Am. J. Hum. Genet. 65:635–644, 1999

Congenital End-Plate Acetylcholinesterase Deficiency Caused by a Nonsense Mutation and an A→G Splice-Donor–Site Mutation at Position +3 of the Collagenlike-Tail–Subunit Gene (*COLQ*): How Does G at Position +3 Result in Aberrant Splicing?

Kinji Ohno,[1] Joan M. Brengman,[1] Kevin J. Felice,[2] David R. Cornblath,[3] and Andrew G. Engel[1]

[1]Department of Neurology and Neuromuscular Research Laboratory, Mayo Clinic, Rochester, MN; [2]Department of Neurology, University of Connecticut Health Center, Farmington; and [3]Department of Neurology and Pathology, Johns Hopkins University, Baltimore

Summary

Congenital end-plate acetylcholinesterase (AChE) deficiency (CEAD), the cause of a disabling myasthenic syndrome, arises from defects in the *COLQ* gene, which encodes the AChE triple-helical collagenlike-tail subunit that anchors catalytic subunits of AChE to the synaptic basal lamina. Here we describe a patient with CEAD with a nonsense mutation (R315X) and a splice-donor–site mutation at position +3 of intron 16 (IVS16+3A→G) of *COLQ*. Because both A and G are consensus nucleotides at the +3 position of splice-donor sites, we constructed a minigene that spans exons 15–17 and harbors IVS16+3A→G for expression in COS cells. We found that the mutation causes skipping of exon 16. The mutant splice-donor site of intron 16 harbors five discordant nucleotides (at −3, −2, +3, +4, and +6) that do not base-pair with U1 small-nuclear RNA (snRNA), the molecule responsible for splice-donor–site recognition. Versions of the minigene harboring, at either +4 or +6, nucleotides complementary to U1 snRNA restore normal splicing. Analysis of 1,801 native splice-donor sites reveals that presence of a G nucleotide at +3 is associated with preferential usage, at positions +4 to +6, of nucleotides concordant to U1 snRNA. Analysis of 11 disease-associated IVS+3A→G mutations indicates that, on average, two of three nucleotides at positions +4 to +6 fail to base-pair, and that the nucleotide at +4 never base-pairs, with U1 snRNA. We conclude that, with G at +3, normal splicing generally depends on the concordance that residues at +4 to +6 have with U1 snRNA, but other *cis*-acting elements may also be important in assuring the fidelity of splicing.

Received April 7, 1999; accepted for publication June 30, 1999; electronically published August 9, 1999.

Address for correspondence and reprints: Dr. Andrew G. Engel, Department of Neurology, Mayo Clinic, Rochester, MN 55905. E-mail: age@mayo.edu

0002-9297/1999/6503-0008$02.00

Introduction

Acetylcholinesterase (AChE; E.C.3.1.1.7) rapidly hydrolyzes acetylcholine (ACh) released at cholinergic synapses. AChE limits the number of collisions between ACh and the acetylcholine receptor (AChR) and hence limits the duration of the synaptic response (Katz and Miledi 1973). Two classes of AChE are present in mammalian skeletal muscle (Massoulié et al. 1993; Rotundo and Fambrough 1994): (i) homomeric or globular forms that consist of monomers (G_1), dimers (G_2), or tetramers (G_4) of the T isoform of the catalytic subunit ($AChE_T$) and (ii) heteromeric or asymmetric forms that consist of $AChE_T$ subunits linked to a collagenlike-tail subunit composed of three ColQ strands. The function of the collagenlike tail is to anchor catalytic subunits to the basal lamina. The proline-rich attachment domain of each ColQ strand can bind an $AChE_T$ tetramer, allowing the formation of asymmetric A_4, A_8, and A_{12} moieties (Bon et al. 1997). Asymmetric AChE is concentrated at the end plate, where it is the predominant species of AChE (Hall 1973; Younkin et al. 1982). Congenital end plate AChE deficiency (CEAD; MIM 603034) results in a disabling myasthenic syndrome. The prolonged synaptic currents cause repetitive muscle-fiber action potentials, and neither these nor the clinical symptoms are altered by AChE inhibitors (Engel et al. 1977; Hutchinson et al. 1993).

We recently cloned human *COLQ* cDNA (MIM 603033), determined the structure of the gene, and identified six recessive truncation mutations in six patients with CEAD (Ohno et al. 1998b). We also expressed each mutant *COLQ* cDNA, along with the wild-type $ACHE_T$ cDNA, in COS cells. First, a mutation that truncates ColQ before the proline-rich attachment domain prevents association of ColQ with $AChE_T$. Second, truncation mutations in the collagenic domain of ColQ generate a mutant ~10.5 S species of AChE composed of one $AChE_T$ tetramer and a truncated ColQ strand. These

mutations prevent the assembly of the normal asymmetric forms of AChE. Donger et al. (1998) also cloned *COLQ* cDNA and reported a recessive missense mutation in the C-terminal region of ColQ that caused partial CEAD without preventing the assembly of the normal asymmetric forms of AChE. Feng et al. (1999) recently reported that ColQ-deficient mice exhibit abnormalities similar to those observed in patients with CEAD.

Here we trace a case of CEAD to two heterozygous mutations in the C-terminal region of ColQ: a nonsense mutation (R315X) and a splice-donor–site mutation (IVS16+3A→G). Since the intron nucleotide at +3 is an A in 49% and a G in 46% of human genes (see table 1), we constructed a *COLQ* minigene that spans exons 15–17 and harbors IVS16+3A→G. Expression of the minigene in COS cells revealed that IVS16+3A→G causes skipping of exon 16. Critical examination of the nucleotides surrounding IVS16+3 in *COLQ*, of other disease-associated A→G mutations at position +3, and of native splice-donor sites leads to the conclusion that, with G at position +3, normal splicing generally depends on concordance of residues at +4 to +6 with U1 snRNA, the molecule responsible for splice-donor–site recognition (Mount et al. 1983).

Patient and Methods

Patient

A 46-year-old male had had both progressive weakness increased by exertion and difficulty in walking, since age 8 years. On examination at age 43 years, the patient had weakness of facial, respiratory, neck, and proximal-limb muscles. An electromyogram showed a decremental response on repetitive stimulation and a repetitive compound muscle-fiber action potential in response to a single nerve stimulus. The patient has no anti–acetylcholine receptor (AChR) antibodies, and neither his physiological abnormalities nor his symptoms respond to AChE inhibitors. His two sisters are similarly affected. His parents are unaffected.

