Van Khanh Tran · Zhujun Zhang · Mariko Yagi
Atsushi Nishiyama · Yasuaki Habara
Yasuhiro Takeshima · Masafumi Matsuo

A novel cryptic exon identified in the 3' region of intron 2 of the human dystrophin gene

Received: 30 May 2005 / Accepted: 20 June 2005 / Published online: 30 August 2005
© The Japan Society of Human Genetics and Springer-Verlag 2005

Abstract The dystrophin gene, which is mutated in Duchenne muscular dystrophy (DMD), is the largest known human gene and is characterized by the huge size of its introns. Intron 2, the second largest intron, is 170-kb long and has been shown to include a 140-bp cryptic exon (exon 2a) in its 5' region. The rest of this intron has no known function. In this study, we find that another cryptic exon, located in the 3' region of intron 2, is activated in a promoter- or tissue-specific manner. An unknown 98-bp insertion precisely between exons 2 and 3 was identified in one of the dystrophin mRNAs from lymphocytes of a DMD patient with a duplication of exon 2. This 98-bp sequence, located in the 3' region of intron 2, was found to possess a branch point, acceptor and donor splice-site consensus sequences, and an exonic splicing enhancer sequence, and thus is a novel exon, which we named "exon 2b." In lymphocytes, exon 2b incorporation was detected in the muscle-specific, promoter-driven transcript. Five of 20 normal human tissue mRNAs, including cardiac and skeletal muscle mRNAs, were confirmed to contain a fragment extending from exon 1 to exon 2b by reverse transcription PCR amplification, indicating that exon 2b is activated in a tissue-specific manner. This provides a clue to a novel cause of dystrophinopathy.

Introduction

The human dystrophin gene, which is defective in patients with Duchenne or Becker muscular dystrophy

(DMD/BMD), spans approximately 3,000 kb of the X-chromosome and encodes a 14-kb transcript consisting of 79 exons (Ahn and Kunkel 1993; Nishio et al. 1994). Genomic structural analysis disclosed at least eight alternative promoters over the entire dystrophin gene, producing tissue-specific dystrophin isoforms (Ahn and Kunkel 1993; Nishio et al. 1994). Consequently, more than 99% of the gene sequence is comprised of introns and has been considered functionless. An alternative promoter identified within the largest intron, the 250-kb intron 44, regulates the expression of a tissue-specific dystrophin isoform, and this isoform has been suggested to be required for normal intellectual development (Bardoni et al. 2000; Felisari et al. 2000). In contrast, the second largest intron, the 170-kb-long intron 2, has been shown to contain a cryptic exon, exon 2a, in its 5' region, but the physiological role of exon 2a is still unknown (Dwi Pramono et al. 2000). Recently, a part of the 5' region of intron 2 was shown to be incorporated into dystrophin mRNA due to an activating mutation in the splice donor site of an embedded weak exon (exon p2a) (Yagi et al. 2003). So far, two tiny segments of the huge intron 2 have been shown to be incorporated into dystrophin mRNA, leaving nearly 170 kb uncharacterized.

Splicing is the process that removes introns from pre-mRNA thereby producing mature mRNA consisting of only exons. The presence of well-defined cis elements, namely, the 5' and 3' splice sites and the branch point, is necessary but not sufficient to define intron–exon boundaries in pre-mRNA (Senapathy et al. 1990). Pseudoexons that match splice-site consensuses have been identified in introns, but their inclusion in mRNA is prevented by silencer elements (Sironi et al. 2004). However, unconventional splicing defects often occur at exons with weak homology to canonical splicing sequences, leading to dystrophinopathies (Tuffery-Giraud et al. 2003; Yagi et al. 2003).

Complex patterns of alternative splicing of the 5' region of the dystrophin gene have been reported (Chelly et al. 1991; Reiss and Rininsland 1994; Torelli

V.K. Tran · Z. Zhang · M. Yagi · A. Nishiyama · Y. Habara
Y. Takeshima · M. Matsuo (✉)
Department of Pediatrics,
Kobe University Graduate School of Medicine,
7-5-1 Kusunokicho, Chuo,
Kobe 650-0017, Japan
E-mail: matsuo@kobe-u.ac.jp
Tel.: +81-78-3826080
Fax: +81-78-3826098

and Muntoni 1996; Surono et al. 1997). The translational reading frame rule explains genotype–phenotype correlation in dystrophinopathy; i.e., out-of-frame deletion of the dystrophin gene results in severe DMD while in-frame deletions result in mild BMD (Chelly et al. 1991; Winnard et al. 1992, 1995). However, many dystrophinopathy cases with deletions in the 5' region of the dystrophin gene have been shown to be exceptions to this rule, and alternative splicing has been considered to be a factor leading to such an exception by changing the translational frame (Muntoni et al. 1994).

