Kobe J. Med. Sci., Vol. 52, No. 3, pp. 61-75, 2006

Novel Cryptic Exons Identified in Introns 2 and 3 of the Human Dystrophin Gene with Duplication of Exons 8-11

KAZUTO ISHIBASHI, YASUHIRO TAKESHIMA, MARIKO YAGI, ATSUSHI NISHIYAMA, and MASAFUMI MATSUO

Division of Pediatrics, Department of Development and Aging, Kobe University Graduate School of Medicine

Received 6 February 2006 /Accepted 16 February 2006

Key words: dystrophin, cryptic exon, splicing, Duchenne muscular dystrophy

The dystrophin gene, which is mutated in Duchenne muscular dystrophy, is the largest known human gene and characterized by the huge size of its introns. Intron 2 has been shown to include cryptic exons termed exons 2a and 2b, while intron 3 has been shown to include a cryptic exon designated exon 3a. In the present study, we identified 2 and 1 additional cryptic exons in introns 2 and 3, respectively. A previously unknown 157-bp insertion was identified between exons 2 and 3 of a dystrophin mRNA isolated from the lymphocytes of a dystrophinopathy patient with duplication of exons 8-11. Since this sequence exhibited the typical characteristics of a genomic exon, we designated it "exon 2c-l". A more detailed examination revealed that a position 4 bp downstream from the 5' end of exon 2c-l was also used as a splice acceptor site, and this exon was designated "exon 2c-s". In the same patient, a 357-bp insertion was identified between exons 3 and 4. Since this sequence also showed the typical characteristics of an exon, and its 3' end was the same as the splice donor site of exon 3a, we designated the novel cryptic exon "exon 3a-l", and changed the name of the previously reported exon 3a to "exon 3a-s". Among these novel cryptic exons, exon 3a-l was also incorporated into the dystrophin mRNA from normal lymphocytes, whereas exons 2c-l and 2c-s were not. The physiological or pathophysiological roles of these novel cryptic exons remain to be clarified.

The human dystrophin gene, which is defective in patients with Duchenne or Becker muscular dystrophy (DMD/BMD), encodes a dystrophin protein that is located along the plasma membrane of muscle cells (3). It spans approximately 3,000 kb of the X-chromosome and encodes a 14-kb transcript consisting of 79 exons (2,21). Genomic structural analyses have revealed at least eight alternative promoters over the entire dystrophin gene, producing tissue-specific dystrophin isoforms (2,21). Consequently, more than 99% of the gene sequence is comprised of introns which have been considered to be functionless. To date, four cryptic exons (exon 1a from intron 1, exons 2a and 2b from intron 2, and exon 3a from intron 3) have been reported to be incorporated into dystrophin mRNA (7,23,30,36). Furthermore, we previously identified six novel alternative splicing patterns in the 5' region of the dystrophin gene in addition to the six already-known patterns (33), indicating that exon usage in the 5' region of the dystrophin gene is highly complicated.

Splicing is the process that removes introns from pre-mRNAs, thereby producing mature mRNAs that only consist of exons. The presence of well-defined cis elements, namely the 5' and 3' splice sites and a branch point, are necessary but not sufficient to define the intron-exon boundaries in pre-mRNAs (26). Pseudoexons that match the splice-site

Phone: +81-78-382-6090 Fax: +81-78-382-6099 E-mail: takesima@med.kobe-u.ac.jp

consensus sequences have been identified in introns, but their inclusion in mRNA is prevented by silencer elements (29). However, unconventional splicing defects often occur at exons with weak homology to canonical splicing sequences, leading to dystrophinopathies (37,41).

The differences between DMD and BMD can be explained by the frameshift theory. Genomic mutations that disrupt the translational open reading frame lead to a truncated nonfunctional dystrophin protein that results in DMD, while mutations that maintain the open reading frame can create a partly functional protein that results in the milder BMD (19). However, many dystrophinopathy cases with deletions in the 5' region of the dystrophin gene have been revealed to be exceptions to this rule, and alternative splicing is considered to be a factor leading to such exceptions by changing the translational open reading frame (20).

We have analyzed the dystrophin mRNAs expressed in peripheral lymphocytes from more than 100 dystrophinopathy cases. Here, we identified unknown sequences inserted into a dystrophin transcript in a case with duplication of exons 8-11 of the dystrophin gene, and found that they represent novel cryptic exons (designated exons 2c-l, 2c-s and 3a-l). Exon 2c-l is located downstream from exon 2b in intron 2. Exon 2c-s is located in the same region as exon 2c-l, although its splice acceptor site is 4 bp downstream from that of exon 2c-l. Exon 3a-l is located in intron 3, and has the same splice donor site as exon 3a and a splice acceptor site 295 bp upstream from that of exon 3a. Hence, we changed the name of the previously reported exon 3a to "exon 3a-s". The cryptic exon 3a-l is incorporated into the mRNA of normal lymphocytes, whereas exons 2c-l and 2c-s are not.

PATIENT AND METHODS

Case

A 2-year-old Japanese boy was referred to Kobe University Hospital for genetic diagnosis of his dystrophinopathy. He was the first-born boy, and his family history revealed no muscular diseases. During an examination for meconium aspiration syndrome, he was found to have an extremely high level of serum CK at birth, which continued as he grew older (1,887-84,060 IU/L; normal: 56-248 IU/L). He showed Gowers' sign, but no walk disturbance at 3 years of age. DQ was 100, and chest X-ray and ECG were normal. All analyses were carried out after informed consent was obtained from his parents.

Analysis of genomic DNA

Genomic DNA was isolated from lymphocytes obtained from the index patient using a previously described method (18). To examine the entire dystrophin gene, the patient's genomic DNA was analyzed by Southern blotting using a dystrophin cDNA as a probe and *Hind*III restriction enzyme-digested DNA as a template (Mitsubishi Co., Tokyo, Japan), as described previously (14). The genomic regions encompassing the 157-bp inserted sequence (exon 2c-l) and 357-bp inserted sequence (exon 3a-l) were amplified using primers derived from the flanking sequences (g2cF and g2cR for exon 2c-l; g3a-lF and g3a-lR for exon 3a-l; Table). PCR was performed essentially as described previously (18).

