**European Journal of Human Genetics (1999) 7,** 34–44

http://www.stockton-press.co.uk/ejhg

ARTICLES

# Sanfilippo type B syndrome (mucopolysaccharidosis III B): allelic heterogeneity corresponds to the wide spectrum of clinical phenotypes

Birgit Weber[1], Xiao-Hui Guo[1], Wim J Kleijer[2], Jacques JP van de Kamp[3], Ben JHM Poorthuis[4] and John J Hopwood[1]

[1]*Lysosomal Diseases Research Unit, Department of Chemical Pathology, Women's and Children's Hospital, North Adelaide, Australia*
[2]*Department of Clinical Genetics, University Hospital, Rotterdam*
[3]*Leiden University Medical Centre, Leiden*
[4]*University of Leiden, Department of Pediatrics, Leiden, The Netherlands*

**Sanfilippo B syndrome (mucopolysaccharidosis IIIB, MPS IIIB) is caused by a deficiency of α-N-acetylglucosaminidase, a lysosomal enzyme involved in the degradation of heparan sulphate. Accumulation of the substrate in lysosomes leads to degeneration of the central nervous system with progressive dementia often combined with hyperactivity and aggressive behaviour. Age of onset and rate of progression vary considerably, whilst diagnosis is often delayed due to the absence of the pronounced skeletal changes observed in other mucopolysaccharidoses. Cloning of the gene and cDNA encoding α-N-acetylglucosaminidase enabled a study of the molecular basis of this syndrome. We were able to identify 31 mutations, 25 of them novel, and two polymorphisms in the 40 patients mostly of Australasian and Dutch origin included in this study. The observed allellic heterogeneity reflects the wide spectrum of clinical phenotypes reported for MPS IIIB patients. The majority of changes are missense mutations; also four nonsense and nine frameshift mutations caused by insertions or deletions were identified. Only five mutations were found in more than one patient and the observed frequencies are well below those observed for the common mutations in MPS IIIA. R643C and R297X each account for around 20% of MPS IIIB alleles in the Dutch patient group, whilst R297X, P521L, R565W and R626X each have a frequency of about 6% in Australasian patients. R643C seems to be a Dutch MPS IIIB allele and clearly confers the attenuated phenotype. One region of the gene shows a higher concentration of mutations, probably reflecting the instability of this area which contains a direct repeat. Several arginine residues seem to be 'hot-spots' for mutations, being affected by two or three individual base pair exchanges.**

**Keywords: Sanfilippo type B syndrome; mucopolysaccharidosis IIIB; MPS; common mutations; α-N-actylglucosaminidase**

Correspondence: Dr John J. Hopwood, Lysosomal Diseases Research Unit, Department of Chemical Pathology, Women's and Children's Hospital, 72 King William Road, North Adelaide, South Australia 5006, Australia. Tel: 61 8 8204 6153; Fax: 61 8 8204 7100; E-mail: jhopwood@medicine.adelaide.edu.au
Received 6 May 1998; revised 9 June 1998; accepted 19 June 1998

## Introduction

In Sanfilippo syndrome, also known as mucopolysaccharidosis III (MPS III), the impaired lysosomal degradation and subsequent lysosomal storage of heparan sulphate causes severe central nervous system degeneration resulting in progressive dementia often combined with hyperactivity and aggressive behaviour.[1] Clinical onset is usually marked by delayed speech development by the age of two to three years, together with sleep disturbance, hirsutism and diarrhoea. MPS IIIB is one of the four recognised biochemical subtypes of this disease. Each of the MPS III types is inherited as an autosomal recessive disorder with some differences among the four types.[1] MPS III may be underdiagnosed in patients with less severe mental retardation since in contrast to other MPS disorders, Sanfilippo syndrome patients often have little somatic disease with the coarse facial features that are prominent in the other MPS often absent in adults. The high incidence of false negative results in the urinary screening test for mucopolysaccharides by some methods increases the difficulties in diagnosing the rare less severe and slowly evolving cases.[1] These slower-progressing disease phenotypes should more accurately be described as the attenuated form of Sanfilippo syndrome.[2,3]

MPS IIIA patients are reported to have the most severe of the four Sanfilippo subtypes with earlier onset of symptoms and more rapid progression,[1] whereas MPS IIIB patients display a wider clinical heterogeneity with attenuated and severe cases reported even within the same sibship.[2] The combined incidence of all four Sanfilippo subtypes has been estimated at 1:24 000[4] and MPS IIIB is the most common subtype in Greece,[5,6] whereas in Northern Europe MPS IIIA seems to be more predominant.[7] In Australia MPS IIIA and IIIB have incidences of about 1 in 114 000 and 1 in 211 000, respectively (Meikle P, 1998 unpublished observations), reflecting the pattern of migration from Northern Europe (predominantly Great Britain and Ireland) and from Southern Europe (mainly Italy and Greece).

MPS III may result from a deficiency in one of four enzymes involved in the degradation of heparan sulfate, with α-N-acetylglucosaminidase (NAGLU, EC 3.2.1.50) being deficient in MPS IIIB. A partial cDNA clone had been characterised earlier[8] and recently full-length cDNA clones encoding NAGLU were isolated (GenBank U 40846[9] and GenBank U43572[10]). The structure of the corresponding NAGLU gene on chromosome 17q21.1 was predicted by alignment of the cDNA sequence with the genomic sequence of the 5'-flanking region of the β-*HSD* gene (GenBank U34879). The presence of different genotypes, seen to segregate between black and white families[11,12] suggested the presence of NAGLU polymorphism. Vance *et al* pointed out that alleles with high enzymatic activity could interfere with the identification of heterozygotes[13] and this might have occurred in the case described by Pande *et al*.[14] In this family some of the heterozygotes had normal levels of NAGLU activity presumably due to a 'hyperactive' allele.

Recently several mutations causing MPS IIIB have been published.[3,10,15,16] We now report the analysis of the NAGLU alleles of 40 MPS IIIB patients revealing strong heterogeneity with 25 novel mutations and two polymorphisms identified, including one mutation (R643C) common amongst Dutch patients that is clearly associated with the attenuated form of the disease.

