Proc. Natl. Acad. Sci. USA
Vol. 87, pp. 8531–8535, November 1990
Medical Sciences

Hunter syndrome: Isolation of an iduronate-2-sulfatase cDNA clone and analysis of patient DNA

(mucopolysaccharidosis type II/lysosomal storage disorder/X chromosome-linked mutations/sulfatase sequence homology)

P. J. WILSON, C. P. MORRIS, D. S. ANSON, T. OCCHIODORO, J. BIELICKI, P. R. CLEMENTS, AND J. J. HOPWOOD*

Lysosomal Diseases Research Unit, Department of Chemical Pathology, Adelaide Children's Hospital, North Adelaide S.A. 5006, Australia

Communicated by Elizabeth F. Neufeld, August 2, 1990

ABSTRACT **Iduronate 2-sulfatase (IDS, EC 3.1.6.13) is required for the lysosomal degradation of heparan sulfate and dermatan sulfate. Mutations causing IDS deficiency in humans result in the lysosomal storage of these glycosaminoglycans and Hunter syndrome, an X chromosome-linked disease. We have isolated and sequenced a 2.3-kilobase cDNA clone coding for the entire sequence of human IDS. Analysis of the deduced 550-amino acid IDS precursor sequence indicates that IDS has a 25-amino acid amino-terminal signal sequence, followed by 8 amino acids that are removed from the proprotein. An internal proteolytic cleavage occurs to produce the mature IDS present in human liver shown to contain a 42-kDa polypeptide N-terminal to a 14-kDa polypeptide. The IDS sequence has strong sequence homology with other sulfatases (such as sea urchin arylsulfatase, human arylsulfatases A, B, and C, and human glucosamine 6-sulfatase), suggesting that the sulfatases comprise an evolutionarily related family of genes that arose by gene duplication and divergent evolution. The arylsulfatases have a greater homology with each other than with the non-arylsulfatases (IDS and glucosamine 6-sulfatase). The IDS cDNA detected RNA species of 5.7, 5.4, 2.1, and 1.4 kilobases in human placental RNA and revealed structural alterations and gross deletions of the IDS gene in many of the clinically severe Hunter syndrome patients studied.**

Iduronate 2-sulfatase (IDS, EC 3.1.6.13) acts as an exosulfatase in lysosomes to hydrolyze the C2-sulfate ester bond from non-reducing-terminal iduronic acid residues in the glycosaminoglycans heparan sulfate and dermatan sulfate (1). IDS is one of a family of at least nine sulfatases that hydrolyze sulfate esters in human cells. They are all lysosomal enzymes that act on sulfated monosaccharide residues in a variety of complex substrates with the exception of microsomal steroid sulfatase (or arylsulfatase C), which acts on sulfated 3β-hydroxysteroids (1, 2). Each sulfatase displays absolute substrate specificity, making the sulfatase family an attractive model to investigate the molecular requirements for substrate binding and the catalysis of sulfate ester hydrolysis.

A deficiency in the activity of IDS in humans leads to the lysosomal accumulation of heparan sulfate and dermatan sulfate fragments and their excretion in urine (1). This storage results in the clinical disorder Hunter syndrome (mucopolysaccharidosis type II, MPS-II) in which patients may present with variable phenotypes from severe mental retardation, skeletal deformities, and stiff joints to a relatively mild course (1). It has been postulated that this clinical heterogeneity reflects different mutations at the IDS locus affecting enzyme expression, stability, or function. MPS-II is one of the most common mucopolysaccharidoses and is the only one that is X chromosome-linked (1).

We have reported (3) the purification of IDS from human liver and other tissues and in this communication we report the nucleotide sequence for a full-length cDNA clone for IDS† from human endothelial cells. We present data demonstrating sequence homology with other sulfatases and the detection of deletions and rearrangements in the IDS gene of severely affected MPS-II patients.

MATERIALS AND METHODS

Materials. Form A of IDS was purified from human liver as described (3). Restriction endonucleases, polynucleotide kinase, T4 DNA ligase, the Klenow fragment of DNA polymerase I, and M13 sequencing kits were from Boehringer Mannheim. GeneScreen*Plus* nylon filters were from DuPont/NEN. [γ-^{32}P]ATP (5000 Ci/mmol; 1 Ci = 37 GBq), [α-^{32}P]dCTP (3000 Ci/mmol), and Multiprime DNA labeling kit were from Amersham. Oligo(dT)-cellulose and Sephadex G-50 were from Pharmacia P-L Biochemicals. The X chromosome genomic library LA0XNL01 was from the American Tissue Culture Collection, and the λgt10 random-primed human colon cDNA library (1.5×10^6 independent clones) and the λgt11 human endothelial cDNA library (2.1×10^6 independent clones) were from Clontech.

