Am. J. Hum. Genet. 49:298–310, 1991

Direct Detection of Dystrophin Gene Rearrangements by Analysis of Dystrophin mRNA in Peripheral Blood Lymphocytes

Roland G. Roberts, Teresa F. M. Barby, Elizabeth Manners, Martin Bobrow, and David R. Bentley

Paediatric Research Unit, Division of Medical and Molecular Genetics, United Medical and Dental Schools, Guy's Hospital, London

Summary

Using 10 overlapping nested sets of primers and using peripheral blood lymphocyte (PBL) total RNA as template, we have developed a system, based on PCR, which allows the rapid production of double-stranded cDNA corresponding to the entire coding sequence of the dystrophin gene. The product can be visualized on native minigels by ethidium staining and directly sequenced after gel purification. We have used this system to analyze the structures of PBL dystrophin mRNA in 26 Duchenne, Becker, or intermediate muscular dystrophy patients who have gross rearrangements of the dystrophin gene. In each case, the effect that the genomic rearrangement has on the structure of the transcript—and, by inference, on the dystrophin protein—has been determined, and the results confirm the frameshift hypothesis. The study also identifies a series of alternatively spliced transcripts which are specific to the rearranged genotypes and which seem therefore to arise following the alteration in the context of the splice signal. The system has been used for unambiguous identification of carrier females. Furthermore, the rapid production of microgram quantities of dystrophin cDNA from a readily accessible tissue makes point-mutation screening a practical proposition.

Introduction

Duchenne muscular dystrophy (DMD) and Becker muscular dystrophy (BMD) are X-linked diseases which both originate from mutations in the dystrophin gene. This clinical division into two diseases stems from a bimodal distribution of severity and age at onset, whereby substantial numbers of severe, early-onset cases (DMD) and fewer mild, late-onset cases (BMD) are separated by a small number of intermediate cases (reviewed in Emery 1987).

In approximately 66% of cases, the gene is disrupted by deletion or duplication of an integral number of exons (Koenig et al. 1987; den Dunnen et al. 1989). A strong correlation has been observed between the shift in translational reading frame of a notional transcript from such a disrupted gene and the severity of the patient's disease, while the size and position of the disruption appear less important (the frameshift hypothesis; Monaco et al. 1988). Thus a mutation which causes a translational frameshift results in a truncated protein and a DMD phenotype, while one in which the reading frame is maintained results in a protein bearing an interstitial deletion or duplication and a BMD phenotype.

Although the frameshift hypothesis holds for the majority of cases, there are patients for whom this model appears not to be valid—both DMD patients with in-frame mutations and BMD patients with frameshifts have been observed (Gillard et al. 1989; Koenig et al. 1989). There are also cases where lack of knowledge of exon boundaries precludes prediction of the reading frame. As mRNA has undergone transcription and splicing from the gene and represents the actual template for translation, it is one step further along the route from the defective gene to the protein whose defect causes a disease. It was anticipated that for some types of change the nature of the defect at the mRNA level might correlate better with disease severity than does the notional mRNA structure ex-

Received January 14, 1991; revision received April 3, 1991.

Address for correspondence and reprints: Roland Roberts, Paediatric Research Unit, Division of Medical and Molecular Genetics, UMDS, Guy's Hospital, London SE1 9RT, England.

0002-9297/91/4902-0006$02.00

trapolated from genomic data. Other types of mutation may have effects on the translation product (e.g., altered stability, degradation rate, intracellular location, or tendency to aggregate) which are not readily predictable from the mRNA structure.

Dystrophin mRNA is found at significant levels only in muscle tissues and the brain, neither of which can readily be obtained from patients for research purposes. Extremely low levels of transcript (of the order of 1 copy /1,000 cells) have been reported in a number of other tissues (Chelly et al. 1988). We have developed a system based on reverse transcription and nested PCR (Saiki et al. 1988) which allows rapid generation of microgram quantities of double-stranded cDNA from peripheral blood lymphocyte (PBL) total RNA. The entire dystrophin coding sequence is amplified in 10 overlapping nested reactions averaging almost 1.2 kb in length, yielding products which can be visualized directly by gel electrophoresis and ethidium staining (fig. 1*a* and *b*). This approach has been used to analyze the structure of transcripts in 26 affected males and their relatives (patients selected were those who attended clinic and were known to have rearrangements of the dystrophin gene). Direct sequencing (Green et al. 1989) of appropriate products permits identification of deletion and duplication mutations by characterization of splice junctions in transcripts. In addition, some alternative splicing is observed which is specific to the rearranged genotype.

Methods

RNA Preparation

Lymphocytes were separated from up to 10 ml of whole blood (up to 4 d old) by centrifugation through a layer of Histopaque 1077 (Sigma) followed by rinsing in cold PBS. The resulting cellular pellet was subjected to the acid guanidinium thiocyanate-phenol-chloroform RNA extraction method (Chomczynski and Sacchi 1987) followed by precipitation in 5 M LiCl (– 20°C for 16 h) and a further ethanol precipitation. The RNA was stored as an ethanol precipitate at – 20°C until used. The LiCl and the TE (10 mM Tris-Cl pH 7.4, 0.1 mM EDTA) which was used for preparing reagents were pretreated with diethylpyrocarbonate.

