THE MOLECULAR GENETICS OF MUCOPOLYSACCHARIDOSIS TYPE I

by

HAMISH STEELE SCOTT B.Sc.(Hons.)

A thesis submitted for the degree of

Doctor of Philosophy

in

The University of Adelaide

(Faculty of Medicine)

Lysosomal Diseases Research Unit,
Department of Chemical Pathology
Adelaide Children's Hospital,
South Australia.

and

Department of Pathology,
School of Medicine,
University of Adelaide,
South Australia.

October 6, 1992

DEDICATION

This thesis is dedicated to my parents. For always being there.

And to Teresa. Who will always be there.

TABLE OF CONTENTS

TABLE OF CONTENTS

CHAPTER 2

MATERIALS AND METHODS

TABLE OF CONTENTS

CHAPTER 2

MATERIALS AND METHODS

TABLE OF CONTENTS

CHAPTER 2

MATERIALS AND METHODS

TABLE OF CONTENTS

CHAPTER 3

ISOLATION OF THE α-L-IDURONIDASE GENE

TABLE OF CONTENTS

CHAPTER 3

ISOLATION OF THE α-L-IDURONIDASE GENE

TABLE OF CONTENTS

CHAPTER 4

MOLECULAR CHARACTERIZATION OF MPS-I PATIENTS

TABLE OF CONTENTS

CHAPTER 4

MOLECULAR CHARACTERIZATION OF MPS-I PATIENTS

Results and Discussion

TABLE OF CONTENTS

CHAPTER 5

LIST OF ABBREVIATIONS

ASO	allele specific oligonucleotide
BMT	bone marrow transplantation
bp	base pair
CC	chemical cleavage
CHO cells	Chinese hamster ovary cells
CNS	central nervous system
cpm	counts per minute
DMSO	dimethylsulphoxide
DNA	deoxyribonucleic acid
DS	dermatan sulphate
ER	endoplasmic reticulum
GAG	glycosaminoglycans
H	hydroxylamine
HD	Huntington disease or chorea
hr	hour
HS	heparan sulphate
IDUA	α-L-iduronidase
IDUA	the α-L-iduronidase gene or locus
kb	kilobase
kDa	kilo-Dalton
M6P	mannose-6-phosphate
min	minute
MPS	mucopolysaccharidoses
MPS-I	mucopolysaccharidosis type I
MPS-IH	Hurler syndrome
MPS-IH/S	Hurler/Scheie syndrome
MPS-IS	Scheie syndrome
mRNA	messenger RNA
N-terminal	amino-terminal
nt	nucleotide
LSD	lysosomal storage disorder
OT	osmium tetroxide
PCR	polymerase chain reaction
pfu	plaque forming units
RNA	ribonucleic acid
s	second
USP	universal sequencing primer

THESIS SUMMARY

Mucopolysaccharidosis type I (MPS-I, eponyms: Hurler, Hurler/Scheie and Scheie syndromes) is an autosomal recessive genetic disease caused by a deficiency of the glycosidase α-L-iduronidase (IDUA) which is required for the lysosomal degradation of the glycosaminoglycans heparan sulphate and dermatan sulphate. Patients with MPS-I store these partially degraded glycosaminoglycans in their lysosomes. MPS-I patients have a wide range of clinical presentations (Hurler (severe clinical phenotype), Hurler/Scheie (intermediate) and Scheie (mild) syndromes), which makes it difficult to predict patient phenotype that is necessary for genetic counselling and also impedes the selection and evaluation of patients undergoing therapy such as bone marrow transplantation. This work was performed with the aim of cloning the gene for IDUA, to enable molecular genetic diagnosis of mutations causing MPS-I for more accurate disease prognosis, and make possible new therapy protocols such as enzyme replacement therapy and gene replacement therapy in the near future. An understanding of the enzymology of IDUA and lysosomal biogenesis may also be obtained. The following results were obtained.

1.) Peptide data obtained from purified IDUA was used to design oligonucleotide probes which were used to screen cDNA and genomic libraries. An EMBL3 genomic clone positive to the oligonucleotide probe ID47 (λID475) was confirmed to be IDUA by showing that it contained nucleotide sequence colinear to the peptide sequence of purified IDUA.

2.) *IDUA* had previously been localized to chromosome 22. To further confirm the identity of the genomic clone, chromosomal localization studies were performed by *in situ* hybridization to human metaphase chromosomes, Southern blot analysis of human-mouse cell hybrids, and immunocapture of human IDUA from human-mouse cell hybrids. These studies revised the localization of *IDUA* from chromosome 22 to chromosome 4p16.3.

3.) Screening of a random-primed colon cDNA library with a 1.6-kb *Pst* I subclone of λID475 produced a single positive with a 729-bp insert which contained nucleotide sequence colinear to 6 peptide sequences of purified IDUA. Screening of seven other cDNA libraries produced a further 58 positives, the longest of which (λE8A) was 1765-bp long. Of these clones, 45 ended at the same 5' base as λE8A.

