Am. J. Hum. Genet. 56:597–607, 1995

Molecular Diagnosis of Mucopolysaccharidosis Type II (Hunter Syndrome) by Automated Sequencing and Computer-Assisted Interpretation: Toward Mutation Mapping of the Iduronate-2-Sulfatase Gene

Jon J. Jonsson,[1,2,*] Elena L. Aronovich,[1,2] Stephen E. Braun,[1,3,†] and Chester B. Whitley[1,2,3]

[1]Institute of Human Genetics, [2]Department of Pediatrics, and [3]Department of Genetics and Cell Biology, University of Minnesota Medical School, Minneapolis

Summary

Virtually all mutations causing Hunter syndrome (mucopolysaccharidosis type II) are expected to be new mutations. Therefore, as a means of molecular diagnosis, we developed a rapid method to sequence the entire iduronate-2-sulfatase (IDS) coding region. PCR amplicons representing the IDS cDNA were sequenced with an automatic instrument, and output was analyzed by computer-assisted interpretation of tracings, using Staden programs on a Sun computer. Mutations were found in 10 of 11 patients studied. Unique missense mutations were identified in five patients: H229Y (685C→T, severe phenotype); P358R (1073C→G, severe); R468W (1402C→T, mild); P469H (1406C→A, mild); and Y523C (1568A→G, mild). Nonsense mutations were identified in two patients: R172X (514C→T, severe) and Q389X (1165C→T, severe). Two other patients with severe disease had insertions of 1 and 14 bp, in exons 3 and 6, respectively. In another patient with severe disease, the predominant (>95%) IDS message resulted from aberrant splicing, which skipped exon 3. In this last case, consensus sequences for splice sites in exon 3 were intact, but a 395C→G mutation was identified 24 bp upstream from the 3′ splice site of exon 3. This mutation created a cryptic 5′ splice site with a better consensus sequence for 5′ splice sites than the natural 5′ splice site of intron 3. A minor population of the IDS message was processed by using this cryptic splice site; however, no correctly spliced message was detected in leukocytes from this patient. The mutational topology of the IDS gene is presented.

Received May 5, 1994; accepted for publication December 20, 1994.

Address for correspondence and reprints: Dr. Chester B. Whitley, Box 446 UMHC, University of Minnesota, 420 Delaware Street, SE, Minneapolis, MN 55455.

* Present address: Department of Genetics, Yale University School of Medicine, New Haven, CT.

† Present address: Walther Oncology Center, Indiana University School of Medicine, Indianapolis.

0002-9297/95/5603-0007$02.00

Introduction

Hunter syndrome (mucopolysaccharidosis type II) is an X-linked inborn error of lysosomal glycosaminoglycan metabolism, resulting from deficiency of iduronate-2-sulfatase (IDS) enzyme activity (i.e., iduronate-2-sulfate 2-hydrolase; E.C.3.1.6.13) and the consequent systemic accumulation of heparan sulfate and dermatan sulfate (for review, see Neufeld and Muenzer 1989; Hopwood and Morris 1990; Whitley 1992). All patients with this disorder are believed to have a defect of a single gene localized to Xq28 (Le Guern et al. 1990).

Most children with Hunter syndrome have a relatively severe form with early somatic abnormalities including skeletal deformities (dysostosis multiplex), hepatosplenomegaly, and progressive cardiopulmonary deterioration. A prominent clinical feature is neurological damage that presents as developmental delay and hyperactivity but progresses to mental retardation and dementia. Patients with severe forms of Hunter syndrome die before 15 years of age, usually as a result of obstructive airway disease or cardiac failure (Young and Harper 1982; Young et al. 1982). In contrast, those with a relatively mild form of Hunter syndrome may survive into adulthood, with attenuated somatic complications and often without mental retardation. This phenotypic variability is presumed to reflect allelic heterogeneity of IDS mutations; however, current assays for IDS catalytic activity do not distinguish the increments of residual activity postulated to account for the variable clinical phenotype.

Recent cloning and sequencing of the human IDS coding region (Wilson et al. 1990) has made possible the characterization of specific molecular genetic defects. Mutations described thus far have been heterogeneous as expected for X-linked recessive, lethal conditions (Haldane 1947). Cumulated results of work in this laboratory and several others have suggested that many patients (14 of 62 studied) have major gene deletions or rearrangements, as detected by Southern analysis (Wilson et al. 1990; Wilson et al. 1991; Wraith et al. 1991; Beck et al. 1992; Bunge et al. 1992; Gal et al. 1992; Palmieri et al. 1992; Steen-Bondeson et al. 1992; Wehnert et al. 1992; Annella et al. 1993; Froissart et al. 1993; Schroder et al. 1993; Sukegawa et al. 1993;

Yamada et al. 1993) or by reverse transcription (RT) linked to PCR (Crotty et al. 1992; Flomen et al. 1992). Importantly, such major gene defects are found exclusively in patients with the severe form of disease including brain involvement. Further genotype-phenotype correlation has been only anecdotal; however, such information would be extremely helpful in making decisions regarding management, especially when considering high-risk or experimental treatments (e.g., allogeneic bone marrow transplantation, enzyme replacement therapy, and gene therapy [Braun et al. 1993]). To identify mutations efficiently and with increased sensitivity, we developed a method that is based on direct sequencing of the entire IDS coding region in cDNA amplicons. Sequences were read by an automated instrument, and mutations were identified with computer-assisted visual scanning of electrophorofluorograms.

Subjects and Methods

Subjects

Blood samples and clinical information were obtained from patients with Hunter syndrome, who were evaluated at the University of Minnesota and other institutions. Informed consent was obtained according to the policies of the Institutional Review Board at the University of Minnesota. Clinical phenotype was assessed according to established clinical criteria that distinguish mild from severe forms on the basis of neurological involvement (Young and Harper 1982; Young et al. 1982).

Isolation and RT of RNA

Leukocytes were prepared by differential sedimentation in dextran (Lichtman 1990). Total cellular RNA was isolated, using the method of Chomczynski and Sacchi (1987), from leukocytes or lymphoblastoid cell lines (LCL, transformed with Epstein-Barr virus). IDS cDNA was generated by reverse transcribing the IDS message in an RNA sample corresponding to ~1.2 ml of the original blood. The 20-μl reaction contained 1 mM of each dNTP, 50 pmol primer IDS6 (table 1), 15 U RNasin (Promega), and 10 mM DTT. Moloney murine leukemia virus RT (200 U) and buffer were from GIBCO/BRL. Temperature during RT was 37°C for 30 min, followed by a linear gradient from 37°C to 55°C, over 15 min.

Amplification of IDS cDNA

The cDNA preparation was used as template for PCR, in order to amplify the entire IDS coding region in a reaction mixture containing primers IDS5 and IDS6 (50 pmol each) (table 1), 5 U *Taq* polymerase, 1 × buffer (Boehringer Mannheim), and the following conditions: denaturation at 94°C for 8 min, followed by 40 cycles of 94°C for 1 min, 60°C for 1 min 40 s, and 72°C for 3 min. Aliquots (5 μl) of the primary PCR reaction were reamplified in two separate 100-μl reactions by using 5 U *Taq* polymerase, 1 × buffer, and nested primers IDS11 with IDS12 or IDS3 with IDS20, to generate 5′ and 3′ secondary amplicons, respectively. PCR conditions were 94°C for 5 min, then 30 cycles of 94°C for 1 min, 58°C for 1 min, and 72°C for 2 min 30 s. Secondary amplicons were analyzed by electrophoresis of reaction aliquots (8 μl) in 0.7% agarose gels containing ethidium bromide (0.5 μg/ml), followed by visualization of products under UV light. Concentrations of amplicons were determined by comparison with a λ*Hin*dIII marker (BRL).