Analysis of dystrophin transcripts

Total RNA was isolated from peripheral lymphocytes as previously described (18). cDNA fragments encompassing exons 1-5 and exons 10-12 of the dystrophin mRNA were analyzed by reverse-transcription (RT)-nested PCR. For exons 1-5, the first PCR amplified the region comprising exons 1-8 using primers 1A and c8R (Table), and the second PCR amplified a fragment comprising exons 1-5 using primers 1C and c5R, as described

previously (Table) (36). For exons 10-12, a fragment was amplified using primers c7F and c12R for the first PCR, and then primers 1E and c12R were used for the second PCR (Table).

To examine the efficiency of exon 2c-l, 2c-s and 3a-l activation in normal lymphocytes and muscle tissues, cDNA fragments spanning exons 1-2c and exons 1-3a were amplified. The amplifications were carried out by semi-nested PCR for lymphocyte cDNA and single PCR for muscle cDNA. For the semi-nested PCR amplification of exons 1-2c, primers 1A and c2cR were used for the first PCR, and primers 1C and c2cR were used for the second PCR. For the semi-nested PCR amplification of exons 1-3a-l, primers 1A and c3a-lR were used for the first PCR, and primers 1C and c3a-1R were used for the second PCR. The primer sequences are shown in the Table and the PCR conditions were described previously (36).

DNA sequencing

For DNA sequencing, the amplified products were separated by electrophoresis in low-melting-point agarose gels. The bands of the amplified products were excised from the gels, and the DNA was purified. The purified DNA was subcloned into the pT7 vector (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). In some experiments, the nucleotide sequences were determined by a direct sequencing method using the same automated DNA sequencer.

Table. Primer sequences

Target region	Forward primer	Reverse primer
gDNA		
exon 2c	g2cF: AGCATAGCAGTGACAGACGT	g2cR: CATCTATGAGTACTCAGTTC
exon 3a-l	g3a-lF: CATATGCACATGTGTACATT	g3a-lR: GCTAAAGCCATCTTACTCTC
mRNA		
exon 1	1A: TTTTTATCGCTGCCTTGATATACA 1C: ATGCTTTGGTGGGAAGAAGTAG	
exon 2c		c2cR: CAATTACCCTACGGTGCAGT
exon 3a-l		c3a-lR: TTTTCCTCCAAGAATGTCAG
exon 5		c5R: TGCCAGTGGAGGATTATATTCCAA
exon 7	c7F: GTGGTTTGCCAGCAGTCAGCCACA	
exon 8		c8R: TGTTGAGAATAGTGCATTTGATG
exon 10	1E: TTGCAAGCACAAGGAGAGATT	
exon 12		c12R: CAAGAGGCTCTTCCTCCATTTTCC

RESULTS

Detection of the responsible dystrophin mutation in the index case

The Southern blot analysis revealed abnormally high densities of bands corresponding to exons 4 and 10-11, indicating duplication of these exons (Fig. 1a). This duplication of two noncontiguous regions of the dystrophin gene represents a very rare kind of mutation, although DMD patient with duplication of exons 45-48 in combination with duplication of exons 54-55 has been reported (13). In order to examine these duplications at the mRNA level, the dystrophin mRNA expressed in peripheral lymphocytes obtained from the patient was analyzed. When the region encompassing exons 10-12 was amplified by RT-nested PCR, a single band corresponding to a cDNA fragment encompassing exons 10-12 was amplified in normal lymphocytes, whereas a major large band and two barely visible weak bands were obtained in lymphocytes from the patient (Fig. 1b). Each of the bands was sequenced after subcloning. Sequencing of the smallest fragment produced the sequence of exons 10-12. On the other hand, sequencing of the largest fragment revealed that it was composed of exons 10, 11, 8, 9, 10, 11 and 12 (Fig. 1c), while sequencing of the middle-sized band revealed skipping of exon 9 compared to the largest product, which is usually observed due to alternative splicing in lymphocyte mRNA (22). These results demonstrate that the 3' end of exon 11 was directly joined to the 5' end of exon 8 in the mRNA, indicating duplication of exons 8-11. In the Southern blot analysis, the *Hind*III fragment corresponding to exons 8-9 was 7.2 kb in the normal mRNA. However, the genomic rearrangement altered the length of the *Hind*III fragment corresponding to exons 8-9, which has been identified as a junction fragment, to 8.4 kb and this aberrant *Hind*III fragment comigrated with the exon 4 fragment (8.4 kb). The junction fragment therefore leads to misinterpretation of the Southern blot analysis as a double duplication, and the true mutation of the index case is duplication of exons 8-11.

Detection of novel cryptic exons

To exclude the possible duplication of exon 4 completely, the region encompassing exons 1-5 was amplified by RT-nested PCR using lymphocyte mRNA as a template. Two major bands and one additional faint band were observed in the control mRNA (Fig. 2a). Sequencing of these bands revealed that the two major bands corresponded to the normal cDNA fragments encompassing exons 1-5 and the exon 1a-inserted product, and that the faint bands was non-specific products. In the patient's lymphocytes, the same three bands observed in the control were present (Fig. 2a). To determine the nucleotide sequences, the PCR-amplified products were subcloned into pT7 vector, and four kinds of product were subcloned. Sequencing of the small-sized clone revealed the normal sequence encompassing exons 1-5, corresponding to the small band. The middle-sized clone contained the same exons as the small-sized clone, but, remarkably, a 157-bp sequence was found to be inserted between the tandem exons 2 and 3 (Fig. 3a). Since the PCR product with the 157-bp insertion was almost the same size as the exon 1a-inserted product (162 bp), which was also detected in the patient, the product containing this inserted sequence comigrated with the exon 1a-inserted product and could not be separated by gel electrophoresis. Furthermore, the sequence of the large-sized clone revealed the same exons as the small-sized clone, but an unidentified 357-bp sequence was found to be precisely inserted between exons 3 and 4 (Fig. 3c), corresponding to the large band in the gel electrophoresis.

