ONLINE MUTATION REPORT

Mucopolysaccharidosis IVA: identification of mutations and methylation study in GALNS gene

S Tomatsu, T Nishioka, A M Montaño, M A Gutierrez, O S Pena, K O Orii, W S Sly, S Yamaguchi, T Orii, E Paschke, S G Kircher, A Noguchi

J Med Genet 2004;**41**:e98 (http://www.jmedgenet.com/cgi/content/full/41/7/e98). doi: 10.1136/jmg.2003.018010

Mucopolysaccharidosis IVA (MPS IVA; Morquio A disease) is an autosomal recessive disorder caused by a deficiency of the lysosomal N-acetylgalactosamine-6-sulfate sulfatase (GALNS; E.C.3.1.6.4; OMIM# 253000). GALNS is one of the sulfatases required to degrade glycosaminoglycans (GAGs), keratan sulfate (KS), and chondroitin-6-sulfate (C6S). GALNS deficiency results in lysosomal storage disease. As in other mucopolysaccharidoses, MPS IVA patients have a broad spectrum of clinical severity. Phenotypes vary from the classical form with severe bone dysplasia (spondyloepiphyseal dysplasia), short trunk dwarfism, coxa valga, odontoid hypoplasia, corneal opacity, and a life span of 20–30 years, to a milder (attenuated) form. The patients with a milder form may have a normal quality of life and mild bone and visceral organ involvement. The broad spectrum of clinical phenotypes seen in MPS IVA is presumed to be generated by multiple different GALNS mutations. Investigations of the molecular nature of MPS IVA have been facilitated by purifying the enzyme, and by isolating and characterising both the GALNS gene and the full-length cDNA encoding the human GALNS protein.[1–3] The cDNA contains an open reading frame of 1566 bp which encodes a 522-residue polypeptide. The gene spans approximately 50 kb, and contains 14 exons (GDB Accession ID: 129085).

Molecular analyses of over 100 MPS IVA patients with diverse ethnic or geographic origins (Australian, British, Colombian, Finnish, German, Irish, Japanese, and Turkish) have been reported.[4–11] To date, around 90 different mutations responsible for various phenotypes have been identified, documenting the allelic heterogeneity in the GALNS gene, which correlates with the clinical variability within MPS IVA. Most of the known mutations (85–95%) in the GALNS gene have been detected by using PCR of genomic DNA for each exon (14 amplicons) and single-strand conformation polymorphism (SSCP),[6–12] a reliable screening method.

The variety (missense or nonsense), frequency, and location of point mutations causing human genetic disease are highly non-random. This non-randomness at the DNA level has been attributed to variations in the local DNA sequence environment, for example, the presence or absence of CpG dinucleotides. DNA methylation at the cytosine residue of CpG dinucleotides produces 5-methylcytosine, which results in a cytosine-to-thymine transitional change via methylation-mediated deamination. That CpG methylation in mammalian cells has important implications in the etiology of genetic diseases is suggested by the fact that 10–60% of point mutations causing human diseases result from transitions at CpG dinucleotides.[13 14] Distribution of CpG mutations within a gene is uneven. This may be explained by differences in methylation in germ-line DNA of each gene, differences among exons of the same gene, and variable numbers of CpG sites in different genes.

Transitions at CpG dinucleotides account for over 20% of point mutations that cause MPS IVA phenotypes and two polymorphisms. However, no transition at CpG sites has been detected in exons 1, 9, 12, or 14, another example of the non-random distribution of the CpG transitions. Although recent advances have led to a better understanding of the methylation pattern in the 5′-region of housekeeping genes, few studies have been reported on the methylation pattern of an entire gene.[15]

The present study applied a new strategy using long genomic PCR with five amplicons followed by direct sequencing to sequence the complete coding regions of the GALNS gene. We report mutation analyses in 28 patients from various ethnic populations. In addition, analysis of the methylation pattern at CpG cytosines of the GALNS gene allowed us to clarify the correlation between transitions at CpG sites and the methylation status of CpG sites.

Key points

- Mucopolysaccharidosis type IVA (MPS IVA) is an autosomal recessive disorder caused by deficiency of lysosomal N-acetylgalactosamine-6-sulfate sulfatase (GALNS: OMIM# 253000).
- Mutation screening by long genomic PCR with five amplicons and direct sequence analyses in 28 MPS IVA patients was performed and the methylation pattern of the GALNS coding region including 78 CpG sites was analysed using a sensitive bisulfite-based technique.
- We identified 32 different mutations. Transitional mutations at CpG dinucleotides accounted for 23% of all single base substitutions leading to missense and nonsense mutations in the coding region (10 of 44 alleles). Methylation of individual CpG cytosines was extensive within exons 2–14 while CpG cytosines in exon 1 were completely unmethylated. All transitional mutations at CpG sites were located between exons 2 and 14. Non-methylation at CpG sites correlated with the absence of transitional mutations in exon 1.
- Our data provide further evidence for extensive allelic heterogeneity in GALNS mutations and provide another example of the correlation between methylation status of CpG sites and distribution of transitional mutations.

Abbreviations: C6S, chondroitin-6-sulfate; GAGs, glycosaminoglycans; GALNS, N-acetylgalactosamine-6-sulfate sulfatase; KS, keratan sulfate; SSCP, single-strand conformation polymorphism

METHODS

Patients

MPS IVA patients were diagnosed at each collaborating institute and the Department of Pediatrics, Saint Louis University (St. Louis, MO). Twenty-eight MPS IVA patients were analysed. GALNS deficiency was demonstrated in all cases by enzyme assay. Plasma and urine KS concentrations were measured as described previously, and the reagents were obtained from Seikagaku (Tokyo, Japan).[16] The subjects studied included 16 Americans, four Canadians, two Austrians, two Pakistanis, two Japanese, one Mexican, and one New Zealander. Clinical data and the racial background of each patient are reported in table 1. Peripheral blood samples were collected from the patients after obtaining an informed consent at each collaborating institute.

