

## Distal transcript of the dystrophin gene initiated from an alternative first exon and encoding a 75-kDa protein widely distributed in nonmuscle tissues

J. P. HUGNOT\*, H. GILGENKRANTZ\*, N. VINCENT\*, P. CHAFEY\*, G. E. MORRIS†, A. P. MONACO‡, Y. BERWALD-NETTER§, A. KOULAKOFF§, J. C. KAPLAN\*, A. KAHN\*, AND J. CHELly\*\*‡

\*Institut Cochin de Génétique Moléculaire, Institut National de la Santé et de la Recherche Médicale U129, 24 rue du Fg Saint Jacques, 75014 Paris, France; †The North East Wales Institute, Deeside, Clwyd, CH5 4BR, Wales, United Kingdom; ‡Imperial Cancer Research Fund, Institute of Molecular Medicine, J. Radcliffe Hospital, Oxford OX3 9DU, United Kingdom; and §Biochimie Cellulaire, Collège de France, 75231 Paris Cedex 05, France

Communicated by Jean Dausset, March 19, 1992

**ABSTRACT** A transcript generated by the distal part of the Duchenne Muscular Dystrophy (DMD) gene was initially detected in cells where the full size 14-kilobase (kb) messenger RNA is not found at a significant level. This transcript, ≈4.5 kb long, corresponds to the cysteine-rich and carboxyl-terminal domains of dystrophin. It begins with a novel 80- to 100-nucleotide exon containing an ATG start site for a new coding sequence of 17 nucleotides in-frame with the consecutive dystrophin cDNA sequence from exon 63. This result suggests the existence of a third promoter that would be localized about 8 kilobases upstream from exon 63 of the *DMD* gene. The distal transcript is widely distributed but is absent in adult skeletal and myometrial muscle. It is much more abundant in fetal tissues. With an antibody directed against the dystrophin carboxyl terminus, the protein corresponding to this transcript was detected as a 70- to 75-kDa entity on Western blots. It was found in all tissues analyzed except in skeletal muscle. It was not found in lymphoblastoid cells from a Duchenne patient with a complete deletion of the dystrophin gene. The role and subcellular localization of this protein is not known. It may explain extramuscular symptoms exhibited by some Duchenne patients.

The huge X-linked *DMD* gene, which is abnormal in Duchenne and Becker muscular dystrophies (1, 2), is controlled by different promoters in different tissues (3, 4). A promoter active in skeletal muscle, heart, and smooth muscle and a neuron-specific promoter (5, 6), which is located 90 kilobase pairs (kbp) upstream of the muscle-specific promoter (7), control the production of two 14-kilobase (kb) mRNA species (8) that is translated into 427-kDa proteins (9) differing in their N-terminal ends. In addition, several forms generated by alternative splicing, some of them being tissue-specific or developmentally regulated, have been detected (3, 10, 11). Most of them involve the 3' coding part of the transcripts.

In this paper, we report the characterization and expression analysis of a newly discovered transcript of the *DMD* gene, probably similar to the species recently reported by Bar *et al.* (12). This mRNA, referred to as the "distal transcript," is about 4.5 kb long and shares with the dystrophin mRNA those exons that are located downstream of exon 62. Thus, it encodes the last two domains of the protein—i.e., the cysteine rich and C-terminal domains. Its 5' end consists of a new first exon spliced to the dystrophin exon 63 and located between exons 62 and 63 of the *DMD* gene. This alternative first exon contains a consensus ATG translation initiation codon in phase with the dystrophin gene sequence. The open reading frame codes for a 70- to 75-kDa protein that is

detected in several tissues. Transcripts containing the alternative first exon are widely distributed in adult and fetal tissues and cell lines, but are absent from adult skeletal muscle and myometrium.

### MATERIAL AND METHODS

**RNA Extraction.** Total RNA was extracted from different cell lines and frozen tissues by the method of Chirgwin *et al.* (13). The amount of RNA used for reverse transcription and amplification reactions (cDNA PCR) was determined by spectrophotometry and corrected by computerized image analysis (Vilber Lourmat, Marne la Vallée, France) after loading on an agarose electrophoresis gel stained with ethidium bromide.

**Quantitative cDNA PCR.** PCR was performed as described in detail by Chelly *et al.* (10, 14). The internal standard used here was the L-pyruvate kinase mRNA amplified from rat liver.

**Nuclease S1 Protection Assay.** This assay was performed with a single-stranded, quasi-end-labeled probe as described by Cognet *et al.* (15).

