ASA E382K disrupts a potential exonic splicing enhancer and causes exon skipping, but missense mutations in *ASA* are not associated with ESEs

VORASUK SHOTELERSUK, TAYARD DESUDCHIT and SIRAPRAPA TONGKOBPETCH

Department of Pediatrics, Faculty of Medicine, Chulalongkorn University, Bangkok 10330, Thailand

Received May 17, 2004; Accepted July 26, 2004

Abstract. Metachromatic leukodystrophy (MLD) is an autosomal recessive disorder caused by mutations in the *arylsulfatase A* (*ASA*) gene. We identified a Thai boy with typical late-infantile MLD and found that he was a compound heterozygote for a novel mutation, g.IVS3-2A>G causing c.679-696del inherited from his father, and a previously reported missense mutation, g.1144G>A causing c.1102-1204del inherited from his mother. The g.1144G>A mutation was located in the middle of exon 7 and previously assumed to be deleterious by causing an amino acid change, E382K. We, herein, found that its actual pathogenic effect was splicing-related by disrupting a potential exonic splicing enhancer (ESE) and causing a complete exon 7 skipping. This is the first missense mutation in the *ASA* gene that is deleterious from disrupting a potential ESE. The results prompted us to investigate pathogenic effects of other reported missense mutations in the *ASA* gene. Unlike pathogenic missense mutations in some other genes, those in the *ASA* gene do not colocalize with ESE sites suggesting that pathogenic effects of majority of them are not splicing-related.

Introduction

Metachromatic leukodystrophy (MLD, MIM# 250100) is an autosomal recessively inherited disorder in which cerebroside sulfate degradation is defective. The degradation depends on the combined action of ASA and saposin B, a nonenzymatic activator protein. The cerebroside sulfate mainly occurs in the myelin sheaths in the central and peripheral nervous system. In patients with MLD, the sulfated glycolipids accumulate in lysosomes of these tissues and are responsible for the demyelination, which dominates the clinical manifestations. MLD can be divided into three major clinical forms: late-infantile, juvenile and adult, based primarily on clinical onset. Gait disturbance and mental regression are among the earliest signs in all variants. The disorder can be fatal in a few years after onset in the more severe form or may progress slowly over several decades in the less severe form of the disease. Major laboratory findings are the demonstration of demyelination by magnetic resonance imaging (MRI) and reduced nerve conduction velocity (NCV). Diagnosis of MLD is based on the clinical symptoms and laboratory findings caused by the demyelination and on demonstration of the deficiency of ASA or saposin B (1).

The ASA gene consists of 8 exons covering 3.2 kb of genomic DNA on chromosome 22q13.31. The primary transcript yields a 2.1-kb mRNA encompassing 363 nucleotides of 5' untranslated region, followed by an open reading frame of 1521 nucleotides and 130 nucleotides of 3' untranslated region. The ASA mRNA is translated into a polypeptide of 507 amino acids containing three potential N-glycosylation sites (1).

One hundred and eleven different MLD-causing mutations have been characterized in the *ASA* gene (2-6). Of these, 79 are missense mutations. The deleterious effects of some of these missense mutations have been supposedly attributed to their impact on primary amino acid sequence and protein structure.

We studied a Thai family with MLD and found that the affected boy was compoundly heterozygous for a novel mutation, g.IVS3-2A>G inherited from his father and a previously reported missense mutation, g.1144G>A (p.E382K) (7) inherited from his mother. The latter was predicted to be pathogenic because of the glutamic acid to lysine substitution at amino acid 382. But we demonstrated here that it actually disturbed splicing presumably from disrupting a potential exonic splicing enhancer (ESE) causing skipping of the whole exon 7. This prompted us to investigate pathogenic effects of other reported missense mutations in the *ASA* gene. We found that they do not colocalize with ESE sites, suggesting that their pathogenic effects are not splicing-related.

Materials and methods

Patient description. The patient is a Thai male with late-infantile MLD who was well until the age of 2 years, when

Correspondence to: Dr Vorasuk Shotelersuk, Medical Genetics and Metabolism, Department of Pediatrics, Sor Kor Building, 11th floor, King Chulalongkorn Memorial Hospital, Bangkok 10330, Thailand

E-mail: vorasuk.s@chula.ac.th

Key words: metachromatic leukodystrophy, arylsulfatase A, missense mutations, exonic splicing enhancer

Table I. Primers used in *ASA* gene analysis.

