Office de la Propriété Intellectuelle du Canada
Un organisme d'Industrie Canada

Canadian Intellectual Property Office
An agency of Industry Canada

CA 2099503 C 2007/10/09

(11)(21) 2 099 503

(12) BREVET CANADIEN
CANADIAN PATENT

(13) C

(86) Date de dépôt PCT/PCT Filing Date: 1992/11/12
(87) Date publication PCT/PCT Publication Date: 1993/05/27
(45) Date de délivrance/Issue Date: 2007/10/09
(85) Entrée phase nationale/National Entry: 1993/06/30
(86) N° demande PCT/PCT Application No.: AU 1992/000611
(87) N° publication PCT/PCT Publication No.: 1993/010244
(30) Priorité/Priority: 1991/11/14 (AUPK9490/91)

(51) Cl.Int./Int.Cl. *C12N 15/56* (2006.01), *A61K 38/47* (2006.01), *A61K 48/00* (2006.01), *C12N 15/62* (2006.01), *C12N 5/10* (2006.01), *C12N 9/26* (2006.01), *C12P 21/02* (2006.01)

(72) Inventeurs/Inventors:
SCOTT, HAMISH STEELE, AU;
ORSBORN, ANNETTE MARIE, AU;
NELSON, PAUL VICTOR, AU;
CLEMENTS, PETER ROY, AU;
MORRIS, CHARLES PHILLIP, AU;
HOPWOOD, JOHN JOSEPH, AU;
ANSON, DONALD STEWART, AU

(73) Propriétaire/Owner:

(54) Titre : ALPHA-L-IDURONIDASE SYNTHETIQUE ET SEQUENCES DE GENES LA CODANT
(54) Title: SYNTHETIC ALPHA-L-IDURONIDASE AND GENETIC SEQUENCES ENCODING SAME

(57) Abrégé/Abstract:
The present invention relates generally to α-L-iduronidase and to genetic sequences encoding same. More particularly, the present invention provides an isolated nucleic acid molecule comprising a sequence of nucleotides which encodes or are complementary to a sequence which encodes a mammalian α-L-iduronidase or fragment or derivative thereof and to the recombinant enzyme encoded thereby. These molecules are useful in the investigation, diagnosis and treatment of subjects suspected of or suffering from α-L-iduronidase deficiency.

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 : C12N 15/56, A61K 37/54 — A1

(11) International Publication Number: WO 93/10244

(43) International Publication Date: 27 May 1993 (27.05.93)

2099503

(21) International Application Number: PCT/AU92/00611

(22) International Filing Date: 12 November 1992 (12.11.92)

(30) Priority data:
PK 9490 14 November 1991 (14.11.91) AU

(71) Applicant *(for all designated States except US)*: WOMEN'S AND CHILDREN'S HOSPITAL [AU/AU]; 72 King William Road, North Adelaide, S.A. 5006 (AU).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)* : SCOTT, Hamish, Steel [AU/AU]; 29 Wood Street, Millswood, S.A. 5034 (AU). ANSON, Donald, Stewart [GB/AU]; 12 Ross Street, Thebarton, S.A. 5031 (AU). ORSBORN, Annette, Marie [AU/AU]; 50 Quailo Avenue, Hallett Cove, S.A. 5158 (AU). NELSON, Paul, Victor [AU/AU]; 48 Salisbury Crescent, Colonel Light Gardens, S.A. 5041 (AU). CLEMENTS, Peter, Roy [AU/AU]; 3 Powell Court, West Lakes, S.A. 5021 (AU). MORRIS, Charles, Phillip [AU/AU]; 17 Thomas Street, Unley, S.A. 5061 (AU). HOPWOOD, John, Joseph [AU/AU]; 2 Monarto Court, Stonyfell, S.A. 5066 (AU).

(74) Agents: SLATTERY, John, M. et al.; Davies Collison Cave, 1 Little Collins Street, Melbourne, VIC 3000 (AU).

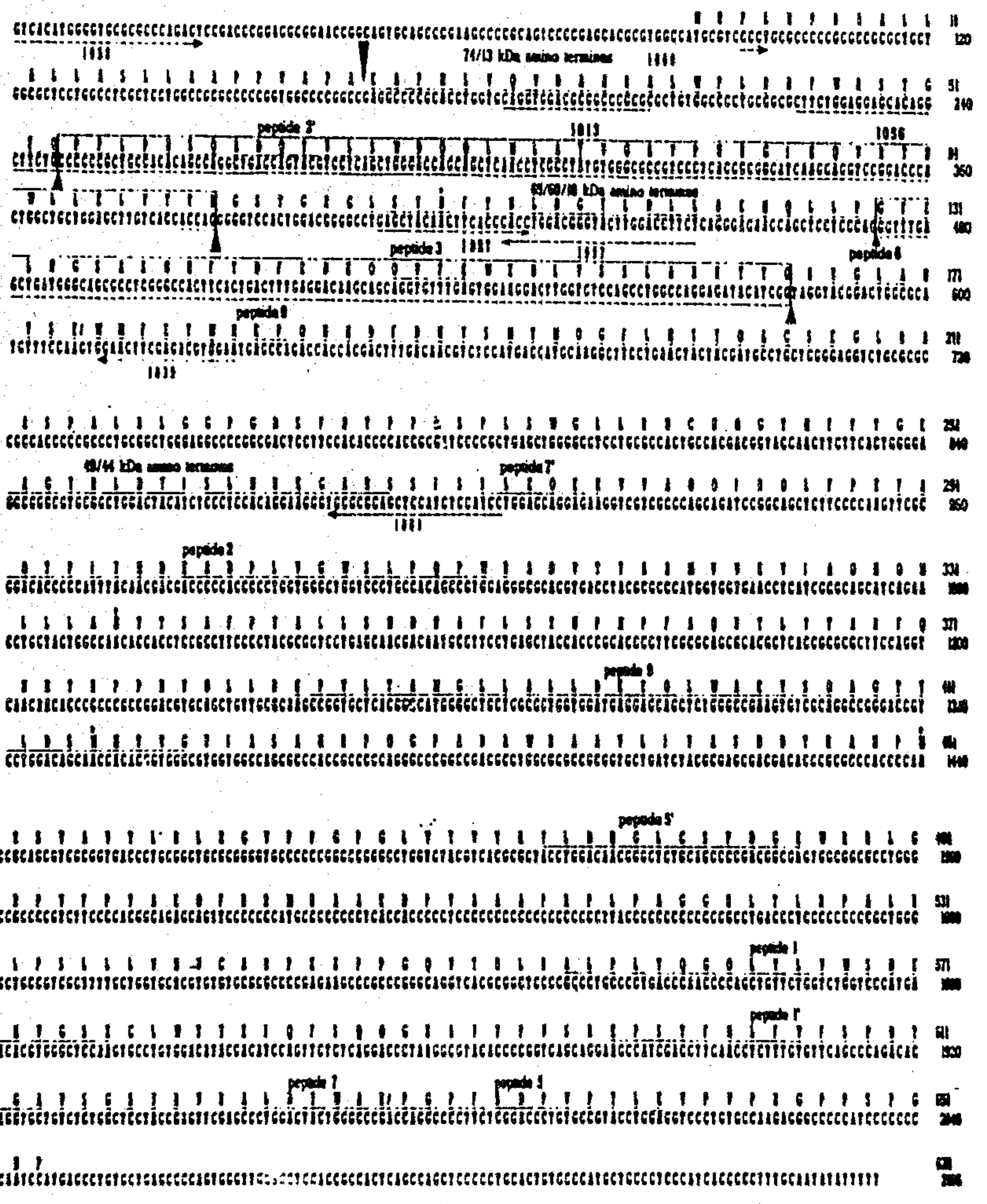
(81) Designated States: AU, CA, JP, KR, US, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, SE).

Published
With international search report.

(54) Title: SYNTHETIC ALPHA-L-IDURONIDASE AND GENETIC SEQUENCES ENCODING SAME

(57) Abstract

The present invention relates generally to α-L-iduronidase and to genetic sequences encoding same. More particularly, the present invention provides an isolated nucleic acid molecule comprising a sequence of nucleotides which encodes or are complementary to a sequence which encodes a mammalian α-L-iduronidase or fragment or derivative thereof and to the recombinant enzyme encoded thereby. These molecules are useful in the investigation, diagnosis and treatment of subjects suspected of or suffering from α-L-iduronidase deficiency.



SYNTHETIC α-L-IDURONIDASE AND GENETIC SEQUENCES ENCODING SAME

5

The present invention relates generally to α-L-iduronidase and to genetic sequences encoding same and to the use of these in the investigation, diagnosis and treatment of subjects suspected of or suffering from α-L-iduronidase deficiency.

10

The lysomal enzyme α-L-iduronidase (IDUA; glycosaminoglycan α-L-iduronohydrolase, EC 3.2.1.76) hydrolyzes the nonreducing terminal α-L-iduronide glycosidic bonds in the glycosaminoglycans heparan sulfate and dermatan sulfate (1,2). IDUA has served as a model for process and maturation events undergone
15 by lysosomal enzymes (3-8). A deficiency of IDUA in humans results in the lysosomal storage disorder mucopolysaccharidosis type I (MPS-I; cp-onyms, Hurler, Hurler/Scheic, and Scheic syndromes), which is inherited as an autosomal recessive disease and shows wide variation of clinical presentation. Severely affected patients have mental retardation, somatic tissue complications and a
20 reduced life span, while mildly affected patients may have only mild somatic complications and a normal life span. Multiple different mutant alleles at the IDUA locus are thought to be responsible for the spectrum of clinical phenotypes (1,9), but biochemical characterisation of the residual IDUA activity has enabled discrimination only between the extremes of clinical phenotypes (10-12). In work
25 leading up to the present invention, the isolation of the IDUA gene was undertaken to provide a DNA probe for molecular analysis of mutations in MPS-I patients and for use in enzyme and gene therapy experiments in the canine model (1,3) of MPS-I.

2099503

Accordingly, the present invention provides an isolated nucleic acid molecule comprising a sequence of nucleotides which encodes, or are complementary to a sequence which encodes, a mammalian α-L-iduronidase (IDUA) or fragment or
5 derivative thereof or its like molecule.

Preferably, the mammal is a human, livestock animal, companion animal, wild animal or laboratory test animal (e.g. rabbit, rat, mouse or guinea pig). Most preferably, the mammal is a human. Conveniently, the IDUA is isolatable from
10 the liver. However, the present invention extends to all mammalian IDUA enzymes and from any anatomical or cellular source and/or any biological fluid source, such as but not limited to plasma, serum, cell extract or lymph fluid.

