMD_DHPLC_GLA_Ex10	DMD_45	DMD Ex10@57	0	DMD Ex10@58	0.5
DMD_DHPLC_GLA_Ex11	DMD_55	DMD Ex11@57	0	DMD Ex11@59	1.5
DMD_DHPLC_GLA_Ex12	DMD_55	DMD Ex12@57	1	DMD Ex12@58	1
DMD_DHPLC_AMER_Ex13	DMD_45	DMD Ex13@55	0.5	DMD Ex13@58	1.5
DMD_DHPLC_AMER_Ex14and15	DMD_45	DMD Ex14/15@55	0	DMD Ex14/15@58	1
DMD_DHPLC_AMER_Ex16	DMD_45	DMD Ex16@57	0	DMD Ex16@59	1
DMD_DHPLC_GLA_Ex17	DMD_55	DMD Ex17@57	0	DMD Ex17@59	1
DMD_DHPLC_AMER_Ex18	DMD_45	DMD Ex18@55	0.5	DMD Ex18@59	2
DMD_DHPLC_AMER_Ex19	DMD_45	DMD Ex19@56	0.8	DMD Ex19@60	2
DMD_DHPLC_AMER_Ex20	DMD_45	DMD Ex20@58	0.8	DMD Ex20@59	0.8
DMD_DHPLC_GLA_Ex21	DMD_55	DMD Ex21@58	0	DMD Ex21@59	0.5
DMD_DHPLC_GLA_Ex22	DMD_55	DMD Ex22@59	0.5	DMD Ex22@60	1.4
DMD_DHPLC_GLA_Ex23	DMD_45	DMD Ex23@58	1	DMD Ex23@59	1
DMD_DHPLC_GLA_Ex24	DMD_55	DMD Ex24@58	0.8		
DMD_DHPLC_GLA_Ex25	DMD_55	DMD Ex25@58	0.5	DMD Ex25@59	1
DMD_DHPLC_GLA_Ex26	DMD_55	DMD Ex26@55	0	DMD Ex26@58	0.5
DMD_DHPLC_AMER_Ex27	DMD_45	DMD Ex27@56	0	DMD Ex27@59	1
DMD_DHPLC_AMER_Ex28	DMD_55	DMD Ex28@56	1.3	DMD Ex28@58	1.5
DMD_DHPLC_GLA_Ex29	DMD_55	DMD Ex29@57	0.5	DMD Ex29@58	1
DMD_DHPLC_AMER_Ex30	DMD_45	DMD Ex30@60	1		
DMD_DHPLC_GLA_Ex31	DMD_55	DMD Ex31@57	0	DMD Ex31@58	0.7
DMD_DHPLC_GLA_Ex32	DMD_55	DMD Ex32@56	0.5	DMD Ex32@58	1.5
DMD_DHPLC_GLA_Ex33	DMD_45	DMD Ex33@57	1	DMD Ex33@58	1
DMD_DHPLC_GLA_Ex34	DMD_45	DMD Ex34@58	1		
DMD_DHPLC_GLA_Ex35	DMD_45	DMD Ex35@57	0.5	DMD Ex35@58	0.5
DMD_DHPLC_AMER_Ex36	DMD_45	DMD Ex36@56	0.5	DMD Ex36@57	1
DMD_DHPLC_AMER_Ex37	DMD_45	DMD Ex37@60	0	DMD Ex37@59	0
DMD_DHPLC_AMER_Ex38	DMD_45	DMD Ex38@55	0.5	DMD Ex38@57	1
DMD_DHPLC_AMER_Ex39	DMD_45	DMD Ex39@55	0	DMD Ex39@56	0.8
DMD_DHPLC_GLA_Ex40	DMD_45	DMD Ex40@56	0.7	DMD Ex40@57	1

TABLE 2.2 (continued) A list of amplicons with dHPLC temperatures and elution time shifts using Navigator Software.

Amplicon	PCR Program	dHPLC Temp	Time Shift	dHPLC Temp	Time Shift
DMD_DHPLC_GLA_Ex41	DMD_55	DMD Ex41@59	0	DMD Ex41@62	1.5
DMD_DHPLC_AMER_Ex42	DMD_45	DMD Ex42@56	0	DMD Ex42@57	0
DMD_DHPLC_AMER_Ex43	DMD_45	DMD Ex43@57	1	DMD Ex43@59	2
DMD_DHPLC_AMER_Ex44	DMD_45	DMD Ex44@56	0.5	DMD Ex44@58	1.5
DMD_DHPLC_AMER_Ex45	DMD_45	DMD Ex45@59	0.5		
DMD_DHPLC_AMER_Ex46	DMD_45	DMD Ex46@56	0	DMD Ex46@55	0.5
DMD_DHPLC_AMER_Ex47	DMD_45	DMD Ex47@58	0		
DMD_DHPLC_AMER_Ex48	DMD_45	DMD Ex48@57	0	DMD Ex48@55	0.5
DMD_DHPLC_AMER_Ex49	DMD_45	DMD Ex49@57	0	DMD Ex49@58	0
DMD_DHPLC_AMER_Ex50	DMD_45	DMD Ex50@58	0	DMD Ex50@60	0.8
DMD_DHPLC_AMER_Ex51	DMD_45	DMD Ex51@59	0	DMD Ex51@56	0.8
DMD_DHPLC_GLA_Ex52	DMD_55	DMD Ex52@59	0	DMD Ex52@60	0.5
DMD_DHPLC_GLA_Ex53	DMD_55	DMD Ex53@58	1	DMD Ex53@59	1.5
DMD_DHPLC_AMER_Ex54	DMD_55	DMD Ex54@57	0.5	DMD Ex54@61	2
DMD_DHPLC_AMER_Ex55	DMD_45	DMD Ex55@59	0.5	DMD Ex55@60	0.5
DMD_DHPLC_AMER_Ex56	DMD_45	DMD Ex56@58	0.5	DMD Ex56@59	1.3
DMD_DHPLC_AMER_Ex57	DMD_45	DMD Ex57@61	0	DMD Ex57@62	1
DMD_DHPLC_AMER_Ex58	DMD_45	DMD Ex58@58	0.3	DMD Ex58@60	1
DMD_DHPLC_AMER_Ex59	DMD_55	DMD Ex59@61	0.5	DMD Ex59@63	1.8
DMD_DHPLC_AMER_Ex60	DMD_45	DMD Ex60@60	1	DMD Ex60@62	1
DMD_DHPLC_AMER_Ex61	DMD_45	DMD Ex61@58	1.5	DMD Ex61@60	1.5
DMD_DHPLC_AMER_Ex62	DMD_45	DMD Ex62@60	0.5	DMD Ex62@63	3
DMD_DHPLC_AMER_Ex63s	DMD_45	DMD Ex63@58	0.5	DMD Ex63@60	1
DMD_DHPLC_GLA_Ex64	DMD_45	DMD Ex64@59	0.5	DMD Ex64@60	1.5
DMD_DHPLC_GLA_Ex65	DMD_55	DMD Ex65@59	0	DMD Ex65@60	0.5
DMD_DHPLC_GLA_Ex66	DMD_45	DMD Ex66@57	0.5	DMD Ex66@59	0.8
DMD_DHPLC_GLA_AMER_Ex67	DMD_45	DMD Ex67@59	0.5	DMD Ex67@60	1
DMD_DHPLC_AMER_Ex68	DMD_55	DMD Ex68@60	0.5	DMD Ex68@63	2
DMD_DHPLC_GLA_Ex69	DMD_45	DMD Ex69@58	0	DMD Ex69@60	1
DMD_DHPLC_GLA_Ex70	DMD_45	DMD Ex70@58	0	DMD Ex70@61	1.5
DMD_DHPLC_GLA_Ex71	DMD_45	DMD Ex71@60	0.5	DMD Ex71@61	1
DMD_DHPLC_GLA_Ex72	DMD_55	DMD Ex72@59	0.5	DMD Ex72@61	2
DMD_DHPLC_GLA_Ex73	DMD_45	DMD Ex73@56	0	DMD Ex73@57	0.5
DMD_DHPLC_GLA_Ex74	DMD_55	DMD Ex74@57	0	DMD Ex74@60	1
DMD_DHPLC_AMER_Ex75	DMD_55	DMD Ex75@61	1.2	DMD Ex75@63	1.7
DMD_DHPLC_GLA_Ex76	DMD_45	DMD Ex76@60	0	DMD Ex76@63	1
DMD_DHPLC_AMER_Ex77	DMD_45	DMD Ex77@59	0	DMD Ex77@61	0.3
DMD_DHPLC_GLA_Ex78	DMD_45	DMD Ex78@57	0	DMD Ex78@60	0.5
DMD_DHPLC_GLA_Ex79	DMD_45	DMD Ex79@57	0	DMD Ex79@59	0.5

2.5 **Semi-automated sequencing of PCR products**

Sequencing was performed on selected amplicons from patients. Genomic DNA was amplified by PCR as described in method 2.4.2. 3µl of this reaction and 2µl of sucrose loading mix was loaded onto a 2% agarose gel and checked for size and intensity alongside 4µl of 1Kb ladder (GibcoBRL). The remaining 22µl of the PCR was added to 4u of Shrimp Alkaline Phosphatase - which dephosphorylates 5' phosphates on nucleotides (Promega), 8u of Exonuclease I (New England Biolabs) - which removes nucleotides from single stranded DNA and 4µl of 10X *Exo* I buffer (supplied with enzyme). This reaction was heated in a PCR machine to 37°C for 30 minutes and then the enzymes were deactivated by incubation at 72°C for 15 minutes. 2µl of the reaction products were then added to a 10µl sequencing reaction, one containing the sense primer and one containing the antisense primer - for amplicon under investigation. Sequencing reactions used the following conditions in 200µl microtubes (Abgene).

2µl (SAP/*Exo* I treated) PCR product
2µl BigDye Version 3 (Applied Biosystems)
2µl of Better Buffer (Microzone Ltd)
3µl of sH_2O
1µl of primer (5µM)

Cycled on a GRI PTC-0200 Thermocycler through 25 cycles of

96°C 15 seconds
50°C 5 seconds
60°C 4 minutes

The products of this reaction were transferred to a 96 well MicroAmp®plate (Applied Biosystems) and precipitated out of solution by addition of 5µl 125mM EDTA and 60µl of 100% ethanol and then left for 45 minutes. The products were then pelleted at the bottom of the wells by centrifuging the plate at 1500g for 45 minutes at 10°C. The supernatant was removed by inversion of the plate and 150µl of 70% ethanol added to wash the pellet. The pellet was then centrifuged again at 1500g for 20 minutes at 10°C, the supernatant removed by inversion and the pellet allowed to air dry for 1 hour. Finally the pellet was resuspended in 12µl of Hi-Dye formamide (Applied Biosystems). The products were then injected onto an Applied Biosystems 3100 capillary automated sequencer using POP-4 matrix (Applied

Biosytems) to resolve and detect the products. The data generated was automatically analysed by the machine's software to produce a trace file containing normalised sequence.

2.6 **Staden package analysis of sequence data**

The Staden package V1.4 (http://staden.sourceforge.net) was used to assemble the sequence traces into a database for mutation analysis (Bonfield *et al.*, 1998). For each patient a database was constructed using Pregap4 containing the EMBL format reference file described in method 2.4.1. Sequencing files from both the patient and normal control were clipped to quality level 20 (Phred 1 error in 1000) and then added into this database, the software aligned the sequences against the reference trace in both sense and antisense directions. Mutation analysis was performed by visual inspection of the trace subtraction line - this represents the difference between the sample and control traces. Each exon was visually inspected for mutation, twice in each direction.

2.7 **dHPLC methods**

Elution gradients for each amplicon were determined by calculation of amplicon length and sequence content using the Navigator software (Transgenomic). The elution profile was empirically checked for each amplicon, adjusting the time shift parameter to ensure the fragment eluted between 1.5 and 2.5 minutes. Table 2.2 lists the methods and time shift used for each amplicon.

2.7.1 Manual heteroduplexing and dHPLC analysis in Dp71

Amplicons for exons 63-79 were generated as detailed in method 2.4 using the primers listed in table 2.1. Each amplicon amplified from a patient needed to be duplexed with another allele before it could be analysed by dHPLC, thus a PCR from normal control DNA was always amplified alongside the patient's sample. Cycling conditions were as shown in Table 2.2. 3µl of PCR products plus 2µl of sucrose loading buffer were run on a 2% agarose gel and the remainder of the patient and control samples were mixed together in a 96 well thermosprint® plate (Thermo Electron Corporation) in equal quantities based on the

intensity of the products on the check gel. The plate was then sealed with a rubber mat (Thermo Electron Corporation) placed in a thermocycler and cycled though a programme that heats the products and then slowly cools them down encouraging the formation of heteroduplexes between different alleles.

Programme for forming heteroduplexes:

95°C for 5 minutes then 90 cycles of 95°C for 20 seconds (-0.8°C at every cycle)

The plate was then unsealed and 30µl of dHPLC low temp standard (Transgenomic) was added to an empty well, this was then placed in the loading tray of a 3500HT dHPLC machine (Transgenomic).The dHPLC machine was controlled by a pre-defined worksheet using Navigator software (Transgenomic). The software specified the method (Table 2.2) which was automatically applied to an injection of PCR products from the plate. The machine requires a constant supply of four solutions made up with dHPLC grade H_2O in 2 litre volumes:

0.1mM TEAA (Transgenomic)
0.1mM TEAA:25% Acetonitrile (Rathburn, UK)
75% Acetonitrile (Rathburn, UK)
8% Acetonitrile (Rathburn, UK)

Data from the dHPLC machine was stored on the controlling computer and was analysed after the samples had run through the machine. dHPLC profiles from all the samples were compared against each other, amplicon by amplicon using the Navigator software.

2.7.2 Robotic PCR and dHPLC of Exon 70

An MWG RoboAmp4200 was used to automatically amplify dystrophin exon 70 from 46 patient DNA samples. A single arm with a lid remover was used as well as liquid sensing disposable tips that were automatically replaced after every step involving a change of solution. The robot aliquoted 25µl of a mastermix containing: 240µl sense primer (5µM), 240µl antisense primer (5µM) and 1700µl megamix-w (Microzone Ltd) into all the wells of a 96 well thermosprint® plate (Thermo Electron Corporation). The robot then aliquoted

300ng of DNA from 46 individual patient samples each into a separate well and then aliquoted 300ng of DNA from a single control DNA sample into a further 46 wells on the plate. The plate was then transferred to the robot's thermocycler and cycled through the DMD_45 PCR programme, once finished 3µl of sucrose loading mix was aliquoted into a 96 well thermosprint® plate and a 4µl aliquot of each PCR reaction transferred onto the plate. The prepared products were then run out on a 96 well Electrofast® (Abgene) gel to check for amplification efficiency. The robot then mixed 24µl of control PCR with each patient PCR into a 96 well thermosprint® plate this plate was heated and cooled in the robot thermocycler as in method 2.7.1. 30µl of dHPLC low temp mutation standard (Transgenomic) was added to an empty well on the plate and this was then placed in the loading tray of the 3500HT dHPLC machine (Transgenomic). 5µl of each sample was injected under the two elution methods specific for exon 70 (Table 2.2).

2.8 **Robotic PCR and dHPLC of dystrophin exons 1-79**

2.8.1 Robotic PCR

An MWG RoboAmp4200 was used to automatically amplify all amplicons listed in table 2.2. A single arm with a lid remover was used as well as liquid sensing disposable tips that were automatically replaced after every step involving a change of solution. The process is summarised in Figure 3.3.1 on page 70. Two reagent plates were used: one containing 150µl of each sense primer (5µM) aliquoted into a separate well on a 96 well thermosprint® plate (Thermo Electron Corporation) and the other 96 well thermosprint® plate with 150µl of each antisense primer (5µM) in the same orientation. The robot was programmed to transfer 25µl of negative control from 1500µl of megamix-w (Microzone Ltd) to an empty well in a new 96 well NCC PCR plate. The robot then mixed 500ng of the sample DNA with the mastermix, which was then aliquoted in 23µl volumes into 50 of the 96 wells on the plate. The robot then aliquoted out 3µl of the sense primers – only those optimised for DMD_45 – individually into the wells containing DNA and mastermix on the PCR plate, followed by the same process for the antisense primers. The plate was then transferred to the PCR block and cycled through the DMD_45 PCR programme, once finished the plate was removed. The second round of PCR repeated the above process, only with 800µl

megamix-w being distributed into 28 wells and using the primers optimised for DMD_55, this plate was cycled through the DMD_55 programme. Once finished 3µl of sucrose loading mix was aliquoted into a 96 well thermofast plate (Web scientific) and a 4µl aliquot of each PCR reaction transferred in gene order onto the plate. This was then run out on a 96 well Electrofast® (Abgene) gel (Method 2.3) to check for amplification efficiency. This entire process was repeated to amplify the control DNA.

2.8.2 Heteroduplexing and dHPLC

24µl of sample PCR products and 24µl of control PCR products for each amplicon were mixed in turn by the robot and placed onto a 96 well thermofast plate (Web scientific). This plate was then was then heated and cooled in the robot thermocycler (Method 2.7.1). 30µl of dHPLC low temp standard (Transgenomic) was added to an empty well on the plate and this was then placed in the loading tray of the dHPLC machine (Transgenomic). The Navigator software (Transgenomic) was programmed to make injections of 17µl using the elution methods defined for each amplicon (Table 2.2). Traces were analysed using the Navigator software normalisation option.

2.9 **Polyacrylamide gel electrophoresis**

2.9.1 Resolving products on a polyacrylamide gel

8% Polyacrylamide gels were made using the Bio-Rad Protean apparatus with 1mm spacers and a 20 well comb. 30ml of dH_2O, 10ml of 19:1 Acrylamide (National Diagnostics), 10ml 5xTBE, 250µl of freshly made ammonium persulphate (APS) and 40µl of N,N,N-tetramethylethylenediamine (TEMED) were mixed and the gel quickly poured into the gel frame. Once polymerised the gel was immersed in 1xTBE within its running tank. 4µl of samples were mixed with 3µl of sucrose loading mix and loaded onto the gel and electrophoresed for between 2.5 and 6 hours at a constant 340V (Table 2.3).

2.9.2 Silver staining

After electrophoresis silver staining was carried out. The gel was fixed in a solution of 10% ethanol/0.5% acetic acid for 10 minutes, the solution was discarded and the gel was stained in 0.1% silver nitrate for 15 minutes, after which it was briefly rinsed twice in dH_2O. The bands were developed in a solution of 1.5% sodium hydroxide/0.1% formaldehyde for 20 minutes or until the bands were of the desired intensity. The gel was fixed in 0.75% sodium carbonate for 10 minutes, sealed in a plastic bag and photographed for storage.

DMD LINKED MARKERS

MARKER	LOCATION	A.T.	SIZE RANGE	ACRYL.GEL	FORWARD PRIMER	REVERSE PRIMER
DYS I	3.5 KB 5' EX.1 (PROMOT	55	177-185 BP	350V 4HRS	ACTGTAAATGAAATTGTTTTCTAAGTGCC	CTTAACAAAATGTCCTTCAGTTCTATCC
DYS II B	1.2 KB 5' EX.1 (PROMOT	55	80-94	350V 3HRS	TGAGTACTTGCACACAAAGC	TAGTGTTTTCCTAAGGGGTT
5' 5N1	INTRON 1	55	210-220	350V 6HRS	CACGACAGATTATATGAGTCAT	AAGCCTTATGAATACTGGCAT
STR 7A	INTRON 7	55	218-235	350V 6HRS	TTCTGGTTTTCTGGTCTG	GAATCAATCTCTCTGTCAAG
STR 7B	INTRON 7	55	227-241	350V 6HRS	AGCTATTATCTGAGAAAGTC	TGAGGTGAATTTATTAAGGG
87-8 TAQI	INTRON 13	55	155>145/74+71 BP	AGAROSE	GTCAGTTGGTCAGTAAAAGCC	CAGATCAGTCGACCAATTAAAACCACAGCAG
87-15 XMNI	INTRON 17	55	740>730/520+210 BP	AGAROSE	GACTGGAGCAAGGGTCGCC	CTGATGAACAATTTCCCTTTCATTCCAG
5' 7N4	INTRON 25	55	163-169	350V 4HRS	GTGAAGCTACAAAAATATTAGAG	CAACAATATCTCACCATACTTG
STR 45	INTRON 45	50	156-184	350V 4HRS	GAGGCTATAATTCTTTAACTTTGGC	CTGTTTCCCTCTTTATTCATGTTAC
E48 MSE1	EXON 48	55	249>108/85+23 BP	AGAROSE	AAGCTTGAAGACCTTGAAGAGC	CCTGAATAAAGTCTTCCTTACCACAC
DXS 997	INTRON 48	55	109-117	350V 3HRS	TGGCTTTATTTTAAGAGGAC	GTTTTCAGTTTCCTGGGT
STR 50	INTRON 50	60	233-251	350V 6HRS	AAGGTTCCTCCAGTAACAGATTTGG	TATGCTACATAGTATGTCCTCAGAC
DXS 1036	INTRON 51	55	145-151	350V 4HRS	TGCAGTTTATTATGTTTCCACG	GCCATTGATAAGTGCCAGAT
1C-2 CA	INTRON 59	50	245-255	350V 6HRS	TGTCTGTCTTCAGTTATATG	ATAACTTACCCAAGTCATGT
J66	INTRON 60	61	1.25, 1.2, 1.1 KB	AGAROSE	GCAGCTATATGTTTCCCAAGATTGA	GAGGTTCTTTGGAGGAATAC
3' 19N8	INTRON 63	60	140-150	350V 4HRS	AGCCCCATTCTGTACATCAAAT	AACGACTTCCCCCACTCTGT
STR H1	INTRON 64	50	90-102	350V 3HRS	ACGACAAGAGTGAGACTCTG	ATATATCAAATATAGTCACTTAGG
3'DYS	3' UTR	55	127-135	350V 3HRS	GAAAGATTGTAAACTAAAGTGTGC	GGATGCAAAACAATGCGCTGCCTC

Table 2.3
Selected intragenic dystrophin microsatellite markers and SNPs

Shown is marker name, location, Annealing Temperature for PCR (A.T.), size in base pairs, time to run on an gel and the primer sequences used.

2.10 **Restriction enzyme analysis**

Restriction sites were mapped within an amplicon using a web based programme, webcutter (http://www.firstmarket.com/cutter/cut2.html). Primers used to generate a restriction site in cases where no natural site was present are listed below:

Primer	Sequence (mismatch highlighted)	Enzyme	PCR
Ex70as_10171C>T	ACAGTCTGCACTGGCAGGTAGCCCACTC	*Dde* I	55
Ex36as_5089 C>T	CTCTGATTCATCCAAAAGTGTGTCAGCTT	*Tru9* I	45
Ex11s_1150-1 G>T	GTTCTAATTACAATTGTTAACTTCCTCC	*Bsl* I	45
Ex23as_3036_3037delinsC	CAAATTCTGATTGATATTTCCGGCTAATCT	*Dde* I	45
Ex24as_3201delT	GGCCATTCCTCCTTCAGAAAAACATTA	*Tru9* I	45

Table 2.4 A list of the primers used to incorporate restriction site in amplicons to differentiate between the normal and mutant alleles. Shown are the primer names, sequences (with mismatched bases in grey) the enzyme used and the PCR programme (Method 2.4.2)

Restriction enzyme digestion of PCR products was performed using a variety of enzymes (New England Biolabs). Enzyme reactions were set up in 500µl tubes (Abgene) as follows:

10µl PCR product
2µl Buffer (New England Biolabs)
1µl Enzyme (New England Biolabs)
5µl sH_2O
2µl Bovine Serum Albumin (New England Biolabs) or sH_2O

Digests were incubated at either 37˚C or 65˚C overnight.

2.11 **Mitochondrial sequencing**

Primers to amplify the HV1 region (16009-16390) were taken from (Pfeiffer *et al* 1999)

F15971: (TTAACTCCACCATTAGCACC)
R16410: (GAGGATGGTGGTCAAGGGAC)

PCR conditions were as follows:

18µl Megamix-without BSA – (Microzone ltd)
2.5µl F15971 primer (5µM)
2.5µl R16410 primer (5µM)
2µl DNA (50ng/µl)

95°C	3min
30 cycles of	
94°C	30sec
55°C	30sec
72°C	45sec

The PCR programme was run on a GRI PTC-0200 thermocycler. Sequencing was then conducted as in (Method 2.5). The revised cambridge reference sequence was obtained from http://www.mitomap.org/mitomap/mitoseq.html.