4.) Regional mapping of the 1.6-kb *Pst* I genomic fragment had further localized the gene for IDUA to 1100-kb from the telomere of chromosome 4p and 1000-kb from either of the two most likely regions for the Huntington disease gene and within a cosmid clone A157.1. Using a revised peptide model for IDUA, it was hypothesized that the 74/13-kDa N-terminal peptide sequence was 5' to the sequence that had thus far been obtained. An oligonucleotide probe to this peptide sequence was used to screen the cosmid A157.1 and a 2.8-kb *Bam*HI fragment positive to this probe was subcloned and partially sequenced. An exon encoding an initiating methionine, a signal peptide, the 74/13-kDa N-terminal sequence and part of a tryptic peptide from IDUA was identified.

5.) Oligonucleotide primers made to exon I of IDUA and the 5' end of the cDNA clone λE8A were used to PCR amplify IDUA mRNA sequence from cDNA, and two PCR products were obtained. The major PCR product contained exon II of IDUA including sequence that encoded the last unlocated tryptic peptide for IDUA and the minor product was shown to be the product of alternative splicing of exon II of the IDUA gene. Alternative splicing was also found to occur for exon IV of IDUA.

6.) A full-length cDNA construct for IDUA was constructed using PCR, cloned into the expression vector pRSVN.07 and transfected into Chinese hamster ovary cells. A 160-fold increase in IDUA activity was observed and monoclonal antibodies were used to prove that the transfected cell line was producing human IDUA of a normal specific activity. This was the final proof that the sequence obtained was indeed the sequence that encoded human IDUA. Thus, IDUA is translated as a 653 amino acid precursor with a 26 amino acid signal peptide that is cleaved immediately prior to the 74/13-kDa

N-terminus. The protein sequence contains six potential N-glycosylation sites. Northern blot analysis revealed only a single 2.3-kb mRNA.

7.) Mapping and sequencing of the EMBL3 genomic clone λID475 and the cosmid clone A157.1 showed that the gene for IDUA was split into 14 exons spread over 20-kb. The potential promoter for IDUA has only GC box type consensus sequences consistent with a housekeeping promoter and is bounded by an Alu-type repeat sequence. The first 2 exons are separated by an intron of 566-bp followed by a large intron of approximately 13-kb (intron 2). The last 12 exons are clustered in a 4.5-kb segment. No consensus polyadenylation signals were found although two variant polyadenylation signals are proposed.

8.) Two polymorphisms were found to exist in the IDUA gene, a *Kpn* I polymorphism in exon III and a variable number of tandem repeats (VNTR) polymorphism with three common alleles in intron 2. This VNTR polymorphism is the well characterized *D4S111* polymorphic locus used in Huntington disease diagnosis. The analysis of allele and haplotype frequencies for these two polymorphisms in the normal population and in MPS-I patients revealed the presence of linkage disequilibrium. The frequency of the 2,2 (VNTR, *Kpn* I) allele in MPS-I patients was 57% compared to only 37% in the normal population. This implied the presence of a major MPS-I allele. There was also slight disequilibrium between the 2,1 haplotype and MPS-I implying a second common mutation.

9.) On the basis of this haplotyping, a group of patients with the 2,2 haplotype and some with the 2,1 haplotype were selected for mutation analysis by chemical cleavage and direct PCR sequencing. Several patients with mild phenotypes and unusual biochemistry were also selected for mutation analysis. A G to A substitution was detected in exon IX of IDUA that altered a tryptophan codon at position 402 (TGG) to a stop codon (TAG), thus resulting in premature termination of translation approximately two thirds of the way through the 653 amino acid IDUA protein ($W_{402}X$). $W_{402}X$ was found associated with the 2,2 haplotype and accounted for 33% of MPS-I alleles in a group of 76 patients studied. Patients homozygous for $W_{402}X$ have an extremely

severe Hurler-like clinical phenotype. Patients heterozygous for $W_{402}X$ have the full range of clinical phenotypes found in MPS-I.

10.) A C to T substitution in exon II which altered a glutamine codon at position 70 (CAG) to a stop codon (TAG) was found associated with the 2,1 haplotype and accounted for 15% of MPS-I alleles in the patient group ($Q_{70}X$). $Q_{70}X$ is also associated with a very severe Hurler-like clinical phenotype as can be seen in patients with the $W_{402}X$, $Q_{70}X$ genotype. Thus two mutations accounting for 48% of MPS-I allele and therefore 23% of MPS-I genotypes have been defined. The present clinical phenotype of patients who have had both alleles defined as either $W_{402}X$ or $Q_{70}X$, and have undergone bone marrow transplantation provides encouraging evidence that MPS-I is treatable by bone marrow transplantation and therefore also possibly treatable by enzyme replacement and gene replacement therapy in the future.

11.) Three amino acid substitutions, one single base deletion and two splicing mutations have also been found. One splicing mutation, $g_{1487}a$, was observed in three mildly affected patients. It is hypothesized that a small amount of normal mRNA is produced from this mutant allele leading to the extremely mild phenotype seen in these patients, two of whom are heterozygous for $W_{402}X$. Neither the biochemical or clinical outcome of $t_{1978}c$ splicing mutation is known. One of the amino acid substitution mutations, $R_{89}Q$, is also associated with a mild phenotype. These are the first observations to explain the vast difference in the extremes of the clinical spectrum for MPS-I. An amino acid mutation, $D_{349}N$, may be involved in active site catalysis and/or processing of IDUA.