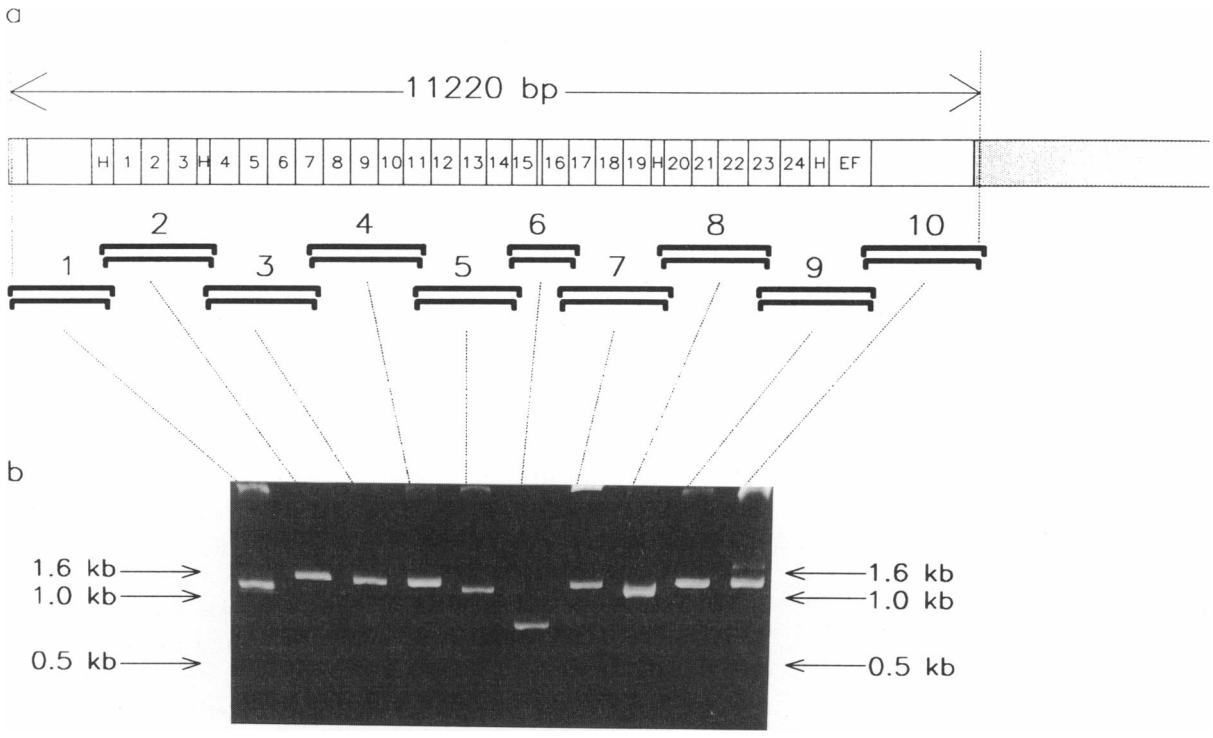


Figure 1 *a*, Schematic diagram of dystrophin mRNA, showing positions of 10 principal nested PCR reactions. Shaded boxes denote untranslated regions; open box denotes coding region, showing revised protein domain structure (after Koenig and Kunkel 1990) (H = hinge segments; 1–24 = repeat motifs; EF = EF-hand-like region); horizontal bars denote inner and outer PCR products. *b*, Amplification products, from PBL RNA by using 10 nested primer sets, run on ethidium-stained polyacrylamide minigel.

Reverse Transcription

Samples (200–500 ng) of total lymphocyte RNA in TE were incubated in a volume of 7.5 μl with 50 ng of primer DMDXb or DMDXe (X = 1–10; see below for details of primer nomenclature) at 65°C for 10 min, topped with 50 μl light paraffin. The sample was then snap-chilled on ice and made to total 20 μl with a premix containing 4 μl 5 × reverse transcriptase buffer (BRL), 2 μl 100 mM DTT, 200 units MMLV reverse transcriptase (BRL), 5 μl 5 mM dNTPs, and 25 units RNase inhibitor (Boehringer Mannheim). The reaction was incubated at 42°C for 1 h.

Nested PCR (Saiki et al. 1988; Holding and Monk 1989)

A 30-μl mixture containing 5 μl 10 × PCR buffer (670 mM Tris-HCl pH 8.8, 166 mM $(NH_4)_2SO_4$, 67 mM $MgCl_2$, 1.7 mg BSA (Sigma)/ml, and 100 mM 2-mercaptoethanol), 500 ng primer DMDXa, 450 ng primer DMDXb or DMDXe, and 5 units Amplitaq *Taq* polymerase (Perkin Elmer–Cetus) was added to the reverse transcription reaction. Thirty cycles of PCR (at 93°C for 1 min, at 60°C for 1 min, and at 72°C for 7 min, on a Techne PHC-2 Programmable Heating Block) were performed on the sample, followed by an incubation at 72°C for 5 min. Samples (1 μl) of the product were added to a 50-μl mixture containing 5 μl 10 × PCR buffer, 5 μl 5 mM dNTPs, 500 ng primer DMDXc, 500 ng primer DMDXd or DMDXf, and 5 units Amplitaq *Taq* polymerase. PCR was repeated as above. Portions (1–8 μl) of the final product were electrophoresed in a 4% polyacrylamide minigel containing ethidium bromide.

Primer Sequences

Sequences of the main set of 40 oligonucleotides are presented in table 1 (Nomenclature: DMDXa and DMDXb = outer set, 5′ and 3′, respectively; DMDXc and DMDXd = inner set, 5′ and 3′, respectively). Additional 3′ primers used are as follow (nucleotide numbering is as in Koenig et al. 1988):

DMD1e–CTTTAGGTGGCCTTGGCAAC (967–986);
DMD1f–CAGGATCGAGTAGTTTCTC (836–854);
DMD7e–GCTTTTCTTTTAGTTGCTGC (6937–6956);
DMD7f–GTTGCTGCTCTTTTCCAGGTTC (6923–6944);
DMD8e–GTAACAGGACTGCATCATCGG (8514–8534);
DMD8f–GAATGCTTCTCCAAGAGGC (8210–8228).

Categorization of Patients

As elsewhere (Hodgson et al. 1989; Dubowitz 1990), we have accepted the characterization of patients as DMD if they are confined to a wheelchair at or before the age of 12 years and as BMD if they are still ambulant at age 16 years. Intermediate cases fall between these two categories. The categorization of young boys with no family history is inevitably less precise but was based on clinical and pathological assessments of severity of disease. Phenotypes of some patients have been described in more detail by Hodgson et al. (1989) (these phenotypes are underlined in table 3).

Direct Sequencing

PCR products were purified from 4% polyacrylamide minigels by electroeluting individual bands sideways onto pieces of NA45 DEAE membrane (no. 417082; Schleicher and Schuell). The DNA was eluted by incubating the membrane in 1 M NaCl at 70°C for 20 min. After two ethanol precipitations the products were directly sequenced by the method of Green et al. (1989) with the inclusion of dimethyl sulfoxide (Winship 1989). Primers were those used for PCR or internal oligonucleotides (sequences available on request).

Results

Detection of Dystrophin Transcripts in PBL RNA

PBLs isolated from 10 ml of peripheral blood generally yielded approximately 15–40 μg of total RNA, sufficient for approximately 80 reverse transcriptions. Satisfactory results were obtained from samples up to 4 d old on receipt. In many cases PBLs had been stored at −20°C as frozen pellets for up to 2 mo. Only one of the 56 samples tested failed to be amplified. The rRNA bands were not discernible in the failed sample on an agarose gel, and the RNA was assumed to be degraded.

Ten overlapping reverse-transcription/nested PCRs (reactions 1–10; fig. 1*a* and table 1) were used to obtain dsDNA spanning the entire coding region of the gene. The products of a typical analysis of PBL RNA from a normal individual are shown in figure 1*b*. All 26 patients used in the present study had transcripts capable of sustaining amplification. All 10 reactions reproducibly amplified mRNA from patients, rela-