2.12 **Splice site analysis and ESE identification**

2.12.1 Splice site analysis

Splice site analysis was performed using a web based programme http://home.snafu.de/probins/Splice/splice.html; the mutant sequence was given a score based on the consensus sequence defined by Shapiro and Senapathy (1987).

2.12.2 ESE analysis

ESE site analysis was performed according to Cartegni *et al.* (2003). The sequence was screened for consensus ESE motifs using the described methods and programme.

2.13 SNP analysis

Entrez SNP (http://www.ncbi.nlm.nih.gov/projects/SNP/) was used to determine the location and identity of putative Single Nucleotide Polymorphisms (SNPs). The primers designed for amplification of these SNPs are listed below:

dbSNP:1317642 F-TTTGAAGTGCTGGCAAATGT
dbSNP:1317642 R-CACAATGAAAGGCAAGGTGA

dbSNP:2178539 F-GGAGGGTGGCTTCATACAAA
dbSNP:2178539 R-ACCCCAGCAACTCTCCTACC

dbSNP:1921392 F-CTTTTGGATTAACCTTAGGAA
dbSNP:1921392 R-GGCTAAACCAGGGTCCTCTC

PCR and sequencing was as in (method 2.11 and 2.5)

2.14 Robotic sequencing of the dystrophin gene in control DNA

Amplification of dystrophin exons 1-79 was performed as described in method 2.8. The robot aliquoted to each well 8µl of exonuclease mix and incubated the reaction as in method 2.5. The robot then aliquoted 7µl of a sequencing mix consisting of 200µl Bigdye V.3 (Applied Biosystems) 200µl Better Buffer (Microzone Ltd) and 300µl sH_2O into 78 wells of a 96 well thermosprint® plate (Thermo Electron Corporation). Then 2µl of sense primers (5µM) were aliquoted into the plate followed by 2µl of the PCR products in the same orientation as the primers. This was then put though the sequencing cycles in the robot's thermocycler and prepared for sequencing and analysis as previously described (Method 2.5).

2.15 Microsatellite and SNP analysis

Primers for the microsatellite and SNP alleles analysed are listed in Table 2.3, including the annealing temperature required for amplification modifying method 2.11 and the time required for PAGE analysis (Method 2.9). SNP alleles in this group were tested were analysed on a 1% agarose gel (Method 2.3).

2.16 **Patients and controls**

In total 423 pedigrees have been identified in a database of DNA collected from affected members of Scottish families with DMD and BMD from 1982-2004. All individuals had contributed samples for testing at the Department of Medical Genetics, Yorkhill with the intention of identifying a causative genetic mutation. Patients with a detectable deletion or duplication using a routine fluorescent multiplex assay (Yau *et al.*, 1996) were excluded, this identified a group of 76 pedigrees as potential cases for point mutation analysis. From this group 46 patients were chosen for analysis of exon 70 alone on the basis of a referral for DMD/BMD testing being made at a regional neurology/genetics clinic. An additional subset of 10 patients were selected based on the individual having learning difficulties of a degree to warrant special schooling. In Scotland this correlates with an IQ score of approximately 70 or less. Finally 18 samples were chosen from the initial group of 76 for point mutation analysis of the Dp427m coding sequence based on a confirmed clinical diagnosis of DMD, including serum creatine kinase activity over 9000 and/or positive family history. 16 control samples were taken from consenting healthy individuals all living in the West of Scotland.

CHAPTER THREE

RESULTS

3.0 Establishing protocols for dHPLC analysis of dystrophin exons 1-79

Each of the exons in the dystrophin gene was amplified by PCR; amplicons contained the full exonic coding sequence and flanking intronic sequence. Primers were designed to amplify at least 40bp of flanking intronic sequence to allow efficient sequencing of the coding region (Method 2.4). Each primer pair was subjected to PCR of three different DNA samples, performed under two conditions, the resulting products were evaluated against each other on an agarose gel. The most efficient of the reactions was used in all subsequent reactions. When amplicons failed or were of low yield an alternative set of primers were designed. Some primers were taken from the publication by Bennet *et al.* (2001) and the remainder were designed using the web-based programme Primer3 according to the default criteria (Method 2.4). Each amplicon sequence was then entered into the Navigator software and dHPLC elution gradients were chosen for optimal detection of mutation within the coding region (Table 2.2). dHPLC profiles from homozygous (single allele) samples were saved to a database to form a consensus profile for each exon, this was then used to compare against the sample profiles for the presence of any variance (Method 2.8.2).

A robotic protocol was devised to amplify all the dystrophin gene amplicons using two PCR conditions, extensive practice with water coloured with loading buffer in place of reagents ensured that the robot consistently transferred the required volumes for each step. When using PCR reagents and control DNA the robot produced products for all amplicons (Method 2.8). These products were subsequently used to evaluate the elution gradients theoretically determined by the Navigator software. Automated sequencing was used to confirm that all dystrophin exons were correctly and specifically amplified by the robotic PCR process. A second robotic programme was devised to mix the PCR products from two samples together and then heat and slowly cool the mixes (Method 2.8.2). A spreadsheet was also produced to contain all the dHPLC parameters required for analysis of the robotic PCR products. Following this a sample could be robotically amplified and mixed with PCR from a control sample before being injected into the dHPLC machine using a set template.

Sequencing analysis was designed to use the Staden package of programmes. This aligned sequencing traces from a control and patient sample against a reference sequence of the entire dystrophin coding region and flanking sequences. In order to test

the sensitivity of dHPLC: exons 63-79 of the dystrophin gene from ten samples were manually amplified and analysed by automated dHPLC, subsequent sequencing and trace subtraction analysis showed that dHPLC detected all the changes observed by trace subtraction in the coding region of all samples (Results 3.1).

The above protocols were used to analyse the dystrophin coding region in an initial group of four individuals with known dystrophin mutations (Results 3.3). Shifts observed by dHPLC were noted, these were then sequenced and analysed by trace comparison against a normal control sequence using the Staden package. Each of the four mutations were correctly identified amongst a total of twenty five nucleotide changes observed in this group. Based on this a total of 23 patients with a diagnosis of DMD or BMD were analysed for variance by dHPLC. This protocol detected a potential mutation in 96% of the samples, only one sample failed to show any evidence of a mutation. If possible, a restriction digest was designed to confirm a mutation result with the introduction of a restriction site (by mismatch in an alternative primer) being required in some cases (Figure 3.2.7 and Methods 2.10). All digests showed the predicted patterns of bands based on the mutation result. Sequence flanking the mutation was analysed for motifs believed to affect mutagenesis. An adaptation of the above methods was also used to test the theory that cognitive impairment may be associated with a point mutation in Dp71 in a cohort of ten individuals with DMD/BMD (Results 3.1).

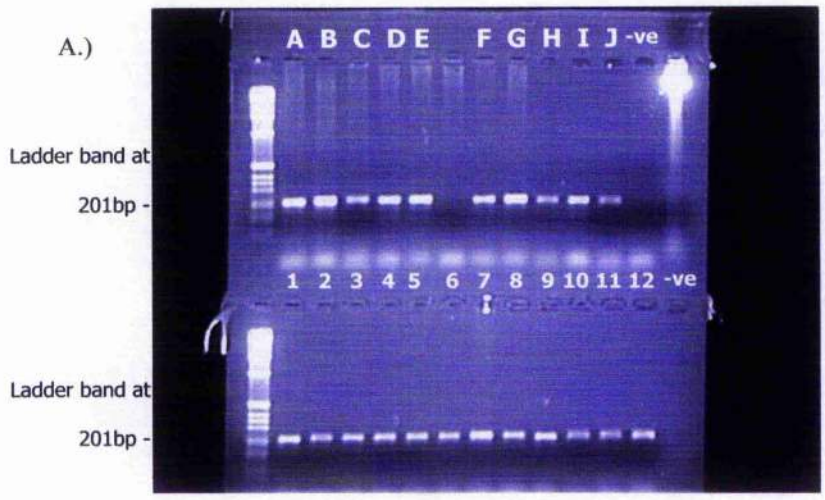
The inheritance of a recurrent mutation was analysed using published primers for a number of well characterised intragenic polymorphisms and some newly identified SNPs. The repeat polymorphisms were analysed using polyacrylamide gel analysis followed by silver staining using established methods (Method 2.9). The SNPs were analysed by restriction digest and agarose gel if already published or direct sequencing if unpublished (Methods 2.13,2.15). The mitochondrial HV1 region was also sequenced in three patients by a method adapted from Pfeiffer *et al.*, (1999) (Method 2.1).

3.1 **Point mutation in Dp71 and cognitive impairment**

3.1.1 PCR and dHPLC of the dystrophin exons 63-79

As the samples being examined were from patients with a muscular dystrophy, the region of interest in this study constitutes the coding regions of Dp71 that are shared with the muscle transcript Dp427m (exons 63-79). Primers for PCR were designed according to method 2.4 and optimised to one of two PCR programmes. PCR design was performed with the aim of simplifying subsequent automation of the process. In total 17 amplicons were designed, containing exons 63 to 79 of the Dp427m transcript. Using these primers PCR products from 10 patient DNA samples with cognitive impairment (Method 2.16) and 12 reactions for a normal control were produced for each amplicon (Method 2.7.1). An aliquot of each PCR reaction was run out on an agarose gel to check reaction quality within the run. Figures 3.1.1 and 3.1.2 show photographs of the agarose gels of PCR products from patient and control sample DNAs. No PCR fails were detected at this stage in the analysis. The samples were then mixed, heated then cooled to allow heteroduplexes to form and injected onto a dHPLC column using the predetermined parameters for mutation analysis of the coding region contained within the fragment (Method 2.7.1). Figure 3.1.1. shows the results of dHPLC analysis of the mixed PCRs from the 10 patients for an amplicon containing exon 79.

Two recognised polymorphisms are known to exist within exons 63-79, these are: 9649+15C>T and 9974+13_9974+23(A)9-12, their frequency in the UK is unknown but is high in other populations. Given this, the number of polymorphisms in exons 63-79 is expected to be low, thus dHPLC profiles were compared against each other within the group as a measure of variance. dHPLC analysis of amplicons for exons 63-79 using the methods described (Method 2.7) showed that exon 70 was the only fragment to show shifts using dHPLC in this group of 10 patients - labelled A to J (Figure 3.1.2).



B.)

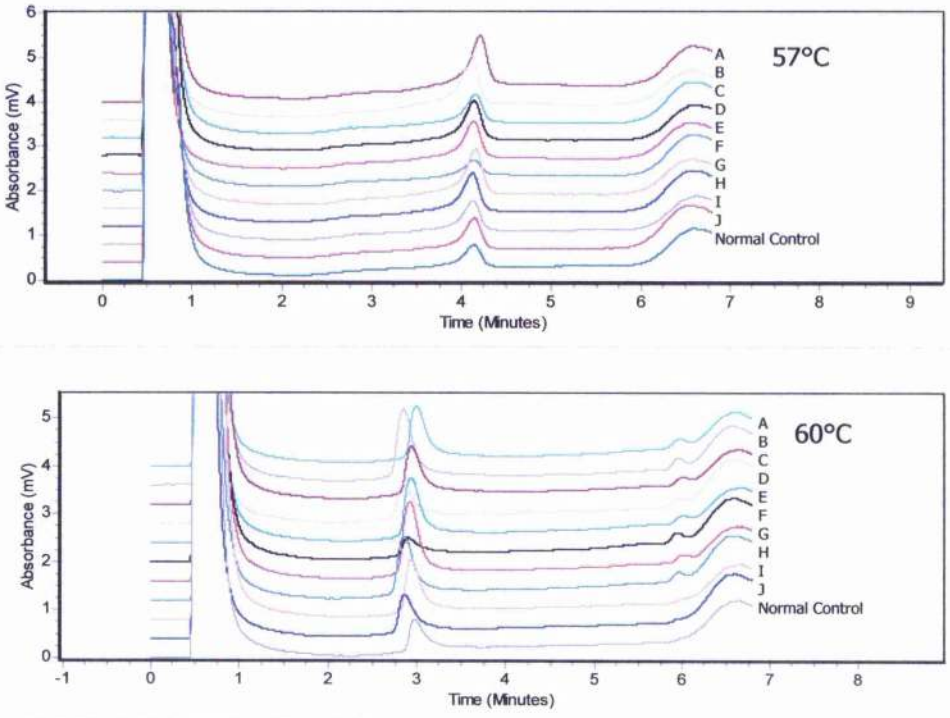


Figure 3.1.1 PCR and dHPLC analysis of exon 79 for patients A-J. A.) A 2% agarose gel photograph showing the 224bp PCR products obtained using primers for exon 79, 3µl of product plus 2µl of sucrose loading mix was run in each lane alongside 4µl of 1kb ladder. In the upper half are PCR products for patients A-J and in the lower half are PCR products for exon 79 from 12 reactions using the same control DNA. Each patient PCR was then combined with an aliquot of control DNA in equal quantities based on the intensity on the agarose gel, this mix was heated and then slowly cooled to allow heteroduplexes to form. 5µl of each PCR mix was then injected into the dHPLC machine. B.) Shown are the dHPLC profiles for each of the patient samples mixed with the control PCR (plus an unmixed control homoduplex profile) under two separate denaturing conditions, 57°C and 60°C. Samples are compared against each other for differences within the group – no shifts in exon 79 were observed under these two conditions.

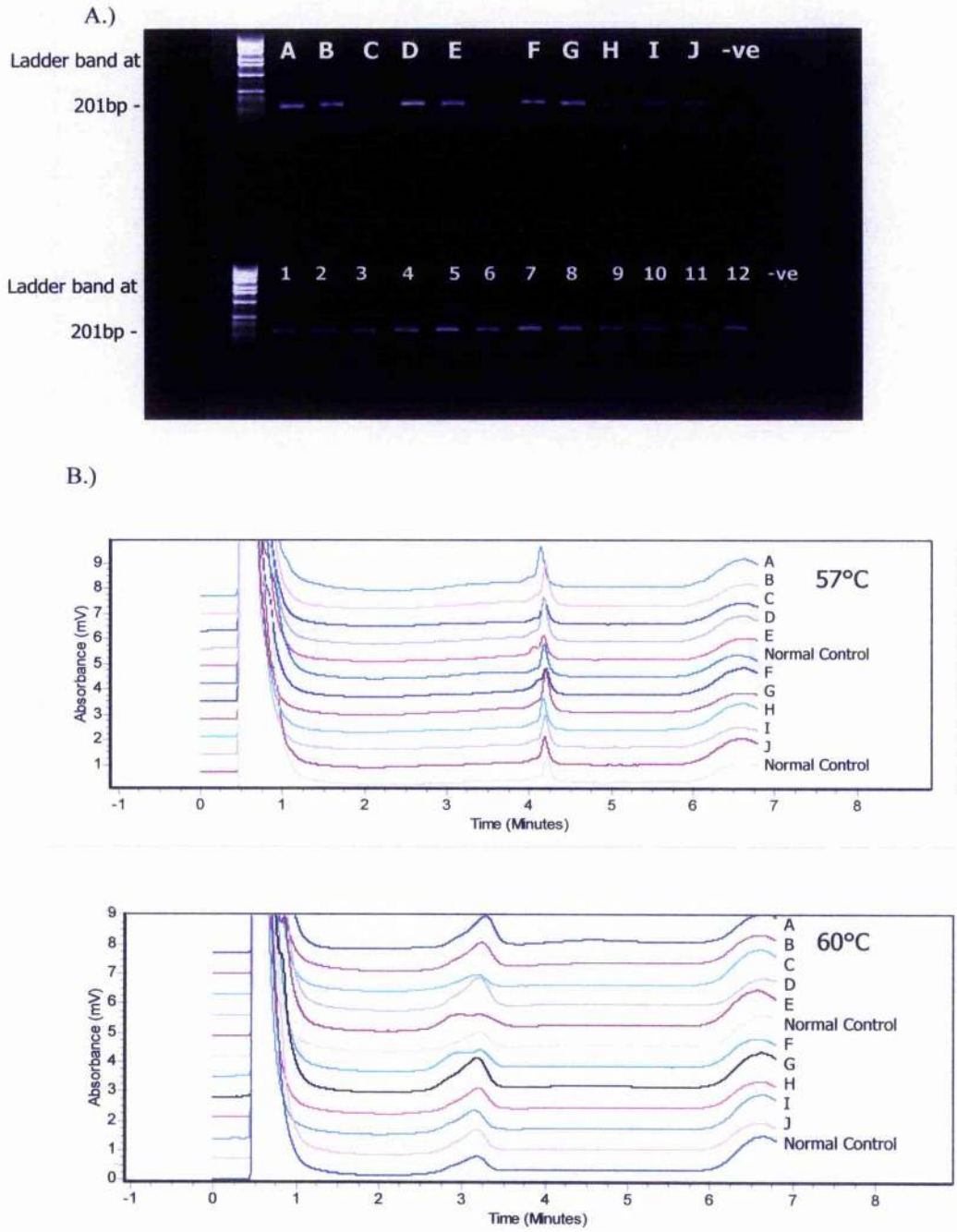


Figure 3.1.2 PCR and dHPLC analysis of exon 70 for patients A-J. A.) A 2% agarose gel photograph showing the 233bp PCR products obtained using primers for exon 70, 3µl of product plus 2µl of sucrose loading mix was run in each lane alongside 4µl of 1kb ladder. In the upper half of the gel are PCR products for patients A-J and in the lower half are PCR products for exon 70 from 12 reactions using the same control DNA Each patient PCR was then combined with an aliquot of control DNA in equal quantities based on the intensity on the agarose gel, this mix was heated and then slowly cooled to allow heteroduplexes to form. 5µl of each PCR mix was then injected into the dHPLC machine. B.) Shown are the dHPLC profiles for each of the patient heteroduplexes (plus two unmixed control homoduplex profiles) under two separate denaturing conditions, 57°C and 60°C. Samples are compared against each other for differences within the group. Shifts can be seen at both temperatures in samples E and F.

3.1.2 Characterising variants

A shift in the profiles of exon 70 for patients E and F was evidence for the presence of heteroduplexes. To further characterise these shifts, bi-directional sequencing of exon 70 was performed. Comparison of patient sequence traces against that from the normal control revealed differences in the amplicon sequence in both patient samples (Figure 3.1.3):

In patient E a single nucleotide deletion within the coding region of exon 70 was observed: 10413delA (Figure 3.1.3). The deletion of a single base causes a shift in the reading frame at codon 3402 (normally coding for a glycine) resulting in a stop codon 10 amino acids downstream predicted to result in a truncated protein of 3412 amino acids in length (G3402fsX3412). The patient was wheelchair bound aged 10 years and had a muscle biopsy with abnormal dystrophin staining typical of DMD. The variant 10413delA has not been previously reported but is predicted to result in a truncated protein, this is consistent with the patient's phenotype and is considered to be the causative mutation in this case. The mutant base change removes the only restriction site for the enzyme *Mnl* I in the exon 70 amplicon (Figure 3.1.4). A study of other individuals in the pedigree show that the mutation is present in the mother's DNA and has not been inherited by the proband's two female siblings, consistent with haplotype data from the region (Figure 3.1.5). As grandparental samples were unavailable the origin of the mutation could not be confirmed.

In patient F sequencing showed the presence of a single base substitution of a C to T within intron 70: 10223+6T>C (Figure 3.1.3). This patient was walking aged 12 years and had a muscle biopsy with abnormal dystrophin staining typical of BMD. 10223+6T>C is a novel base change and its position within the intron makes interpretation difficult. The consensus splice site for the human 5' donor is AG/gtaagt, Shapiro and Senapathy (1987) defined this as 100% using a scoring system based on observations of 1446 5' splice sites in Genbank. The dystrophin exon 70 5' donor differs from the consensus by the substitution of a terminal C in the exon and a G at position 10223+3, underlined in the following sequence: AC/gtgagt, this reduces the Shapiro-Senapathy score to 83.2. Introduction of the variant reduces the Shapiro-Senapathy score to 77.4, the significance of this result is unclear and other splicing factors, such as cryptic splice sites, have not been considered in this calculation. Furthermore, no

mutation within the dystrophin gene has been reported at +6 within the 5' splice site of any exon although similar changes have been shown to be pathogenic in other human disorders (Scholl *et al.*, 1999). This variant remains unique and further investigation, such as mRNA studies, would be necessary to clarify any pathogenic effect.

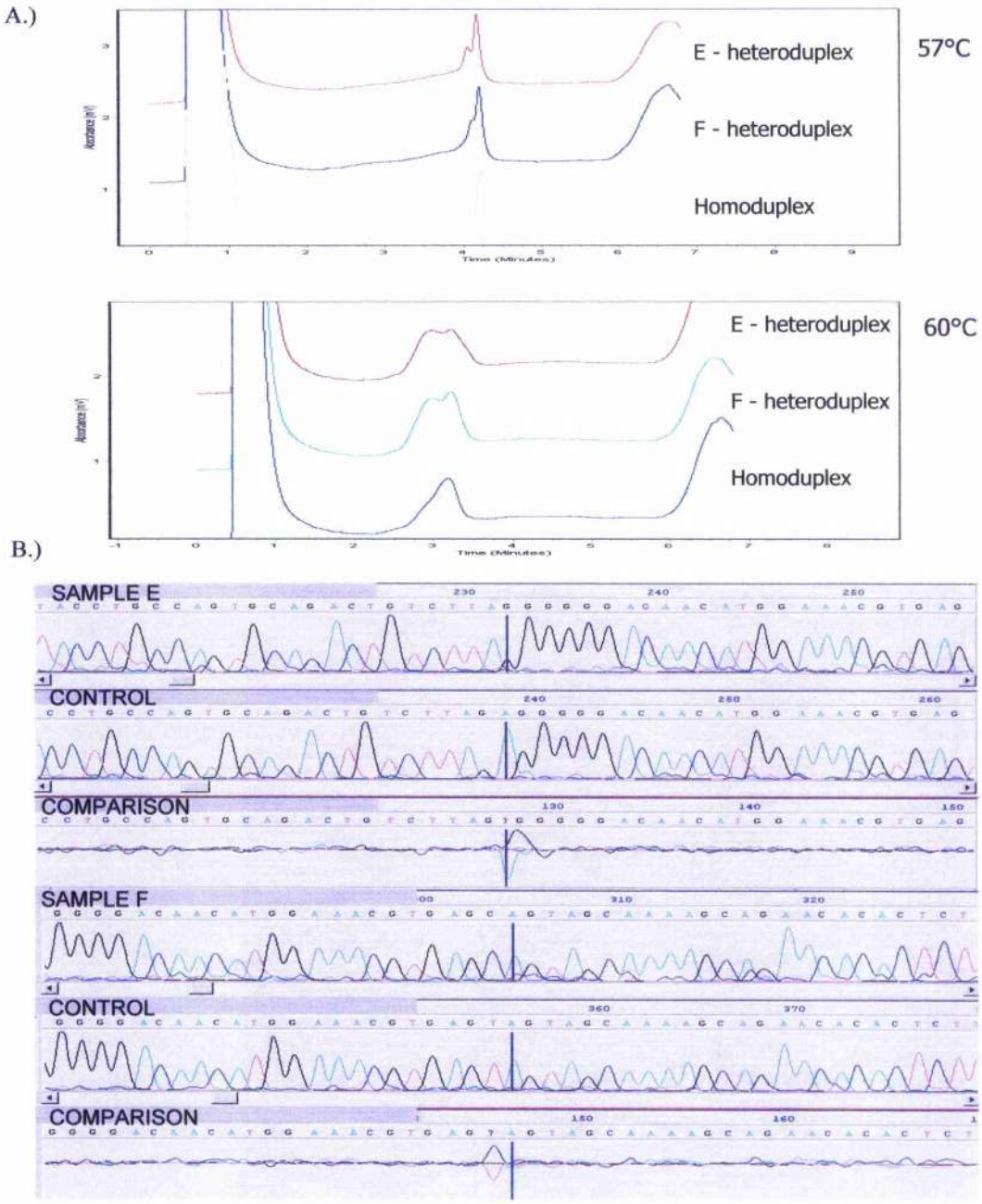
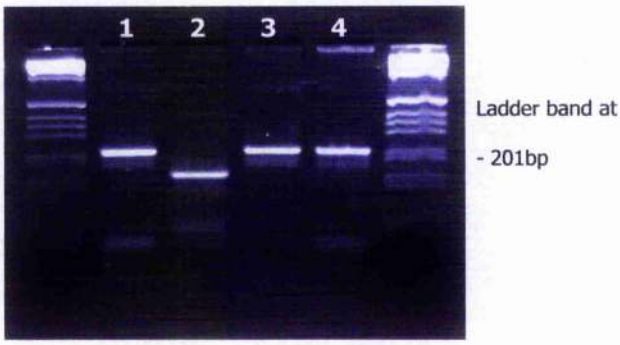


Figure 3.1.3 dHPLC and sequencing of variants in exon 70. A.) dHPLC profiles of the exon 70 heteroduplexes seen in patients E and F alongside the normal homoduplex control. Shown are the profiles at two denaturing conditions (57°C and 60°C) demonstrating that under each condition the different variants had different profiles. B.) Details of sequencing data in Gap4. 2µl of SAP/*Exo* I treated PCR products from each patient were sequenced in a 10µl sequencing reaction using the exon 70 sense primer, products were then precipitated in ethanol:EDTA, centrifuged and the residual nucleotides washed off. Sequencing products were then re-dissolved in formamide and run on an ABI 3100. In each panel the sample sequence is aligned against the normal sequence read, below this is a comparison where the traces are subtracted from one another, any differences are highlighted by peaks above and below the comparison line. The vertical blue line marks the centre of the viewing window in Gap4. Sample E shows a deletion of a single A nucleotide (Dystrophin 10413delA, G3402fsX3412) and sample F shows a T to C base substitution (Dystrophin 10223+6T>C).