Sequencing Techniques

Secondary amplicons were washed twice in 2 ml deionized-distilled H_2O by using Centricon-100 centrifugal concentrators (Amicon). Templates (typically, 100 ng) were sequenced using the PRISM Ready Reaction Dye Deoxy Terminator cycle sequencing kit, which is based on *Taq* polymerase cycle sequencing and fluorescent dideoxynucleotide terminators (Applied Biosystems). Conditions for sequencing were 25 cycles at 96°C for 30 s, 50°C for 15 s, and 60°C for 4 min. Products of sequencing reactions were purified by Biogel P-10 (Bio-Rad) size-exclusion centrifugal column chromatography and were fractionated by electrophoreses in a 6% denaturing polyacrylamide gel. Fluorescent signals from sequencing gels were read by an Applied Biosystems 373A automated sequencer.

Computer-Assisted Analysis of Sequences

Electronic data (i.e., electrophoretic fluorograms and nucleotide assignments generated by the automated sequencer) were transferred to a computer (Sun0S, release 4.1.3; SPARC station IPC) and were reformatted by using the transfer and editing subroutine for further analysis with the Staden comprehensive sequence analysis package (Dear and Staden 1991). Sequences were analyzed as described in Results.

Isolation and Analysis of Genomic DNA

Genomic DNA was isolated from peripheral blood or LCL cells by using the Puregene DNA isolation kit as described by the manufacturer (Gentra Systems). PCR amplification was carried out by using 1 μg genomic DNA as template, 10 pmol primers (see fig. 2 legend), 5 U *Taq* polymerase, and 1 × buffer in 50-μl reactions. Conditions were 94°C for 5 min, then 30 cycles at 94°C for 1 min, 58°C for 1 min, and 72°C for 2 min 30 s. Aliquots (8 μl) of reactions were analyzed by agarose gel electrophoresis, as described above.

Results

Generation of Amplicons Representing the Entire IDS Coding Sequence

The IDS mRNA containing the 1,650-bp coding sequence appears to be an extremely rare species, as expected

Table 1

Locations and Sequences of Oligodeoxynucleotides

Primer	Orientation	Location[a]	Sequence	Reference
IDS5	Sense	−92 to −73	ACGAGGAGGTCTCTGTGGCT	This report
IDS11	Sense	−65 to −46	CTGCTAACTGCGCCACCTGC	Primer C; Flomen et al. 1992
IDS1	Sense	−14 to +5	CGCGTCGAAttCGAAATGCC[b]	Crotty and Whitley 1992
IDS16	Antisense	+184 to +165	GGGACCTCACCAGCTTATCC	Crotty et al. 1992
IDS13	Sense	+276 to +295	CACTGGCAGGAGACCTGACA	This report
IDS8	Antisense	+366 to +347	GTACTGGGGGATGGTGGAGA	Crotty et al. 1992
IDS14	Antisense	+534 to +515	GAGTTCTCCATCTGGCCCTC	Crotty et al. 1992
IDS3	Sense	+634 to +653	ATGAAAACGTCAGCCAGTCC	Crotty and Whitley 1992
IDS4	Antisense	+707 to +688	TTGGGGTATCTGAAGGGGAT	Crotty and Whitley 1992
IDS18	Antisense	+910 to +891	CAGAGGCAAAGTAGCTCTGG	This report
IDS9	Sense	+1008 to +1027	GTGGGCTCTAGGTGAACATG	This report
IDS12	Antisense	+1083 to +1064	GAATATCAGGGGAACATGGG	Crotty et al. 1992
IDS7	Sense	+1257 to +1276	ACCTCGCTGCCCCGTTCCTT	Crotty et al. 1992
IDS10	Antisense	+1352 to +1333	GGATCCTCTTCCAAGTCACG	Crotty et al. 1992
IDS2	Antisense	+1679 to +1660	CATTTGCCATggATGGTTGG[b]	Crotty et al. 1992
IDS20	Antisense	+1698 to +1679	GCTGGAAGGGAGCACATCAC	Primer D; Flomen et al. 1992
IDS6	Antisense	+1718 to +1699	TCTAgaTCCTCCTCTCACCA[b]	Crotty and Whitley 1992
SIN2	Sense	−53 from exon 3	GGCAGACATGTTTTGCTGTG	This report
ASIN3	Antisense	+60 from exon 3	GAATGCTGGATTCAGACACC	This report

[a] Counting from translation initiation site.

[b] Lowercase letters represent changes from normal sequence included for genetic engineering purposes (Braun et al. 1993).

for a message encoding a lysosomal enzyme (Wilson et al. 1990). To generate amplicons for direct sequencing, we reverse transcribed IDS mRNA by using a primer for the 3′-UTR, and the cDNA was subsequently amplified by using the same primer with the 5′-UTR primer (fig. 1). This reaction produced a quantity of the 1,810-bp amplicon insufficient for visualization in ethidium bromide–stained agarose gels or for sequencing. Therefore, two overlapping secondary amplicons were generated by using nested primers and the product of the primary PCR reaction as a template (fig. 1). This approach consistently resulted in a unique amplicon (1.0–1.5 μg each) of the correct size (fig. 2*A*).

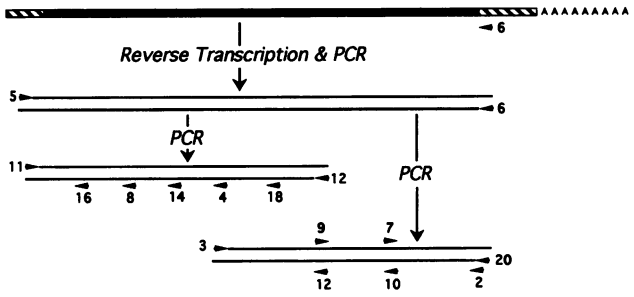

Figure 1 Strategy to generate IDS cDNA, which was then amplified in sequential PCR reactions to generate 5′ and 3′ secondary PCR products (examples shown in fig. 2*A*). IDS coding sequence is represented by blackened rectangle, and UTRs are represented by cross-hatched rectangles. The cDNA is illustrated with a single line, and PCR products are illustrated with double lines. Primers used for amplification and sequencing are indicated with arrowheads and are numbered as indicated in table 1.

Sequencing of the IDS Coding Region

The secondary amplicons were sequenced directly with five primers for each of the templates (fig. 1). Each sequencing reaction generated an average of ~270 bp of clearly interpretable sequence. The 10 sequences combined from both amplicons represented the entire IDS coding sequence, with overlapping sequences covering 60% of the region. However, a substantial number of errors or indeterminate bases were present in each tracing.