Table 1

Primers Used for 10 Principal Nested Reactions

Name, Sequence (5′→3′)	Position	Product	
DMD1a, CTTTCCCCCTACAGGACTCAG.....	13–33	1207	Reaction 1 (E1–E10)
DMD1b, CTCTCCATCAATGAACTGCC.....	1199–1218		
DMD1c, CTGGGAGGCAATTACCTTCGG.....	36–56	1159	
DMD1d, GACTTGTCTTCAGGAGCTTG.....	1175–1194		
DMD2a, CGATTCAAGAGCTATGCCTAC	1091–1111	1329	Reaction 2 (E9–E18)
DMD2b, GCGAGTAATCCAGCTGTGAAG.....	2399–2419		
DMD2c, CCTCTGACCCTACACGGAGC.....	1134–1153	1263	
DMD2d, CAGTTATATCAACATCCAACC.....	2376–2396		
DMD3a, CATGCTCAAGAGGAACTTCC.....	2300–2319	1320	Reaction 3 (E17–E25)
DMD3b, CTGAGTGTTAAGTTCTTTGAG.....	3599–3619		
DMD3c, GCAGATTACTGTGGATTCTG	2344–2363	1251	
DMD3d, GTCTCAAGTCTCGAAGCAAAC	3574–3594		
DMD4a, CAATTCAGCCCAGTCTAAAC.....	3507–3526	1344	Reaction 4 (E25–E33)
DMD4b, CAAAGCTGTTACTCTTTCATC.....	4832–4852		
DMD4c, GTGTCAATGAAGGTGGGCAG.....	3528–3547	1281	
DMD4d, CTGCTTTTTCTGTACAATCTG	4790–4810		
DMD5a, GTCTGAGTGAAGTGAAGTCTG.....	4740–4760	1198	Reaction 5 (E33–E40)
DMD5b, CCTTTCATCTCTGGGCTCAG	5919–5938		
DMD5c, GAAATGGTGATAAAGACTGG	4766–4785	1131	
DMD5d, CAATGTCATCCAAGCATTTC.....	5878–5897		
DMD6a, CTCTAGAAATTTCTCATCAG.....	5823–5842	801	Reaction 6 (E40–E44)
DMD6b, GCATGTTCCCAATTCTCAGG	6605–6624		
DMD6c, GGTATCAGTACAAGAGGCAG	5844–5863	739	
DMD6d, CTGTTCAGCTTCTGTTAGCC	6564–6583		
DMD7a, GCAACGCCTGTGGAAAGGGTG.....	6404–6424	1311	Reaction 7 (E43–E51)
DMD7b, GTCACCCACCATCACCCTCTG.....	7694–7714		
DMD7c, CAGGAAGCTCTCTCCCAGC.....	6431–6449	1227	
DMD7d, GGTAAGTTCTGTCCAAGCCCGG...	7636–7657		
DMD8a, CTAGAAATGCCATCTTCCTTG.....	7583–7603	1307	Reaction 8 (E51–E58)
DMD8b, CTCAGGAGGCAGCTCTCTGG.....	8871–8890		
DMD8c, CTGCTCTGGCAGATTTCAAC.....	7617–7636	1252	
DMD8d, CTCCTGGTAGAGTTTCTCTAG.....	8849–8869		
DMD9a, GGGCCTTCAAGAGGGAATTG.....	8754–8773	1316	Reaction 9 (E57–Ey)[a]
DMD9b, CCAGTCTCATCCAGTCTAGG	10051–10070		
DMD9c, CTAAAGAACCTGTAATCATG.....	8778–8797	1268	
DMD9d, GGGCCGCTTCGATCTCTGGC.....	10027–10046		
DMD10a, GGTGAAGTTGCATCCTTTGG	9953–9972	1298	Reaction 10 (Ex–3′ UTR)[a]
DMD10b, CATGACTGATACTAAGGACTC....	11308–11328		
DMD10c, GGCAGTAACATTGAGCCAAG.....	9974–9993	1245	
DMD10d, CCAAATCATCTGCCATGTGG.....	11277–11296		

[a] Ex represents the (hitherto unnumbered) last exon of the EF-hand region; Ey represents the immediately subsequent exon.

tives, and normal subjects. The yield range was approximately 750 ng–6 μg/50-μl reaction.

Restriction digestion of products of all reactions from normal samples yielded patterns consistent with their structure being colinear with the published cDNA (Koenig et al. 1988), with two exceptions: (1) In all cases, approximately 50% of products of reaction 1 were missing exon 9 (encoding 1040–1168 of

the mRNA). This is shown in figure 2*a*, in which the full-length product of reaction 1 (1,159 bp; left-hand lane) is accompanied by a shorter product of 1,030 bp. Alternative splicing of exon 9 (which does not affect the reading frame but removes a large portion of the first hinge segment; Koenig and Kunkel 1990) has not to our knowledge been reported previously in muscle from normal individuals (although Winnard et al. [1990] report alternative splicing of exon 9 in several BMD patients). (2) Loss of a 39-bp exon (encoding 10432–10470 of the mRNA) from reaction 10 was observed in approximately 50% of products from all individuals tested. (The doublet is not resolved in fig. 1*b*, although a higher band, presumed to represent a heteroduplex between products with and without this exon, is observed. The difference was resolved by restriction digestion and was sequenced following gel purification.) The latter phenomenon has been reported in muscle tissue, and it has been proposed that the alternatively spliced transcript encodes an isoform of dystrophin (Feener et al., 1989). The only deviation from the published sequence (Koenig et al. 1988) encountered was a $T_{7936} \rightarrow C_{7936}$ transition in patient 24 (a silent change in an Asn codon).

In addition to the 10 principal reactions, combinations of nested primers from different reactions were used in cases where deletions were large and included priming sites. For example, primers 2a and 4b, followed by 2c and 4d, were used for amplification of RNA from patient 05. Further 3′ primers (DMDXe and DMDXf; see Methods section above) were also used in combination with 5′ primers to generate reactions 6.5 and 7.5 (see fig. 2*d* and *f*) and for the "diverging nest" used for patients 04 and 02 (see below). The largest product shown here is that amplified from a normal sample by using reaction 7.5 (2,131-bp outer product and 1,798-bp inner product; see fig. 2*f*). Products of the outer reaction may be stored at −20°C for at least 6 mo, providing an archive source of template for the inner reaction. Reverse transcription and the first round of PCR can be multiplexed in two reactions each containing five primer sets ("odds" and "evens"). Although this saves time and RNA, a significant increase in background smear was observed, which would reduce the quality of direct sequencing.

Analysis of PBL RNA was carried out in 26 patients and the carrier sister of an additional patient. Figure 2 shows the results of PCR; each panel includes a normal sized product, the size of the full-length product for each reaction being shown in the left-hand lane. The results are summarized in table 3. Results of direct sequencing are shown in table 2.

Direct Confirmation of Frameshift Hypothesis

In 10 deletion cases where the reading frame is predictable from genomic data and supports the frameshift hypothesis, the PBL mRNAs were found to have the expected structures (cases 06, 12, 22–24, and 26 have frameshifted transcripts and DMD phenotype, while cases 18–21 have in-frame transcripts and BMD phenotype); for example, amplification of PBL RNA from patient 26 with reaction 7 results in the appearance of a PCR product of 833 bp in size, indicating loss of 394 bp compared with the normal, full-length product (fig. 2*e*). This size is consistent with loss of exons 48–50 as indicated in the diagram (fig. 2*e*, bottom). Direct sequence analysis of the reaction 7 PCR products from a normal sample and patient 26 demonstrated that the sequence of the RNA in the normal sample is colinear and spans the boundary between exons 47 and 48, whereas in the patient exon 47 was spliced directly to exon 51, resulting in a frameshift (see table 2) and hence in an in-frame termination signal (TGA) nine codons into exon 51. This result is consistent with the observed phenotype of DMD (see last column of table 2). As a second example, case 21 is a BMD patient. Analysis of his PBL mRNA with reaction 7 (fig. 2*e*) and by direct sequence analysis of the PCR products demonstrates deletion of exons 45–48 and correct splicing of exon 44 to exon 49. This results in retention of the correct reading frame, which is consistent with his phenotype.