Primers	Sequences
MLDF1	5'-GTCTGCGGTATCGGAAAGAG-3'
MLDF2	5'-TGTACCCTGGCGTCCTGGTG-3'
MLDF3	5'-TTCGCCCATGACCTCATGGC-3'
MLDF4	5'-TTCTGGCCAGGTCATATCGC-3'
MLDR1	5'-AAACCCCCTCCAGACACCTG-3'
MLDR2	5'-TCATGAGCAGTCAGAGAGCT-3'
MLDEx4R	5'-GTGAAGATGACCAGCGTCTC-3'
MLDEx7F	5'-TTGGATGGCTTTGACCTCAG-3'

he began to experience difficulty walking and mental regression. Magnetic resonance imaging of the brain showed demyelination, nerve conduction velocity was slowed, and ASA activity measured in white blood cells (8) was reduced. He is still alive at the age of 10 years.

ASA gene analysis. After informed consent was obtained, 3 ml of peripheral blood from the boy and his parents were obtained. Total RNA was isolated from white blood cells using QIAamp® RNA blood mini kit (Qiagen, Valencia, CA, USA). Reverse transcription was performed using ImProm-II™ reverse transcriptase (Promega, Madison, WI, USA), according to the company recommendations. PCR amplification of the *ASA* cDNA exons 1-8 was performed using primers MLDF1 and MLDR1 shown in Table I. We

Table II. Pathogenic missense mutations found in the *ASA* gene

No.	Codon	Nucleotide	ESE	Nucleotide change	Amino acid change	Reference
1	32	94	1	GGC-AGC	Gly-Ser	2
2	43	129	0	AGC-AGA	Ser-Arg	5
3	68	203	1	CTG-CCG	Leu-Pro	2
4	82	245	1	CCG-CTG	Pro-Leu	2
5	84	251	0	CGG-CAG	Arg-GIn	2
6	84	250	0	CGG-TGG	Arg-Trp	2
7	86	257	0	GGC-GAC	Gly-Asp	2
8	94	280	0	CCC-GCC	Pro-Ala	2
9	95	284	1	AGC-AAC	Ser-Asn	2
10	96	287	1	TCC-TTC	Ser-Phe	2
11	99	296	0	GGC-GAC	Gly-Asp	2
12	99	296	0	GGC-GTC	Gly-Val	2
13	119	355	1	GGA-AGA	Gly-Arg	2
14	122	364	1	GGC-AGC	Gly-Ser	2
15	135	404	0	CTG-CCG	Leu-Pro	2
16	136	407	0	CCC-CTC	Pro-Leu	2
17	136	406	0	CCC-TCC	Pro-Ser	2
18	143	427	1	CGA-GGA	Arg-Gly	2
19	148	443	0	CCG-CTG	Pro-Leu	2
20	152	454	1	GAC-TAC	Asp-Tyr	2
21	153	459	0	CAG-CAC	GIn-His	2
22	154	461	1	GGC-GAC	Gly-Asp	2
23	155	464	1	CCC-CGC	Pro-Arg	2
24	155	464	1	CCC-CTC	Pro-Leu	3
25	167	500	1	CCT-CGT	Pro-Arg	2
26	169	505	1	GAC-AAC	Asp-Asn	2
27	172	515	0	TGT-TAT	Cys-Tyr	2
28	179	536	0	ATC-AGC	Ile-Ser	2
29	181	542	1	CTG-CAG	Leu-Gin	3
30	190	570	1	CAG-CAC	GIn-His	2
31	191	571	1	CCC-ACC	Pro-Thr	2
32	201	602	1	TAC-TGC	Tyr-Cys	2
33	212	635	0	GCC-GTC	Ala-Val	2
34	224	671	1	GCC-GTC	Ala-Val	2
35	227	679	0	CAC-TAC	His-Tyr	2
36	231	691	0	CCT-ACT	Pro-Tyr	2

Table II. Continued.