To improve the accuracy of sequencing, electronic files, containing tracings of the electrophoretic fluorograms and base sequence "calls" determined by the automated sequencer, were transferred to a Sun computer and were aligned with the normal IDS coding sequence (Wilson et al. 1990) by using the Staden program. This computer-assisted approach allowed for processive scanning of the entire coding region at each site by relying primarily on the tracing with the highest quality and by referring to overlapping sequences from the same patient for comparison, as necessary. The heights of individual peaks in the tracings were variable. However, tracings generated with the same primer, but using templates from different individuals, were very similar in shape. This comparison of tracings from patients with those of normal individuals greatly aided interpretation of problematic sequences (fig. 3). An average of 30 indeterminate bases and sequencing errors per patient were reviewed this way. This computer-assisted approach made possible an unambiguous evaluation of the IDS coding sequence for all 11 patients studied and required ~2 h/patient.

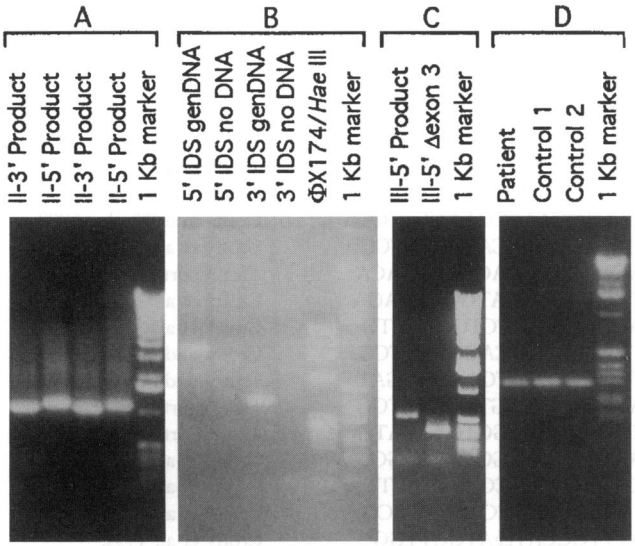

Figure 2 Ethidium bromide–stained PCR products visualized after electrophoresis, as described in Subjects and Methods. *A*, Two examples of overlapping 1,148-bp 5′ and 1,064-bp 3′ secondary PCR products covering the entire coding sequence (fig. 1). *B*, PCR products amplified from genomic DNA around the 5′ (1,100 bp) and 3′ (422 bp) termini of the IDS coding sequence, generated by using primers IDS1 with IDS16 and IDS7 with IDS2, respectively. Complete reactions with genomic DNA template (genDNA) and negative control reactions lacking template (no DNA) are shown. *C*, 5′ Tertiary PCR products from a patient with skipping of exon 3, compared with the 548-bp tertiary product of a normal control. Note the predominant 370-bp product representing messages with skipping of the entire exon 3 (178 bp) and the minor 522-bp product resulting from activation of a cryptic splice site in exon 3 (fig. 4). PCR products were generated by using primers IDS1 with IDS14 and 5′ secondary PCR products as templates, under conditions for generation of secondary products described in Subjects and Methods. *D*, Products (291 bp) of the amplification of exon 3 and flanking intron sequences in genomic DNA from a patient with skipping of exon 3 at the mRNA level and from normal controls for comparison. Primers SIN2 and ASIN3 were used for the amplifications.

Screening for Large Deletions in the IDS Gene

To confirm extensive deletions involving the IDS gene in these patients for whom RT-PCR products could not be generated, we also developed two PCRs that amplified genomic sequences around the 5′ and 3′ termini of the IDS coding sequence. Potential contamination of material containing the IDS cDNA was avoided by amplifying across the 0.9-kb intron 1 in generating the 5′ product. The PCRs using genomic DNA as a template resulted in two products of the correct sizes in normal individuals and in patients with base-pair substitutions (fig. 2*B*). No products were generated from genomic DNA of a patient previously determined to have a deletion of the IDS gene (Crotty and Whitley 1992).

Mutation Analysis in 11 Patients with MPS II

Using the method described above, we studied 11 patients with Hunter syndrome (table 2). Secondary amplicons were generated from the IDS message in all instances. We identified isolated missense mutations in five patients. Two patients, with mutations in exon 5 (H229Y) and exon 8 (P358R), had severe disease. Three patients with missense mutations in exon 9 had mild disease (P469H; Y523C; R468W, previously encountered in one other patient; see table 2). Two patients had nonsense mutations (i.e., in exon 5 [R172X; this mutation had also been described previously in two other patients; see table 2] and in exon 8 [Q389X]). Two patients with severe disease had short insertions of 1 nt and 14 nt, in exons 3 and 6, respectively. Both of these insertions caused frameshift and truncation of the protein. No mutation in the IDS coding sequence was identified in one patient with severe disease.

The 5′ secondary amplicon from one patient was slightly shorter than the expected length and shorter than the 3′ secondary PCR product (data not shown). Sequencing with antisense primer IDS14, which hybridized to exon 5, revealed skipping of the entire 178-bp exon 3 in the IDS message. However, sequencing of this patient's 5′ secondary amplicon with the antisense primer IDS8, which hybridized to exon 3, resulted in a weak but detectable signal, representing normal sequence. This demonstrated that a

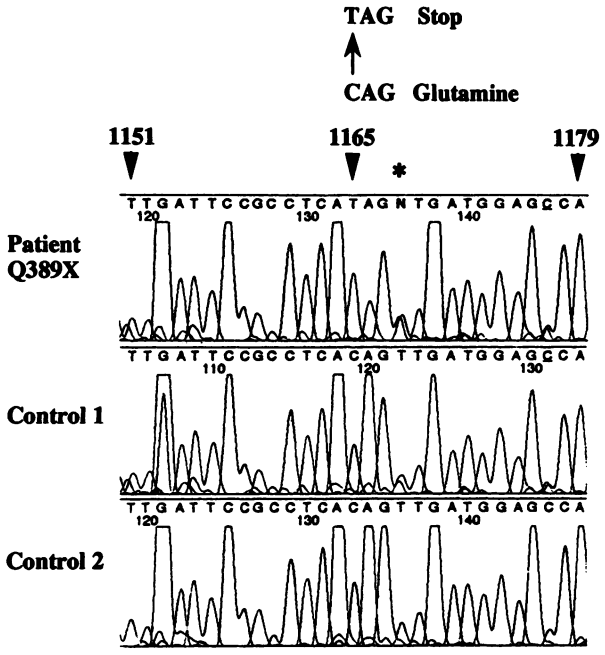

Figure 3 Comparison of tracings from different individuals in mutation analysis. Illustrated are tracings generated by the automated sequencer using primer IDS9 to sequence a 3′ secondary product containing a mutation (1165C→T resulting in Q389X) and two different 3′ secondary products (normal sequence controls). The location within the IDS coding sequence is indicated above the tracings. Note the 1165C→T mutation in the tracing from the patient and the difference in height of the first downstream base 1166A between mutant and normal sequences. Base 1168T could not be identified in the uppermost tracing. However, all three tracings appear very similar for this base (called "T" in the lower two tracings) and for the next base downstream (i.e., 1169T), suggesting that base 1168 in the uppermost tracing is also a T. This interpretation was confirmed by overlapping sequence generated from primer IDS10 (data not shown).

very minor population of the original IDS mRNA had contained exon 3 sequences. Further amplification of the 5′ secondary amplicon from this patient, with nested primers IDS1 and IDS14, produced, in addition to the predominant 370-bp product lacking exon 3, a minor population of a tertiary amplicon that was slightly smaller than the 548-bp normal control product (fig. 2C). Direct sequencing of the tertiary amplicons, with the sense primer IDS13 (which hybridized to the 5′ half of exon 3), revealed aberrant splicing with activation of a cryptic 5′ splice site in exon 3 at position +393 in the coding sequence.