## Materials and Methods

### *Sample Preparation and PCR Amplification*

Skin fibroblasts from patients and normal controls were cultured and genomic DNA isolated as described elsewhere.[17] Clinical diagnosis was confirmed by skin fibroblasts having very low or non-detectable levels of NAGLU activity.[18,19] Exons were amplified from genomic DNA as shown in Table 1 which summarises primers, PCR and SSCP gel electrophoresis conditions. PCR reactions were carried out in 1 × Biotech buffer (67 mM Tris-HCl pH 8.8, 16.6 mM $NH_4SO_4$, 0.45% Triton X-100, 0.2% mg/ml gelatin) 2.5 mM $MgCl_2$, 400 μM dNTPs, 1 U Taq polymerase (Boehringer Mannheim, Germany), 200 ng of each primer and with or without (see Table 1) 10% dimethylsulfoxide using the tube control option of a Hybaid OmniGene Thermocycler. After initial denaturation of 2 min at 95°C, 35 cycles of 30 s at 95°C, 45 s at annealing temperature (see Table 1) and 30 s of extension at 72°C followed, with a final extension step for 2 min at 72°C. Around 30 ng of genomic DNA were used as template.

### *SSCP Analysis*

Fragment 5 was digested with Bcl I, fragment 6 II/III with Ava II and fragment 6 V with Hind II before SSCP analysis. 5 μl of PCR product and 4 μl of sample buffer (95% formamide, 20 mM EDTA, 0.005% Bromophenol blue, 0.05% Xylene cyanol FF) were mixed, heated to 95°C for 4 min and kept on ice before being loaded on to an 8%, 0.8 mm thick polyacrylamide gel (acrylamide/bisacrylamide ratio: 41.5:1 with or without 5% glycerol, see Table 1). Electrophoresis of the fragments was performed in 1 × TBE buffer on an IBI sequencing unit for 15 h at 20 W/450 V. DNA bands were visualised by staining with silver.

### *Sequencing*

Fragments with an altered band pattern were identified and the corresponding exon amplified by PCR. After purification

**Table 1** PCR and SSCP conditions

| *Exon* | *Primer* | *Position* | *Sequence* | *Product (bp)* | *Annealing temp. (°C)* | *DMSO 10%* | *SSCP gel 10% glycerol* |
|---|---|---|---|---|---|---|---|
| 1 I | ng 1 | 10861–10883 | 5′-CCTTCGGGTCACGTGGCCGGAG-3′ | 280 | 65 | + | – |
| | NAG 45 | 11141–11118 | 5′-GGTGTCCAAGCCCGGCTTGGCAGC-3′ | | | | |
| 1 II | NAG 38* | 11086–11110 | 5′-CAACTTCTCCGTGTCGGTGGAGCG-3′ | 329 | 55 | + | – |
| | ng 2 | 11415–11395 | 5′-GCTCAGCCTCCCGATTTGGGT-3′ | | | | |
| 2 | ng 3 | 12014–12033 | 5′-TTGTTCCAGGGCCGTGGACCC-3′ | 277 | 58 | – | + |
| | ng 4 | 12291–12268 | 5′-GGGCGGGTGAAAACACCTACGGTG-3′ | | | | |
| 3 | ng 5 | 12944–12968 | 5′-GAATGGTTGTTGAATGAATGA-3′ | 304 | 55 | – | – |
| | ng 6 | 13248–13227 | 5′-AGCCCTGGGGCCGGCCTAATAC-3′ | | | | |
| 4 | ng 7 | 13269–13289 | 5′-GCGTGTATCCTGGGAGATGAG-3′ | 248 | 58 | – | + |
| | ng 8 | 13517–13496 | 5′-CCAGAGCTTAAGTTTTTAAGTC-3′ | | | | |
| 5 | NAGit | 15574–15590 | 5′-GGTGAACACTATGGCGG-3′ | 407 | 55 | – | + |
| | ng 9 | 15981–15961 | 5′-GGCATTATTCCAGTCCCTCTG-3′ | | | | |
| 6 I | ng 10 | 17641–17660 | 5′-GAACATTCCCTGGGCCCTCTG-3′ | 325 | 60 | – | + |
| | NAG 12 | 17966–17943 | 5′-TCTGGGCCTCCGTTCACAGCCTCT-3′ | | | | |
| 6 II | NAG 42 | 17862–17882 | 5′-CACTGCCTCCTTCCAGGGCCA-3′ | 306 | 60 | – | – |
| | NAG 49 | 18168–18148 | 5′-CCTCCTCGCTGCCCCTGCGTC-3′ | | | | |
| 6 III | NAG 42 | 17862–17882 | 5′-CACTGCCTCCTTCCAGGGCCA-3′ | 477 | 60 | – | – |
| | CNBr 3n | 18339–18316 | 5′-GGTGGCCAGGGAGGGAGCAGATGT-3′ | | | | |
| 6 IV | CNBr 3c | 18310–18330 | 5′-CTGCTGCTCACATCTGCTCCCTCC-3′ | 310 | 60 | – | + |
| | CNBr 3′ | 18620–18597 | 5′-AGCTGGTAGCGGCTGTTCTGCTCG-3′ | | | | |
| 6 V | NAG 44 | 18568–18588 | 5′-GCAGTCAGTGAGGCCGAGGCC-3′ | 422 | 60 | – | –/+ |
| | NAG 48 | 18990–18970 | 5′-GATGTCTGTCCAGCTCTGGGC-3′ | | | | |
| 6 VI | NAG 47 | 18918–18939 | 5′-GATTCGCCACCACTGGGCCTTG-3′ | 263 | 60 | – | – |
| | ng 11 | 19181–19157 | 5′-CCCAAGCGTGGCAGCAGTGACCTTC-3′ | | | | |
| 1 | ng 1 | 10861–10883 | 5′-CCTTCGGGTCACGTGGCCGGAG-3′ | 554 | 55 | – | sequencing |
| | ng 2 | 11415–11395 | 5′-GCTCAGCCTCCCGATTTGGGT-3′ | | | | only |
| 6 | ng 10 | 17641–17660 | 5′-GAACATTCCCTGGGCCCTCTG-3′ | 1540 | 65 | – | sequencing |
| | ng 11 | 19181–19157 | 5′-CCCAAGCGTGGCAGCAGTGACCTTC-3′ | | | | only |

Positions of primers are given according to database entryGenBank U34879. *Oligonucleotide NAG 38 is mismatched at the second position due to a sequencing error but worked in the PCR.

of the PCR product on QIAquick PCR purification columns (Qiagen, Victoria, Australia) exons were sequenced using the Dye terminator Cycle Sequencing kit (Perkin Elmer) according to the manufacturer's instructions. Direct sequencing of fragments used for SSCP analysis was performed for exon 2, 4 and 5. Sequencing products were analysed on an ABI Prism sequencer model 737A by the University of Queensland DNA Sequencing Facility, Brisbane. In most cases nested primers were used for sequencing except for exons 2 and 4, where only the PCR primers were available. Sequence alterations were verified by sequencing the opposite strand and, where possible, restriction digest of an independent second PCR product or allele specific oligonucleotide hybridisation for missense and nonsense mutations. Deletions and insertions were confirmed by sequencing.