A.)



B.)

ttagttttgaaatcatcctgtcctaaatctgatctcaccatgatctccctttagACTACATCA
GGAGAAGATGTTCGAGACTTTGCCAAGGTACTAAAAAACAAATT
TCGAACCAAAAGGTATTTTGCGAAGCATCCCCGAATGGGCTAC
CTGCCAGTGCAGACTGTCTTAGAGGGGGACAACATGGAAACtga
gtagtagcaaaagcagaacacactcttgtttgatgtat

Figure 3.1.4. Restriction digest of exon 70 amplicon using *Mnl*I. A.) A 2% agarose gel showing the results from an *Mnl* I restriction digest reaction of PCR products. 100ng of genomic DNA was amplified with primers for exon 70, 8µl of this reaction was then digested with 20 units of *Mnl* I enzyme in a 20µl reaction volume for 12hrs at 37°C and 15µl run out on a 2% agarose gel alongside a 1kb ladder. The restriction enzyme cuts the normal product at 165bp whilst the mutant product remains uncut at 233bp. Lane 1 is Patient E, lane 2 is a normal control, lane 3 is patient E and lane 4 is an uncut sample.

B.) Below the gel photo is the 233bp sequence of the exon 70 amplicon taken from Genbank accession no: AC079143. Highlighted in red at position 165bp is the base deleted in patient E (10205delA). Exonic sequence is in upper case and intronic sequence is in lower case. Sense and antisense primers are highlighted in grey. The restriction enzyme *Mnl* I recognition sequence (GAGG) is underlined, this is removed in the mutant sequence - the mutant base is highlighted in red. *Mnl* I digest products in the normal sequence, are 165bp and 68bp.

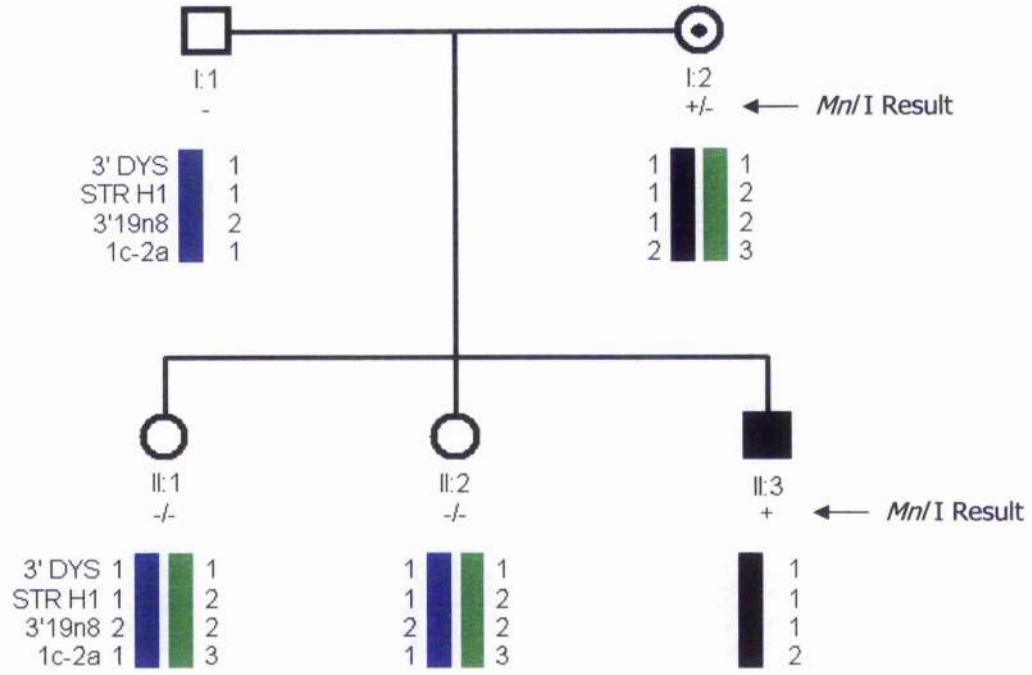
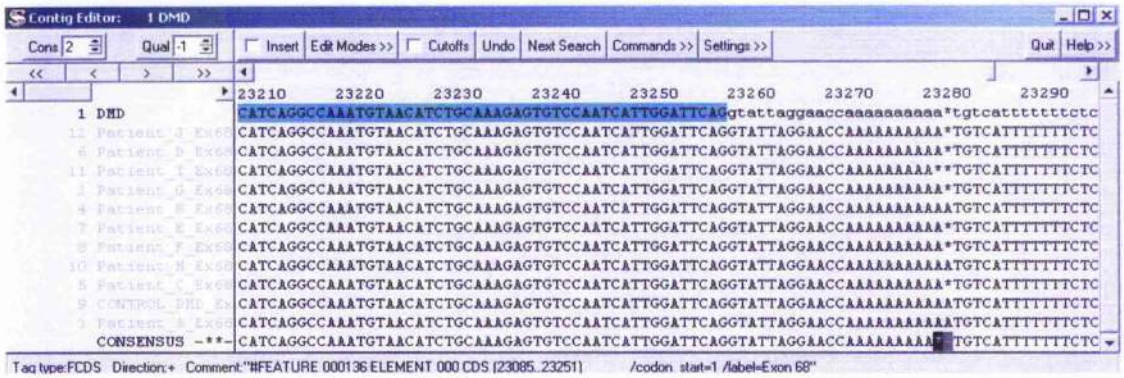


Figure 3.1.5 Pedigree data for family E. Haplotype data from four markers in the 3' region of the dystrophin gene, showing the immediate family of patient E (10205delA). Included are results for patient E (II:3), his two sisters (II:1 and II:2) their father (I:1) and mother (I:2). Each sample was amplified by PCR at 55°C using the published primers, 4µl of this reaction was run on an 8% polyacrylamide gel and detected by silver staining. Alleles are numbered according to their relative size, the largest being 1. Markers STRH1 to 1c-2a are informative for this pedigree. STR H1 is located in intron 64 and 1c-2a is located in intron 59. Also included is the result from the *Mnl* I restriction digest for the familial mutation in exon 70, shown as – for a normal allele and + for the mutated allele. This data shows that neither of the sisters have inherited the mutation, this is corroborated by the marker results showing inheritance of the low risk maternal haplotype for the region.

3.1.3 Sequencing of dystrophin exons 63-79

To determine the sensitivity of dHPLC in this study, patient samples were re-amplified and sequenced in one direction (Method 2.5). Sequence data was then compared against the normal control for variance within the coding region of each amplicon. No further change in the coding sequence of exons 63-79 could be identified in any of the ten patients using this method. The sequence data agrees with the dHPLC data for the regions screened. The two known polymorphisms with the Dp71 region were analysed by sequencing 9974+13_9974+23(A)9-12 showed a frequency of 1/11(9%) for 9A, 6/11(55%) for 10A and 4/11(36%) for 11A (Figure 3.1.6), this was not detected by dHPLC (Figure 3.1.7). The 9649+15C>T was not observed in any of the samples when sequenced (Figure 3.1.8).

A.)



B.)

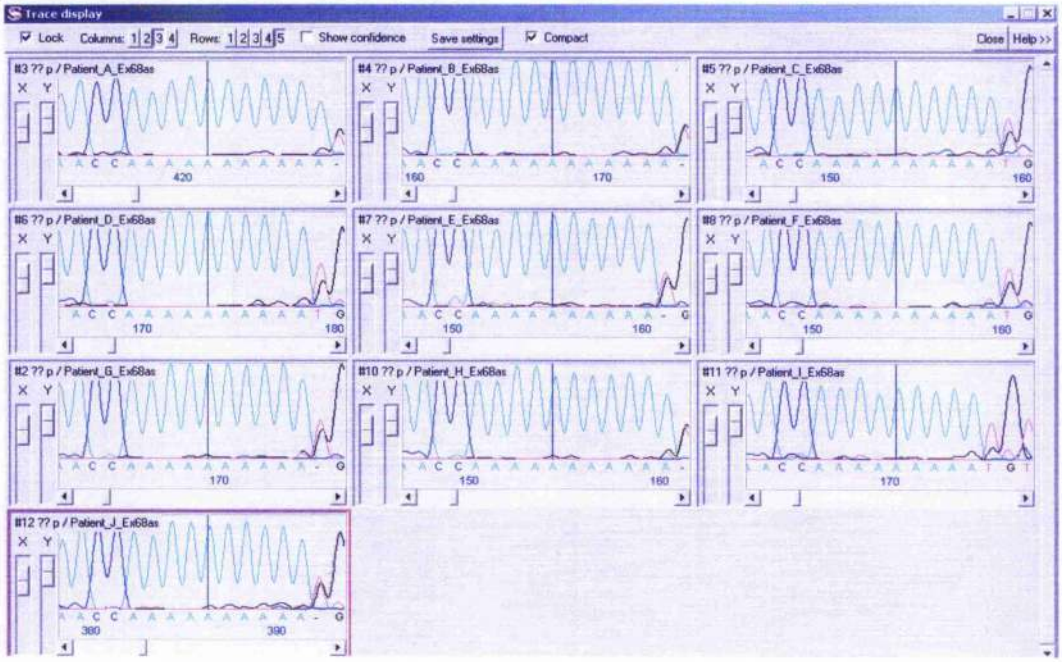


Figure 3.1.6 Sequencing of exon 68 for patients A-J. Details of sequencing data in Gap4. 2µl of SAP/*Exo* I treated PCR products from each patient were sequenced in a 10µl sequencing reaction using the exon 68 antisense primer, products were then precipitated in ethanol:EDTA, centrifuged and the residual nucleotides washed off. Sequencing products were then re-dissolved in formamide and run on an ABI 3100. A.) Sequence alignment as shown by Gap4. Data presented is from patients A-J aligned against each other and a reference sequence (AC078958). The data shown is a section of the exon 68/intron 68 boundary including the 9974+13_9974+23(A)9-12 polymorphism (between bases 23270 and 23280). Missing bases in the alignment are represented by a star (*). B.) Sequence traces from the same database as shown by Gap4. The vertical blue line marks the centre of the viewing window in the software. Data is from each patient showing the polymorphic string of A nucleotides seen at position 9974+13_9974+23(A)9-12.

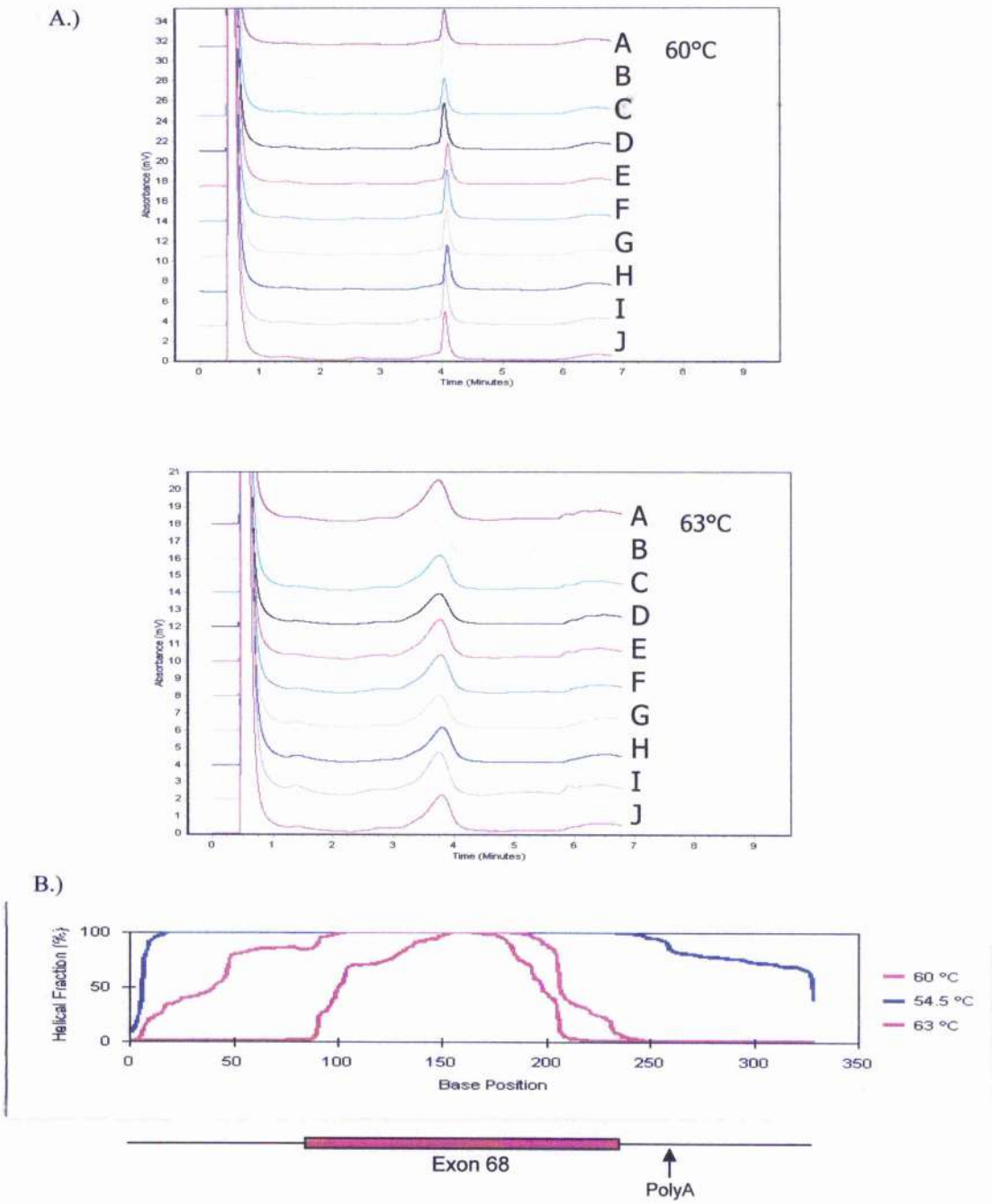


Figure 3.1.7 dHPLC of exon 68 for patients A-J. A.) dHPLC profiles from 5µl injections of patient:control exon 68 amplicon PCR mixes for patients A-J. Shown are profiles under the two denaturing temperatures of 60°C and 63°C. There was no discernable difference between the samples under these conditions although sequencing showed the samples to vary at the poly-A site. B.) An output from the WAVEmaker software, demonstrating the predicted melting properties of the exon 68 amplicon under 3 denaturing conditions (54.5°C, 60°C and 63°C). The graph displays helical fraction (%) against base position in the fragment, below this is a diagram of the gene structure within the fragment highlighting the position of the exon in red and the poly A repeat at position 9974+13_9974+23(A)9-12. The software shows the fragment to be 100% melted at the polymorphic position when at 60°C and 63°C.

A.)

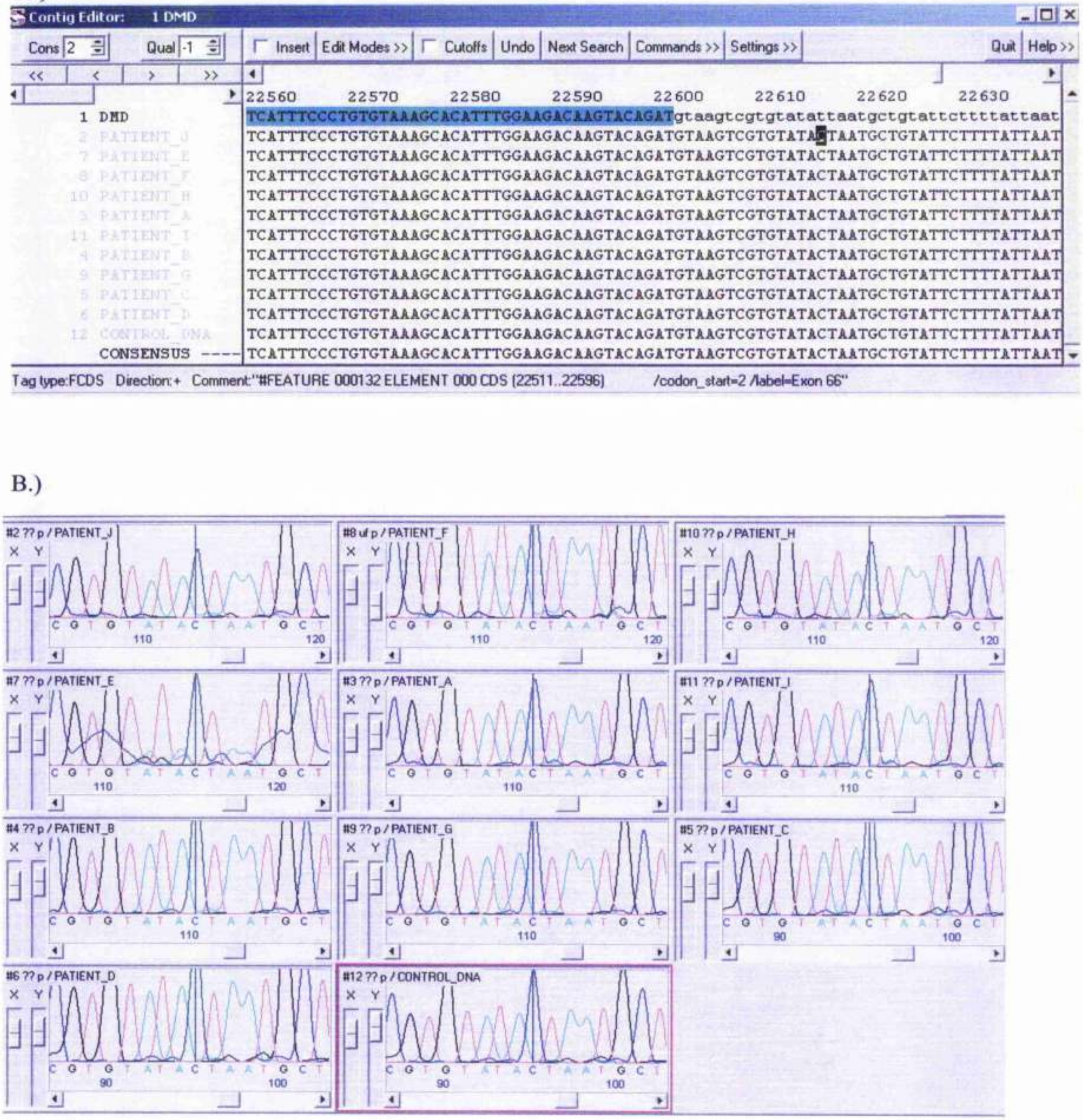


Figure 3.1.8 Sequencing of exon 66 for patients A-J. Details of sequencing data in Gap4. 2µl of SAP/*Exo* I treated PCR products from each patient were sequenced in a 10ul sequencing reaction using the exon 66 antisense primer, products were then precipitated in ethanol:EDTA, centrifuged and the residual nucleotides washed off. Sequencing products were then re-dissolved in formamide and run on an ABI 3100. A.) Sequence alignment as shown by Gap4. Data presented is from patients A-J aligned against each other and a reference sequence (AC078958). The data shown is the exon 66/intron 66 boundary including the 9649+15C>T polymorphism. B.) Sequence traces from the same database as shown by Gap4. Data is from each patient showing the 9649+15C>T polymorphism highlighted by the vertical blue line.
All samples have the C allele at position 9649+15.

3.2 **A Recurrent mutation in exon 70**

3.2.1 Results for robotic PCR and dHPLC

Robotic PCR of exon 70 from 46 patient DNA samples was performed and the resulting products checked on a 96 well agarose gel (Figure 3.2.1). Wild type control DNA was also amplified in 46 wells alongside the patient samples and the resultant PCR products pooled by the robot. Amplification efficiency varied within the group of samples, this is most likely due to the presence of contaminants carried over from sample extraction. As the protocol involved automated mixing of PCR products from patient and control samples, the sensitivity of the dHPLC machine to a range of sample:control PCR concentrations was tested using a known mutation. The results show that detection of a heteroduplex is possible even when the ratio of alleles is 5:1 (Figure 3.2.2). It was decided that a minimum ratio of 2:1 was to be used routinely. Aliquots of pooled control DNA were diluted 1:1 with distilled water and run alongside the sample PCRs on the check gel. All 46 patient PCRs gave bands of at least the same intensity as the diluted control DNA (i.e. a ratio above 2:1). Subsequently the PCR samples were mixed, heated and cooled slowly to allow heteroduplexes to form. Samples were then aliquoted onto a plate for dHPLC by the robot (Method 2.7.2). The robot, once programmed took eight hours to prepare the samples. dHPLC analysis under two denaturing conditions identified 4 samples with a shift from the normal trace.

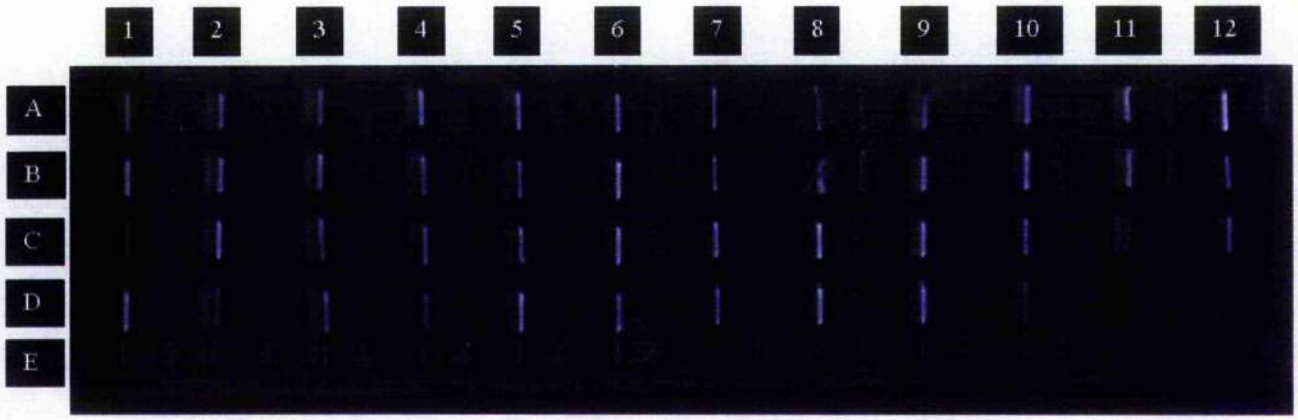


Figure 3.2.1 Robotic PCR of exon 70. Robotic PCR of exon 70 amplicon from 46 patient DNA samples (lanes A1-D10). 3µl of each PCR product was aliquoted onto a separate well on a 96 well plate and mixed with 3µl loading buffer. These aliquots were then loaded onto a 1.5% 96 well agarose gel and run at 80V limiting for 20 minutes. Lane D12 contains the No DNA control PCR and lane D11 is a blank well. A further 46 reactions for exon 70 using control DNA was performed alongside and then pooled by the robot. 3µl aliquots of the pooled control DNA PCR (diluted 1:1 with water) are loaded into lanes E1-12, this allows for a check of PCR efficiency and size of product. Bands can be seen for all patient PCR reactions, the bands are the same size as the control PCR band and are of equal or greater intensity.