Improvement by RNA Studies

In four cases (05, 08, 09, and 27; see fig. 2 *b, c, c,* and *g,* respectively) one or both ends of the deletions lay in regions where positions of exon boundaries had not previously been established. In addition, the precise extent of deletions at the genomic level in patients 05 and 09 was not known. Hence, on the basis of studies at the DNA level, the reading frame in each case was unknown ("U" in fourth column of table 3). From the sequence of the amplification products from PBL mRNA, however, it was possible to establish the positions of exon boundaries and hence the effect that the deletions had on the translational reading frame. In case 05, direct sequence analysis of the PCR product of reaction 2/4 showed that the 3′ end of exon 9 (G_{1168}) was spliced directly to A_{3812}. The reading frame is conserved, and this correlates with the observed BMD

phenotype (see table 3). This result identifies an exon boundary between G_{3811} and A_{3812} (this lies in repeat 8 [as revised in Koenig and Kunkel 1990] at the conserved exon boundary position). In case 08, exon 30 is deleted, and the conservation of reading frame is consistent with the BMD phenotype; this result also identifies the exon boundary G_{4279}/G_{4280} at the beginning of exon 30. Analysis of case 09 shows loss of reading frame, consistent with the observed severe DMD phenotype, and this result provides independent confirmation of the exon boundary G_{4279}/G_{4280}. The deletion in case 27 also results in loss of reading frame and in occurrence of the DMD phenotype; the exon boundary identified is A_{9432}/C_{9433} (in hinge segment 4; Koenig and Kunkel 1990). Note that 05 lacks 17 exons (the translation product is expected to be missing hinge segment 2 and the first 8.5 repeats) but has BMD, while patient 27, whose dystrophin should have the entire "rod-like" region intact, lacking only the C-terminal domain, has a severe disease.

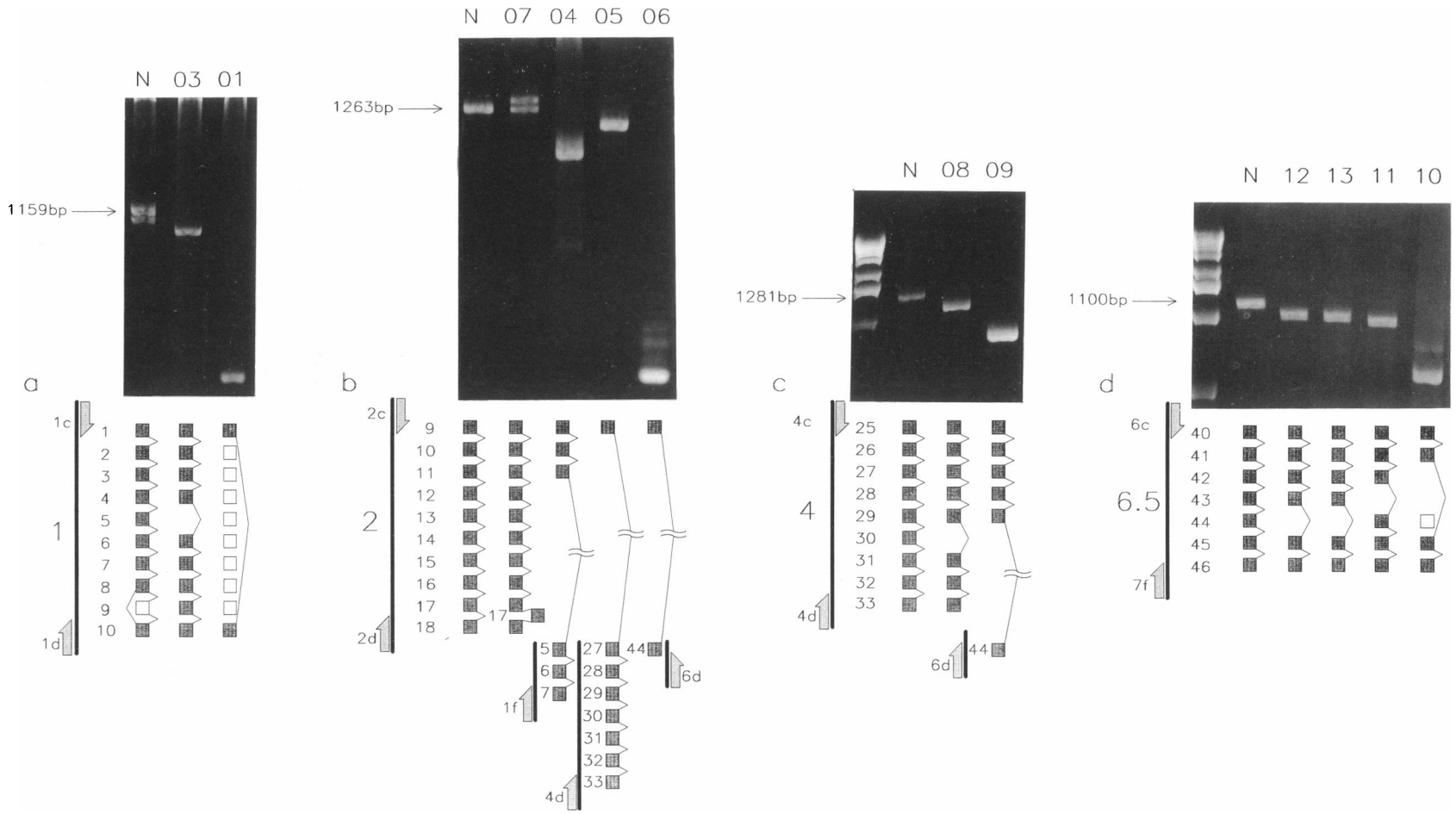
Three patients were analyzed who had duplications of small portions of the gene. The duplication in patient 04 was detected initially by the presence of both a junction fragment and increased dosage of several exons on Southern blots (S. Abbs, personal communication). The extent of the duplication at the genomic level was not known with confidence. At the mRNA level the 04 duplication was analyzed by using a "diverging nest" of DMD2a and DMD1e (outer) and DMD2c and DMD1f (inner), which will only amplify when there are two copies of exons 7–9. A PCR product of 788 bp was obtained (see fig. 2*b*). Sequencing of this product showed that in 04 a copy of exon 11 is followed directly by a second copy of exon 5, suggesting a duplication of exons 5–11. Patient 02 was analyzed using the same reaction and is described in detail below. The duplication in patient 07 was initially detected by dosage alone (S. Abbs, personal communication). As PBLs from the patient were not available, a sample from a carrier sister was analyzed. The 07 duplication was detected by the appearance of a larger product with reaction 2, in addition to the normal band (fig. 2*b*). Sequencing of the abnormal product from the carrier sister showed that a copy of exon 17 was spliced to a further copy of the same exon. In each case the mRNA study shows that duplication is both tandem and direct and that it causes a frameshift consistent with the DMD phenotypes (table 3).