### *Allele-Specific Oligonucleotide Hybridisation*

ASO hybridisation was performed as described previously.[17] The oligonucleotides and washing conditions are summarised in Table 2.

### *Restriction Analysis*

A 15 μl aliquot of a 50 μl PCR reaction was incubated with the appropriate restriction enzyme according to manufacturer's instructions. Fragments were separated on 2% agarose gels and stained with ethidium bromide. Table 3 shows the different enzymes used for identification of the mutations.

### *Patients and Control Groups*

The patients included in this study, their origin, the age at diagnosis and their clinical phenotype are summarised in Table 4. Altogether 40 patients were analysed, 13 of them were from The Netherlands and 25 from Australasia, predominantly Caucasians. One patient was from Norway[14] and one from the USA.[20] As control, the *NAGLU* gene of 72 unaffected individuals was amplified and screened for the identified sequence changes. The control group consisted of 28 individuals from Australasia and 44 from The Netherlands.

### *Sequence Alignments*

A non-redundant database compiled from the entries of GenBank and EMBL databases, including updates (as of 17 February 1998) at the Australian National Genomic Information Service (ANGIS), was searched for NAGLU-related sequences with the TBLASTN local alignment tool[21] which screens the nucleotide database entries translated into all six reading frames with the protein sequence of the query. The number of alignments per clone varied between 1 and 17. In the analysis only alignments with the same strand in all three reading frames were included to allow for frameshifts by introns and/or sequencing errors. In case of overlapping or

**Table 2** ASO primers and washing conditions

| *Mutation* | *ASO primer* | *Washing conditions* |
|---|---|---|
| F48L | wt 5-CCGACTTCTCCGTGTC-3<br>mut 5-CCGACTTATCCGTGTC-3 | 1SSC/0.1% SDS 60°C |
| G292R | wt 5-CCATCATCGGGAGCCTC-3<br>mut 5-CCATCATCAGGAGCCTC-3 | 1SSC/0.1% SDS 65°C |
| R297X | wt 5-CTTCCTGCGAGAGCTGAT-3<br>mut 5-C6TTCCTGTGAGAGCTGAT-3 | 1SSC/0.1% SDS 65°C |
| V334F | wt 5-CACTGCCGTCTATGAG-3<br>mut 5-CACTGCCTTCTATGAG-3 | 1SSC/0.1% SDS 50°C |
| W404X | wt 5-CCTTCATCTGGTGCATGC-3<br>mut 5-CCTTCATCTAGTGCATGC-3 | 1SSC/0.1% SDS 60°C |
| F410S | wt 5-CACAACTTTGGGGGAAAC-3<br>mut 5-CACAACTCGGGGGAAAC-3 | 1SSC/0.1% SDS 55°C |
| W494X | wt 5-GGCAGCGTGGAGGCTAC-3<br>mut 5-GGCAGCGTGAAGGCTAC-3 | 0.5SSC/0.1% SDS 65°C |
| P521L | wt 5-CAGGCGGCCGTCCCTAC-3<br>mut 5-CAGGCGGCTGTCCCTAC-3 | 0.5SSC/0.1% SDS 65°C |
| L560P | wt 5-CTACGACCTGCTGGAC-3<br>mut 5-CTACGACCCGCTGGAC-3 | 1SSC/0.1% SDS 45°C |
| R565P | wt 5-CCTCACTCGGCAGGCAG-3<br>mut 5-CCTCACTCCGCAGGCAG-3 | 0.5SSC/0.1% SDS 65°C |
| R565W | wt 5-CCTCACTCGGCAGGCAG-3<br>mut 5-CCTCACTTGGCAGGCAG-3 | 0.5SSC/0.1% SDS 65°C |
| R643C | wt 5-CAGAACAGCCGCTACCAG-3<br>mut 5-CAGAACAGCTGCTACCAG-33 | 0.5SSC/0.1% SDS 65°C |
| R674C | wt 5-CACCCCTCGCTGGCGGC-3<br>mut 5-CACCCCTTGCTGGCGGC-3 | 0.5SSC/0.1% SDS 60°C |
| R676P | wt 5-CTCGCTGGCGGCTTTTCCTG-3<br>mut 5-CTCGCTGGCCGCTTTTCCTG-3 | 0.5SSC/0.1% SDS 65°C |

alternative alignments the one with the higher score was chosen. Smallest sum probabilities of 0–1.2e-5 were considered significant.

# Results

## *Mutation Screening*

The six exons of the *NAGLU* gene including the intron/exon boundaries and 5'- and 3'- non-translated regions were amplified by PCR in 12 overlapping PCR fragments. Non-radioactive SSCP analysis and subsequent sequencing of the DNA of 40 MPS IIIB patients revealed 33 different primary sequence changes. Under the given conditions, all missense mutations produced mobility shifts of the heteroduplex bands. However, some of them – R565P, L560P, P521L, F410S, V334F, G292R, H248R, G69S and F48L, as well as R297X and R626X – failed to produce a different pattern of single stranded DNA bands, whereas mobility shifts of the single-stranded DNA bands were observed in all cases of deletions and insertions.

The majority of changes lead to non-conservative exchanges of amino acids (Table 3), one (F48L) in the pro-peptide and the rest in the mature protein. Two of the amino acid exchanges (H248R and H414R) are conservative.

Four insertions ranging in size from 2 to 25 bp were identified as well as four deletions of 1–10 bp length, and four nonsense mutations. Two of these (R297X and R626X) were described previously;[10] another (W404X) was also found in a group of European patients (Bunge S, 1998 personal communication). The fourth nonsense mutation (W494X) is novel. All insertions result in frameshifts and subsequent premature termination, similar to the majority of deletions. However, delA2100 and del2bp2171 lead to frameshifts that mask the original stop codon and would elongate the gene product by 43 and 80 aa respectively.