Although a preferred embodiment of the present invention contemplates the use
15 of human IDUA or genomic or recombinant genetic sequences encoding same in the investigation, diagnosis and/or treatment of human subjects (i.e. homologous system), one skilled in the art will appreciate that the enzyme or genetic sequences encoding same from a non-human animal may also be useful. Such a heterologous system is encompassed by the present invention.
20

The "nucleic acid molecule" of the present invention may be RNA or DNA (eg. cDNA), single or double stranded and linear or covalently closed. The nucleic acid molecule may also be genomic DNA corresponding to the entire gene or a substantial portion thereof or to fragments and derivatives thereof. The
25 nucleotide sequence may correspond to the nautrally occurring nucleotide sequence or may contain single or multiple nucleotide substitutions, deletions and/or additions. All such modifications encode the IDUA-like molecules contemplated by the present invention. The length of the nucleotide sequence may vary from a few bases, such as in nucleic acid probes or primers, to a full
30 length sequence.

In one aspect of the invention, there is provided a purified α-L-iduronidase (IDUA) wherein the IDUA is secreted and purified from mammalian cells in culture transfected with a DNA sequence encoding human IDUA.

In another aspect of the invention, there is provided an isolated nucleic acid
5 molecule comprising a genomic nucleotide sequence wherein said genomic nucleotide sequence encodes a human α-L-iduronidase (IDUA) having an amino acid sequence substantially identical to an amino acid sequence of Figure 2 or a sequence complementary thereto.

The nucleic acid molecule of the present invention may constitute solely the nucleotide sequence encoding IDUA or like molecule or may be part of a larger nucleic acid molecule and extends to the genomic clone of IDUA. The non-IDUA encoding sequences in a larger nucleic acid molecule may include vector, promoter, terminator, enhancer, replication or signal sequences or non-coding regions of the genomic clone.

The present invention is particularly directed to the nucleic acid in cDNA form and particularly when inserted in an expression vector. The expression vector may be replicable in a eukaryotic or prokaryotic cell and may either produce mRNA or the mRNA may be subsequently translated into IDUA or like molecule. Particularly preferred eukaryotic cells include CHO cells but may be in any other suitable mammalian cells or cell lines or non-mammalian cells such as yeast or insect cells.

The present invention is further directed to synthetic IDUA or like molecule. The term "synthetic" includes recombinant forms and molecules produced by the sequential addition of amino acid residues, or groups of amino acid residues, in defined order. In a most preferred embodiment, the invention relates to recombinant IDUA or like molecule encoded by or expressed from the nucleic acid molecules as hereinbefore described.

The synthetic or recombinant IDUA may comprise an amino acid sequence corresponding to the naturally occurring amino acid sequence or may contain single or multiple amino acid substitutions, deletions and/or additions. The length of the amino acid sequence may range from a few residues to a full length molecule. Accordingly, this aspect of the present invention contemplates a proteinaceous molecule comprising an amino acid sequence corresponding to the full length mammalian IDUA enzyme or to a like molecule. The like molecule, therefore, comprises parts, derivatives and/or portions of the IDUA enzyme whether functional or not. Preferably, the mammal is human but may be of non-

human origin as contemplated above.

Advantageously, the recombinant IDUA is a biologically pure preparation meaning that it has undergone some purification away for other proteins and/or
5 non-proteinacous material. The purity of the preparation may be represented as at least 40% of the enzyme, preferably at least 60%, more preferably at least 75%, even more preferably at least 85% and still more preferably at least 95% relative to non-IDUA material as determined by weight, activity, amino acid homology or similarity, antibody reactivity or other convenient means.
10

Amino acid insertional derivatives of IDUA of the present invention include amino and/or carboxyl terminal fusions as well as intra-sequence insertions of single or multiple amino acids. Insertional amino acid sequence variants are those in which one or more amino acid residues are introduced into a
15 predetermined site in the protein although random insertion is also possible with suitable screening of the resulting product. Deletional variants are characterised by the removal of one or more amino acids from the sequence. Substitutional amino acid variants are those in which at least one residue in the sequence has been removed and a different residue inserted in its place. Typical substitutions
20 are those made in accordance with the following Table 1:

TABLE 1

Suitable residues for amino acid substitutions

5

Original Residue	Exemplary Substitutions
Ala	Ser
Arg	Lys
Asn	Gln; His
Asp	Glu
Cys	Ser
Gln	Asn
Glu	Asp
Gly	Pro
His	Asn; Gln
Ile	Leu; Val
Leu	Ile; Val
Lys	Arg; Gln; Glu
Met	Leu; Ile
Phe	Met; Leu; Tyr
Ser	Thr
Thr	Ser
Trp	Tyr
Tyr	Trp; Phe
Val	Ile; Leu

10 15 20 25

2099503

Where the enzyme is derivatised by amino acid substitution, the amino acids are generally replaced by other amino acids having like properties such as hydrophobicity, hydrophilicity, electronegativity, bulky side chains and the like.
5 Amino acid substitutions are typically of single residues. Amino acid insertions will usually be in the order of about 1-10 amino acid residues and deletions will range from about 1-20 residues. Preferably, deletions or insertions are made in adjacent pairs, i.e. a deletion of two residues or insertion of two residues.

10 The amino acid variants referred to above may readily be made using peptide synthetic techniques well known in the art, such as solid phase peptide synthesis (Merrifield synthesis) and the like, or by recombinant DNA manipulations. Techniques for making substitution mutations at predetermined sites in DNA having known or partially known sequence are well known and include, for
15 example, M13 mutagenesis. The manipulation of DNA sequence to produce variant proteins which manifest as substitutional, insertional or deletional variants are conveniently elsewhere described such as Sambrook *et al*, 1989 *Molecular Cloning: A Laboratory Manual* Cold Spring Harbor Laboratories, Cold Spring Harbor, NY.

20

The derivatives or like molecules include single or multiple substitutions, deletions and/or additions of any component(s) naturally or artificially associated with the IDUA enzyme such as carbohydrate, lipid and/or other proteinaceous moieties. For example, the present invention extends to glycosylated and non-
25 glycosylated forms of the molecule. All such molecules are encompassed by the expression "mutants", "derivatives", "fragments", "portions" and "like" molecules. These molecules may be active or non-active and may contain specific regions, such as a catalytic region. Particularly, preferred derivative molecules include those with altered glycosylation patterns relative to the naturally occurring
30 molecule. Even more particularly, the recombinant molecule is more highly glycosylated than the naturally occurring molecule. Such higly glycosylated derivatives may have improved take-up properties and enhanced half-lives.

The present invention also extends to synthetic IDUA or like molecules when fused to other proteinaceous molecules. The latter may include another enzyme, reporter molecule, purification site or an amino acid sequence which facilitates transport of the molecule out of a cell, such as a signal sequence.

5

In a most preferred embodiment, the present invention has an amino acid or corresponding IDUA cDNA nucleotide sequence substantially as setforth in Figure 2 or genomic nucleotide sequence substantially as set forth in Figure 4A and 4B or having at least 40% similarity, preferably at least 60% similarity
10 thereto or more preferably at least 80% or 85-90% similarity thereto.

The present invention further contemplates antibodies to synthetic IDUA or like molecule. The antibodies may be polyclonal or monoclonal, naturally occurring or synthetic (including recombinant, fragment or fusion forms). Such antibodies
15 will be useful in developing immunoassays for IDUA.

A further aspect of the present invention contemplates a method of screening for abberations in the IDUA gene. Such a method may be accomplished in a number of ways including isolating a source of DNA to be tested or mRNA
20 therefrom and hybridising thereto a nucleic acid molecule as hereinbefore described. Generally, the nucleic acid is probe or primer size and polymerase chain reaction is a convenient means by which to analyse the RNA or DNA. Other suitable assays include the ligation chain reaction and the strand displacement amplification methods. The IDUA sequence can also be
25 determined and compared to the naturally occurring sequence. Such methods may be useful in adults and children and may be adapted for a pre-natal test. The DNA to be tested includes a genomic sample carrying the IDUA gene, a cDNA clone and/or amplification product.

30 In accordance with this aspect of the present invention there is provided a method for screening for abberations in the IDUA gene including the absence of such a gene or a portion or a substantial portion thereof comprising isolating a

sample of DNA or mRNA corresponding to a region of said DNA and contacting same with an oligonucleotide probe capable of hybridising to one or more complementary sequences within the IDUA gene and then detecting the hybridisation, the extent of hybridisation or the absence of hybridisation.
5 Alternatively, the probe is a primer and capable of directing amplification of one or more regions of said IDUA gene and the amplification products and/or profile of amplification products is compared to an individual carrying the full gene or to a reference date base. Conveniently, the amplification products are sequenced to determine the presence or absence of the full gene.
10

The present invention further extends to a method of treating patients suffering from IDUA deficiency, such as in MPS-I, said method comprising administering to said patient an effective amount of IDUA or active like form thereof. Preferably, the IDUA is in recombinant form. Such a method is referred to as
15 "enzyme therapy". Alternatively, gene therapy can be employed including introducing an active gene (i.e. a nucleic acid molecule as hereinbefore described) or to parts of the gene or other sequences which facilitate expression of a naturally occurring IDUA gene.

20 Administration of the IDUA for enzyme therapy may be by oral, intravenous, suppository, intraperitoneal, intramuscular, intranasal, intradermal or subcutaneous administration or by infusion or implantation. The IDUA is preferably as hereinbefore described including active mutants or derivatives thereof and glycosylation variants thereof. Administration may also be by way of
25 gene therapy including expression of the gene by inclusion of the gene in viral vectors which are introduced into the animal (e.g. human) host to be treated. Alternatively, the gene may be expressed in a bacterial host which is then introduced and becomes part of the bacterial flora in the animal to be tested.

30 Still yet another aspect of the present invention is directed to a pharmaceutical composition comprising synthetic (e.g. recombinant) IDUA or like molecule, including active derivatives and fragments thereof, alone or in combination with

other active molecules. Such other molecules may act synergistically with the enzyme or facilitates its entry to a target cell. The composition will also contain one or more pharmaceutically acceptable carriers and/or diluents. The composition may alternatively comprise a genetic component useful in gene
5 therapy.