Polypeptide Isolation and Sequencing. Approximately 20 μg of form A liver IDS was subjected to SDS/polyacrylamide gel electrophoresis and transferred to an Immobilon P membrane (Millipore) (4) with modifications of overnight preelectrophoresis of the SDS/polyacrylamide gel and the addition of 200 μl of 100 mM sodium thioglycollate to the cathode buffer chamber before electrophoresis. The 42-kDa and the 14-kDa polypeptides were excised and directly amino-terminal sequenced by Bresatec (Adelaide, Australia).

Library Screening. A 49-mer oligonucleotide sequence (3′-ACTAGTAGCACCTGCTGGACGCCGGGAGGGACCCGCTGATGCTGCTGCA-5′) was designed from the amino-terminal amino acid sequence (using residues 8–24 of TSALNVLLIIVDDLRPSLGDYDDVL) of the 42-kDa IDS polypeptide. T4 polynucleotide kinase and [γ-^{32}P]ATP were used to end-label the 49-mer to a specific activity greater than 10^7 cpm/μg for screening of the X chromosome library. The bacterial host used was NM538 and 2×10^5 recombinants were screened at a density of 30,000 plaque-forming units per plate. Positive clones were plaque-purified, DNA was isolated from lysates, and the inserts were analyzed in 1% agarose gels by Southern hybridization using the labeled 49-mer. A 49-mer positive 1.6-kilobase (kb) *Hin*dIII genomic

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: IDS, iduronate 2-sulfatase; MPS-II, mucopolysaccharidosis type II.

*To whom reprint requests should be addressed.

†The sequence reported in this paper has been deposited in the GenBank data base (accession no. M38371).

DNA fragment was labeled with [α-^{32}P]dCTP using a Multiprime DNA labeling kit and used to screen the human colon cDNA library. Approximately 5×10^5 recombinants were screened at a density of 55,000 plaque-forming units per plate using the bacterial host C600. A 300-base-pair (bp) *Hin*dIII–*Eco*RI fragment from the 3′ end of a 1.5-kb colon cDNA clone (λc2S15) was labeled and used to screen the human endothelial cDNA library. The bacterial host used was NM538 and 5×10^5 recombinants were screened at a density of 40,000 plaque-forming units per plate.

Nucleotide Sequencing. Sonicated DNA fragments generated from the 1.5-kb cDNA insert were subcloned into M13mp19 for nucleotide sequence analysis by the dideoxynucleotide chain-termination method by using the Klenow fragment of DNA polymerase I at 45°C (5). Some internal regions of the 1.5-kb cDNA were sequenced using primers labeled at their 5′ ends with [γ-^{32}P]ATP with single-stranded DNA templates generated by asymmetric polymerase chain reactions. The remaining coding sequence and the 5′ and 3′ untranslated regions present on the 2.3-kb endothelial cDNA were sequenced using specific primers on M13 subclones.

Southern Blot Analysis of MPS-II Patients. DNA from MPS-II patients and normal control cultured fibroblasts was prepared and digested with *Pst* I (6) and transferred to GeneScreen*Plus* nylon membrane. The cDNA fragment λc2S15 was radiolabeled using the Multiprime DNA labeling kit and purified by gel filtration on a 1-ml Sephadex G-50 column. The nylon filter was prehybridized, hybridized, and washed according to the manufacturer's instructions.

RNA Isolation and Northern Blot Analysis. Total RNA was isolated from placental tissue by using a single-step guanidinium thiocyanate method (7). Poly(A)$^+$ RNA was obtained by oligo(dT)-cellulose chromatography and characterized by Northern Blot analysis carried out after electrophoresis in a 0.8% agarose/2.2 M formaldehyde gel and transfer to GeneScreen*Plus* nylon membrane. Prehybridization, hybridization, and washing were performed according to the manufacturer's instructions. Radiolabeled λc2S15, prepared and purified as described above, was used in all hybridization experiments.