Exon 3 was PCR amplified from genomic DNA of this patient by using primers hybridizing to flanking intron sequences. The resulting amplicon was of correct size (fig. 2D). Sequencing of both strands by using primers SIN2, IDS13, IDS8, and ASIN3 (table 1) revealed that recognized consensus sequences for splice sites (Shapiro and Senapathy 1987) in exon 3 were unchanged. However, a single base-pair change TCG→TGG was identified at position +395 of the coding sequence, 24 bp from the 3′ end of exon 3. This transversion caused a missense mutation (i.e., S132W). To understand why the 395C→G mutation was associated with skipping of exon 3, we applied the method of Shapiro and Senapathy (1987) to determine how closely 5′ splice sites of the IDS gene (and, in particular, the natural 5′ splice site of intron 3) resembled consensus sequences for 5′ splice sites. The 5′ splice site of intron 3 was ordinated relatively low, with a score of 62 (the highest possible score being 100). This was the lowest score of all 5′ splice sites in the IDS gene, the other 5′ splice sites having an average score of 79. The 395C→G mutation also generated a sequence in exon 3, which resembled consensus sequences for 5′ splice sites more closely than either the normal coding sequence at that site or the natural 5′ splice site of intron 3 (fig. 4). The 395C→G mutation adversely affected recognition of the natural 5′ splice site of intron 3, a site with a poor consensus sequence for 5′ splice sites. With inactivation of the natural 5′ splice site of exon 3, the predominant form of splicing (>95%) was skipping of the entire exon 3. The open reading frame in this message contained a frameshift after 80 codons and termination after 152 codons. A minor population of aberrantly spliced message, which used the cryptic 5′ splice site in exon 3, created by the 395C→G substitution, was also present. The coding sequence in this message contained a frameshift at codon 132, resulting in a truncated 134-amino-acid protein. Presumably, this patient did not produce any functional IDS enzyme, consistent with his severe phenotype.

Discussion

Hunter syndrome is an inborn error of lysosomal metabolism for which the spectrum of clinical severity is thought to result from allelic heterogeneity and for which low levels of residual enzymatic activity in some individuals are believed to account for the relatively mild form. However, assays of IDS enzymatic activity have not been sufficiently sensitive to differentiate the very low activity levels of mutant gene products that would discriminate between mild and severe disease. IDS enzyme assays are limited for a number of reasons, including the low concentrations of artificial substrate used and the presumption that in vitro conditions cannot adequately simulate the microphysiological environment of the lysosome. That current enzyme assays have not been able to detect IDS activity in patients with mild disease, therefore, does not weaken the hypothesis that some patients with Hunter syndrome have slight residual IDS enzyme activity, accounting for the attenuated phenotype. However, if this hypothesis is correct, there should be a correlation between the predicted functional consequences of the mutations on the IDS enzyme and the clinical course. A further problem is that the inaccuracy of IDS assays, in combination with random lyonization of X chromosomes, results in a significant error rate (10%–20%) in carrier detection (Zlotogora and Bach 1984). An efficient and reliable method to detect the heterogeneous mutations in the IDS gene was therefore needed to study genotype-phenotype correlations further in patients with Hunter syndrome, as well as to identify more accurately female carriers of mutant IDS alleles.

Detection of Mutations by Direct Sequence Analysis

Molecular diagnosis on the basis of detection of known mutations in genes has, for many diseases, moved from research laboratories into clinical laboratories (Chehab 1993). Currently, unknown mutations in a gene are generally localized by screening gene segments for SSCP or by duplex stability analysis (Landegren 1992; Grompe 1993). These screening methods detect mutations in an indirect way (i.e., by analysis of physical structure). Each mutation must be characterized by sequencing the gene segment implicated by the screening study. Interpretation of results can therefore be difficult, and these methods have limited sensitivity. Consequently, identification of unknown mutations is still done primarily in research laboratories. This has limited the applicability of molecular diagnosis in diseases where mutations are very heterogeneous and are often unique to each patient. This situation typically occurs in autosomal dominant diseases with high mutation rates and in X-linked recessive diseases where affected patients have markedly reduced reproductive fitness. Improved methods to identify such mutations would be a crucial advancement in diagnostic medicine.

One goal of this study was to determine whether automated direct sequencing of coding sequence in amplicons, in conjunction with computer-assisted analysis of the results, is an efficient way to identify heterogeneous mutations. By sequencing the entire IDS coding region, we were able to determine mutations in 10 of the 11 patients studied. The single patient for whom no mutation was found awaits further analysis aimed at noncoding elements. Direct sequencing of the coding region, using this approach,

Table 2

Mutations of the IDS Gene

Base[a]	Codon	Exon[b]	Nucleotide Change	Amino Acid Substitution	Phenotype	Reference
Missense mutations:						
143	48	II	CGC→CCC[c]	Arg→Pro	Mild	Sukegawa et al. 1993
257	86	III	CCG→CGG	Pro→Arg	Severe	Bunge et al. 1993
281	94	III	GGC→GAC	Gly→Asp	Mild	Bunge et al. 1993
359	120	III	CCC→CGC	Pro→Arg	Severe	Bunge et al. 1993
395	132	III	TCG→TGG[d,e]	Ser→Trp	Severe	This report
404	135	III	AAA→AGA	Lys→Arg	Intermediate	Bunge et al. 1992
479	160	IV	CCT→CGT	Pro→Arg	No data	Flomen et al. 1992
587	196	V	TTG→TCG[c]	Leu→Ser	Mild	Sukegawa et al. 1993
662	221	V	CTG→CCG	Leu→Pro	Intermediate	Bunge et al. 1993
673	225	V	TAT→GAT[c]	Tyr→Asp	Intermediate	Sukegawa et al. 1993
679	227	V	AAG→CAG	Lys→Gln	Severe	Hopwood et al. 1993
685	229	V	CAC→TAC	His→Tyr	Severe	This report
998	333	VII	TCG→TTG[e]	Ser→Leu	Severe	Flomen et al. 1992; Sukegawa et al. 1993
1009	337	VIII	TGG→CGG[c]	Trp→Arg	Intermediate	Sukegawa et al. 1993
1073	358	VIII	CCC→CGC	Pro→Arg	Severe	This report
1264	422	IX	TGC→GGC	Cys→Gly[f]	Mild	Bunge et al. 1992
1402	468	IX	CGG→TGG[e]	Arg→Trp	MIld	Crotty et al. 1992; this report
1402	468	IX	CGG→GGG	Arg→Gly	Severe	Hopwood et al. 1993
1403	468	IX	CGG→CAG	Arg→Gln	Unknown[g]	Whitley et al. 1993
1406	469	IX	CCT→CAT	Pro→His	Mild	This report
1505	502	IX	TGG→TCG	Trp→Ser	No data	Flomen et al. 1992
1568	523	IX	TAT→TGT	Tyr→Cys	Mild	This report
Nonsense mutations, small deletions, and insertions:						
43	15	I	CTG→-TG	Truncated 16 aa	Mild	Hopwood et al. 1993
181-188	61–63	II	8-bp del	Truncated 61 aa	No data	Goldenfum et al. 1993
314/315	105	III	ins of C	Truncated 143 aa	Severe	This report