We have analyzed dystrophin mRNA expressed in peripheral lymphocytes from more than 100 cases of dystrophinopathy. Here, we identify an unknown sequence inserted into a dystrophin transcript in a case with exon 2 duplication of the dystrophin gene, and we find that the sequence is a novel cryptic exon (exon 2b) located in the 3' region of intron 2 of the dystrophin gene. Exon 2b is incorporated into mRNA in a promoter- or tissue-specific manner. This provides a clue to a novel cause of dystrophinopathy.

Patient and methods

Case

A 5-year-old Japanese boy was referred to the Kobe University Hospital for the genetic diagnosis of DMD. He was the first-born boy, and his family history disclosed no muscular disease. At age 4, he was shown to have an extremely high level of serum CK (13,750 IU/l, normal: 56–248 IU/l) and was clinically diagnosed as DMD. Physical examination disclosed mild calf hypertrophy, and he showed Gowers' sign. Chest X-ray and ECG were normal. All analysis was done after obtaining informed consent from his parents.

Analysis of genomic DNA

Genomic DNAs were isolated from lymphocytes of DMD patients and a normal male individual using a Wizard genomic DNA extraction kit (Promega Corporation, Madison, WI, USA). Conventional PCR amplification was employed to find deletion mutations in 19 deletion-prone exons of the dystrophin gene (Chamberlain et al. 1988; Beggs et al. 1990). PCR was performed essentially, as described previously (Matsuo et al. 1991). To examine the entire dystrophin gene, Southern blot analysis of the patient's genomic DNA was performed using *Hae*III-digested cDNA fragments as probe. A genomic region encompassing the 98-bp inserted sequence (exon 2b) was amplified using primers derived from the flanking sequences (Table 1). The copy number of exons was assessed by semiquantitative, multiplex PCR. Seven segments in the 5' region of the dystrophin gene, including exon 1; exon 1a; exon 2, a pseudoexon (exon p2a) in intron 2 (Yagi et al. 2003); exon 2a; exon 2b; and exon 3 were amplified in one PCR reaction together with the exon-19-encompassing region. Amplification was carried out in a total volume of 20 µl containing 400 ng of genomic DNA, 2 µl 10X Ex Taq Buffer (Takara Bio Inc., Kyoto, Japan), 2 µl of 2.5 mM dNTPs, 5 pmol of each primer, and 1U of Ex Taq Polymerase (Takara Bio Inc., Kyoto, Japan). PCR cycling conditions were as follows: an initial denaturation at 94°C for 5 min followed by 20 cycles of denaturation at 94°C for 45 s, annealing at 60°C for 45 s, extension at 72°C for 2 min, and a final extension at 72°C for 5 min. To quantify the amplified products, 1 µl of each reaction mixture mixed with 5 µl of the loading buffer solution containing size markers (15 and 1,500 bp) was analyzed by capillary electrophoresis (Agilent 2001 Bioanalyzer with DNA 1000 Lab Chips, Agilent Technologies, Palo Alto, CA, USA). The amount of each PCR product was quantified by measuring the peak area and calculating the ratio of this area to that of exon 19. The sequences of the primers used in this study are listed in Table 1. All

Table 1 Primer sequences

Target region	Forward primer	Reverse primer
Exon 1	SQQPmF: TAGACAGTGGATACATAACAAATGCATG	SQQPmR: TTCTCCGAAGGTAATTGCCTCCCAGATCTGAGTCC
Exon 1a	SQQ1aF: GAGCTATTTGCCACTTTTACCG	SQQ1aR: GGCCTGTTAGAAAGTGACATTC
Exon 2	SQQ2F: AAAAGAAAACATTCACAAAATGGG	SQQ2R: GTGTATCTTTGCCATATCTTCTGC
Exon P2a	int2SKF: TTCCATTTTCTCCGCAGCCC	int2SKR: GCATCATCAGCAAAACCTTCCG
Exon 2a	g2aF: TAGAGTTATCCTAGAGAGGTGG	g2aR: TCACGTGCATCATCCAGCAAC
Exon 2b	g2bF: AAAGGCTTGATACACATGGATA	g2bR: AGGTAGGGCAGGATAAATCGT
Exon 3	SQQ3F: TCATCCGTCATCTTCGGCAGATTAA	SQQ3R: CAGGCGGTAGAGTATGCCAAATGAAAATCA
Exon 19	SQQ19F: TTCTACCACATCCCATTTTCTTCCA	SQQ19R: GATGGCAAAAGTGTTGAGAAAAAGTC

PCR oligonucleotide primers were synthesized off site (Hokkaido System Science Co. Ltd., Sapporo, Japan).

Analysis of dystrophin transcripts

Total RNA was isolated from peripheral lymphocytes, as previously described (Matsuo et al. 1991). A fragment encompassing exons 1–5 of dystrophin mRNA was analyzed by reverse-transcription (RT), seminested PCR. The first PCR was done to amplify the region comprising exons 1–8 using primers located in each exon (M1: ATGCTTTGGTGGGAAGAAGTAG and c8R: TGTTGAGAATAGTGCATTTGATG, respectively) followed by the second amplification of a fragment comprising exons 1–5 (primer c5R: TGCCAGTG-GAGGATTATATTCCAA), as described previously (Suminaga et al. 2002).

