Protein Transcripts of Dysferlin

Alternate Start Exons

Pramono *et al.* [Hum Genet (2006) 120:410–419] identified an alternate human dysferlin isoform, designated DYSF_v1 (accession number DQ267935), of nearly the same size as the previously characterized dysferlin transcript (accession number AF075575). The new isoform differs from the previously described dysferlin protein in that it uses a different initial exon (located in the intron between Exons 1 and 2) and has a significantly different amino acid sequence in the N-terminal region. The N-terminal amino acid sequence of DYSF_v1 closely resembles that of the dysferlin protein characterized for *M. musculus*, in contrast to the N-terminal sequence of DYSF, which has little homology with the murine sequence.

A second murine start sequence, which is analogous to the "original" human sequence, is found on a contig (Accession number AC153607) from mouse chromosome 6, which contains the mouse dysferlin gene. This region of similarity is located on the same strand as the standard mouse dysferlin Exon 1, approximately 10.8 kb upstream. Comparing the murine and human sequences, of the 30 amino acids encoded by Exon 1, 28 are identical, and the other two are similar.

Therefore, mice appear to possess the same two full-length dysferlin isoforms as humans. For convenient comparison across species, we designate the upstream start exon (the "original" one in humans and the "alternate" one in mice, as Exon 1, and the downstream start exon (located between Exon 1 and Exon 2) as Exon 1a.

In both humans and mice, Exon 1 has the initial amino acid sequence MLRV…, while Exon 1a has the initial sequence MLCC… Exon 1a is located downstream of Exon 1 by 10.8 kb in mice and 12.8 kb in humans.

Exon 1 human-mouse alignment:

```
Mouse MLRVFILFAENVHTPDSDISDAYCSAVFAG
      MLRVFIL+AENVHTPD+DISDAYCSAVFAG
Human MLRVFILYAENVHTPDTDISDAYCSAVFAG
```

Predicted dysferlin protein sequences for *Rattus norvegicus* (accession numbers XP_232123 and XP_001069038) contain initial amino acid sequences which are identical to those of the two start exons of *M. musculus* (XP_232123 to mouse Exon 1, and XP_001069038 to mouse Exon 1a).

In their analysis of the 5'UTR of the human dysferlin gene, Foxton et al. [Eur. J. Hum. Genetics (2004) 12, 127–131], identified a number of possible upstream open reading frames. They suggested that transcription of these ORFs might regulate gene expression. They noted that the mouse 5'UTR was completely unlike the human sequence. In retrospect, this is a result of the human sequence known at the time, Exon 1, being compared to the mouse sequence Exon1a. Comparing the UTRs of human and mouse Exon 1, regions of homology are found.

Reversed and complemented bases 142801-144000 of mouse sequence AC153607

```
   1 tggcgagttg ggggcgcgcg ctcgagaggc agtcaatgca aattgtcagt tacacacatt
  61 tataagttcc gaggggagcc ggattggtaa atatccgaat cttaacagtt tttgttttga
 121 tttgttttaa agaaccagag gtgagttgtc tgtcgtttgt gtaaaagtgc ttctgaaact
 181 gtgagtcgtg acccctttgg ggggtcaaat aaccattcta cagaagtcga acatcatata
 241 tcttgcatat cagttattta cattacaact cataacagta gcaaaattac agttatgaag
 301 tagcaacgaa aataatttta tgattagggg gtcaccacaa catggttatt atcattatta
 361 ttattaattt tattattagt attaaagggc cgcagcatta ggaaggttga gaaccactgt
 421 gttaaagcat ttgttacttt tacagtcagg gaaggaagca aacttttaa agccgctgct
 481 gtccaccaat gaaattctga gctctcagtg aaagcgccag aggtgcttct gctcctagct
 541 ctaagccctg gatctctctt tctccctggg gctgtctgac cagtttctga ggagctatca
 601 cgtccccgtg gccactgcaa gcgccacgcg tagccaagcg tatcacagct cctgaacaga
 661 ggtggaaggt caggggtgga gcccaacttt cttctgtccc ggagagagat ctggtctgca
 721 ctgcgtgcct gaagctctgg tcctccttca GGCCATTGCG GCCGCCGCCC AGCCcgcagc
 781 ttcagggtct aagGCAAAGT GCCGTGTCAT TGGGAaagct ggtggcgggg cattgaatta
 841 caattccatg gagctggagt acaactGCGG GGGTGGGAAA TGAACAGAAT CCCCTGTTCT
                                              M   N  R  I   P  C  S
 901 Cctcgcaacg cactctgact agcggggtga ggccgtccga ggcgggggcc cactggggcc
     P  R  N  A   L  *
 961 gcctggacta gcttgttcct cttcagggca acacctgtga gccggcagcc attcatccaa
1021 gtcggcctcg ctcctggggc gagccctccg cctgcgcctt gaccctgccc ctccagctta
1081 gcgATGCTGC GAGTCTTCAT CCTTTTTGCG GAGAATGTCC ACACCCCGGA CTCCGACATC
        M  L  R   V  F  I   L  F  A   E  N  V  H   T  P  D   S  D  I
1141 AGCGATGCCT ACTGCTCCGC GGTGTTTGCA GGTAGgaccg ccggggagac cctgccaggc
     S  D  A  Y   C  S  A   V  F  A   G
```