Southern analysis of patient 01 shows the presence of all exons plus an additional "junction fragment" recognized by one of exons 1–11 (S. Abbs, personal communication). Such results are usually interpreted as indicative of a duplication. mRNA analysis of case 01 by using reaction 1 reproducibly showed correct splicing of exon 1 to exon 10 in all products, resulting in a frameshift which would terminate translation after only 15 amino acids (a result consistent with the DMD phenotype). It is suggested that a more complex genomic rearrangement (e.g., inversion or translocation of exons 2–9 to another portion of the genome) has occurred in this family. This example particularly illustrates the value of studying the mutation directly at the level of the transcribed RNA.

Alternative Splicing Events in Rearranged Genotypes

In one patient whose genotype/phenotype relationship was at variance with the frameshift hypothesis, unexpected alternative splicing was found, which, if occurring in muscle tissue, would explain the anomalous phenotype. At the DNA level, patient 10 appeared to violate the frameshift hypothesis. The deletion of exons 42 and 43 should cause a frameshift but the patient has a phenotype which is apparently on the mild side of the intermediate range (the patient was still able to run at age 8 years although subsequently was rendered paraplegic by unrelated causes). Amplification using reaction 6 gave the expected products, but, as DMD6b and DMD6d lie in exon 44 (immediately 3′ to the deletion), alternatively spliced transcripts would not be detected. When PBL mRNA was amplified using reaction 6.5, the minor product was 733 bp, a size which is consistent with the loss of exons 42 and 43, while the major product was 585 bp (fig. 2*c*). The sequence of the major product showed that exon 44 (known to be present in the genomic DNA) was also missing from these transcripts. The alternative splicing which gives rise to this transcript restores the translational reading frame (table 2). On the assumption that this alternative splicing also occurs in muscle, this result would explain the mild/intermediate phenotype that is observed in this patient and thus would resolve the violation of the frameshift hypothesis.

Cases 14–17 have phenotypes in the DMD range (case 14 was only just mobile at age 9 years, and cases 15–17 were off their feet at ages 10 years, 12 years, and 11 years, respectively). This is expected, as analysis of genomic DNA indicates that they have deletions of exon 45, which would result in frameshifts. Approximately half of the amplification products from


N 03 01
1159bp
a
1c
1
1d
N 07 04 05 06
1263bp
b
2c
2
2d
17
1f
4d
5
6
7
27
28
29
30
31
32
33
44
6d
N 08 09
1281bp
c
4c
4
4d
6d
N 12 13 11 10
1100bp
d
6c
6.5
7f
40
41
42
43
44
45
46

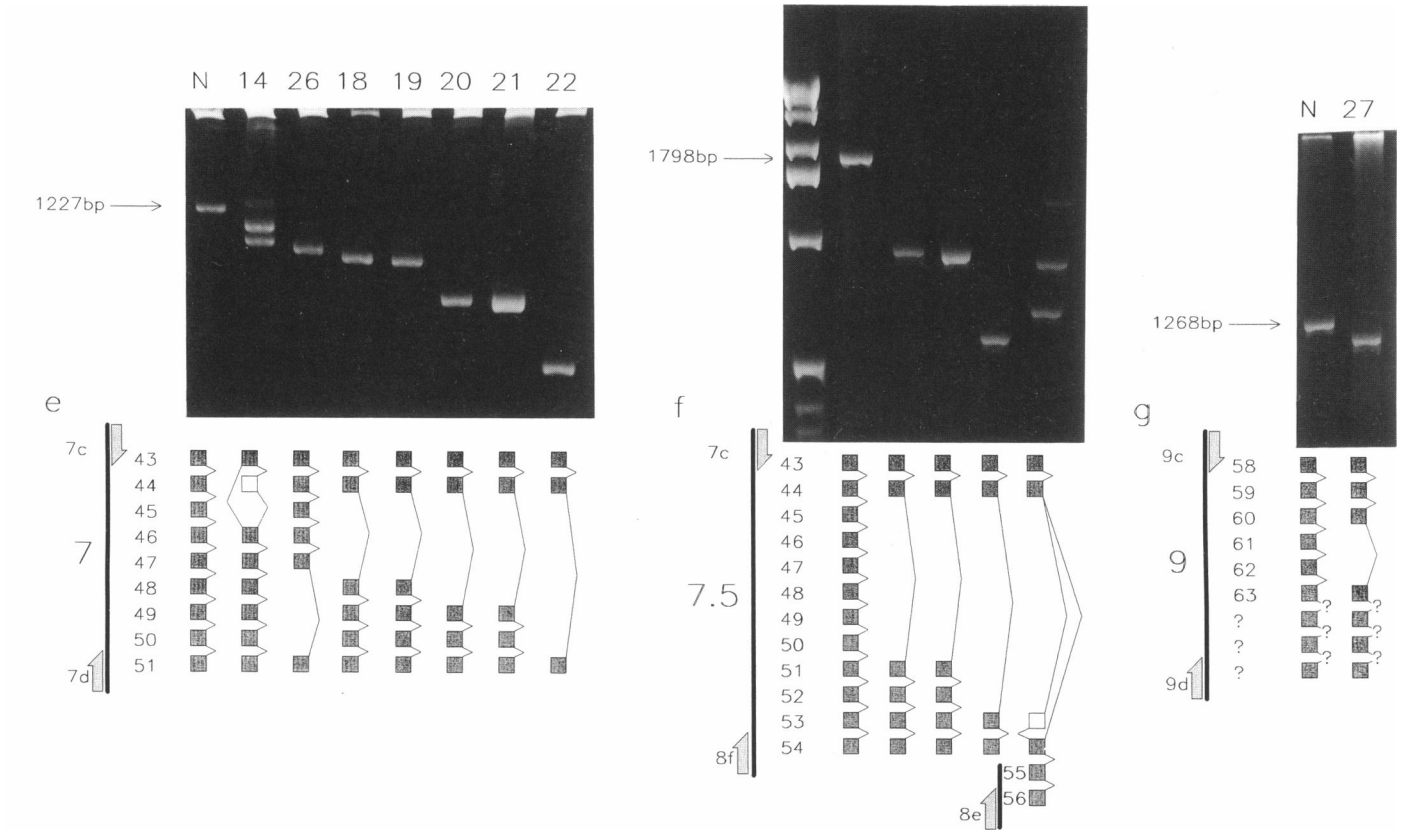


Figure 2 Products of amplification from PBL RNA (from normal subjects, patients, and carrier female). The structure of each product (derived from restriction and sequence data) is shown beneath its gel track. Shaded arrows denote inner primers used for PCR; filled squares denote exons present in all products: connected open squares denote exons present in a proportion of products; unconnected open squares denote exons known to be present in the genome but absent from products; N = normal sized products. The left-hand tracks in parts *c*, *d*, and *f* contain 1 kb DNA ladder (Bethesda Research Laboratories). Note that case 07 in *b* is represented by a carrier sister of two deceased affected boys, hence the additional presence of the normal sized band.