To examine the promoter specificity of exon 2b incorporation, fragments stretching from promoter-specific exon 1 to exon 2b were amplified from lymphocyte cDNA. PCR primers were designed to detect promoter-specific transcripts. PCR detection of transcripts from the L, M (exon 1), C, or P promoters was performed using different exon-1-specific forward primers (L1: ACTGACACATAGAGTAAC, C1: TTGATTTGTTA-CAGCAGCCAACTTAT, M1, and P1: CCAGGTTTACCATACCCCATAGA, respectively). A reverse primer for exon 2b (ex 2b: GGAGGTTGCATTGAGTTGAG) was used in combination with one of each of the unique exon-1-specific primers. cDNA corresponding to 0.2 µg of the RNA samples was subjected to PCR amplification.

To examine the efficiency of exon 2b activation in different tissues, fragments spanning from exon 1 to exon 2b and from exon 1 to exon 5 were amplified from cDNA prepared from total RNA from 20 human tissues (adrenal gland, brain cerebellum, whole brain, fetal brain, fetal liver, heart, kidney, liver, whole lung, placenta, prostate, salivary gland, skeletal muscle, spleen, testis, thymus, thyroid gland, trachea, uterus, and spinal cord; BD Biosciences, San Jose, CA, USA). cDNA corresponding to 0.2 µg of each RNA sample was subjected to PCR amplification. Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) cDNA was amplified as a control. PCR reactions (20 µl) contained 1 µl of cDNA and 200 nM of each primer and 250 µM dNTPs. PCR was performed using 2 µl of 10× Ex Taq buffer (Takara Bio Inc., Kyoto, Japan) and 1 U of Ex Taq polymerase (Takara Bio Inc., Kyoto, Japan). A 5-min, 94°C denaturation step was followed by 30 cycles of PCR (94°C denaturation for 0.5 min, 60°C annealing for 0.5 min, 72°C extension for 0.5 min) followed by extension at 72°C for 7 min. A 10-µl sample of each PCR reaction was separated on an agarose gel containing 0.2 mg/ml of ethidium bromide, prior to photography.

DNA sequencing

For DNA sequencing, amplified products were separated by electrophoresis in low-melting-point agarose

gels. Bands of amplified products were cut out, and the DNA was purified. The purified DNA was subcloned into vector pT7 (Novagen, Inc., Madison, WI, USA) and the inserted DNA was sequenced using an automated DNA sequencer (model 373A, Perkin-Elmer Applied Biosystems Inc., Norwalk, CT, USA).

Results

PCR amplification of the selected 19 exons of the dystrophin gene disclosed neither deletion mutations in the index case nor any deletions or duplications detected by the conventional Southern blot analysis. However, PCR amplification of the region encompassing exon 2 appeared to result in a larger amount of amplified product from the case than from the control. This was quantified using capillary electrophoresis. In the coamplified products of the exon 2 and 19 encompassing regions, the ratio of exon 2 peak area to exon 19 peak area of the index case was twice that of the control (0.76 versus 0.34; Fig. 1). This indicated that exon 2 was duplicated in the genome of the index case.

In order to examine the duplication of exon 2 at the mRNA level, dystrophin mRNA expressed in the patient's peripheral lymphocytes was analyzed. The region encompassing exons 1–5 was amplified by RT nested PCR. Remarkably, one barely visible, weak band and as well as two major, equally dense bands were obtained (Fig. 2). Each of the bands was sequenced after subcl-

Fig. 1 Quantification of PCR products. Capillary electrophoretic patterns of PCR products are shown. Eight genomic regions were coamplified in one PCR reaction, and the products were separated using capillary electrophoresis. The position of each amplified product of exons 1, 1a, 2, p2a, 2a, 2b, 3, and 19 is marked above its peak (lower). The peak area of exons 2 and p2a is nearly double in the patient (upper). IS refers to 1,500-bp marker

Fig. 2 Reverse transcription (RT)-PCR amplification of a fragment comprising exons 1–5 from lymphocyte mRNA. **a** RT-PCR products: Three PCR products were visualized on the gel from the index case—one barely visible, and two others clearly visible and with equally dense bands (*P*). One discrete amplified product was obtained from the control (*C*). A schematic representation of exon organization in the amplified fragments is shown at the right of the products. *C* and *P* refer to control and index case, respectively. **b** A partial sequence of each product is shown. Sequences joining the

double exon 2 are shown (*upper*). The 3' terminal sequence of exon 2 (CTAAG) is joined precisely to the 5' end of the sequence of exon 2 (ATGAA). Partial sequences at the junctions of the inserted sequence and its flanking authentic exons are shown (*lower*). The terminal sequence of exon 2 (CTAAG) is joined precisely to the 5' end of the 98-bp inserted sequence (GCTAG), and at its 3' end (TACAG), the unknown sequence is joined precisely to the 5' end of the sequence of exon 3 (TTTGG)

oning. Sequencing of the smallest fragment revealed a sequence of tandem exon 2 sequences between exons 1 and 3 (Fig. 2). Sequencing of the middle-sized band revealed an insertion of exon 1a between exon 1 and 2 in addition to duplication of exon 2. The sequence of the largest fragment revealed the same exons as in the middle-sized one, but, remarkably, an unidentified 98-bp sequence was found to be inserted precisely between tandem exon 2 and exon 3 (Fig. 2). Since all three dystrophin mRNAs contained tandem exon 2 sequences, we concluded that the index case had a duplication of exon 2. This mutation created a premature stop codon at 15th codon of the duplicated exon 2 sequence and was determined as a cause of DMD.