Sequence Analysis. The nucleotide sequence was screened against the GenBank nucleotide sequence data base (Release 62.0, December 1989) and the encoded protein sequence was screened against the National Biomedical Research Foundation protein data base (Release 23.0, December 1989). General sequence analysis and the multiple protein sequence

```
CGGCTGTGTTGCGCAGTCTTCATGGGTTCCCGACGAGGAGGTCTCTGTGGCTGCGGCGGCTGCTAACTGCGCCACCTGCTGCAGCCTGTCCCCGCCGCTCTGAAGCGGCCGCGTCGAAGC  120

   M  P  P  P  R  T  G  R  G  L  L  W  L  G  L  V  L  S  S  V  C  V  A  L  G  S  E  T  Q  A  N  S  T  T  D  A  L  N  V  39
CGAAATGCCGCCACCCCGGACCGGCCGAGGCCTTCTCTGGCTGGGTCTGGTTCTGAGCTCCGTCTGCGTCGCCCTCGGATCCGAAACGCAGGCCAACTCGACCACAGATGCTCTGAACGT  240

   L  L  I  I  V  D  D  L  R  P  S  L  G  C  Y  G  D  K  L  V  R  S  P  N  I  D  Q  L  A  S  H  S  L  L  F  Q  N  A  F  A  79
TCTTCTCATCATCGTGGATGACCTGCGCCCCTCCCTGGGCTGTTATGGGGATAAGCTGGTGAGGTCCCCAAATATTGACCAACTGGCATCCCACAGCCTCCTCTTCCAGAATGCCTTTGC  360

   Q  Q  A  V  C  A  P  S  R  V  S  F  L  T  G  R  R  P  D  T  T  R  L  Y  D  F  N  S  Y  W  R  V  H  A  G  N  F  S  T  I  119
GCAGCAAGCAGTGTGCGCCCCGAGCCGCGTTTCTTTCCTCACTGGCAGGAGACCTGACACCACCCGCCTGTACGACTTCAACTCCTACTGGAGGGTGCACGCTGGAAACTTCTCCACCAT  480

   P  Q  Y  F  K  E  N  G  Y  V  T  M  S  V  G  K  V  F  H  P  G  I  S  S  N  H  T  D  D  S  P  Y  S  W  S  F  P  P  Y  H  159
CCCCCAGTACTTCAAGGAGAATGGCTATGTGACCATGTCGGTGGGAAAAGTCTTTCACCCTGGGATATCTTCTAACCATACCGATGATTCTCCGTATAGCTGGTCTTTTCCACCTTATCA  600

   P  S  S  E  K  Y  E  N  T  K  T  C  R  G  P  D  G  E  L  H  A  N  L  L  C  P  V  D  V  L  D  V  P  E  G  T  L  P  D  K  199
TCCTTCCTCTGAGAAGTATGAAAACACTAAGACATGTCGAGGGCCAGATGGAGAACTCCATGCCAACCTGCTTTGCCCTGTGGATGTGCTGGATGTTCCCGAGGGCACCTTGCCTGACAA  720

   Q  S  T  E  Q  A  I  Q  L  L  E  K  M  K  T  S  A  S  P  F  F  L  A  V  G  Y  H  K  P  H  I  P  F  R  Y  P  K  E  F  Q  239
ACAGAGCACTGAGCAAGCCATACAGTTGTTGGAAAAGATGAAAACGTCAGCCAGTCCTTTCTTCCTGGCCGTTGGGTATCATAAGCCACACATCCCCTTCAGATACCCCAAGGAATTTCA  840

   K  L  Y  P  L  E  N  I  T  L  A  P  D  P  E  V  P  D  G  L  P  P  V  A  Y  N  P  W  M  D  I  R  Q  R  E  D  V  Q  A  L  279
GAAGTTGTATCCCTTGGAGAACATCACCCTGGCCCCCGATCCCGAGGTCCCTGATGGCCTACCCCCTGTGGCCTACAACCCCTGGATGGACATCAGGCAACGGGAAGACGTCCAAGCCTT  960

   N  I  S  V  P  Y  G  P  I  P  V  D  F  Q  R  K  I  R  Q  S  Y  F  A  S  V  S  Y  L  D  T  Q  V  G  R  L  L  S  A  L  D  319
AAACATCAGTGTGCCGTATGGTCCAATTCCTGTGGACTTTCAGCGGAAAATCCGCCAGAGCTACTTTGCCTCTGTGTCATATTTGGATACACAGGTCGGCCGCCTCTTGAGTGCTTTGGA  1080

   D  L  Q  L  A  N  S  T  I  I  A  F  T  S  D  H  G  W  A  L  G  E  H  G  E  W  A  K  Y  S  N  F  D  V  A  T  H  V  P  L  359
CGATCTTCAGCTGGCCAACAGCACCATCATTGCATTTACCTCGGATCATGGGTGGGCTCTAGGTGAACATGGAGAATGGGCCAAATACAGCAATTTTGATGTTGCTACCCATGTTCCCCT  1200

   I  F  Y  V  P  G  R  T  A  S  L  P  E  A  G  E  K  L  F  P  Y  L  D  P  F  D  S  A  S  Q  L  M  E  P  G  R  Q  S  M  D  399
GATATTCTATGTTCCTGGAAGGACGGCTTCACTTCCGGAGGCAGGCGAGAAGCTTTTCCCTTACCTCGACCCTTTTGATTCCGCCTCACAGTTGATGGAGCCAGGCAGGCAATCCATGGA  1320

   L  V  E  L  V  S  L  F  P  T  L  A  G  L  A  G  L  Q  V  P  P  R  C  P  V  P  S  F  H  V  E  L  C  R  E  G  K  N  L  L  439
CCTTGTGGAACTTGTGTCTCTTTTTCCCACGCTGGCTGGACTTGCAGGACTGCAGGTTCCACCTCGCTGCCCCGTTCCTTCATTTCACGTTGAGCTGTGCAGAGAAGGCAAGAACCTTCT  1440

   K  H  F  R  F  R  D  L  E  E  D  P  Y  L  P  G  N  P  R  E  L  I  A  Y  S  Q  Y  P  R  P  S  D  I  P  Q  W  N  S  D  K  479
GAAGCATTTTCGATTCCGTGACTTGGAAGAGGATCCGTACCTCCCTGGTAATCCCCGTGAACTGATTGCCTATAGCCAGTATCCCCGGCCTTCAGACATCCCTCAGTGGAATTCTGACAA  1560

   P  S  L  K  D  I  K  I  M  G  Y  S  I  R  T  I  D  Y  R  Y  T  V  W  V  G  F  N  P  D  E  F  L  A  N  F  S  D  I  H  A  519
GCCGAGTTTAAAAGATATAAAGATCATGGGCTATTCCATACGCACCATAGACTATAGGTATACTGTGTGGGTTGGCTTCAATCCTGATGAATTTCTAGCTAACTTTTCTGACATCCATGC  1680

   G  E  L  Y  F  V  D  S  D  P  L  Q  D  H  N  M  Y  N  D  S  Q  G  G  D  L  F  Q  L  L  M  P                              550
AGGGGAACTGTATTTTGTGGATTCTGACCCATTGCAGGATCACAATATGTATAATGATTCCCAAGGTGGAGATCTTTTCCAGTTGTTGATGCCTTGAGTTTTGCCAACCATGGATGGCAA  1800
ATGTGATGTGCTCCCTTCCAGCTGGTGAGAGGAGGAGTTAGAGCTGGTCGTTTTGTGATTACCCATAATATTGGAAGCAGCCTGAGGGCTAGTTAATCCAAACATGCATCAACAATTTGG  1920
CCTGAGAATATGTAACAGCCAAACCTTTTCGTTTAGTCTTTATTAAAATTTATAATTGGTAATTGGACCAGTTTTTTTTTTAATTTCCCTCTTTTTAAAACAGTTACGGCTTATTTACTG  2040
AATAAATACAAAGCAAACAAACTCAAGTTATGTCATACCTTTGGATACGAAGACCATACATAATAACCAAACATAACATTATACACAAAGAATACTTTCATTATTTGTGGAATTTAGTGC  2160
ATTTCAAAAAGTAATCATATATCAAACTAGGCACCACACTAAGTTCCTGATTATTTTGTTTATAATTTAATAATATATCTTATGAGCCCTATATATTCAAAATATTATGTTAACATGTAA  2280
TCCATGTTTCTTTTTCC                                                                                                          2297
```