Mutation analysis of GALNS gene

Genomic DNA was extracted by a standard method. Five sets of primers covering all 14 exons and their exon–intron boundaries were assigned to produce five amplicons (0.3–5.8 kb in length). In each amplicon (F1–F5), 200 ng of genomic DNA were mixed with a set of primers in a total volume of 50 µl, including dNTPs, DTT, dimethyl sulfoxide (DMSO), and Taq polymerase. Except for F1 (0.3 kb) for exon 1, PCR fragments included multiple exons, amplified by a long PCR amplification method according to the manufacturer's instructions (Gibco BRL, Rockville, MD) with the addition of 2.5–5% of DMSO or 1 µl Perfect Match (Stratagene, La Jolla, CA). All PCR amplification reactions were performed on a Perkin-Elmer 9700 Thermal Cycler. The PCR products were directly sequenced using fluorescent-labelled dideoxynucleotides (ABI, Foster City, WA). The positions of PCR primers and the sizes of PCR product are shown in fig 1 together with the picture of the agarose gel. The sets of primers for genomic PCR are as follows:

- F1 forward (GMO75): 5′ gccccactggtcacgaggcagtcca 3′;
- F1 reverse (GMO11R): 5′ cccaccccggccctgccccgtcc caccgcccgcactca 3′;
- F2 forward (GMO4): 5′ tggcttccacggtccccgacacgctctgg cacca 3′;
- F2 reverse (GMO5R): 5′ atgtcccttggaaccaaggccaggaagtg gatgga 3′;
- F3 forward (GMO13): 5′ gtggtatctgttgctgctcagaacttcc gagtgtc 3′;
- F3 reverse (GMO7R): 5′ cgcacacaccctgggatggctgcaggcctg gacct 3′;
- F4 forward (MO25): 5′ tgtgagcatgtatgcatatctgtagaccca 3′;
- F4 reverse (GMO9R): 5′ ccaccaagcacgtgtgggtatgaatagcaa cagca 3′;
- F5 forward (GMO10): 5′ agtgactgctcactgtggttctcagcccgtta gag 3′;
- F5 reverse (TOMF14R): 5′ ctgcgtctgcaggtgctgtctgtctggctt 3′.

Methylation study of the GALNS coding region

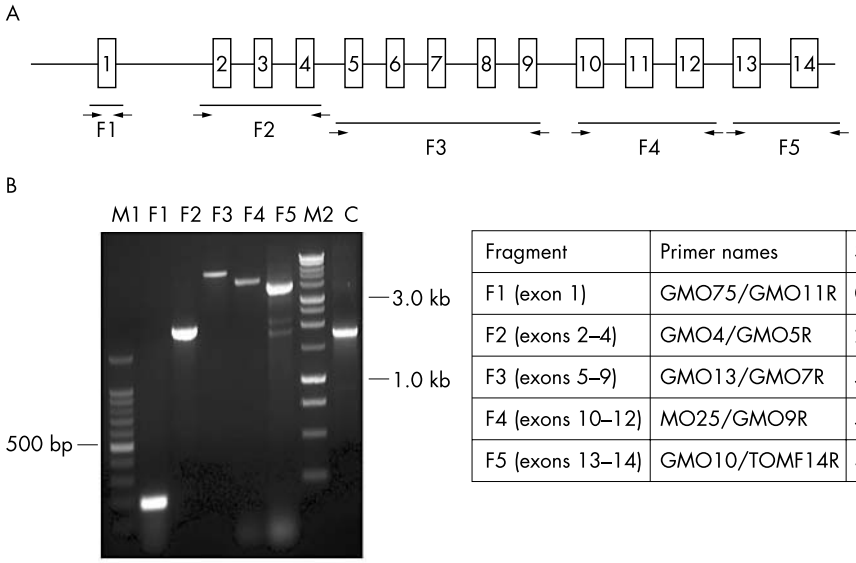
Sodium bisulfite and sequencing

Sodium bisulfite catalyses the conversion of cytosine to uracil residues in single-stranded DNA, whereas methylated cytosines remain unreactive under these conditions. Upon PCR amplification of the genomic region of interest, the converted uracil residues are replicated as thymines instead of cytosines. A remaining cytosine in the sequence of the PCR product, therefore, demonstrates that this site is methylated on the template DNA.