Mutation Analysis

Genomic DNA was isolated from proteinase/SDS digestion of blood (Sambrook et al. 1989). Since a repetitive compound muscle-action potential can also occur in the slow-channel congenital myasthenic syndrome caused by a mutation in an AChR subunit gene, we first sequenced all 44 exons of four AChR subunit genes, as well as their flanking regions (Ohno et al. 1995), but we identified no mutation. We next sequenced the 17 constitutive and 2 alternatively transcribed exons, as well as their flanking regions in *COLQ* (Ohno et al. 1998b). Sequencing analysis was performed with an ABI 377 DNA sequencer (Perkin-Elmer).

For tracing the R315X mutation in family members, we amplified a 342-bp fragment of *COLQ*, using primers *5*′-AGGCATACCCTGAAATGCTG-*3*′, in intron 12, and 5′-GGATTTTACAAAGCCCCATA-3′, in intron 13. The R315X mutation results in loss of an *Aci*I site (New England Biolabs), so that the wild-type allele gives rise to 249- and 93-bp fragments, whereas the mutant allele remains undigested.

For tracing IVS16+3A→G in family members and in 100 normal controls, we employed allele-specific PCR. The sense primer was 5′-AGGGGTGCAGCACTTA-

Table 1

Nucleotide Frequencies at −3 to +6 in Human Splice-Donor Sites

	% of Nucleotide at Position								
Splice Sites and Nucleotide	−3	−2	−1	+1	+2	+3	+4	+5	+6
Total splice sites (1,801 sites)									
A	33.3	*59.6*	8.8	0	0	*48.6*	*72*	7.2	15.4
C	*36.6*	13.4	3.3	0	0	2.6	8.2	4.6	17.3
G	18.4	13.1	*80.3*	*100*	0	46.4	11.4	*83.3*	20.9
T	11.7	13.8	7.5	0	*100*	2.4	8.4	4.9	*46.4*
Splice sites with A at position +3 (876 sites)									
A	33.7	*61.4*	10.3	0	0	*100*	*65.5*	11.3	18.5
C	*37*	10	2.7	0	0	0	9.5	7.2	15.6
G	16.8	12.9	*78.5*	*100*	0	0	12	*72.6*	22.5
T	12.6	15.6	8.4	0	*100*	0	13	8.9	*43.4*
Splice sites with G at position +3 (835 sites)									
A	32.2	*55.8*	8.1	0	0	*0*	*80.6*	3	12.7
C	*36.2*	17.6	4.2	0	0	0	7.1	2	19.2
G	20	14.1	*80.4*	*100*	0	100	9.5	*94*	18.3
T	11.6	12.5	7.3	0	*100*	0	2.9	1	*49.8*

Note.—The data were compiled from the GENIO gene database, which represents 1,801 introns of 382 human multiexon genes. Nucleotides complementary to the 5′ end of U1 snRNA are italicized.

GGT-3′. The respective wild-type and mutant antisense primers were 5′-GAGGGAGGGGAGTCATCgCT-3′ and 5′-GAGGGAGGGGAGTCATCgCC-3′, where "g" represents a deliberately introduced mismatch to avoid annealing of the primer to the opposite allele. The size of the expected PCR product was 235 bp.

Construction of Minigenes to Analyze Aberrant Splicing

To evaluate the effects that IVS16+3A→G has on pre-mRNA splicing, we constructed a *COLQ* minigene spanning exons 15–17 (fig. 1). Since introns 15 and 16 are large (~2.4 and ~2.9 kb, respectively), we trimmed the intronic sequences. Three DNA fragments were amplified from control genomic DNA to include exons 15, 16, and 17, respectively, using the Expand High Fidelity PCR System (Boehringer Mannheim). The three PCR products were mixed and double-digested with *Bam*HI and *Sac*I (Boehringer Mannheim). *Bam*HI recognition sites are at IVS15+229 and IVS15−90. *Sac*I recognition sites are at IVS16+266 and IVS16−322. The restriction fragments were purified with the QIAquick PCR Purification Kit (Qiagen) to eliminate small restriction fragments and then were ligated at the *Bam*HI sites in intron 15 and at the *Sac*I sites in intron 16. Ligated products were used as a template for the second-step PCR with one primer located within exon 15 and the other in the 3′ noncoding region of exon 17. The PCR product was excised from 1% SeaPlaque agarose gel (FMC BioProducts), purified with the Wizard PCR Preps (Promega), and ligated to pGEM-T vector (Promega). The entire insert was sequenced to confirm that the expected fragment was cloned and that there was no PCR artifact. The insert was then transferred to the cytomegalovirus-based expression vector, pRBG4 (Lee et al. 1991).

Next, we introduced IVS16+3A→G into the minigene in pRBG4, using the QuickChange Site-Directed Mutagenesis kit (Stratagene). We similarly introduced C, A, A, and T nucleotides at positions −3, −2, +4, and +6, respectively. The presence of the desired mutations and the absence of unwanted mutations were confirmed by sequencing the entire insert.

Transfection of Minigene Constructs into COS Cells and Reverse-Transcription–PCR Analysis of Cytoplasmic RNA

COS-7 cells were transfected, 1 d after plating, by the DEAE-dextran method (Selden et al. 1986) with 1 μg of a minigene construct per 35-mm dish. The cells were collected 3 days after transfection, and cytoplasmic RNA was extracted by the RNeasy Mini kit (Qiagen), according to the manufacturer's instructions. Isolated RNA was treated with 40 units of DNase to eliminate plasmid DNA and was purified again with the RNeasy Mini kit. One-third of the isolated cytoplasmic RNA was used for cDNA synthesis with an oligo-dT primer and the cDNA Cycle kit (Invitrogen). One-tenth of the synthesized cDNA was used for RT-PCR with primer 5′-CAGCTGACCCCTTTCTACCC-3′ in exon 15 and with primer 5′-AGCGGCAGGGCGTGGAGT-3′ in exon 17. With these primers, a normal transcript would yield a 286-bp fragment, whereas a transcript skipping exon 16 (103 bp) would give rise to a 183-bp fragment.