Table 2

Sequences across cDNA Breakpoints

Patient(s)	5′ Exon	3′ Exon	5′ Sequence / 3′ Sequence[a]
01	1	10	..GTA GAG GAC TGT T / CAT TTG GAA GCT..
03	4	6	..CAG AAC AAT AAT / GTC AAA AAT GTA..
02[b]	9	2	..TTT CCT TCA CAG / AT GGA AGA GAA..
	9	3	..TTT CCT TCA CAG / TTT GGG AAG CAG..
05	9	27	..TTT CCT TCA CAG / AGA GCT AAA GAA..
06	9	44	..TTT CCT TCA CAG / G CGA TTT GAC..
04	11	5	..AAA CAA AGC AA / GTT GAT TTA GTG..
07	17	17	..ATT AGG AAA AG / ATT TCA CAG GCT..
08	29	31	..CTA CAT GAA GAG / AAA ATC CAA TCT..
09	29	44	..CTA CAT GAA GAG / G CGA TTT GAC..
10	41	45	..TTT GCA CAA ATT / GAA CTC CAG GAT..
11	42	44	..GAG TCT CTG AAG / G CGA TTT GAC..
12 and 13	43	45	..GAC CGA CAA GG / GAA CTC CAG GAT..
14–17[b]	43	46	..GAC CGA CAA GG / G CTA GAA GAA..
	44	46	..TGG TAT CTT AAG / G CTA GAA GAA..
18 and 19	44	48	..TGG TAT CTT AAG / GTT TCC AGA GCT..
20 and 21	44	49	..TGG TAT CTT AAG / GAA ACT GAA ATA..
22 and 23	44	51	..TGG TAT CTT AAG / CT CCT ACT CAG..
24	44	53	..TGG TAT CTT AAG / TT GAA AGA ATT..
25[b]	44	53	..TGG TAT CTT AAG / TT GAA AGA ATT..
	44	54	..TGG TAT CTT AAG / CAG TTG GCC AAA..
26	47	51	..CAG TGG ATA AAG / CT CCT ACT CAG..
27	60	63	..AAG CTT CTG CAG / C CAC GAG ACT..

[a] Triplets are presented in wild-type phase (to show effect on reading frame), *not* as translated in patients.

[b] These patients each yield two distinct products; both are presented.

PBL mRNA, however, also lack exon 44, as shown by the appearance of two bands (1,051 bp and 903 bp) when reaction 7 is used (fig. 2*e*). Loss of exon 44 from transcripts already deleted for exon 45 should restore the translational reading frame, but any expected mitigating effects on the phenotype must be very slight. We have observed in a number of cases that PCR on two competing targets strongly favors the shorter one (Roberts et al. 1990). Hence in these cases an additional band of intensity equal to that of the expected product may represent a small minority of target. This might also explain why case 10 (see above), who appears to have a large excess of in-frame product, has an intermediate phenotype rather than BMD. Alternatively, significant levels of the in-frame product may not be found in muscle in patients 14–17. Patients with deletions of exon 45 alone have widely ranging phenotypes, from severe DMD to a BMD patient still mobile at age 18 years (Gillard et al., 1989; Hodgson et al., 1989). The frequency of loss of exon 44 from the transcripts of patients deleted for exon 45 may vary from case to case, resulting in a continuum of phenotype severity.

Two other patients exhibit alternative splicing which is at odds with the frameshift hypothesis. Patient 25, like patient 24, has a deletion which comprises exons 45–52. This should disrupt the reading frame, and the patient has DMD (and was wheelchair bound by age 7 years). Alternative 3′ primers were used which are farther away from the deletion (DMD8b as outer and DMD8e as inner). Two PCR products were obtained (fig. 2*f*). One is 882 bp, which corresponds to the transcript with the exon 45–52 deletion; the other is 670 bp, which corresponds to an alternatively spliced transcript which lacks exon 53 and has an intact reading frame. Patient 02 was proposed to have a duplication at the 5′ end of the gene, on the basis of a junction fragment detected by cDNA probe 9–7 on Southern blots. Amplification of PB RNA by using the same "diverging nest" of primers as was used for patient 04 yielded two products. The larger of these is 650 bp, with exon 9 followed by a second copy of exon 2, while the smaller is 588 bp, with exon 9 followed by a second copy of exon 3. These results are consistent with a duplication of exons 2–9, with alternative splicing of the second copy of

Table 3

Genotype/Phenotype Correlation at RNA Level

Patient (clinic number[a])	DNA mut[b]	RNA mut[b,c]	DNA Frame[d]	RNA Frame[d]	Phenotype D/I/B[e]	Age[f] (years)
01 (44698)	dup?	1d2-9	U	O	D	9
02 (57939)	dup?	1dup2-9,3-9	U	O/I	D	9
03 (50506)	d5	1d5	I	I	D	5[g]
04 (49449)	dup?	2/1dup5-11	U	O	D	7
05 (59755)	d10-?	2/4d10-26?	U	I	B	3[g]
06 (43052)	d10-43	2/6d10-42	O	O	D	20
07 (18317)	dup17	2dup17	O	O	D	Deceased
08 (48940)	d30?[h]	4d30?	U	I	B	14
09 (43484)	d?-43	4/6d30-43	U	O	D	10
10 (39487)	d42-43	6.5d42-44	O	O/I	I	12
11 (45049)	d43	6.5d43	O	O	I	14
12 (52382)	d44	6.5d44	O	O	D	8
13 (41708)	d44	6.5d44	O	O	I	14
14 (40335)	d45	7d45,44-45	O	O/I	D	10
15 (33378)	d45	7d45,44-45	O	O/I	D	14
16 (33216)	d45	7d45,44-45	O	O/I	D	16
17 (33217)	d45	7d45,44-45	O	O/I	D	14
18 (38604)	d45-47	7d45-47	I	I	B	23
19 (45480)	d45-47	7d45-47	I	I	B	9
20 (58815)	d45-48	7d45-48	I	I	B	30
21 (46405)	d45-48	7d45-48	I	I	B	15
22 (46038)	d45-50	7.5d45-50	O	O	D	13
23 (45976)	d45-50	7.5d45-50	O	O	D	12
25 (40106)	d45-52	7.5d45-52	O	O	D	11
25 (45929)	d45-52	7/8d45-52,53	O	O/I	D	15
26 (58264)	d48-50	7d48-50	O	O	D	3
27 (56741)	d61,62?[i]	9d61,62?	U	O	D	3

[a] Underlines denote those patients described by Hodgson et al. (1989).
[b] d = deletion; dup = duplication (followed by extent as exon number); ? = limits unknown or exon number provisional.
[c] Notation: reaction number, d/dup, exon number(s).
[d] I = in frame; O = out of frame; U = unknown; O/I = combination;
[e] D = DMD; B = BMD; I = intermediate.
[f] At next birthday.
[g] Phenotype of elder brother(s) known.
[h] Deletion of 4.7-kb *Hin*dIII fragment detected by probe 30-1.
[i] Deletion of 2.8-kb and 12-kb *Hin*dIII fragments detected by probe 63-1(10).

exon 2 (although detection of a normal sized reaction 1 product in patient 02 suggests that exon 1 and the promoter region may also be duplicated). The larger product has a disrupted reading frame, which is restored in the smaller product by the loss of exon 2. Patient 02 has DMD. In both case 25 and case 02 the phenotype is consistent with the larger product but not with the presence of the alternatively spliced product, which would be expected to have a dominant rescue effect. As discussed above, this apparent inconsistency could be explained by the tendency of PCR to favor shorter target sequences. If the promoter region is duplicated in patient 02, then the severe phenotype may result from the 3′ copy of the promoter being separated from its upstream control elements. Investigation of the long-range structure of this individual's dystrophin gene may clarify this matter.