Although all intron/exon boundaries as well as the non-translated region 5' and 3' of the coding sequence were included in the analysis, no splice-site, initiation or polyadenylation mutations were detected.

To identify possible polymorphism we screened more than 140 control chromosomes for these sequence changes by ASO hybridisation or, where possible, by

**Table 3** Mutations and polymorphisms

| *Mutation* | *Number of alleles found (frequency in % in **A**ustralasian and **D**utch patients)* | *Sequence alteration* | *Effect* | *Restriction enzyme* | *aa conserved in putative α-N-acetylglucosaminidase genes of other species* |
|---|---|---|---|---|---|
| c54g | 2 (homozygous patient) | GCC→GCG | silent mutation | | |
| F48L | 1 | TTC→TTA | non-conservative aa exchange in the propeptide | – | F in *mouse* |
| G69S | 1 | GGC→AGC | non-conservative aa exchange in the mature protein | –Bgl I | G in *mouse* |
| H227P | 1 | CAC→CCC | non-conservative aa exchange in the mature protein | +Xma I | H in *mouse*; K in *A. thaliana* and *C. elegans* |
| H248R | 1 | CAT→CGT | conservative aa exchange in the mature protein | –Nla III | H in *mouse*; N in *A. thaliana*; F in *C. elegans* |
| G292R[a] | 1 | GGG→AGG | non-conservative aa exchange in the mature protein | – | G in *mouse, A. thaliana* and *C. elegans* |
| R297X[b] | 10 (**A** 8%; **D** 23%) | CGA→TGA | premature termination after 296 of 743 aa | – | |
| V334F | 2 (homozygous patient) | GTC→TTC | non-conservative aa exchange in the mature protein | – | V in *mouse* and *A. thaliana* |
| W404X[a] | 2 (homozygous patient) | TGG→TAG | premature termination after 403 of 743 aa | +Bfa I | |
| F410S | 1 | TTT→TCT | non-conservative aa exchange in the mature protein | – | F in *mouse, A. thaliana, C. elegans* and *rice* |
| H414R | 1 | CAT→CGT | conservative aa exchange in the mature protein | –Nco I | H in mouse; R in C. elegans; I in A. thaliana and rice |
| W494X | 2 (homozygous patient) | TGG→TGA | premature termination after 494 of 743 aa | | |
| P521L[c] | 3 (**A** 6%) | CCG→CTG | non-conservative aa exchange in the mature protein | – | P in *mouse* |
| L560P | 1 | CTG→CCG | non-conservative aa exchange in the mature protein | – | L in *mouse*; M in *A. thaliana* |
| R565P | 1 | CGG→CCG | non-conservative aa exchange in the mature protein | – | R in *mouse* and *A. thaliana* |
| R565W | 3 (**A** 6%) | CGG→TGG | non-conservative aa exchange in the mature protein | – | |
| L617F | 1 | TTG→TTC | non-conservative aa exchange in the mature protein | +BstN I | L in *C. elegans* and *rice* |
| R626X[b] | 3 (**A** 6%) | CGA→TGA | premature termination after 625 of 743 aa | –Ava I | |
| R643C | 5 (**D** 19%) | CGC→TGC | non-conservative aa exchange in the mature protein | – | R in *mouse; A. thaliana, D. melanogaster* and *Dictyostelium;* G in *C. elegans* |
| G650E | 1 | GGG→GAG | non-conservative aa exchange in the mature protein | –Asu I | G in *mouse; C. elegans, D. melanogaster* and *Dictyostelium* |
| R674C[c] | 1 | CGC→TGC | non-conservative aa exchange in the mature protein | – | R in *mouse; C. elegans, D. melanogaster, Dictyostelium* and *rice* |
| R676P | 1 | CGG→CCG | non-conservative aa exchange in the mature protein | +Eae I | R in *mouse; D. melanogaster*; Q in *C. elegans* and *Dictyostelium*; S in *rice* |
| G737R | 4 (**A** 4%; **D** 4%) | GGC→CGC | non-conservative aa exchange in the mature protein | –Ncl I | A in *mouse* |
| delG59 | | GGGGGCG→GGGG_CG | frameshift, premature termination at position 120 after 101 altered aa | | |
| del10bp503b | | del[GGAGCGGCCA] | frameshift, premature termination at position 184 after 13 altered aa | | |
| delTG1035 | | GCTGTGTG→GCTG_TG | frameshift, premature termination at position 386 after 38 altered aa | | |
| delA1317 | | GGTAG→GG_TG | frameshift, premature termination at position 474 after 35 altered aa | | |
| delA2100 | | CAAAAAT→CAAAA_T | frameshift, shifted termination to position 823 after 44 altered and 80 additional aa | | |
| del2bp2171 | | ACTGTGGA→ACTG_GA | frameshift, shifted termination to position 786 after 19 altered and 43 additional aa | | |
| ins25bp48 | | [GGGGC.....CGACG] duplicated | direct repeat of bp 49–73, frameshift, premature termination at position 120 after 107 altered aa | | |
| ins7bp48 | | [GGGGCCG] duplicated | direct repeat of bp 49–55, frameshift, premature termination at position 120 after 102 altered aa | | |
| ins5bp209 | | [GCGGC] duplicated | direct repeat of bp 210–214, frameshift, premature termination at position 120 after 46 altered aa | | |
| insAA950 | | GATGC→GATAAGCA | frameshift, premature termination at position 338 after 22 altered aa | | |

Polymorphisms are in italics. [a]mutations also identified in a group of European patients (Bunge, S. 1998 personal communication); [b]identified previously [10]; [c]identified previously [3].