Human dysferlin 5'UTR and Exon 1 (Genbank AJ566204)

```
3001 gggacacaga ctgcattcta cactcagata tataataaat actgcaattt acatgtgtgt
3061 atacattcag aggagaaagc ctgggctgcc aaatacccaa atgttaatat actatttttt
3121 cagaaccaga gatttctttc tttatgcttt ttgttttaaa atggcatgtg ttgcttttaa
3181 agccagaaaa agggggcaaa gtttttaagg cagctgcctt gcttgtcaat gaaattctca
3241 actccaggct tctcttaggg aaagaggcct ccccaggatc ccccgtccta ccccgggcgg
3301 tcgggttgag ttctggagag actgctccaa tccccgaggc ggaaggaggc aaccgatttg
3361 gcgcagcact cagccagggg gtagaagctc aggggaggag ccgagccttt ctcctgtcca
3421 agagcgagat ctgggctacg ccgggcgccc ggagccctag tccagccccc GGCCATCGCG
3481 GCCGCCGCCC AGCCaggtGC AAAATGCCGT GTCATTGGGA gactccgcag ccggagcatt
                            M  P  C   H  W  E   T  P  Q   P  E  H  *
3541 agattacagc tcgacggagc tcgggaaggg cgGCGGGGGT GGAAGATGAG CAGAAGCCCC
                                                 M  S   R  S  P
3601 TGTTCTCgga acgccggctg acaagcgggg tgagcgcagc cggggcgggg acccagccta
     C  S  R  N   A  G  *
3661 gcccactgga gcagccgggg gtggcccgtt ccccttaag agcaactgct ctaagccagg
3721 agccagagat tcgagccggc ctcgcccagc cagccctctc cagcgagggg acccacaagc
3781 ggcgcctcgg ccctcccgac ctttccgagc cctctttgcg ccctgggcgc acggggccct
3841 acacgcgcca agcATGCTGA GGGTCTTCAT CCTCTATGCC GAGAACGTCC ACACACCCGA
                    M  L  R   V  F  I   L  Y  A   E  N  V  H   T  P  D
3901 CACCGACATC AGCGATGCCT ACTGCTCCGC GGTGTTTGCA GGTAGgaggg gccgaccacc
      T  D  I   S  D  A  Y   C  S  A   V  F  A   G
```

Comparison of DNA sequence of the mouse and human dysferlin Exon 1. Areas of similarity are indicated in bold caps. Start codons are indicated in pink, other codons in blue with the coded amino acids shown below, and the acceptor splice site at the end of the exon is shown in red. Homologous 5' UTR regions are highlighted in identical colors in the two sequences. Upstream ORFs within the 5' UTR regions of homology are underlined.

BLAST matches between human Exon 1-MLRV (Genbank AJ566204--including the entire 5' UTR) and the mouse chromosome 6 contig (Genbank AC153607) show four regions of similarity. In addition to the coding region, there are also three portions of the 5' UTR which are highly conserved.

Within the regions of high homology between the 5' murine and human sequences, there are two ORFs in the human sequence, which would encode 12 and 11 amino acids, respectively. In the mouse sequence, only one of these ORFs occurs, due to a A-G substitution between humans and mice, which changes the ATG start codon of the second human ORF to GTG in the mouse sequence. The 5' UTR upstream ORF which the human and mouse sequences share are quite similar in their encoded amino acids.

Exons present in only some isoforms:

The dysferlin protein has three exons which are expressed in only some, but not all "full-length" transcripts. These include Exon 17, and two exons not included in the originally described 55-exon human dysferlin sequence: Exon 5a, located between exons 5 and 6, and Exon 40a, located between Exons 40 and 41. All three of these "optional" exons occur between C2 domains, so their presence or absence does not change dysferlin's conserved domain structure (see figure below). All three exons are expressed in EST in both humans and mice.

There are a total of 16 possible isoforms resulting from use of either of the two start exons and inclusion or exclusion of Exons 5a, 17, and 40a. Of these, 14 human sequences have been submitted to Genbank as of 2008; the only two which have not been described are the isoforms containing all three Exons 5a, 17, and 40a, with either start exon. A table listing which combination of exons each isoform in Genbank contains is given below.

Exon 5a

The mouse sequences NP_001071162 and NP_067444 includes an additional exon, 5a, between Exons 5 and 6 of the human sequence. The human Exon 5' is not included in sequence O75923, but is contained in clone DQ976379—which contains Exons 5,5a, 6, and 7. The human sequence below is taken from DQ976379.