The active ingredients of the pharmaceutical composition comprising the synthetic or recombinant IDUA or mutants or fragments or derivatives thereof are contemplated to exhibit excellent activity in treating patients with a deficiency in
10 the enzyme when administered in an amount which depends on the particular case. The variation depends, for example, on the patient and the IDUA used. For example, from about 0.5 ug to about 20 mg of enzyme per animal body or, depending on the animal and other factors, per kilogram of body weight may be administered. Dosage regima may be adjusted to provide the optimum
15 therapeutic response. For example, several divided doses may be administered daily, weekly, monthly or in other suitable time intervals or the dose may be proportionally reduced as indicated by the exigencies of the situation. Accordingly, alternative dosages in the order of 1.0 μg to 15 mg, 2.0 μg to 10 mg or 10μg to 5mg may be administered in a single or as part of multiple doses. The
20 active compound may be administered in a convenient manner such as by the oral, intravenous (where water soluble), intramuscular, subcutaneous, intranasal, intradermal or suppository routes or implanting (eg using slow release molecules). Depending on the route of administration, the active ingredients which comprise a synthetic (e.g. recombinant) IDUA or fragments, derivatives or mutants thereof
25 may be required to be coated in a material to protect same from the action of enzymes, acids and other natural conditions which may inactivate said ingredients. For example, the low lipophilicity of IDUA will allow it to be destroyed in the gastrointestinal tract by enzymes capable of cleaving peptide bonds and in the stomach by acid hydrolysis. In order to administer the vaccine by other than
30 parenteral administration, the enzyme will be coated by, or administered with, a material to prevent its inactivation. For example, the enzyme may be administered in an adjuvant, co-administered with enzyme inhibitors or in

liposomes. Adjuvant is used in its broadest sense and includes any immune stimulating compound such as interferon. Adjuvants contemplated herein include resorcinols, non-ionic surfactants such as polyoxyethylene oleyl ether and n-hexadecyl polyethylene ether. Conveniently, the adjuvant is Freund's Complete or
5 Incomplete Adjuvant. Enzyme inhibitors include pancreatic trypsin inhibitor, diisopropylfluorophosphate (DEP) and trasylol. Liposomes include water-in-oil-in-water CGF emulsions as well as conventional liposomes.

The active compound may also be administered in dispersions prepared in
10 glycerol, liquid polyethylene glycols, and/or mixtures thereof and in oils. Under ordinary conditions of storage and use, these preparations contain a preservative to prevent the growth of microorganisms.

The pharmaceutical forms suitable for injectable use include sterile aqueous
15 solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersion. In all cases the form must be sterile and must be fluid to the extent that easy syringability exists. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of
20 microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyethylene glycol, and the like), suitable mixtures thereof, and vegetable oils. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the
25 required particle size in the case of dispersion and by the use of superfactants. The prevention of the action of microorganisms can be brought about by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, sorbic acid, thirmerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars or sodium chloride. Prolonged
30 absorption of the injectable compositions can be brought about by the use in the compositions of agents delaying absorption, for example, aluminum monostearate and gelatin.

Sterile injectable solutions are prepared by incorporating the active compound in the required amount in the appropriate solvent with various of the other ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the various sterilized active ingredient(s) into a sterile vehicle which contains the basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, the preferred methods of preparation are vacuum drying and the freeze-drying technique which yield a powder of the active ingredient plus any additional desired ingredient from previously sterile-filtered solution thereof.

When the IDUA of the present invention is suitably protected as described above, the composition may be orally administered, for example, with an inert diluent or with an assimilable edible carrier, or it may be enclosed in hard or soft shell gelatin capsule, or it may be compressed into tablets, or it may be incorporated directly with the food of the diet. For oral therapeutic administration, the active compound may be incorporated with excipients and used in the form of ingestible tablets, buccal tablets, troches, capsules, elixirs, suspensions, syrups, wafers, and the like. Such compositions and preparations should contain at least 1% by weight of active compound. The percentage of the compositions and preparations may, of course, be varied and may conveniently be between about 5 to about 80% of the weight of the unit. The amount of active compound in the vaccine compositions is such that a suitable dosage will be obtained. Preferred compositions or preparations according to the present invention are prepared,so that an oral dosage unit form contains between about 0.5 ug and 20 mg of active compound.

The tablets, troches, pills, capsules and the like may also contain the following: a binder such as gum gragacanth, acacia, corn starch or gelatin; excipients such as dicalcium phosphate; a disintegrating agent such as corn starch, potato starch, alginic acid and the like; a lubricant such as magnesium stearate; and a sweetening agent such a sucrose, lactose or saccharin may be added or a flavoring agent such as peppermint, oil of wintergreen, or cherry flavouring. When the

dosage unit form is a capsule, it may contain, in addition to materials of the above type, a liquid carrier. Various other materials may be present as coatings or to otherwise modify the physical form of the dosage unit. For instance, tablets, pills, or capsules may be coated with shellac, sugar or both. A syrup or elixir may
5 contain the active compound, sucrose as a sweetening agent, methyl and propylparabens as preservatives, a dye and flavoring such as cherry or orange flavor. Of course, any material used in preparing any dosage unit form should be pharmaceutically pure and substantially non-toxic in the amounts employed. In addition, the active compound may be incorporated into sustained-release
10 reparations and formulations.

As used herein "pharmaceutically acceptable carriers and/or diluents" include any and all solvents, dispersion media, aqueous solutions, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like. The
15 use of such media and agents for pharmaceutical active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active ingredient, use thereof in the pharmaceutical compositions is contemplated. Supplementary active ingredients can also be incorporated into the compositions.
20

The present invention further relates to the use of IDUA or active fragment, mutant or derivative thereof in the manufacture of a medicament for the treatment of patients suffering from a deficiency in the naturally occurring enzyme (e.g. MPS-1).
25

The present invention is further described with reference to the following non-limiting figures and examples.

With reference to the figures:

FIGURE 1 is a schematic representation showing a model to connect the seven
5 major polypeptides in immune purified human liver IDUA present after
SDS/PAGE as shown on the left, with the polypeptide sizes indicated in kDa (7).
The three amino-terminal sequences present are represented by the letters A, B
or C next to the polypeptides. The proteolytic sites cleaved to produce the seven
polypeptides from the 74-kDa polypeptide are numbered 1, 2 and 3.
10

FIGURE 2 is a representation of a compiled nucleotide sequence for IDUA
cDNA and the deduced amino acid sequence of the protein. The amino acid
sequence is shown in single letter code above the cDNA sequence. Nucleotide
and amino acid numbers are in the right margin. The probable site of signal
15 peptide peptidase cleavage is shown by a large arrow, and small arrows indicate
exon junctions. Exons II and IV, which are alternatively spliced in some RNA
transcripts, are boxed. Amino acids colinear with either amino-terminal peptide
data or tryptic peptides are underlined and named above the sequence. Potential
N-glycosylation sites are asterisked. Oligonucleotides used in this study are
20 underlined below the nucleotide sequence with the arrows indicating either sense
(→) or antisense (←). The cDNA clone λRPCI extended from base 541 to base
1269 and λE8A extended from base 391 to the 3' end of the sequence shown.

FIGURE 3 is a representation of reverse-transcribed normal fibroblast RNA
25 showing the alternative splicing of exons II and IV. Lane 1, PCR between 1D56
and 1D57, howing a major 225-bp product and a minor 84-bp product: lane 2,
pUC19 Hpa II markers: lane 3, PCR between IDNT and ID39, showing a major
222-bp product and a minor 114-bp product. Partial sequences of the two minor
products and their encoded amino acid are at the left and right of the figures.
30 The position of the missing exon is indicated by the arrow labelled "Exon
junction".

FIGURE 4 shows the sequence of the human genomic IDUA gene. Primers were made every 200 to 400 bp to completely sequence areas of interest in both directions. The coding region of the exons are in uppercase letters; untranslated sequence and introns are in lowercase letters. (A) Exons I and II of the human IDUA gene are shown in the 1.8 kb segment. The *Alu* repeat sequence and the four best potential OC boxes in the promoter region of IDUA are boxed. Potential transcription start sites are underlined. (B) Exons III to XIV of the human IDUA gene are shown in this 4.5 kb segment. Potential polyadenylation signals are underlined.

EXAMPLE

1. MATERIALS AND METHODS

Polypeptide Isolation and Sequencing.
All seven major polypeptides of IDUA (7) were directly sequenced from their amino termini as previously described (17). Tryptic peptide sequences from 150μg of purified human liver IDUA were generated as previously described (18).

Oligonucleotides and Primers.
All oligonucleotides were synthesised on an Applied Biosystems 391 DNA synthesiser. ID47, 5'-AACTTCGAGACCTGGAACGAGCCCGACCAGCACGACTTCGACAACGT-3', designed from residue 2 to residue 17 of peptide 8 (see Figure 2), was used for initial library screening. ID13, 5'-GCCCGGGCGGCA/GTCCACC/TTG-3' (a mixture of four sequences; nucleotides separated by / are options at the same position), designed from residue 13 to residue 7 of the 74/13-kDa amino-terminal amino acid sequence (see Figure 2), was used to screen Southern blots of the cosmid clone A157.1 (15). IDUA-specific primers used for PCR from cDNA were IDNT, ID39, ID56, ID57, ID58, ID60 and ID61 (see Figure 2).

Library Screening.

All libraries screened were of human origin and were purchased from Clontech. They were a leukocyte genomic DNA in EMGL3 (catalogue number HL1006)
5 and the following cDNA libraries: colon (random primed, HL1034a), unbilical endothelial (HL1024b), umbilical endothelial 5' stretch (HL1070b), and T-cell 5' stretch (HL1068b). All libraries were plated at a density of between 40,000 and 55,000 plaques per 140 mm plate. The host cells used for each library were NM538 for the EMBL3 genomic library, C600 for the λgt11 cDNA libraries.
10 Probes were either labelled at the 5' end (19) or labelled by primer extension of random oligonucleotide primers (20) and the Colony/Plaquescreen filters (DuPont/NEN) were prehybridised, hybridised, and washed according to the manufacturer's instructions.

15 **Sequencing.**

Specific oligonucleotides were made every 200-400 base pairs (bp) to fully sequence fragments in both directions (21). Compressed areas of G+C-rich sequence were resolved by using 7-deazaguanosine (22). Direct PCR sequencing was by the linear PCR method (23).

20

RNA Isolation and Northern Blot Analysis.

Total RNA was isolated from normal human placental, liver and kidney tissue or cultured normal human fibroblasts as previously described (24). Poly (A)+ RNA was obtained (25) from placental RNA and Northern blotting was carried out on
25 40 μg of total RNA and 10 and 40 μg of poly (a)+ RNA as described (17).

cDNA Synthesis.