**Table 4** Genotype and phenotype of patients

| | *1. Allele* | *2. Allele* | *Polymorphism* | *Phenotype* | *Age at diagnosis (years/month)* |
|---|---|---|---|---|---|
| 1 | – | – | G737R | severe | 6 |
| 2 | H227P | P521L | | attenuated | 9 |
| 3 | F48L | R297X | c54/c54 | attenuated | 12 |
| 4 | R297X | – | | severe | 19 |
| 5 | V334F | V334F | | severe | 4 |
| 6 | R565W | – | | severe | 1 |
| 7 | R297X | del2bp2171 | | severe | 2/4 |
| 8 | P521L | P521L | | severe | 5 |
| 9 | del10bp503 | H414R | | severe | 3/2 |
| 10 | R297X | G69S | c54/c54 | attenuated | 21/8 |
| 11 | G292R | R565W | | severe | 2/9 |
| 12 | – | – | | severe | 7 |
| 13 | L560P | | | attenuated | 21/5 |
| 14 | – | – | | severe | 2/6 |
| 15 | L617F | – | | severe | 7/9 |
| 16 | W404X | W404X | | severe | <1 |
| 17 | delG59 | R626X | | severe | <1 |
| 18 | R565W | – | | severe | 5 |
| 19 | delTG1035 | R565P | | severe | 10 |
| 20 | R674C | G650E | | severe | 14 |
| 21 | F410S | – | | severe | 8/8 |
| 22 | – | – | | severe | 4/5 |
| 23 | R626X | R626X | | severe | 3/5 |
| 24 | insAA950 | – | G737R | severe | 3 |
| 25 | ins7bp48 | delA2100 | | severe | 4/6 |
| 26 | ins25bp48 | delA1317 | | severe? | ? |
| 27 | – | – | | severe | 4 |
| 28 | R643C | – | | attenuated | 29 |
| 29 | R676P | R297X | | severe | 5 |
| 30 | R297X | R297X | c54g/c54g | severe | 6 |
| 31 | R643C | R643C | | attenuated | 26 |
| 32 | – | – | | attenuated | 22 |
| 33 | – | – | | attenuated? | 5 |
| 34 | ins5bp209 | R297X | G737R | severe | 9 |
| 35 | R297X | – | | severe | 6 |
| 36 | R297X | – | c54/c54 | attenuated | 32 |
| 37 | – | – | | attenuated | 63 |
| 38 | R643C | R643C | | attenuated | 28 |
| 39 | H248R | – | | attenuated | 45 |
| 40 | W494X | W494X | G737R/G737R | severe | <1 |

– not known; ? insufficient data. Patients 1 to 15 and 17 to 26 are from Australasia. Patients 27 to 39 are from The Netherlands. Patient 40 is from Norway.[14] Patient 16 is from the United States.[20]

digestion with an appropriate restriction enzyme. Eighty of the controls were from Dutch individuals to allow the detection of population specific differences. The G737R allele was found on 12 control chromosomes and is therefore considered non-pathogenic.

The substitution of Cytosine54 by a Guanosine is clearly a polymorphism since it does not alter the encoded amino acid (Ala18). It was found in homozygous form in a patient with R297X/R297X genotype (patient 30). The c54g polymorphism is not linked to R297X, since other patients with this allele (patients 3, 10 and 36) clearly have the wildtype sequence at position 54.

## Genotype/Phenotype Correlation

Classifying patients into different phenotypes is problematic since much of the available data are observations of individual clinicians and not directly comparable. We are therefore only able to distinguish between a severe and an attenuated phenotype depending on onset of symptoms, language development and onset of

regression. Within these two groups there is still considerable variability in the disease progression.

Twenty-one of the Australasian MPS IIIB developed patients the classical severe Sanfilippo phenotype, whereas four patients developed the attenuated form of the disorder with later onset and/or slower regression. As expected, all mutations resulting in premature termination either by a nonsense mutation or insertions and deletions lead to the severe phenotype (patients 17, 23, 26, 30 and 40). Similarly, the deletions masking the original stop codon and presumably elongating the gene product (del2bp2171 and delA2100 in patients 7 and 25) confer a severe phenotype in combination with R297X and ins7bp48. Missense mutations V334F and P521L also cause the severe phenotype when present on both alleles in patients 5 (V334F/V334F) and 8 (P521L/P521L). However, in heterozygous form in combination with H227P (patient 2) P521L leads to the attenuated clinical phenotype as does G69S in combination with R297X (patient 10). The other two patients with attenuated phenotype (patients 3 and 13) have F48L/R297X and L560P/unknown genotypes.

The majority of Dutch patients show clinical symptoms consistent with the attenuated form of the disease. Patients 37 and 39 were only recently diagnosed at the age of 63 and 47 years respectively. Two of the attenuated cases (patients 31 and 38), who are apparently not related, were found to be homozygous for R643C, an allele that was also identified in an attenuated patient (patient 28) with an unknown second allele. Two patients (patients 35 and 36) with R297X/unknown genotype have the severe and attenuated phenotype respectively, implying that the unknown alleles modify the severity of the clinical phenotype.

## Unstable Regions

A plot of the mutations on to the gene structure (Figure 1) revealed a reasonably even distribution of mutations throughout the gene with the exception of a region that contains a direct repeat encoding the leader peptide and the putative signal peptidase cleavage site (bp 49–73). Three mutations (ins25bp48, ins7bp48 and delG59), as well as a polymorphism (c54g), are located in this area. There also appears to be a cluster of mutations in the last third of exon 6. Two codons (arginine 565 and 643) also show some instability with different mutations affecting the same residue: R565P and R565W, as well as R643H[10] and R643C.

## Frequencies

Due to the observed strong allelic heterogeneity, frequencies for mutations are generally low. The most common mutation, R297X, has an overall frequency of 12.5% (10 out of 80 alleles) and seems to be more common in the Dutch patient group where it accounts for 23% (6 out of 26 alleles), compared with 8% for the Australasian patients. Three alleles with the P521L mutation were identified, giving this amino acid exchange a frequency of 6% among Australasian patients. The second nonsense mutation, R626X, was identified in two patients of South European and Asian background in homozygote and heterozygote form respectively, giving it a frequency of 6% in the Australasian group similar to R565W, which was