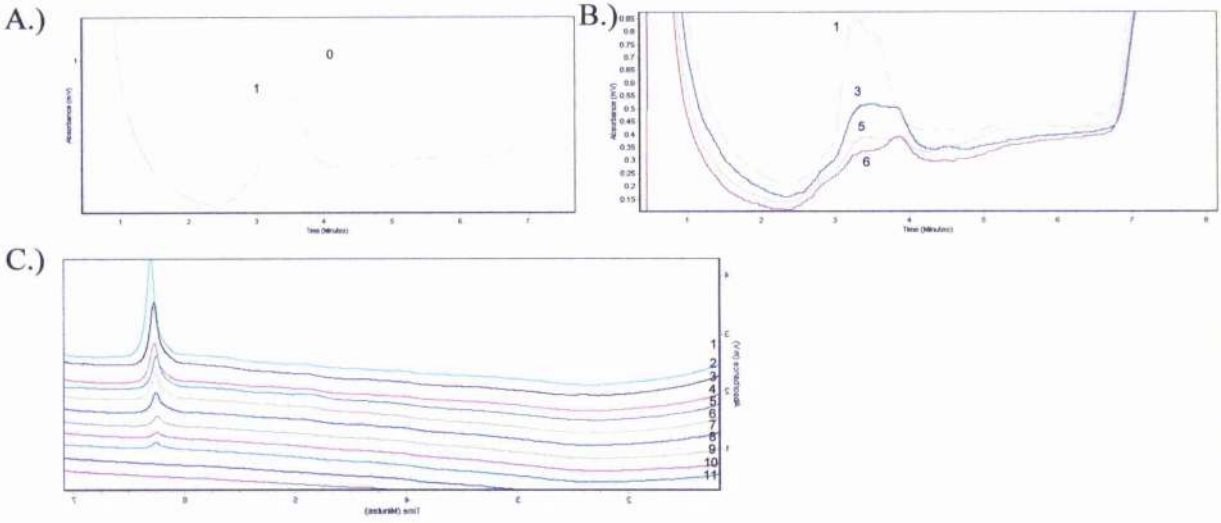


Figure 3.2.2 Sensitivity of dHPLC to heteroduplex concentration. Results from experiments designed to test the sensitivity of the dHPLC machine. The amplicon for Exon 70 was amplified from a sample with a known mutation (10205delA, G3402fsX3412) and a control sample with no variant. To study the effects of mixing heteroduplexes at varying concentrations (mimicking a reduced efficiency of the patient PCR) the samples were mixed according to the table below.

Lane	Normal Control PCR	Mutant PCR	dH_2O
0	10	0	0
1	5	5	0
2	5	4	1
3	5	3	2
4	5	2	3
5	5	1	4
6	5	0	5
7	4	0	6
8	3	0	7
9	2	0	8
10	1	0	9
11	0	0	10

The mixes were heated then slowly cooled and 5µl of each mix was injected onto the dHPLC column under denaturing conditions (58°C) and non-denaturing conditions (50°C). A.) Shows lanes 1 and 0 under denaturing conditions, demonstrating that a 1:1 heteroduplex mix reduces the intensity of signal slightly. B.) Lanes 1, 3, 5 and 6, (ratios of 1:1, 5:3, 5:1 and complete homozygote at 50% concentration) under denaturing conditions. These results show that at a ration of 5:1 plus dilution of signal (mimicking poor PCR) a shift in dHPLC trace from normal is observed. However the signal at this level is weak, therefore a minimum ratio of 2:1 was chosen as a suitable threshold for the control:patient PCR concentrations. C.) Profiles of lanes 1-11 under non-denaturing conditions showing the effect of mixing lower concentrations of PCR products on signal intensity.

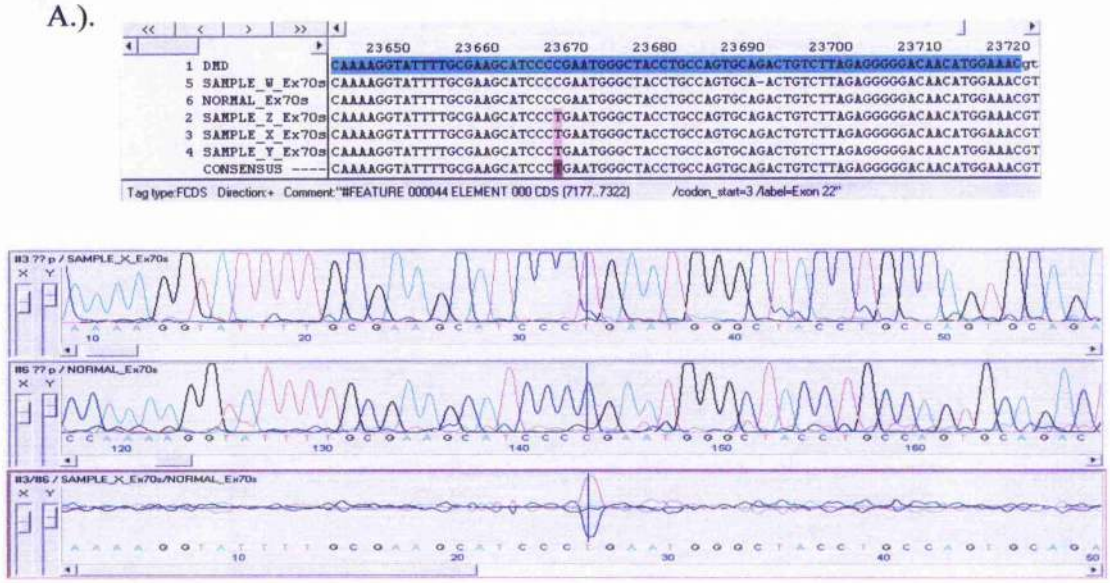
3.2.2 Sequence analysis of exon 70 in samples with abnormal dHPLC traces

PCR and sequencing of dystrophin exon 70 from the DNA samples with abnormal dHPLC profiles showed the presence of a variant within each of the samples (Figure 3.2.3).

ID	Mutation	Reported previously	CK / age	Age of diagnosis	Residence	Familial cases
W	10099_10101 delGAA	Yes	8440/ 7yr	7 yrs	N. Scot.	N
X	10171C>T	Yes	9500/6yr	6yrs	E. Scot.	Y
Y	10171C>T	Yes	11,420/4yr	4yrs	SW.Scot	Y
Z	10171C>T	Yes	9890/ 6yr	6yrs	NW. Scot.	N

The variant 10099_10101delGAA; E3367del has been reported before as present in a case of intermediate muscular dystrophy. An in-frame deletion of an amino acid is not usually associated with the Duchenne phenotype but this variant lies within a highly conserved region involved in protein binding in which missense changes have been associated with a severe phenotype (Figure 3.2.4). The mutation 10171C>T is a nonsense mutation changing an arginine to a stop codon (R3391X) and has been reported several times before (www.dmd.nl).

A.).



B.)

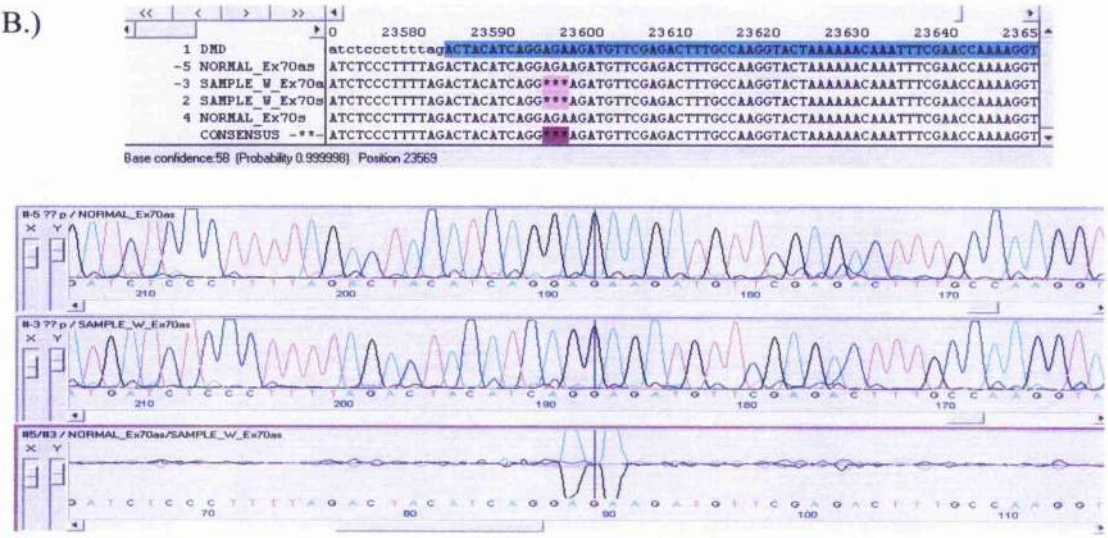


Figure 3.2.3 Sequencing of exon 70 for patients W-Z. Sequence output from the Staden package showing data from the 4 dHPLC variants and a abnormal control aligned against each other and a reference sequence (AC078958). A.) The data shown in the upper panel is a sequence alignment of sequence reads from patients W, X, Y, Z and a normal control showing sequence obtained using the Exon 70 sense primer. This section covers the exon 70/intron 70 boundary (exon highlighted blue in the reference sequence), the variant 10171C>T (R3391X) in samples X,Y and Z is highlighted in orange. Below the sequence alignment is a comparison where the traces are subtracted from one another, any differences are highlighted by peaks above and below the comparison line, this data shows the C to T base substitution in sample X. B.) Sense and antisense sequences from the normal control and sample W. Shown in this section is the exon 70/intron 69 boundary (exon highlighted blue in the reference sequence), the variant 10099_10101del Glu3367del is highlighted in orange. The trace subtraction shows that several bases are different between the samples.

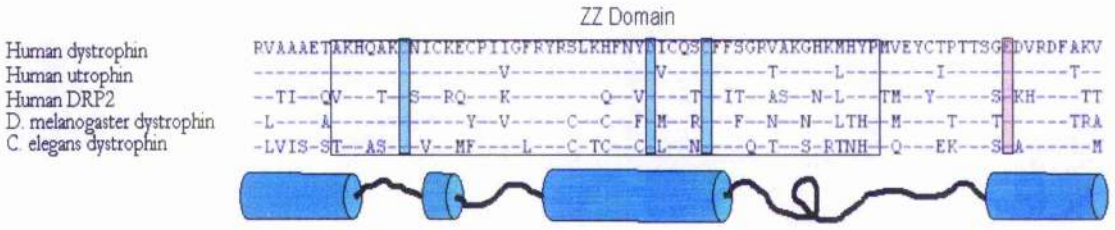


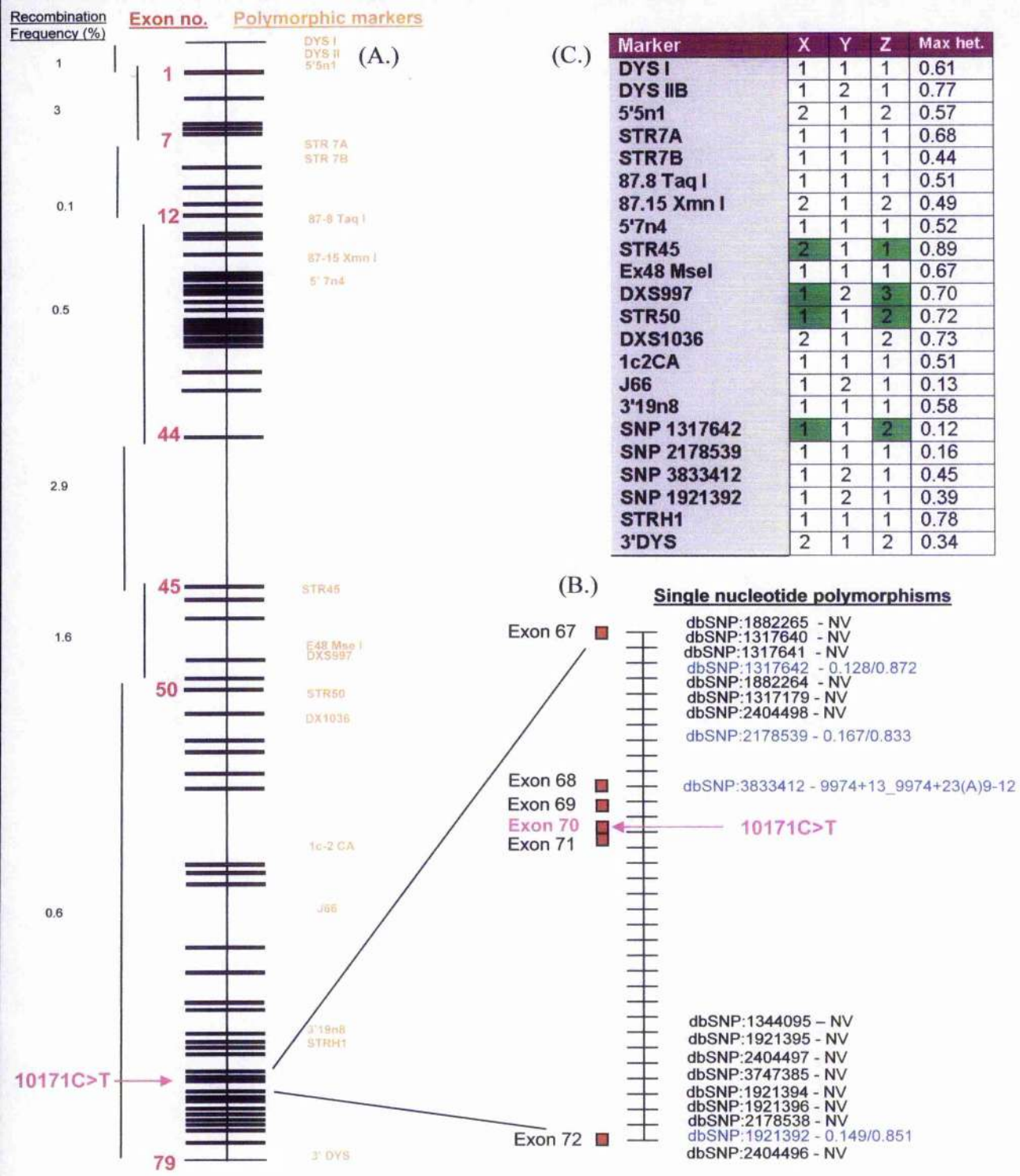
Figure 3.2.4 Sequence alignment and protein structure of E3367del. An image based on data presented in Becker *et al.* (2003). The figure represents an alignment of amino acids 3300-3375 from human dystrophin and the related sequences in utrophin and DRP2 from human and dystrophin from *Drosophila melanogaster* and *Caenorhabditis elegans*. Differences to the dystrophin amino acid sequence in the other proteins are shown. The E3367del variant is highlighted in red. Highlighted in green are reported missense mutations, each affects a conserved amino acid believe to have structural significance in the region: C3313F (Flanigan *et al.*, 2003), D3335H (Goldberg *et al.*, 1998) and C3340Y (Lenk *et al.*, 1996). Below the sequence comparison is a representation of the protein secondary structure showing predicted helix structures (green cylinders). It is predicted that E3367del removes a conserved amino acid at the amino-terminal of an α-helix suggesting an important role for this uncharacterised domain in dystrophin function.

3.2.3 Haplotype analysis of a common mutation

The identification of the same mutation (10171C>T R3391X) in three individuals raised the possibility that the mutation in these individuals has a common origin. This prompted pedigree analysis in each of the three cases to identify any shared haplotype between the individuals as described in Method 2.15.

Haplotype analysis for 18 polymorphic markers across the dystrophin gene showed that two of the samples (X and Z) shared a haplotype at the 3' and 5' regions of the gene but did not share a haplotype between markers STR50 and STR45. This result could be due to recent recombination within the DMD gene. To help resolve whether the mutated alleles shared a common ancestor a database search for single nucleotide polymorphisms (SNPs) within the dystrophin exon 70 region revealed 18 SNPs located in a 25kb region flanking exon 70. Four of the SNPs had been previously verified in dbSNP (www.ncbi.nlm.nih.gov) therefore these four SNPs were sequenced in the samples from the three patients with the same mutation. Three of the SNPs were heterozygous and added new haplotype information to the microsatellite data. Summarising the exon 70 region SNP data: two patients (X and Z) had similar haplotypes that differed by only one SNP which was within the region of the mutation. The third patient (Y) had a haplotype that differed from the other haplotypes at two SNPs (Figure 3.2.5).

Combining the data generated by the SNP and the polymorphic markers it appears that the three patients have different intragenic haplotypes but with two of the patients sharing 'blocks' of alleles in certain regions of the gene. This result could be due to recent intragenetic recombination within a common ancestral gene. To investigate this mitochondrial haplotyping was carried out.



(C.)

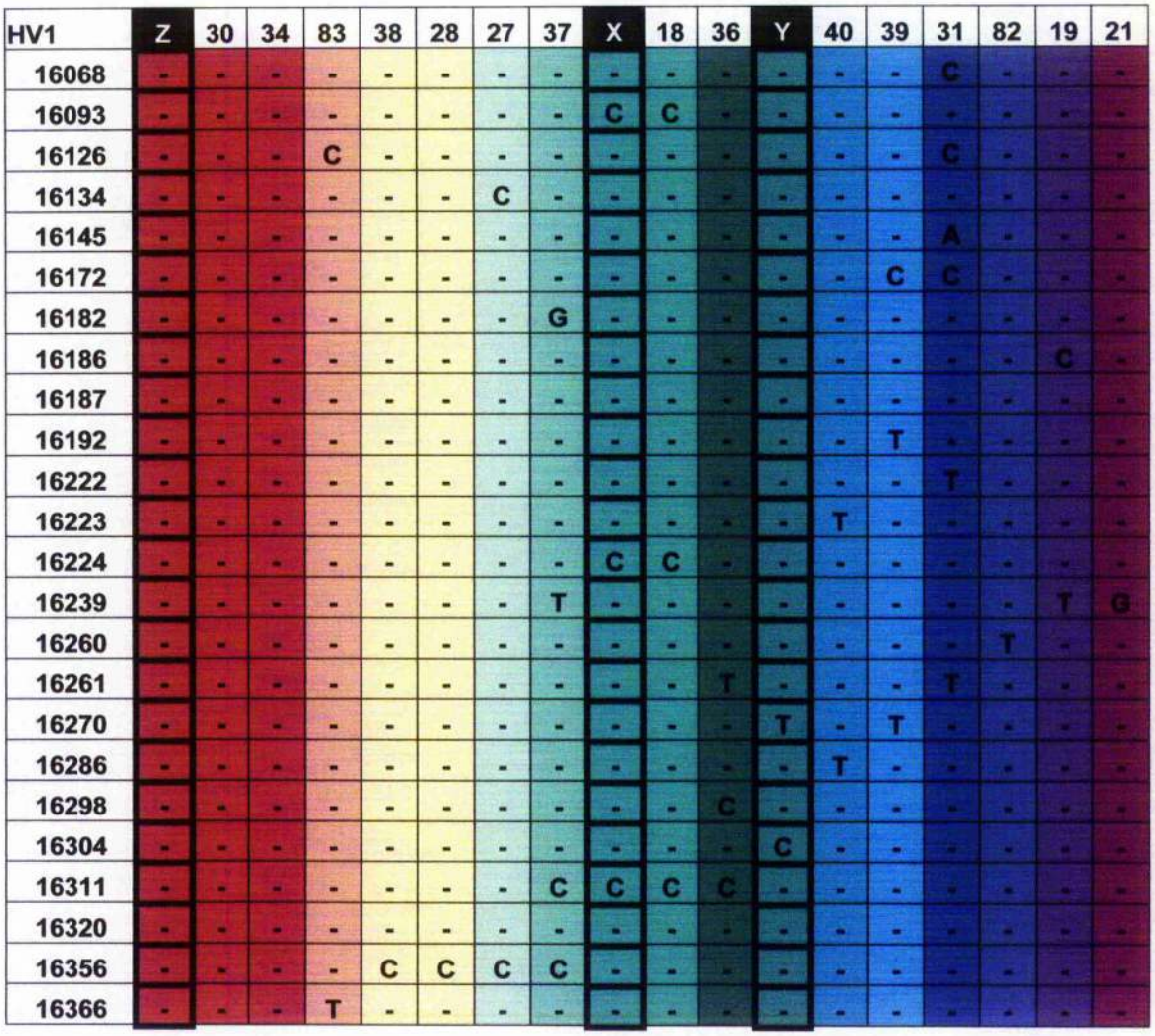
Marker	X	Y	Z	Max het.
DYS I	1	1	1	0.61
DYS IIB	1	2	1	0.77
5'5n1	2	1	2	0.57
STR7A	1	1	1	0.68
STR7B	1	1	1	0.44
87.8 Taq I	1	1	1	0.51
87.15 Xmn I	2	1	2	0.49
5'7n4	1	1	1	0.52
STR45	2	1	1	0.89
Ex48 MseI	1	1	1	0.67
DXS997	1	2	3	0.70
STR50	1	1	2	0.72
DXS1036	2	1	2	0.73
1c2CA	1	1	1	0.51
J66	1	2	1	0.13
3'19n8	1	1	1	0.58
SNP 1317642	1	1	2	0.12
SNP 2178539	1	1	1	0.16
SNP 3833412	1	2	1	0.45
SNP 1921392	1	2	1	0.39
STRH1	1	1	1	0.78
3'DYS	2	1	2	0.34

Figure 3.2.5 Microsatellite and SNP haplotyping (A.) Depiction of the dystrophin gene (2.4Mb) showing recombination frequency between labelled exons. In orange are the polymorphic markers, positioned in their relative location in the gene. The position of mutation 10171 C>T is highlighted in pink. (B.) Detail of the region between exons 68 and 72 showing the SNPs identified in genebank. SNPs are identified by their Genbank number followed by the registered heterozygosity. NV = not validated in the database. SNPs used have been highlighted in blue. (C.) Haplotypes for patients X,Y and Z for all the markers studied. Alleles are labelled by relative size, 1 being largest. Markers where patients X and Z differ are highlighted in green.

3.2.4 Mitochondrial hypervariable region 1 (HV1) haplotyping.

Mitochondrial hypervariable regions have been shown to be capable of establishing a common genetic origin via the maternal (therefore X-linked) lineage (Richards and Macaulay, 2001). The hypervariable region 1 (HV1) of mitochondrial DNA is a commonly used region to establish mitochondrial ancestry. This region was therefore analysed in the three patients and also in fifteen control samples (method 2.11).

The results for all samples analysed revealed 24 base changes which allowed the identification of 14 haplotypes within the HV1 region (Figure 3.2.6). The patients who carried the identical DMD mutation each had different haplotypes. Two of the patient haplotypes were also found in the control samples. The consensus mitochondrial sequence was the most common haplotype in the study group, being present in 1 in 6 of the samples tested. Two other haplotypes (16356 T>C and 16093T>C:16224 T>C: 16311 T>C) were seen at a frequency of 1 in 9. This frequency is similar to that found in the Welsh population where HV1 haplotypes have been shown to be relatively homogeneous (Richards *et al.*, 1996). The HV1 data shown here suggests a similarly low level of mitochondrial diversity exists in the control sample population.