No.	Codon	Nucleotide	ESE	Nucleotide change	Amino acid change	Reference
37	244	730	1	CGC-TGC	Arg-Cys	2
38	244	731	1	CGC-CAC	Arg-His	2
39	245	733	1	GGG-AGG	Gly-Arg	2
40	250	749	0	TCC-TAC	Ser-Tyr	2
41	255	763	1	GAT-CAT	Asp-His	2
42	274	821	0	ACG-ATG	Thr-Met	2
43	281	841	1	GAC-TAC	Asp-Tyr	2
44	286	856	0	ACC-CCC	Thr-Pro	2
45	288	862	1	CGT-TGT	Arg-Cys	2
46	288	863	1	CGT-CAT	Arg-His	2
47	295	884	1	TCC-TAC	Ser-Tyr	2
48	298	893	1	TTG-TCG	Leu-Ser	2
49	299	895	0	CGG-TGG	Arg-Trp	5
50	300	899	1	TGT-TTT	Cys-Phe	2
51	306	916	1	TAC-CAC	Tyr-His	3
52	308	923	0	GGC-GAC	Gly-Asp	2
53	308	923	0	GGC-GTC	Gly-Val	2
54	309	925	0	GGT-AGT	Gly-Ser	2
55	311	932	0	CGA-CAA	Arg-GIn	2
56	312	936	0	GAG-GAT	Glu-Asp	2
57	314	940	0	GCC-ACC	Ala-Thr	2
58	325	973	1	GGC-AGC	Gly-Ser	3
59	327	980	1	ACC-ATC	Thr-Ile	2
60	329	985	1	GAG-AAG	Glu-Arg	5
61	335	1004	0	GAC-GTC	Asp-Val	2
62	367	1101	0	AAG-AAC	Lys-Asn	2
63	370	1109	0	CGG-CAG	Arg-GIn	2
64	370	1108	0	CGG-TGG	Arg-Trp	2
65	377	1130	1	CCG-CTG	Pro-Leu	2
66	382	1144	1	GAG-AAG	Glu-Lys	2
67	384	1150	1	CGT-TGT	Arg-Cys	2
68	390	1169	1	CGG-CAG	Arg-GIn	2
69	390	1168	1	CGG-TGG	Arg-Trp	2
70	391	1172	1	ACT-AGT	Thr-Ser	2
71	397	1189	1	CAC-TAC	His-Tyr	2
72	408	1223	1	ACC-ATC	Thr-Ile	2
73	409	1226	1	ACT-ATT	Thr-Ile	2
74	425	1273	1	CCC-ACC	Pro-Thr	2
75	426	1277	1	CCG-CTG	Pro-Leu	2
76	428	1283	0	CTC-CCC	Leu-Pro	2
77	429	1286	0	TAT-TCT	Tyr-Ser	3
78	464	1391	0	GCC-GTC	Ala-Val	2
79	488	1462	0	TGC-CGC	Cys-Arg	4

used 1 µl of first-strand cDNA, 1X PCR buffer (Promega, Madison, WI, USA), 1.5 mM $MgCl_2$, 0.2 mM dNTPs, 0.5 µM of each primer, and 0.5 U Taq DNA polymerase (Promega) in a total volume of 20 µl. PCR products were treated with ExoSAP-IT (USP Corporation, Cleveland, OH, USA), according to the company recommendations, and sent for direct sequencing, using primers MLDF1, MLDF2, MLDF3, MLDF4, MLDR1, and MLDR2 as shown in Table I, at the Macrogen Inc., Seoul, Korea. Genomic DNAs of the proband and his parents were obtained from whole blood using a