**RNA Blot-Hybridization Analysis.** After electrophoresis in a 20 mM methylmercury(II)/1% (wt/vol) agarose gel, Northern blot analysis was carried out as described by Bayley *et al.* (16).

**Anchored PCR.** The 5' sequence of the distal transcript was obtained by anchored PCR as described (17). cDNA synthesis was performed by primer extension of an oligonucleotide complementary to nucleotides 10,055–10,075. The cDNA was then 3'-tailed by terminal deoxynucleotidyltransferase (Boehringer Mannheim), under conditions designed to add about 15–30 dGMP residues. PCR amplification was performed with a 3' complementary oligonucleotide spanning nucleotides 9885–9904 of the dystrophin sequence. The 5' PCR primers and conditions have been described by Loh *et al.* (17). The PCR products were size-fractionated by 1% (wt/vol) agarose gel electrophoresis, electroeluted, digested with *Not I* restriction enzyme, and then cloned in the *Not I/HincII* sites of Bluescript KSII(–) plasmid (Stratagene).

**Mapping and Cloning of a Genomic Fragment Including the Alternative Exon.** Cosmids, isolated by using 63.1 dystrophin cDNA clone (position 7880 to the end) (18) as a hybridization probe on Imperial Cancer Research Fund flow-sorted chromosome X cosmid filter (19), were used to map and clone the genomic region of interest. Analysis of these cosmids by hybridization with the oligonucleotide probe allowed isolation of two overlapping cosmids (DMD3'-8 and DMD3'-12) that contained the sequence specific to the distal transcript. To define the position of the alternative exon, the cosmids were restriction-mapped with *Sac I*, *Sac II*, *HindIII*, and *EcoRI*. Southern blots were hybridized with probes derived from the exon corresponding to the 12-kb *HindIII* genomic

fragment detected by the 63.1 dystrophin cDNA probe and the alternative exon, respectively (see *Results*). To sequence the genomic fragment, *Hind*III fragments of cosmid DMD3'-8 were cloned into *Hind*III-cut Bluescript KSII(-) plasmid. Recombinant colonies hybridizing with a distal transcript-specific oligonucleotide probe were further analyzed and sequenced.

**Cell Culture.** Cultures of mouse brain neurons and astroglial cells were prepared and maintained as described (5, 20, 21).

**Protein Analysis.** Immunoblotting was performed essentially as described by Nicholson *et al.* (22) with the following modifications. The samples were fractionated on 1-mm SDS/polyacrylamide gel by using a 3% stacking gel and a 5–15% gradient-resolving gel. Mandra-1 is a primary monoclonal antibody against the end of the C-terminal domain of dystrophin (amino acid residues 3558–3684) (G.E.M., unpublished data).

## RESULTS

### Detection of a Different Transcript of the Dystrophin Gene.

To investigate simultaneously several parts of the dystrophin gene transcript in different tissues and cell lines, we coamplified by cDNA PCR two fragments of the dystrophin mRNA, one corresponding to exons 1 and 3 (fragment A, positions 209–331) and the other encompassing several exons of the distal part (fragment B, positions 9811–10,516) (Fig. 1 *Upper*). Quantitative analysis performed after 15 cycles of amplification showed that the 5' fragment A was detected only, as expected, in muscle and in spleen (in this tissue, the expression of dystrophin is related to the smooth muscle content of blood vessels), whereas the distal fragment B was reproducibly detected in addition in cultured hepatoma and lymphoblastoid cells, in which dystrophin is not expressed