415	139	III	CCT→-CT	Truncated 211 aa	Severe	Bunge et al. 1993
475-477	159	IV	CAT→---	del 1 aa	Intermediate	Bunge et al. 1993
509+510	170	V	ACA→A--	Truncated 196 aa	Intermediate	Bunge et al. 1992
514	172	V	CGA→TGA[e]	Arg→term	Severe	Flomen et al. 1992; Bunge et al. 1993; this report
693	231	V	CCC→CC-	Truncated 278 aa	Severe	Bunge et al. 1993
829/830	277	VI	14 bp ins[h]	Truncated 279 aa	Severe	This report
1034 or 35	345	VIII	TGG→TAG or TGA	Trp→term	Severe	Sukegawa et al. 1993
1044	348	VIII	TAC→TA-	Truncated 347 aa	Intermediate	Bunge et al. 1993
1123	375	VIII	GAG→TAG	Glu→term	Severe	Hopwood et al. 1993
1165	389	VIII	CAG→TAG	Gln→term	Severe	This report
1312	438	IX	CTT→-TT	Truncated 438 aa	No data	Bunge et al. 1992
1327	443	IX	CGA→TGA	Arg→term	Intermediate	Bunge et al. 1992; Sukegawa et al. 1992; Froissart et al. 1993[i]
1425	475	IX	TGG→TGA	Trp→term	Intermediate	Bunge et al. 1992
1591	531	IX	CAG→TAG[e]	Gln→term	Mild	Sukegawa et al. 1993
Aberrant splicing:						
241-418	80–139	III	395C→G[d]	Exon skipping, truncated 152 aa	Severe	This report
419-2	140	In3/ExIV	aagG→agGG	1-nt ins, truncated 143 aa	Intermediate	Bunge et al. 1993
880-2	294	In6/ExVII	agC→ggc	51-nt del in ExVII	Intermediate	Bunge et al. 1993
1006, 1006+1	336	ExVII/In7	TGgt→TTTT	22-nt ins, truncated 347 aa	Severe	Bunge et al. 1993
1007-8	336	In7/ExVIII	tatgaactagGGT→tagGAACTAGGGT	7-nt ins, truncated 343 aa	Mild	Hopwood et al. 1993
1007-1	336	In7/ExVIII	agG→acg	12-nt del in ExVIII	Intermediate	Bunge et al. 1993
1122	374–394	ExVIII/In8	AGGC→AGgt	60-nt del in ExVIII	Mild	Flomen et al. 1992; Bunge et al. 1992; Bunge et al. 1993;[i] Hopwood et al. 1993

Note.—Nucleotide changes are underlined. Abbreviations: aa = amino acids; del = deletion; Ex = exon; In = intron; ins = insertion.

[a] Counting from translational initiation site (Wilson et al. 1990).

[b] Exon structure as reported by Flomen et al. (1993).

[c] Mutations inferred from amino acid changes.

[d] This patient had incomplete skipping of entire exon 3 (see text and fig. 3).

[e] Mutation due to C→T transition of CpG dinucleotide.

[f] Missense mutation involving an amino acid that is not conserved between human and murine sequences.

[g] Clinical phenotype was indeterminable, because of a coexistent chromosomal translocation.

[h] AAGATGTTTAAGGG.

[i] In this study, two patients with the mutation were noted.

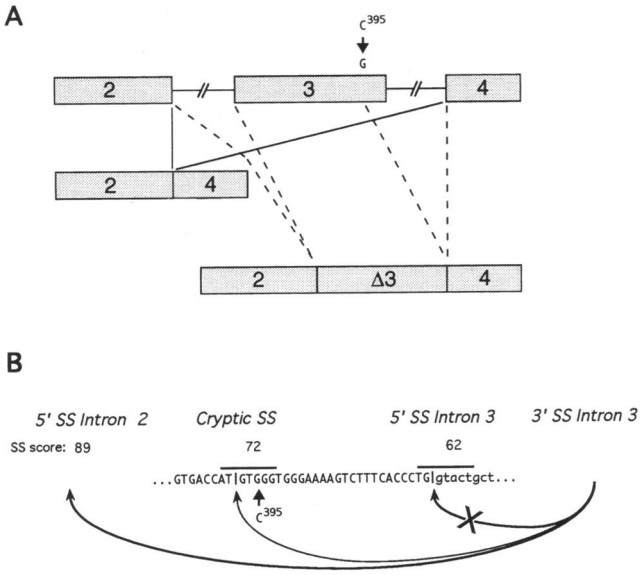

Figure 4 *A*, Diagram of aberrant splicing and exon 3 skipping in a patient with a mutation 395C→G. Exons are illustrated with boxes and are labeled. The predominant form of splicing is indicated with solid lines, and the minor population generated by using a cryptic splice site (SS) in exon 3 is indicated with dashed lines. *B*, Sequences around the 5′ SS of intron 3. Normal exon 3 sequences are represented by capital letters, and intron 3 sequences are represented by lowercase letters (Flomen et al. 1993). Locations of splice junctions are denoted with vertical bars. Consensus sequences for 5′ SSs, as defined by the method of Shapiro and Senapathy (1987), for the natural 5′ SS of intron 3 and for the fortuitous 5′ SS created by the 395C→G transversion (*arrowhead*) are indicated with a bar. Consensus SS scores, determined by the same method, are shown above each site. The normal sequence corresponding to the cryptic SS had a score of 67.

should be more sensitive and was, in our experience, more efficient than diagnostic methods that are based on analysis of physical structure followed by sequencing of the region shown to contain a variation. In addition, direct sequencing of coding regions with this automated system did not use radioactive materials, an important factor for high-volume clinical laboratories. The major drawback is the high cost of sequencing instrumentation. However, increased application of molecular diagnostics should result in high-volume use of the instruments, and, with expected improvements in sequencing technology driven by the Human Genome Project, the cost of direct sequencing should be reduced. We have demonstrated here the feasibility of using automated direct sequencing of the IDS coding region, combined with computer-assisted analysis, to identify mutations of the IDS gene. We anticipate that this approach will become more common in molecular diagnosis of other genetic diseases.