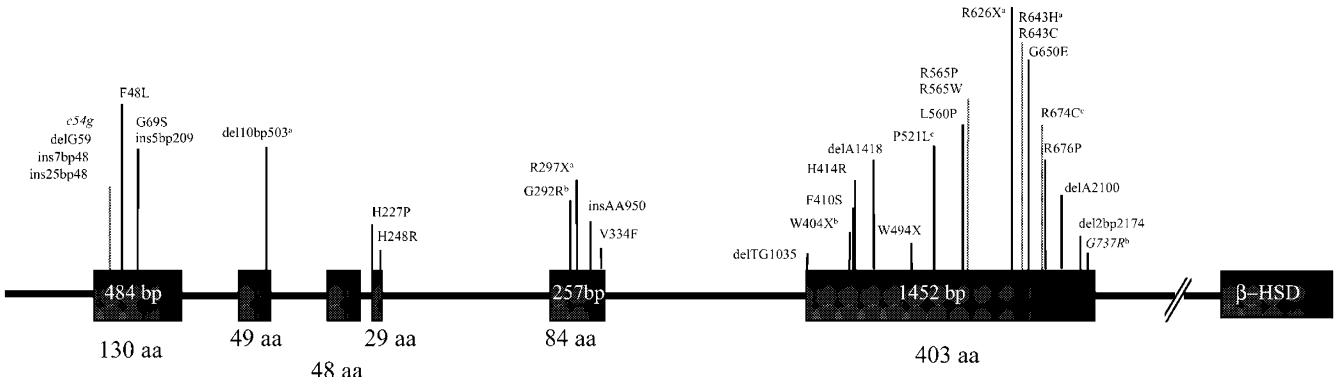


**Figure 1** *Novel polymorphisms are printed cursive. Exons are represented by boxes with the length in bp given for exon 1, 5 and 6. The number of amino acids encoded by the individual exons is printed underneath.* [a]*Mutations previously identified [10].* [b]*Mutations identified in a group of European patients (Susanna Bunge, personal communication).* [c]*Mutations identified previously [3]. Mutation 'hot-spots' are highlighted*

identified in three heterozygote patients. R643C accounts for 19% of the MPS IIIB alleles in the Dutch patients with two patients being homozygous and one being heterozygous, respectively. All other mutations were only found in a single patient.

The G737R polymorphism was identified on 12 out of 142 control alleles. Australasian and Dutch controls carry this polymorphism with similar frequencies, 7% and 9% respectively, whereas the frequencies in the two patient groups are slightly lower. Only 4% of Australasian and Dutch MPS IIIB patients carried the G737R allele.

### *Putative α-N-acetylglucosaminidase Genes*

With the TBLASTN local alignment tool[21] the NAGLU protein sequence was used to search a non-redundant database compiled of the entries from GenBank and EMBL databases, including updates as of 17 February 1998 at the Australian National Genomic Information Service (ANGIS) for related sequences. Apart from near-perfect alignments with various human genomic and cDNA clones isolated from different tissues and developmental stages and the mouse homolog (GenBank U85247), significant homologies [$P(N) < 1.2e\text{-}5$] were found to different clones from *Caenorhabditis elegans, Drosophila melanogaster, Arabidopsis thaliana, Dictyostelium discoideum* and *rice*. These putative *NAGLU* genes appear to have highly conserved amino-acid sequence in several regions of the protein. Table 3 summarises the alignments in respect to the mutations and polymorphisms found in patients. No homologies to yeast genes were found, suggesting that yeast does not have any α-N-acetylglucosaminidase activity, or that a different type of enzyme is catalysing the reaction.

## Discussion

We were able to identify 66% of the MPS IIIB alleles in the 40 patients studied. Since all intron/exon boundaries were included in the analysis, it is unlikely that splice-site mutations were missed. Although no changes were observed in the SSCP patterns of PCR product 6 VI, which consists of the 3' non-translated region including the polyA-signal, we cannot exclude mutations in this region, since the sensitivity of SSCP decreases with increased length of the analysed fragment.[22] Similarly, no changes were identified in the 63 bp of 5' non-translated region included in fragment 1 I. Mutations further upstream of the gene or within the introns might influence the transcription rate and/or stability of the mRNA and can only be detected by determining the mRNA levels in patient cell lines.

No highly common mutations were identified in either of the two patient groups. Only six mutations/polymorphism (R297X, P521L, R565W, R626X, R643C and G737R) were found in more than one patient, and R643C is restricted to the Dutch patient group, whilst R626X was only found in Australasian patients. However, R626X was reported earlier in an American patient and the deletion of 10 bp at position 503 was identified in an Italian patient[10] as well as being found on one allele in an Australasian patient. The P521L mutation, which has a frequency of 6% in Australasian patients, has been described in an American family.[3] Additionally, G292R and W404X have been identified in a European group of patients (Bunge S, personal communication), and R674C was found in an American family.[3] All other reported mutations appear to be unique so far.

The allelic heterogeneity that we observe in MPS IIIB patients compared with MPS IIIA patients is consistent with the greater clinical variability of MPS IIIB. MPS IIIA is considered to be generally more severe than MPS IIIB.[1] In MPS IIIA one mutation (R245H) accounts for over 30% of the alleles in Australasian[23] and for nearly 60% of the alleles in Dutch patients, conferring the severe phenotype.[24] Frequencies for mutations in MPS IIIB are much lower, with R297X accounting for 8% and 23% in the Australasian and Dutch patients, respectively, explaining the greater severity as well as the reduced clinical variability MPS IIIA compared to MPS IIIB. The low frequency of individual mutations makes it unlikely that mutation screening will improve the accuracy of the initial diagnostic procedures based on the analysis of metabolites and the determination of enzyme activity, the latter being sensitive enough for prenatal diagnosis.[25,26] However, mutation analysis will allow carrier-testing for siblings, especially if a partner is consanguineous, and will enable a more accurate prognosis for families of newly diagnosed patients, thereby improving genetic counselling. Mutation analysis is also of great importance for the selection of patients for trials of enzyme or gene replacement therapies and the evaluation of the efficacy of the protocols. Knowing the genotype of each treated patient might give an insight into the normal clinical

variations and adverse immunological reactions and thus improve the interpretation of the results of these trials.

R643C seems to be a Dutch MPS IIIB allele. The frequency of 23% is similar to the frequency of the R297X allele in this patient cohort but, in contrast to the truncated protein, R643C mutant protein seems to have some residual activity that leads to the attenuated clinical phenotype. Attempts to measure residual activities with fluorogenic substrate in skin fibroblasts were not successful and will have to be done with either natural substrate or with overexpressed mutant protein. Patients 31 and 38 are homozygous for the R643C allele. They were reported to have slow early development, but learned to speak. Regression of intellectual abilities and motor skills was not observed until they were in their late teens and early twenties, respectively. The two families, although apparently not related, are from the same geographical area which suggests a founder effect. Detailed clinical descriptions of patient 31 and her affected siblings and cousins are available.[27,28]

Although most of the mutations in the Australasian MPS IIIB patient cohort seem to result in the severe phenotype, there are four attenuated cases with H227P/P521L (patient 2), F48L/R297X (patient 3) G69S/R297X (patient 10) and L560P/unknown (patient 13) genotype. Since P521L and R297X cause a severe phenotype in homozygous form (patients 8 and 30) it is likely that F48L, G69S and H227P allow enough active enzyme to be produced to delay onset and progression of the disease. The other novel missense mutations (V334F, F410S, H414R, R565W/P, L617F, G650L and R676P) are associated with severe clinical phenotypes.