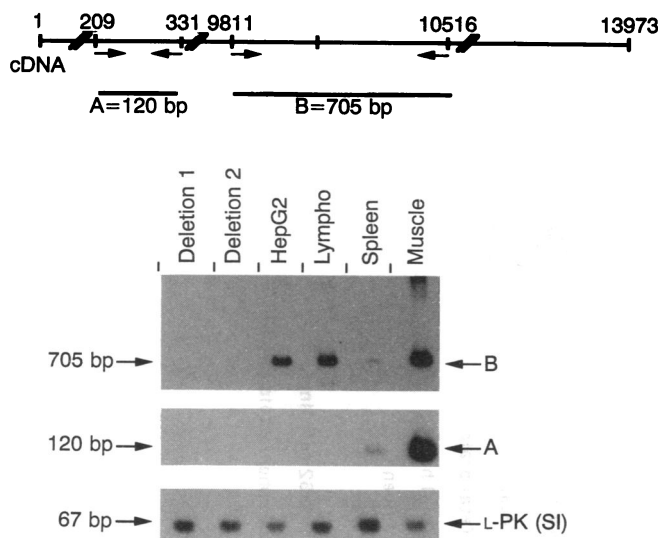


FIG. 1. Characterization of a nonmuscular transcript produced by the distal part of the dystrophin gene. (*Upper*) Position of the different primers on dystrophin mRNA and size of the amplified fragments. Nucleotide position numbers are on top. (*Lower*) Autoradiographs of a Southern blot of the cDNA PCR products obtained after 15 cycles. Hybridization was performed with an internal 5'-radiolabeled oligonucleotide probe. Total RNA was extracted from lymphoblasts of a patient with a chromosome X deletion that removed the whole dystrophin gene (lane Deletion 1; patient MJ), lymphoblasts from a patient with a chromosome X deletion that removed the distal part of the dystrophin gene (lane Deletion 2), HepG2 hepatoma cell line (lane HepG2), control lymphoblasts (lane Lympho), spleen (lane Spleen), and skeletal muscle (lane Muscle). SI, internal standard; L-PK, L-pyruvate kinase transcript.

Table 1. Position of dystrophin mRNA fragments analyzed by cDNA PCR in nonmuscular cells

Fragment	Nucleotide position	Result after amplification
1	Brain exon 1–exon 2	–
2	209–331	–
3	1017–1197	–
4	6465–6670	–
5	7366–7450	–
6	9365–9490	–
7	9811–9907	+
8	9811–10,516	+
9	10,520–10,855	+
10	10,785–11,307	+

(Fig. 1 *Lower*). The absence of amplification of these two fragments from two lymphoblastoid cell lines of patients bearing different deletions—a deletion of the entire dystrophin gene (patient MJ, deletion 1) and a deletion from exon 8 to the end of the gene (deletion 2)—ruled out the possibility of a contamination or of the amplification of a related transcript encoded by another gene. Since the limits between exons in this region are not precisely known, we also verified that no genomic DNA could be amplified with these primers. This indicates that they are indeed located in two different exons (data not shown). These results suggest that the distal part of the dystrophin gene is transcribed in significant amounts in the two cell lines, whereas the proximal part is not.

To further characterize the divergence region between the dystrophin mRNA and this new transcript, cDNA PCR amplification of different fragments of the dystrophin mRNA was carried out with several couples of primers (see Table 1). With couples of primers localized in the spectrin-like and the N-terminal regions, fragments of the expected size were observed only in muscle and spleen. Similar results were obtained with primers that allowed amplification of the sequence corresponding to the junction between the spectrin-like region and the cysteine-rich domain (positions 9365–9490). In contrast, fragments localized in the cDNA sequence corresponding to the cysteine-rich and C-terminal domains of dystrophin (fragments 7, 8, 9, and 10; Table 1) gave strong signals not only in muscle but also in lymphoblastoid and hepatoma cells. Altogether, these results indicate that the divergence between these two transcripts is located between positions 9365 and 9811. In addition, cDNA PCR amplification showed in a lymphoblast sample the presence of a signal of the same size as in muscle as well as shorter species that resulted from alternative splicing, suggesting the existence of a distal transcript-specific species. One of these alternatively

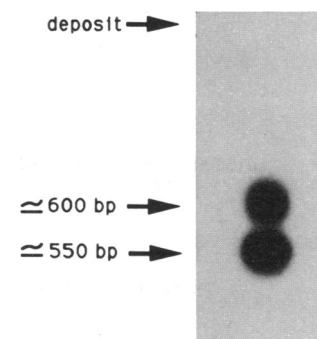


FIG. 2. Detection of the anchored PCR products. Autoradiograph of a Southern blot of the anchored PCR products electrophoresed on a 1% (wt/vol) agarose gel. Hybridization was performed with a 5'-radiolabeled dystrophin primer (positions 9461–9482).