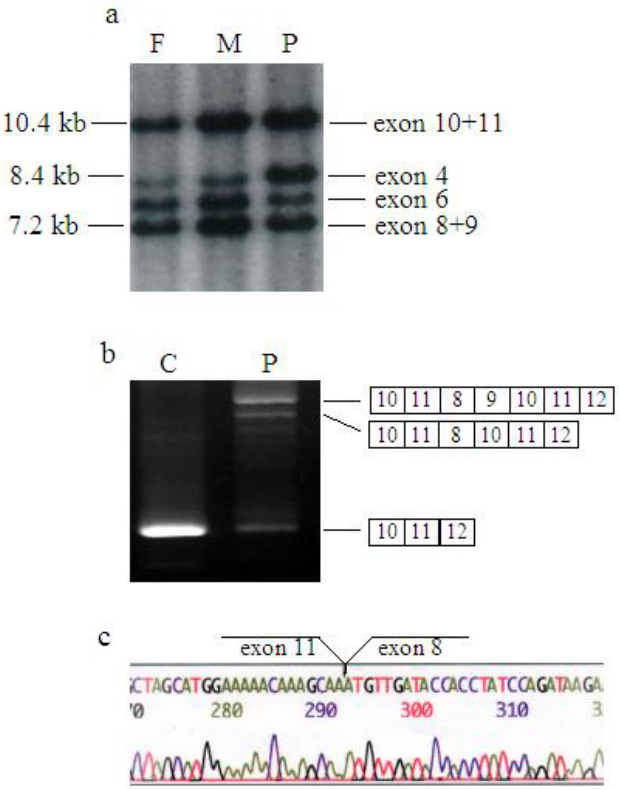

Fig. 1. Analysis of the dystrophin mutation in the index case. (a) Southern blot analysis of the patient (lane P), his father (F) and his mother (M). In lane P, the densities of the 10.4-kb and 8.4-kb bands, corresponding to exons 10-11 and 4, respectively, are high compared with those in lane F (the dystrophin gene of the father should be normal). (b) RT-PCR amplification of a fragment comprising exons 10-12 from lymphocyte mRNA. A single amplified product is obtained from the control mRNA (C), whereas three PCR products are observed for the index case. A schematic representation of the exon organization in the amplified products is shown on the right of the products. (c) Partial sequence of the largest-sized product. The 3' terminal sequence of exon 11 is precisely joined to the 5' end of the sequence of exon 8.

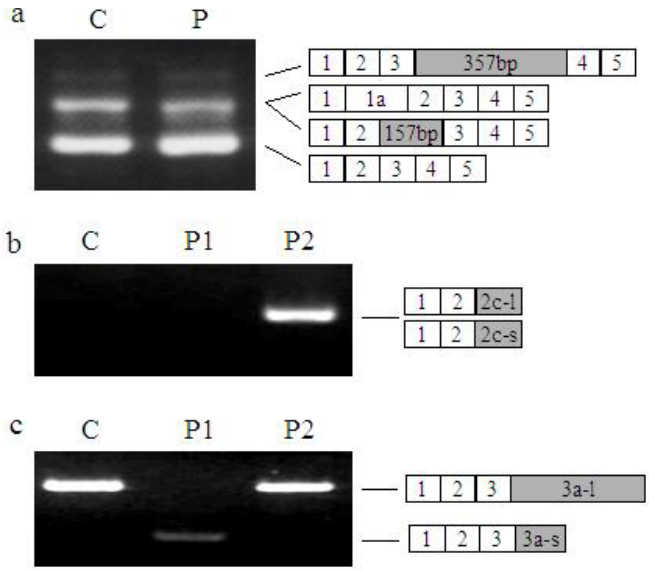

Fig. 2. RT-PCR amplifications of lymphocyte mRNA containing the cryptic exons. (a) RT-PCR product of a fragment comprising exons 1-5. C and P refer to the control and index case, respectively. A band of the expected size and two additional bands are obtained in both the control and the index case. A schematic representation of the exon organization in the amplified products is shown on the right of the products. In the index case, the middle-sized band consists of two types of cDNA, one including exon 1a and the other including an unknown 157-bp sequence. Furthermore, the largest band of the index case consists of a cDNA that includes an unknown 357-bp sequence. On the other hand, cDNAs including these unknown sequences are not obtained after subcloning in the control. The open and shaded boxes represent authentic and cryptic exons, respectively. (b) RT-PCR products of fragments comprising exons 1-2c-l and 1-2c-s. C refers to the control, while P1 and P2 refer to different aliquots of lymphocyte cDNA in the index case. An amplified product including exons 1, 2 and 2c is observed in lane P2. Subcloning and sequencing of this product revealed two types of sequence, one containing exon 2c-l and the other with the first 4 nucleotides of exon 2c-l deleted (see Fig. 3b). (c) RT-PCR product of a fragment comprising exons 1-3a-l. In the control, a cDNA including exon 3a-l is amplified (lane C). A cDNA including exon 3a-s is amplified in one aliquot of the patient's cDNA (lane P1), while a cDNA including exon 3a-l is amplified in another aliquot of the patient's cDNA (lane P2).

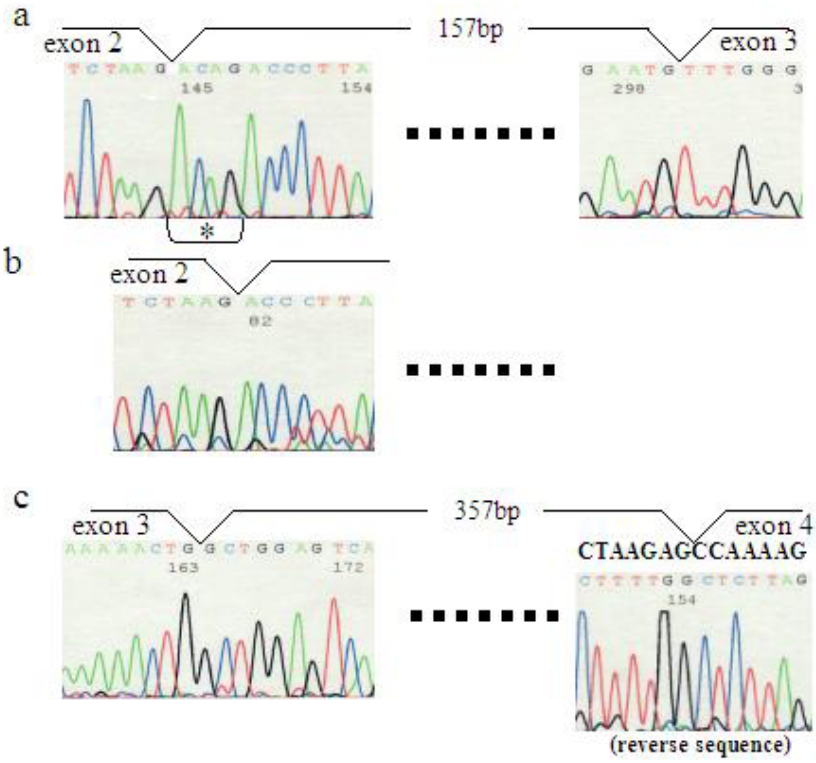

Fig. 3. Partial nucleotide sequences of cDNAs including the cryptic exons.
(a) Nucleotide sequence of the 157-bp exon 2c-l inserted between exons 2 and 3, corresponding to the middle-sized band in Fig. 2a lane P. The 3' terminal sequence of exon 2 is precisely joined to the 5' end of exon 2c-l, while its 3' end is precisely joined to the 5' end of the sequence of exon 3. The asterisk in panel (a) indicates the 4 nucleotides deleted from exon 2c-s in panel (b).