Table 1 Clinical, biochemical, and genetic data of the MPS IVA patients

Case	Sex	Phe*	Age (years)	Present weight (kg)	Present height (cm)	Stop growing	Surgery†	Allele 1	Allele 2	Racial background	Plasma KS‡	Urine KS§
1	M	S	13.4	20.9 (−3.8 SD)	106.7 (−7 SD)	+	+	I113F	R361G	British/German**	NA	NA
2	M	S	10.7	18.1 (−3.5 SD)	92.7 (−7.1 SD)	+	+	c.899-2A>G	ND	Mexican/Lebanese**	1388	13.6
3	F	S	37	24.9 (−5.1 SD)	106.7 (−9.1 SD)	+	+	600 Gdel	ND	Caucasian	NA	NA
4	F	S	9.5	31.7 (0.2 SD)	107 (−4.9 SD)	+	+	S287L	L352P	Irish/Italian/Polish**	NA	NA
5	M	S	14.5	27.4 (−3.1 SD)	119.4 (−5.9 SD)	+	−	G301C	ND	French††	802	3.3
6	F	M	11.5	30.4 (−1.6 SD)	110.4 (−5.3 SD)	+	+	M391V	Q338X	Irish/German/ Swedish**	760	19.4
7	F	M	27	46.7 (−1.7 SD)	146.6 (−2.7 SD)	+	+	T312S	154–165del	Irish	NA	NA
8¶	M	S	20	27 (−5.1 SD)	92.7 (−12.4 SD)	+	+	G47R	G47R	Honduran**	NA	NA
9	M	S	10	13 (−9.1 SD)	90 (−7.6 SD)	+	+	A291T	c.758+1G>C	Japanese	379	13.9
10	F	S	12.5	22.6 (−3.6 SD)	119.3 (−5.3 SD)	+	+	G155E	ND	Dutch/Irish/ Scottish**	899	12.8
11	F	S	17	22.7 (−6.2 SD)	95 (−11.6 SD)	+	+	G301C	G301C	French**	617	3.5
12	M	S	9.5	20 (−2.8 SD)	104.3 (−5.1 SD)	+	+	M41L	M41L	Acadian**	629	NA
13	M	S	14.2	15.9 (−5.2 SD)	88.9 (−9.5 SD)	+	+	R376Q	I113F	Irish**	NA	NA
14	F	S	6	14 (−2.8 SD)	92 (−4.8 SD)	+	+	R90W	G340D	Austrian	NA	NA
15	M	S	7.7	15.8 (−3.3 SD)	88.9 (−7.1 SD)	+	+	c.898+1G>C	c.898+1G>C	Czech/German/ Irish/Polish**	NA	NA
16	M	S	8.2	13.6 (−4.3 SD)	86.4 (−8.4 SD)	+	−	D233N	L345P	German/Polish**	NA	NA
17	M	M	22.3	54.5 (0.5 SD)	129.5 (−5.4 SD)	+	−	G139S	G247D	Irish/German**	NA	NA
18	F	M	10.6	22.2 (−2.6 SD)	122 (−3 SD)	−	+	M391V	M391V	French††	582	13.1
19	M	M	8.1	31.7 (0.0 SD)	123 (−1.0 SD)	−	−	M391V	M391V	French††	577	6
20	M	S	16.7	27 (−4.7 SD)	111 (−9.1 SD)	+	+	R386C	ND	Mexican	NA	8.7
21	M	M	38	50 (−1.2 SD)	142 (−5 SD)	+	+	R94G	N204K	Japanese	310	0.9
22	F	S	26.5	24.4 (−5.1 SD)	106.6 (−9.1 SD)	+	+	H166Q	ND	Irish/Danish††	714	4.7
23	M	S	NA	NA	NA	NA	NA	P77R	P77R	Caucasian**	NA	NA
24	M	M	7.6	26.3 (0.7 SD)	119.5 (−1.2 SD)	−	−	H398D	H398D	Pakistani	NA	NA
25	F	M	24	40 (−2.7 SD)	140 (−3.8 SD)	+	−	H398D	H398D	Pakistani	NA	NA
26	M	M	18	48 (−2.5 SD)	165 (−1.6 SD)	+	−	S287L	S295F	Austrian	NA	NA
27	M	M	24.7	47.7 (−2.7 SD)	127 (−7.4 SD)	+	−	M41L	M41L	Acadian**	422	0.4
28	M	S	10	40.8 (−2.2 SD)	99.1 (−6.5 SD)	+	+	F69V	G301C	Colombian**	NA	NA

*Phe: phenotype: m, milder (attenuated); s, severe; †surgery: cervical fusion, knee, hip, and leg surgery involving the bone deformity; ‡plasma KS: normal value KS (20–240 ng/ml); §urine KS: normal value KS (0–0.17 mg/g creatinine); ¶consanguinity; **US resident; ††resident in Canada.
Patients 7 and 8 and patients 9 and 10 are sibling cases. NA, not available; ND, not determined.



Fragment	Primer names	Size
F1 (exon 1)	GMO75/GMO11R	0.3 kb
F2 (exons 2–4)	GMO4/GMO5R	2.0 kb
F3 (exons 5–9)	GMO13/GMO7R	5.8 kb
F4 (exons 10–12)	MO25/GMO9R	5.0 kb
F5 (exons 13–14)	GMO10/TOMF14R	3.7 kb

Figure 1 Positions of PCR primers and size of the PCR product. (A) Locations of PCR fragments and primers used for genomic PCR amplification of human GALNS gene. The arrows represent the position of each primer and the horizontal bar shows the position of the respective PCR fragment. The numbered boxes represent each exon. (B) Gel electrophoresis of genomic PCR fragments. Lane indicated by: M1, 100 bp marker; F1, exon 1; F2, exons 2–4; F3, exons 5–9; F4, exons 10–12; F5, exons 13–14; M2, 1.0 kb marker; C, cDNA of GALNS. The names of the primers and the sizes of the PCR product are described in the right panel.

Normal genomic DNAs from two males and two females were isolated from peripheral blood samples. The DNA samples were used to determine the distribution of cytosine methylation at the GALNS gene locus covering the coding region (exons 1–14) and each exon–intron boundary. Bisulfite conversion of genomic DNA was performed using the protocol described previously.[15 17] Briefly, genomic DNA was digested overnight with *Eco*RV, boiled, and denatured by adding freshly prepared 3 M NaOH to a final concentration of 0.3 M and incubated at 42℃ for 30 min. Final concentrations of 3.4 M sodium bisulfite and 1 mM hydroquinone were added to the denatured DNA. The DNA was gently mixed in this sodium bisulfite/hydroquinone solution and incubated at 55℃ for 6 h. Unbound bisulfite was removed from the DNA using microspin S-200HR columns (Pharmacia Biotech, Piscataway, NJ). The purified DNA sample was subsequently mixed and incubated with freshly prepared NaOH (0.3 M final concentration) at 37℃ for 20 min. NaOH was removed using microspin S-200HR columns and the flow-through (about 100 µl) contained the converted DNA ready for amplification.