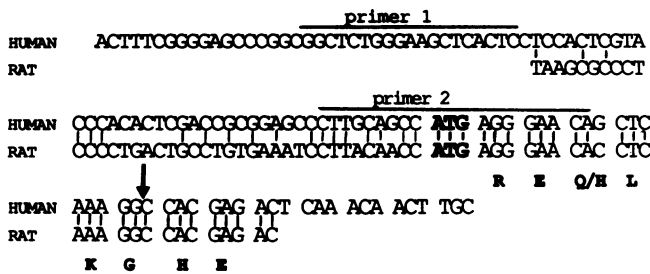


FIG. 3. Sequences of the human and rat distal alternative exon. The putative initiation codon is represented in boldface letters. The arrow indicates the exonic limit between the cysteine-rich domain coding region and the additional exon. Sequence identity between the two species is indicated by vertical bars. The sequences of the distal exon-specific primers 1 and 2 are indicated by horizontal lines. The deduced amino acid sequence beginning at the ATG codon is represented in the one-letter code.

spliced species corresponds to the 32-nucleotide exon deletion found by Bar *et al.* (12) in one liver cDNA clone.

To confirm the presence of a transcript in lymphoblasts, we performed a nuclease S1 protection assay with cloned fragment B (Fig. 1 *Upper*) as a probe. This fragment was protected not only in muscle but also in lymphoblasts (data not shown).

In Northern blot analysis with fragment B (Fig. 1 *Upper*) as a probe, we detected in lymphoblastoid and hepatoma cells a transcript migrating just below 28S RNA, which corresponds to a length of about 4.5 kb.

**Cloning of the 5' End of the Distal Transcript.** To obtain more details on the 5' end of this distal transcript, we used the anchored PCR technique (17) on a lymphoblast RNA sample. The amplification products consisted mainly of two bands hybridizing with an internal oligonucleotide specific to the dystrophin gene (Fig. 2). After cloning and sequencing, we found that the largest fragment contained the dystrophin

cDNA until position 9431, preceded by a 80- to 100-nucleotide-long new sequence (Fig. 3). The sequence of the second band revealed that it was artifactual, since it did not contain the poly(dGMP) stretch and, thus, was most likely generated by internal priming of the anchored primer during the PCR procedure.

The new sequence seems to be an additional exon rather than an unspliced intron, since the sequence at the junction does not show an intron acceptor site. An ATG codon in good agreement with the Kozak consensus initiation sequence (23) and in frame with the dystrophin reading frame is seen 20 nucleotides upstream from the divergence point.

**Expression of the Distal Transcript.** cDNA PCR with a primer situated in the additional exonic sequence was negative in patient MJ, indicating that it corresponds to a part of the dystrophin gene (Fig. 4 *Upper*).

We used the same couple of primers to quantify the distal transcript in various adult and fetal tissues, and cell cultures. The results are shown in Fig. 4 *Lower*, and the quantification is reported in Table 2. The distal transcript is particularly abundant in fetal tissues (brain, heart, liver, and stomach). It is also present to a lesser extent in various adult tissues and cultured cells but is absent from adult myometrium and leg skeletal muscle.

**Mapping of the Alternative Exon Within the Dystrophin Gene and Nucleotide Sequence Surrounding This Exon.** Southern blot of a cosmid DMD3'-8 *Hind*III digest hybridized with oligonucleotide probes specific to the exon 63 (GenBank M86887) of the dystrophin cDNA, and the alternative exon showed that this cosmid contains the two exons (data not shown). Restriction mapping of cosmids DMD3'-8 and DMD3'-12, combined with oligonucleotide probe-specific hybridization with both probes, allowed localization of the new exon about 8 kb upstream from the exon 63 (Fig. 5 *Left*).

The nucleotide sequence of the region downstream from the alternative exon shows a donor splice site in agreement with the consensus sequence. The sequence upstream from