(b) Nucleotide sequence of the cDNA including exon 2c-s, corresponding to the amplified product in Fig. 2b lane P2. The 3' terminal sequence of exon 2 is joined to a site 4 bp downstream from the 5' end of exon 2c-l. Since this product was amplified using primers in exons 1 and 2c, the 3' end sequence of this inserted sequence was not obtained, but the same splice donor site as exon 2c-l is likely to be used.

(c) Nucleotide sequence of the 357-bp exon 3a-l inserted between exons 3 and 4, corresponding to the large-sized band in Fig. 2a lane P. The 3' terminal sequence of exon 3 is precisely joined to the 5' end of exon 3a-l, while its 3' end is precisely joined to the 5' end of the sequence of exon 4. The sequence obtained for the junction between exons 3a-l and 4 is shown in the reverse direction, and the corresponding forward sequence is described above.

Characterization of the novel cryptic exons

The fact that the 157-bp and 357-bp sequences were precisely inserted between exons 2 and 3, and exons 3 and 4, respectively, led us to speculate that these sequences could be retained introns or unknown exons. A BLAST search of the 157-bp sequence revealed an identical sequence in the 3' region of intron 2 (bases 197992-197836 of GenBank AL096699). The 157-bp sequence was located 156 kb downstream from exon 2 and 14 kb upstream of exon 3 (Fig. 4a). Remarkably, the AG and GT dinucleotides that are absolutely conserved at the splice acceptor and donor sites, respectively, of all introns were identified immediately adjacent to the 5' and 3' ends of the 157-bp sequence (Fig. 5a). The Shapiro's probability scores for the splice acceptor and donor sites were both 0.72 (27) (Fig. 4a). Furthermore, the sequence TTTTGAC, a perfect match for a branch-point consensus sequence (YNYTRAC, where Y = C or T, R = G or A and N = any base), was identified 86 bp upstream of the novel sequence (Fig. 5a, inverted triangle). A polypyrimidine tract was also identified between the putative branch point and the splice acceptor site (Fig. 5a), although its length was relatively short at 6 bp. Splicing enhancer sequences are critical for the proper incorporation of exons into mRNAs (25), and have been identified in many exons including the dystrophin gene (28,34,35). The possible presence of a splicing enhancer sequence within the 157-bp sequence was examined using ESE Finder (4), and the analysis revealed 15 possible binding sites for splice factors (data not shown). Since the 157-bp sequence exhibited all the typical characteristics of a genomic exon and was inserted between authentic dystrophin exons, we designated it exon 2c-l.

Since the exon 2c-l insertion was postulated to be due to a genomic mutation, the intron region covering the inserted 157-bp sequence and its flanking regions was sequenced. In total, a 435 bp sequence of the genomic DNA was PCR-amplified, and found to be completely normal (data not shown). Therefore, no genomic mutation contributing to the activation of exon 2c-l was found.

A BLAST search of the 357-bp sequence revealed an identical sequence in intron 3 (bases 181055-180699 of GenBank AL096699). The 357-bp sequence was located 2750 bp downstream from exon 3 and 1762 bp upstream of exon 4 (Fig. 4b). AG and GT dinucleotides were identified immediately adjacent to the 5' and 3' ends of the 357-bp sequence, respectively (Fig. 5b). Interestingly, the 3' end of the 357-bp sequence was the same as that of exon 3a (30), whereas the 5' end was 295 bp upstream from that of exon 3a (Fig. 4b). The Shapiro's probability scores for the splice acceptor and donor sites were 0.88 and 0.80, respectively (27) (Fig. 4b). Although no putative branch point sequence that was a perfect match for the consensus sequence was found within the 150-bp region from the 5' end of the 357-bp sequence, the sequence TAATTAC, with two nucleotide mismatches to the branch point consensus sequence, was identified 113 bp upstream of the novel sequence (Fig. 5b, inverted triangle). A polypyrimidine tract was also identified between the putative branch point and the splice acceptor site (Fig. 5b). The possible presence of a splicing enhancer sequence within the 357-bp sequence was examined using ESE Finder (4), and the analysis revealed 40 possible binding sites for splice factors (data not shown). Since the 357-bp sequence exhibited the typical characteristics of a genomic exon and used the same splice donor site as exon 3a, we designated it exon 3a-l, and changed the name of the previously reported exon 3a to exon 3a-s.

Since we postulated that a genomic mutation located near exon 3a-l activated the incorporation of exon 3a-l in the index case, we examined the genomic sequence encompassing exon 3a-l. In total, a 765 bp sequence of the genomic DNA was PCR-amplified, and found to be completely normal (data not shown). Therefore, no genomic mutations contributing to the activation of exon 3a-l were found.