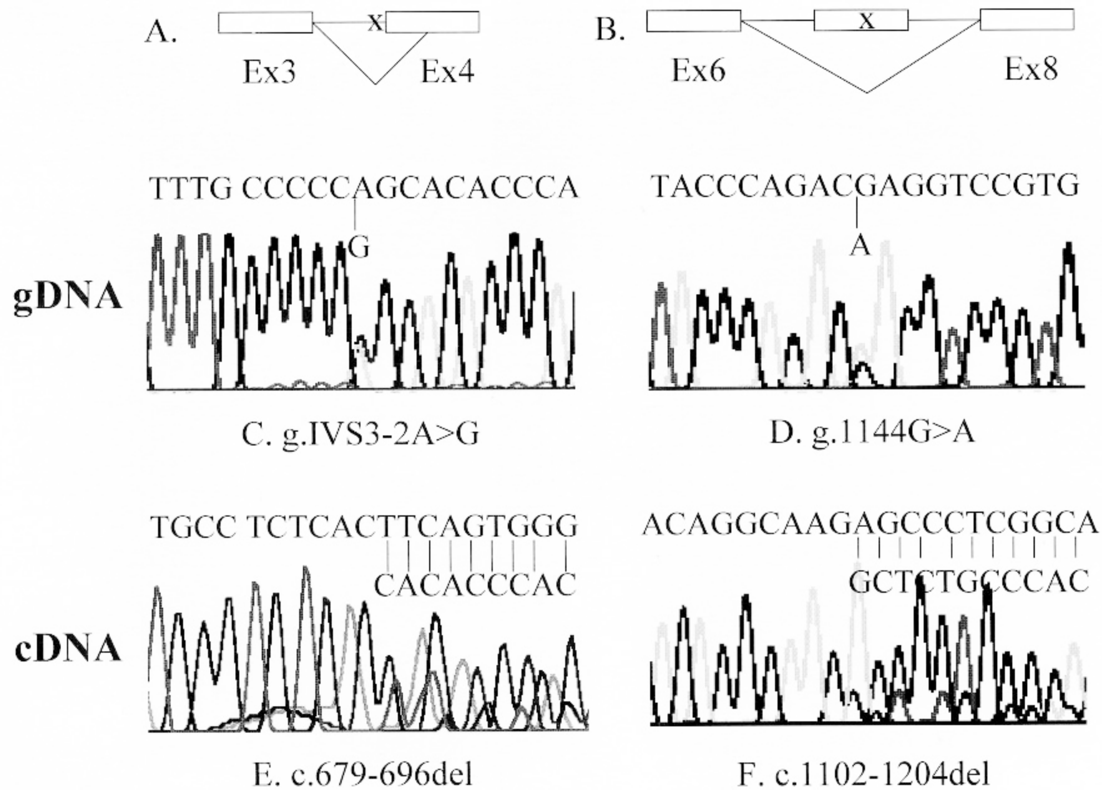

Figure 1. The *ASA* gene analysis of the proband. (A and B) Diagrams showing the locations (X) of the mutations in gDNA and their effects on splicing. Ex stands for exon. (C and D) Partial sequences of the gDNA showing single base pair substitutions. (E and F) Partial sequences of the cDNA showing the start of deletions.

standard extraction method. 3' end of exon 3, intron 3 and 5' end of exon 4 of the *ASA* gene was amplified using primers MLDF3, MLDEx4R and 3' end of exon 6, intron 6, exon 7, intron 7 and 5' end of exon 8 using primers MLDEx7F and MLDR2 as shown in Table I. The products were sequenced directly, as above.

Study of association of pathogenic missense mutations and ESE sites in the ASA gene. Published data on pathogenic missense mutations of the *ASA* gene were used and shown in Table II.

We searched the coding regions of the *ASA* gene for the presence of ESE motifs with ESEfinder software (http://exon.cshl.org/ESE/index.html). To lower the number of false-positive results, we used a more-stringent-than-recommended threshold value of 3.0 for all four types of ESE motifs as recommended in a previous study (9). ESEs in exon/exon boundaries were excluded. We calculated the percentage of sequences that were potential ESEs, which provided us with the proportion of missense mutations expected to be in ESE motifs. We, then, classified reported missense mutations into those in ESE motifs and those do not. Comparison of the observed and expected frequencies of missense mutations in ESE sites was performed using standard Chi-square and p-value by a program available at http://www.unc.edu/~preacher/.

Results

RNA was prepared from leukocytes of the proband and each parent, reverse-transcribed and the cDNAs sequenced. The proband had two heterozygous nucleotide changes: c.679-696del and c.1102-1204del (Fig. 1E and F). Sequence of his father's cDNA showed the former change and of his mother's showed the latter. The genomic DNA of the relevant regions of the proband and his parents were amplified by PCR and directly sequenced. The sequences showed that the proband was a compound heterozygote for IVS3-2A>G and 1144G>A (Fig. 1C and 1D), his father was heterozygous for the former and his mother heterozygous for the latter.

Among the 112 disease-causing mutations reported in the literature (2-6) including one novel IVS3-2A>G mutation identified in this study, 79 mutations are missense mutations (Table II).