Total RNA (3 μg) from normal fibroblasts was added to a reaction mix containing 1x Moloney murine leukaemia virus (Mo-MLV) reverse transcriptase
30 buffer (BRL), 40 units of RNAsin (Promega), 500 ng of random octamers, 0.5 mM deoxynucleotides (Boehringer Mannheim), and 200 units of Mo-MLV reverse transcriptase (BRL) to a final reaction volume of 50 μl. Incubation at 37°C for 1

m was followed by hydrolysis of the RNA by the addition of 5μl of 3 M NaOH and further incubation at 37° for 30 min. The NaOH was neutralised by the addition of 1.25 μl of 10.3 M HCl, and the cDNA was precipitated and resuspended in 50 μl of water. Each PCR used 5μl of cDNA.

5

PCR.

PCR reagents were as described by Saiki *et al.* (26) except that the final concentrations of deoxynucleotides were 400 μM and 10% v/v dimethyl sulfoxide was present in the reaction mix. Forty cycles of denaturation at 94 °C for 45 s,

10 annealing at 58 °C for 43 s, and elongation at 72 °C for 2 min were carried out. PCR products were analysed on 4% w/v Nusieve GTG agarose (FMC) gels.

Construction of Full-Length IDUA cDNA.

cDNA from a mixture of normal human fibroblast cell lines was used for PCR as

15 described, using the primers ID60 and ID6L. ID60 spans the initiating ATG codon and has a *Hin*dIII restriction site with a 4bp GC clamp on the 5' end. ID61 is ≈ 100 bp 3' of a unique *Kpn*I restriction iste (bases 818-823, see Figure 2). Utilizing the *Hin*dIII and the *Kpn*I sites, the PCR product was directionally cloned in a pTZ19 vector that contained the rest of the IDUA coding sequence from the

20 *Kpn*I site to the *Eco*RI cloning site of the clone λE8A. In all, 48 clones were analysed and only one was found to be correct (full length). This insert was excised with *Hin*dIII and *Eco*RI and was directionally cloned in the expression vector pRSVN.07 (which drives expression of the insert from the Rous sarcoma virus long terminal repeat) to give pPSVNID7I. This full length IDUA cDNA

25 insert was also subcloned in M13 and sequenced between the *Hin*dIII and *Kpn*I restriction sites, using IDUA-specific oligonucleotide primers to determine if any errors were present in the sequence.

Expression of IDUA.

CHO (Chinese hamster ovary) cells (strain DKI) were grown in Ham's F12 medium (GIBCO), 10% v/v fetal calf serum (GOBCO), penicillin at 100 μg/ml streptomycin sulfate at 100 μg/ml, and kanamycin sulfate at 120 μg/ml at 37 °C in a 5% v/v CO_2 atmosphere. CHO cells (1.2×10^7) were electroporated at 0 °C by using a BRL Cell-Porator at a pulse of 330 μF and 275 V in the presence 15 μg of pRSVNID21. Cells were grown in nonselective medium for 48 hr and then 1:20 and 1:100 dilutions of the electroporated cells were selected in G418 sulfate (Geneticin; GIBCO) at 750 μg/ml. A bulk culture of resistant cells was extracted (14) and assayed for IDUA activity with the fluorogenic substrate 4-methylumbelliferyl α-L-iduronide (Calbiochem) (6). The Bio-Rad protein assay was used to quantitate the amount of protein in each sample according to the manufacturer's instructions. The monoclonal antibody Id1A was used for immunocapture (14) and immunoquantification in conjunction with a polyclonal antibody (12) to assay the specific activity of the expressed IDUA (7).

2. RESULTS

All seven polypeptides of IDUA were subjected to direct amino-terminal sequencing, and three different amino-terminal sequences were found to be present. The 65-, 60-, and 18-kDa species have a common amino-terminal amino acid sequence, the 49- and 44-kDa another, and the 74- and 13-kDa species another. Assuming that all seven species represent part of a single IDUA polypeptide, a model (Figure 1), is proposed showing three sites of proteolytic processing of the 74-kDa polypeptide to produce the seven major species of IDUA.

After tryptic digestion and separation by HPLC (18) of immunopurified IDUA, nine major peptides were sequenced. One tryptic peptide was the same as the 65/60/18-kDa amino-terminal sequence, and one of the two tryptic peptide species present in part 3 were contained within the 49/44-kDa amino-terminal

sequence. Incorporating choices based on human codon usage and assuming that the undetermined amino acid at position 16 of peptide 8 was a glycosylated asparagine residue (see Figure 2) the sequence was used to design a 74-mer oligonucleotide (ID47) for library screening.

5

Using ID47 as a probe, 500,000 clones were screened of the EMBL3 human genomic library and obtained 8 clones. A genomic clone, ID-475, was purified and an ID47-positive 1.6 kilobase (kb) *PstI* fragment was subcloned in pUC19 to produce pID89 (14). This 1.6-kb insert was then used to screen a number of
10 cDNA libraries, this screening yielded only 1 clone, which contained an insert of 729 bp (λRPC1, bases 541-1269; see Figure 2) from the λgt10 random-promed human colon cDNA library. The sequence of this clone was colinear with six peptide sequences, including the 49/44-kDa amino-terminal sequence, but the clone ended within peptide 9.

15

The λRPCI insert was then used to screen a λgt11 human endothelial cDNA library. Twenty clones were isolated, and the insert of the longest clone, λE8A, was fully sequenced. The 11765-bp insert contained an open reading frame starting just before the position of the 65/60/18-kDa amino terminus (base 391 in
20 Figure 2) to a stop codon (base 2048). Six further tryptic peptides were matched to the translated DNA sequence but, significantly, the sequence of the 74/13-kDa amino terminus, a secondary tryptic peptide (peptide Z'), a signal peptide, and an initiating methronine were not present in this clone. Of the other clones, 7 ended at the same base at the 5' end, while all the others were shorter. A 5' probe
25 derived from λE8A was used to screen another seven cDNA libraries. No clones were obtained from the screening of five of these cDNA libraries. Screening of two 5' "stretch" cDNA libraries (umbilical endothelial and T cell) resulted in a further 38 clones. PCR analysis of these clones showed that all ended at the same 5' base as λE8A. Major secondary structures present in the IDUA mRNA
30 may be responsible for the premature termination of these clones at their 5' ends.

Using the polypeptide model for IDUA (Figure 1) it was hypothesised that the 74/13-kDa amino-terminal peptide sequence lay at the 5' end of the IDUA mRNA. A mixed oligonucleotide, ID13, made to the 74/13-kDa amino-terminal sequence was used to probe Southern blots of the cosmid A157.1, which spans the area of the IDUA gene (15). A 2.8 kb *Bam*HI fragment was isolated and partially sequenced. The sequence contained an initiating methionine, a signal peptide, 74/13 kDa amino terminus, and the start of the last unmatched tryptic peptide (peptide 2' in figure 2). A number of oligonucleotides were made to this exon and PCR used to amplify normal fibroblast cDNA. A major PCR product was obtained between ID58 and ID61, and the oligonucleotides ID56 and ID57, was directly sequenced (23). The collated DNA sequence (Figure 2) encodes a protein containing all amino-terminal and tryptic peptide sequences obtained from purified IDUA and is consistent with the model for IDUA (Figure 1).

PCR of normal fibroblast cDNA at the 5' end of the IDUA mRNA, using the oligonucleotides ID58 and ID61, produced a major product representing the sequence described (Figure 2) and several minor products that also hybridised to an internal oligonucleotide, ID56. This indicates that the minor products were representative of alternative mRNA species from the IDUA gene, as has been reported for a number of other genes, including lysosomal hydrolases (27-29).

PCR of normal fibroblast cDNA using the oligonucleotide pairs ID56 to ID57 and IDNT to ID39 produced two products per reaction. The smaller products were isolated and directly sequenced; they showed alternative splicing of exons II and IV of IDUA (Figure 3). The polypeptides from these alternatively spliced IDUA mRNA species would maintain the translation frame for the IDUA protein (see Figure 3) leaving the primary sequence of the translated peptide identical to that of the deduced IDUA peptide except for the omission of 47 and 36 amino acids, respectively. Thus, the alternatively spliced mRNA species individually missing exons II and IV would produce peptide products of 606 and 617 amino acids, respectively.

Using the insert of λESA as a probe against total placental RNA and poly(A)+ RNA, a single 2.3 kb band only was detected when 40 μg of poly(A)+ RNA was loaded in a single track. The strength of the signal also indicated that the mRNA for IDUA has a considerably lower abundance than the iduronate-2-sulfatase
5 mRNA in placental RNA (16). Multiple PCR products of the same relative intensity were observed when reverse-transcribed liver, kidney, or placental RNA was used as template, indicating that this splicing does not appear to be tissue specific and that these products may be minor mRNA species not detectable by Northern blot analysis. The alternative splicing of exon II introduces a
10 tryptophan residue into the amino acid sequence at the splice junction, and the alternative splicing of exons II and IV both interrupt reported peptide sequences (peptide 2' and the 65/60/18 kDa amino terminus of IDUA, respectively, see Figure 2). Thus, it was thought that the major PCR product was most likely to represent the full-length mRNA encoding IDUA. Expression of this putative full-
15 length mRNA would establish that the nucleotide sequence presented here in Figure 2 encodes enzymically active IDUA.

PCRs were performed with reverse-transcribed fibroblast RNA as template and the primers ID60 and ID61. The 840 bp PCR product was subcloned in the
20 pTZ19 vector to produce a "full-length" IDUA cDNA clone. Sequence analysis of this full-length insert found four nucleotides that were different from the previously determined sequence. The differences, numbered as in Figure 2, were A to C (base 276), G to A (base 402), T to C (base 440), and T to C (base 631). The first two differences alter the amino acid residues coded for by the cDNA
25 from Gln to Pro (amino acid 63) and Arg to Gln (amino acid 105), respectively. The T to C (base 440) is a silent change that alters a Leu (amino acid 118) codon from TTG to CTG and introduces a second *KpnI* site into the cDNA. Thus, the cloned PCR product presumably resulted from partial digestion with *KpnI* or the ligation of three fragments. The last change T to C (base 631) is a silent change
30 in the third base of an Asn (amino acid 181) codon. All of these differences may be polymorphic, but as two change amino acids, they may be transcription errors introduced by *Taq* DNA polymerase during PCR in the presence of high

concentrations of dNTPs (400 µM) for 40 cycles (30). However, these conditions were essential to produce enough PCR product to conduct the experiment.