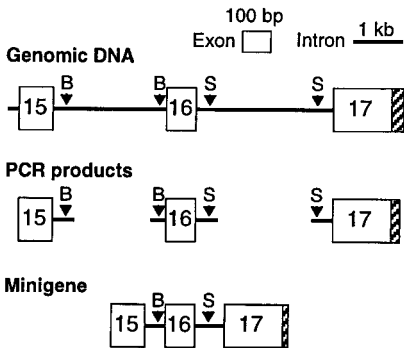

Figure 1 Construction of a minigene spanning exons 15–17. Three PCR products were amplified from genomic DNA to include exons 15–17. The PCR products were digested with *Bam*HI (B) and *Sac*I (S), and ligated to make a minigene. Unblackened boxes show exons and their numbers. Hatched areas indicate 3′ noncoding region. Exons and introns are drawn to scale.

Faint bands below the aberrantly transcribed 183-bp fragment were excised from 4% NuSieve agarose (FMC Biolabs), purified with the Wizard PCR Preps, and ligated into pGEM-T vector. To determine the origin of the faint bands, eight pGEM-T clones were sequenced.

Expression Studies of AChE Molecules in COS Cells

In a study reported elsewhere (Ohno et al. 1998b), we cloned human $ACHE_T$ cDNA and *COLQ* cDNA and introduced them into the pTargeT expression vector (Promega). We engineered R315X into *COLQ* cDNA, using the QuickChange Site-Directed Mutagenesis kit. To eliminate exon 16 from *COLQ* cDNA, we used the aberrant transcript that skips exon 16 (see Results). The RT-PCR product amplified from the aberrant transcript served as a megaprimer for the megaprimer-based modification of the QuickChange Site-Directed Mutagenesis kit (Ohno et al. 1998a). The presence of the desired mutation and the absence of unwanted mutations were confirmed by sequencing the entire insert.

COS-7 cells were transfected, 1 d after plating, by the DEAE-dextran method with 5 μg of $AChE_T$ cDNA and 5 μg of DNA, encoding the wild-type or mutant *COLQ* cDNA, per 10-cm dish. We extracted AChE molecules 3 d after transfection, fractionated AChE in a 5%–20% sucrose density gradient, and determined AChE activity

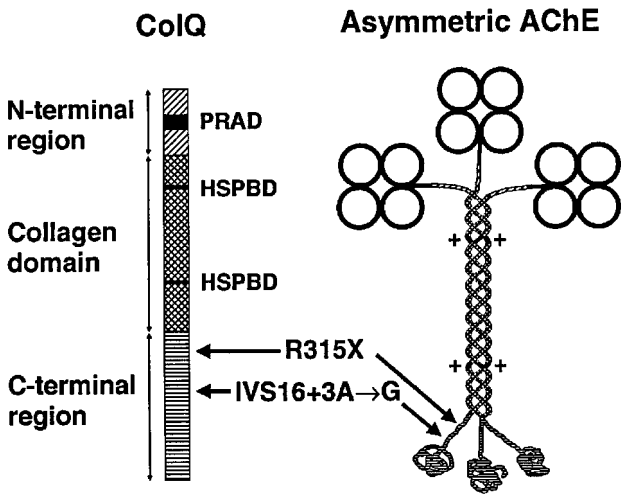

Figure 2 Schematic presentation of positions of *COLQ* mutations. Both R315X and IVS16+3A→G are at the C-terminal region. Arrows point to the C-terminal end of the wild-type residues. PRAD = proline-rich attachment domain to bind $AChE_T$ tetramer; HSPBD = heparan sulfate proteoglycan-binding domain to anchor ColQ in the synaptic basal lamina.

in the gradient fractions, by the Ellman method (Ohno et al. 1998b).

Analysis of 1,801 Native Splice-Donor Sites in Humans and of 10 Additional Disease-Associated A→G Mutations at Position +3

The GENIO database, a nonredundant gene database derived from GenBank release 102, lists 382 multiexon genes in humans (see Electronic-Database Information). These genes harbor 1801 introns, and we extracted 1,801 splice-donor–site sequences at positions −3 to +6 from these introns.

Disease-associated A→G mutations at position +3 have been reported at 10 splice-donor sites of eight human genes. For seven mutations, we obtained the splice-donor–site sequences at positions −3 to +6 either directly from published reports or from the gene database (Higuchi et al. 1990; Carstens et al. 1991; Bidichandani et al. 1994; Purandare et al. 1994; Brackett et al. 1995; Richard et al. 1995; Zolezzi et al. 1997). For two mutations in the retinitis pigmentosa GTPase regulator (*RPGR*) gene, splice-donor–site sequences of introns 4 and 10 were kindly provided by A. Swaroop (Buraczynska et al. 1997; Fujita et al. 1997). For a mutation in the ferrochelatase gene, we determined the entire sequence of intron 10, using published PCR primers (Wang et al. 1995).

Results

Mutation Analysis

Direct sequencing of the 17 constitutive and 2 alternatively transcribed *COLQ* exons and their flanking regions revealed two heterozygous mutations in the C-terminal region of ColQ (fig. 2). The first mutation is a C→T transition at nucleotide 943, which converts an arginine codon to a TGA stop codon at position 315 (R315X). R315X eliminates 86% of the residues of the C-terminal region. The second mutation is an A→G transition at the third nucleotide of the splice-donor site of intron 16 (IVS16+3A→G). IVS16+3A→G was not found in 100 normal controls. Three affected siblings carry both mutations. Unaffected family members are either heterozygous for a single mutation or carry no mutation (fig. 3). Thus, the two mutations are heteroallelic and recessive.

Reverse-Transcription–PCR Analysis of COS Cells Transfected with Minigenes

Since both A and G nucleotides are commonly observed at position +3 of native splice-donor sites in humans, an A→G substitution at position +3 could be a rare polymorphism. We therefore analyzed the effects of IVS16+3A→G on pre-mRNA splicing, by examining COS cells transfected with a minigene construct harboring either the wild-type A or the mutant G nucleotide at position +3. RT-PCR of cytoplasmic RNA isolated from transfected COS cells shows that the wild-type splice-donor site of intron 16 results in normal splicing plus trace amounts of an aberrant transcript that skips

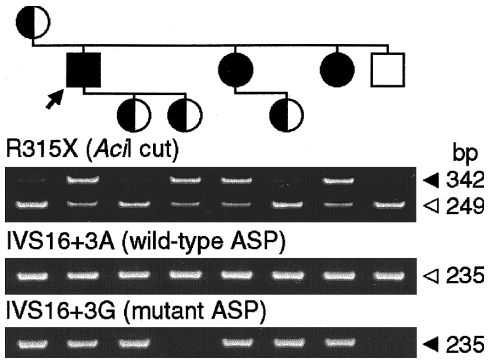

Figure 3 Restriction analysis of PCR products and allele-specific PCR (ASP) of genomic DNA isolated from blood of nuclear family members. The R315X mutation results in loss of an *Aci*I site. The wild-type allele gives rise to 249- and 93-bp fragments, whereas the mutant allele remains undigested at 342 bp. The 93-bp fragment is not shown. The IVS16+3A→G mutation is detected by ASP. Blackened symbols represent affected individuals; half-blackened symbols represent asymptomatic carriers. Unblackened and blackened arrowheads indicate wild-type and mutant fragments, respectively.