Potential ESE motifs found in the *ASA* gene are listed in Table III. After excluding ESEs in exon/exon boundaries, we found potential ESE motifs encompassing 810 bp out of the 1521 bp (53%) of *ASA* coding region. Of the 79 pathogenic missense mutations, 44 are in ESE (56%) (Table II). Using Chi-square test, we found that missense mutations do not colocalize with ESEs (χ^2=0.18, df=1, p=0.67)

Discussion

The proband had typical clinical presentation of MLD. His cDNA was sequenced and showed that he was a compound heterozygote for c.679-696del inherited from his father and c.1102-1204del inherited from his mother. Sequencing of gDNA showed that the c.679-696del was caused by a novel g.IVS3-2A>G mutation. The nucleotide change eliminated the consensus sequence of the splice acceptor site of intron 3. Spliceosome presumably recognized g.695-696AG as a

Table III. ESE motifs found in the *ASA* gene.

SF2/ASF ESE Motifs				SC35 ESE Motifs				SRp40 ESE Motifs				SRp55 ESE Motifs			
Position in ORF				Position in ORF				Position in ORF				Position in ORF			
L	R	Motif	Score	L	R	Motif	Score	L	R	Motif	Score	L	R	Motif	Score
83	89	CCGACGA	4.6	5	12	GGGCACCG	3.0	22	28	CTCCTGG	3.0	53	58	CCCGTC	3.0
121	127	CACCCCA	3.0	14	21	GGTCCCTC	3.6	134	140	CCACTCC	4.6	64	69	AACATC	3.0
158	164	CGGCGGG	3.6	33	40	GGCTGCTG	5.3	176	182	TCACAGA	3.9	187	192	TACGTG	3.8
169	175	CTGCGGT	3.3	45	52	GGCCGTTG	3.8	212	218	CCTCTAG	4.1	195	200	TGTGTC	4.6
179	185	CAGACTT	3.9	75	82	GATCTTTG	3.0	230	236	TGACCGG	3.8	268	273	GGCGTC	3.3
234	240	CGGCCGG	4.2	88	95	GACCTCGG	3.0	301	307	CCCCTGG	3.3	346	351	TACCTC	3.2
241	247	CTCCCGG	3.7	109	116	GGCTGCTA	4.9	350	356	TCACAGG	6.3	422	427	TCCATC	3.2
285	291	CTCCCGG	3.7	121	128	CACCCCAG	3.0	388	394	CCTGAGG	3.2	481	486	TGCTTC	3.8
351	357	CACAGGA	6.0	167	174	GGCTGCGG	3.0	430	436	TTTCTAG	4.2	502	507	TGCGAC	3.5
389	395	CTGAGGG	3.9	173	180	GGTTCACA	3.3	448	454	TCCCACG	3.6	566	571	CGCAGC	3.9
449	455	CCCACGA	4.9	174	181	GTTCACAG	3.7	494	500	CCACTCC	4.6	601	606	TACATG	3.2
585	591	CGGACTA	3.8	181	188	GACTTCTA	4.7	516	522	TGACCAG	3.2	669	674	TGCCTC	3.8
839	845	CAGACAA	3.8	197	204	TGTCTCTG	3.1	538	544	CCACTGT	3.3	714	719	TGCAGA	4.3
882	888	CTCCGGT	3.5	239	246	GGCTCCCG	5.6	573	579	CCCCTGG	3.3	765	770	TGCAGC	4.7
968	974	CTCCCGG	3.7	261	268	GTACCCTG	4.8	589	595	CTAGAGG	3.9	837	842	TGCAGA	4.3
980	986	CCCACGA	4.9	283	290	AGCTCCCG	3.8	600	606	CTACATG	3.6	860	865	TGCGTA	5.8
1092	1098	CACAGGC	3.8	299	306	TGCCCCTG	4.4	673	679	TCTCACC	4.1	864	869	TATGTC	4.0
1139	1145	CAGACGA	6.3	321	328	GGCCGAAG	3.3	675	681	TCACCAC	3.2	898	903	TGTGGA	3.4
1168	1174	CGGACTG	3.0	360	367	GGCCGGCA	3.5	678	684	CCACACC	4.8	1007	1012	TGCTGC	3.0
1199	1205	CCCAGGG	3.5	386	393	GGCCTGAG	3.8	684	690	CCACTAC	4.3	1010	1015	TGCCTA	3.5
1213	1219	CACAGTG	3.1	418	425	GGCTTCCA	4.4	698	704	TCAGTGG	4.4	1082	1087	TGCTGC	3.0
1344	1350	CACCCCA	3.0	428	435	GATTTCTA	4.1	785	791	TGACAGC	4.7	1128	1133	CCCGTC	3.0
1349	1355	CAGAGGT	5.4	460	467	GGCCCCTG	6.5	791	797	CCATAGG	3.6	1166	1171	TGCGGA	5.0
				492	499	GGCCACTC	3.7	833	839	TCACTGC	5.2	1227	1232	TGCAGA	4.3
				528	535	GGTCCCCA	4.8	842	848	ACAATGG	3.5	1245	1250	CGCCTC	3.0
				587	594	GACTAGAG	3.4	888	894	TCTCTTG	3.0	1332	1337	TGTGGC	3.8
				595	602	GCCCGCTA	3.1	948	954	CTTCTGG	4.4	1355	1360	TGCTGC	3.0
				606	613	GGCTTTCG	3.1	997	1003	TCCCTGG	3.7	1401	1406	CGCAGC	3.9
				612	619	CGCCCATG	3.1	1039	1045	CCACTGC	4.9	1414	1419	TTCGGC	3.0
				627	634	GGCCGACG	3.6	1050	1056	TGTCACC	3.3	1454	1459	TGCAGA	4.3
				696	703	GTTCAGTG	3.8	1071	1077	CCTCAGC	4.1	1495	1500	TGCTGC	3.0
				727	734	GGCCGCGG	3.2	1091	1097	GCACAGG	3.8				
				762	769	GGATGCAG	3.3	1125	1131	CTACCCG	3.6				
				776	783	GGACCCTG	5.4	1138	1144	CCAGACG	3.9				
				806	813	GGCTGCTT	3.1	1188	1194	TCACTTC	3.2				
				848	855	GACCTGAG	3.4	1194	1200	CTTCACC	3.5				
				867	874	GTCCCGAG	3.9	1212	1218	CCACAGT	3.6				
				877	884	GGCTGCTC	3.1	1223	1229	CCACTGC	4.9				
				887	894	GTCTCTTG	3.8	1259	1265	TGACTGC	4.4				
				912	919	GACCTACG	3.4	1314	1320	CTACAAC	4.2				
				946	953	GCCTTCTG	3.4	1343	1349	CCACCCC	3.1				
				966	973	CGCTCCCG	3.5	1348	1354	CCAGAGG	4.2				
				978	985	GACCCACG	4.2	1375	1381	CTTCAGC	3.8				
				995	1002	GCTCCCTG	3.8	1396	1402	TTAGACG	4.0				
				1034	1041	GGGCCCCA	3.0								
				1035	1042	GGCCCCAC	3.2								
				1069	1076	GACCTCAG	4.3								
				1076	1083	GCCCCCTG	4.4								