HV1	Z	30	34	83	38	28	27	37	X	18	36	Y	40	39	31	82	19	21
16068	-	-	-	-	-	-	-	-	-	-	-	-	-	-	C	-	-	-
16093	-	-	-	-	-	-	-	-	C	C	-	-	-	-	-	-	-	-
16126	-	-	-	C	-	-	-	-	-	-	-	-	-	-	C	-	-	-
16134	-	-	-	-	-	-	C	-	-	-	-	-	-	-	-	-	-	-
16145	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-
16172	-	-	-	-	-	-	-	-	-	-	-	-	-	C	C	-	-	-
16182	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	-
16186	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	C	-
16187	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
16192	-	-	-	-	-	-	-	-	-	-	-	-	-	T	-	-	-	-
16222	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	-	-	-
16223	-	-	-	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-
16224	-	-	-	-	-	-	-	-	C	C	-	-	-	-	-	-	-	-
16239	-	-	-	-	-	-	-	T	-	-	-	-	-	-	-	-	T	G
16260	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	-	-
16261	-	-	-	-	-	-	-	-	-	-	T	-	-	-	T	-	-	-
16270	-	-	-	-	-	-	-	-	-	-	-	T	-	T	-	-	-	-
16286	-	-	-	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-
16298	-	-	-	-	-	-	-	-	-	-	C	-	-	-	-	-	-	-
16304	-	-	-	-	-	-	-	-	-	-	-	C	-	-	-	-	-	-
16311	-	-	-	-	-	-	-	C	C	C	C	-	-	-	-	-	-	-
16320	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
16356	-	-	-	-	C	C	C	C	-	-	-	-	-	-	-	-	-	-
16366	-	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Figure 3.2.6 Mitochondrial HV1 haplotyping. Sequencing of the mitochondrial HV1 region (base pairs 16009 to 16390) in 15 control samples and patients X,Y and Z. Only bases where a change was observed from the cambridge reference sequence (http://www.mitomap.org/mitomap/mitoseq.html) have been listed. Base numbers are listed in the column on the far left and sample ID numbers are in the first row. Individual results are in columns and have been aligned in order of first base change from the consensus sequence in the direction 16390 to 16009. Columns have been shaded according to the haplotype generated. Individuals with the same haplotype have the same shading. Patient haplotypes are emphasised in bold. Samples Z, 30 and 34 share the same consensus haplotype. Samples 38 and 28 have the same haplotype (16356 T>C) and samples X and 18 have the same haplotype (16093T>C, 16224 T>C, 16311 T>C), all other individuals have a unique HV1 haplotype in this sample set.

The conclusion from the haplotype analysis is that the mutation 10171C>T has independently arisen on at least three different occasions. The three families therefore do not represent a founder effect for this mutation.

To study the origins of the mutation in each of the pedigrees, samples from family members were analysed for the presence of the mutation.

3.2.5 Mutation analysis in family members.

To test for the 10171C>T mutation in other family members a PCR-based analysis of the mutation was developed. A primer was designed that incorporated a *Dde* I restriction site in the presence of the mutation. The primer has a C to A mismatch at the 3rd 3' base and incorporates a G/C base in the final PCR product (Figure 3.2.7). Samples from family members were collected and a study into the inheritance of the mutation in each pedigree was performed. The results show that in the pedigree of individual Z the mutation is present in the patient's mother but not in his maternal grandmother (Figure 3.2.8). In this family the mutation appears to be a new mutation either in the germline of the grandparents or an early embryonic mutation in the mother. From the haplotype data: Individual Z shared several alleles with individual X (Figure 3.2.5). In the pedigree of individual Y the mutation was shown to have originated in a generation previous to that for which DNA samples were available. In this large pedigree several males were affected with one manifesting carrier diagnosed aged 5 years (Figure 3.2.9). Samples for pedigree X were limited to the immediate family, showing that the mutation was present in the affected patient's mother (Figure 3.2.10) and no samples were available for pedigree analysis of patient W.

The sporadic nature of the mutation is shown by the results from the family study of patient Z. Patient Z had a similar dystrophin haplotype to patient X (Figure 3.2.5). The other pedigree results did not provide evidence as to the origin of the mutation in these families.

A.)

Wild type sequence: CATCCCCGAATG
Mutant sequence: CATCCCTGAATG

Base pairs 10165 to 10176
Accession no. NM_004006.1

Antisense primer introduces a mismatch at base no.10174, highlighted in green:

Polymerase extension
Primer sequence: ← CTCACCCGATGGACGGTCACGTCTGACA
Genomic sequence: CATCCCNGAATGGGCTACCTGCCAGTGCAGACTGT

Amplicon sequence with incorporated *Dde* I restriction site in the mutant allele:

Wild type sequence: CATCCCCGAGTG
Mutant sequence: CATCCCTGAGTG *Dde* I restriction site C'TNAG in red

New amplicon size: 168bp
Digest products: wild type allele: 168bp mutant allele: 140bp/28bp

B.)

Figure 3.2.7 Generation of a *Dde* I restriction site and digest results. A.) The mutation 10171C>T R3391X does not alter a natural restriction site so a primer was designed to incorporate a *Dde* I restriction site in the presence of the mutant base. A new antisense primer of 28 oligonucleotides in length introduces a mismatched base at base no.10174 (Accession no. NM_004006.1) used with the original sense primer it produces an amplicon of 168bp in length. *Dde* I will excise the terminal 28 base pairs in the mutant amplicon. Digestion products were resolved on a 3% agarose gel. B.) An agarose gel showing the results from a *Dde* I restriction digest reaction of PCR products of the amplicon described above. 100ng of genomic DNA was amplified using the mismatched antisense primer and the exon 70 dHPLC sense primer, 8µl of this reaction was then digested with 20 units of *Dde* I enzyme in a 20µl reaction volume for 12hrs at 37°C. 15µl of this reaction was run out on a 3% agarose gel alongside a 1kb ladder. Lane 1: 10171T>C affected patient, Lane 2. 10171T>C obligate carrier, Lane 3: 10171T>C affected patient, Lane 4: Normal control, Lane 5: No DNA control. Lane 6: 1 kb ladder.

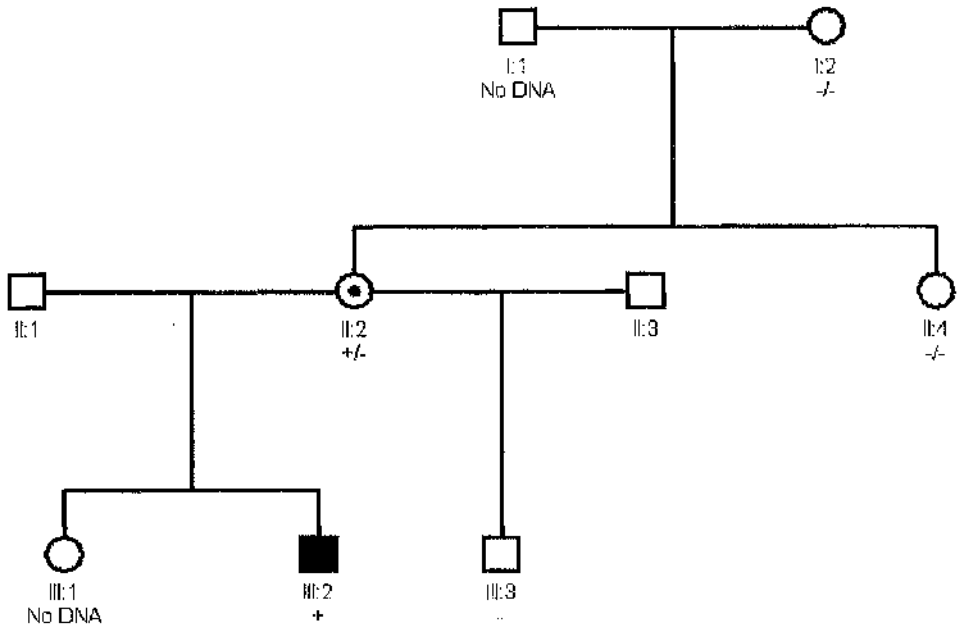
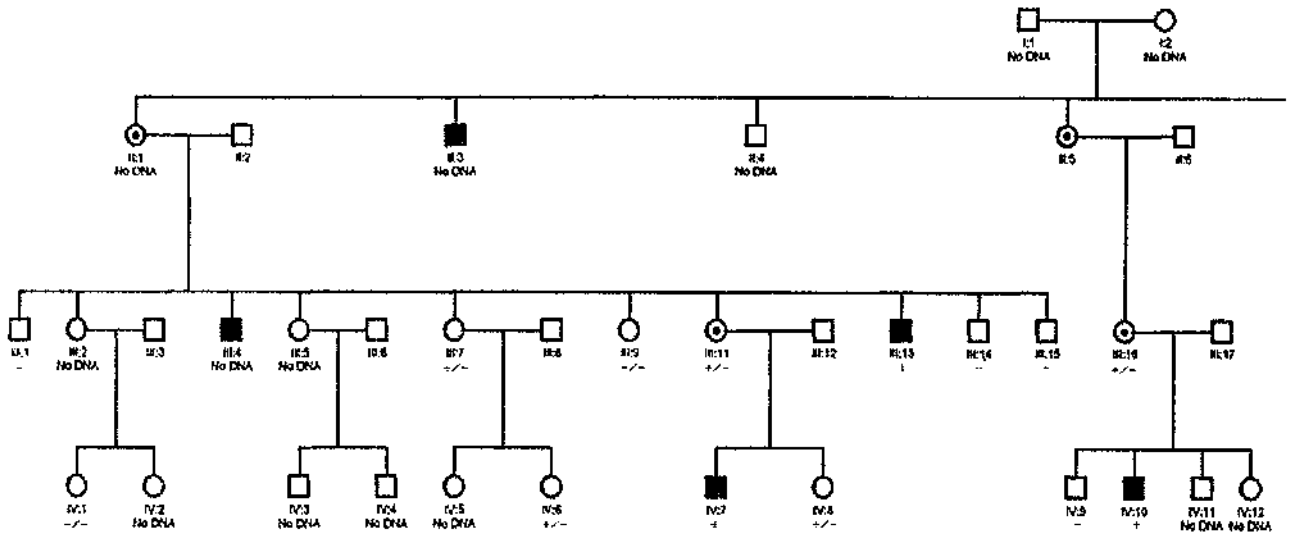


Figure 3.2.8 Pedigree data for family Z. Pedigree for patient Z (III:2) affected by DMD mutation 10171C>T. Availability of DNA samples for all individuals related to the proband's mother is indicated. The result from the PCR followed by *Dde* I restriction digest is shown by a + for the presence of the mutant allele and a – for the presence of a normal allele. When a DNA sample was unavailable the individual is marked as: No DNA. The mutation is seen in the proband's mother (II:2) but not in his half brother (III:3). The proband's maternal aunt (II:4) does not carry the mutation nor does his maternal grandmother (I:2). This data shows that the mutation originates from the grandparental germline - haplotype studies were not conducted.


I:1
No DNA
I:2
No DNA
II:1
No DNA
II:2
II:3
No DNA
II:4
No DNA
II:5
II:6
II:7
+/−
II:8
II:9
−/−
II:11
+/−
II:12
II:13
+
II:14
−
II:15
−
II:16
+/−
II:17
IV:1
−/−
IV:2
No DNA
IV:3
No DNA
IV:4
No DNA
IV:5
No DNA
IV:6
+/−
IV:7
+
IV:8
+/−
IV:9
−
IV:10
+
IV:11
No DNA
IV:12
No DNA

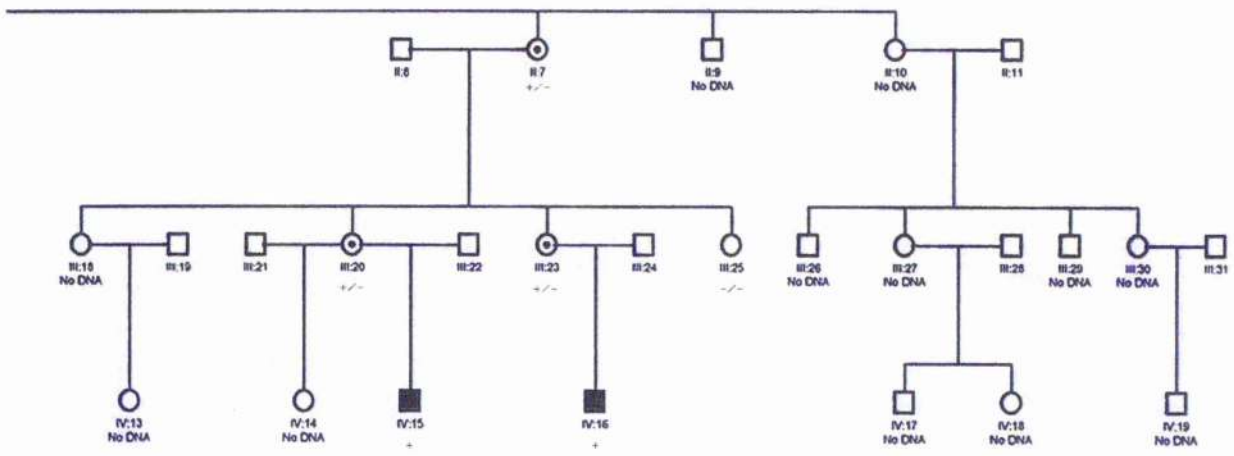


Figure 3.2.9 Pedigree data for family Y. Pedigree for patient Y (IV:16) affected by DMD mutation 10171C>T. Availability of DNA samples for all individuals related to the obligate carrier mother indicated. The result from the PCR followed by *Dde* I restriction digest is indicated by a + for the presence of the mutant allele and a – for the presence of a normal allele. When a DNA sample was unavailable the individual is marked as: No DNA. There are seven affected males in this pedigree identifying several obligate carriers. This data shows that the mutation can be traced back to the great-grandparents of patient Y, its origin cannot be determined. IV:8 is a manifesting carrier with symptoms of muscle myopathy.

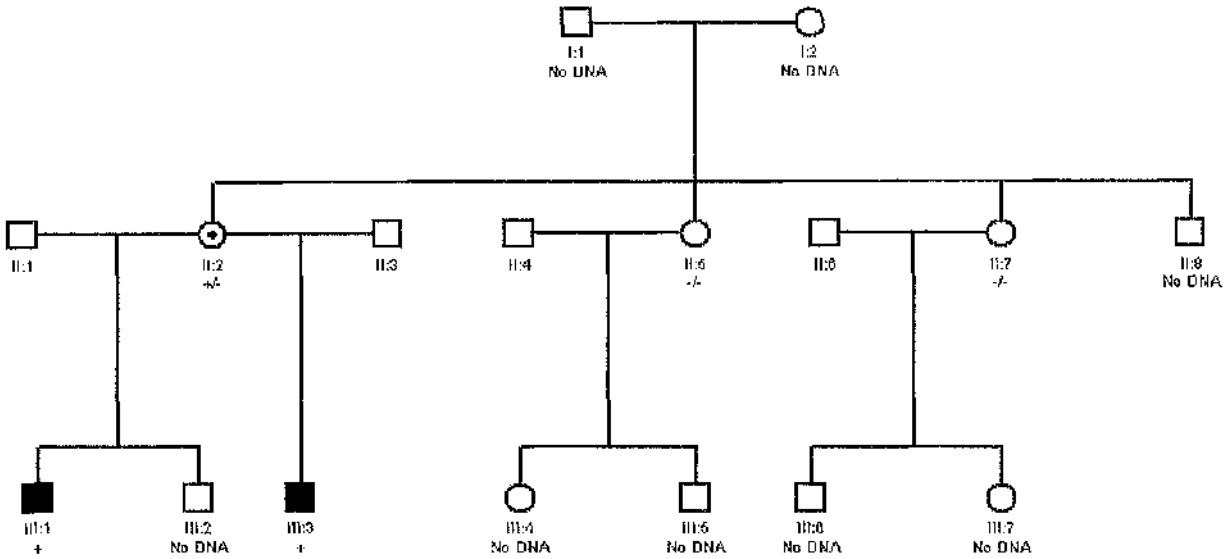


Figure 3.2.10 Pedigree data for family X. Pedigree for patient X (III:1) affected by DMD mutation 10171C>T. Availability of DNA samples for all individuals related to the proband's mother indicated. The result from the PCR followed by *Dde* I restriction digest is indicated by a + for the presence of the mutant allele and a – for the presence of a normal allele. When a DNA sample was unavailable the individual is marked as: No DNA. The mutation is seen in the proband's half brother (III:3) and mother (II:2). Neither of the two maternal aunts carried the mutation. No DNA sample was available from the maternal grandparents. Haplotype studies were not conducted.

3.3 dHPLC analysis of dystrophin exons 1-79

3.3.1 Assay design

3.3.1.1 Primer selection

Results from the primers designed to screen dystrophin exons 63-79 by dHPLC has already been described in Results 3.1. Primers for each of the remaining dystrophin exons 1-62 were subjected to PCR at two conditions (Methods 2.4.2) with the most efficient of the two reactions chosen as the final programme. Where a reaction failed new primer sets were optimised to produce a single (exon containing) amplicon at one of the two PCR programmes.

3.3.1.2 Robotic PCR and dHPLC

Robotic programming allowed amplification of each amplicon from a single DNA aliquot within a single programme (Figure 3.3.1).

Robotic manipulation was then used to mix PCR products (exons 1-79) from the patient and control DNA samples (Figure 3.3.2) which were subsequently heated and cooled to allow heteroduplexes to form. The resultant mixes were then automatically injected onto a dHPLC column under conditions specific to mutation detection within the exon of each amplicon (Methods 2.8.2).

As a preliminary test to confirm that all the exons had been correctly amplified using the robotic protocol: automated unidirectional sequencing of all PCR amplicons from the control DNA sample was performed. Subsequent construction of a Gap4 database of these sequencing traces in the Staden package showed that each exon had been amplified correctly (Figure 3.3.3).

A.)

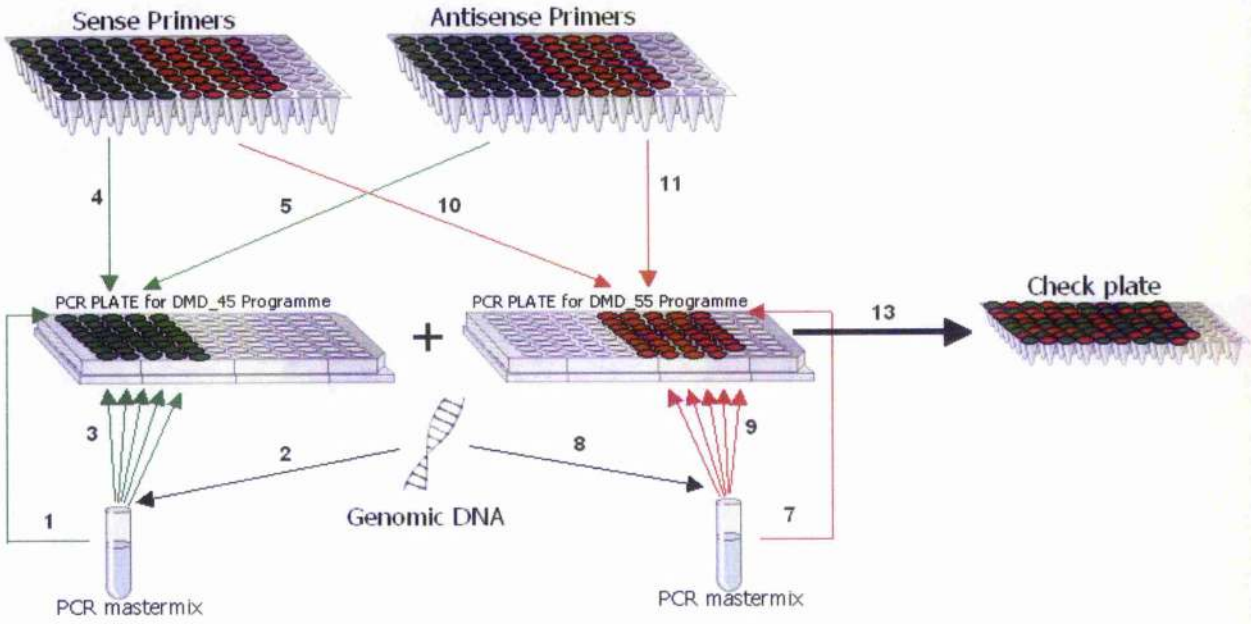
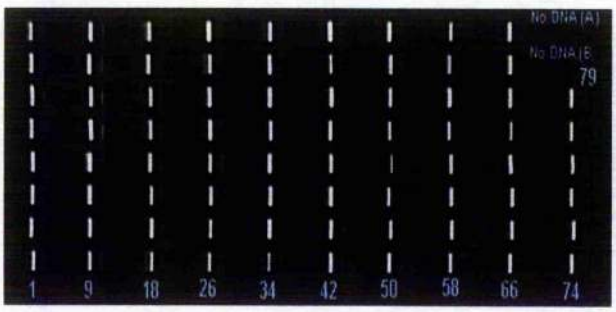
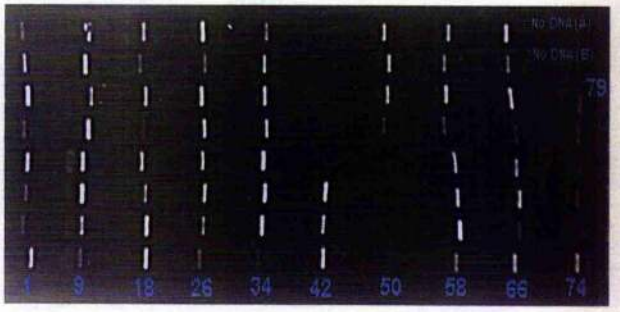


Figure 3.3.1 Robotic PCR of dystrophin exons 1-79. A schematic representation of the robotic steps involved in automated PCR of the dystrophin gene. Liquid transfers are represented by numbers, in the order they are performed on the robot. 1) A 25µl aliquot of (negative control) mastermix is transferred to the DMD_45 PCR plate. 2) 500ng of sample DNA is mixed into the mastermix. 3) 23µl aliquots distributed into the PCR plate. 4) 3µl of sense primers (amplicons optimised to DMD_45) are added. 5) 3µl of antisense primers (amplicons optimised to DMD_45) are added. 6) plate is cycled through DMD_45 programme. 7) A 25µl aliquot of (negative control) mastermix is transferred to the DMD_55 PCR plate. 8) 500ng of sample DNA is mixed into the mastermix. 9) 23µl aliquots distributed into the PCR plate. 10) 3µl of sense primers (amplicons optimised to DMD_55) are added. 11) 3µl of antisense primers (amplicons optimised to DMD_55) are added. 12) plate is cycled through DMD_55 programme. 13) 4µl aliquots from each PCR reaction are arranged onto a check plate in the order the amplicons are found in the gene. The products in the check plate are then run on a 96 well 1% agarose gel. Below are two photos of such gels. B.) A gel image showing correct amplification of all amplicons from the normal control, exon numbers are given for the bottom row. Negative control lanes for DMD_45 (A) and DMD_55 (B) are also highlighted. C) A gel image from a patient with a dystrophin deletion of exons 44-53, shown by the absence of bands for these exons.

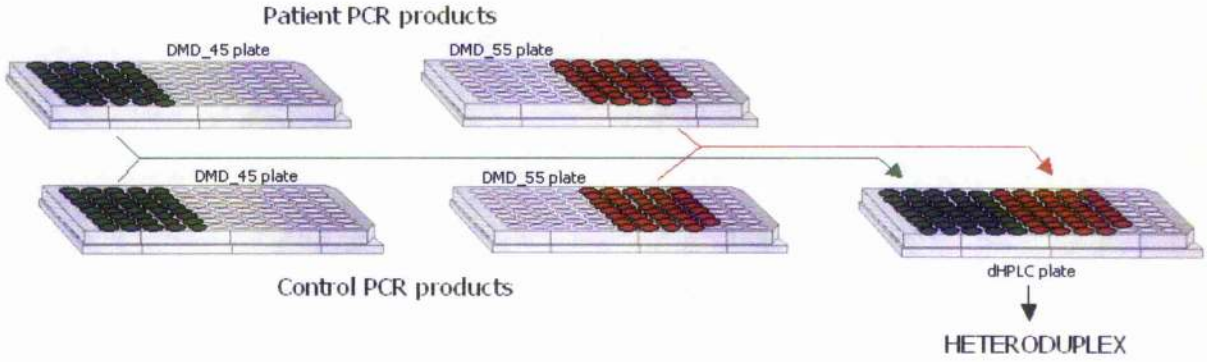
B.)



C.)



A.)



B.)