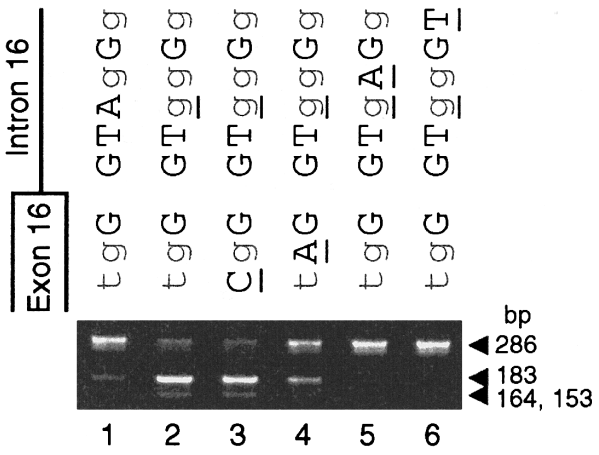

Figure 4 Size fractionation of RT-PCR products amplified from cytoplasmic RNA of COS cells transfected with minigene constructs harboring the indicated splice-donor–site sequences at the exon16/intron16 boundary of *COLQ*. The 286-bp band represents a normal transcript; the 183-bp band represents an aberrant transcript that skips exon 16. Faint 164- and 153-bp bands in lanes 2 and 3 are due to activation of cryptic splice sites in exon 15 (see fig. 5). Uppercase and lowercase letters indicate, respectively, accordant and discordant nucleotides for U1 snRNA. The patient's mutation and artificial nucleotide substitutions are underlined. Lane 1, Wild-type *COLQ*. Lane 2, IVS16+3A→G in the patient. Lane 3, IVS16+3A→G plus a T→C substitution at −3. Lane 4, IVS16+3A→G plus a G→A substitution at −2. Lane 5, IVS16+3A→G plus a G→A substitution at +4. Lane 6, IVS16+3A→G plus a G→T substitution at +6.

exon 16 (fig. 4, lane 1). By contrast, IVS16+3A→G yields a transcript with skipped exon 16 plus trace amounts of normal transcript (fig. 4, lane 2). Skipping of exon 16 eliminates 34% of the native residues of the C-terminal region and results in 57 missense codons followed by a stop codon. We previously reported seven alternative *COLQ* transcripts but none with skipped exon 16 (Ohno et al. 1998b). Thus, that some skipping of exon 16 occurs with the normal minigene either is due to the heterologous expression system or is peculiar to the minigene. A low consensus value of the native splice-donor site of intron 16 could also contribute to the aberrant splicing (fig. 5). That some normal transcript is detected with the IVS16+3A→G mutant may or may not be unique to the heterologous expression system employing a minigene. It is also possible that traces of the normally spliced transcript are expressed at the patient's end plates, which would mitigate the consequences of the splice-site mutation.

We also detected faint bands below the aberrant transcript that skips exon 16 (fig. 4, lanes 2 and 3). Sequencing of the cloned faint bands revealed activation of two cryptic splice sites within exon 15 (fig. 5). The activated cryptic splice sites are 30 and 19 bp upstream of the native splice-donor site of intron 15. The activation of cryptic splice sites is likely due to the low consensus value of the native splice-donor site of intron 15 (Krawczak et al. 1992; Nakai and Sakamoto 1994). Since both cryptic splice sites result in frameshifts, these aberrant transcripts cannot mitigate the consequences of skipped exon 16.

We next asked why a G nucleotide at position +3 causes aberrant splicing in *COLQ*, even though the consensus nucleotide at position +3 is either an A or a G. The wild-type sequence of the splice-donor site of *COLQ* intron 16 is tgG\GTAgGg, where the lowercase letters at positions −3, −2, +4, and +6 represent discordant nucleotides against the 5′ end of U1 snRNA (fig. 6). The mutant splice-donor site, tgG\GTggGg, harbors five discordant nucleotides, at positions −3, −2, +3, +4, and +6. We hypothesized that the number of discordant nucleotides in the mutant donor site exceeds the critical threshold and that the positions of the discordant nucleotides play a pivotal role in skipping of exon 16. To test this notion, we engineered accordant nucleotides at positions −3, −2, +4, and +6, one by one, into the construct that contained IVS16+3A→G and transfected them into COS cells (fig. 4, lanes 3–6). Introduction of the accordant C nucleotide at position −3 did not prevent aberrant splicing (fig. 4, lane 3). Introduction of the accordant A nucleotide at position −2 partially restored normal splicing (fig. 4, lane 4). Interestingly, a G→A substitution at position +4 completely rectified aberrant splicing (fig. 4, lane 5). This artificial construct functioned even better than the wild-type construct, since we detected no skipping of exon 16. Introduction of the accordant T nucleotide at position +6 also prevented aberrant splicing (fig. 4, lane 6). To summarize, in the presence of IVS16+3A→G, introduction of an accordant nucleotide at position −3 or −2 has no or little beneficial effect, whereas introduction of an accordant nucleotide at position +4 or +6 completely abolishes aberrant splicing. This implies that both the number and position of accordant nucleotides determine normal splicing. Therefore, a G nucleotide at position +3 is acceptable at the splice-donor site of intron 16 of

Figure 5 Nucleotide sequence of intron/exon boundaries of the wild-type *COLQ*. Boldface letters indicate invariant GT or AG dinucleotides at the native or cryptic splice sites. The activated cryptic splice sites are 30- and 19-bp upstream to the native splice-donor site in intron 15. The consensus value (see Discussion) for each splice site is shown above the invariant dinucleotide. Note that the consensus values for the cryptic splice sites are as high as that of the native splice-donor site in intron 15.