The fact that the 98-bp sequence was inserted precisely between exons 2 and 3 led us to speculate that the sequence could be a retained intron or an unknown exon. A BLAST search of the 98-bp sequence revealed an identical sequence in the 3' region of intron 2 (bases 10151–10054 of GenBank AL121880). The 98-bp sequence was located 82 kb downstream from exon 2a and 29 kb upstream of exon 3 (Fig. 3). Remarkably, the AG and GT dinucleotides that are absolutely conserved at

the splice acceptor and donor sites of all introns, respectively, were identified immediately adjacent to the 5' and 3' ends of the 98-bp sequence (Fig. 3). The Shapiro probability scores for splice acceptor and donor sites were 0.77 and 0.74, respectively (Shapiro and Senapathy 1987). Furthermore, the sequence TATTAAT, a perfect match to the branch-point consensus sequence, was identified 84 bp upstream of the novel sequence (Fig. 3). A polypyrimidine tract was also identified between the putative branch point and the splice acceptor site (Fig. 3). Splicing enhancer sequences have been identified in exon sequences and are critical for proper incorporation of exons into mRNA (Schaal and Maniatis 1999). The splicing enhancer sequence within the 98-bp sequence was examined by ESE Finder (Cartegni et al. 2003). The heptanucleotide CTCCCGG in the middle of exon 2b has a score (3.74) higher than the threshold score (1.95) for SF2/ASF binding, and we consider it a strong candidate for a splicing enhancer. Since the 98-bp-inserted sequence exhibited all of the characteristics typical of a genomic exon and was inserted between authentic dystrophin exons, we refer to it as the novel exon 2b.

Fig. 3 Genomic structure and sequence. **a** Schematic description of intron 2 of the dystrophin gene. Intron 2 is the second largest intron (170 kb). One cryptic exon (exon 2a) is located 59 kb downstream from the 3' end of exon 2. One pseudoexon (exon p2a) is 5 kb downstream of exon 2. The 98 bp of the unidentified sequence is identical to the intron 2 sequence (nt 10151–10054, GenBank AL121880) located 82 kb downstream of exon 2a and 29 kb upstream of exon 3. Remarkably, two nucleotides, both upstream and downstream of the sequence, are AG and GT dinucleotides that are conserved at splicing acceptor and donor sites, respectively. Their Shapiro probability scores for splice donor and acceptor sites are 0.77 and 0.74, respectively. *Boxes and lines* indicate exons and introns, respectively. The *dotted box* indicates the novel cryptic exon, *brackets* indicate the size, and the *numbers*

under the boxes show the Shapiro's splicing probability scores at splicing sites. **b** Genomic nucleotide sequences of the inserted 98-bp sequence and its flanking introns are shown. The 98 bp of exon 2b are shown in *uppercase* letters, and the 139 and 184 bp upstream and downstream of the exon 2b sequence, respectively, are shown in *lower case* letters. Absolutely conserved AG and GT (*bold*) dinucleotides are present at the boundaries between exon 2b and its flanking regions. The branch point was identified 84 bp upstream of exon 2b within the consensus sequence tattaat (*inverted triangle*). Between exon 2b and the branch point, a polypyrimidine tract was identified. The exonic splicing enhancer sequence is *boxed*. *Horizontal arrows* indicate the locations and directions of primers. *Superscripted reference numbers* indicate numbered nucleotides of AL121880 (GenBank)

As we supposed that a genomic mutation located near exon 2b activated the incorporation of exon 2b in the index case, we examined the genomic sequence encompassing exon 2b. A total of 421 bp of genomic DNA was PCR amplified, but the sequence of the product was completely normal (data not shown), thus, no genomic mutation contributing to the activation of exon 2b was found.

Although tandem exon 2 sequences were identified in the case's dystrophin mRNA, only one exon 2b was identified. This fact suggested two possibilities: (1) a single copy of exon 2b was present in his genome, or (2) exon 2b was duplicated but only one copy was incorporated. To explore this, the copy numbers of exon 2b and nearby exons were examined by coamplification of the regions encompassing exons 1, 1a, 2, p2a, 2a, 2b, 3, and 19 (Fig. 1). The amplified products were quantified. The ratio of the peak area of exon 2b to that of exon 19

from the case was the same as that of the male control, indicating the presence of a single copy of exon 2b in the case's genomic DNA. We conclude that a single exon 2b downstream of the duplicated exon 2 was activated. In addition, the exon p2a to exon 19 peak area ratio was twice that of the control while those of exons 1a and 2a were the same as the control. This indicated that the duplicated region extended from exon 2 to exon p2a but not to exon 1a or 2a. The size of the duplication was calculated to be at least 5.5 kb (Fig. 3).