PCR conditions

PCR amplifications were performed as previously described[15] in a 50-µl reaction mixture containing 2–5 µl bisulfite-treated genomic DNA, 200 µM dNTPs, 1 µM primers, 10 mM Tris-HCl, pH 8.3, 1.5 mM $MgCl_2$, and 0.4 µl Taq 2000 DNA polymerase (Stratagene, La Jolla, CA). The amplification was performed under the following conditions: 94℃ for 5 min for one cycle; 95℃ for 45 s, 60–65℃ for 30 s, 72℃ for 45–90 s for 35 cycles; 72℃ for 7 min for one cycle. The PCR products were purified using a column purification kit (Roche, Indianapolis, IN). The second PCR proceeded in a 2–4-µl reaction mixture from the first PCR, 200 µl dNTPs, 10 mM Tris-HCl (pH 8.3), 1.5 mM $MgCl_2$, 1 µM primers, and 0.4 µl Taq 2000 polymerase. The primers used were a new set of two nested primers. Namely, four sets of primers were synthesised to obtain one fragment and PCR amplifications were performed for both strands by using each strand-specific primer. To include exons 1–14, 14 PCR fragments were prepared for one strand (56 primers of PCR amplification for each strand) and in total 112 different primers were used for PCR amplification (fig 2). The primer sequences designed here are available upon request. The final PCR fragments generated with the nested primer pair were isolated and ligated into easy T-vector (Promega, Madison, WI). After transformation, five randomly picked clones (in total 20 independent clones for each fragment) were sequenced using fluorescent-labelled dideoxynucleotides (ABI, Foster City, WA). The status of methylation at individual CpG sites was determined from the frequency of clones having C at each site in the sequence analysis.

RESULTS

Biochemical and physical diagnoses of the MPS IVA patients

In this study, we have analysed the GALNS gene of 28 MPS IVA patients from various ethnic populations using long genomic PCR and a direct sequencing method. The clinical characteristics, biochemical data, and identified genotypes of the patients included in this study are summarised in table 1. On physical examination, all patients were found to have growth retardation, joint laxity, and bone dysplasia and some were found to have heart valvular disease and deafness (data not shown). The final heights of the patients, who had all stopped growing, were between 86.4 cm (−8.4 SD) and 165 cm (−1.6 SD). All patients were neurologically normal. Ten patients had a milder form of the disease, while the other 18 had a severe form based upon physical examination and biochemical analysis. The plasma KS concentrations in MPS IVA patients examined were between 310 and 1388 ng/ml. These values were higher than those in age-matched controls (20–240 ng/ml) (table 1).[16] Urine KS concentrations in MPS IVA patients (0.94–19.4 mg/g creatinine) were also significantly higher than those of age-matched controls (0.03–0.67 mg/g creatinine).[16]

Genomic PCR and sequence analyses of the GALNS gene

The PCR products (F1–F5) of the GALNS gene were amplified from genomic DNA (fig 1). The complete coding region and

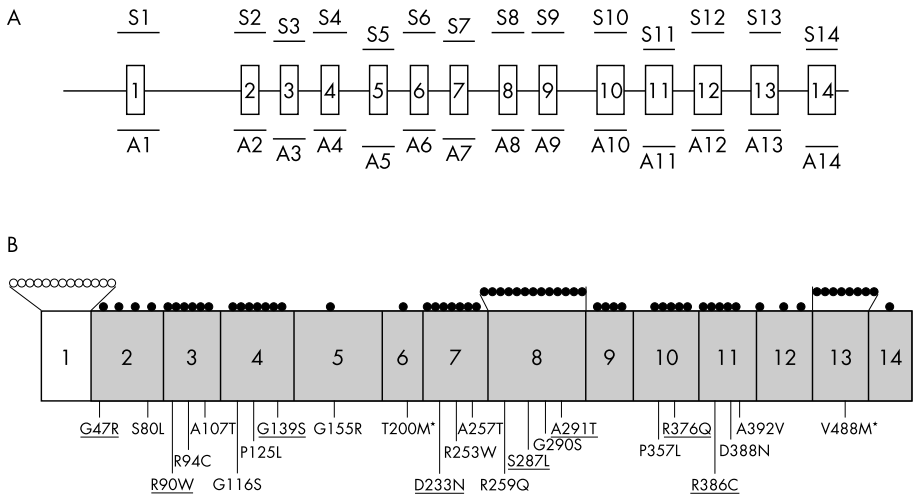


Figure 2 Locations of PCR fragments used for study of the methylation status of the human GALNS gene. (A) Locations of PCR fragments used in genomic sequencing analysis of methylation status of the human GALNS gene. The numbered boxes represent each exon and the horizontal bars show the positions of respective PCR fragments. The bars above and below the numbered exons indicate the positions of the PCR fragments used for sequencing following bisulfite treatment, which converts unmethylated Cs to Ts. The upper and lower bars indicate the sense strand (S1–S14) and the antisense strand (A1–A14), respectively. The primer sequences used for the methylation study are provided upon request. (B) Methylation pattern of the coding region and the transitional mutations sites at CpG dinucleotides. The numbered boxes represent each exon and the vertical bars show the positions of the respective transitional mutations at the CpG site. The mutations found out in this study are underlined. A schematic representation of methylation profiles across the human GALNS coding region is shown. The open box for exon 1 indicates unmethylated DNA sequences, and the shadow boxes for exons 2–14 indicate methylated sequences. The methylation profiles described reflect the data obtained from PCRs (S1–S14 and A1–A14). The open circles indicate the unmethylated CpG sites and the filled circles show the methylated sites. The asterisks indicate the polymorphism.