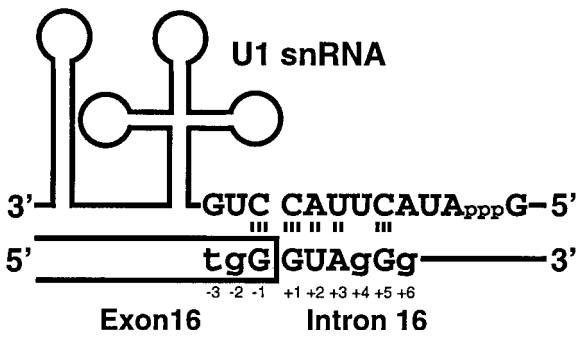

Figure 6 Schematic presentation of base pairing between the 5′ end of U1 snRNA and the wild-type splice-donor site of *COLQ* intron 16. Short vertical lines between the two molecules represent hydrogen bonds. Uppercase and lowercase letters indicate accordant and discordant nucleotides, respectively.

COLQ only when at least two of three nucleotides at positions +4 to +6 are complementary to the U1 snRNA sequence.

Expression Studies of AChE Molecules in COS Cells

We next investigated the effects of two *COLQ* mutations on the expression and assembly of the AChE molecules. We transfected wild-type *COLQ* cDNA or mutant *COLQ* cDNA harboring either R315X or skipped exon 16, along with the wild-type $ACHE_T$ cDNA. Extracts of COS cells transfected with wild-type $ACHE_T$ and wild-type *COLQ* show three asymmetric forms (A_{12}, A_8, and A_4) and three globular forms (G_4, G_2, and G_1) (fig. 7, *top*), as described elsewhere (Ohno et al. 1998b). Cotransfection of the R315X mutant with the wild-type $ACHE_T$ produces low peaks of the asymmetric and globular species of AChE and a prominent ~10.5 S mutant (M) peak composed of a catalytic tetramer and a truncated single strand of ColQ (Ohno et al. 1998b) (fig. 7, *center*). The sedimentation coefficient of the ~10.5 S mutant peak is slightly greater than that of G_4 peak (Ohno et al. 1998b), and, therefore, the mutant peak may include a small contribution from the G_4 peak. Cotransfection of the *COLQ* mutant harboring skipped exon 16 along with the wild-type $ACHE_T$ produces a sedimentation profile similar to that obtained after cotransfection of the wild-type *COLQ* and wild-type $ACHE_T$ (fig. 7, *bottom*), but the ~10.5 S peak is slightly higher than the wild-type G_4 peak, suggesting that a small amount of the ~10.5 S mutant species might also be present.

Discussion

AChE Molecules Formed in COS Cells

We have described two heterozygous truncation mutations in the *COLQ* gene in a patient with CEAD. The R315X mutation eliminates 86% of the C-terminal region of ColQ. The IVS16+3A→G mutation causes frameshifting skipping of exon 16, which eliminates 34% of the C-terminal region of ColQ.

Expression of the R315X mutant with the wild-type $ACHE_T$ reveals reduced asymmetric and globular forms of AChE and a prominent ~10.5 S mutant peak con-

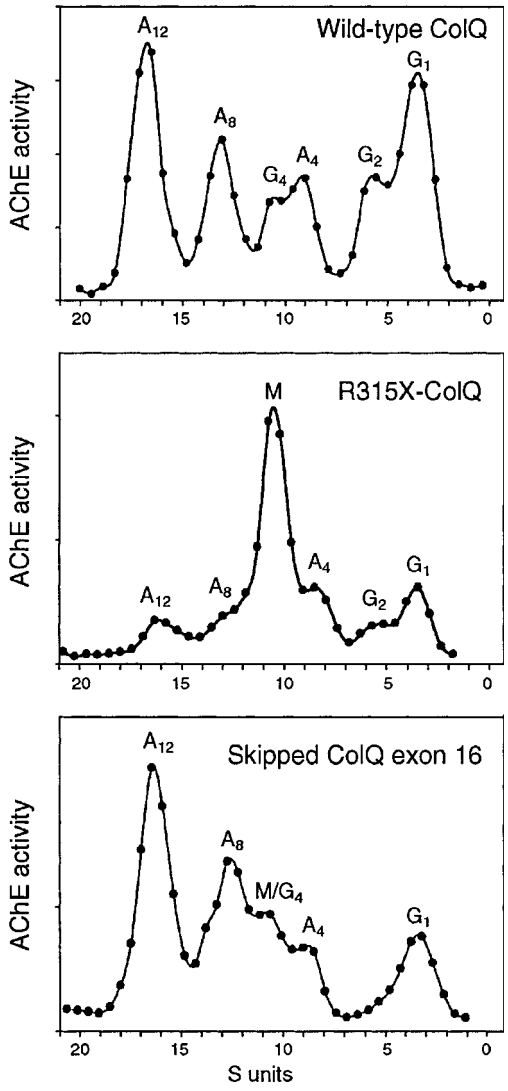

Figure 7 Sedimentation profiles of AChE species extracted from COS cells transfected with the indicated *COLQ* cDNA along with the wild-type $ACHE_T$ cDNA. G_1, G_2, and G_4 are globular forms; A_4, A_8, and A_{12} are asymmetric forms; M is a ~10.5 S mutant peak composed of a catalytic tetramer bound to a truncated single strand of ColQ.

Table 2

Effect that Nucleotide at Intron Position +3 Has on Complementarity to U1 snRNA

Nucleotide at Position +3	No. of Sites	Mean ± SD CV (and DN) at Segment[a]		
		−3 to −1	+4 to +6	−3 to +6
Any nucleotide	1,801	.728 ± .289 (1.2 ± .9)	.737 ± .277 (1.2 ± .9)	.843 ± .087 (2.2 ± .9)
Native A	876	.726 ± .296 (1.2 ± .9)	.660 ± .306 (1.2 ± .9)	.825 ± .088 (2.4 ± .9)
Native G	835	.715 ± .288 (1.3 ± .9)	.821 ± .213* (.8 ± .7*)	.867 ± .078* (3 ± .9*)
Mutant G	11	.839 ± .172 (.9 ± .7)	.343 ± .227* (2 ± .6*)	.753 ± .074* (3.9 ± 1*)

*$P < .005$; native G is compared to native A; mutant G is compared to native G.