Mutational Topology of the IDS Gene

Mutations described in this laboratory, along with those described by others, have been compiled in table 2: 22 different missense mutations, 7 nonsense mutations, 10 small insertions/deletions, and 7 mutations causing aberrant splicing are shown. In analyzing the different mutations collectively, we made two assumptions. First, we believe that patients described in the literature are unrelated. Second, we assumed the causality of the described mutations in obliterating IDS activity. This can only be demonstrated directly by either finding these mutations repeatedly in unrelated patients (and never in normal alleles) or by expression studies of the mutant coding sequences in cultured cells. Currently, the only mutations in the coding sequence that have been found repeatedly in unrelated patients are the C→T transitions in CpG nucleotides at bases 514, 998, 1327, and 1402. The only mutations tested with expression studies are the missense mutation at R468 (Crotty et al. 1992) and the nonsense mutation R443X (Sukegawa et al. 1992) and Q531X (Sukegawa et al. 1993). Mutations were plotted (fig. 5) with respect to the recently described exon organization of the gene (Flomen et al. 1993; Wilson et al. 1993). Only one mutation each has been identified in exon 1 (104 bp) and exon 6 (171 bp). Only two mutations have been identified in exon 2 (137 bp). Exon 1 is very poorly conserved between the human and the murine IDS genes (Daniele et al. 1993) (fig. 6). This suggests that many residues in this exon are not critical for function of the enzyme. In contrast, the relatively low number of mutations found in exons 2 and 6 cannot be explained in this way, insofar as these regions are highly conserved between human and murine coding sequences (fig. 6). Furthermore, exon 2 contains many residues that are highly conserved between the various sulfatases (Tomatsu et al. 1991). Most mutations in the IDS gene were located in 178-bp exon 3 (9 patients), 201-bp exon 5 (10 patients), 174-bp exon 8 (9 patients), and 471-bp exon 9 (13 patients).

We queried whether the severity of disease in patients with Hunter syndrome corresponds to the location of mutations in the IDS coding sequence. Four missense mutations (where the phenotype was known) in exons 2–8 were associated with mild disease, four were associated with intermediate disease, and seven were associated with severe disease. In exon 9, in contrast, four missense mutations were associated with mild disease, and only one was associated with severe disease. Nonsense mutations have been identified in exons 5, 8, and 9. The four different nonsense mutations in exons 5 and 8 were all found in patients with severe disease. Three of four nonsense mutations identified in exon 9 were found in patients with intermediate disease, and one was found in a patient with mild disease. Such point mutations of exon 9 tended to result in mild disease more often than mutations in other exons. Mutations altering the length of the coding sequence (i.e., small insertions, small deletions, or aberrant splicing) were also associated with various phenotypes (fig. 5). The pattern was comparable to that observed with single base substitutions, insofar as frameshifts and truncations more localized to 5′ in the gene were more often associated with a severe phenotype (fig. 5).

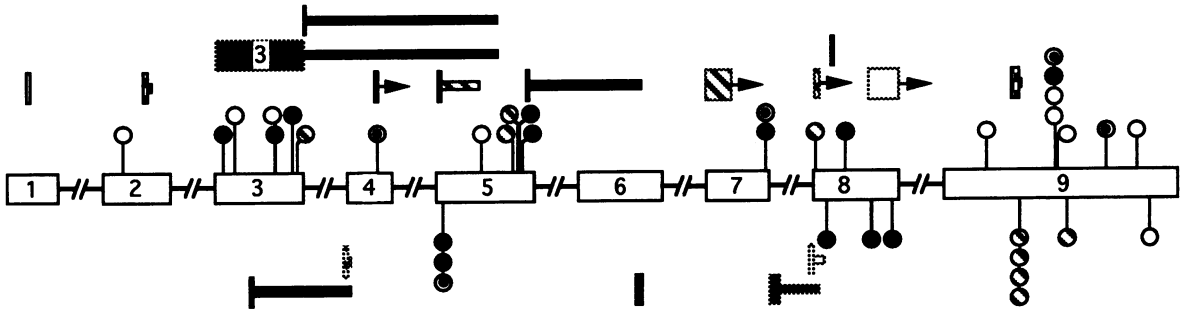

Figure 5 Location and type of mutations in the human IDS gene. The exons of the IDS gene are illustrated with open boxes and numbered 1 through 9. Deletions and insertions in the coding sequence are indicated with thicker bars, above and below the diagram, respectively. Narrower bars indicate extent of aberrant amino acid residues incorporated 3′ of frameshift mutations. Bars representing in-frame deletions are indicated with a horizontal arrow to the right. Bars representing deletions or insertions resulting from aberrant splicing have a stippled outline. Base substitutions are indicated as circles, with missense mutations above and nonsense mutations below the diagram. Phenotypes associated with mutations are indicated with patterns: blackened circles represent severe Hunter syndrome; hatched circles represent intermediate Hunter syndrome; open circles represent the mild form; and a dotted circle or stippled box denotes that phenotype was not known.

The C→T transitions in CpG dinucleotides are known to be relative "hotspots" for mutation, according to the methylation-associated deamination model (Cooper and Krawczak 1990). In the IDS gene, C→T transitions at CpG dinucleotides occurred 18 times (seven different mutations) in 46 mutations (57 patients) (i.e., in 32% of the patients) (table 2). This percentage was comparable to the relatively high frequency of such mutations in other genes (Cooper and Krawczak 1990). In addition, hotspots for mutation in the IDS gene, defined as those that have been identified in two or more unrelated patients, were all C→T transitions in CpG dinucleotides (see table 2, bases 514, 998, 1327, and 1402; and fig. 5).

We had previously identified two patients with missense mutations at codon 468 (i.e., R468W [1402C→T] [Crotty et al. 1992] and R468Q [1403G→A] [Whitley et al. 1993]). Both of these mutations result from C→T transitions in a CpG dinucleotide in the template strand or in the transcribed strand, respectively (table 2). In the interim, Hopwood et al. (1993) described a patient with the R468G (1402C→G) mutation, and, in this paper, we identified two new additional patients with mutations at this site (i.e., R468W [1402C→T] and P469H [1406C→A]). The patient with the R468W mutation was unrelated to the first patient we described. In addition, the coding sequence of our two R468W patients differed at the C-versus-T silent polymorphism at position 438 (Hopwood et al. 1993), further confirming the independent origin of these two mutations. Interestingly, the R468G and the P469H mutations were not C→T transitions in CpG dinucleotides. Thus, codons 468 and 469 constitute a hotspot for mutations, and the corresponding amino acid residues also appear to be critical for function of the IDS enzyme.

All mutations described to date, except the C422G mutation found in a patient with a very mild disease (Bunge et al. 1992), involved amino acids that are conserved between the human and the murine IDS enzymes (fig. 6). In contrast, missense mutations of the IDS gene generally did not involve amino acid residues previously identified as being highly homologous among seven sulfatases compared by Tomatsu et al. (1991). The only exceptions were three of five missense mutations in exon 3 (fig. 6); however, exon 3 contains the highest number of highly homologous residues (10) of all the exons of the IDS gene (Tomatsu et al. 1991). Interesting in this regard is the concentration of four mutations associated with intermediate or severe disease in a short segment at the 3′ end of exon 5, which is unique among sulfatases to IDS (fig. 6). This segment is fully conserved between human and murine sequences, and none of these four mutations involve C→T transitions of CpG dinucleotides. Taken together, these findings suggest that this unique segment in exon 5 is of critical importance for IDS function.