FIG. 1. Compiled nucleotide sequence of the IDS cDNA clones and the deduced amino acid sequence of the encoded protein. Amino acid sequence is shown in the one-letter code above the nucleotide sequence. Nucleotide and amino acid numbers are depicted on the right margin. Possible sites for peptidase cleavage of the signal peptide are indicated with arrows. Underlined amino acids are colinear with amino-terminal sequences (14 kDa, Pro-Arg-Glu-Leu-Ile-Ala-Tyr-Ser-Xaa-Tyr-Pro-Arg-Xaa-Xaa-Ile-Pro, determined by direct sequence analysis). Potential N-glycosylation sites are starred. A potential polyadenylylation signal is doubly underlined.

alignment were performed using programs from Reisner and Bucholtz (8) and Lipman *et al.* (9), respectively.

RESULTS AND DISCUSSION

IDS from human liver can be purified to two major forms (A and B) which have different pI values and contain both 42-kDa and 14-kDa polypeptides (3). The 42-kDa and 14-kDa polypeptides in form A were subjected to direct amino-terminal amino acid sequencing and a region of low codon redundancy in the 42-kDa amino-terminal sequence was used to design a single 49-mer oligonucleotide sequence incorporating choices based on human codon usage (10). The 49-mer detected 14 clones when used to screen an X chromosome-enriched genomic library. Two overlapping clones were analyzed in more detail and found to contain the same 1.6-kb 49-mer positive *Hin*dIII fragment. This fragment was shown to give a positive signal when used to probe DNA from a human–mouse cell hybrid that contained the tip of the long arm of the X chromosome (Xq26–ter) (unpublished data). This result increased our confidence that the clones represented the IDS gene, since the IDS gene has been localized to this small portion of the human X chromosome (1).

The 1.6-kb *Hin*dIII genomic DNA fragment was then used to screen a human colon cDNA library. Eighteen clones were detected and their inserts were sized. The clone with the longest insert (λc2S15) was fully sequenced and found to contain an initiating methionine and a continuous open reading frame that included a sequence that was colinear with the 42-kDa and the 14-kDa amino-terminal amino acid sequences. However, the reading frame did not extend to include a stop codon or any 3′ untranslated region. A 300-bp *Hin*dIII–*Eco*RI restriction fragment from the 3′ end of the λc2S15 was then used to screen a cDNA library constructed from human endothelial cells. Twenty-seven clones were isolated; 5 of which were also positive to the amino-terminal-specific 49-mer. Of the five, the clone that contained the longest insert (2.3 kb; λc2S23) was sequenced in combination with λc2S15.