[a] CV = consensus value; DN = number of discordant nucleotides.

sisting of a single ColQ strand attached to an $AChE_T$ tetramer (fig. 7, *center*). Formation of triple-helical ColQ clusters positive charges at heparan sulfate proteoglycan binding domains and is essential for anchoring the collagenlike tail to the basal lamina (Deprez and Inestrosa 1995). Therefore, the single-stranded ~10.5 S mutant species is unlikely to be anchored to the synaptic basal lamina. Moreover, we have shown elsewhere that no AChE is anchored to the synaptic basal lamina with truncation mutations in the collagenic domain of ColQ that produce prominent ~10.5 S peaks (Ohno et al. 1998b). The small amount of asymmetric A_4, A_8, and A_{12} species detected with this mutation may not be sufficient to effect hydrolysis of ACh at the synapse; in addition, truncation of the C-terminal region of ColQ likely prevents anchoring of the asymmetric species to the basal lamina.

The IVS16+3A→G mutation causes skipping of exon 16. Expression of *COLQ* cDNA lacking exon 16 with the wild-type $ACHE_T$ cDNA indicates that the skipped exon 16 does not prevent the assembly of the asymmetric species of AChE (fig. 7, *bottom*). Therefore, the AChE deficiency at the end plate must owe entirely to a defect in the anchoring of the truncated asymmetric species to the basal lamina. Casanueva et al. (1998) recently reported that two collagenous polypeptides (140 and 195–215 kD) in the synaptic basal lamina bind asymmetric AChE, but the binding domain(s) in ColQ were not identified. Additional studies will be required to determine whether the C-terminal region of ColQ binds to these polypeptides.

That more asymmetric species of AChE are formed when exon 16 of *COLQ* is skipped than when the R315X mutant is present may be because of preservation of a longer stretch of the C-terminal region with the former mutation, for the C-terminal region of ColQ, like that of other collagen molecules, likely promotes the triple-helical association of individual collagen strands (Dölz et al. 1988; Prockop and Kivirikko 1995). It is interesting to note, however, that only 14% of the C-terminal region is able to induce triple-helical association of ColQ, although at a markedly reduced efficiency (compare top and center panels of fig. 7). On the other hand, when 66% of the C-terminal region is retained, the triple-helical association of ColQ is unhindered (compare the top and bottom panels of fig. 7).

How Does a G Nucleotide at Intron Position +3 Result in Aberrant Splicing?

The IVS16+3A→G mutation is intriguing because both A and G nucleotides are commonly observed at the +3 position of native splice-donor sites in humans (table 1). Several site-directed mutagenesis studies have investigated the effects of the splice-donor–site sequence on splicing (Grabowski et al. 1991; Kuo et al. 1991; Kister et al. 1993; Tsukahara et al. 1994), but none has addressed the consensus A and G nucleotides at +3. Our transfection experiments using minigene constructs indicate that accordant nucleotides at positions +4 and +6 in intron 16 restore normal splicing when the nucleotide at intron position +3 is G, whereas accordant nucleotides at position −3 or −2 in exon 16 had no or little effect on splicing (see fig. 4). To evaluate the generality of this observation, we examined the nucleotide frequencies at positions −3 to +6 of the 1,801 native splice-donor sites of 382 human multiexon genes (table 1). When the nucleotide at position +3 is G, the frequencies of accordant nucleotides at positions +4 to +6 are higher than those observed in splice-donor sites with A at position +3. On the other hand, the frequencies of the accordant nucleotides at positions −3 to −1 do not depend on whether position +3 is occupied by an A or a G nucleotide. It is interesting to note that the frequency of an accordant T at +6 is only slightly higher with G at +3 than with A at +3, whereas the introduction of a T at +6 restores normal splicing in our transfection experiments. This may imply that, for fidelity of splicing, the number of accordant nucleotides at +4 to +6 is more crucial than their position.

Next, using 1,801 human slice donor sites, we calculated the consensus values (CVs) at positions −3 to −1 ($CV_{-3\text{ to }-1}$), positions +4 to +6 ($CV_{+4\text{ to }+6}$), and all positions from −3 to +6 ($CV_{-3\text{ to }+6}$), in two categories

Table 3

A→G Mutations at Intron Position +3

Gene	Mutated Sequence[a]	Intron	Reference
Coagulation factor VIII[b]	CAG:GTgtGT	6	Bidichandani et al. 1994
Coagulation factor VIII[c]	CAG:GTgtGa	14	Higuchi et al. 1990
Ornithine transcabamylase[d]	gAG:GTgtGc	7	Carstens et al. 1991
Collagen α2I[d]	CgG:GTggGT	26	Zolezzi et al. 1997
β-hexosaminidase A[d]	CtG:GTgtGa	8	Richard et al. 1995
Trifunctional protein α[d]	aAG:GTgtcT	3	Brackett et al. 1995
Ferrochelatase[d]	aAG:GTgtcT	10	Sarkany and Cox 1995; Wang et al. 1995
Neurofibromin[d]	tAG:GTgtaT	11	Purandare et al. 1994
RPGR[d]	CAG:GTgtag	4	Buraczynska et al. 1997
RPGR[e]	agG:GTgcaa	10	Fujita et al. 1997
COLQ[d,e]	tgG:GTggGg	16	Present study

[a] Uppercase and lowercase letters indicate, respectively, complementarity and discordance with regard to U1 snRNA. Colons indicate exon/intron boundaries.

[b] Exons 5 and 6 are skipped.

[c] mRNA is not analyzed.

[d] Preceding exon is skipped.

[e] Cryptic splice site is activated.

with A and G at position +3 (table 2). The CV reflects the similarity of a given splice site to the consensus sequence deduced from a comprehensive collation of native genes (Shapiro and Senapathy 1987). Consistent with the observations in table 1, $CV_{+4\ to\ +6}$ with G at position +3 is significantly higher, and the number of discordant nucleotides is significantly lower, than the corresponding values with A at position +3.

Ten disease-associated A→G splice-donor–site mutations have been reported at intron position +3, and aberrant splicing was documented for nine of these mutations (table 3). Including the currently reported IVS16+3A→G, calculation of CVs at 11 mutant splice-donor sites reveals that the $CV_{+4\ to\ +6}$ of the mutant sites is significantly lower, and that the number of discordant nucleotides is significantly higher, than it is with a native G at position +3 (table 2). Ketterling et al. (1999) recently proposed that splice-donor–site mutations not involving GT-invariant nucleotides cause disease when only five or six of eight nucleotides at −2 to +6 match the consensus sequence (the "5-6 hypothesis"). Of 11 A→G mutations at +3, 10 conform to this hypothesis. Our studies also point to the positional significance of accordant nucleotides at splice-donor sites.