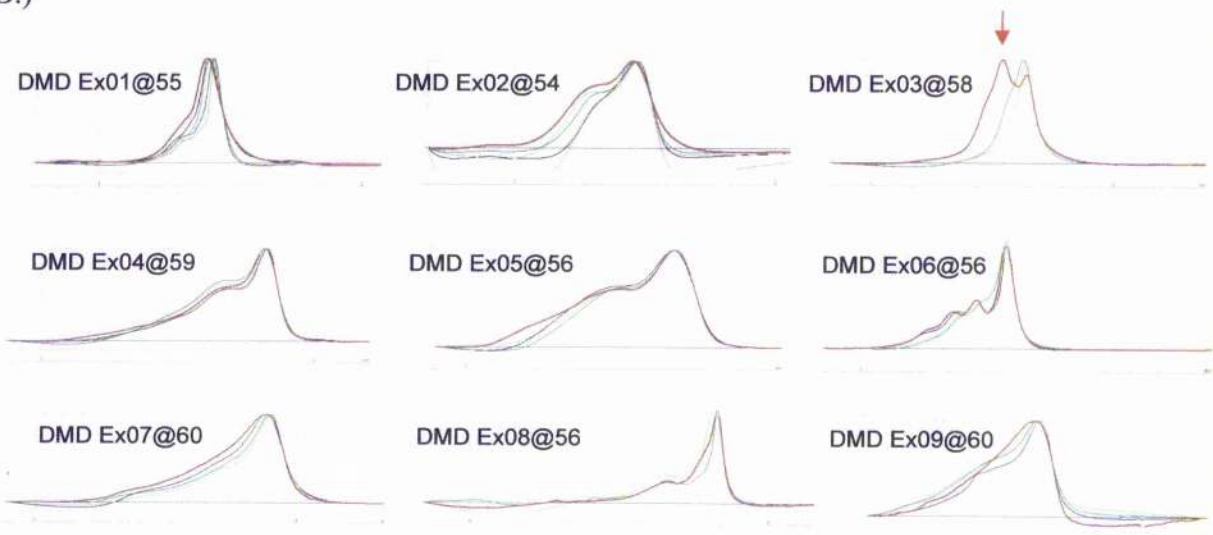


Figure 3.3.2 Forming heteroduplexes and dHPLC analysis. A.) Schematic representation of the robotic steps involved in automated mixing of PCR products from the patient sample and the control sample. Products from the two samples for each amplicon are mixed (24µl each) and placed in a 96 well plate. This is sealed, heated and then cooled slowly, allowing heteroduplexes to form. 17µl of each amplicon mix is then injected onto the dHPLC column using the elution profiles and temperatures determined by the Navigator software. B.) Examples of dHPLC traces for amplicons containing DMD exons 1 –9. Each pane shows the results from one elution method for sample ID No.5. The sample result is highlighted in brown, the other traces are normal control (homozygous) results from previous runs and were used for comparison. The traces have been normalised to account for any differences in intensities between the samples. These results show that sample No.5 does not show any difference from known normal controls in all the traces shown here apart from exon 3. The shift observed in exon 3 (identified by the red arrow) was shown by sequencing to be a heteroduplex caused by the presence of the 94-9dupT polymorphism in the control PCR, it was not present in the patient PCR.

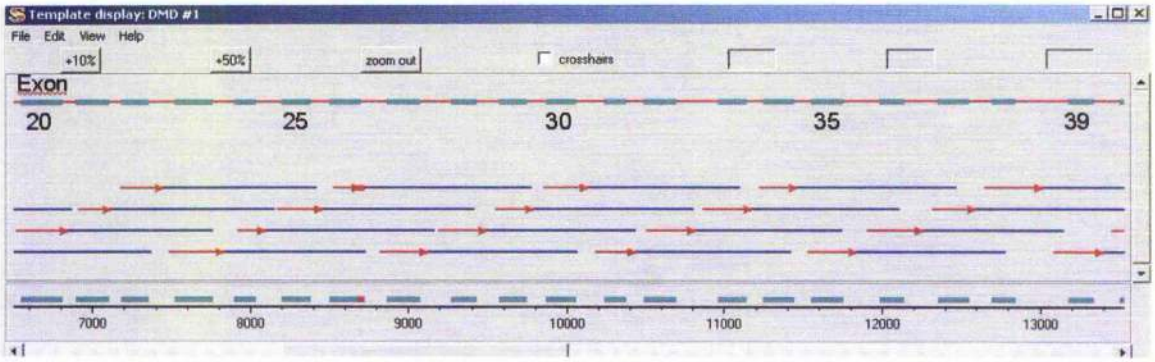
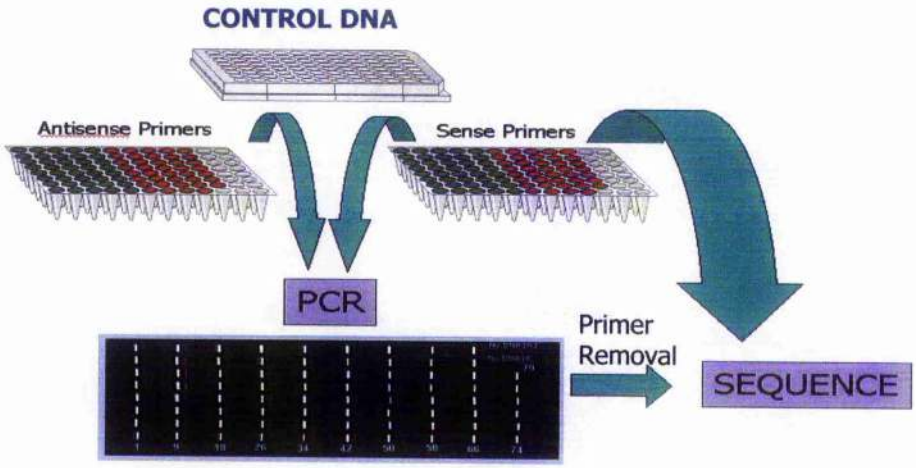


Figure 3.3.3 Automated sequencing of dystrophin exons 1-79. Above is a detail from an alignment of sequences in Gap4. The data is from unidirectional sequencing of dystrophin exons 1-79 in control DNA. Amplicons for each exon were robotically amplified and then purified using SAP/*Exo* I. These purified products were then checked on an agarose gel and 3µl was robotically added to a sequencing reaction containing the sense primer specific to the amplicon. The resulting products were resolved on an ABI3100 analyser. Shown in this detail are alignments for exons 20 to 39 from the normal control. Exons are highlighted in blue on the red upper reference line, clipped sequence that matches the reference sequence is shown by the red arrows, the blue lines represent the (clipped) remainder of each sequence read (this extends past the sequence of the PCR product and includes background from the capillary). The red highlight at exon 26 shows the presence of variance at the poly A tract 5' to exon 26 (3603+15). Comparison of sequence reads against the reference sequence (Method 2.4.1) showed several variants: 94-9dupT, 1993-37G>T, 2645G>A, 3603+15dupA, 8027+11C>T and 8669-75C>G.



Above is a schematic diagram showing the automated steps in the robotic sequencing of dystrophin exons 1-79 for dHPLC analysis. 500ng of control DNA is mixed with a PCR mastermix and aliquoted onto a plate. The DNA is amplified in two PCR programmes (red/green wells) using aliquots of sense and antisense primers (Figure 3.3.1). 4µl of the resulting PCR products are resolved on a 1% agarose check gel. An SAP/*Exo* I digest is used to remove residual primers and free nucleotides from the remainder of the PCR reactions. A 3µl aliquot of each reaction - arranged in the same order as the PCR plate - is then added to a 7µl aliquot of sequencing reagents on a sequencing plate. The sense primers are then aliquoted into the sequencing plate using the same alignment as the initial PCR, this ensures that the primers correspond to the amplicons in the reaction. Finally the robot performs a cycle sequencing reaction and the products are then manually prepared for resolving on an automated sequencer.

3.3.2 Point mutation analysis in dystrophin exons 1-79.

3.3.2.1 Establishing consensus traces and identification of common polymorphisms

Initially four samples (with a previously identified mutation) were analysed by dHPLC for the presence of heteroduplexes within exons 1-79 of the dystrophin gene. dHPLC analysis was performed under the described conditions and profiles were compared using the Navigator software to a control sample mixed with itself to form homoduplexes (Method 2.8.2). Profile comparison indicated the presence of heteroduplexes at the sites of all four mutations and also highlighted the presence of several polymorphic variants within the patient samples that were confirmed by sequencing (Table 3.3.1). Consensus traces of each exon profile where no variance was observed were normalised using the Navigator software and saved to a database. The consensus trace for each exon was then used as a reference against which further sample profiles were compared.

3.3.2.2 Analysis of patient samples

In total DNA samples from 21 cases of DMD, 1 case of intermediate DMD and 1 case of BMD were analysed for point mutation in dystrophin exons 1-79 by robotic PCR followed by dHPLC. All amplicons for each patient could be amplified by PCR, excluding exon deletion in all cases. Sequencing of amplicons in which a shift was identified was performed on the patient and control samples (Method 2.5) and traces were visually compared against each other using the Staden package of programmes. In all cases where a pathogenic mutation was identified the sample was re-amplified and (if possible) an alternative method was used to confirm the sequencing results. The results from this analysis are listed in Table 3.2.

Amplicon	Nucleotide	Amino acid	Frequency	Database
3	94-9dupT	-	0.09	0.10 EUR
6	358-80T>C	-	0.04	-
13	1483-123G>T	-	0.22	0.25 AMER
13	1483-110A>G	-	0.22	0.48 AMER
14	1635A>G	R454R	0.17	0.10 EUR
14	1704+51T>C	-	0.17	0.20 EUR
17	2160+13T>C	-	0.52	0.80 EUR
17	1993-37G>T	-	0.26	0.30 EUR
21	2645G>A	G882D	0.17	0.15 EUR
31	4234-13A>G	-	0.09	0.10 EUR
37	5134G>A	R1745H	0.17	0.30 EUR
38	5448+169A>T	-	0.43	-
41	5922+77_5922+78del	-	0.04	0.20 AMER
43	6118-63_6118-62dup	-	0.13	0.10 EUR
43	6290+27T>A	-	0.04	-
44	6291-115G>A	-	0.17	0.20 EUR
48	7096C>A	K2366Q	0.17	0.20 EUR
49	7200+53C>G	-	0.26	0.30 EUR
53	7728 T>C	N2576N	0.09	0.20 EUR
54	8027+11C>T	-	0.22	0.15 EUR
59	8669-75C>G	-	0.87	-
59	8810A>G	Q2937R	0.04	0.10 AMER
59	8729_8734delinsTGGTCG	[E2910G; N2912D]	0.09	-

Table 3.3.1 A list of the polymorphisms found in this study. The first column indicates the amplicon (exon) in which the change is found, the second column is the nucleotide description - nomenclature is according to Antonarakis (1998) and the reference sequences are as given in Method 2.4.1. The next column is any amino acid changes that are caused by the polymorphism. Then the next column lists the frequency of the change in this study of 23 individuals and next to that, the observed frequency taken from the locus specific database http://www.dmd.nl. The label EUR indicates that this was taken from a report on a European population and the AMER label shows this was from a report on an American population. (–) indicates that the polymorphism is either novel or reported as rare (once per study) in the database.

Patient ID	Exon	Pathogenic variant	Protein effect	Phenotype	Novel	Origin of Mutation		Alternative detection: Enzyme / RG-PCR / Sequence
						Mother	Grandparent Haplotype	
1	61	9100C>T	R3034X	DMD	Y	nd	nd	*Dde* I
X	70	10171C>T	R3391X	DMD	N	Y	nd	*Dde* I – RG-PCR
Y	70	10171C>T	R3391X	DMD	N	Y	GM	*Dde* I – RG-PCR
Z	70	10171C>T	R3391X	DMD	N	Y	nd	*Dde* I – RG-PCR
2	37	5159T>A	L1720X	DMD	Y	Y	GM	*Dra* I
3	08	754G>T	E252X	DMD	Y	Y	nd	*Bsl* I
4	37	5255 T>G	L1752X	DMD	Y	Y	nd	*Tsp*509 I
5	12	1357C>T	Q453X	DMD	Y	N	-	*Xba* I
6	36	5089C>T	Q1697X	DMD	Y	nd	nd	*Tru*9 I – RG-PCR
7	05	313A>T	K105X	DMD	N	nd	nd	*Mse* I
8	64	9337C>T	R3133X	DMD	N	Y	GM	*Acu* I
9	11	1150-1G>T	Splice site	DMD	Y	Y	nd	*Bsl* I – RG-PCR
E	70	10205delA	G3402fsX3412	DMD	Y	nd	nd	*Mnl* I
W	70	10099_10101delGAA	E3367del	IMD*	N	nd	nd	*Mbo* II
10	06	482_484delCCA	T161del	BMD	Y	nd	nd	Sequence
11	57	8438delC	S2813fsX2823	DMD	Y	nd	nd	Sequence
12	57	8405delC	A2802fsX2823	DMD	Y	nd	nd	*Bsr* I
13	23	3036_3037delinsC	E1013fsX1043	DMD	Y	Y	GM	*Dde* I – RG-PCR
14	10	1021_1022insC	L341fsX359	DMD	Y	nd	nd	*Bsa* JI
15	24	3201delT	V1067fsX1071	DMD	Y	Y	GP	*Tru*9 I
16	12	1332-9A>G	Splice?	DMD	N	nd	nd	Sequence
17	48	7098+105_7098+106 insTATTTAATACTTTG	Splice?	DMD	Y	Y	GM	Sequence
18	-	None	-	DMD	-	-	-	-

Table 3.3.2 A table listing the potentially pathogenic variants found in the 23 patients analysed, in grey are the variants recognised as pathogenic mutations. Splice site = recognised splice site mutation. Splice? = a potential splice variant. No pathogenic variant was found in Patient ID. 18. Nomenclature is according to Antonarakis 1998 and the reference sequences are as given in Method 2.4.1. Novel: N=mutation reported previously, Y=Not previously reported. Origin of Mutation: Mother: Y=mutation is present in maternal blood, N=Mutation not present in maternal blood. Grandparent haplotype: (Method 2.15) Informative haplotype evidence for at least two microsatellite markers flanking the mutation shows that the mutation is on a GP=grandpaternal haplotype GM=grandmaternal haplotype. Alternative Detection: Listed are enzymes used according to Method 2.10, Sequencing was according to Methods 2.5 and 2.6. *IMD: the phenotype of the individual was given as Duchenne Muscular Dystrophy but with a slower course (the individual was still ambulatory at 11 years)

3.3.3 Sensitivity of analysis and location of variants:

Of the 23 patient samples studied a recognised cause of the disorder was found in 18 (78%). When the two single amino acid deletions and a further two likely causative splice site mutations the detection rate rises to 22/23 (96%) in this group of patients. This is similar to a previous report for dHPLC sensitivity on dystrophin (75% in eight patients, the remainder having duplications) (Bennett *et al.*, 2001). 18 of the 21 cases with a diagnosis of DMD (85%) had a recognised mutation. Only one of the mutations was seen more than once (1017C>T) and the remainder were distributed throughout the gene sequence. More than one mutation was found in exons 70, 37 and 57 with exon 70 having 4 recognised mutations and one amino acid deletion (1099_10101delGAA).

3.3.4 Mutation type and mechanisms of mutagenesis

3.3.4.1 Base substitutions

The variant nucleotides and local DNA sequence surrounding each potentially pathogenic variant are listed in Figure 3.3.4: highlighted are features in the DNA sequence that may have participated in the generation of the variant. 13/23 of the variants are single base substitutions, with 11 of these 13 generating a stop codon. The remaining two substitutions are believed to affect splicing. 5 of the stop codons generated were CGA>TGA, which has been implicated as a likely target for the antibiotic nonsense read-through therapies. The same 5 CGA>TGA mutations made up all the substitutions located in CpG dinucleotides (38% of all substitutions) which is higher than previously reported in dystrophin (Todorova and Danieli., 1997) but similar with that of a genome wide study (Cooper and Krawczak., 1990). The transition:transversion ratio of all the substitution mutations is 2:1 which is the same as previously reported figures. Small repeat sequences were frequently seen in the immediate flanking sequence of the substitution mutations, it has been suggested that these mediate structures that encourage substitution mutations (Todorova and Danieli., 1997).

Mutation	Protein	Sequence	Repeats	Palindrome/symmetry
10205delA	G3402fsX3412	TGCAGACTGTCTTAGAGGGGGACAACATGGA	run of G	
10099_10101delGAA	E3367del	TTAGACTACATCAGGAGAAGATGTTCGAGACTT	AGA	AGAAGA
482_484delCCA	T161del	ATGTAATCAACTTCACCACCAGCTGGTCTGATGG CC	CCA	ACCAGCTGGT
8438delC	S2813fsX2823	AGCGTCTGCACCTTTCTCTGCAGGAACTTCT	GCA / CT / run of T	-
8405delC	A2802fsX2823	gGTCCCATTTGGAAGCCAGTTCTGACCAGTG	CCA	-
3036_3037delinsC	E1013fsX1043	GAAGAAAGCGCCCTCTGAAATTAGCCGGAAAT	run of C, run of A	-
1021_1022insC	L341fsX359	AGAGTGAAGTAAACCCTGGACCGTTATCAAA	ACC / run of C	-
3201delT	V1067fsX1071	ATGGATGGCTGAAGTTGATGTTTTTCTGAAG	GTT/ CTGAAG	AGTTGA

Figure 3.3.4 Potential pathogenic variants and surrounding sequence. The potential pathogenic variants found in this study highlighted in grey and their surrounding DNA sequences (+/-15bp), repeat motifs are highlighted in yellow and boxed or underlined. Exon sequences are in upper case and intronic sequences in lower case. Repeated stretches of a single nucleotide are also noted. Symmetrical and palindromic elements are listed in the last column. Shown on this page are the deletion and insertion changes, the following page are the nucleotide substitutions and the potential splice site variants are covered separately.

Nucleotide	Protein	Sequence	Repeats	Palindrome/symmetry
9100C>T	R3034X	GTGGCCGTCGAGGACCGAGTCAGGCAGCTGC	GTC / GAG	-
10171C>T	R3391X	TTTGCGAAGCATCCCCGAATGGGCTACCTGC	AT / run of C	-
5159T>A	L1720X	atggaatatagCGTTTAAAGGCAGAACTGAA	GAA / TA / run T and A	TTTAAA
754G>T	E252X	GCCATCCAGGAAGTGGAAATGTTGCCAAGGC	GT / CCA / GAA / run of A	-
5255 T>G	L1752X	ACCACTGCAGGAAATTAGTAGAGCCCCAAAT	AAA / AG	AATT
1357C>T	Q453X	GTTTTAATGGATCTCCAGAATCAGAAACTGA	TC / AAT	-
5089C>T	Q1697X	ACAAAGTGGATCATTCAGGCTGACACACTTT	GG / 5' run of A and T	-
313A>T	K105X	GTAGATGGAAATCATAAACTGACTCTTGGTT	AAA / GG	-
9337C>T	R3133X	ACTGCCATGAAACTCCGAAGACTGCAGAAGG	CT / GAA	-
1150-1G>T	Splice site	cttccttctttgtcagGGGTACATGATGGAT	run of G	

Figure 3.3.4 (cont).

Variant	Protein	Sequence	Repeats	Palindrome/symmetry
1332-9A>G	Splice?	gaggtcataataggcttctttcaaattttcagTTTACAT	TTTCAG / run of A	-
7098+105_7098+106 insTATTTAATACTTTG	Splice?	taaacattttggcttatgccttgagaattatttacctttttaaaat gtattttcctttcagGTTTCCAGAGCTTTACCTGAGAAACAAGGAG AAATTGAAGCTCAAATAAAAGACCTTGGGCAGCTTGAAAAAAAGCT TGAAGACCTTGAAGAGCAGTTAAATCATCTGCTGCTGTGGTTATCT CCTATTAGGAATCAGTTGGAAATTTATAACCAACCAAACCAAGAAG GACCATTTGACGTTAAGgtagggaacttttgctttaaatattttt gctccttttttaagaaaaatggcaatatcactgaattttctcattt ggtatcattattaaagacaaaatattactttatttaatactttggt taaagtgtggtaaggaagactttattcaggataacacaataggcac aggcacca	runs of A and T within intron	-

Figure 3.3.4 (cont).

Splice?= Potential splice site variant.

The 1332-9A>G variant found in patient ID. 16. highlighted in grey is the location of the A>G substitution. 6bp of exon 12 5' acceptor splice site are repeated in the presence of the A>G substitution. Splice site predictions show that the normal 5' acceptor splice site has a consensus score of 85.16 and that the repeated site at 1332-8 in the presence of the variant has a consensus score of 86.48.

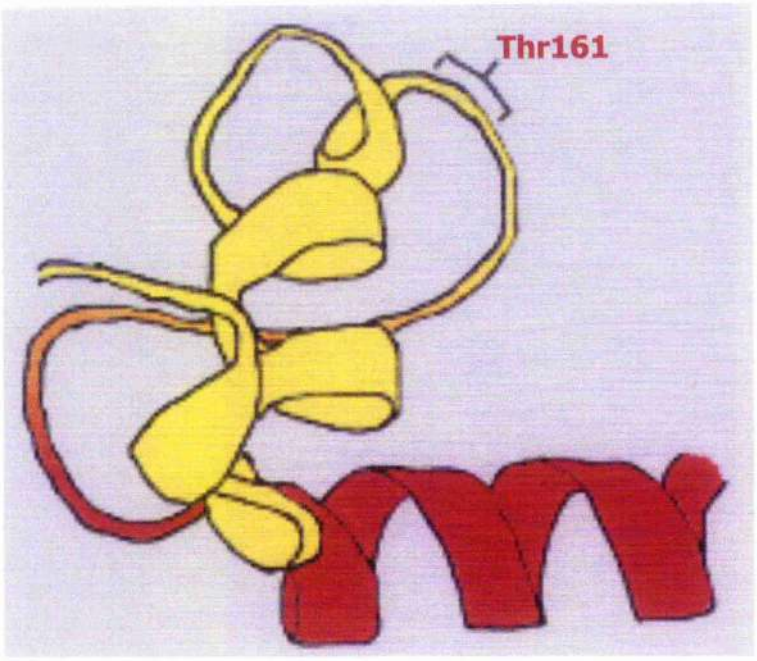
The 14 bp insertion in intron 48 (highlighted in grey) is shown in relation to exon 48 (in upper case).

3.3.4.2 Base deletions/insertions

Deletion or insertion variants were seen in 9 individuals, with the deletion or insertion of one or more base pairs leading to a predicted shift in the reading frame occurring in 6 cases. All deletion or insertion variants were located within or adjacent to repeated sequence; the slipped-mispairing model of mutagenesis explains this type of mutation (Chuzhanova *et al.*, 2003). One delins mutation was seen (3036-3037delinsC) where two nucleotides are replaced by a single nucleotide, causing a shift in the reading frame. This mutation is close to a run of C nucleotides that may be involved in the mutational mechanism.

In two cases an in-frame deletion was seen. In one individual with DMD the loss of a glutamic acid with 10307_10309delGAA was the only potential mutation observed, this has been reported once before and was associated with the milder DMD phenotype plus considerable retention of dystrophin by immunohistochemical staining (Becker *et al.*, 2003). The case in this study had a similar phenotype; a more detailed description of the amino acid sequence context of this variant is given in Figure 3.2.4. This mutation occurs within an AGA repeat and is also within a palindrome, both types of sequence are implicated in generating mutations through slipped-mispairing of DNA during replication. The second in-frame deletion was the only potential mutation observed in a patient with a diagnosis of BMD, this was a novel deletion of a threonine (482_484delCCA T161del) within the actin binding domain of the dystrophin molecule (Figure 3.3.5). Several missense variants at bases within an adjacent α-helix have been associated with BMD. The crystal structure shows the deleted amino acid is located before a small α-helix, close to the site of actin binding in the molecule. This report of an in-frame deletion is consistent with the BMD phenotype. Although only a single amino acid is lost in the protein the location of this small deletion suggests that it may disturb the structural arrangement of the actin-binding site in that part of the protein. This deletion is within a CCA repeat with a symmetrical sequence directly adjacent, both are DNA sequences believed to influence slipped-mispairing during replication.

A.)



B.)