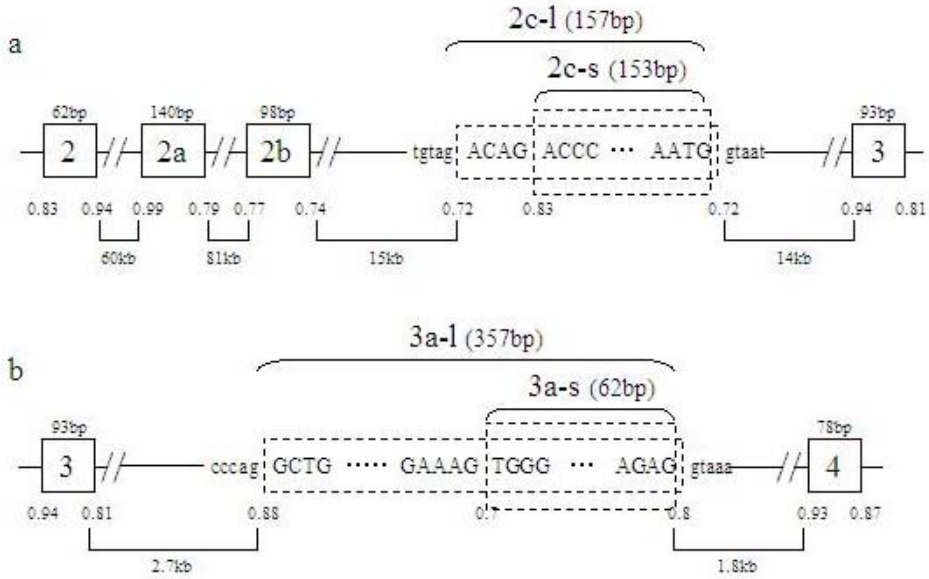

Fig. 4. Genomic structures of the cryptic exons.
(a) Schematic representation of intron 2 of the dystrophin gene.
Intron 2 is the second largest intron at 170 kb. One cryptic exon, exon 2a, is located 60 kb downstream from the 3' end of exon 2, while the other cryptic exon, exon 2b, is located a further 81 kb downstream from exon 2a. The 157 bp unknown sequence is identical to an intron 2 sequence (bases 197992-197836 of GenBank AL096699) located 15 kb downstream from exon 2b and 14 kb upstream of exon 3. Remarkably, two nucleotides present upstream and downstream of the sequence are the AG and GT dinucleotides that are conserved at splicing acceptor and donor sites, respectively. The Shapiro's probability scores for the splice donor and acceptor sites are both 0.72. Remarkably, a site 4 bp downstream from exon 2c is also used as a splice acceptor site, and its Shapiro's score is 0.83. The former and latter cryptic exons are designated exons 2c-l and 2c-s, respectively. The boxes and lines indicate exons and introns, respectively. The dotted boxes indicate the novel cryptic exons, while the brackets indicate the exon sizes and the numbers under the boxes show the Shapiro's splicing probability scores at the splicing sites.

(b) Schematic representation of intron 3 of the dystrophin gene.
One cryptic exon, exon 3a-s, located 3045 bp downstream from the 3' end of exon 3 has already been reported. The 357-bp unknown sequence is identical to an intron 3 sequence (bases 181055-180699 of GenBank AL096699) located 2.7 kb downstream from exon 3 and 1.8 kb upstream of exon 4. Two nucleotides present upstream and downstream of the sequence are the AG and GT dinucleotides that are conserved at splicing acceptor and donor sites, respectively. The Shapiro's probability scores for the splice donor and acceptor sites are 0.88 and 0.80, respectively. Remarkably, the 3' end of the 357-bp inserted sequence is the same as that of exon 3a-s, whereas its 5' end is 295 bp upstream of that of exon 3a-s. Therefore, the 357-bp sequence has been designated exon 3a-l. The dotted boxes indicate cryptic exons 3a-s and 3a-l.

a

```
agcatagca  gtgacagac  gttctcctt  taaatattt  tatcctctg  tgacttttg  acagaagat
cttctttaa  atattgtat  cctatgtga  catttatct  tttctgtgt  ttccaattt  ctttgattt
ggaacttcc  tgtagACAG  ACCCTTACA  GGCATGGAA  GAAGAAAGA  ATGAATAAA  CCAAGGATG
           (tgtagacag ACCCTTACA)
ACTTCCCAG  TAGGTGGAG  GATGGGAAT  AATTAGGAA  GAAGCTGTG  TCTTGTCAC  CTATATTGT
CCATACACA  ACTGCACCG  TAGGGTAAT  TGAGAAATT  ATAAGAATG  gtaataaat  acgttttac
attcattaa  ttaaagtca  tctttggat  gccctttgt  ccataatgt  ttcctctac  tagatctga
agcctaaat  atcactttt  cttctctgc  aatttccct  ggaactgag  tactcatag  atg
```

b

```
catatgcac  atgtgtaca  tttaatgtt  tgaatatga  tacaaagga  aaattttgc  taaatcttg
tgactatat  aattacata  attgaaagg  aaatatgtg  gctttggtc  taactacta  ggcagagaa
ctgatttga  gtaattatt  ttcaaaact  ttaaagtcc  tataattac  cattaaact  tgtatactg
tcaatctat  atatacaca  agaagattg  gttatttgc  tgtttttct  ttcatgtat  tttttttt
attttgatg  gaatcttac  tctgtcgcc  cagGCTGGA  GTCAGCGAC  ATGATCTTG  GCTCACTGC
AACCTCTGC  CTCCTGGGT  TCAAGCAAT  TTTCCTGCC  TCGGCCTCC  CAAGTAGCT  GAGACTACA
GGCATGTTC  CATCATGCC  TGGCTAAGT  TTTGTATTT  TTAGTAGAC  ATGGGGTTT  CACCATGTT
GGCCAGGCT  GGTCTTGAG  CTCCTGATC  TCAATGAGC  CACCTGCCT  CAGCCTCCC  CAAATGTCT
GGCTGTTGC  TTTTTTTTT  TTTTTTTTT  TTTTTTAAA  TCTAAAGTC  TTATTTTTT  TCCTCTTTT
TGGTGGAAA  GTGGGAGAA  ATATCAGAA  TGTAAACCA  ACATCATTC  TGACATTCT  TGGAGGAAA
ATCTAAGAG  gtaaaaatg  gagattctg  gttcataag  gttagcctt  aatgttttg  tgaaatcaa
aggaacaag  atgtcaata  ctgtgttat  aaaggatat  caggcaagt  taccttgga  gagtaagat
ggctttagc
```

Fig. 5. Genomic sequences of the cryptic exons.

(a) Genomic nucleotide sequences of exons 2c-l and 2c-s, as well as their flanking introns. The 157-bp exon 2c-l sequence is shown in uppercase letters, while the sequences 140-bp upstream and 138-bp downstream of exon 2c-l are shown in lowercase letters. Absolutely conserved AG and GT dinucleotides, shown in italics, are present at the boundaries between exon 2c-l and its flanking regions. The splice acceptor site of exon 2c-s is shown in parentheses. A putative branch point is present 86 bp upstream of exon 2c-l (inverted triangle). A polypyrimidine tract is present between exon 2c-l and the branch point, although its length is relatively short at 6 bp. The horizontal arrows indicate the locations and directions of primers g2cF and g2cR.