Interestingly, at all 11 mutant splice-donor sites the nucleotides at position +4 are always discordant (table 3). A T nucleotide is used at position +4 at 8 of 11 mutant sites, yet T is the least frequent nucleotide at this position in native splice sites when G is at +3 (table 1). Indeed, the "GT" dinucleotides at positions +3 and +4 are the only discordant nucleotides in the mutant splice-donor site of intron 6 of the coagulation factor VIII gene (Bidichandani et al. 1994). Therefore, discordance of nucleotides at positions +3 and +4 may be necessary for aberrant splicing. Discordance at +3 and +4 alone, however, may not be sufficient for aberrant splicing, because a similar minigene construct in our expression system resulted in normal splicing (see fig. 4, lane 6), and because discordant nucleotides at +3 and +4 were observed at 11% of native splice-donor sites. That there is no critical threshold in the splice-donor–site sequence that differentiates normal from aberrant splicing implies that additional *cis*-acting elements also play a role in assuring the fidelity of splicing. The additional *cis*-acting elements in splicing include a splice-acceptor site in the preceding intron (Berget 1995), polypyrimidine tract (Roscigno et al. 1993; Anderson and Moore 1997), intronic (A/U)GGG consensus sequence (Sirand-Pugnet et al. 1995), branch-point sequence (Berglund et al. 1997), exonic splicing enhancers (Watakabe et al. 1993; Tanaka et al. 1994), secondary structures at the splice site (Clouet d'Orval et al. 1991), and exon size (Dominski and Kole 1991, 1992). For example, suppressor mutations remote from the splice-donor site can rescue aberrant splicing caused by splice-donor–site mutations at positions +3, +4, or +5 in the dihydrofolate reductase gene (Carothers et al. 1993). Zhang (1998) recently reported that A, rather than G, is preferentially used at +3 at human splice-donor sites at low G+C loci; this also implicates exonic *cis*-acting elements in splicing.

The importance of discordance of a G nucleotide at intron position +3 to U1 snRNA has been underestimated, because both A and G at +3 are consensus nucleotides. Our findings imply that nucleotides at +4 to +6 need to be accordant to U1 snRNA when there is a G nucleotide at +3. Our observations should also help to determine whether an A→G substitution at position +3 is a mutation or a polymorphism and should aid in the search for exons in human genomic sequences.

Acknowledgments

This work was supported by National Institutes of Health grant NS6277 (to A.G.E.) and Muscular Dystrophy Association grants (to A.G.E. and K.O.).

Electronic-Database Information

Accession numbers and URLs for data in this article are as follows:

GENIO database, http://ipvr2.informatik.uni-stuttgart.de/GENIO/seq/

Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim/ (for CEAD [MIM 603034] and the human *COLQ* gene [MIM 603033])

References

Anderson K, Moore MJ (1997) Bimolecular exon ligation by the human spliceosome. Science 276:1712–1716

Berget SM (1995) Exon recognition in vertebrate splicing. J Biol Chem 270:2411–2414

Berglund JA, Chua K, Abovich N, Reed R, Rosbach M (1997) The splicing factor BBP interacts specifically with the pre-mRNA branchpoint sequence UACUAAC. Cell 89:781–787

Bidichandani SI, Shiach CR, Lanyon WG, Connor JM (1994) A novel splice donor mutation affecting position +3 in intron 6 of the factor VIII gene. Hum Mol Genet 3:651–653

Bon S, Coussen F, Massoulié J (1997) Quaternary associations of acetylcholinesterase. II. The polyproline attachment domain of the collagen tail. J Biol Chem 272:3016–3021

Brackett JC, Sims HF, Rinaldo P, Shapiro S, Powell CK, Bennett MJ, Strauss AW (1995) Two alpha subunit donor splice site mutations cause human trifunctional protein deficiency. J Clin Invest 95:2076–2082

Buraczynska M, Wu W, Fujita R, Buraczynska K, Phelps E, Andreasson S, Bennett J, et al (1997) Spectrum of mutations in the RPGR gene that are identified in 20% of families with X-linked retinitis pigmentosa. Am J Hum Genet 61:1287–1292

Carothers AM, Urlaub G, Grunberger D, Chasin LA (1993) Splicing mutants and their second-site suppressors at the dihydrofolate reductase locus in Chinese hamster ovary cells. Mol Cell Biol 13:5085–5098

Carstens RP, Fenton WA, Rosenberg LR (1991) Identification of RNA splicing errors resulting in human ornithine transcarbamylase deficiency. Am J Hum Genet 48:1105–1114

Casanueva OI, Deprez P, Garciahuidobro T, Inestrosa NC (1998) At least two receptors of asymmetric acetylcholinesterase are present at the synaptic basal lamina of *Torpedo* electric organ. Biochem Biophys Res Commun 250:312–317

Clouet d'Orval B, d'Aubenton Carafa Y, Sirand-Pugnet P, Gallego M, Brody E, Marie J (1991) RNA secondary structure repression of a muscle-specific exon in HeLa cell nuclear extracts. Science 252:1823–1828

Deprez PN, Inestrosa NC (1995) Two heparin-binding domains are present on the collagenic tail of asymmetric acetylcholinesterase. J Biol Chem 270:11043–11046

Dölz R, Engel J, Kühn K (1988) Folding of collagen IV. Eur J Biochem 178:357–366

Dominski Z, Kole R (1991) Selection of splice sites in pre-mRNAs with short internal exons. Mol Cell Biol 11: 6075–6083

——— (1992) Cooperation of pre-mRNA sequence elements in splice site selection. Mol Cell Biol 12:2108–2114

Donger C, Krejci E, Pou Serradell A, Eymard B, Bon S, Nicole S, Chateau D, et al (1998) Mutation in the human acetylcholinesterase-associated collagen gene, COLQ, is responsible for congenital myasthenic syndrome with end-plate acetylcholinesterase deficiency (type Ic). Am J Hum Genet 63:967–975

Engel AG, Lambert EH, Gomez MR (1977) A new myasthenic syndrome with end-plate acetylcholinesterase deficiency, small nerve terminals, and reduced acetylcholine release. Ann Neurol 1:315–330

Feng G, Krejci E, Molgo J, Cunningham JM, Massoulié J, Sanes JR (1999) Genetic analysis of collagen Q: roles in acetylcholinesterase and butyrylcholinesterase assembly and in synaptic structure and function. J Cell Biol 144: 1349–1360

Fujita R, Buraczynska M, Gieser L, Wu W, Forsythe P, Abrahamson M, Jacobson SG, et al (1997) Analysis of the RPGR gene in 11 pedigrees with the retinitis pigmentosa type 3 genotype: paucity of mutations in the coding region but splice defects in two families. Am J Hum Genet 61:571–580