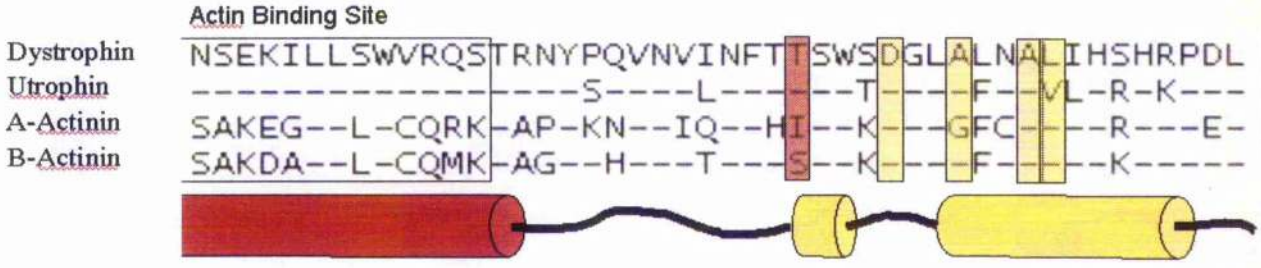


Figure 3.3.5 Sequence alignment and protein structure of T161del. A.) The crystal structure of this region has been determined and a representation of the tertiary structure described in Norwood *et al.* (2000) has be re-drawn here. The actin binding site is shown in red and the yellow indicates internal hydrophobic helices. Missense changes in the yellow helices are believed to destabilise the tertiary structure of the region and suppress actin binding. This data shows that the T161del variant is located immediately before a small internal helix bend. B.) An alignment of a portion of the calponin homology domain of dystrophin (amino acids 135-180) against other human proteins known to share the same domain structure, Utrophin, A-Actinin and B-Actinin. Only amino acids that are different are shown. Dystrophin T161 and equivalent amino acids in related species are highlighted in red, and positions of reported missense mutations are in yellow: D165V, A168D, A171P, L172H (all associated with BMD). The actin binding site in this region is boxed and helical structure is indicated below the sequence.

3.3.4.3 Variants with a potential effect on mRNA splicing

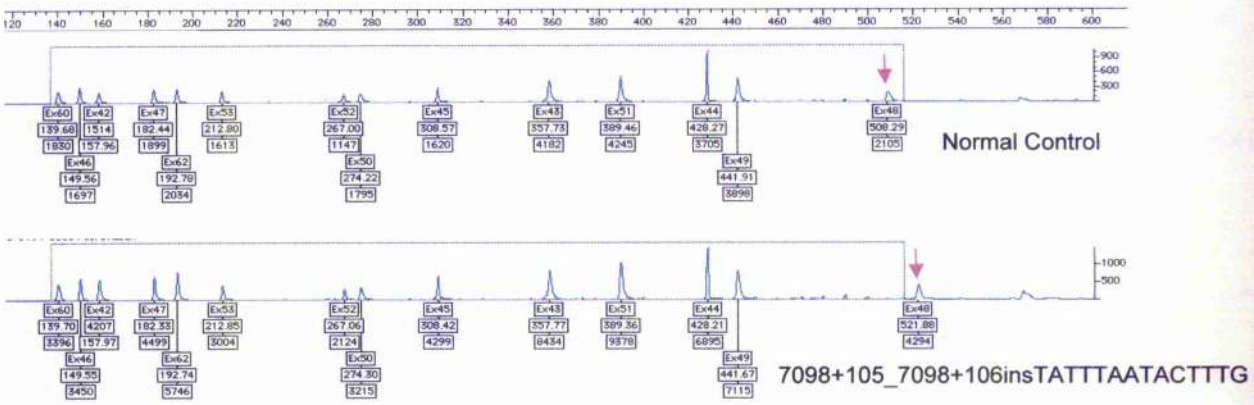
Variants affecting splicing have been described as pathogenic in a significant proportion of both DMD and BMD. There are several recognised types of splicing mutations, altering: Exonic Splicing Enhancer/Silencer (ESE/ESS) binding sites, the consensus splice site, branch sites within the intron and the introduction of cryptic splice sites. The consensus splice site was altered in one patient in this study. The G to T substitution at the consensus -1 position of exon 11 (1150-1G>T) is believed to cause dysfunctional splicing. In this case a reduction in spliceosome affinity for the 5' acceptor splice site of exon 11 is expected to influence the processing of mRNA with an out of frame deletion of exon 11 consistent with the DMD phenotype in the patient. Changes in the consensus AG at the 3' end of the intron are recognised as a cause of the disorder. However, as the 1150-1G>T variant is novel to this study, the precise effect of this change has not been previously characterised at the mRNA level.

A further variant that could affect splicing is 1332-9A>G, this change has been reported before in one American patient and is believed to affect the 5' acceptor splice site of exon 12 (Mendell *et al.*, 2001). Dystrophin 1332-9A>G was the only potential pathogenic variant found in patient ID.16. Using the Shapiro and Senepathy consensus scores this mutation does not significantly alter the splice score at the exon 12 5' acceptor splice site, the variant increases the score from 85.16 to 85.31. However the variant is shown to generate a possible cryptic splice site at 1332-8 with a consensus score of 86.48, higher than the normal exon 12 5' acceptor splice site (Figure 3.3.4). If alternative splicing were to occur this would introduce 8 nucleotides into exon 12, producing a shift in the reading frame resulting in a truncated transcript, consistent with a diagnosis of DMD in the patient.

One intronic change was not associated with a change in the consensus splice site of the adjacent exon, this was: 7098+105_7098+106insTATTTAATACTTTG (adjacent to exon 48), the only novel change in patient ID.17. This novel variant can be detected by fluorescent PCR used in routine clinical testing and was not observed in a retrospective audit of over 500 chromosomes from the same population (data from the West of Scotland Regional Genetics Laboratory, Glasgow, UK). Analysis for consensus scores in the intron shows two potential 3' donor splice sites (score>70) the original exon 48 splice site having a score of 84.67 and a second splice site having a higher score of

87.59. The second predicted splice site at 7098+116 is 11 bp 3' from the insertion, presence of the inserted sequence does not change the consensus score of this potential splice site (Figure 3.3.6). ESE sites are binding sites for proteins involved in the splicing of RNA. Mutations altering ESE sites have been shown to alter splicing in DMD (Tuffery-Giraud *et al.*, 2004). Analysis for potential ESE sites shows that the insertion variant introduces an SC35 ESE site before the potential 3' donor splice site. ESE sites identified by this programme have been shown to cluster in regions where sites have been experimentally mapped (Figure 3.3.7).

A.)



B.)

Taaacattttggcttatgccttgagaattatttacctttttaaaatgtattttccttt
cagGTTTCCAGAGCTTTACCTGAGAAACAAGGAGAAATTGAAGCTCAAATAAAAGACC
TTGGGCAGCTTGAAAAAAAGCTTGAAGACCTTGAAGAGCAGTTAAATCATCTGCTGCT
GTGGTTATCTCCTATTAGGAATCAGTTGGAAATTTATAACCAACCAAACCAAGAAGGA
CCATTTGACGTTAAGgtagggaacttttgctttaaatatttttgctccttttttaag
aaaaatggcaatatcactgaattttctcatttggtatcattattaaagacaaaatatt
actttatttaatactttggttaaagtgtggtaaggaagactttattcaggataacaca
ataggcacaggcacca

Figure 3.3.6 Analysis of 7098+105_7098+106insTATTTAATACTTTG. A.) Results from the analysis of patient ID. 17 and a normal control using a fluorescent multiplex PCR of 14 dystrophin exons (Yau *et al.*, 1996). Observed fluorescence (Y axis) is shown against size of fragment (X axis). Under each peak is an exon label with an automatic calculation of size (bp) and peak area. Exon 48 in each sample is shown by the red arrow and is reported as 13.59bp larger in patient ID. 17 than the normal control exon 48 fragment. Sequencing showed a 14bp insertion in patient ID.17 (7098+105_7098+106insTATTTAATACTTTG). B.) Exon 48 amplicon sequence, The exon sequence is in upper case and the inserted sequence is highlighted in grey. A potential cryptic splice site is highlighted in red. The cryptic splice site has a consensus score of 87.59 whilst the conventional donor splice site has a consensus score of 84.67 - the presence of the insertion does not alter these scores.

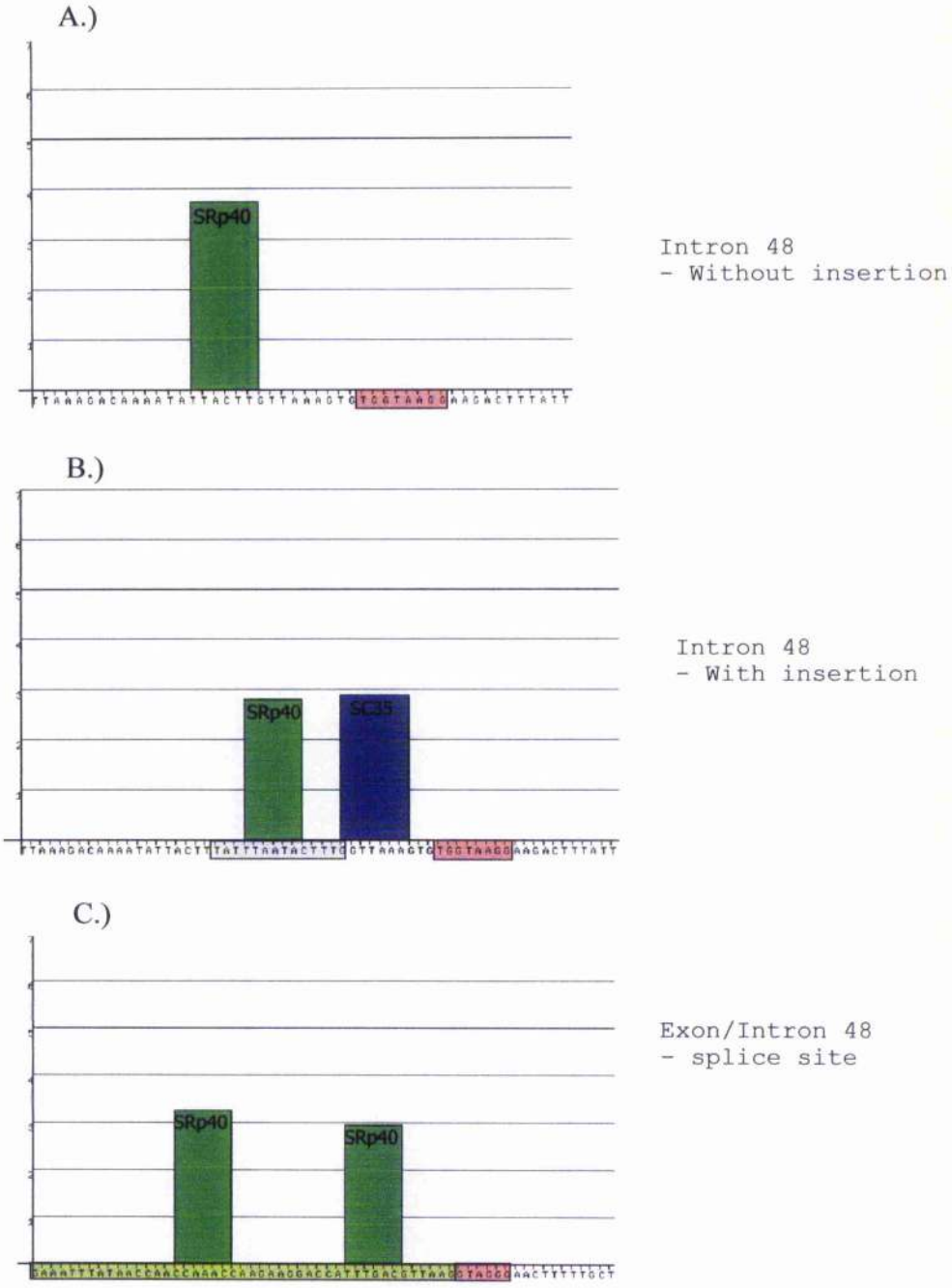


Figure 3.3.7 Output from ESE finder for intron 48. A.) Output from ESE finder (http://exon.cshl.edu/ESE/) showing the predicted cryptic 3' donor splice site at 7098+116 within intron 48, shown in red is the consensus splice site. Significant motif scores for exon splicing enhancer SRp40 were present in the non-inserted intronic sequence. B.) output from ESE finder showing the predicted cryptic 3' donor splice site at 7098+116 within intron 48, shown in red is the consensus splice site. The inserted sequence is shown in grey. Significant motif scores for exon splicing enhancer SRp40 and an additional SC35 sequence were present in the intronic sequence with the insertion. C.) output from ESE finder showing the conventional Exon 48 3' donor splice site, the coding sequence is highlighted in green and the splice site in red. Two significant motif scores for SRp40 were observed.

3.3.5 Polymorphisms

Unidirectional sequencing showed several variants from the reference sequence in the normal control (Table 3.3.1), these have all been reported as common polymorphisms in the locus specific database for DMD (www.dmd.nl). Table 3.3.1 lists the polymorphism location and frequency in this study and in others. As some amplicons contain more than one common polymorphism and some polymorphisms were present in the control, these amplicons needed sequencing in the majority of cases. Only one variant (358-80T>C) found in Patient ID.2 has not been seen before, as this was observed with a recognised pathogenic mutation on the same allele in this patient it is most likely represents a rare polymorphism. Two of the polymorphisms although unique to this study were observed in several individuals (5448+169A>T; 8669-75C>G). All other variants found in the patient samples have been reported as polymorphic in other studies. The frequency of polymorphisms within the group of patients showed that some variants were observed at different frequencies to that reported in other european studies. One polymorphism 6290+27T>A in patient ID.1 has only been reported twice before, in this instance the patient had a nonsense mutation on the same allele and was of asian origin. A second rarely reported variant was seen twice in this study: 8729_8734delinsTGGTCG: this is a compound of the reported polymorphisms 8734A>G (Asn2912Asp) and 8729A>T (Glu2910Val), neither were seen independently in this study. When this change was first observed in this study only 8729A>T had been identified as a polymorphism and two previous reports of 8729_8734delinsTGGTCG classified the variant as associated with a severe form of BMD. In this study 8729_8734delinsTGGTCG is seen once with a nonsense change on the same allele and once with potential splice site change on the same allele. Combining the occurrence in cis with a nonsense mutation with the relatively recent report of 8734A>G being a polymorphism this data suggests a non-pathogenic role for the 8729_8734delinsTGGTCG variant.

3.3.6 Origin of mutations

For 12 of the 23 probands a maternal sample was available and carrier testing had been requested in all cases. Only 1 of the 12 cases was shown to be a new mutation, this lower than expected (30%) new mutation rate may be explained by the ascertainment bias for family history in some of the cases. Haplotype analysis performed in 6 of the pedigrees showed that 1 of the mutations occurred on a grandpaternal haplotype and 5 on a grandmaternal haplotype, once again selection of cases with positive family history could have influenced these results.

4.0 Discussion and future perspectives

4.1 Cognitive impairment and mutation in Dp71

In results 3.1 the reported possible phenotype genotype relationship between cognitive impairment and mutations in the Dp71 coding sequence was investigated. If mutations in this region were found more frequently in DMD/BMD patients with cognitive impairment than those without this would be an important finding. A phenotype genotype correlation is indicative of functional significance and would contribute to understanding dystrophin biology. The correlation could also have identified a mutation detection protocol for this sub-set of patients. In total a cohort of 10 patients who had been diagnosed with DMD/BMD and cognitive impairment were studied. The percentage of the cohort in whom recognised mutations were found in the Dp71 region was 10%. In the subsequent study of 23 DMD/BMD patients where no cognitive impairment had been noted (results 3.3), 1 recognised mutation was found in 3 unrelated patients in the Dp71 coding sequence (13%), this mutation has been previously reported in a patient with cognitive impairment (Moizard *et al.*, 2000) and in other patients where no neurological examinations have been reported (Leiden dystrophin mutation database http://www.dmd.nl). The comparison between the results presented here does not support the suggestion of a genotype-phenotype correlation within this region of the gene. We suggest that the distribution of cognitive impairment amongst affected individuals is too broad to increase the sensitivity of a limited screen based on the presence of learning difficulties. The reports of point mutation in Dp71 associated with mental retardation indicate that the individuals with mutation were all severely neuropsychologically impaired (Lenk *et al.*, 1993; Kekou *et al.*, 1999; Moizard *et al.*, 2000). It is possible that the neurological phenotype may only be associated with specific mutations or that the mutations which predispose to the additional phenotype may have a variable penetrance. Any genotype phenotype association in dystrophin is likely to be better defined with an increase in the number of reported mutations, in particular repeated reports of the same mutation.

To definitively answer the question of whether mutations in the Dp71 coding sequence are associated with DMD and cognitive impairment further investigation needs to be carried out on larger cohorts of rigorously diagnosed patients. As cognitive abilities in DMD can be affected by many factors, guidelines would need to be agreed for the

accurate study of cognitive dysfunction to ensure a comparable cohort was being investigated.

4.2 The sensitivity and efficiency of dHPLC for analysing mutation in Dp71

In results 3.1 the sensitivity of dHPLC was equal to that of the sequencing, all variants within the coding region were detected by both methods. Although the polymorphism 9974+13_9974+23(A)9-12 was not detected by dHPLC (Figure 3.1.7) the parameters used for dHPLC analysis detection were not specific to this region. The poly-A tract lies within the intron and is predicted to be fully melted at the temperatures used in this study. Avoiding a positive signal from the intronic polymorphism by dHPLC was efficient as it reduced the amount of post-dHPLC sequencing required.

4.3 Analysis of exon 70 in a cohort of 46 patients with DMD/BMD

In results 3.2 robotic amplification of exon 70 followed by dHPLC identified four variants within a total of 46 patients who had been referred for DMD/BMD analysis and have no detectable deletion or duplication. This is a higher than expected level of detection for exon 70 if considered by proportion of coding sequence alone: 6.5% confirmed mutations in 1.2% of coding sequence screened. Based on this exon 70 could be considered as a candidate exon for a preliminary screen.

4.4 Investigations into the inheritance of a recurrent mutation in exon 70

In results 3.2 the R3391X (10171C>T) nonsense mutation was identified on a background of three different haplotypes and was shown to be a new mutation in at least one case. The mutation lies within a CGA codon, a sequence shown to have a 20 fold incidence of nonsense mutation compared to any other potential nonsense mutation site in DMD (Prior *et al.*, 1995). This data is therefore a report of recurrent mutation within one of these sites and when the three cases here are included, makes this mutation one of the most frequently recorded point mutations in DMD. This data suggests that regional hotspots for mutation may exist in the dystrophin gene, however these may be difficult to identify as there are many potential sites for mutation within the dystrophin gene. Combinations of recurrent mutation within a single exon could identify exons

with an increased likelihood of mutation. This should become more apparent as more mutation data becomes available.

The mitochondrial HV1 studies in results 3.2 demonstrate that the local control population studied had a relatively low level of HV1 diversity, further studies may establish a more accurate estimation of this. There was also a high level of similarity in the dystrophin intragenic markers used and although only three individuals were analysed this suggests that the distribution of alleles in the Scottish population might be smaller than that reported in other populations.

4.5 dHPLC analysis of dystrophin exons 1-79 in 23 individuals

The large amount of data presented in results 3.3 represents a comprehensive study of the dystrophin gene sequence by dHPLC in 23 individuals. Robotic amplification by PCR and analysis by dHPLC and sequencing demonstrates a semi-automated approach that could be used to detect point mutation in any gene. This approach is relatively rapid and demonstrated a high level of sensitivity for variance in the samples studied. Automation improved the accuracy and speed of the PCR and dHPLC process whilst reducing user time. Generating the dHPLC and sequence data in patient specific databases rapidly completed the analysis for each patient in turn. This approach may be suitable to the analysis of genes where the frequency of referrals makes waiting to complete a batch of samples impractical. This is pertinent to DMD/BMD where the number of referrals may be small (in Scotland estimates are 3-18 per year) but requests for urgent testing are possible (e.g. the pregnant mother of a sporadic case could be offered prenatal mutation analysis based on the rapid detection of the mutation in her son).

4.6 Sensitivity of the technique

Confirmation of the dHPLC results by sequencing in results 3.1 showed that the technique had a high level of sensitivity for point mutations. Subsequent detection of novel variants and polymorphisms across the gene sequence in the group of 23 patients also demonstrated its ability to identify a range of variants (results 3.3). The proportion of mutations seen in the group of patients reported here is similar to those reported in other whole gene studies (Mendell *et al.*, 2001;Bennet *et al.*, 2001). The overall

sensitivity was 96% for either recognised or potentially pathogenic mutations. The sensitivity of dHPLC is however dependent on sequence specific factors implying that some variants may not be detected by the technique. One example is the poly-A tract in dystrophin exon 68 where no variance by dHPLC was observed, this was shown to be most likely affected by the low local GC content, if this situation is present in the coding sequence this may reduce the sensitivity of dHPLC at this point. This may be particularly relevant given the repeated occurrence of mutations after a run of the same nucleotide. All patients with truncating mutations had a clear diagnosis of DMD. The one patient with DMD who had no detectable mutation could either represent a lack of sensitivity in the method, a mutation outside the region covered, or a duplication not detected by the fluorescent dosage multiplex PCR. The use of MLPA or mRNA studies could help determine the cause of mutation in this case. A combination of the multiplex PCR (Yau *et al.*, 1996) and dHPLC point mutation analysis should detect 90-94% of mutations in DMD (85-95% dHPLC sensitivity in DMD in 21 patients - depending on identity of variants). Introduction of MLPA would detect a further 1% (missed duplications) and mRNA studies are expected to detect the remainder (Yan *et al*, 2004).

4.7 **Efficiency of the technique**

Although the Staden package facilitated the process, sequencing and data analysis is still costly and laborious. Compared to complete sequencing of the gene, dHPLC reduced the amount of sequence analysis by a factor of 10 and reduced costs by approximately 70-80%. The time taken to analyse the whole gene by automated sequencing or dHPLC was equal in the single example given here. The use of a heteroduplex based approach required the characterisation of the control allele sequence as well as that of the sample. Polymorphism accounted for a large proportion of the sequencing required with the control's polymorphisms contributing significantly to the workload for each patient (Figure 3.3.2). Mixing the patient PCR products with PCR products containing each allele of the common polymorphisms would distinguish the patient allele by dHPLC resolving both homoduplexes and heteroduplexes as well as screening for mutation at the same time. For example, screening for the polymorphisms occurring above a frequency of 0.2 in this study would require an additional 8 PCR products (which could be achieved using redundant wells on the 96 well plate). This would be expected to reduce the amount of sequencing required, further improving the efficiency of this dHPLC protocol. Advances in sequencing software are also expected,

with an overall aim of fully automated analysis. Combining the automated sequencing of the whole gene as shown in figure 3.3.3 and a reduction in sequencing costs it may become possible that automatic sequencing of the entire coding region using these protocols will be equally (or more) efficient.

4.8 **Distribution and types of mutation**

The distribution of the mutations observed covered exons from across the gene. More than one recognised mutation was found in exons 70, 37 and 57. In exon 70 the mutation 10171C>T was found in three separate pedigrees. The three 10171C>T mutations were shown to have occurred independently by the use of intragenic markers and mitochondrial haplotypes (Results 3.2). Although this suggests that certain exons may have a raised likelihood of mutation, the distribution of reported mutations within the gene show that mutations are found in virtually all exons of the gene (www.dmd.nl). Thus sensitive analysis of the gene should consider the entire coding sequence.

The mechanism of each mutation was considered: showing that CpG dinucleotides and runs of nucleotides are common features of the mutations reported. Based on studies like this it may be possible to identify particular regions of the gene with an increased chance of a certain type of mutation. The selection of samples with more than one affected family member introduced potential bias into studies on the origin of the mutations. With only one mutation shown to have occurred in the maternal germline, this is suggestive of a grandparental origin for many of the mutations however we could not determine the true grandparental ratio due to the aforementioned selection in certain pedigrees. Further haplotype studies to determine if there is a sex bias to the origin of point mutations would help contribute to estimates of point mutation carriers in the general population.