This full-length cDNA construct was subcloned in the expression vector pRSVN.07 to produce the construct pRSVNID2L CHO cells were electroporated in the in the presence of pRSVNID21, and G418-resistant colonies were selected and grown as a mass culture. Cellular extracts from control CHO cells, mixed normal human skin fibroblasts, and pRSVNID21 transfected cells were assayed for total IDUA activity by using the IDUA-specific fluorogenic substrate. CHO cell extract contained a low level of IDUA activity. Cellular extract from CHO cells transfected with pRSVNID21 gave a total activity 160-fold greater than the control normal human fibroblast activity (Table 2). To compare the specific activities of the recombinant and fibroblast IDUA serial dilutions of the cellular extracts were assayed in parallel, using human IDUA-specific IDIA monoclonal antibody based immunocapture (14) and ELISA assays (12). The CHO cell extract gave sero background in both assays. The ELISA result was normalized to the normal fibroblast extract and showed a 12.7 fold higher expression of human IDUA in the pRSVNID21 transfected CHO cells. The immunocapture assay showed that this results in an almost proportional increase in IDUA activity in the transfected CHO cells, demonstrating that the normal and recombinant enzymes have similar specific activities (Table 2). These results prove that the IDUA sequence used in this experiment codes for a protein that has a specific activity similar to the IDUA activity present in normal cultured human skin fibroblasts.

2099503

TABLE 2

Expression of IDUA

Cell Type	IDUA activity[1] Total	IDUA activity[1] Captured[4]	Relative IDUA protein[2]	Relative IDUA specific activity[3]
CHO	1	ND	ND	-
CHO with pRSVNID21	160	152	12.7	12.0
Normal human fibroblasts	16	12.6	1	12.6

ND none detected

1 Activity is in pmol x 10^{-2} per min per mg of cell protein

2 The amount of human IDUA protein captured in the ELISA assay per mg of cell protein normalised against human fibroblasts.

3 Expressed as IDUA activity relative to IDUA protein.

4 IDUA activity captured in the immunocapture assay.

A further expression construct was made such that the normal 5' non-coding sequence of the IDUA mRNA, was found in the full length cDNA clone described, was replaced with 30 bp of the 5' non-coding sequence of the rat preproinsulin mRNA (5'-
5 AACCATCAGCAAGCAGGTCATTGTTCCAACGCGTGGCC-3'). At the same time, the four nucleotide differences noted in the PCR-produced 840 bp portion of the original cDNA used for expression (A→C, bp 276; G→A, bp 402; T→C bp 440; T→C bp 631) were corrected. This ensures efficient mRNA translation (34) and has been shown to lead to high-level expression of other
10 lysosomal enzymes in CHO cell expression systems (32,33). This modification also led to greatly enhanced expression of IDUA in CHO-K1 cells. The original expression plasmid was also modified such that the RSV-LTR promoter element was replaced with the human elongation factor 1α gene promoter from pEFBOS (35). This promoter is 5 times more efficient in CHO-K1 cells than the RSV-
15 LTR.

The total coding sequence, therefore, for IDUA has an open reading frame of 1959 bp encoding a peptide of 653 amino acids. A signal peptide of 26 amino acids with a consensus cleavage site (31) was present immediately adjacent to the
20 mature amino terminus of the protein (74/13 kDa amino terminus). Thus, the mature human IDUA protein of 627 amino acids has a molecular mass of 70,029 Da, which is consistent with the previous estimates of IDUA size after allowing for post-translational modifications (5-8). All major peptide species sequences are present in the translation of the open reading frame, totalling 234 amino acids
25 (42%) of the 627 amino acids of the mature IDUA. This includes several peptides that were present as minor sequences in peptide peaks (secondary peptides, e.g. peptide 7'). The presence of all three amino-terminal sequences from purified human liver IDUA in the peptide sequence presented in Figure 2 supports the hypothesised model of proteolytic processing of the 74 kDa IDUA
30 polypeptide (Figure 1). Of six potential stes in the 65/60/18 kDa amino-terminal sequence and peptide 8 was not detected in sequencing and may, therefore, be

glycosylated. The potential glycosylation site at the very end of peptide 9 was also not defected, but this may be due to a weak signal towards the end of the sequence rather than a glycosylated residue. No significant homology was found between the human IDUA amino acid sequence and proteins in the GenBank, National Biomedical Research Foundation, or Swiss-Prot data bases (all releases of May, 1991).

Having determined the cDNA sequence, the genomic sequence was then sought. The IDUA genomic sequence is valuable for defining mutations in MPS-1 patients, for defining diagnostically useful polymorphisms for MPS-1 and Huntington's disease and for refining the genetic and physical map of the IDUA gene. The genomic sequence is shown in Figure 4A and B as two segments.

The gene for IDUA is split into 14 exons spaning approximately 19 kb. The first 2 exons are separated by a 566 bp intron and the last 12 exons are separated by a 566 bp intron and the last 12 exons are clustered in a 4.2 kb region. Two variant polyadenylation signals consistent with a 2.3 kb mRNA transcript are underlined in Figure 4B. From the position of the proposed polyadenylation signals, the mRNA produced would be 2203 and 2285 bp with an additional 20-30 prior to the poly(A) tail.

Accordingly, the potential promoter for IDUA is bounded by an *Alu* repeat sequence and has only GC box type concensus sequences (Figure 4A).

The full length cDNA and genomic sequence described herein for human IDUA makes it possible to characterise MPS-I mutations and to determine how much of the clinical variability reflects different mutations and how much reflects other genetic or environmental influeneces. Furthermore, large-scale expression of IDUA will provide enzyme for evaluation of enzyme therapy, for example in the dog model for MPS-I and the cDNA in the appropriate vectors may be used for experimental gene therapy in the same model.

Those skilled in the art will appreciate that the invention described herein is susceptible to variations and modifications other than those specifically described. It is to be understood that the invention includes all such variations and modifications. The invention also includes all of the steps, features, compositions and compounds referred to or indicated in this specification, individually or collectively, and any and all combinations of any two or more of said steps or features.

REFERENCES:

1. Neufeld, EF & Muenzer, J, (1989) *The Metabolic Basis of Inherited Disease*, pp 1565-1587.
2. Hopwood, JJ, (1989) *Heparin: Chemical and Biological Properties, Clinical Applications*, pp 190-229.
3. Shapiro, LJ, Hall, CW, Leder IJ, & Neufeld, EF, (1976) *Arch. Biochem. Biophys. 172:* 156-161.
4. Rome, IH, Garvin, A, Neufeld, EF, (1978) *Arch. Biochem. Biophys. 189:* 344-353.
5. Myerowitz, R, & Neufeld EF, (1981) *J. Biol. Chem. 256:* 3044-3048.
6. Clements, PR, Brooks, DA, Saccone, GTP, & Hopwood, JJ, (1985) *Eur, J. Biochem. 152:* 21-28.
7. Clements, PR, Brooks, DA, McCourt, PAG, & Hopwood, JJ, (1989) *Biochem. J. 259:* 199-208.
8. Taylor, JA, Gibson, GJ, Brooks, DA & Hopwood, JJ, (1991) *Biochem. J. 274:* 263-268.
9. Hopwood, JJ, & Morris, CP, (1990) *Mol. Biol. Med. 7:* 381-404.
10. Hopwood, JJ, & Muller, V, (1979) *Clin. Sci. 57:* 265-272.
11. Muller, V, & Hopwood, JJ, (1984) *Clin. Genet. 26:* 414-421.
12. Ashton, LJ, Brooks, DA, McCourt, PAG, Clements, PR, & Hopwood, JJ, (1991) *Am. J. Humn. Genet.* in press
13. Spellacy, E, Shull, RM, Constantopoulos, G, & Neufled, EF (1983) *Proc. Natl. Acad. Sci. USA 80:* 6091-6095.
14. Scott, HS, Ashton, LJ, Eyre, HJ, Baker, E, Brooks, DA, Callen, DF, Sutherland, GR, Morris, CP & Hopwood, JJ, (1990) *Am. J. Hum. Genet. 47:* 802-807.
15. MacDonald, ME, Scott, HS, Whaley, WL, Phol, T, Wasmuth, JJ, Lehrach, H, Morris, CP, Frischuaf, AM, Hopwood, JJ, & Gusella, JF (1991) *Somatic Cell Mol. Genet. 17:* 421-425.

16. Stoltzfus, LY, Uhrhammer, N, Sosa-Pineda, B, Teplow, DB, & Neufeld, EF, (1990) *Am. J. Hum. Genet.* *47:* A147 (abstract 655).

17. Wilson, PJ, Morris, CP, A..son, DS, Occhiodoro, T, Bielicki, J, Clements, PR & Hopwood, JJ, (1990) *Proc. Natl. Acad. Sci. USA* *87:* 8531-8535.

18. Robertson, DA, Freeman, C, Nelson, PV, Morris, CP, & Hopwood, JJ, (1988) *Biochem. Biophys. Res. Commun.* *157:* 218-224.

19. Chaconas, G, & van de Sande, JH, (1980) *Methods Enzymol.* *65:* 75-88.

20. Feinberg, AP, & Vogelstein, B, (1983) *Anal. Biochem.* *132:* 6-13.

21. Sanger, F, Nicklen, S, & Coulson, AR, (1977) *Proc. Natl. Acad. Sci. USA* *74:* 5463-5467.

22. Mizusawa, S, Nishimura, S, & Seela, F, (1986) *Nucleic Acids Res.* *14:* 1319-1324.

23. Murray, V, (1989) *Nucleic Acids Res.* *17:* 8889.

24. Chomezynski, P, & Sacchi, N, (1987) *Anal. Biochem.* *162:* 156-159.

25. Kingston, RE, (1987) *Current Protocols in Molecular Biology* pp 4.5.1-4.5.3.

26. Saiki, RK, Gelfand, DH, Stoffel, S, Scharf, SJ, Higuchi, R, Horn, GT, Mullis, KB, & Erlich, HA, (1988) *Science* *239:* 487-491.

27. Osnima, A, Kyle, JW, Miller, RD, Hoffman, JW, Powell, P, Grubb, JH, Sly, WS, Tropak, M, Guise, S, & Gravel, RA (1987) *Proc. Natl. Acad. Sci. USA* *84:* 685-689.

28. Morreau, H, Galjart, NJ, Gillemans, N, Willemsen, R, van der Horts, GTJ, & d'Azzo, A, (1989) *J. Biol. Chem.* *264:* 20655-20663.

29. Quintern, LE, Schuchman, EH, Levran, O, Suchi, M, Ferlinz, K, Reinke, H, Sandhoff, K, & Desnick, RJ, (1989) *EMBO J.* *8:* 2469-2473.