L, left; R, right.

Table III. Continued.

SF2/ASF ESE Motifs				SC35 ESE Motifs				SRp40 ESE Motifs				SRp55 ESE Motifs			
Position in ORF				Position in ORF				Position in ORF				Position in ORF			
L	R	Motif	Score	L	R	Motif	Score	L	R	Motif	Score	L	R	Motif	Score
				1082	1089	TGCTGCTG	3.3								
				1146	1153	GGTCCGTG	5.0								
				1155	1162	GGTTTTTG	3.2								
				1203	1210	GGGCTCTG	3.4								
				1221	1228	TACCACTG	3.6								
				1230	1237	AGACCCTG	3.7								
				1246	1253	GCCTCCAG	3.2								
				1253	1260	GCTCTCTG	3.1								
				1271	1278	AGCCCCCG	4.1								
				1299	1306	GGACCCTG	5.4								
				1341	1348	GGCCACCC	3.1								
				1344	1351	CACCCCAG	3.0								
				1417	1424	GGCCCCAG	5.4								
				1431	1438	GGCCCGGG	3.2								
				1443	1450	GGACCCCG	4.8								
				1510	1517	GATCCCCA	4.4								

L, left; R, right.

splice acceptor site resulting in deletion of the 18 nucleotides at the 5' end of the exon 4 and subsequently deletion of codons 227-332.