(b) Genomic nucleotide sequences of exon 3a-l and its flanking introns. The 357 bp exon 3a-l sequence is shown in uppercase letters, while the sequences 282-bp upstream and 126-bp downstream of exon 3a-l are shown in lowercase letters. Absolutely conserved AG and GT dinucleotides, shown in italics, are present at the boundaries between exon 3a-l and its flanking regions. A putative branch point is present 113 bp upstream of exon 3a-l (inverted triangle). A polypyrimidine tract is present between exon 3a-l and the branch point. The exon 3a-s sequence is underlined. The horizontal arrows indicate the locations and directions of primers g3a-lF and g3a-lR.

Incorporation of the cryptic exons in normal tissues

Although we have analyzed dystrophin mRNAs in peripheral lymphocytes obtained from more than 100 dystrophinopathy cases, incorporation of exon 2c-l or 3a-l has never been identified previously (1,11,18,32,41). Therefore, in order to investigate whether activation of these exons is detectable in normal dystrophin mRNA, we analyzed the expressions of these cryptic exons in lymphocytes or muscle tissues from normal individuals using RT-PCR. Since we postulated that the amount of mRNA including these exons would be very low, if present at all, specific primers located in these cryptic exons were used for the PCR amplification (Table). Amplification of a fragment encompassing exons 1-2c revealed a clear band corresponding to exons 1-2c-l in one cDNA aliquot of the patient's lymphocytes, whereas no amplified product was observed in normal lymphocytes (Fig. 2b). Surprisingly, a novel transcript showing deletion of 4 bp of the 5' end of exon 2c-l was detected after subcloning (Fig. 3b). This result indicated that a site 4 bp downstream from the 5' end of exon 2c-l was also used as a splice acceptor site, and this cryptic exon was designated exon 2c-s (Fig. 4a). On the other hand, the amplified product from exons 1-3a-l was detected in normal lymphocyte mRNA (Fig. 2c, lane 1). Furthermore, exons 3a-s and 3a-l were both amplified in the lymphocyte mRNA of the index case (Fig. 2c). Due to the very low amounts of these transcripts, exons 3-s and 3a-l were amplified using different aliquots of lymphocyte cDNA (Fig. 2c, lanes 2 and 3). We concluded that cryptic exon 3a-l was also incorporated into normal lymphocyte mRNA, whereas exons 2c-l and 2c-s were not.

Next, we analyzed muscle mRNA. However, specific products corresponding to the incorporation of these novel exons were not amplified in the normal tissues by single PCR amplification (data not shown). These results indicate that activation of these cryptic exons is very low or absent in muscle.

Protein-coding capacities of the cryptic exons

The protein-coding capacities of the novel dystrophin transcripts retaining exon 2c-l, 2c-s or 3a-l were examined. Exon 2c-l encoding the 157 bp sequence disrupted the open reading frame. Exon 2c-s encoding the 153 bp sequence maintained the open reading frame, but contained many in-frame termination codons. Although exons 2c-l and 2c-s contained an in-frame ATG codon at the 3' end of the exon sequence, the upstream sequence was not compatible with the Kozak sequence (Fig. 5a) (15), which represents the consensus sequence required for the initiation of translation. Therefore, it is unlikely that a transcript containing exons 2c-l or 2c-s would direct the synthesis of a novel protein. Exon 3a-l encoding the 357 bp sequence maintained the translational open reading frame. However, since exon 3a-l contained many in-frame termination codons and does not possess an in-frame ATG codon after the last termination codon, it is also unlikely that exon 3a-l allows the initiation of protein synthesis. However, the transcript would be expected to allow reinitiation of translation at the downstream ATG codon (17) or be of other unknown biological significance (8,10).

DISCUSSION

It is well established that DMD and BMD are caused by mutations in the dystrophin gene, which is one of the largest human genes. More than 99% of the gene sequence is composed of introns, and these huge introns have been considered to be functionless. To date, five cryptic exons have been identified in introns of the dystrophin gene. The large introns 1 and 2 have been shown to contain a cryptic exon 1a and cryptic exons 2a and 2b, respectively (7,23,36). On the other hand, the relatively small intron 3 has also been reported to contain a cryptic exon (30). In the present study, we identified novel cryptic exons 2c-l and 2c-s in the

3' region of intron 2 and novel cryptic exon 3a-l in the central region of intron 3. These exons possess all of the characteristic sequences required for exon recognition and are incorporated into dystrophin mRNA.

Exons 2c-l and 2c-s are the third and fourth cryptic exons discovered within intron 2, which is the second largest intron at 170 kb. Exons 2a and 2b are 140 and 98 bp long and located 60 kb and 141 kb downstream from exon 2, respectively. Neither of the exons contains an in-frame ATG codon after the last termination codon, and it is therefore unlikely that transcripts containing these exons will be translated (7,36). On the other hand, the last three nucleotides of exons 2c-l and 2c-s are ATG. However, since the upstream sequences of these initiation codons are not compatible with the Kozak sequence (15), it is also unlikely that transcripts containing these exons would direct the synthesis of novel proteins.

Exon 3a-l is the second cryptic exon discovered within the 4.9-kb intron 3, which has already been reported to contain exon 3a-s (30). The splice donor site of exon 3a-l is the same as that for exon 3a-s, whereas its acceptor site is 295 bp upstream of that of exon 3a-s. Due to the absence of an in-frame ATG codon after the last termination codon, exon 3a-l is also nonfunctional for producing a protein.

Although exons 2c-l, 2c-s and 3a-l have similar structures to real exons, they have not previously been described. This may be due to their low Shapiro's splicing probability scores or unsuitable branch sites. The scores for exon 2c-l are 0.72 for both the splice acceptor and donor sites, while those of exon 2c-s are 0.83 and 0.72, respectively, and the putative branch point is 86 bp upstream of the splice acceptor site of exon 2c-l, which is too far to act as a suitable branch point. The scores for exon 3a-l are also relatively low (0.88 for the splice acceptor site and 0.80 for the splice donor site) compared with those of the authentic exon 2 (0.83 and 0.94, respectively) and exon 3 (0.94 and 0.81, respectively), and the putative branch point sequence differs from the consensus sequence by two nucleotides. It remains to be clarified why these cryptic exons are activated in the index case with duplication of exons 8-11 in the dystrophin gene. It is likely that the gross genomic rearrangement may modify the splicing environment, although further studies are necessary. The newly identified exons are the sixth, seventh and eighth examples of cryptic exons embedded in introns of the dystrophin gene, and it is possible that additional cryptic exons will be uncovered within the introns of this gene.