exon–intron boundaries of the GALNS gene were analysed by direct sequencing. In all, 32 different mutations were identified in 28 MPS IVA patients (table 2). Thirteen of the mutations (M41L, 209-220del, F69V, G155E, 655delG, D233N, c.813+1G>C, c.953+1G>C, c.954-2A>G, G340D, L345P, L352P, and H398D) were described here for the first time in MPS IVA patients. The other 19 mutations (G47R, P77R, R90W, R94G, I113F, G139S, H166Q, N204K, G247D, S287L, A291T, S295F, G301C, T312S, Q338X, R361G, R376Q, R386C, and M391V) were previously reported[4 6–12 18]. Sixteen out of 19 reported mutations were observed in multiple ethnic groups, while I113F and T312S or N204K were found only in British/Irish or Japanese subjects, respectively. In this study, no mutation accounted for over 10% of all alleles investigated. None of these mutations was found in over 100 normal control chromosomes. Control DNAs were of European ancestry.

Fifty of the 56 disease alleles were defined and accounted for 88% of the mutant alleles in this group of MPS IVA patients. Eight out of 27 missense and nonsense mutations (30%) resulted from transitions at CpG dinucleotide sites. When the recurrent mutational event was added, the transitional mutations at CpG dinucleotides accounted for 23% of all point mutations (10 of 44 alleles).

The severity of each mutation depends on its location, the nature of the amino acid change, and the effect that it has on the catalytic activity of the protein. Residues conserved among different sulfatases might be expected to be more important and mutations in these residues found more often in MPS IVA patients. In this study, 13 out of 26 missense mutations occurred in conserved amino acid residues among sulfatase genes. Of the 26 missense mutations, 19 had a non-conservative amino acid change (table 2). Two previously reported polymorphisms (A231G and A393S) were found among the MPS IVA patients investigated.

A phenotype–genotype relation analysis was attempted, taking into account the clinical phenotype of the patient, homozygosity for the mutation, the amount of residual enzyme activity of the patient's fibroblast (or leukocyte), and the amount observed following in vitro mutagenesis and expression studies described previously. From this analysis, 23 mutations (17 missense, three splicing, two deletions, and one nonsense mutation) were associated with a severe phenotype, while eight missense mutations were associated with a milder phenotype (table 2).

Methylation status at CpG sites in the coding region and correlation with transitional events at CpG sites (fig 2)

In four individuals analysed in this study, complete methylation was observed from exons 2 through 14 (65 CpG sites) while the region of exon 1 (13 CpG sites) was completely unmethylated. The methylation status of the sense strand correlated well with that of the antisense strand, suggesting symmetrical methylation in the coding region. We could not find any correlation between the methylation status and sex. Altogether, a total of 90 missense and nonsense mutations were identified in MPS IVA patients in the previous and current studies. Among them, 21 mutations (23.3%) resulted from transitions at CpG sites (fig 2). In addition, two out of seven polymorphisms that changed an amino acid residue derived from transitions at CpG dinucleotide (V488M and T200M).[6 7] Taken together, all 23 CpG sites involving transitions were in exons with complete methylation and distributed throughout the coding region except for exons 1, 9, 12, and 14. Thirteen CpG dinucleotides in exon 1 comprised 16.7% of total CpG dinucleotides. Despite occupying the largest fraction of CpG dinucleotides in all exons, no transitional event at a CpG site in exon 1 has been found. Paucity of transitions at CpG sites was also observed in exons 9, 12, and 14 with complete methylation. The number of CpG sites for these exons was four, three, and one, respectively.

DISCUSSION

The aim of this study was to characterise the MPS IVA mutations in various ethnic populations and to examine the correlation between transition events and the methylation