The mutant allele inherited from the mother, c.1102-1204del, corresponded to skipping of the whole 103-bp exon 7. Before sequencing his gDNA, we reasoned that the mutation at the gDNA level should involve nucleotides around the exon-intron junctions of either intron 6 or 7. To our surprise, sequencing of the gDNA of the proband and his mother from the 3' end of exon 6 to the 5' end of exon 8 showed only a single base pair substitution, 1144G>A (7). The nucleotide 1144G is in the middle region of exon 7. The 1144G>A mutation was previously reported and predicted to be pathogenic by causing a glutamic acid to lysine substitution at amino acid 382 (E382K) (7). Here we showed that its pathogenic effect is not attributed to its impact on primary amino acid sequence and protein structure, but is splicing-related.

Human genes are typically composed of several exons interrupted by introns. To generate correct mature mRNAs, the exons must be identified and joined together precisely. ESEs are discrete motifs of 6 to 8 nucleotides located inside exons and are target sequences for the family of conserved essential splicing factors - the serine/arginine-rich (SR) proteins (10). ESEs play an important role in exon recognition. Nucleotide substitutions in ESEs can result in failure of SR proteins to bind to the ESE, which leads to failure of spliceosome machinery to recognize the sequence as exon and causes exon skipping (11). Using ESEfinder software to identify potential ESE motifs in the *ASA* gene, we found that the nucleotide 1144G is in the ESE motifs of SF2/ASF and SRp40 SR proteins with scores of 6.3 and 3.9, respectively. The 1144G>A mutation reduces the score of the site being a potential ESE for SF2/ASF SR protein to 3.7 and eliminates it as a potential ESE for SRp40 SR protein. This result suggested that the 1144G>A mutation was pathogenic because it disrupted an ESE motif and caused complete exon 7 skipping. The skipping is expected to result in changes starting from codon 368 downwards and the subsequent changes of 18 amino acids and truncation at amino acid 385 because of the frame shift.

Some previously reported missense mutations in the *ASA* gene were predicted to be pathogenic by disrupting normal splicing because they changed the conserved splice donor or acceptor sequences. The c.459G>C (p.Gln153His) (12), c.973G>A (p.Gly325Ser) (3), c.1101G>C (p.Lys367Asn) (13) mutations were all changes of the last nucleotides of exons 2, 5 and 6, respectively. These have been predicted to cause aberrant splicing due to their changes of the conserved sequences of the exon-intron boundaries. To our knowledge, the 1144G>A mutation is the first missense mutation in the *ASA* gene that its pathogenic effect is splicing-related, caused by disrupting ESE.

Missense mutations disrupting ESEs have previously been reported in several diseases including spinal muscular atrophy (14) and an autosomal dominant growth hormone deficiency (15). The mechanism is even claimed to be the most common one in neurofibromatosis type 1 (16) and hereditary nonpolyposis colorectal cancer (HNPCC) due to mutations in *hMSH2* and *hMLH1* genes (9). This prompted

us to investigate the possibility of other reported missense mutations in the *ASA* gene being deleterious because they disturbed ESEs.

First, we searched the coding regions of the *ASA* gene for the presence of ESE motifs with ESEfinder software, using a threshold value of 3.0 for all four types of ESE motifs. Potential ESEs, excluding ESEs in exon/exon boundaries, are listed in Table III and encompass 53% (810/1521) of the *ASA* coding region. Fifty-six percent (44/79) of reported pathogenic missense mutations are in these potential ESEs. Using Chi-square test, we found that, unlike missense mutations in *hMSH2* and *hMLH1* genes, those in the *ASA* gene do not colocalize with ESEs (p=0.67). This result suggests that pathogenic effects of the majority of missense mutations in the *ASA* gene are not splicing-related but through other mechanisms eg. structural changes, RNA instability or misplacing. The *ASA* gene encodes an enzyme; therefore, its protein structure may have to be more specific than the gene products of the *hMSH2* and *hMLH1* genes. In other words, missense mutations in the *ASA* gene that cause amino acid changes may be as deleterious as those disrupt ESEs. On the contrary, this result may be spurious if the potential ESEs identified by the software are not the real ESEs.