30. Eckert, KA, & Kunkel, TA (1990) *Nucleic Acids Res.* *18:* 3739-3744.

31. von Heijne, G, (1986) *Nucleic Acids Res.* *14:* 4683-4690.

32. Anson, DS, *et al* (1992) *Biochem. J.* *284:* 789-794.

33. Bielicki, J, *et al* (1992) *Biochem. J.* (in press).

34. Cullen, BJ, (1988) *DNA* *7:* 645-650.

35. Mizishima, S, & Nagata, S, (1990) *BAR* *18:* 5322.

CLAIMS:

1. A purified α-L-iduronidase (IDUA) wherein the IDUA is secreted and purified from mammalian cells in culture transfected with a DNA sequence encoding human IDUA substantially identical to the amino acid sequence of Figure 2.

2. The purified IDUA according to claim 1 wherein the mammalian cells are Chinese Hamster Ovary (CHO) cells.

3. The purified IDUA according to claim 1 fused to another proteinaceous molecule.

4. The purified IDUA according to claim 3 wherein said other proteinaceous molecule is at least one of an enzyme, reporter molecule, purification moiety or signal sequence.

5. A pharmaceutical composition comprising purified α-L-iduronidase (IDUA) and one or more pharmaceutically acceptable carriers or diluents wherein said IDUA is secreted and purified from mammalian cells in culture transfected with a DNA sequence encoding human IDUA substantially identical to the amino acid sequence of Figure 2.

6. The pharmaceutical composition according to claim 5 wherein the mammalian cells are CHO cells.

7. The purified IDUA of any one of claims 1, 2, 3 or 4 wherein the purified IDUA comprises an enzymatically active fragment thereof.

8. The pharmaceutical composition of claim 5 or 6 wherein said purified IDUA comprises an enzymatically active fragment thereof.

9. The purified IDUA of any one of claims 1, 2, 3 or 4 having a purity of at least 95% relative to non-IDUA material as determined by at least one parameter selected from weight, activity, and antibody reactivity.

10. The pharmaceutical composition of claim 5 or 6 wherein said purified IDUA has a purity of at least 95% relative to non-IDUA material as determined by at least one parameter selected from weight, activity, and antibody reactivity.

11. An isolated nucleic acid molecule comprising a genomic nucleotide sequence wherein said genomic nucleotide sequence encodes a human α-L-iduronidase (IDUA) having an amino acid sequence substantially identical to an amino acid sequence of Figure 2, or a complementary nucleic acid molecule that is complementary to said nucleic acid molecule comprising said genomic nucleotide sequence.

12. The isolated nucleic acid molecule according to claim 11 wherein the nucleotides are deoxyribonucleotides.

13. The isolated nucleic acid molecule according to claim 12 wherein said molecule comprises a nucleotide sequence substantially identical to a nucleotide sequence of Figure 4a, Figure 4b, Figure 4c or Figure 4d.

14. A vector which comprises the nucleic acid molecule according to any one of claims 11, 12 or 13 wherein the vector is capable of replicating in an eukaryotic cell or a prokaryotic cell.

15. The vector according to claim 14 wherein the vector is an expression vector.

16. A host cell which comprises the vector of claim 14.

17. The host cell according to claim 16 wherein the host cell is a mammalian, yeast or insect cell.

18. The host cell of claim 17 wherein the mammalian cell is a Chinese Hamster Ovary (CHO) cell.

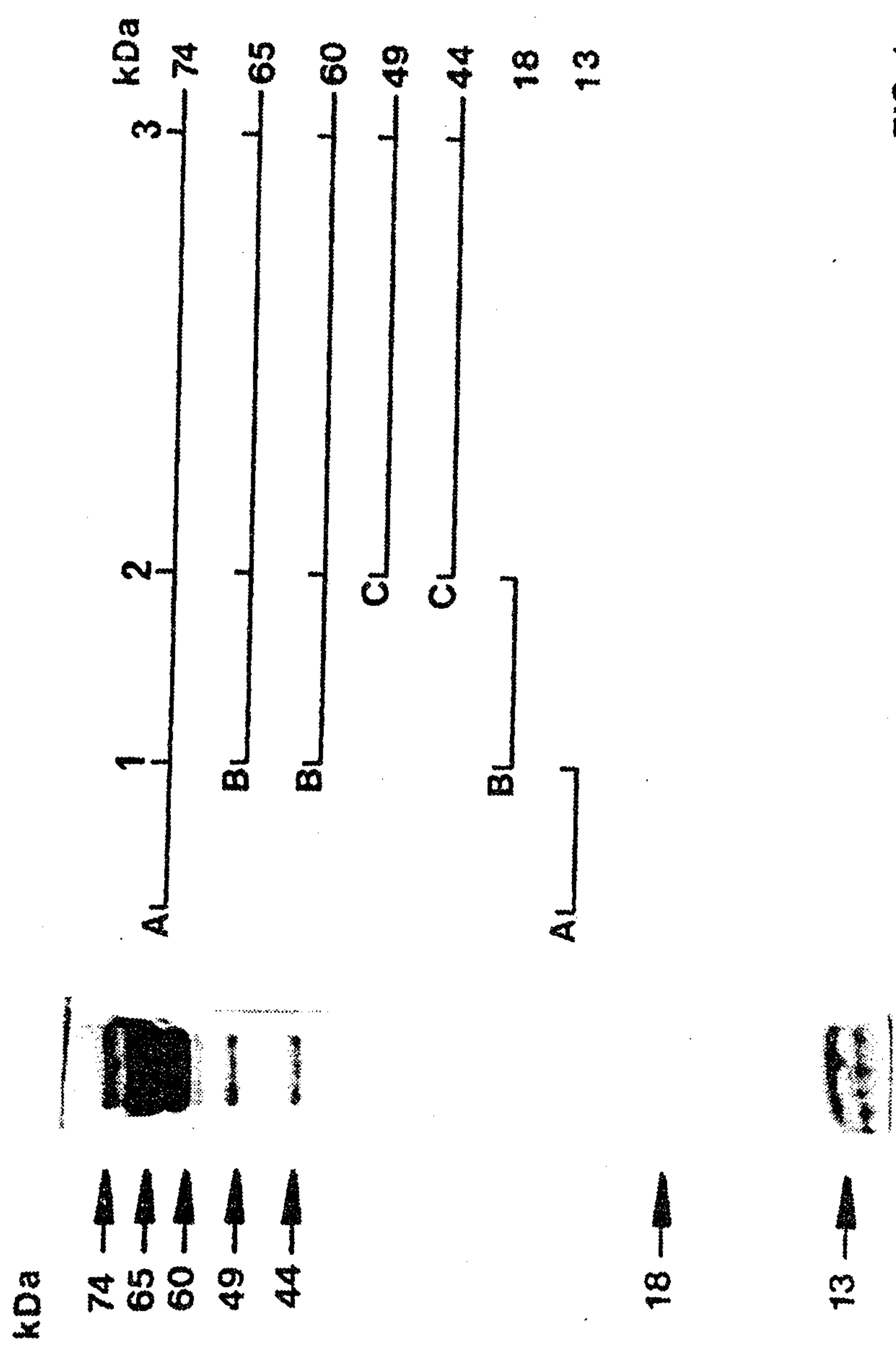


FIG 1

SUBSTITUTE SHEET

FIGURE 2A

FIGURE 2B

FIGURE 2C

```
                                                                                         M  R  P  L  R  P  R  A  A  L  L    11
GTCACATGGGGTGCGCGCCCAGACTCCGACCCGGAGGCGGAACCGGCAGTGCAGCCCGAAGCCCCGCAGTCCCCGAGCACGCGTGGCCATGCGTCCCCTGCGCCCCCGCGCCGCCTGCT   120
      1058                                              74/13 kDa amino terminus     1060

 A  L  L  A  S  L  L  A  A  P  P  V  A  P  A  E  A  P  H  L  V  Q  V  D  A  A  R  A  L  W  P  L  R  R  F  W  R  S  T  G   51
GCCGCTCCTGGCCTCGCTCCTGGCCGCGCCCCCGGTGGCCCCGGCCGAGGCCCCGCACCTGGTGCAGGTGGACGCGGCCCGCGCGCTGTGGCCCCTGCGGCGCTTCTGGAGGAGCACAGG   240

                      peptide 2'                                    1013                              1056
 F  Q  P  P  L  P  H  S  Q  A  D  Q  Y  V  L  S  W  D  Q  Q  L  M  L  A  Y  V  G  A  V  P  H  R  G  I  K  Q  V  R  T  H   91
CTTCTGCCCCCCGCTGCCACACAGCCAGGCTGACCAGTACGTCCTCAGCTGGGACCAGCAGCTCAACCTCGCCTATGTGGGCGCCGTCCCTCACCGCGGCATCAAGCAGGTCCGGACCCA   360

                                                                     65/60/18 kDa amino terminus
 W  L  L  E  L  V  T  T  R  G  S  T  G  R  G  L  S  Y  N  F  T  H  L  D  G  Y  L  D  L  L  R  E  H  Q  L  L  P  G  F  E   131
CTGGCTGCTGGAGCTTGTCACCACCAGGGGGTCCACTGGACGGGGCCTGAGCTACAACTTCACCCACCTGGACGGGTACTTGGACCTTCTCAGGGAGAACCAGCTCCTCCCAGGGTTTGA   480
                                                  peptide 3   IDHT           1057                           peptide 6

 L  H  G  S  A  S  G  H  F  T  D  F  E  D  K  Q  Q  V  F  E  W  K  D  L  V  S  S  L  A  R  R  Y  T  G  R  Y  G  L  A  H   171
GCTGATGGGCAGCGCCTCGGGCCACTTCACTGACTTTGAGGACAAGCAGCAGGTGTTTGAGTGGAAGGACTTGGTCTCCAGCCTGGCCAGGAGATACATCGGTAGGTACGGACTGGCGCA   600

                               peptide 8
 V  S  K/ W  M  F  E  T  W  H  E  P  Q  H  H  D  F  D  N  V  S  M  T  M  Q  G  F  L  H  Y  Y  Q  A  C  S  E  G  L  R  A   211
TGTTTCCAAGTGGAACTTCGAGACGTGGAATGAGCCAGACCACCACGACTTTGACAACGTCTCCATGACCATGCAAGGCTTCCTGAACTACTACGATGCCTGCTCGGGAGGTCTGCCGCGC   720
           1039
```