Table 2 Mutations detected in GALNS gene of 28 MPS IVA patients

Base change*	aa change‡	Exon/intron	Conserved aa residue	Conservative change	Population identified§	Phenotype**	Frequency (%)††
c.121A>T	M41L	E2	Conserved	Conservative	Acadian	Milder	7.1
c.139G>A†	G47R‡‡	E2	Conserved	Non-conservative	Honduran, Po	Severe	3.6
c.154_165del	4 aa deletion	E2			UK	Severe	1.8
c.206T>G	F69V	E2	Conserved	Non-conservative	Co	Severe	1.8
c.230C>G†	P77R	E2	Conserved	Non-conservative	Indian, USA	Severe	3.6
c.268C>T†	R90W‡‡	E3	Conserved	Non-conservative	Au, Pa	Severe	1.8
c.280C>G†	R94G	E3	Conserved	Non-conservative	Ja, Po	Undefined	1.8
c.337A>T†	I113F	E4	Non-conserved	Conservative	UK	Severe	3.6
c.415G>A†	G139S‡‡	E4	Conserved	Conservative	Ar, Br, UK, USA	Severe	1.8
c.464G>A†	G155E	E5	Conserved	Non-conservative	USA	Severe	1.8
c.478C>G†	H166Q	E5	Non-conserved	Non-conservative	Ca, UK	Severe	1.8
600delG	Premature stop	E6			New Zealand	Severe	1.8
c.612C>G†	N204K	E6	Non-conserved	Non-conservative	Ja	Milder	1.8
c.697G>A	D233N‡‡	E7	Non-conserved	Non-conservative	Ge	Severe	1.8
c.740G>A†	G247D	E7	Conserved	Non-conservative	Ge, No, USA	Milder	1.8
c.758+1G>C	IVS 7+1 G>C	I7			Ja	Severe	1.8
c.860C>T†	S287L‡‡	E8	Non-conserved	Non-conservative	Au, Po, USA	Severe	3.6
c.871G>A†	A291T‡‡	E8	Non-conserved	Conservative	Finnish, Ja, Pa	Severe	1.8
c.884C>T†	S295F	E8	Non-conserved	Non-conservative	Au, Dutch	Milder	1.8
c.898+1G>C	IVS8+1G>C	I8			Caucasian	Severe	3.6
c.899-2A>G	IVS8-2A>G	I8			UK	Severe	1.8
c.901G>T†	G301C	E9	Conserved	Non-conservative	Co, Ma, Port, UK	Severe	3.6
c.935C>G†	T312S	E9	Non-conserved	Conservative	UK	Milder	1.8
c.1012C>T	Q338X	E10		Non-conservative	No, USA	Severe	1.8
c.1019G>A	G340D	E10	Non-conserved	Non-conservative	Au	Severe	1.8
c.1034T>C	L345P	E10	Conserved	Non-conservative	Ge	Severe	1.8
c.1055T>C	L352P	E10	Conserved	Non-conservative	USA	Milder	1.8
c.1082G>T†	R361G	E10	Conserved	Non-conservative	Ge, UK	Severe	1.8
c.1127G>A†	R376Q‡‡	E10	Non-conserved	Conservative	Ge, UK	Severe	1.8
c.1156C>T†	R386C‡‡	E11	Non-conserved	Non-conservative	Ar, Br, Ch, Co, Ge, It, Ja, Me, Tu, UK, USA	Severe	1.8
c.1171A>G†	M391V	E11	Non-conserved	Conservative	Ca, UK, USA	Milder	8.9
c.1192C>G	H398D	E11	Non-conserved	Non-conservative	Pa	Milder	7.1

*Numbered according to standard nomenclature and +1 is the adenine of the initiation codon, Met; †recurrent mutation; ‡the novel mutation is written in italic; §population in which the mutation was identified: Ar, Argentine; Au, Austrian; Br, Brazilian; Ca, Canadian; Ch, Chilean; Co, Colombian; Fr, French; Ge, German; It, Italian; Ja, Japanese; Ma, Moroccan; Me, Mexican; No, Norwegian; Pa, Pakistani; Po, Polish; Port, Portuguese; Tu, Turkish; UK, British/Irish; USA, resident in USA; **phenotype: clinical severity associated with each mutation; ††allele frequency of each mutation in this study; ‡‡the mutation of a C to T transition in the CpG dinucleotide; aa; amino acid.

status of CpG sites of the GALNS gene. We have found 32 different mutations including 13 novel mutations. Our studies support the extensive allelic heterogeneity of MPS IVA and also show that the methylation status of the GALNS gene affects the distribution of transitional mutations at CpG sites.

Missense and nonsense mutations constitute the largest portion (82 out of 102 different mutations; 80%) of known alterations in the GALNS gene causing the MPS IVA phenotype. While identification of 13 novel mutations in this study documents the extensive allelic heterogeneity of MPS IVA, around 60% of mutations occurred recurrently. Among the recurrent mutations in this study, the most prevalent are R386C, G301C, I113F, T312S, and A291T. These five account for 8.0, 6.5, 6.5, 2.1, and 1.8%, respectively, of all missense and nonsense mutations compiled in MPS IVA.[4 5 8 11] It is difficult to determine whether these apparently recurrent mutations are identical-by-descent or truly recurrent since the origin of most of these mutations remains unclear. However, I113F and T312S mutations have been defined only in British/Irish population and 80% of G301C are identified in Colombian patients, while R386C and A291T have been detected randomly in various ethnic populations. I113F, T312S, and G301C are derived from transversion at non-CpG sites, while R386C and A291T are from transitions at CpG sites. These mutations that are population specific could be associated with a common founder effect, while the others could be due to mutational recurrences. The relationship between mutation types and clinical phenotypic features is influenced by multiple factors including the extent of the evolutionary conservation of the amino acid residue and the nature of the amino acid exchange. Half of missense mutations (13 out of 26) occurred on amino acid residues conserved among human sulfatase genes. Nineteen out of 26 (73%) missense mutations derived from a non-conservative (large) amino acid change (table 2). One might expect that a dramatic change of amino acid structure is more likely to cause disease.[19–21]

Of the total number of point mutations in the coding region that we have compiled in MPS IVA patients, 29.1% occur by transitions at the CpG dinucleotides. This percentage is slightly lower than the compiled data from many genes described previously,[13 14] but represents an 18-fold higher probability of transitional mutations at these sites compared with the expected value for the number of CpG sites. Fourteen out of 21 transitional mutations at CpG sites were recurrently observed across different ethnic groups.