4.9 **Uncharacterised variants**

The majority of variants detected were predicted to cause a truncated protein by their effect on the amino acid coding sequence and are all associated with a DMD phenotype. Only one of the intronic variants observed was a recognised cause of disorder whilst five changes identified in this study could not be confidently characterised as pathogenic by DNA results alone. These are the intronic variants 1332-9A>G,

7098+105_7098+106insTATTTAATACTTTG, 10223+6T>C and the two in-frame deletions of 10099_10101delGAA and 482_484delCCA. Only two of these changes (1332-9A>G and 10099_10101delGAA) have been reported before and were associated with disease in both cases. It has not yet been confirmed that small in-frame deletions are mutations of structural significance or rare deleterious polymorphism. The patient with the 10099_10101delGAA variant had a diagnosis of a less severe DMD phenotype that is similar to the other reported case in the literature. Unfortunately no immunohistochemistry results were available as this may have revealed if, as in the previous report, dystrophin was produced and observed at the cell membrane. The similarity of phenotypes between the reported patient and the one in this study with this mutation is evidence to suggest that this variant does have a distinct effect on dystrophin function. The patient with the 482_484delCCA variant had a diagnosis of BMD, this could be expected from the suggested alteration in protein structure. No immunohistochemical results were available from this individual but the deletion was shown to be adjacent to an actin binding domain. As actin binding occurs in several domains of the protein, a local change in structure in a single domain may not be sufficient to fully remove binding but capable of lowering actin affinity to a level expected to cause a BMD phenotype. Investigation into the effects of the intronic variants (1332-9A>G, 7098+105_7098+106insTATTTAATACTTTG and 10223+6T>C) on spliceosome recognition sites (Figures 3.3.6 and 3.3.7) was suggestive of an effect on splicing but remains inconclusive. Further work including mRNA studies would be an appropriate method for further characterisation of the intronic changes whilst the single codon deletions could both be candidate mutations for binding assays. The 10223+6T>C variant identified in results 3.1 may also be a candidate for mRNA analysis, unfortunately insufficient DNA was available for further analysis of the coding region in this sample. In all the cases with an unconfirmed variant, further family studies could exclude the change as a mutation if it is observed in unaffected males or suggest a pathogenic effect if it is shown to have arisen in a recent transmission.

4.10 Analysis in BMD

Only one case of BMD was analysed for point mutation in this study (results 3.3). Determining the contribution of dystrophin point mutation in BMD is expected to reveal a more diverse range of mutation types (i.e. splicing, missense, deletion/insertions and nonsense/frameshift mutations). There will be challenges in interpreting dystrophin

missense changes and this may require investigations into their effect on the dystrophin structure. Furthermore, a significant proportion of BMD mutations have been found to have an unpredictable effect on RNA splicing. Some BMD mutations are shown to be nonsense mutations that do not lead to nonsense-mediated decay but (as has also been reported in other genes) affect splicing of in-frame exons in the mRNA (Ginjaar *et al.*, 2000; Fakusova *et al.*, 2001). This splicing removes the mutation-containing sequence producing a shortened mRNA. Also, the incidence of non-consensus intronic mutation (cryptic splice site/branch site mutation) has been reported at 44% of all point mutations in a cohort of 9 BMD patients (Yau *et al.*, 2003) and these have only been detected by PTT. Combined, these reports indicate that mRNA studies would be required to detect and characterise a larger proportion of mutations in BMD than in DMD.

4.11 **Polymorphism within the study population**

A variety of polymorphisms were detected within the group of patients. The frequencies of several of the polymorphisms varied from that reported in other studies; this could be a result of the geographical differences between populations. Some polymorphisms were novel to this study 358-80T>C, 5448+168A>T and 8669-75C>G with the latter two being observed in several individuals. These results suggest that this may be the first time these polymorphisms have been included in amplicons within a study of the dystrophin gene. In two instances a rare polymorphism (each reported twice before in separate studies) was observed on the same allele as a nonsense mutation providing evidence for their non-pathogenic status. The first of these was 6290-27T>A seen in an individual of Asian origin. The second was the more complex change of 8729_8734delinsTGGTCG which we believe is a compound of the 8729A>T and 8734A>G polymorphisms, this had been previously associated with a severe form of BMD, but we believe this should be considered a polymorphism.

References

Abbs S, Yau SC, Clark S, Mathew CG, Bobrow M. (1991) A convenient multiplex PCR system for the detection of dystrophin gene deletions: a comparative analysis with cDNA hybridisation shows mistypings by both methods. J Med Genet. 28:304-11.

Abbs S, Roberts RG, Matthew CG, Bentley DR, Bobrow M (1990) Accurate assessment of intragenic recombination frequency within the Duchenne muscular dystrophy gene. Genomics 7:602-6

Angelini C, Fanin M, Freda MP, Martinello F, Miorin M, Melacini P, Siciliano G, Pegoraro E, Rosa M, Danieli GA.(1996) Prognostic factors in mild dystrophinopathies. J Neurol Sci. 142:70-8.

Antonarakis SE. (1998) Recommendations for a nomenclature system for human gene mutations. Nomenclature Working Group.Hum Mutat. 11:1-3.

Azofeifa J, Voit T, Hubner C, Cremer M. (1995) X-chromosome methylation in manifesting and healthy carriers of dystrophinopathies: concordance of activation ratios among first degree female relatives and skewed inactivation as cause of the affected phenotypes. Hum Genet. 96:167-76.

Bakker E, Van Broeckhoven C, Bonten EJ, van de Vooren MJ, Veenema H, Van Hul W, Van Ommen GJ, Vandenberghe A, Pearson PL. (1987) Germline mosaicism and Duchenne muscular dystrophy mutations. Nature. 329:554-6.

Becker K, Robb SA, Hatton Z, Yau SC, Abbs S, Roberts RG. (2003) Loss of a single amino acid from dystrophin resulting in Duchenne muscular dystrophy with retention of dystrophin protein. Hum Mutat. 21:651.

Becker PE, Keiner F. (1955) Eine neue X-chromosomale Muskeldystrophie. Archiv fur Psychiatrie und Zeitschrift Neurologie. 193:427-48

Beggs AH, Koenig M, Boyce FM, Kunkel LM. (1990) Detection of 98% of DMD/BMD gene deletions by polymerase chain reaction. Hum Genet. 86:45-8.

Bennett RR, den Dunnen J, O'Brien KF, Darras BT, Kunkel LM. (2001) Detection of mutations in the dystrophin gene via automated DHPLC screening and direct sequencing. BMC Genet. 2:17.

Blake DJ, Kroger S. (2000) The neurobiology of Duchenne muscular dystrophy: learning lessons from muscle? Trends Neurosci. 23:92-9.

Bonfield JK, Rada C, Staden R. (1998) Automated detection of point mutations using fluorescent sequence trace subtraction. Nucleic Acids Res. 26:3404-9.

Bonsett CA (1969) Studies of pseudohypertrophic muscular dystrophy. Thomas, Springfield.

Bradley WG, Jones MZ, Mussini JM, Fawcett PR.(1978) Becker-type muscular dystrophy. Muscle Nerve. 1:111-32.

Bushby KM, Gardner-Medwin D. (1993) The clinical, genetic and dystrophin characteristics of Becker muscular dystrophy. I. Natural history. J Neurol. 240:98-104.

Byers TJ, Lidov HG, Kunkel LM. (1993) An alternative dystrophin transcript specific to peripheral nerve. Nat Genet. 4:77-81.

Cartegni L, Wang J, Zhu Z, Zhang MQ, Krainer AR. (2003) ESEfinder: A web resource to identify exonic splicing enhancers. Nucleic Acids Res. 31:3568-71.

Chan YM, Bonnemann CG, Lidov HG, Kunkel LM. (1998) Molecular organization of sarcoglycan complex in mouse myotubes in culture. J Cell Biol. 143:2033-44.

Chuzhanova NA, Anassis EJ, Ball EV, Krawczak M, Cooper DN. (2003) Meta-analysis of indels causing human genetic disease: mechanisms of mutagenesis and the role of local DNA sequence complexity. Hum Mutat. 21:28-44.

Comi GP, Prelle A, Bresolin N, Moggio M, Bardoni A, Gallanti A, Vita G, Toscano A, Ferro MT, Bordoni A, *et al.* (1994) Clinical variability in Becker muscular dystrophy. Genetic, biochemical and immunohistochemical correlates. Brain. 117:1-14.

Cooper DN, Krawczak M. (1990) The mutational spectrum of single base-pair substitutions causing human genetic disease: patterns and predictions. Hum Genet. 85:55-74.

D'Souza VN, Nguyen TM, Morris GE, Karges W, Pillers DA, Ray PN. (1995) A novel dystrophin isoform is required for normal retinal electrophysiology. Hum Mol Genet 4:837-42.

den Dunnen JT, Antonarakis SE. (2000) Mutation nomenclature extensions and suggestions to describe complex mutations: a discussion. Hum Mutat. 15:7-12.

Ehmsen J, Poon E, Davies K. (2002) The dystrophin-associated protein complex. J Cell Sci. 115:2801-3.

Emery AE, Skinner R. (1976) Clinical studies in benign (Becker type) X-linked muscular dystrophy. Clin Genet. 10:189-201.

Emery AE, (1993) Duchenne Muscular Dystrophy (2nd edn.), Oxford University Press, Oxford.

Emery AE, (2001) Duchenne muscular dystrophy or Meryon's disease. In: Emery AE (ed) The Muscular Dystrophies Oxford University Press, Oxford. pp55-71

Emery A (2002).Duchenne and other X-linked muscular dystrophies In: Principles and Practice of Medical Genetics, 3rd edn. Eds. Rimoins DL, Connor JM, Pyeritz RI. New York, Churchill Livingston, pp 3266-84

England SB, Nicholson LV, Johnson MA, Forrest SM, Love DR, Zubrzycka-Gaarn EE, Bulman DE, Harris JB, Davies KE. (1990) Very mild muscular dystrophy associated with the deletion of 46% of dystrophin. Nature. 343:180-2.

Fabbrizio E, Nudel U, Hugon G, Robert A, Pons F, Mornet D. (1994) Characterization and localization of a 77 kDa protein related to the dystrophin gene family. Biochem J. 299:359-65.

Fajkusova L, Lukas Z, Tvrdikova M, Kuhrova V, Hajek J, Fajkus J. (2001) Novel dystrophin mutations revealed by analysis of dystrophin mRNA: alternative splicing suppresses the phenotypic effect of a nonsense mutation. Neuromuscul Disord. 11:133-8.

Finsterer J, Stollberger C. (2003) The heart in human dystrophinopathies. Cardiology. 99:1-19.

Flanigan KM, von Niederhausern A, Dunn DM, Alder J, Mendell JR, Weiss RB. (2003) Rapid direct sequence analysis of the dystrophin gene. Am J Hum Genet. 72:931-9.

Ginjaar IB, Kneppers AL, van der Meulen JD, Anderson LV, Bremmer-Bout M, van Deutekom JC, Weegenaar J, den Dunnen JT, Bakker E. (2000) Dystrophin nonsense mutation induces different levels of exon 29 skipping and leads to variable phenotypes within one BMD family. Eur J Hum Genet. 8:793-6.

Goldberg LR, Hausmanowa-Petruscwicz I, Fidzianska A, Duggan DJ, Steinberg LS, Hoffman EP. (1998) A dystrophin missense mutation showing persistence of dystrophin and dystrophin-associated proteins yet a severe phenotype. Ann Neurol 44: 971-976.

Grimm T, Meng G, Liechti-Gallati S, Bettecken T, Muller CR, Muller B. (1994) On the origin of deletions and point mutations in Duchenne muscular dystrophy: most deletions arise in oogenesis and most point mutations result from events in spermatogenesis. J Med Genet. 31:183-6.

Helderman-van den Enden AT, Ginjaar HB, Kneppers AL, Bakker E, Breuning MH, de Visser M. (2003) Somatic mosaicism of a point mutation in the dystrophin gene in a patient presenting with an asymmetrical muscle weakness and contractures. Neuromuscul Disord. 13:317-21.

Hilleren P, Parker R. (1999) mRNA surveillance in eukaryotes: kinetic proofreading of proper translation termination as assessed by mRNP domain organization? RNA. 5:711-9.

Hofstra RM, Mulder IM, Vossen R, de Koning-Gans PA, Kraak M, Ginjaar IB, van der Hout AH, Bakker E, Buys CH, van Ommen GJ, van Essen AJ, den Dunnen JT. (2004) DGGE-based whole-gene mutation scanning of the dystrophin gene in Duchenne and Becker muscular dystrophy patients. Hum Mutat. 23:57-66.

Kekou K, Mavrou A, Florentin L, Youroukos S, Zafiriou DI, Skouteli HN, Metaxotou C. (1999) Screening for minor changes in the distal part of the human dystrophin gene in Greek DMD/BMD patients. Eur J Hum Genet. 7:179-87.

Kerr TP, Sewry CA, Robb SA, Roberts RG. (2001) Long mutant dystrophins and variable phenotypes: evasion of nonsense-mediated decay? Hum Genet. 109:402-7.

Kostakow St, Derix F. (1937) Familienforschung in einer muskeldystrophischen Sippe und die Erbprognose ihrer Mitglieder. Deutsche Archiv fur klinischen Medizin. 180:585-6.

Lenk U, Hanke R, Thiele H, Speer A. (1993) Point mutations at the carboxy terminus of the human dystrophin gene: implications for an association with mental retardation in DMD patients. Hum Mol Genet. 2:1877-81.

Lenk U, Oexle K, Voit T, Ancker U, Hellner KA, Speer A, Hubner C. (1996) A cysteine 3340 substitution in the dystroglycanbinding domain of dystrophin associated with Duchenne muscular dystrophy, mental retardation and absence of the ERG bwave. Hum Mol Genet 5: 973-975.

Lidov HG, Selig S, Kunkel LM. (1995) Dp140: a novel 140 kDa CNS transcript from the dystrophin locus. Hum Mol Genet. 4:329-35.

Lu QL, Morris GE, Wilton SD, Ly T, Artem'yeva OV, Strong P, Partridge TA (2000) Massive idiosyncratic exon skipping corrects the nonsense mutation in dystrophic mouse muscle and produces functional revertant fibers by clonal expansion. J Cell Biol. 148:985-96.

Manzur AT (2001) Medical Management and treatment of muscular dystrophy In: Emery AE (ed) The Muscular Dystrophies Oxford University Press, Oxford. pp223-246

Mehler MF. (2000) Brain dystrophin, neurogenetics and mental retardation. Brain Res Rev. 32:277-307.

Mendell JR, Buzin CH, Feng J, Yan J, Serrano C, Sangani DS, Wall C, Prior TW, Sommer SS. (2001) Diagnosis of Duchenne dystrophy by enhanced detection of small mutations. Neurology. 57:645-50.

Moizard MP, Billard C, Toutain A, Berret F, Marmin N, Moraine C. (1998) Are Dp71 and Dp140 brain dystrophin isoforms related to cognitive impairment in Duchenne muscular dystrophy? Am J Med Genet. 80:32-41.

Moizard MP, Toutain A, Fournier D, Berret F, Raynaud M, Billard C, Andres C, Moraine C. Severe cognitive impairment in DMD: obvious clinical indication for Dp71 isoform point mutation screening. Eur J Hum Genet. (2000) 8:552-6.

Monaco AP, Bertelson CJ, Leichti-Gallati S, Moser H Kunkel LM. (1988) An explanation for the phenotypic differences between patients bearing partial deletions of the DMD locus. Genomics 2:90-95

Moser H, Emery AE. (1974) The manifesting carrier in Duchenne muscular dystrophy. Clin Genet. 5:271-84.

Muntoni F, Melis MA, Ganau A, Dubowitz V. (1995) Transcription of the dystrophin gene in normal tissues and in skeletal muscle of a family with X-linked dilated cardiomyopathy. Am J Hum Genet. 56:151-7.

Muntoni F. (2001) Is a muscle biopsy in Duchenne dystrophy really necessary? Neurology. 57:574-5.

Muntoni F, Torelli S, Ferlini A. (2003) Dystrophin and mutations: one gene, several proteins, multiple phenotypes. Lancet Neurol. 2:731-40.

Nevo Y, Muntoni F, Sewry C, Legum C, Kutai M, Harel S, Dubowitz V. (2003) Large in-frame deletions of the rod-shaped domain of the dystrophin gene resulting in severe phenotype. Isr Med Assoc J. 5:94-7.

Newey SE, Benson MA, Ponting CP, Davies KE, Blake DJ. (2000) Alternative splicing of dystrobrevin regulates the stoichiometry of syntrophin binding to the dystrophin protein complex. Curr Biol. 10:1295-8.

Nicholson LV, Bushby KM, Johnson MA, den Dunnen JT, Ginjaar IB, van Ommen GJ. (1992) Predicted and observed sizes of dystrophin in some patients with gene deletions that disrupt the open reading frame. J Med Genet. 29:892-6.

Norwood FL, Sutherland-Smith AJ, Keep NH, Kendrick-Jones J. (2000) The structure of the N-terminal actin-binding domain of human dystrophin and how mutations in this domain may cause Duchenne or Becker muscular dystrophy. Structure Fold Des. 8:481-91.

Perkins KJ, Davies KE. (2002) The role of utrophin in the potential therapy of Duchenne muscular dystrophy. Neuromuscul Disord. Suppl 1:S78-89.

Pfeiffer H, Brinkmann B, Huhne J, Rolf B, Morris AA, Steighner R, Holland MM, Forster P. (1999) Expanding the forensic German mitochondrial DNA control region database: genetic diversity as a function of sample size and microgeography. Int J Legal Med. 112:291-8.

Pozzoli U, Elgar G, Cagliani R, Riva L, Comi GP, Bresolin N, Bardoni A, Sironi M. (2003) Comparative analysis of vertebrate dystrophin loci indicate intron gigantism as a common feature. Genome Res. 13:764-72.

Prior TW, Papp AC, Snyder PJ, Sedra MS, Western LM, Bartolo C, Moxley RT, Mendell JR. (1994) Heteroduplex analysis of the dystrophin gene: application to point mutation and carrier detection. Am J Med Genet. 50:68-73.

Prior TW, Bartolo C, Pearl DK, Papp AC, Snyder PJ, Sedra MS, Burghes AH, Mendell JR. (1995) Spectrum of small mutations in the dystrophin coding region. Am J Hum Genet. 57:22-33.

Rando TA. (2001) The dystrophin-glycoprotein complex, cellular signaling, and the regulation of cell survival in the muscular dystrophies. Muscle Nerve. 24:1575-94.

Richards M, Macaulay V. (2001) The mitochondrial gene tree comes of age. Am J Hum Genet. 68:1315-20

Richards M, Corte-Real H, Forster P, Macaulay V, Wilkinson-Herbots H, Demaine A, Papiha S, Hedges R, Bandelt HJ, Sykes B. (1996) Paleolithic and neolithic lineages in the European mitochondrial gene pool. Am J Hum Genet. 59:185-203.

Roberts RG, Gardner RJ, Bobrow M. (1994) Searching for the 1 in 2,400,000: a review of dystrophin gene point mutations. Hum Mutat. 4:1-11.

Roest PA, Roberts RG, van der Tuijn AC, Heikoop JC, van Ommen GJ, den Dunnen JT. (1993) Protein truncation test (PTT) to rapidly screen the DMD gene for translation terminating mutations. Neuromuscul Disord. 3:391-4.

Rozen S, Skaletsky H. (2000) Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol. 132:365-86.

Sadoulet-Puccio HM, Kunkel LM. (1996) Dystrophin and its isoforms. Brain Pathol. Jan;6(1):25-35.

Sadoulet-Puccio HM, Rajala M, Kunkel LM. (1997) Dystrobrevin and dystrophin: an interaction through coiled-coil motifs. Proc Natl Acad Sci U S A. 94:12413-8.

Sander M, Chavoshan B, Harris SA, Iannaccone ST, Stull JT, Thomas GD, Victor RG. (2000) Functional muscle ischemia in neuronal nitric oxide synthase-deficient skeletal muscle of children with Duchenne muscular dystrophy. Proc Natl Acad Sci U S A. 97:13818-23.

Scholl T, Pyne MT, Russo D, Ward BE. (1999) BRCA1 IVS16+6T-->C is a deleterious mutation that creates an aberrant transcript by activating a cryptic splice donor site. Am J Med Genet. 85:113-6.

Sewry CA. (2000) Immunocytochemical analysis of human muscular dystrophy. Microsc Res Tech. 48:142-54.

Shapiro MB, Senapathy P. (1987) RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. Nucleic Acids Res. 15:7155-74.

Sigurgardottir S, Helgason A, Gulcher JR, Stefansson K, Donnelly P. (2000) The mutation rate in the human mtDNA control region. Am J Hum Genet. 66:1599-609.

Smith TA, Yau SC, Bobrow M, Abbs SJ.(1999) Identification and quantification of somatic mosaicism for a point mutation in a Duchenne muscular dystrophy family. J Med Genet. 36:313-5.

Sutherland-Smith AJ, Moores CA, Norwood FL, Hatch V, Craig R, Kendrick-Jones J, Lehman W. (2003) An atomic model for actin binding by the CH domains and spectrin-repeat modules of utrophin and dystrophin. J Mol Biol. 329:15-33.

ten Kate LP. (1984) The significance of new mutations for the genetic epidemiology of Duchenne muscular dystrophy. In: ten Kate LP, Pearson PL, Stadhouders Am (eds) Research into the origin and treatment of muscular dystrophy. Excerpta Medica, Amsterdam pp 3-6.

Tennyson CN, Klamut HJ, Worton RG. (1995) The human dystrophin gene requires 16 hours to be transcribed and is cotranscriptionally spliced. Nat Genet. 9:184-90.

Todorova A, Danieli GA. (1997) Large majority of single-nucleotide mutations along the dystrophin gene can be explained by more than one mechanism of mutagenesis. Hum Mutat. 9:537-47.

Toffolatti L, Cardazzo B, Nobile C, Danieli GA, Gualandi F, Muntoni F, Abbs S, Zanetti P, Angelini C, Ferlini A, Fanin M, Patarnello T. (2002) Investigating the mechanism of chromosomal deletion: characterization of 39 deletion breakpoints in introns 47 and 48 of the human dystrophin gene. Genomics. 80:523-30.

Torelli S, Ferlini A, Obici L, Sewry C, Muntoni F. (1999) Expression, regulation and localisation of dystrophin isoforms in human foetal skeletal and cardiac muscle. Neuromuscul Disord. 9:541-51.

Tuffery-Giraud S, Chambert S, Demaille J, Claustres M. (1999) Point mutations in the dystrophin gene: evidence for frequent use of cryptic splice sites as a result of splicing defects. Hum Mutat. 14:359-68.

Tuffery-Giraud S, Saquet C, Chambert S, Claustres M. (2003) Pseudoexon activation in the DMD gene as a novel mechanism for Becker muscular dystrophy. Hum Mutat. 21:608-14.

Tuffery-Giraud S, Saquet C, Chambert S, Echenne B, Marie Cuisset J, Rivier F, Cossee M, Philippe C, Monnier N, Bieth E, Recan D, Antoinette Voelckel M, Perelman S, Lambert JC, Malcolm S, Claustres M. (2004) The role of muscle biopsy in analysis of the dystrophin gene in Duchenne muscular dystrophy: experience of a national referral centre. Neuromuscul Disord. 14:650-8.

van Deutekom JC, van Ommen GJ. (2003) Advances in Duchenne muscular dystrophy gene therapy. Nat Rev Genet. 4:774-83.

van Essen AJ, Kneppers AL, van der Hout AH, Scheffer H, Ginjaar IB, ten Kate LP, van Ommen GJ, Buys CH, Bakker E. (1997) The clinical and molecular genetic approach to Duchenne and Becker muscular dystrophy: an updated protocol. J Med Genet. 34:805-12.

Winder SJ. (1997) The membrane-cytoskeleton interface: the role of dystrophin and utrophin. J Muscle Res Cell Motil. 18:617-29.

Winnard AV, Mendell JR, Prior TW, Florence J, Burghes AH. (1995) Frameshift deletions of exons 3-7 and revertant fibers in Duchenne muscular dystrophy: mechanisms of dystrophin production. Am J Hum Genet. 56:158-66.

Yan J, Feng J, Buzin CH, Scaringe W, Liu Q, Mendell JR, den Dunnen J, Sommer SS. (2004) Three-tiered noninvasive diagnosis in 96% of patients with Duchenne muscular dystrophy (DMD). Hum Mutat. 23:203-4.

Yau SC, Bobrow M, Mathew CG, Abbs SJ. (1996) Accurate diagnosis of carriers of deletions and duplications in Duchenne/Becker muscular dystrophy by fluorescent dosage analysis. J Med Genet. 33:550-8.

Yau SC, Hatton Z, Hihalani V, Roberts RG, Abbs S, (2003) Analysis of dystrophin mRNA shows nonsense and cryptic splice sites cause Becker muscular dystrophy. Poster presented at American Society of Human Genetics Annual Meeting, Los Angeles. November 4-8.