The 395C→G substitution resulting in skipping and aberrant splicing of exon 3 was particularly interesting. Creation of cryptic 5′ splice sites has been described in other systems (Krawczak et al. 1992). These cryptic 5′ splice sites have in common that the sequence resembles the consensus sequences for 5′ splice sites more closely than the downstream natural 5′ splice sites for which they substitute. Even more intriguing was the observation that the 395C→G mutation apparently caused exon skipping, also. The association of nonsense mutations with exon skipping has been described and has been postulated to be associated with recognition of the nonsense codon before RNA splicing (Dietz et al. 1993). The association of missense mutations in our patient and another (Wakamatsu et al. 1992) with skipping of exons illustrates that exon skipping can be due to missense mutations, as well, and suggests that changes in RNA secondary structure resulting from sequence changes may have contributed to the aberrant splicing.

It has been shown that deletions involving the IDS gene result in severe disease (Hopwood et al. 1993). All mutations in the coding sequence likely to be "null mutations" were associated with severe disease, and mutations in pa-

tients with mild disease were theoretically compatible with slight residual IDS activity. Importantly, when recurrent mutations are encountered in unrelated patients, the phenotype appeared to be similar. All current information on mutations of the IDS gene, including mutations of the coding sequence, is therefore in accordance with the hypothesis that variation in severity of disease reflects primarily allelic heterogeneity at the IDS locus. Other possible causes for variation in clinical manifestations in patients with Hunter syndrome include microdeletion syndromes with loss of gene(s) contiguous to the IDS locus, as well as differences in genetic backgrounds and environmental factors interacting with mutant IDS alleles. Contributions from these factors, if any, to the phenotype have not been characterized to date.

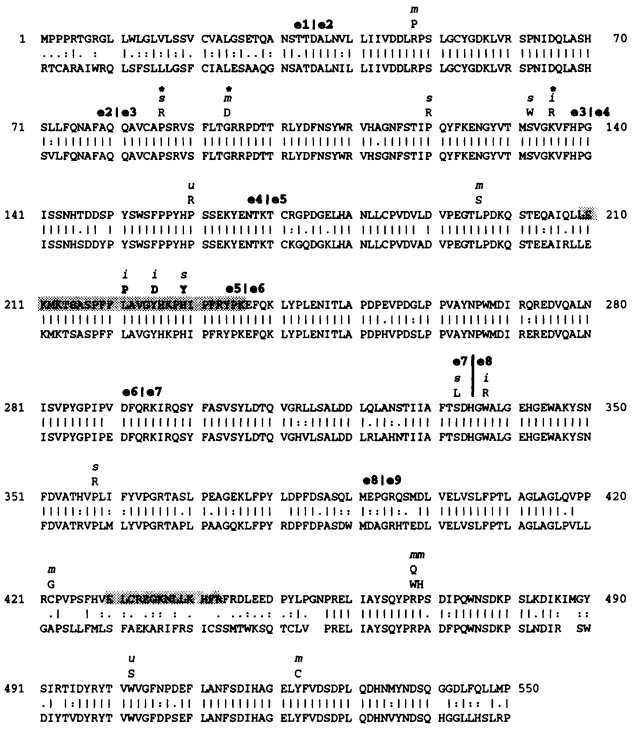

Figure 6 Interspecies conservation of amino acid residues involved in missense mutations in patients with Hunter syndrome. Amino acid sequences (single-letter abbreviation) of human (*upper*) and murine (*lower*) coding regions are shown. Homology plot is modified from Daniele et al. (1993). Residues R464 and H507 in the murine sequence, which have no human counterparts, have been deleted. Murine residues 231 and 495 are indicated as alanine and serine, respectively (i.e., as conserved between the human and murine coding sequences) (J. J. Jonsson, R. Decker, and C. B. Whitley, unpublished data), rather than as glycine and threonine, as reported by Daniele et al. (1993). Exon (e) junctions in the human sequence are shown. Sequences unique to human IDS among the homologous sulfatases as defined by Tomatsu et al. (1991) are shaded. Missense mutations are displayed above the human sequence, with phenotype indicated with italic letters as follows: *m* = mild phenotype; *i* = intermediate phenotype; *s* = severe phenotype; and *u* = phenotype unknown. Missense mutations involving highly homologous amino acids, as defined from analysis of seven sulfatase sequences by Tomatsu et al. (1991), are indicated with an asterisk (*).

Acknowledgments

We thank Dr. Eric D. Eccleston and Sheila St. Cyr for helpful advice and technical expertise. This work was supported by a grant from the Ronald McDonald Children's Charity and by National Institutes of Health grant DK39891.

References

Annella T, Daniele A, DiNatale PD (1993) Heterogeneity of DNA and RNA in Hunter patients. Hum Genet 92:350–352

Beck M, Steglich C, Zabel B, Dahl N, Schwinger E, Hopwood JJ, Gal A (1992) Deletion of the Hunter gene and both DXS466 and DXS304 in a patient with mucopolysaccharidosis type II. Am J Med Genet 44:100–103

Braun SE, Aronovich EL, Anderson RA, Crotty PL, McIvor RS, Whitley CB (1993) Metabolic correction and cross-correction of mucopolysaccharidosis type II and expression of human iduronate-2-sulfatase. Proc Natl Acad Sci USA 90:11830–11834

Bunge S, Steglich C, Beck M, Rosenkranz W, Schwinger E, Hopwood JJ, Gal A (1992) Mutation analysis of the iduronate-2-sulfatase gene in patients with mucopolysaccharidosis type II (Hunter syndrome). Hum Mol Genet 1:335–339

Bunge S, Steglich C, Zuther C, Beck M, Morris CP, Schwinger E, Schinzel A, et al (1993) Iduronate-2-sulfatase gene mutations in 16 patients with mucopolysaccharidosis type II (Hunter syndrome). Hum Mol Genet 2:1871–1875

Chehab FF (1993) Molecular diagnostics: past, present, and future. Hum Mutat 2:331–337

Chomczynski P, Sacchi N (1987) Single-step method of RNA isolation by acid guanidinium thiocyante-phenol-chloroform extraction. Anal Biochem 162:156–159

Cooper DN, Krawczak M (1990) The mutational spectrum of single base-pair substitutions causing human genetic disease: patterns and predictions. Hum Genet 85:55–74

Crotty PL, Braun SE, Anderson RA, Whitley CB (1992) Mutation R468W of the iduronate-2-sulfatase gene in mild Hunter syndrome (mucopolysaccharidosis type II) confirmed by *in vitro* mutagenesis and expression. Hum Mol Genet 1:755–757

Crotty PL, Whitley CB (1992) Assessment of iduronate-2-sulfatase mRNA expression in Hunter syndrome (mucopolysaccharidosis type II). Hum Genet 90:285–288

Daniele A, Faust CJ, Herman GE, di Natale P, Ballabio A (1993) Cloning and characterization of the cDNA for the murine iduronate sulfatase gene. Genomics 16:755–757

Dear S, Staden R (1991) A sequence assembly and editing program for efficient management of large projects. Nucleic Acids Res 19:3907–3911

Dietz HC, Valle D, Francomano CA, Kendzior RJ Jr, Pyeritz RE, Cutting GR (1993) The skipping of constitutive exons in vivo induced by nonsense mutations. Science 259:680–683

Flomen RH, Green EP, Green PM, Bentley DR, Giannelli F (1993) Determination of the organisation of coding sequences within the iduronate sulphate sulphatase (IDS) gene. Hum Mol Genet 2:5–10