Fig. 1 shows the nucleotide sequence of the 2297-bp insert from λc2S23, which encodes the entire amino acid sequence of IDS. Except for a few differences, the deduced amino acid sequence was colinear with the determined amino-terminal amino acid sequence of the 42-kDa and 14-kDa polypeptides. We believe that the amino acid discrepancies (residues 35, 53, 55, and 57) between the direct and predicted amino acid sequence data reflect amino acid sequencing errors resulting from the low signal obtained toward the end of the amino acid sequencing run. The detection of gene deletions and rearrangements in DNA from a group of severely affected MPS-II patients when hybridized with λc2S15 established that these cDNA clones encoded IDS (Fig. 2*A*). Of the 23 MPS-II patients analyzed, 7 had structural alterations including deletions of the entire λc2S15 coding region. These 7 patients also revealed similar Southern patterns indicative of structural alterations of the IDS gene when their DNA was digested with *Hin*dIII, *Stu* I, and *Taq* I and probed with λc2S15 (unpublished data). Sixteen patients had identical patterns to normal controls, suggesting the presence of small deletions or point mutations responsible for the MPS-II biochemical and clinical phenotype. The two patients, in which the entire IDS gene had been removed (Fig. 2*A*) had the most severe clinical phenotype of the large group of MPS-II patients studied, raising the possibility that these patients may also have codeletions of contiguous genes to IDS (unpublished data).

The sequence of λc2S23 shown in Fig. 1 contains an open reading frame from the initiation codon at position 125 to the termination codon (TGA) at position 1775. This 1650-bp sequence encodes a polypeptide of 550 amino acids as shown.

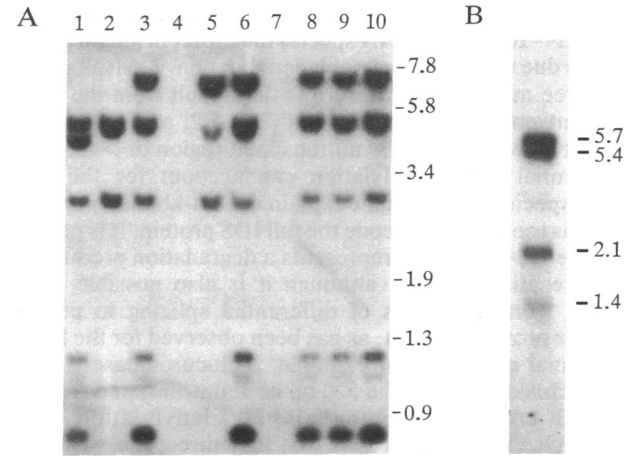

FIG. 2. (*A*) Southern blot analysis of MPS-II DNA for deletions and rearrangement of the IDS gene. λc2S15 was used to probe a Southern blot of *Pst* I-digested DNA samples from a normal male and female (lanes 9 and 10, respectively) and from severely affected MPS-II patients (lanes 1–8). The sizes (kb) of DNA molecular mass standards are shown in the right margin. (*B*) Northern blot of RNA from human placenta. The size (kb) of each RNA species is shown in the right margin.

The sequence flanking the ATG codon at bp 125 is in agreement with the consensus sequence for initiator codons (11). The first 25 amino acids at the amino terminus of the deduced protein have features characteristic of a signal sequence (12). Two putative sites for cleavage between the signal sequence and mature protein are indicated by arrows (Fig. 1). It would appear that eight amino acids are removed from the IDS precursor immediately after the most favored signal peptidase cleavage site (12) between residues 25 and 26. The 14-kDa polypeptide amino-terminal amino acid sequence was identified at amino acid residue 456, giving a total of 95 amino acids to the carboxyl terminus. The full sequence contains eight possible N-glycosylation sites (Asn-Xaa-Ser/Thr, Fig. 1). The molecular weight of the deduced polypeptide for the 14-kDa component was calculated as 11,093, suggesting that the two potential N-glycosylation sites may be glycosylated with a complex type oligosaccharide of ≈2 kDa to give a total calculated molecular weight of 15,093. The 14-kDa polypeptide does not contain cysteine residues, which is compatible with the finding that the 42-kDa and the 14-kDa polypeptide are not linked by disulfide bonds (3). The number of potential N-glycosylation sites used in the 42-kDa polypeptide is not known. The first N-glycosylation site (residue 31) is not contained within IDS form A since this asparagine residue is removed during amino-terminal processing. The molecular weight of the deduced peptide for the 42-kDa component was calculated as 47,404, suggesting that the value determined by SDS/polyacrylamide gel electrophoresis (3) may be in error or that additional amino acids are lost during internal proteolytic cleavage of the IDS precursor. These results suggest that posttranslational proteolytic processing of IDS is restricted to cleavage of a signal peptide, removal of the amino-terminal 8 amino acids, and internal cleavage to produce the observed 42- and 14-kDa polypeptides in human liver, kidney, lung, and placenta (3). This is a commonly observed polypeptide maturation process for lysosomal enzymes that are generally synthesized as larger precursors and then converted to their mature forms by a limited number of proteolytic steps shortly before or after their transfer into lysosomes (13).