Though we have analyzed dystrophin mRNAs in peripheral lymphocytes obtained from more than 100 dystrophinopathy cases, exon 2b incorporation has never before been identified (Matsuo et al. 1991; Hagiwara et al. 1994; Surono et al. 1999; Adachi et al. 2003; Yagi et al. 2003). In order to see whether exon 2b activation is common to all exon 2 duplication mutations, we analyzed dystrophin mRNA from lymphocytes of

Fig. 4 Promoter specificity of exon 2b incorporation. Dystrophin transcripts from four different upstream promoters were analyzed for incorporation of exon 2b in lymphocyte mRNA. Exon 2b incorporation was examined by amplifying a region from tissue-specific exon 1 to exon 2b by reverse transcriptase (RT)-nested

PCR. Only one product extending from muscle exon 1 to exon 2b was amplified (*M*) while fragments extending from lymphocyte (*L*)-, cortical (*C*)-, and Purkinje cell (*P*)- specific exons 1–2b could not be amplified. *C* and *P* refer to the control and the index case, respectively

another case with an exon 2 duplication. The sequences of the subcloned, amplified products encompassing exons 1–5 revealed duplication of exon 2 but no incorporation of exon 2b (data not shown). We concluded that the incorporation of exon 2b was specific to the index case. The result indicated that duplication of exon 2 is not by itself an activator of exon 2b incorporation.

The dystrophin gene has four alternative promoters at its 5' end, and each promoter is activated in a tissue- or developmentally specific manner. To assess the promoter specificity of exon 2b activation, dystrophin transcripts from each promoter were analyzed for incorporation of exon 2b. The region encompassing promoter-specific exon 1 to exon 2b was amplified by RT-PCR from mRNA from the case's lymphocytes. As expected, a fragment extending from the muscle-promoter-specific exon 1 to exon 2b was amplified from the index case and the control (Fig. 4). However, transcripts extending from the lymphocyte (*L*)-, cortical (*C*) - or Purkinje cell (*P*)-specific exon 1 to exon 2b could not be amplified, indicating that exon 2b is not incorporated into other promoter-specific transcripts. We conclude that exon 2b incorporation is dependent upon the muscle-specific promoter.

The tissue specificity of exon 2b activation was examined by analysis of RNAs derived from 20 different

human tissues. RT-PCR products encompassing exons 1–2b were obtained from heart, lung, prostate, salivary gland, and skeletal muscle (Fig. 5) while a fragment extending from exon 1 to exon 5 was amplified from all tissues (Fig. 5). This indicates that exon 2b incorporation is under the control of one or more tissue-specific factors.

The protein-coding capacity of the novel dystrophin-transcript-retaining exon 2b was examined. Exon 2b encoding 98 bp disrupted the open reading frame. Furthermore, as exon 2b does not contain an in-frame ATG codon after the last termination codon, it is unlikely that a transcript containing exon 2b would direct the synthesis of a novel polypeptide. The transcript would be expected to allow reinitiation of translation at a downstream ATG codon (Malhotra et al. 1988) or to be of other, unknown biological significance (Galante et al. 2004; Graveley 2005).

Discussion

It is well known that DMD and BMD are caused by mutations in the dystrophin gene, as are X-linked dilated cardiomyopathy and abnormality of electroretinogram (D'Souza et al. 1995; Ferlini et al. 1999). Mental retardation is observed in one third of patients with DMD,

Fig. 5 Tissue specificity of exon 2b incorporation. A cDNA fragment extending from exon 1 to exon 2b was amplified from RNA from 20 different human tissues (*upper panel*) and, for comparison, a cDNA fragment stretching from exon 1 to exon 5 of the dystrophin transcript (*middle panel*) and a fragment of GAPDH cDNA were also amplified (*lower panel*). A product including exons

1 and 2b was obtained from heart, lung, prostate, salivary gland, and skeletal muscle, as were products corresponding to the dystrophin transcript from exon 1 to exon 5 (*middle panel*) and to the GAPDH transcript (*lower panel*). *Mk* represents a size marker (*HincII*-digested f X174 phage DNA; Toyobo Co., Osaka, Japan)

indicating a physiological role for dystrophin in brain function (Bardoni et al. 2000; Felisari et al. 2000; Giliberto et al. 2004). The functional diversity of the dystrophin gene is now becoming apparent, but the role of its unusually huge introns is still unknown. The identification of exons within the introns may shed light on the diverse functions of dystrophin. In this report, the novel cryptic exon 2b was identified in the 3' region of intron 2 (Fig. 3); exon 2b maintains all of the characteristic sequences necessary for exon recognition and is incorporated into dystrophin mRNA (Figs. 1 and 3).