The identification of a homozygous nonsense mutation in the *NAGLU* gene of patient 16,[20] also proves that MPS IIIB and the glycogen storage disease, from which this patient also suffered, are not linked to each other, nor is the MPS IIIB phenotype due to the 12:20 translocation. This was obvious after the *NAGLU* gene was located on chromosome 17.[8–10]

Since the three-dimensional structure of NAGLU is unknown, speculation about the effect of these amino acid exchanges on the structure is not possible. Furthermore, we cannot exclude that some changes may be rare polymorphisms, although none of them were found on a relatively large number of control chromosomes. Also, except for H414R, the amino acid exchanges are non-conservative and differences in amino acid size and charge are expected to have an impact on the enzyme. Ultimately, expression studies are necessary to prove the detrimental effect of these amino acid exchanges on enzyme activity and/or stability.

The G737R polymorphism is clearly non-pathogenic since it was found on 12 of the 142 control chromosomes. The slightly lower frequency of this allele in the two patient groups is due to the fact that those mutations which were identified in more than one patient (R643C, R297X, P521L and R565W) have occurred on the wildtype background. It remains to be seen whether the exchange of G737 by arginine has any modifying effect on the enzyme function. However, G737R can be ruled out as a candidate for the proposed 'hyperactive' allele(s), since it was identified in homozygous form in patient 40, described by Pande *et al*.[14] In this family some of the obligate carriers had enzyme activity within the normal range, thought to be due to a putative 'hyperactive' allele. Identification of allele(s) with higher enzymatic activity might be important with respect to the development of therapies as well as giving insights into the catalytically important structures of the enzyme.

Although the second polymorphism, c54g, was not systematically screened for, it was found in homozygous form in patient 30 (R297X/R297X). No linkage of c54g to the R297X allele was found, since other patients with this nonsense mutation have the wildtype cytosine at position 54. c54 is located within a region of high mutation frequency and it seems likely that the exchange of c54 occurred on the R297X allele. Two insertions (direct repeats) and a deletion were identified within a stretch of 21 uninterrupted CpGs containing a direct repeat [CCGGGG]. CpG dinucleotides are known[29] to be hot spots for mutations similar to repeats within the sequence. In case of the latter, this is presumably caused by a slipped strand mispairing mechanism.[30]

The relative instability of CpG dinucleotides might also contribute to some mutations arising independently in the same codon which was found in R565 W/P, R643 C/H[10] and R674 C/H.[3] Similar multiple mutations were also observed in MPS IIIA (amino acids R74 and R377),[23,31,32], MPS I,[33] and disorders other than lysosomal storage diseases.[34–36] Two of the six arginine codons constitute CpG dinucleotides and in the above amino acid exchanges these two codons are mutated.

The apparent clustering of mutations in the last third of exon 6 (amino acids 612 to 737) might be coincidental, since no significant differences in this area

compared with the neighbouring sequence were found with respect to GC content or the number of CpG dinucleotides. Alternatively the apparent clustering might reflect a higher efficiency of the mutation detection method (SSCP) in this area, since mutations detected by direct sequencing of patient genomes are not clustered in this region.[3] Another possible explanation is that the DNA structure leaves this area more vulnerable to mutagenic events than other parts of the gene. A suggested[16] crucial role of the carboxy end of the protein in enzyme structure/function cannot necessarily be assumed.

A database search for homologous proteins, in order to identify motifs and conserved regions of the enzyme, revealed a number of putative *NAGLU* genes of different species, ranging from the slime mould *Dictyostelium* to mice. Unfortunately, the NAGLU protein sequence aligned to partial cDNAs, genomic clones or ESTs and therefore it is not possible to distinguish between regions that are not conserved and those which are not in the database. Thus predictions of important protein functions from evolutionary conserved regions cannot be made.

The attenuated clinical phenotype is thought to result from the presence of a low level of residual enzyme activity delaying the storage of heparan sulphate and thereby the onset and progression of symptoms. Mutations like R643C that are associated with this phenotype are therefore expected to produce active enzyme, possibly with strongly altered properties. Expression of mutants with some residual activity and analysis of enzyme stability, localisation and catalytic properties might enable some predictions of structure/function relationship and will be useful in interpreting future crystallographic data.

## Acknowledgements

We thank the patients' families for their support and Greta Richardson for tissue culture. The referral of patient cell lines and clinical data by colleagues from Australasia, by Dr U Moog, Clinical Genetics Centre, Maastricht, The Netherlands (patients 37–39), and Dr G Hug, The Children's Hospital Medical Center, Cincinnati (patient 16), is greatly acknowledged. This work was supported by grants from the National Health and Research Council of Australia, the Women's and Children's Hospital Research Foundation and the MPS Society (USA).