Northern blot analysis of placental poly(A)$^+$ RNA with λc2S15 revealed three major RNA species (5.7, 5.4, and 2.1 kb) and one minor species (1.4 kb) (Fig. 2*B*). It is likely that

IDS, like other lysosomal enzymes [e.g., arylsulfatase A, B, and C (14–16)], has mRNA species that differ in length at their 3′ ends due to differential polyadenylylation. Arylsulfatase C has three major RNA transcripts that result from the use of different polyadenylylation sites (2.7, 5.2, and 7.0 kb), the longest of which has a 3′ untranslated region of >4 kb (16). Differential polyadenylylation can account for the three major species but it cannot explain the 1.4-kb minor species, which is too small to encode the full IDS protein. It is possible that the 1.4-kb species represents a degradation product or a cross-reacting species, although it is also possible that it results from a process of differential splicing to produce another protein product, as has been observed for the human lysosomal enzymes, for example, β-glucuronidase (17) and β-galactosidase (18). The 520 bp of 3′ untranslated region in λc2S23 contains a potential polyadenylylation signal (AATAAA) at position 2041 that may direct the position of polyadenylylation for the observed 2.1-kb mRNA species. If this is the case, the 124 bp of 5′ untranslated region in λc2S23 is sufficient to account for most, if not all, of the 5′ untranslated region expected for the 2.1-kb mRNA species [allowing for 50–100 residues of poly(A) tail].

Fig. 3 shows an alignment of IDS amino acid sequence with sequence of other human-derived sulfatases and a sea urchin arylsulfatase. This analysis reveals many areas of identical and conserved amino acid matches within the arylsulfatase group (galactose 3-sulfatase, *N*-acetylgalactosamine 4-sulfatase, and steroid sulfatase) and the two nonarylsulfatase sequences (unpublished data), IDS and glucosamine 6-sulfatase. Sea urchin arylsulfatase is also aligned and has sequence homology with the other five human sulfatases. A multiple sequence alignment of the amino acid sequence of these six sulfatases has the highest level of homology in the amino-terminal third of each sulfatase (Fig. 3). The human arylsulfatase group has conserved blocks of up to six identical amino acid residues, for example, Cys-Thr-Pro-Ser-Arg and Gly-Lys-Trp-His-Leu-Gly (Fig. 3). On the other hand, only part of these sequences are conserved in the two nonarylsulfatases, IDS and glucosamine 6-sulfatase. We propose that these sequences may represent regions of the arylsulfatases that enable the relatively nonspecific hydrolysis of arylsulfates. All five human sulfatases have significant sequence homology with the amino acid sequence of sea urchin arylsulfatase (Fig. 3). By taking account of conservative amino acid substitutions (23), there are even larger areas of homology within these six sulfatases. This high level of sequence conservation further supports the suggestion that these five human sulfatases are evolutionarily related to a common ancestral gene (14, 15, 19).

There are several regions in Fig. 3 where peptide inserts appear to be unique to a particular sulfatase. For instance, the microsomal membrane-bound steroid sulfatase contains two membrane-spanning regions (Fig. 3) (21). IDS also contains an amino acid sequence insert in the same region as the second membrane-spanning region of steroid sulfatase (Fig. 3). A second peptide insert in IDS is present just before the