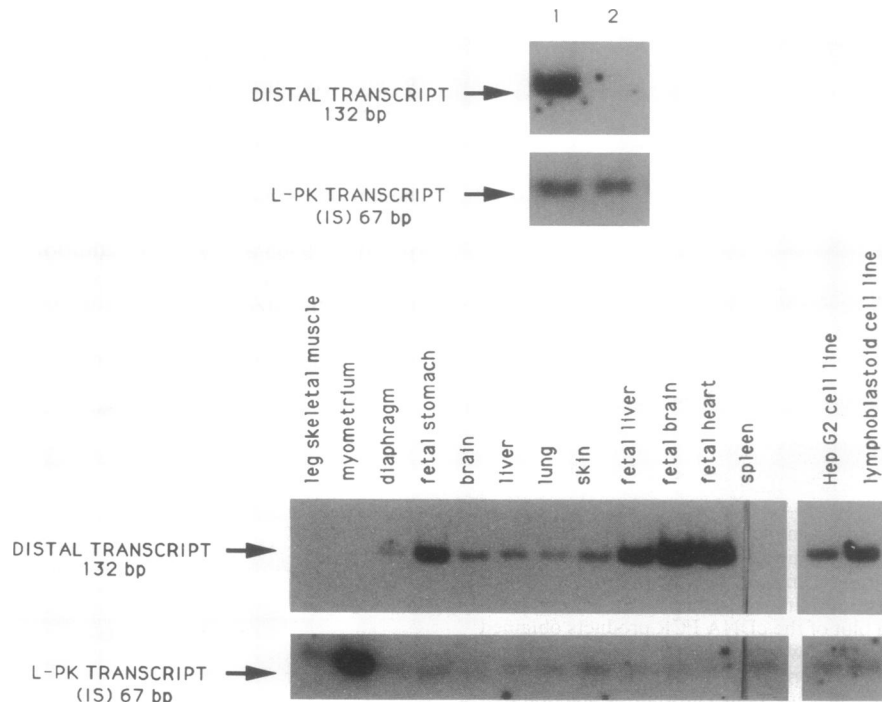


FIG. 4. Expression of the distal transcript in human tissues and cell lines. Autoradiographs of Southern blots of the cDNA PCR products obtained after 15 cycles are shown. The distal-transcript-specific primer 1 (see Fig. 3) and a primer spanning nucleotides 9461-9482 were used for the amplification. The oligoprobe used is the 5' radiolabeled distal primer 2 (see Fig. 3). IS, internal standard; L-PK, L-pyruvate kinase. (*Upper*) Total RNA was extracted from lymphoblastoid cell lines of a normal subject (1) and patient MJ (2). (*Lower*) Total RNA was extracted from indicated tissues and cell cultures. The fetal tissues are from a 3.5-month-old fetus.



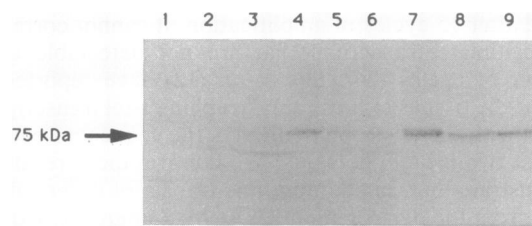


FIG. 6. Detection of a 70- to 75-kDa protein by immunoblot analysis with monoclonal antibody Mandra 1. The proteins were extracted from: rat leg skeletal muscle (lane 1); lymphoblastoid cells of patient MJ (lane 2); lymphoblastoid cells of a normal subject (lane 3); cultures of mouse brain neuronal cells at 22 days of culture (lane 4), 15 days of culture (lane 5), and 7 days of culture (lane 6); cultures of mouse brain astroglial cells (lane 7); rat brain (lane 8); and mdx (dystrophin-deficient mouse) mouse brain (lane 9). The arrow indicates the 70- to 75-kDa protein encoded by the dystrophin gene.

adult skeletal muscle and in smooth muscle. Transcriptional interference (25) between the muscular and the distal promoter could be one of the mechanisms responsible for this extinction.

The translation product of the distal mRNA is expected to possess 607 amino acid residues shared with dystrophin and a 6-amino-acid-long unique N-terminal peptide with polar characteristics (Arg-Glu-Gln-Leu-Lys-Gly). Its predicted molecular weight (70 kDa) fits well with the mobility of a protein that we have observed on Western blot with a monoclonal antibody specific to the carboxyl domain. This 70- to 75-kDa protein is not recognized by antibodies directed against other dystrophin domains and is found in the same cells and tissues as those accumulating the distal transcript; it is absent in skeletal muscle and present in liver, brain, lymphoblastoid cells, and Hep G2 cells. In addition, both the distal transcript and the 70- to 75-kDa protein are absent in patients with dystrophin gene deletions involving the C-terminal domains. Therefore, the 70- to 75-kDa protein is most likely the product of the distal transcript.

The putative physiological role of the 70- to 75-kDa protein is unknown. Since it is relatively abundant in brain, where, according to Bar *et al.* (12), the distal transcript is expressed at least at the same level as the dystrophin transcript, it may play a role in brain cell functions. In fact, this function could be common in glial cells and neurons. Further studies will establish whether there is a correlation between synthesis of the 70- to 75-kDa protein and the mental retardation recognized in as much as 1:3 of the Duchenne muscular dystrophy patients (26).

**Note Added in Proof.** The products described in this paper are probably identical to the species recently reported by Blake *et al.* (27) and Lederfein *et al.* (28).