In conclusion, we studied a Thai boy with typical late-infantile MLD and found that he was a compound heterozygote for a novel mutation, g.IVS3-2A>G causing c.679-696del inherited from his father and a missense mutation, g.1144G>A causing c.1102-1204del inherited from his mother. The latter mutation disrupted a potential ESE and caused a complete exon 7 skipping. This is the first missense mutation in the *ASA* gene that is deleterious from disrupting an ESE. However, a study of association between reported pathogenic missense mutations in the *ASA* gene and ESE sites showed that the mutations do not colocalize with ESE sites suggesting that pathogenic effects of majority of them are not splicing-related.

Acknowledgements

We wish to thank Professor Piyaratana Tosukhowong for the assay of ASA activity. This work was supported by grants from the Thailand Research Fund and Chulalongkorn University.

References

1. Von Figura K, Gieselmann V and Jaeken J: Metachromatic leukodystrophy. In: The Metabolic and Molecular Bases of Inherited Disease. 8th edition. Scriver C, Beaudet A, Sly W, Valle D, Childs B, Kinzler K and Vogelstein B (eds). McGraw-Hill, New York, pp3695-3724, 2001.
2. Stenson PD, Ball EV, Mort M, *et al*: Human gene mutation database [HGMD(R)]: 2003 update. Hum Mutat 21: 577-581, 2003.
3. Eng B, Nakamura LN, O'Reilly N, *et al*: Identification of nine novel arylsulfatase a (ARSA) gene mutations in patients with metachromatic leukodystrophy (MLD). Hum Mutat 22: 418-419, 2003.
4. Coulter-Mackie MB and Gagnier L: Spectrum of mutations in the arylsulfatase A gene in a Canadian DNA collection including two novel frameshift mutations, a new missense mutation (C488R) and an MLD mutation (R84Q) in cis with a pseudodeficiency allele. Mol Genet Metab 79: 91-98, 2003.
5. Rafi MA, Coppola S, Liu SL, Rao HZ and Wenger DA: Disease-causing mutations in cis with the common arylsulfatase A pseudodeficiency allele compound the difficulties in accurately identifying patients and carriers of metachromatic leuko-dystrophy. Mol Genet Metab 79: 83-90, 2003.
6. Regis S, Corsolini F, Ricci V, Di Duca M and Filocamo M: An unusual arylsulfatase A pseudodeficiency allele carrying a splice site mutation in a metachromatic leukodystrophy patient. Eur J Hum Genet 12: 150-154, 2004.
7. Barth ML, Fensom A and Harris A: Missense mutations in the arylsulphatase A genes of metachromatic leukodystrophy patients. Hum Mol Genet 2: 2117-2121, 1993.
8. Percy AK and Brady RO: Metachromatic leukodystrophy: diagnosis with samples of venous blood. Science 161: 594-595, 1968.
9. Gorlov IP, Gorlova OY, Frazier ML and Amos CI: Missense Mutations in hMLH1 and hMSH2 Are Associated with Exonic Splicing Enhancers. Am J Hum Genet 73: 1157-1161, 2003.
10. Hastings ML and Krainer AR: Pre-mRNA splicing in the new millennium. Curr Opin Cell Biol 13: 302-309, 2001.
11. Cartegni L, Chew SL and Krainer AR: Listening to silence and understanding nonsense: exonic mutations that affect splicing. Nat Rev Genet 3: 285-298, 2002.
12. Tsuda T, Hasegawa Y and Eto Y: Two novel mutations in a Japanese patient with the late-infantile form of metachromatic leukodystrophy. Brain Dev 18: 400-403, 1996.
13. Draghia R, Letourneur F, Drugan C, *et al*: Metachromatic leukodystrophy: identification of the first deletion in exon 1 and of nine novel point mutations in the arylsulfatase A gene. Hum Mutat 9: 234-242, 1997.
14. Cartegni L and Krainer AR: Disruption of an SF2/ASF-dependent exonic splicing enhancer in SMN2 causes spinal muscular atrophy in the absence of SMN1. Nat Genet 30: 377-384, 2002.
15. Moseley CT, Mullis PE, Prince MA and Phillips JA III: An exon splice enhancer mutation causes autosomal dominant GH deficiency. J Clin Endocrinol Metab 87: 847-852, 2002.
16. Ars E, Serra E, Garcia J, *et al*: Mutations affecting mRNA splicing are the most common molecular defects in patients with neurofibromatosis type 1. Hum Mol Genet 9: 237-247, 2000.