FIG 2a

```
  A  S  P  A  L  R  L  G  G  P  G  D  S  F  H  T  P  P  R  S  P  L  S  W  G  L  L  R  H  C  H  O  G  T  H  F  F  T  G  E    251
CGCCACCCCCGCCCTGCGGCTGGGAGGCCCCGGCGACTCCTTCCACACCCCACCGCGATCCCCGCTGAGCTGGGGCCTCCTGCGCCACTGCCACGACGGTACCAACTTCTTCACTGGGGA   840

        49/44 kDa amino terminus                                           peptide 7'
  A  G  Y  R  L  D  Y  I  S  L  H  R  K  G  A  R  S  S  I  S  I  L  E  O  E  K  V  V  A  O  O  I  R  O  L  F  P  K  F  A    291
GCCGGGCGTGCGGCTGGACTACATCTCCCTCCACAGGAAGGGTGCGCGCAGCTCCATCTCCATCCTGGAGCAGGAGAAGGTCGTCGCCCAGCAGATCCGGCAGCTCTTCCCCAAGTTCGC   960
                                          <---------------------
                                                   1061

                 peptide 2
  O  T  P  I  Y  H  D  E  A  D  P  L  V  G  W  S  L  P  Q  P  W  R  A  D  V  T  Y  A  A  M  V  V  K  V  I  A  O  H  O  M    331
GGACACCCCCATTTACAACGACGACCCCCACCCCCTGGTGGGCTGGTCCCTGCCACAGCCGTGGAGGGCGCACGTGACCTACGCGCCCATGGTGGTGAACCTCATCGCGCAGCATCAGAA  1080

             *
  L  L  L  A  H  T  T  S  A  F  P  Y  A  L  L  S  H  D  H  A  F  L  S  Y  H  P  H  P  F  A  Q  R  T  L  T  T  A  R  F  Q    371
CCTGCTACTGGCCAACACCACCTCCGCCTTCCCCTACGCGCTCCTGAGCAACGACAATGCCTTCCTGAGCTACCACCCGCACCCCTTCGCGCAGCGCACGCTCACCGCGCGCTTCCAGGT  1200

                                                                      peptide 9
  H  H  T  R  P  P  H  V  O  L  L  R  K  P  V  L  T  A  M  G  L  L  A  L  L  D  E  E  O  L  W  A  E  V  S  O  A  G  T  Y    411
CAACAACACCCCCCCGCCGGACGTGCAGCTGTTGCGCAAGCCGGTGCTCACGGCCATGGGGCTGCTGGCGCTGGTGGATGAGGAGCAGCTCTGGGCCGAAGTGTCGCAGGCCGGGACCGT  1320

          *                                                                                                          *
  L  D  S  M  H  T  V  G  V  I  A  S  A  H  R  P  O  G  P  A  D  A  W  R  A  A  V  L  I  Y  A  S  D  D  T  R  A  H  P  M    451
CCTGGACAGCAACCACACGGTGGGCGTGGTGGCCAGCGCCCACCGCCCCCAGGGCCCGGCCGACGCCTGGCGCGCCGCGGTGCTGATCTACGCGAGCGACGACACCCGCGCCCACCCCAA  1440
```

FIG 2b

SUBSTITUTE SHEET

5/11

peptide 5'
R S V A V T L R L R G V P P G P G L V Y V T R Y L D H G L C S P D G E W R R L G 491
CCGCAGCGTCGCGGTGACCCTGCGGCTGCGCGGGGTGCCCCCCGGCCCGGGCCTGGTCTACGTCACGCGCTACCTGGACAACGGGCTCTGCAGCCCCGACGGCGAGTGGCGGCGCCTGGG 1560

R P V F P T A E Q F R R M R A A E D P V A A A P R P L P A G G R L T L R P A L R 531
CCGCCCCGTCTTCCCCACGGCAGAGCAGTTCCCCCCCATGCCCCCCCCTCACCACCCCCTCCCCCCCCCCCCCCCCCCCCTTACCCCCCCCCCCCGCCTGACCCTCCCCCCCCCGCTGGG 1680

peptide 1
L P S L L L V H V C A R P E K P P G Q V T R L R A L P L T Q G Q L V L V W S D E 571
GCTGCCCGTCGCTTTTGCTGGTGCACGTGTGCGCGCCCCGAGAAGCCCGCCCGGGCAGGTCACGCGGCTCCCCGCCCTGCCCCTGACCCAACCCCAGCTGTTCTGGTCTGGTCCCATGA 1800

peptide 1'
H V G S K C L W T Y E I Q F S Q Q G K A Y T P V S R K P S T F H L F V F S P D T 611
ACACGTGGGCTCCAAGTGCCTGTGGACATACGACATCCAGTTCTCTCAGGACCCTAAGGCCGTACACCCCGGTCAGCAGGAAGCCATCGACCTTCAACCTCTTTGTGTTCAGCCCAGACAC 1920

peptide 7 peptide 5
G A V S G S Y R V R A L D Y W A R/ P G P F S D P V P Y L E V P V P R G P P S P G 651
AGGTGCTGTCTCTGGCTCCTACCGAGTTCGAGCCCTGGACTACTGGGCCCGACCAGGCCCCTTCTCGGACCCTGTGCCGTACCTGGAGGTCCCTGTGCCAAGAGGGCCCCCATCCCCCC 2040

H P 633
CAATCCATGAGCCTGTGCTGAGCCCCAGTGGGTTGCACCTCCACCGCCACTCAGCCAGCTCCCCCTGCACTGTGCCCATGCTGCCCTCCCATCACCCCCTTTGCAATATATTTTT 2155

FIG 2c

SUBSTITUTE SHEET

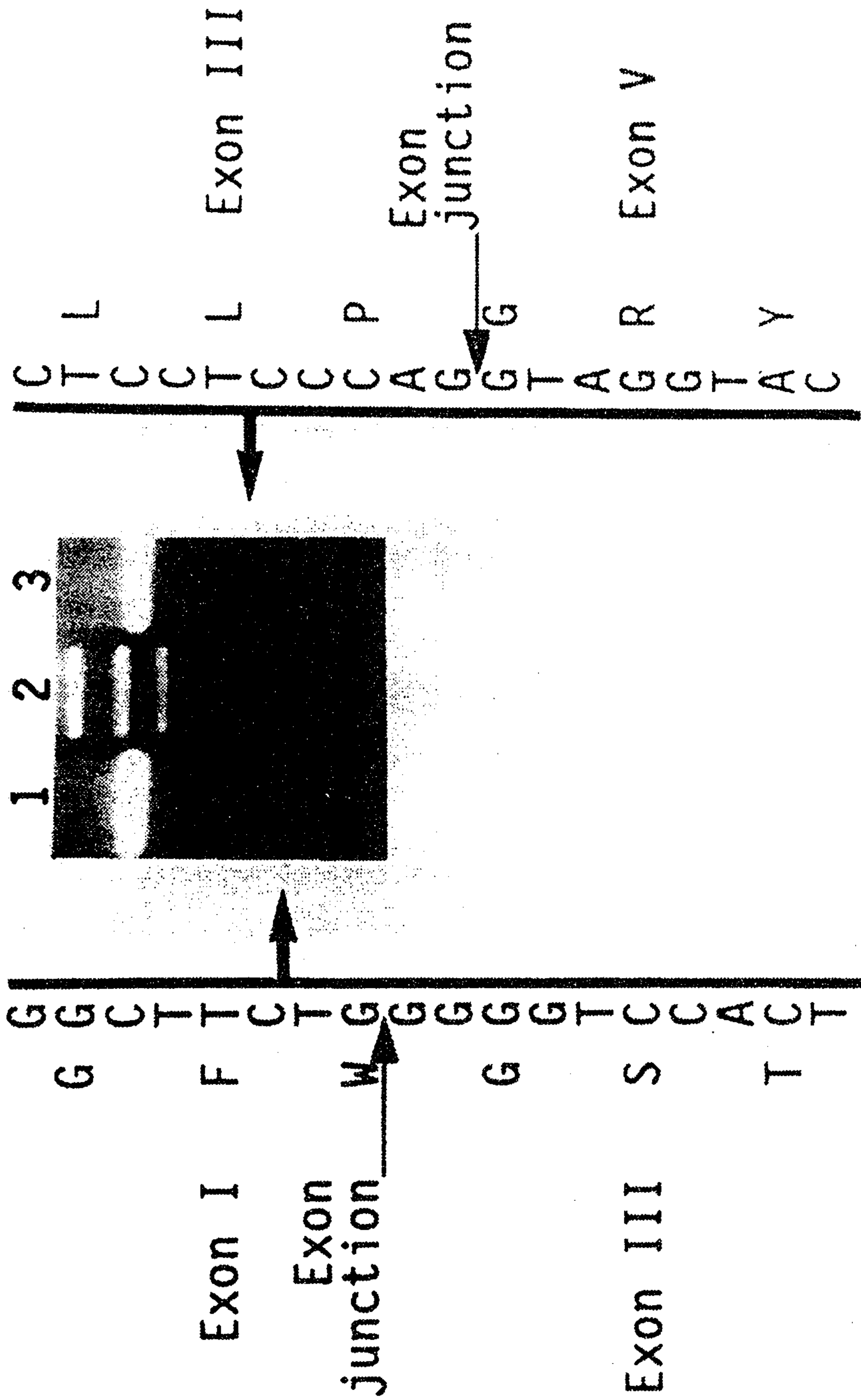


FIG 3

SUBSTITUTE SHEET

FIGURE 4A
FIGURE 4B
FIGURE 4C
FIGURE 4D

A

ttaaatatgtatattcgtattgctttggcttaaacaatatttcagtcctgccttggttgaatccatggatctgaagccaagtgtaaggagggcccagtgtgctcttctggggtaattctc 120

cttcctgctaaagcgcacgctttactcaggaggctggggtgagaaaatcgctgaagccccggagatggaggttgcagtgagctgagatcgcgccactgcacctcagcctgggcgacaaag 240

caagactctgtctcaaaaacacacaaaaacagagaaaaacaagacagtaatggctcaactcacatagcaccaacgggcgaagcgttcttctgagcgctttccgagtcatcggtcctcaga 360

gcagcccctgaggcccgcaaggaagcggggctccaagccctgccgtgctcccggctccccgaggctccccgaggccacccaacccctcccacccggccatcgccccctcaccaaggcccc 480

gccccggcgcggcggtcacatggggtgcgcgcccagactccgacccggaggcggaaccggcagtgcagcccgaagccccgcagtccccgagcacgcgtggccATGCGTCCCCTGCGCCCC 600

CGCGCCGCGCTGCTGGCGCTCCTGGCCTCGCTCCTGGCCGCGCCCCCGGTGGCCCCGGCCGAGGCCCCGCACCTGGTGCAGGTGGACGCGGCCCGCGCGCTGTGGCCCCTGCGGCGCTTC 720