This work was supported by grants from the Association Française contre les Myopathies (AFM). J.P.H. and N.V. are fellows from the AFM.

- Monaco, A. P. & Kunkel, L. M. (1988) *Adv. Hum. Genet.* **17**, 61-98.
- Worton, R. G. & Thompson, M. W. (1988) *Annu. Rev. Genet.* **22**, 601-629.
- Feener, C. A., Koenig, M. & Kunkel, L. M. (1989) *Nature (London)* **338**, 509-511.
- Nudel, U., Robzyk, K. & Yaffe, D. (1988) *Nature (London)* **331**, 635-638.
- Chelly, J., Hamard, G., Koulakoff, A., Kaplan, J. C., Kahn, A. & Berwald-Netter, Y. (1990) *Nature (London)* **344**, 64-65.
- Barnea, E., Zuk, D., Simantov, R., Nudel, U. & Yaffe, D. (1990) *Neuron* **5**, 881-888.
- Boyce, F., Beggs, A., Feener, C. & Kunkel, L. M. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 1276-1280.
- Chamberlain, S. C., Pearlman, J. A., Muzny, D. M., Gibbs, R. A., Ranier, J. E., Reeves, A. A. & Caskey, C. T. (1988) *Science* **239**, 1416-1418.
- Hoffman, E. P., Brown, R. H., Jr., & Kunkel, L. M. (1987) *Cell* **51**, 919-928.
- Chelly, J., Gilgenkrantz, H., Lambert, M., Hamard, G., Chafey, P., Katz, P., de la Chapelle, A., Koenig, M., Ginjaar, I., Fardeau, M., Tomé, F., Kahn, A. & Kaplan, J. C. (1990) *Cell* **63**, 1239-1248.
- Geng, Y., Sicinski, P., Gorecki, D. & Barnard, P. J. (1991) *Neuromuscular Disorder* **1**, 125-133.
- Bar, S., Barnea, E., Levy, Z., Neuman, S., Yaffe, D. & Nudel, U. (1990) *Biochem. J.* **272**, 557-560.
- Chirgwin, J., Przybyla, A., MacDonald, R. & Rutter, W. (1979) *Biochemistry* **18**, 5294-5299.
- Chelly, J., Montarras, D., Pinset, C., Berwald-Netter, Y., Kaplan, J. C. & Kahn, A. (1990) *Eur. J. Biochem.* **187**, 691-698.
- Cognet, M., Lone, Y. C., Vaultont, S., Kahn, A. & Marie, J. (1987) *J. Mol. Biol.* **196**, 11-25.
- Bayley, J. M. & Davidson, N. (1976) *Anal. Biochem.* **70**, 75.
- Loh, E. Y., Elliott, J. F., Cwirla, S., Lanier, L. L. & Davis, M. (1989) *Science* **243**, 217-220.
- Koenig, M., Hoffman, E. P., Bertelson, C. J., Monaco, A. P., Feener, C. C. & Kunkel, L. M. (1987) *Cell* **50**, 509-517.
- Nizetic, D., Zehetner, G., Monaco, A. P., Gellen, L., Young, B. D. & Lehrach, H. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 3233-3237.
- Berwald-Netter, Y., Martin-Moutot, N., Koulakoff, A. & Couraud, F. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 1245-1249.
- Nowak, L., Ascher, P. & Berwald-Netter, Y. (1987) *J. Neurosci.* **7**, 101-109.
- Nicholson, L. V. B., Davison, K., Falkous, G., Harwood, C., O'Donnell, E., Slater, C. R. & Harris, J. B. (1989) *J. Neurol. Sci.* **94**, 125-136.
- Kozak, M. (1987) *Nucleic Acids Res.* **15**, 8125-8148.
- Chelly, J., Concordet, J. P., Kaplan, J. C. & Kahn, A. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 2617-2621.
- Proudfoot, N. J. (1986) *Nature (London)* **322**, 562-565.
- Emery, A. E. H. (1988) *Duchenne Muscular Dystrophy* (Oxford Univ. Press, Oxford).
- Blake, D., Love, D., Tinsley, J., Morris, C., Turloy, H., Gatter, K., Dickson, G., Edwards, Y. & Davies, K. (1992) *Hum. Mol. Genet.* **1**, 103-109.
- Lederfein, D., Levy, Z., Augier, N., Mornet, D., Morris, G., Fuchs, O., Yaffe, D. & Nudel, U. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 5346-5350.