Unexplained Phenotypes

In three further cases the frameshift hypothesis cannot explain the phenotype even when the structure of PBL mRNA is known. The deletion of exon 5 from the mRNA of patient 03 leaves the reading frame intact, but his phenotype is severe. Particular functional

significance of this exon is ruled out by the existence of BMD patients whose deletions include exon 5. Either a transcript-level phenomenon restricted to muscle tissue or the presence of an independent mutation could account for the phenotype.

Patient 11 can still walk and climb stairs at age 13 years 2 mo and is classified as intermediate. His deletion of exon 43 as observed in PBL mRNA should cause a frameshift. It is possible that this apparent violation of the frameshift hypothesis may be resolved by the loss of exon 44 through alternative splicing in muscle only. The removal of exon 44 by alternative splicing in PBLs has been found, in the present study, in association with genomic deletions of exons 42 and 43 (case 10; see above) or of exon 45 (cases 14–17; see above).

Patient 13 is at the severe end of the intermediate range (and is just mobile, with the aid of callipers, at age 12 years 4 mo), but his deletion, which is similar to that of the more severe patient 12, should cause a frameshift (see fig. 2*d* and table 2). At the PBL mRNA level, alternative splicing was not observed, and patient 13's relatively mild phenotype remains unexplained. Perhaps significantly, patient 12 had extremely mild dystrophy until age 7 years, when, in a few months, he deteriorated from being able to walk long distances to having to wear callipers (he was subsequently classified as severe, and shortly he became wheelchair bound).

Discussion

We have demonstrated a rapid method for the generation of microgram quantities of double-stranded dystrophin cDNA from PBL total RNA. This material has been used for the study of structural rearrangements of the dystrophin gene in deletion and duplication patients. The results demonstrate that the transcripts amplified reflect the genomic structure as previously determined from analysis of the DNA. In addition, alternative splicing was observed in a number of cases, although because of the possibility of differential expression of the dystrophin gene in PBL and muscle, some caution must be exercised in the interpretation of the results.

The analysis of rearrangements of the dystrophin gene by the study of PBL transcripts provides direct confirmation of the frameshift hypothesis. Table 3 summarizes the DNA, PBL mRNA, and phenotypic data for the patients studied. Sets of data which best fit the frameshift hypothesis are shaded. The improved fit between patient phenotype and PBL mRNA data compared with the DNA data stems largely from refinement of both the extent of rearrangements and the definition of exon boundaries and supports the validity of using the PBL transcript structure as a means of studying genomic rearrangements.

In addition, the detection of deletion and duplication mutations by analysis of PBL RNA permits direct carrier diagnosis. We have amplified PBL RNA from 20 female relatives of affected boys (Roberts et al., 1990; R. G. Roberts, unpublished data). The anomalous-sized product is found, together with the normal sized product, in females of known high carrier risk (as ascertained by linkage analysis), in a proportion of females of unknown carrier risk, but not in females of known low carrier risk; for example, the second track of figure 2*b* represents the reaction 2 amplification products from a sister of two DMD patients who was assigned a high carrier risk based on linkage analysis. In addition to the normal 1,263-bp product, there is a 1,439-bp product bearing a duplication of exon 17. Presence of the deleted or duplicated product in a female relative constitutes a definitive indication of her carrier status, by using an assay which is rapid, nonradioactive, and qualitative (Roberts et al. 1990). Note that, although a positive diagnosis is conclusive, the certainty of a negative result is restricted by the possibility of nonrandom X inactivation and gonadal mosaicism.

In seven cases, alternative splicing was observed. In each case, the alternative splicing was specific to the rearranged genotype, involved loss of a single exon immediately adjacent to the rearrangement, and was reproducibly detectable. This strongly suggests that the phenomenon arises as a result of the change in context of the splice-donor or -acceptor site.

In particular, exon 44 appears to be alternatively spliced in a number of patients whose deletion starts in either of the adjacent introns. The literature contains a considerable number of patients whose departure from the frameshift hypothesis could be explained by removal of this exon from some or all of their muscle transcripts (Hodgson et al., 1989, case 33174; Gillard et al. 1989, cases 37, 97, 142, 89, and 191; Norman et al. 1990, case 14); and four cases from Koenig et al. (1989). Of these cases, seven are intermediate or BMD patients deleted for exon 45 alone. This said, the extent to which the structure of the PBL mRNA reflects that of the muscle transcript remains to be confirmed. There remain several cases where the phenotype is not explained by the structure of the PBL

mRNA. Case 11 is a candidate for loss of exon 44 in a proportion of muscle transcripts, as this would restore the reading frame and explain the intermediate phenotype (a similar intermediate case reported by Koenig et al. [1989] also has a deletion of exon 43 alone). The severe phenotype of case 03, on the other hand, cannot be explained by muscle-specific alternative splicing to remove a single adjacent exon. Other possibilities, such as either a more extensive muscle-specific alternative splicing or presence of an independent mutation in the same gene, remain to be investigated.

Analysis of PBL mRNA is useful for determining the effect that genotypic alteration has on transcript structure and improves confirmation of the frameshift hypothesis in cases where analysis at the genomic level is inconclusive. Observations of alternative splicing, however, do not always enable explanation of phenotype and could be misleading. Such discrepancies may arise either from genuine transcript-processing differences between muscle and PBLs or from PCR-dependent exaggeration of a small population of alternatively spliced transcripts present in both tissues. In addition, as the dystrophin protein may function as a homodimer, coexistence of truncated (frameshift) and interstitially deleted (in-frame) protein molecules in cells where alternative splicing occurs may result in a phenotype more severe than that expected from the mRNA species observed (see patients 02, 14–17, and 25). The frameshift hypothesis is based on an extrapolation from DNA to mRNA to protein. Study of the structure of mRNA therefore reduces the degree of extrapolation, but the hypothesis essentially deals with only two types of effect on the translation product—namely, (1) interstitial deletion of regions of relatively low importance and (2) truncation leading to essentially complete loss of function (whether due to high functional importance of the carboxy-terminal end or to loss of stability). It must therefore be remembered that any other effects on protein structure are beyond the field of the hypothesis and may manifest as "exceptions." The explanation of such phenotypes would be achieved only through studies at the protein level.