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 (16). Real exons are recognized and spliced cotranscriptionally (40). Therefore, 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. For example, one study reported that authentic splice sites had significantly higher scores than cryptic sites (24), while another study found that negative elements play important roles in distinguishing a real splicing signal from the vast number of false splicing signals (31). 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 (6). Incorporation of exons 2c-l, 2c-s and 3a-l may therefore be regulated in a very specific manner by a number of factors.

The RT-PCR analysis using normal lymphocytes revealed that exon 3a-l is also incorporated into normal lymphocyte mRNA, whereas exons 2c-l and 2c-s are not. Furthermore, amplified products corresponding to the incorporation of these three novel exons were not detected after single PCR amplifications using normal and patient muscle tissues as templates. Therefore, activation of these cryptic exons is very low or absent in

muscles. The physiological or pathophysiological roles of these novel cryptic exons remain to be clarified. It has been suggested that large introns may be removed by a process in which smaller sections are first extracted via intermediate splicing events. Recently, stepwise removal of a large intron was demonstrated to occur via resplicing at a junction between certain joined exons (12). Exons 2c-l, 2c-s and 3a-l, which are embedded in the 140-kb intron 2 and 4.9-kb intron 3, respectively, may be remnants of such stepwise removal of introns. Further studies are required to clarify this possibility.

In dystrophinopathy, the differences between DMD and BMD can be explained by the frameshift theory (19), although many dystrophinopathy cases with deletions in the 5' region of the dystrophin gene have been shown to be exceptions to this rule. Furthermore, there is a wide variety of clinical phenotypes with involvement of skeletal and cardiac muscles and the mental status. The possibility that alternative splicing could modify the clinical phenotypes by editing the translational open reading frame has been pointed out (5,38). We have identified six novel alternative splicing patterns in the 5' region of the dystrophin gene in addition to the six already-known patterns (33). Furthermore, it has been proposed that the cryptic start codon in exon 8 is activated and that its translation product modifies the clinical phenotype (9,39). Therefore, incorporation of cryptic exons may have roles in stabilizing the mRNA or altering the splicing pattern.

In conclusion, we have identified three novel cryptic exons in introns 2 and 3 of the dystrophin gene, one of which is also incorporated into the normal lymphocyte transcript. The physiological and pathophysiological roles of these three cryptic exons in addition to those of the five already-known cryptic exons remain to be clarified.

ACKNOWLEDGEMENTS

This work was supported by grants from the Ministry of Education, Science, Sports and Culture of Japan and a Health and Research Grant for Nervous and Mental Disorders from the Ministry of Health of Japan.

REFERENCES

1. **Adachi, K., Y. Takeshima, H. Wada, M. Yagi, H. Nakamura, and M. Matsuo.** 2003. Heterogous dystrophin mRNA produced by a novel splice acceptor site mutation in intermediate dystrophinopathy. Pediatr Res **53:**125-31.
2. **Ahn, A. H., and L. M. Kunkel.** 1993. The structural and functional diversity of dystrophin. Nat Genet **3:**283-91.
3. **Arahata, K., E. P. Hoffman, L. M. Kunkel, S. Ishiura, T. Tsukahara, T. Ishihara, N. Sunohara, I. Nonaka, E. Ozawa, and H. Sugita.** 1989. Dystrophin diagnosis: comparison of dystrophin abnormalities by immunofluorescence and immunoblot analyses. Proc Natl Acad Sci USA **86:**7154-8.
4. **Cartegni, L., J. Wang, Z. Zhu, M. Q. Zhang, and A. R. Krainer.** 2003. ESEfinder: A web resource to identify exonic splicing enhancers. Nucleic Acids Res **31:**3568-71.
5. **Chelly, J., H. Gilgenkrantz, J. P. Hugnot, G. Hamard, M. Lambert, D. Recan, S. Akli, M. Cometto, A. Kahn, and J. C. Kaplan.** 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:**1161-6.
6. **Cooper, T. A., and W. Mattox.** 1997. The regulation of splice-site selection, and its role in human disease. Am J Hum Genet **61:**259-66.
7. **Dwi Pramono, Z. A., Y. Takeshima, A. Surono, T. Ishida, and M. Matsuo.** 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:**321-8.