Human-mouse alignment:

```
Human:  GGGQSRAETWSLLSDSTMDTRYSGKKWPAPT
        GGGQSRAETWSLLSDSTMDTRYSGKKWP PT
Mouse:  GGGQSRAETWSLLSDSTMDTRYSGKKWPVPT
```

Exon 17

Exon 17 of the human sequence is not included in mouse sequences _001071162 and NP_067444. A DNA sequence similar to human Exon 17 is found between base pairs 55300-55400 on the minus strand of mouse contig AC153607, between Exon 16 (64000, minus strand) and Exon 18 (54100, minus strand). The amino acid sequence predicted for this region is found

on mouse EST CO045564—which also contains Exons 16 and 18. The amino acid sequence below is taken from this EST.
Note, a splice variant of human dysferlin lacking Exon 17 has been reported (Salani et al, Muscle Nerve. 2004 Sep;30(3):366-74). This may account for Exon 17 not being included in the mouse reference sequence

Human-mouse alignment:

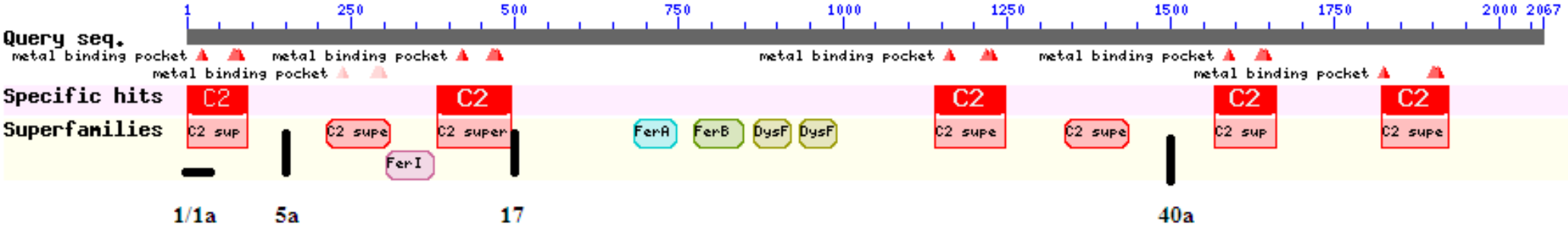
```
Human:  EEPAGVLKSPQATD
        EEPAG +K  +A+D
Mouse:  EEPAGAVKPSKASD
```

Exon 40a
Occurs between Exons 40 and 41 of the "standard" human dysferlin sequence.

Human Exon 40a (from EST EF015906—the sequence contains exons 40, 40a,41). Mouse sequence located on AC153608 bp 42535-42600. Mouse sequence is expressed on EST AK087986—contains exon 40a and surrounding exons).

Human-mouse alignment:

```
Human:  LADGLSSLAPTNTASPPSSPH
        L DGLSSL PTN    PSSPH
Mouse:  LTDGLSSLGPTNLTPSPSSPH
```



Location of the “optional” exons in dysferlin, superimposed on the sequence and conserved domains of Variant V1_3, which lacks Exons 5a, 17, and 40a. The portion of the N-terminal C2A domain comprised by the two alternate start exons, 1 and 1a, is also indicated. The conserved domain identification was performed by the CD tool on the NCBI website www.ncbi.nlm.nih.gov, with the exception that the C2 domain near AA 1400, which is not identified by this tool but is by other CD search algorithms, was added.

Protein transcripts of dysferlin

Known transcripts of dysferlin begin with one of two alternate start exons: Exon 1, whose translated sequence begins MLRV…, or Exon 1a, whose translated sequence begins MLCC… Exons 2-55 are contained in all known transcripts, with the exceptions that Exons 5a (between 5 and 6), 17, and 40a (between 40 and 41) are sometimes present and sometimes skipped. Exons present in each transcript are marked with an X. cDNA's corresponding to transcripts listed in Red are available from the Jain

Organism	Name	Accession #	Exon 1 (MLRV…)	Exon 1a (MLCC…)	Exon 5a	Exon 17	Exon 40a
Human	Dysferlin (= isoform CRA_b)	O75923	X			X	
Human	Variant 2	ACB12752	X		X	X	
Human	Variant 3	ACB12753	X				
Human	Variant 4	ACB12754	X			X	X
Human	Variant 5	ACB12755	X		X		
Human	Variant 6	ACB12756	X				X
Human	Variant 7	ACB12757	X		X		X
Human	Dysferlin_v1	ABB89736		X		X	
Human	Variant V1_2	ACB12758		X	X	X	
Human	Variant V1_3	ACB12759		X			
Human	Variant V1_4 (= isoform CRA_a)	ACB12760		X		X	X
Human	Variant V1_5	ACB12761		X	X		
Human	Variant V1_6	ACB12762		X			X
Human	Variant V1_7	ACB12763		X	X		X
Human	Isoform CRA_c *	EAW99765	X				
Mouse	Isoform 1 (=BAD21394)	NP_067444		X	X		
Mouse	Isoform 2	NP_001071162	X		X		
Mouse		Q9ESD7		X		X	
Mouse		AAG17046		X			
Mouse	(partial sequence)	EDK99114		X		X	

Foundation.

*Predicted C-terminal sequence contains part of Exon 52 and novel C-terminal domain (no transmembrane domain). Does not appear to be supported by ESTs.