The primary DNA sequence is obviously an important factor in determining the location and frequency of point mutations. One major mechanism for the origin of point mutations causing human genetic diseases has been postulated as 5-methylcytosine deamination at CpG dinucleotides, which leads to C-to-T (or G-to-A) conversion. The mutation rate can be also calculated for transitions at other sites. Frequency of transitions at CpG cytosines of the GALNS gene including recurrent mutations (81 mutations in 156 cytosines in both strands) is increased 5.2 times compared with that of transitions at other cytosines (80 mutations in 796 cytosine sites in both strands). As a result, transitions at the 78 CpG dinucleotides account for around 50% of all the transitions. Thus, whatever the roles of DNA methylation in X-inactivation and regulation of gene expression, we pay a heavy price in terms of the increased mutability of the CpG sites, since the deamination of 5-methylcytosine in CpG dinucleotides leads to C-to-T (or G-to-A) transitional

mutations. The methylated CpG dinucleotides are more prone to deamination, whereas those that are unmethylated will be much less likely to mutate to thymidine. Although methylation patterns between and within different genes vary widely, this all-or-none mutational susceptibility conferred by cytosine methylation is well illustrated by analysis of CpG sites within the GALNS gene.

We tested the hypothesis that the hypermutability of the CpG dinucleotides in the GALNS gene is caused by methylation-mediated deamination using bisulfite-treated genomic sequencing to identify methylated CpG sites. No disease-associated transitions were found at CpG dinucleotides in exon 1 in MPS IVA patients although several mutations were identified at other dinucleotides. The methylation status correlated well with the distribution of occurrence of transitions at CpG dinucleotides. For instance, the paucity of transitional events at CpG dinucleotides in exon 1 correlates with the finding of complete non-methylation of exon 1 cytosines. On the other hand, all transitions at CpG sites including 21 disease-causative mutations and two polymorphisms are located in exons 2–14, which are methylated completely. Thus, most excessive transitions in MPS IVA mutations can be attributed to the hypermutability of the CpG dinucleotide. The current results concur with a previous report showing that non-methylation of exon 1 of the human GUS gene correlates with the absence of transitional mutations at CpG sites in exon 1[15] although we can not exclude the possibility that mutations in that part of the gene do not significantly affect the amount or activity of enzyme produced. Absence of transitions at CpG sites in exons 9, 12, and 14, which are also methylated, may be attributed both to the few CpG sites in those exons and to the paucity of conserved amino acid residues with a CpG cytosine.

In conclusion, this study further documents the extent of the allelic heterogeneity of MPS IVA in various ethnic groups, and provides evidence of the importance of the methylation status at CpG sites in the GALNS gene as regards the occurrence of disease causing mutations at these sites.

ACKNOWLEDGEMENTS

We wish to thank Dr Gary S. Gottesman for his review of this manuscript.

Authors' affiliations

S Tomatsu, T Nishioka, M A Gutierrez, O S Pena, A Noguchi, Department of Pediatrics, Saint Louis University, Pediatric Research Institute, St. Louis, MO, USA
A M Montaño, Department of Biosystems Science, The Graduate University for Advanced Studies, Kanagawa, Japan
K O Orii, W S Sly, E.A. Doisy Department of Biochemistry and Molecular Biology, Saint Louis University School of Medicine, St. Louis, MO, USA
S Yamaguchi, Department of Pediatrics, Shimane University, Izumo, Japan
T Orii, Department of Pediatrics, Gifu University, Gifu, Japan
E Paschke, Laboratory of Metabolic Diseases, Department of Pediatrics, University of Graz, Austria
S G Kircher, Institute of Medical Chemistry, University of Vienna, Austria

This work was supported by grants from the Austrian Research Society for Mucopolysaccharidoses and Related Diseases, the German MPS Society, the Italian MPS Society and the International Morquio Organization (Carol Ann Foundation).

Conflict of interest: none declared.

Correspondence to: S Tomatsu, Department of Pediatrics, Saint Louis University, Pediatric Research Institute, 3662 Park Ave, St. Louis, MO 63110-2586, USA; tomatsus@slu.edu