This is the second cryptic exon discovered within the 170-kb long intron 2, but nearly 170 kb still lack any described function. Although exon 2b has a structure similar to the real exon, exon 2b had not been previously described. This may be due to its low Shapiro splicing probability score (Fig. 3) or its tissue-specific incorporation (Fig. 5). Exon 2b is the fifth example of a cryptic exon embedded in an intron of the dystrophin gene. The first example is exon 1a in intron 1 (Roberts et al. 1993), which is incorporated into nearly half of dystrophin mRNAs in lymphocytes, as demonstrated here (Fig. 2). Other cryptic exons have been identified in introns 11, 2, and 3 (Ferlini and Muntoni 1998; Dwi Pramono et al. 2000; Suminaga et al. 2002). It is possible that additional examples of cryptic exons will be uncovered within the unusually huge introns of the dystrophin gene.

In lymphocytes, exon 2b-containing transcripts were identified only in the index case. However, neither a mutation in the genomic sequence near exon 2b nor any gross genomic structural change specific to the index case was identified. The incorporation of exon 2b was limited to a trace amount of the dystrophin transcript (Fig. 2). The exon 2b incorporation was accompanied by the incorporation of another cryptic exon—1a. The concomitant incorporation of exon 1a and exon 2b suggests a common regulatory system for the two cryptic exons, but exon 1a incorporation is not always accompanied by exon 2b incorporation (Fig. 2), indicating that exon 2b incorporation is regulated by a different mechanism from that of exon 1a.

Among four alternative promoters at the 5' end of the dystrophin gene, transcripts containing exon 2b were initiated only at the muscle-specific promoter, indicating that exon 2b incorporation is promoter specific (Fig. 4). However, exon 2b incorporation was detected in mRNAs from only five of the 20 tissues in which the muscle-specific promoter-driven transcript could be detected (Fig. 5). These findings indicate that exon 2b incorporation is under the control of the muscle-specific promoter, but that this is not sufficient for exon 2b incorporation, which requires another tissue-specific factor. This complex pattern of regulation of exon 2b incorporation strongly suggests a physiological role for this cryptic exon.

There are many sequences that match splicing consensus sequences as well as or better than the sequences at real splice sites, yet they are not used for splicing (Krawczak et al. 1992). Real exons are recognized and

spliced cotranscriptionally (Wuarin and Schibler 1994). There must be additional signals that distinguish real splice sites from pseudo sites or vice versa. These additional recognition elements could act either positively or negatively. In one study, authentic splice sites were found to have significantly higher scores than cryptic sites (Roca et al. 2003), but another study found that negative elements play important roles in distinguishing a real splicing signal from the vast number of false splicing signals (Sun and Chasin 2000). Even though Shapiro's probability scores for 5' acceptor and 3' donor splice sites in exon 2b are high, and a perfect branch point sequence is present at the proper position, exon 2b is not a constitutive exon but a cryptic exon. It has been reported that different regulatory programs for splicing run concurrently within the same cell, suggesting that the production of different alternatively spliced pre-mRNAs is regulated by distinct programs that use different sets of cis elements and trans-acting factors (Cooper and Mattox 1997). Incorporation of exon 2b might therefore be regulated in a very specific manner by a number of factors.

Cryptic exons have been shown to be activated by intron mutations that either create or strengthen splice sites or create a branch site (Highsmith et al. 1994; Chillon et al. 1995; Wang et al. 1997; Vervoort et al. 1998; Ars et al. 2000). In addition, an intracryptic exon deletion has been shown to cause erroneous splicing (Pagani et al. 2002; Eng et al. 2004). These observations suggest that cryptic exons are targets for human genetic diseases. Since the typical sequence characteristics of exons have been maintained in exon 2b, we suggest that a sequence acting as a splicing silencer inhibits the incorporation of exon 2b into mRNA. Future experiments may reveal intronic mutations that either disrupt a splicing silencer or activate a splicing enhancer to cause exon 2b incorporation.

Acknowledgements We acknowledge Ms. A. Hosoda for her secretarial help. This work was supported by grants from the Ministry of Education, Science, Sports, and Culture of Japan and the Health and Labor Science Research (Research on Psychiatric and Neurological Diseases and Mental Health), the Research Grant for Nervous and Mental Disorders from the Ministry of Health, Labor, and Welfare, and the Mitsubishi Foundation.