Flomen RH, Green PM, Bentley DR, Giannelli F, Green EP (1992) Detection of point mutations and a gross deletion in six Hunter syndrome patients. Genomics 13:543–550

Froissart R, Blond JL, Maire I, Guibaud P, Hopwood JJ, Mathieu

M, Bozon D (1993) Hunter syndrome: gene deletions and rearrangements. Hum Mutat 2:138–140

Gal A, Beck M, Sewell AC, Morris CP, Schwinger E, Hopwood JJ (1992) Gene diagnosis and carrier detection in Hunter syndrome by the iduronate-2-sulphatase cDNA probe. J Inherit Metab Dis 15:342–346

Goldenfum S, Malcolm S, Young E, Winchester B (1993) Molecular analysis of mucopolysaccharidosis type II (Hunter disease). Paper presented at the Third International Symposium on Mucopolysaccharidosis and Related Diseases, Essen, May 28–31

Grompe M (1993) The rapid detection of unknown mutations in nucleic acids. Nat Genet 5:111–117

Haldane JBS (1947) The mutation rate of the gene for haemophilia and its segregation ratios in males and females. Ann Eugenics 13:262–271

Hopwood JJ, Bunge S, Morris CP, Wilson PJ, Steglich C, Beck M, Schwinger E, et al (1993) Molecular basis of mucopolysaccharidosis type II: mutations in the iduronate-2-sulphatase gene. Hum Mutat 2:435–442

Hopwood JJ, Morris CP (1990) The mucopolysaccharidoses: diagnosis, molecular genetics and treatment. Mol Biol Med 7:381–404

Krawczak M, Reiss J, Cooper DN (1992) The mutational spectrum of single base-pair substitutions in mRNA splice junctions of human genes: causes and consequences. Hum Genet 90: 41–54

Landegren U (1992) Detection of mutations in human DNA. Genet Anal Tech Appl 9:3–8

Le Guern E, Couillin, P, Oberle, I, Ravise N, Boue J (1990) More precise localization of the gene for Hunter syndrome. Genomics 7:358–362

Lichtman MA (1990) Separation of leucocytes. In: Williams WJ, Beutler E, Erslev AJ, Lichtman MA (eds) Hematology, 4th ed. McGraw-Hill, New York, pp 1742–1744

Neufeld EF, Muenzer J (1989) The mucopolysaccharidoses. In: Scriver CR, Beaudet AL, Sly WS, Valle D (eds) The metabolic basis of inherited disease, 6th ed. McGraw-Hill, New York, pp 1565–1587

Palmieri G, Capra V, Romano G, D'Urso M, Johnson S, Schlessinger D, Morris P, et al (1992) The iduronate sulfatase gene: isolation of a 1.2-Mb YAC contig spanning the entire gene and identification of heterogeneous deletions in patients with Hunter syndrome. Genomics 12:52–57

Schroder W, Petruschka L, Wehnert M, Zschiesche M, Seidlitz G, Hopwood JJ, Herrmann FH (1993) Carrier detection of Hunter syndrome (MPS II) by biochemical and DNA techniques in families at risk. J Med Genet 30:210–213

Shapiro MB, Senapathy P (1987) RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. Nucleic Acids Res 15:7155–7174

Steen-Bondeson ML, Dahl N, Tonnesen T, Kleijer WJ, Seidlitz G, Gustavson KH, Wilson PJ, et al (1992) Molecular analysis of patients with Hunter syndrome: implication of a region prone to structural alterations within the IDS gene. Hum Mol Genet 1:195–198

Sukegawa K, Tomatsu S, Tamai K, Ikeda M, Sasaki T, Masue M, Fukuda S, et al (1992) Intermediate form of mucopolysaccharidosis type II (Hunter disease): a C1327 to T substitution in the iduronate sulfatase gene. Biochem Biophys Res Commun 183:809–813

Sukegawa K, Yamada Y, Tomatsu S, Fukuda S, Orii, T (1993) Molecular analysis of mucopolysaccharidosis type II (Hunter disease). Paper presented at the Third International Symposium on Mucopolysaccharidosis and Related Diseases, Essen, May 28–31

Tomatsu S, Fukuda S, Masue M, Sukegawa K, Fukao T, Yamagishi A, Hori T, et al (1991) Morquio disease: isolation, characterization, and expression of full-length cDNA for human N-acetylgalactosamine-6-sulfate sulfatase. Biochem Biophys Res Commun 181:677–683

Wakamatsu N, Kobayashi H, Miyatake T, Tsuji S (1992) A novel exon mutation in the human beta-hexosaminidase beta subunit gene affects 3′ splice selection. J Biol Chem 267:2406–2413

Wehnert M, Hopwood JJ, Schroder W, Herrmann FH (1992) Structural gene aberrations in mucopolysaccharidosis II (Hunter). Hum Genet 89:430–432

Whitley CB (1992) The mucopolysaccharidoses. In: Beighton P (ed) McKusick's heritable disorders of connective tissue, 5th ed. Mosby Year Book, St Louis, pp 367–499

Whitley CB, Anderson RA, Aronovich EL, Crotty PL, Anyane-Yeboa K, Russo D, Warburton D (1993) Caveat to genotype-phenotype correlation in mucopolysaccharidosis type II: discordant clinical severity of R468W and R468Q mutations of the iduronate-2-sulfatase gene. Hum Mutat 2:235–237

Wilson PJ, Meaney CA, Hopwood JJ, Morris CP (1993) Sequence of the human iduronate 2-sulfatase (IDS) gene. Genomics 17:773–775

Wilson PJ, Morris CP, Anson DS, Occhiodoro T, Bielicki J, Clements PR, Hopwood JJ (1990) Hunter syndrome: isolation of an iduronate-2-sulfatase cDNA clone and analysis of patient DNA. Proc Natl Acad Sci USA 87:8531–8535

Wilson PJ, Suthers GK, Callen DF, Baker E, Nelson PV, Cooper A, Wraith JE, et al (1991) Frequent deletions at Xq28 indicate genetic heterogeneity in Hunter syndrome. Hum Genet 86:505–508

Wraith JE, Cooper A, Thornley M, Wilson PJ, Nelson PV, Morris CP, Hopwood JJ (1991) The clinical phenotype of two patients with a complete deletion of the iduronate-2-sulphatase gene (mucopolysaccharidosis II—Hunter syndrome). Hum Genet 87:205–206

Yamada Y, Tomatsu S, Sukegawa K, Suzuki Y, Kondo N, Hopwood JJ, Orii T (1993) Mucopolysaccharidosis type II (Hunter disease): 13 gene mutations in 52 Japanese patients and carrier detection in four families. Hum Genet 92:110–114

Young ID, Harper PS (1982) Mild form of Hunter's syndrome: clinical delineation based on 31 cases. Arch Dis Child 57:828–836

Young ID, Harper PS, Newcombe RG, Archer IM (1982) A clinical and genetic study of Hunter's syndrome. 2. Differences between the mild and severe forms. J Med Genet 19:408–411

Zlotogora J, Bach G (1984) Heterozygote detection in Hunter syndrome. Am J Med Genet 17:661–665