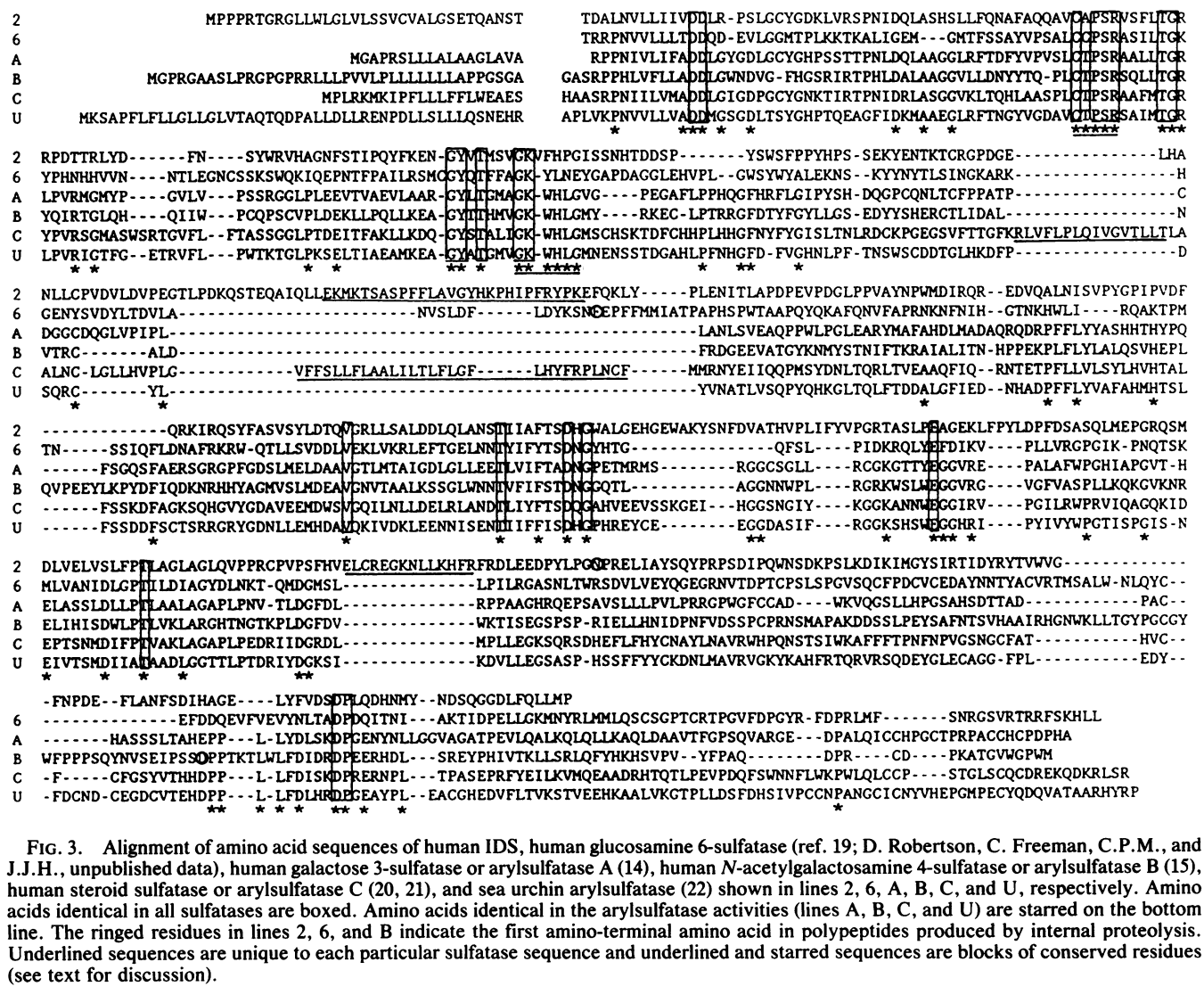

FIG. 3. Alignment of amino acid sequences of human IDS, human glucosamine 6-sulfatase (ref. 19; D. Robertson, C. Freeman, C.P.M., and J.J.H., unpublished data), human galactose 3-sulfatase or arylsulfatase A (14), human *N*-acetylgalactosamine 4-sulfatase or arylsulfatase B (15), human steroid sulfatase or arylsulfatase C (20, 21), and sea urchin arylsulfatase (22) shown in lines 2, 6, A, B, C, and U, respectively. Amino acids identical in all sulfatases are boxed. Amino acids identical in the arylsulfatase activities (lines A, B, C, and U) are starred on the bottom line. The ringed residues in lines 2, 6, and B indicate the first amino-terminal amino acid in polypeptides produced by internal proteolysis. Underlined sequences are unique to each particular sulfatase sequence and underlined and starred sequences are blocks of conserved residues (see text for discussion).

amino-terminal sequence of the 14-kDa polypeptide. The role that these two peptide inserts may have in IDS function is unknown. Interestingly, the sites (ringed in Fig. 3) for internal proteolysis of both glucosamine 6-sulfatase (19) and *N*-acetylgalactosamine 4-sulfatase (unpublished data) also occur near the sequence inserts. Although this may be coincidental, it is tempting to speculate that internal proteolysis may occur at or near these sites for IDS, glucosamine 6-sulfatase, and *N*-acetylgalactosamine 4-sulfatase.

Peters *et al.* (15) have speculated that the histidine residue in the highly conserved Gly-Lys-Trp-His-Leu-Gly region (Fig. 3) of the arylsulfatases, the conserved arginine residues in and nearby the Leu-Cys-Thr-Pro-Ser-Arg region (Fig. 3), and the Thr-Gly-Arg-Xaa-Xaa-Val-Arg region may contain the catalytic residues required for sulfate ester hydrolysis. This hypothesis was supported by evidence from studies reporting that one histidine residue (24) and two or more arginine residues (25) are essential for the catalytic activity of galactose 3-sulfatase (arylsulfatase A). However, the histidine and all but one arginine residue are eliminated from these regions when IDS and glucosamine 6-sulfatase are included in the family of sulfatases.