References

- Adachi K, Takeshima Y, Wada H, Yagi M, Nakamura H, Matsuo M (2003) Heterogenous dystrophin mRNAs produced by a novel splice acceptor site mutation in intermediate dystrophinopathy. *Pediatr Res* 53:125–131
- Ahn AH, Kunkel LM (1993) The structural and functional diversity of dystrophin. *Nat Genet* 3:283–291
- Ars E, Serra E, Garcia J, Krayer H, Gaona A, Lazaro C, Estivill X (2000) Mutations affecting mRNA splicing are the most common molecular defects in patients with neurofibromatosis type 1. *Hum Mol Genet* 9:237–247
- Bardoni A, Felisari G, Sironi M, Comi G, Lai M, Robotti M, Bresolin N (2000) Loss of Dp140 regulatory sequences is associated with cognitive impairment in dystrophinopathies. *Neuromuscul Disord* 10:194–199

- Beggs AH, Koenig M, Boyce FM, Kunkel LM (1990) Detection of 98% of DMD/BMD gene deletions by polymerase chain reaction. *Hum Genet* 86:45–48
- Cartegni L, Wang J, Zhu Z, Zhang MQ, Krainer AR (2003) ESEfinder: a web resource to identify exonic splicing enhancers. *Nucleic Acids Res* 31:3568–3571
- Chamberlain JS, Gibbs RA, Ranier JE, Nguyen PN, Caskey CT (1988) Deletion screening of the Duchenne muscular dystrophy locus via multiplex DNA amplification. *Nucleic Acids Res* 16:11141–11156
- Chelly J, Gilgenkrantz H, Hugnot J, Hamard G, Lambert M, Recan D, Akli S, Cometto M, Kahn A, Kaplan J (1991) Illegitimate transcription. Application to the analysis of truncated transcripts of the dystrophin gene in nonmuscle cultured cells from Duchenne and Becker patients. *J Clin Invest* 88:1161–1166
- Chillon M, Dork T, Casals T, Gimenez J, Fonknechten N, Will K, Ramos D, Nunes V, Estivill X (1995) A novel donor splice site in intron 11 of the CFTR gene, created by mutation 1811+1.6kba → G, produces a new exon: high frequency in Spanish cystic fibrosis chromosomes and association with severe phenotype. *Am J Hum Genet* 56:623–629
- Cooper TA, Mattox W (1997) The regulation of splice-site selection, and its role in human disease. *Am J Hum Genet* 61:259–266
- 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* 4:837–842
- Dwi Pramono ZA, Takeshima Y, Surono A, Ishida T, Matsuo M (2000) A novel cryptic exon in intron 2 of the human dystrophin gene evolved from an intron by acquiring consensus sequences for splicing at different stages of anthropoid evolution. *Biochem Biophys Res Commun* 267:321–328
- Eng L, Coutinho G, Nahas S, Yeo G, Tanouye R, Babaei M, Dork T, Burge C, Gatti RA (2004) Nonclassical splicing mutations in the coding and noncoding regions of the ATM Gene: maximum entropy estimates of splice junction strengths. *Hum Mutat* 23:67–76
- Felisari G, Martinelli Boneschi F, Bardoni A, Sironi M, Comi GP, Robotti M, Turconi AC, Lai M, Corrao G, Bresolin N (2000) Loss of Dp140 dystrophin isoform and intellectual impairment in Duchenne dystrophy. *Neurology* 55:559–564
- Ferlini A, Muntoni F (1998) The 5' region of intron 11 of the dystrophin gene contains target sequences for mobile elements and three overlapping ORFs. *Biochem Biophys Res Commun* 242:401–406
- Ferlini A, Sewry C, Melis MA, Mateddu A, Muntoni F (1999) X-linked dilated cardiomyopathy and the dystrophin gene. *Neuromuscul Disord* 9:339–346
- Galante PA, Sakabe NJ, Kirschbaum-Slager N, de Souza SJ (2004) Detection and evaluation of intron retention events in the human transcriptome. *Rna* 10:757–765
- Giliberto F, Ferreira V, Dalamon V, Szijan I (2004) Dystrophin deletions and cognitive impairment in Duchenne/Becker muscular dystrophy. *Neurol Res* 26:83–87
- Graveley BR (2005) Small molecule control of pre-mRNA splicing. *Rna* 11:355–358
- Hagiwara Y, Nishio H, Kitoh Y, Takeshima Y, Narita N, Wada H, Yokoyama M, Nakamura H, Matsuo M (1994) A novel point mutation (G⁻¹ to T) in a 5' splice donor site of intron 13 of the dystrophin gene results in exon skipping and is responsible for Becker muscular dystrophy. *Am J Hum Genet* 54:53–61
- Highsmith WE, Burch LH, Zhou Z, Olsen JC, Boat TE, Spock A, Gorvoy JD, Quittel L, Friedman KJ, Silverman LM et al (1994) A novel mutation in the cystic fibrosis gene in patients with pulmonary disease but normal sweat chloride concentrations. *N Engl J Med* 331:974–980
- Krawczak M, Reiss J, Cooper DN (1992) The mutational spectrum of single base-pair substitutions in mRNA splice junctions of human genes: causes and consequences. *Hum Genet* 90:41–54
- Malhotra SB, Hart KA, Klamut HJ, Thomas NST, Bodrug E, Burghes AHM, Borrow M, Harper PS, Thompson W, Ray PN, Worton RG (1988) Frame-shift deletions in patients with Duchenne and Becker muscular dystrophy. *Science* 242:755–759
- Matsuo M, Masumura T, Nishio H, Nakajima T, Kitoh Y, Takumi T, Koga J, Nakamura H (1991) Exon skipping during splicing of dystrophin mRNA precursor due to an intraexon deletion in the dystrophin gene of Duchenne muscular dystrophy Kobe. *J Clin Invest* 87:2127–2131
- Muntoni F, Gobbi P, Sewry C, Sherratt T, Taylor J, Sandhu SK, Abbs S, Roberts R, Hodgson SV, Bobrow M, Dubowitz V (1994) Deletions in the 5' region of dystrophin and resulting phenotypes. *J Med Genet* 31:843–847
- Nishio H, Takeshima Y, Narita N, Yanagawa H, Suzuki Y, Ishikawa Y, Minami R, Nakamura H, Matsuo M (1994) Identification of a novel first exon in the human dystrophin gene and of a new promoter located more than 500 kb upstream of the nearest known promoter. *J Clin Invest* 94:1037–1042
- Pagani F, Buratti E, Stuardi C, Bendix R, Dork T, Baralle FE (2002) A new type of mutation causes a splicing defect in ATM. *Nature Genet* 30:426–429
- Reiss J, Rininsland F (1994) An explanation for the constitutive exon 9 cassette splicing of the DMD gene. *Hum Mol Genet* 3:295–298
- Roberts RG, Bentley DR, Bobrow M (1993) Infidelity in the structure of ectopic transcripts: a novel exon in lymphocyte dystrophin transcripts. *Hum Mutat* 2:293–299
- Roca X, Sachidanandam R, Krainer AR (2003) Intrinsic differences between authentic and cryptic 5' splice sites. *Nucleic Acids Res* 31:6321–6333
- Schaal TD, Maniatis T (1999) Selection and characterization of pre-mRNA splicing enhancers: identification of novel SR protein-specific enhancer sequences. *Mol Cell Biol* 19:1705–1719
- Senapathy P, Shapiro MB, Harris NL (1990) Splice junctions, branch point sites, and exons: sequence statistics, identification, and applications to genome project. *Methods Enzymol* 183:252–278
- Shapiro MB, Senapathy P (1987) RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucleic Acids Res* 15:7155–7174
- Sironi M, Menozzi G, Riva L, Cagliani R, Comi GP, Bresolin N, Giorda R, Pozzoli U (2004) Silencer elements as possible inhibitors of pseudoexon splicing. *Nucleic Acids Res* 32:1783–1791
- Suminaga R, Takeshima Y, Adachi K, Yagi M, Nakamura H, Matsuo M (2002) A novel cryptic exon in intron 3 of the dystrophin gene was incorporated into dystrophin mRNA with a single nucleotide deletion in exon 5. *J Hum Genet* 47:196–201
- Sun H, Chasin LA (2000) Multiple splicing defects in an intronic false exon. *Mol Cell Biol* 20:6414–6425
- Surono A, Takeshima Y, Wibawa T, Pramono ZA, Matsuo M (1997) Six novel transcripts that remove a huge intron ranging from 250 to 800 kb are produced by alternative splicing of the 5' region of the dystrophin gene in human skeletal muscle. *Biochem Biophys Res Commun* 239:895–899
- Surono A, Takeshima Y, Wibawa T, Ikezawa M, Nonaka I, Matsuo M (1999) Circular dystrophin RNAs consisting of exons that were skipped by alternative splicing. *Hum Mol Genet* 8:493–500
- Torelli S, Muntoni F (1996) Alternative splicing of dystrophin exon 4 in normal human muscle. *Hum Genet* 97:521–523
- 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* 21:608–614
- Vervoort R, Gitzelmann R, Lissens W, Liebaers I (1998) A mutation (IVS8+0.6kdelTC) creating a new donor splice site activates a cryptic exon in an Alu-element in intron 8 of the human beta-glucuronidase gene. *Hum Genet* 103:686–693