TGGAGGAGCACAGGCTTCTGgtgagcgctccgcggcctccgggacccctggccgcacggggagagctcgggcgccccctgactgcgcactgtgagagcttcagagaccggagctccctc 840

ctctggggccctggctctcccgggcccgcccccgccgtgtttgtgggtgggtcctccacctgagtgggcgccggggcgtgagcctgggccgccccctgcagcccaggccgatgcccggg 960

atcctgctctttgaggtaaaccaggagtctcccctgggagtggacggccctgcagcgggacctggcctgcctgtcccattccttccacctagagctgaggtacccgccttcctggcaggg 1080

ccagggccagggctggcgttggcccctcgtcttactgctgctgccgttccccatgaagatgggacctccccacattcctggccctaagggtcattttattagtcactgaacgcacgggca 1200

gcgcctggatcctgcgcccgggcagtcctgggcttgaacgtgtgtgtcagccgcgctgccagccatgctgaggctcgggactgagccgcccctttgttgtccccagCCCCCCGCTGCCAC 1320

ACAGCCAGGCTGACCAGTACGTCCTCAGCTGGGACCAGCAGCTCAACCTCGCCTATGTGGCGCCGTCCCTCACCGCGGCATCAAGCAGGTCCGGACCCACTGGCTGCTGGAGCTTGTCA 1440

CCACCAGgtgggcggcgggcagggtctgggcgtcccagagccccttacagaggcacagatgggaggggagggctgggggctgctcggaagaccccttgttcccccacctcccgccgaagc 1560

accctgttggggagagcgtgtccttgctggctgtgctggggtgagggctgtgtgctggagggagcccctgcatggggcacggtgggcttcctgcaggtctccctgcaggctcagggttgg 1680

ctgcgccgcacctggctcctggtcacccgtgagcatccctgtgtgtgtctgctggccaggctgggtagggccactgca 1758

FIG 4a

B

tcctgtgcactcatgttgcctcttggggtgtgggaggggaaatggggcactcctgggcctccaggaggtgcagagaaccagggtgaggtgtccaccaggtcctgcctggctcctgacccc 120
tggcccctgctgctcgcgactggcctgcctcgtgccactgagcctcagagccattccgaacccccaccccaagttttccatctcttgatggtgtagggttgggggtctccatgtacaga 240
tactctagttcataccaggccttcatagggttattttccaaggggaagggcccctcgggaagccgggatcggagtcctgtgtggcaccttgcaggctcccacatgctccgttgtggccac 360
ggttccagcctggagcatggagctgtgtgggcaccctgcttcctgacgctgaccgtccttctgcagGGGGTCCACTGGACGGGGCCTGAGCTACAACTTCACCCACCTGGACGGGTACCT 480
GGACCTTCTCAGGGAGAACCAGCTCCTCCCAGgtgagctgtgggctctgccctcccagcccgcctgcacccccttgccctgcccaccctctccctcacccagcccctctgagtccttgga 600
tgtccattcagggctggccttggtgccggagcacaggcctggcagagcatgggtgtggtgtgtggtgggcggtggggcagccctcctgtgttccagGGTTTGAGCTGATGGGCAGCGCCT 720
CGGGCCACTTCACTGACTTTGAGGACAAGCAGCAGGTGTTTGAGTGGAAGGACTTGGTCTCCAGCCTGGCCAGGAGATACATCGgtgggcgagcgcaggccctggggccctggccggggc 840
gggggtactcctgggcaggttgcacccctatcacgcaggctgctgcctggtcaggagatacattggtgggcaggcgcaggcccttgtgggggatgggggtgacaagggataggttggtg 960
gtcggcgcaggccctggggccccaggctgggggtactcctgggcttggtgggtgggcgaaggccctgggcccctggggtgggggtactcctgggcaggctgcacccctatcacccagg 1080
ccgcacccctatcacccaggccgccgcccaggtcttggaccccttgagccagcgcttcctgatgtggggcgggaggctggcctgcatggagatggggttcatcttgagtcagacgccct 1200
tcatcaccttgcaccctccctccgtgggagtcactgaggcgagattcacctgtgctgggggacagcaaggctcctctgcagGTAGGTACGGACTGGCGCATGTTTCCAAGTGGAACTTC 1320
GAGACGTGGAATGAGCCAGACCACCACGACTTTGACAACGTCTCCATGACCATGCAAGgtgtgcaccgcttcctggggtcctgcccggctgaaagggggcagaggaaggcaggagcagag 1440
gctaagccgctcatccccagggcaggtgtagacgcagtgctcccccggcccagGCTTCCTGAACTACTACGATGCCTGCTCGGAGGGTCTGCGCGCCGCCAGCCCCGCCCTGCGGCTGGG 1560
AGGCCCCGGCGACTCCTTCCACACCCCACCGCGATCCCCGCTGAGCTGGGGCCTCCTGCGCCACTGCCACGACGGTACCAACTTCTTCACTGGGGAGGCGGGCGTGCGGCTGGACTACAT 1680
CTCCCTCCACAGGAAGgtgcgccctgcccctccgtccgccccggtgttctgcgccctcagccgctgtgccccgggccgcgctgaccctggtggtgctgaggcggccccgcccgcagGGTG 1800

FIG 4b

```
CGCGCAGCTCCATCTCCATCCTGGAGCAGGAGAAGGTCGTCGCGCAGCAGATCCGGCAGCTCTTCCCCAAGTTCGCGGACACCCCCATTTACAACGACGAGGCGGACCCGCTGGTCGGCT  1920
GGTCCCTGCCACAGCCGTGGAGGGCGGACGTGACCTACGCGGCCATGGTGGTGAAGgtgggccggcccaacgccctgcgcgccccccggccaccttcctcccgagacgggacaggcgagc  2040
ggtggccgcgccacccggtcccagctgccctggacacccgcagGTCATCGCGCAGCATCAGAACCTGCTACTGGCCAACACCACCTCCGCCTTCCCCTACGCGCTCCTGAGCAACGACAA  2160
TGCCTTCCTGAGCTACCACCCGCACCCCTTCGCGCAGCGCACGCTCACCGCGCGCTTCCAGGTCAACAACACCCGCCCGCCGCACGTGCAGCTGTTGCGCAAGCCGGTGCTCACGGCCAT  2280
GGGGCTGCTGGCGCTGCTGGgtgagccggggccgctggggtgggccggccagggccctccaggctggggagcggctcctgcgaaggccccgctgcggggagcgcacttcctccagccgcg  2400
cgcttcccggggtcggcctccgcgtggcggggcctggggactccttcaccaaggggagggggagcgagtggtgggaggcccggccctgggtcgggggcggctgggcaacgacccacgc  2520
ggcgacggcccccccccgccccgcagATGAGGAGCAGCTCTGGCCGAAGTGTCGCAGGCCGGACCGTCCTGGACAGCAACCACACGGTGGCGTCCTGGCCAGCGCCCACCGCCCCCA  2640
GGGCCCGGCCGACGCCTGGCGCGCCGCGGTGCTGATCTACGCGAGCGACGACACCCGCGCCCACCCCAACCGCAGCGTCGCGGTGACCCTGCGGCTGCGCGGGTGCCCCCCGGCCCGGg  2760
taagccggggttccagggaggtctctggccccgctggggctctggagggggcggcccggggagccgaggcctgagtgtcaggccccgcagGCCTGGTCTACGTCACGCGCTACCTGGACA  2880
ACGGGCTCTGCAGCCCCGACGGCGAGTGGCGGCGCCTGGGCCGGCCCGTCTTCCCCACGCAGAGCAGTTCCGGCGCATGCGCGCGGCTGAGgtaggtgggccgcggagggggcgaggggc  3000
cgggccgggccggggtcccgggggggtggggtccggggcgggggctccgaggcggtgtgggtgggaggtggagcggtgggccgggggcgttcgccctgaggtcgggccgagcgtccccag  3120
ctcccctggagaaccctgaggaccggccactgcgcccagGACCCGGTGGCCGCGGCGCCCCGCCCCTTACCCGCCGGCGGCCGCCTGACCCTGCGCCCCGCGCTGCGGCTGCCGTCGCTT  3240
TTGCTGGTGCACGTGTGTGCGCGCCCCGAGAAGCCGCCCGGGCAGgcaagtggcagtcccctaacccgcgccgcggcccggactccccttccccgacgccatcacagcccttccctcccc  3360
cagGTCACGCGGCTCCGCGCCCTGCCCCTGACCCAAGGGCAGCTGGTTCTGGTCTGGTCGGATGAACACGTGGGCTCCAAgtgcgtgagtggggccgcccctccctctgcctggtcctag  3480
gcaggtccctgggtcccgaccccttcacccatgcggtcactcgggccacttgccgtggcccatcggctccctccctcgccgccctgcgtccctgcccttcaccccacacactgtgggcca  3600
```

FIG 4c

```
cgcgccaggccctgccagtggggtgtgggttctcctaggggacatgagatggacattcgggctccagccctctcctgcctgggcaggaagagtgcccaggggctggggaggtgccgccga 3720
ggggcttgagggaatgaggctgtgggtccacgcggcccgtgccctgcctgctcccacctttgaggactgtcttgacccagccttgttcttggcctgacctccccagGTGCCTGTGGACAT 3840
ACGAGATCCAGTTCTCTCAGGACGGTAAGCCGTACACCCCGGTCAGCAGGAAGCCATCGACCTTCAACCTCTTTGTGTTCAGCCCAGgtgcgcccaccacccgctgccctggactcggcc 3960
accccattcttgggcctcagggcagtactgggtgggggcctcgagaagcctggggtcagggggctttcgggtgggggcaggttccggttggcacacatgtccccttgtctccagACACAG 4080
GTGCTGTCTCTGGCTCCTACCGAGTTCGGACCCTGGACTACTGGGCCCGACCAGGCCCCTTCTCGGACCCTGTGCCGTACCTGGAGGTCCCTGTGCCAAGAGGGCCCCCATCCCCGGG-A 4200
ATCCATGAgcctgtgctgagccccagtgggttgcacctccaccggcagtcagcgagctggggctgcactgtgcccatgctgccctcccatcacccccttgcaatatatttttatatttt 4320
attatttt-ttttatatcttggtaccaacgccccctttaaagcggctttgcacaggtcagtctcgggttgaggctctgtggcttggccctgggcacattccagggcagcctccaagggta 4440
aaccccggtggctgatgaggaccagctggagcgaggcct 4480
```

FIG 4d

SUBSTITUTE SHEET