The isolation of cDNA encoding IDS will initiate in-depth studies of mutations causing the Hunter clinical phenotype. The observation that IDS amino acid sequence reduces the number of residues that are conserved in the sulfatase family will aid studies designed to identify catalytic site residues and perhaps also identify those residues involved in determining substrate specificity.

We thank Dan Robertson for providing the amino acid sequence for glucosamine 6-sulfatase before its publication and Dr. D. Callen for donation of human–mouse cell hybrid DNA. We are grateful to Dr. E. Wraith who provided DNA from some of the MPS-II patients used in this study. This work was supported by program grants from the National Health and Medical Research Council of Australia and the Adelaide Children's Hospital Research Foundation.

1. Neufeld, E. F. & Muenzer, J. (1989) in *The Metabolic Basis of Inherited Disease*, eds. Scriver, C. R., Beaudet, A. L., Sly, W. S. & Valle, D. (McGraw-Hill, New York), pp. 1565–1587.
2. Shapiro, L. J. (1989) in *The Metabolic Basis of Inherited Disease*, eds. Scriver, C. R., Beaudet, A. L., Sly, W. S. & Valle, D. (McGraw-Hill, New York), pp. 1945–1964.
3. Bielicki, J., Freeman, C., Clements, P. R. & Hopwood, J. J. (1990) *Biochem. J.*, in press.
4. Matsudaira, P. (1987) *J. Biol. Chem.* **262,** 10035–10038.
5. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
6. Nelson, P. V., Carey, W. F., Morris, C. P. & Pollard, A. C. (1989) *Med. J. Aust.* **151,** 126–131.
7. Chomczynski, P. & Sacchi, N. (1987) *Anal. Biochem.* **162,** 156–159.
8. Reisner, A. H. & Bucholtz, C. (1985) *Nature (London)* **314,** 310.
9. Lipman, D. J., Altschul, S. F. & Kececioglu, J. D. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 4412–4415.
10. Masuyama, T., Gojobori, T., Aota, S. & Ikemura, T. (1986) *Nucleic Acids Res.* **14,** r151–r197.
11. Kozak, M. (1987) *Nucleic Acids Res.* **15,** 8125–8131.
12. von Heijne, G. (1986) *Nucleic Acids Res.* **14,** 4683–4690.
13. Hasilik, A. & von Figura, K. (1984) in *Lysosomes in Biology and Pathology*, eds. Dingle, J. T. & Sly, W. S. (Elsevier, Amsterdam), Vol. 7, pp. 3–16.
14. Stein, C., Gieselmann, V., Kreysing, J., Schmidt, B., Pohlmann, R., Waheed, A., Meyer, H. E., O'Brien, J. S. & von Figura, K. (1989) *J. Biol. Chem.* **264,** 1252–1259.
15. Peters, C., Schmidt, B., Rommerskirch, W., Rupp, K., Zühlsdorf, M., Vingron, M., Meyer, H. E., Pohlmann, R. & von Figura, K. (1990) *J. Biol. Chem.* **265,** 3374–3381.
16. Yen, P. H., Marsh, B., Allen, E., Tsai, S. P., Ellison, J., Connolly, L., Neiswanger, K. & Shapiro, L. J. (1988) *Cell* **55,** 1123–1135.
17. Oshima, A., Kyle, J. W., Miller, R. D., Hoffman, J. W., Powell, P., Grubb, J. H., Sly, W. S., Tropak, M., Guise, S. & Gravel, R. A. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 685–689.
18. Morreau, H., Galjart, N. J., Gillemans, N., Willemsen, R., van Horst, T. J. & d'Azzo, A. (1989) *J. Biol. Chem.* **264,** 20655–20663.
19. Robertson, D. A., Freeman, C., Nelson, P. V., Morris, C. P. & Hopwood, J. J. (1988) *Biochem. Biophys. Res. Commun.* **157,** 218–224.
20. Yen, P. H., Allen, E., Marsh, B., Mohandas, T., Wang, N., Taggart, R. T. & Shapiro, L. J. (1987) *Cell* **49,** 443–454.
21. Stein, C., Hille, A., Seidel, J., Rijnbout, S., Waheed, A., Schmidt, B., Geuze, H. & von Figura, K. (1989) *J. Biol. Chem.* **264,** 13865–13872.
22. Sasaki, H., Yamada, K., Akasaka, K., Kawasaki, H., Suzuki, K., Saito, A., Sato, M. & Shimada, H. (1988) *Eur. J. Biochem.* **177,** 9–13.
23. Dayhoff, M. O., Schwartz, R. M. & Orcatt, B. C. (1978) in *Atlas of Protein Sequence and Structure*, ed. Dayhoff, M. O. (Natl. Biomed. Res. Found., Washington, DC), Vol. 5, Suppl. 3, pp. 345–352.
24. Lee, G. D. & van Etten, R. L. (1975) *Arch. Biochem. Biophys.* **171,** 424–434.
25. James, G. T. (1979) *Arch. Biochem. Biophys.* **197,** 57–62.