## References

1 Neufeld EF, Muenzer J: The mucopolysaccharidoses. In: Scriver CR, Beaudet AL, Sly WS and Valle D (eds). *The Metabolic and Molecular Basis of Inherited Disease* 7th edn. McGraw-Hill: New York, 1995, pp 2465–2494.
2 Andria G, Di Natale P, Del Guidice E, Strisciuglio P, Murino P: Sanfilippo B syndrome (MPS IIIB): attenuated and severe forms within the same sibship. *Clin Genet* 1979; **15**: 500–504.
3 Zhao HG, Aronovich EL, Whitley CB: Genotype–phenotype correspondence in Sanfilippo syndrome type B. *Am J Hum Genet* 1998; **62**: 53–63.
4 van de Kamp JJP, Niermeijer MF, von Figura K, Giesberts MAH: Genetic heterogeneity and clinical variability in the Sanfilippo syndrome (types A, B and C). *Clin Genet* 1981; **20**: 152–160.
5 Beratis NG, Sklower SL, Wilbur L, Matalon R: Sanfilippo disease in Greece. *Clin Genet* 1986; **29**: 129–132.
6 Michelakakis H, Dimitriou E, Tsagaraki S, Giouroukos S, Schulpis K, Bartsocas CS: Lysosomal storage diseases in Greece. *Genet Couns* 1995; **6**: 43–47.
7 OMIM No. 252900 and No. 252920.
8 Zhao H, Lopez R, Rennecker J, Neufeld EF: Sanfilippo syndrome type B: cDNA and gene encoding human α-N-acetylglucosaminidase. *Am J Hum Genet* 1994; **55**: Abstract 1473.
9 Weber B, Blanch L, Clements PR, Scott HS, Hopwood JJ: Cloning and expression of the gene involved in Sanfilippo B syndrome (Mucopolysaccharidosis III B). *Hum Mol Genet* 1996; **5**: 771–777.
10 Zhao HG, Li HH, Bach G, Schmidtchen A, Neufeld EF: The molecular basis underlying Sanfilippo syndrome type B. *Proc Natl Acad Sci USA* 1996; **93**: 6101–6105.
11 Vance JM, Pericak-Vance MA, Elston RC *et al*: Evidence of genetic variation for α-N-acetyl-D-glucosaminidase in black and white populations: A new polymorphism. *Am J Med Genet* 1980; **7**: 131–140.
12 Pericak-Vance MA, Vance JM, Elston RC, Namboodiri KK, Fogle TA: Segregation and linkage analysis of α-N-acetyl-D-glucosaminidase (NAG) levels in a black family. *Am J Med Genet* 1985; **20**: 295–306.
13 Vance JM, Coneally PM, Wappner R, Yu PL, Brandt IK, Pericak-Vance MA: Carrier detection in Sanfilippo syndrome type B: report of six families. *Clin Genet* 1981; **20**: 135–140.
14 Pande H, Chester A, Lie H, Thorsby E, Stormorken H: Concomitant occurence of Mucopolysaccharidosis IIIB and Glanzmann's thrombasthenia. Further evidence of a hyperactive α-N-acetylglucosaminidase-producing allele. *Clin Genet* 1992; **41**: 243–247.
15 Aronovich EL, Zhao HG, Neufeld EF, Whitley CB: Mutation analysis in Sanfilippo syndrome type B by automated sequencing of the NAGLU coding region. *Am J Hum Genet* 1996; **59**: Abstract 1418.
16 Schmidtchen A, Greenberg D, Zhao HG *et al*: NAGLU mutations underlying Sanfilippo syndrome type B. *Am J Hum Genet* 1998; **62**: 64–69.
17 Scott HS, Litjens T, Nelson PV, Brooks DA, Hopwood JJ, Morris CP: α-L-iduronidase mutations (Q70X and P533R) associated with a severe Hurler phenotype. *Hum Mutat* 1992; **1**: 333–339.

18 Marsh J, Fensom AH: 4-methylumbelliferyl α-N-acetylglucosaminidase activity for diagnosis of Sanfilippo B disease. *Clin Genet* 1985; **27**: 258–262.
19 Hopwood JJ, Elliott H: Detection of the Sanfilippo type B syndrome using radiolabelled oligosaccharides as substrates for the estimation of α-N-acetylglucosaminidase. *Clin Chim Acta* 1982; **120**: 77–86.
20 McIntire SC, Bansal V, Wenger SL *et al*: Glycogen storage disease and Sanfilippo syndrome type B in a patient with 12:20 translocation. *Am J Hum Genet* 1993; **53**: Abstract 927.
21 Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: Basic local alignment search tool. *J Mol Biol* 1990; **215**: 403–410.
22 Hayashi K, Yandell DW: How sensitive is PCR-SSCP? *Hum Mutat* 1993; **2**: 338–346.
23 Weber B, Guo X-H, Wraith JE *et al*: Novel mutations in Sanfilippo A syndrome: Implications for enzyme function. *Hum Mol Genet* 1997; **6**: 1573–1579.
24 Weber B, van de Kamp JJP, Kleijer WJ *et al*: Identification of a common mutation (R245H) in Sanfilippo A patients from The Netherlands. *J Inherit Metab Dis* 1998; in press
25 Kleijer WJ, Huijmans JG, Blom W *et al*: Prenatal diagnosis of Sanfilippo disease type B. *Hum Genet* 1984; **66**: 287–288.
26 Minelli A, Danesino C, Lo-Curto F *et al*: First trimester prenatal diagnosis of Sanfilippo disease (MPSIII) type B. *Prenat Diagn* 1988; **8**: 47–52.
27 van de Kamp JJP, van Pelt JF, Liem KO, Giesberts MAH, Niepoth LTM, Staalman CR: Clinical variability in Sanfilippo B disease: a report on six patients in two related sibships. *Clin Genet* 1976; **10**: 279–284.
28 van Schrojenstein-de Valk HMJ, van de Kamp JJP: Follow-up on seven adult patients with mild Sanfilippo B-disease. *Am J Med Genet* 1987; **28**: 125–129.
29 Cooper DN, Krawczak M, Antonarakis SEM: The nature and mechanisms of human gene mutation. In: Scriver CR, Beaudet AL, Sly WS, Valle D (eds). *The Metabolic and Molecular Basis of Inherited Disease*, 7th edn. McGraw-Hill: New York, 1995, pp 259–292.
30 Cooper DN, Krawczak M: Mechanisms of insertional mutagenesis in human genes causing genetic disease. *Hum Genet* 1991; **87**: 409–415.
31 Bunge S, Ince H, Steglich C *et al*: Identification of 16 sulfamidase gene mutations including the common R74C in patients with Mucopolysaccharidosis type IIIA (Sanfilippo A). *Hum Mutat* 1997; **10**: 479–485.
32 Di Natale P, Balzano N, Esposito S, Villani GRD: Identification of molecular defects in Italian Sanfilippo A patients. *Hum Mutat* 1998; in press
33 Scott HS, Bunge S, Gal A, Clarke LA, Morris CP, Hopwood JJ: Molecular genetics of Mucopolysaccharidosis Type I: diagnostic, clinical and biological implications. *Hum Mutat* 1995; **6**: 288–302.
34 Kotze MJ, Loubser O, Thiart R *et al*: CpG hotspot mutations at the LDL receptor locus are a frequent cause of familial hypercholesterolaemia among South African Indians. *Clin Genet* 1997; **5**: 394–398.
35 Munier FL, Korvatska E, Djemai A *et al*: Kerato-epithelin mutations in four 5q31-linked corneal dystrophies. *Nat Genet* 1997; **15**: 224–225.
36 Vihinen M, Belohradsky BH, Haire RN *et al*: BTKbase, mutation database for X-linked agammaglobulinemia (XLA). *Nucleic Acids Res* 1997; **25**: 166–171.