- Wang M, Dotzlaw H, Fuqua SAW, Murphy LC (1997) A point mutation in the human estrogen receptor gene is associated with the expression of an abnormal estrogen receptor mRNA containing a 69 novel nucleotide insertion. *Breast Cancer Res Treat* 44:145–151
- Winnard AV, Jia-Hsu Y, Gibbs RA, Mendell JR, Burghes AH (1992) Identification of a 2 base pair nonsense mutation causing a cryptic splice site in a DMD patient. *Hum Mol Genet* 1:645–646
- Winnard AV, Mendell JR, Prior TW, Florence J, Burghes HM (1995) Frameshift deletions of exons 3–7 and revertant fibers in Duchenne muscular dystrophy: mechanisms of dystrophin production. *Am J Hum Genet* 56:158–166
- Wuarin J, Schibler U (1994) Physical isolation of nascent RNA chains transcribed by RNA polymerase II: evidence for co-transcriptional splicing. *Mol Cell Biol* 14:7219–7225
- Yagi M, Takeshima Y, Wada H, Nakamura H, Matsuo M (2003) Two alternative exons can result from activation of the cryptic splice acceptor site deep within intron 2 of the dystrophin gene in a patient with as yet asymptomatic dystrophinopathy. *Hum Genet* 112:164–170