Received 30 December 2003
Accepted for publication 5 February 2004

REFERENCES

1 **Masue M**, Sukegawa K, Orii T, Hashimoto T. N-acetylgalactosamine-6-sulfate sulfatase in human placenta: purification and characteristics. *J Biochem (Tokyo)* 1991;**110**:965–70.
2 **Tomatsu S**, Fukuda S, Masue M, Sukegawa K, Fukao T, Yamagishi A, Hori T, Iwata H, Ogawa T, Nakashima Y. Morquio disease: isolation, characterization and expression of full-length cDNA for human N-acetylgalactosamine-6-sulfate sulfatase. *Biochem Biophys Res Commun* 1991;**181**:677–83.
3 **Nakashima Y**, Tomatsu S, Hori T, Fukuda S, Sukegawa K, Kondo N, Suzuki Y, Shimozawa N, Orii T. Mucopolysaccharidosis IVA: molecular cloning of the human N-acetylgalactosamine-6-sulfatase gene (GALNS) and analysis of the 5′-flanking region. *Genomics* 1994;**20**:99–104.
4 **Fukuda S**, Tomatsu S, Masue M, Sukegawa K, Iwata H, Ogawa T, Nakashima Y, Hori T, Yamagishi A, Hanyu Y. Mucopolysaccharidosis type IVA. N-acetylgalactosamine-6-sulfate sulfatase exonic point mutations in classical Morquio and mild cases. *J Clin Invest* 1992;**90**:1049–53.
5 **Tomatsu S**, Fukuda S, Cooper A, Wraith JE, Rezvi GM, Yamagishi A, Yamada N, Kato Z, Isogai K, Sukegawa K. Mucopolysaccharidosis IVA: identification of a common missense mutation I113F in the N-acetylgalactosamine-6-sulfate sulfatase gene. *Am J Hum Genet* 1995;**57**:556–63.
6 **Ogawa T**, Tomatsu S, Fukuda S, Yamagishi A, Rezvi GM, Sukegawa K, Kondo N, Suzuki Y, Shimozawa N, Oru T. Mucopolysaccharidosis IVA: screening and identification of mutations of the N-acetylgalactosamine-6-sulfate sulfatase gene. *Hum Mol Genet* 1995;**4**:341–9.
7 **Bunge S**, Kleijer WJ, Tylki-Szymanska A, Steglich C, Beck M, Tomatsu S, Fukuda S, Poorthuis BJ, Czartoryska B, Orii T, Gal A. Identification of 31 novel mutations in the N-acetylgalactosamine-6-sulfatase gene reveals excessive allelic heterogeneity among patients with Morquio A syndrome. *Hum Mutat* 1997;**10**:223–32.
8 **Kato Z**, Fukuda S, Tomatsu S, Vega H, Yasunaga T, Yamagishi A, Yamada N, Valencia A, Barrera LA, Sukegawa K, Orii T, Kondo N. A novel common missense mutation G301C in the N-acetylgalactosamine-6-sulfate sulfatase gene in mucopolysaccharidosis IVA. *Hum Genet* 1997;**101**:97–101.
9 **Yamada N**, Fukuda S, Tomatsu S, Muller V, Hopwood JJ, Nelson J, Kato Z, Yamagishi A, Sukegawa K, Kondo N, Orii T. Molecular heterogeneity in mucopolysaccharidosis IVA in Australia and Northern Ireland: nine novel mutations including T312S, a common allele that confers a mild phenotype. *Hum Mutat* 1998;**11**:202–8.
10 **Terzioglu M**, Tokatli A, Coskun T, Emre S. Molecular analysis of Turkish mucopolysaccharidosis IVA (Morquio A) patients: identification of novel mutations in the N-acetylgalactosamine-6-sulfate sulfatase (GALNS) gene. *Hum Mutat* 2002;**20**:477–8.
11 **Montano AM**, Kaitila I, Sukegawa K, Tomatsu S, Kato Z, Nakamura H, Fukuda S, Orii T, Kondo N. Mucopolysaccharidosis IVA: characterization of a common mutation found in Finnish patients with attenuated phenotype. *Hum Genet* 2003;**113**:162–9.
12 **Tomatsu S**, Fukuda S, Cooper A, Wraith JE, Ferreira P, Di Natale P, Tortora P, Fujimoto A, Kato Z, Yamada N, Isogai K, Yamagishi A, Sukegawa K, Suzuki Y, Shimozawa N, Kondo N, Sly WS, Orii T. Fourteen novel mucopolysaccharidosis IVA producing mutations in GALNS gene. *Hum Mutat* 1997;**10**:368–75.
13 **Antonarakis SE**, Krawczak M, Cooper DN. The nature and mechanisms of human gene mutation. In: Scriver CR, Beaudet AL, Sly WS, Valle D, eds. *The metabolic and molecular bases of inherited disease*. New York: McGraw-Hill, 2001:343–77.
14 **Krawczak M**, Ball EV, Cooper DN. Neighboring-nucleotide effects on the rates of germ-line single-base-pair substitution in human genes. *Am J Hum Genet* 1998;**63**:474–88.
15 **Tomatsu S**, Orii KO, Islam MR, Shah GN, Grubb JH, Sukegawa K, Suzuki Y, Orii T, Kondo N, Sly WS. Methylation patterns of the human beta-glucuronidase gene locus: boundaries of methylation and general implications for frequent point mutations at CpG dinucleotides. *Genomics* 2002;**79**:363–75.
16 **Tomatsu S**, Okamura K, Taketani T, Orii KO, Nishioka T, Gutierrez MA, Velez-Castrillon S, Fachel AA, Grubb JH, Cooper A, Thornley M, Wraith E, Barrera LA, Giugliani R, Schwartz IV, Frenking GS, Beck M, Kircher SG, Paschke E, Yamaguchi S, Ullrich K, Isogai K, Suzuki Y, Orii T, Kondo N, Creer M, Noguchi A. Development and testing of new screening method for keratan sulfate in mucopolysaccharidosis IVA. *Pediatr Res* 2004 (in press).
17 **Clark SJ**, Harrison J, Paul CL, Frommer M. High sensitivity mapping of methylated cytosines. *Nucleic Acids Res* 1994;**22**:2990–7.
18 **Fukuda S**, Tomatsu S, Masuno M, Ogawa T, Yamagishi A, Rezvi GM, Sukegawa K, Shimozawa N, Suzuki Y, Kondo N, Imaizumi K, Kuroki Y, Okabe T, Orii T. Mucopolysaccharidosis IVA: submicroscopic deletion of 16q24.3 and a novel R386C mutation of N-acetylgalactosamine-6-sulfate sulfatase gene in a classical Morquio disease. *Hum Mutat* 1996;**7**:123–34.
19 **Sukegawa K**, Nakamura H, Kato Z, Tomatsu S, Montano AM, Fukao T, Toietta G, Tortora P, Orii T, Kondo N. Biochemical and structural analysis of missense mutations in N-acetylgalactosamine-6-sulfate sulfatase causing mucopolysaccharidosis IVA phenotypes. *Hum Mol Genet* 2000;**9**:1283–90.
20 **Bottema CD**, Ketterling RP, Ii S, Yoon HS, Phillips JA, Sommer SS. Missense mutations and evolutionary conservation of amino acids: evidence that many of the amino acids in factor IX function as ''spacer'' elements. *Am J Hum Genet* 1991;**49**:820–38.
21 **Wacey AI**, Krawczak M, Kakkar VV, Cooper DN. Determinants of the factor IX mutational spectrum in haemophilia B: an analysis of missense mutations using a multi-domain molecular model of the activated protein. *Hum Genet* 1994;**94**:594–608.