Grabowski PJ, Nasim FU, Kuo HC, Burch R (1991) Combinatorial splicing of exon pairs by two-site binding of U1 small nuclear ribonucleoprotein particle. Mol Cell Biol 11: 5919–5928

Hall ZW (1973) Multiple forms of acetylcholinesterase and their distribution in endplate and non-endplate regions of rat diaphragm muscle. J Neurobiol 4:343–361

Higuchi M, Wong C, Kochhan L, Olek K, Aronis S, Kasper CK, Kazazian HH Jr, et al (1990) Characterization of mutations in the factor VIII gene by direct sequencing of amplified genomic DNA. Genomics 6:65–71

Hutchinson DO, Walls TJ, Nakano S, Camp S, Taylor P, Harper CM, Groover RV, et al (1993) Congenital endplate acetylcholinesterase deficiency. Brain 116:633–653

Katz B, Miledi R (1973) The binding of acetylcholine to receptors and its removal from the synaptic cleft. J Physiol (Lond) 231:549–574

Ketterling RP, Drost JB, Scaringe WA, Liao DZ, Liu JZ, Kasper CK, Sommer SS (1999) Reported in vivo splice-site mutations in the factor IX gene: severity of splicing defects and a hypothesis for predicting deleterious splice donor mutations. Hum Mutat 13:221–231

Kister L, Domenjoud L, Gallinaro H, Monique J (1993) A *cis*-acting selector of a 5′ splice site: cooperation between the sequence of the site and an upstream exonic element. J Biol Chem 268:21955–21961

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

Kuo HC, Nasim FH, Grabowski PJ (1991) Control of alternative splicing by the differential binding of U1 small nuclear ribonucleoprotein particle. Science 251:1045–1050

Lee BS, Gunn RB, Kopito RR (1991) Functional differences among nonerythroid anion exchangers expressed in a transfected human cell line. J Biol Chem 266:11448–11454

Massoulié J, Pezzementi L, Bon S, Krejci E, Vallette FM (1993) Molecular and cellular biology of cholinesterases. Prog Neurobiol 41:31–91

Mount SM, Pettersson I, Hinterberger M, Karmas A, Steitz JA (1983) The U1 small nuclear RNA-protein complex selectively binds a 5′ splice site in vitro. Cell 33:509–518

Nakai K, Sakamoto H (1994) Construction of a novel database containing aberrant splicing mutations of mammalian genes. Gene 141:171–177

Ohno K, Anlar B, Özdirim E, Brengman JM, DeBleecker JL, Engel AG (1998*a*) Myasthenic syndromes in Turkish kinships due to mutations in the acetylcholine receptor. Ann Neurol 44:234–241

Ohno K, Brengman JM, Tsujino A, Engel AG (1998*b*) Human endplate acetylcholinesterase deficiency caused by mutations in the collagen-like tail subunit (ColQ) of the asymmetric enzyme. Proc Natl Acad Sci USA 95:9654–9659

Ohno K, Hutchinson DO, Milone M, Brengman JM, Bouzat C, Sine SM, Engel AG (1995) Congenital myasthenic syndrome caused by prolonged acetylcholine receptor channel openings due to a mutation in the M2 domain of the ϵ subunit. Proc Natl Acad Sci USA 92:758–762

Prockop DJ, Kivirikko KI (1995) Collagens: molecular biology, diseases, and potentials for therapy. Annu Rev Biochem 64: 403–434

Purandare SM, Lanyon WG, Connor JM (1994) Characterisation of inherited and sporadic mutations in neurofibromatosis type-1. Hum Mol Genet 3:1109–1115

Richard MM, Erenberg G, Triggs-Raine BL (1995) An A-to-G mutation at the +3 position of intron 8 of the HEXA gene is associated with exon 8 skipping and Tay-Sachs disease. Biochem Mol Med 55:74–76

Roscigno RF, Weiner M, Garcia BM (1993) A mutational analysis of the polypyrimidine tract of introns: effects of sequence differences in pyrimidine tracts on splicing. J Biol Chem 268:11222–11229

Rotundo RL, Fambrough DM (1994) Function and molecular structure of acetylcholinesterase. In: Engel AG, Franzini-Armstrong C (eds) Myology. McGraw-Hill, New York, pp 607–623

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Plainview, NY

Sarkany RP, Cox TM (1995) Autosomal recessive erythropoietic protoporphyria: a syndrome of severe photosensitivity and liver failure. QJM 88:541–549

Selden RF, Howie KB, Rowe ME, Goodman HM, Moore DD (1986) Human growth hormone as a reporter gene in regulation studies employing transient gene expression. Mol Cell Biol 6:3173–3179

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

Sirand-Pugnet P, Durosay P, Brody E, Marie J (1995) An intronic (A/U)GGG repeat enhances the splicing of an alternative intron of the chicken beta-tropomyosin pre-mRNA. Nucleic Acids Res 23:3501–3507

Tanaka K, Watakabe A, Shimura Y (1994) Polypurine sequences within a downstream exon function as a splicing enhancer. Mol Cell Biol 14:1347–1354

Tsukahara T, Casciato C, Helfman DM (1994) Alternative splicing of beta-tropomyosin pre-mRNA: multiple cis-elements can contribute to the use of the 5′- and 3′-splice sites of the nonmuscle/smooth muscle exon 6. Nucleic Acids Res 22:2318–2325

Wang X-H, Poh-Fitzpatrick M, Chen T, Malavade K, Carriero D, Piomelli S (1995) Systematic screening for RNA with skipped exons—splicing mutations of the ferrochelatase gene. Biochim Biophys Acta 1271:358–362

Watakabe A, Tanaka K, Shimura Y (1993) The role of exon sequences in splice site selection. Genes Dev 7:407–418

Younkin SG, Rosenstein C, Collins PL, Rosenberry TL (1982) Cellular localization of the molecular forms of acetylcholinesterase in rat diaphragm. J Biol Chem 257:13630–13637

Zhang MQ (1998) Statistical features of human exons and their flanking regions. Hum Mol Genet 7:919–932

Zolezzi F, Valli M, Clementi M, Mammi I, Cetta G, Pignatti PF, Mottes M (1997) Mutation producing alternative splicing of exon 26 in the COL1A2 gene causes type IV osteogenesis imperfecta with intrafamilial clinical variability. Am J Med Genet 71:366–370