By this method, large (at least 1.8 kb) segments of coding region can be isolated from a readily accessible tissue, in quantities and quality sufficient for direct sequencing and/or other methods for mutation detection (e.g., see Cotton et al., 1988; Montandon et al., 1989; Orita et al., 1989; Sheffield et al., 1989), without the need for cloning. This makes it a particularly useful tool for the study of mutations in tissue-specific genes with large numbers of exons (where genomic studies would be impractical). The use of nested-amplification products, from PBL transcripts, for the detection of point mutations in factor VIIIc by chemical mismatch analysis has already been demonstrated in our laboratory (Naylor et al. 1991). As the phenomenon of "illegitimate transcription" (low-level transcription of tissue-specific genes in noncognate tissues) has been noted in a number of genes (e.g., anti-Müllerian hormone, β-globin, aldolase A, factor VIIIc [Chelly et al., 1989; Berg et al. 1990; Naylor et al. 1991], retinal blue pigment, factor IX, phenylalanine hydroxylase, tyrosine hydroxylase [Sarkar and Sommer 1989], and the cystic fibrosis gene [A. Harris, personal communication]), it is reasonable to expect that many tissue-specific genes will be amenable to the methods described in the present paper.

Acknowledgments

This work was supported by the Medical Research Council, the Muscular Dystrophy Group of Great Britain and Northern Ireland, the Spastics Society, and the Generation Trust. We are grateful to S. Abbs for DNA results and to Professor F. Giannelli for helpful discussions. We are also indebted to Drs. A. C. Berry and S. Hodgson at Guy's Hospital and to Professor V. Dubowitz at the Department of Paediatrics and Neonatal Medicine at the Hammersmith Hospital, for clinical samples.

References

Berg L-P, Wieland K, Millar DS, Schlösser M, Wagner M, Kakker VV, Reiss J, et al (1990) Detection of a novel point mutation causing haemophilia A by PCR/direct sequencing of ectopically-transcribed factor VIII mRNA. Hum Genet 85:655–658

Chelly J, Concordet J-P, Kaplan J-C, Kahn A (1989) Illegitimate transcription: transcription of any gene in any cell type. Proc Natl Acad Sci USA 86:2617–2621

Chelly J, Kaplan J-C, Maire P, Gautron S, Kahn A (1988) Transcription of the dystrophin gene in human muscle and non-muscle tissues. Nature 333:858–860

Chomczynski P, Sacchi N (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal Biochem 162:156–159

Cotton RGH, Rodrigues NR, Campbell RD (1988) Reactivity of cytosine and thymidine in single-base-pair mismatches with hydroxylamine and osmium tetroxide and its application to the study of mutations. Proc Natl Acad Sci USA 85:4397–4401

Den Dunnen JT, Grootscholten PM, Bakker E, Blonden LAJ, Ginjaar HB, Wapenaar MC, van Passen HMB, et al

(1989) Topography of the duchenne muscular dystrophy (DMD) gene: FIGE and cDNA analysis of 194 cases reveals 115 deletions and 13 duplications. Am J Hum Genet 45:835–847

Dubowitz V (1990) The Duchenne dystrophy story: from phenotype to gene and potential treatment. J Child Neurol 4:240–250

Emery AEH (1987) Duchenne muscular dystrophy. Oxford monographs in medical genetics, no 15. Oxford University Press, Oxford

Feener CA, Koenig M, Kunkel LM (1989) Alternative splicing of human dystrophin mRNA generates isoforms at the carboxy terminus. Nature 338:509–511

Gillard EF, Chamberlain JS, Murphy EG, Duff CL, Smith B, Burghes AHM, Thompson MW, et al (1989) Molecular and phenotypic analysis of patients with deletions within the deletion-rich region of the Duchenne muscular dystrophy (DMD) gene. Am J Hum Genet 45:507–520

Green PM, Bentley DR, Mibashan RS, Nilsson IM, Giannelli F (1989) Molecular pathology of haemophilia B. EMBO J 8:1067–1072

Hodgson S, Hart K, Abbs S, Heckmatt J, Rodillo E, Bobrow M, Dubowitz V (1989) Correlation of clinical and deletion data in Duchenne and Becker muscular dystrophy. J Med Genet 26:682–693

Holding C, Monk M (1989) Diagnosis of beta-thalassaemia by DNA amplification in single blastomeres from mouse preimplantation embryos. Lancet 2:532–535

Koenig M, Beggs AH, Moyer M, Scherpf S, Heindrich K, Bettecken T, Meng G, et al (1989) The molecular basis for Duchenne versus Becker muscular dystrophy: correlation of severity with type of deletion. Am J Hum Genet 45:498–506

Koenig M, Hoffman EP, Bertelson CJ, Monaco AP, Feener C, Kunkel LM (1987) Complete cloning of the Duchenne muscular dystrophy (DMD) cDNA and preliminary genomic organization of the DMD gene in normal and affected individuals. Cell 50:509–517

Koenig M, Kunkel LM (1990) Detailed analysis of the repeat domain of dystrophin reveals four potential hinge segments that may confer flexibility. J Biol Chem 265:4560–4566

Koenig M, Monaco AP, Kunkel LM (1988) The complete sequence of dystrophin predicts a rod-shaped cytoskeletal protein. Cell 53:219–228

Monaco AP, Bertelson CJ, Liechti-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

Montandon AJ, Green PM, Giannelli F, Bentley DR (1989) Direct detection of point mutations by mismatch analysis: application to haemophilia B. Nucleic Acids Res 17: 3347–3358

Naylor JA, Green PM, Montandon AJ, Rizza CR, Giannelli F (1991) Detection of three novel mutations in two haemophilia A patients by rapid screening of whole essential region of factor VIII gene. Lancet 337:635–639

Norman AM, Thomas NST, Kingston HM, Harper PS (1990) Becker muscular dystrophy: correlation of deletion type with clinical severity. J Med Genet 27:236–239

Orita M, Suzuki Y, Sekiya T, Hayashi K (1989) Rapid and sensitive detection of point mutations and DNA polymorphisms using the polymerase chain reaction. Genomics 5: 874–879

Roberts RG, Bentley DR, Barby TFM, Manners E, Bobrow M (1990) Direct diagnosis of carriers of Duchenne and Becker muscular dystrophy by amplification of lymphocyte RNA. Lancet 336:1523–1526

Saiki RK, Gelfand DH, Stoffel S, Sharf SJ, Higuchi R, Horn GT, Mullis KB, et al (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239:487–491

Sarkar G, Sommer SS (1989) Access to a messenger RNA sequence or its protein product is not limited by tissue or species specificity. Science 244:331–334

Sheffield VC, Cox DR, Lerman LS, Myers RM (1989) Attachment of a 40-base-pair G + C-rich sequence (GC-clamp) to genomic DNA fragments by the polymerase chain reaction results in improved detection of single-base changes. Proc Natl Acad Sci USA 86:232–236

Winnard A, Mendell JR, Florence J, Brooke M, Burghes AHM (1990) Detection of point mutations and examination of translational frameshift exceptions in Duchenne and Becker dystrophy. Am J Hum Genet 47(Suppl): A242

Winship PR (1989) An improved method for directly amplifying PCR amplified material using dimethyl sulphoxide. Nucleic Acids Res 17:1266