8. **Galante, P. A., N. J. Sakabe, N. Kirschbaum-Slager, and S. J. de Souza.** 2004. Detection and evaluation of intron retention events in the human transcriptome. Rna **10:**757-65.
9. **Gangopadhyay, S. B., T. G. Sherratt, J. Z. Heckmatt, V. Dubowitz, G. Miller, M. Shokeir, P. N. Ray, P. N. Strong, and R. G. Worton.** 1992. Dystrophin in frameshift deletion patients with Becker muscular dystrophy. Am J Hum Genet **51:**562-70.
10. **Graveley, B. R.** 2005. Small molecule control of pre-mRNA splicing. Rna **11:**355-8.
11. **Hagiwara, Y., H. Nishio, Y. Kitoh, Y. Takeshima, N. Narita, H. Wada, M. Yokoyama, H. Nakamura, and M. Matsuo.** 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.
12. **Hatton, A. R., V. Subramaniam, and A. J. Lopez.** 1998. Generation of alternative Ultrabithorax isoforms and stepwise removal of a large intron by resplicing at exon-exon junctions. Mol Cell **2:**787-96.
13. **Janssen, B., C. Hartmann, V. Scholz, A. Jauch, and J. Zschocke.** 2005. MLPA analysis for the detection of deletions, duplications and complex rearrangements in the dystrophin gene: potential and pitfalls. Neurogenetics **6:**29-35.
14. **Koenig, M., E. P. Hoffman, C. J. Bertelson, A. P. Monaco, C. Feener, and L. M. Kunkel.** 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:**509-17.
15. **Kozak, M.** 2002. Pushing the limits of the scanning mechanism for initiation of translation. Gene **299:**1-34.
16. **Krawczak, M., J. Reiss, and D. N. Cooper.** 1992. The mutational spectrum of single base-pair substitutions in mRNA splice junctions of human genes: causes and consequences. Hum Genet **90:**41-54.
17. **Malhotra, S. B., K. A. Hart, H. J. Klamut, N. S. Thomas, S. E. Bodrug, A. H. Burghes, M. Bobrow, P. S. Harper, M. W. Thompson, P. N. Ray, and et al.** 1988. Frame-shift deletions in patients with Duchenne and Becker muscular dystrophy. Science **242:**755-9.
18. **Matsuo, M., T. Masumura, H. Nishio, T. Nakajima, Y. Kitoh, T. Takumi, J. Koga, and H. Nakamura.** 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:**2127-31.
19. **Monaco, A. P., C. J. Bertelson, S. Liechti-Gallati, H. Moser, and L. M. Kunkel.** 1988. An explanation for the phenotypic differences between patients bearing partial deletions of the DMD locus. Genomics **2:**90-5.
20. **Muntoni, F., P. Gobbi, C. Sewry, T. Sherratt, J. Taylor, S. K. Sandhu, S. Abbs, R. Roberts, S. V. Hodgson, M. Bobrow, and et al.** 1994. Deletions in the 5' region of dystrophin and resulting phenotypes. J Med Genet **31:**843-7.
21. **Nishio, H., Y. Takeshima, N. Narita, H. Yanagawa, Y. Suzuki, Y. Ishikawa, R. Minami, H. Nakamura, and M. Matsuo.** 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:**1037-42.
22. **Reiss, J., and F. Rininsland.** 1994. An explanation for the constitutive exon 9 cassette

splicing of the DMD gene. Hum Mol Genet **3:**295-8.
23. **Roberts, R. G., D. R. Bentley, and M. Bobrow.** 1993. Infidelity in the structure of ectopic transcripts: a novel exon in lymphocyte dystrophin transcripts. Hum Mutat **2:**293-9.
24. **Roca, X., R. Sachidanandam, and A. R. Krainer.** 2003. Intrinsic differences between authentic and cryptic 5' splice sites. Nucleic Acids Res **31:**6321-33.
25. **Schaal, T. D., and T. Maniatis.** 1999. Selection and characterization of pre-mRNA splicing enhancers: identification of novel SR protein-specific enhancer sequences. Mol Cell Biol **19:**1705-19.
26. **Senapathy, P., M. B. Shapiro, and N. L. Harris.** 1990. Splice junctions, branch point sites, and exons: sequence statistics, identification, and applications to genome project. Methods Enzymol **183:**252-78.
27. **Shapiro, M. B., and P. Senapathy.** 1987. RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. Nucleic Acids Res **15:**7155-74.
28. **Shiga, N., Y. Takeshima, H. Sakamoto, K. Inoue, Y. Yokota, M. Yokoyama, and M. Matsuo.** 1997. Disruption of the splicing enhancer sequence within exon 27 of the dystrophin gene by a nonsense mutation induces partial skipping of the exon and is responsible for Becker muscular dystrophy. J Clin Invest **100:**2204-10.
29. **Sironi, M., G. Menozzi, L. Riva, R. Cagliani, G. P. Comi, N. Bresolin, R. Giorda, and U. Pozzoli.** 2004. Silencer elements as possible inhibitors of pseudoexon splicing. Nucleic Acids Res **32:**1783-91.
30. **Suminaga, R., Y. Takeshima, K. Adachi, M. Yagi, H. Nakamura, and M. Matsuo.** 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.
31. **Sun, H., and L. A. Chasin.** 2000. Multiple splicing defects in an intronic false exon. Mol Cell Biol **20:**6414-25.
32. **Surono, A., Y. Takeshima, T. Wibawa, M. Ikezawa, I. Nonaka, and M. Matsuo.** 1999. Circular dystrophin RNAs consisting of exons that were skipped by alternative splicing. Hum Mol Genet **8:**493-500.
33. **Surono, A., Y. Takeshima, T. Wibawa, Z. A. Pramono, and M. Matsuo.** 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:**895-9.
34. **Surono, A., T. Van Khanh, Y. Takeshima, H. Wada, M. Yagi, M. Takagi, M. Koizumi, and M. Matsuo.** 2004. Chimeric RNA/ethylene-bridged nucleic acids promote dystrophin expression in myocytes of duchenne muscular dystrophy by inducing skipping of the nonsense mutation-encoding exon. Hum Gene Ther **15:**749-57.
35. **Takeshima, Y., H. Nishio, H. Sakamoto, H. Nakamura, and M. Matsuo.** 1995. Modulation of in vitro splicing of the upstream intron by modifying an intra-exon sequence which is deleted from the dystrophin gene in dystrophin Kobe. J Clin Invest **95:**515-20.
36. **Tran, V. K., Z. Zhang, M. Yagi, A. Nishiyama, Y. Habara, Y. Takeshima, and M. Matsuo.** 2005. A novel cryptic exon identified in the 3' region of intron 2 of the human dystrophin gene. J Hum Genet **50:**425-33.
37. **Tuffery-Giraud, S., C. Saquet, S. Chambert, and M. Claustres.** 2003. Pseudoexon activation in the DMD gene as a novel mechanism for Becker muscular dystrophy. Hum

Mutat **21:**608-14.

38. **Winnard, A. V., C. J. Klein, D. D. Coovert, T. Prior, A. Papp, P. Snyder, D. E. Bulman, P. N. Ray, P. McAndrew, W. King, and et al.** 1993. Characterization of translational frame exception patients in Duchenne/Becker muscular dystrophy. Hum Mol Genet **2:**737-44.
39. **Winnard, A. V., J. R. Mendell, T. W. Prior, J. Florence, and A. H. Burghes.** 1995. Frameshift deletions of exons 3-7 and revertant fibers in Duchenne muscular dystrophy: mechanisms of dystrophin production. Am J Hum Genet **56:**158-66.
40. **Wuarin, J., and U. Schibler.** 1994. Physical isolation of nascent RNA chains transcribed by RNA polymerase II: evidence for cotranscriptional splicing. Mol Cell Biol **14:**7219-25.
41. **Yagi, M., Y. Takeshima, H. Wada, H. Nakamura, and M. Matsuo.** 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:**164-70.