STATEMENT

This thesis contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference is made in the text.

I give consent to this copy of my thesis, when deposited in the University Library, being available for loan and photocopying.

Signed:

Hamish S. Scott

Date: October 6, 1992

ACKNOWLEDGEMENTS

I wish to thank Dr C. Phillip Morris and Professor John J. Hopwood, my supervisors, for their constant encouragement, enthusiasm, advice and friendship during the course of this work. I have learnt a great deal from the opportunity that they gave me to work with them, and for this, I am most grateful. My thanks to Dr E. F. Robertson, the present director, and Dr A.C. Pollard the immediate past director of the Department of Chemical Pathology at the Adelaide Children's Hospital, who are responsible for setting up and maintaining the clinical and departmental infrastructure in which this research took place. Without their continued efforts this work would not have been possible. My thanks to Professor Barrie Vernon-Roberts of the Department of Pathology at the University of Adelaide for enabling me to carry out this degree through his department.

My thanks to Dr Peter Clements and Dr Doug Brooks who were instrumental in the isolation of IDUA leading to production of peptide sequence, without which this study could not have commenced. Annette Orsborn, Xiao-Hui Guo, Lesley Ashton, Teresa Occhiodoro and Dr Don Anson all played a vital and much appreciated role in the isolation, characterization and expression of IDUA. Thanks to Kathy Nelson and Greta Richardson who, undaunted, stored, grew and maintained the large number of normal and MPS-I patient fibroblasts. I also thank Paul Nelson and Tom Litjens for their exceptionally professional, friendly and uncomplaining help, particularly with MPS-I patient analysis. Thanks to all the other members of the Molecular Biology group, the Lysosomal Diseases Research Unit and the Department of Chemical Pathology for their direct and indirect contributions to this work. All these people, and others, were responsible for creating an enjoyable and stimulating work environment.

I am also grateful for the collaboration with the Department of Cytogenetics and Molecular genetics at the Adelaide Children's Hospital for the chromosomal localization of *IDUA*. Thanks to Dr Alan Robbins and Jacqueline Beale of the Department of Biochemistry at the University of Adelaide for the expression vector pRSVN.07 and the CHO cells respectively. Several international collaborations were of

importance to the work presented in this thesis and I would like to thank the people involved. Professor James. F. Gusella and Dr Marcy MacDonald of the Neurogenetics Laboratory at the Massachusetts General Hospital performed the regional mapping study and provided the cosmid clone used in this study. Dr J. Edward Wraith and Mr Alan Cooper of the Willink Biochemical Genetics Unit, Royal Manchester Children's Hospital provided a large number of the MPS-I DNA samples used in the patient study. Dr Wraith also provided a large amount of clinical data which was greatly appreciated. Dr Nina Horn of the Department of Biochemistry and Molecular Genetics at the John F. Kennedy Institute, Denmark, who provided several MPS-I patient samples including a $Q_{70}X$ homozygote. Dr Ashok Vellodi of the Westminster Children's Hospital, who provided clinical data and samples of bone marrow transplanted MPS-I patients. Dr Chester Whitely of the University of Minnesota Medical School, who also provided a large number of DNA samples from BMT MPS-I patients. All these contributions were important to building an overall picture of the state of MPS-I research.

My everlasting gratitude to Teresa Occhiodoro for her patience with me and this thesis during its writing, and her invaluable proof reading. Lastly, and most of all, I would also like to thank my family and friends who were always there to support me, particularly towards the end.

CHAPTER 1

INTRODUCTION

1.1 PRELIMINARY COMMENTS

Since Garrod introduced the concept of inborn errors of metabolism (Garrod, 1909) by his perception of the direct relationship between genes, enzymes and the theories of Gregor Mendel (1866), hundreds of genetic diseases resulting from the deficiency of a single enzyme have been described (Scriver *et al.*, 1989). Of these diseases, 36 are broadly described as lysosomal storage disorders (LSD) (Neufeld, 1991). The concept of LSD was first described by Hers (1965) to explain how the deficiency of α-glucosidase could result in the accumulation of undegraded substrate in the lysosome, leading to the pathology of Pompe's disease. Of the 36 lysosomal storage diseases, ten are the result of deficiency of enzymes specific for the degradation of sulphated glycosaminoglycans (GAG) or mucopolysaccharides and are thus termed mucopolysaccharidoses (MPS). The MPS are characterized primarily on the basis of their clinical phenotype and the type of GAG excreted in the urine. Hurler, Scheie and Hurler/Scheie syndromes are subgroups of mucopolysaccharidosis type I (MPS-I) based on the clinical severity of the disease. The study of the MPS has been the subject of intensive research over the last 30 years. Although each MPS disorder is relatively rare in the human population, they present a significant medical and social problem. In addition, the investigation of hereditary disorders can lead to new knowledge of normal developmental and biochemical mechanisms.

This work was performed with the aim of cloning the gene for α-L-iduronidase (*IDUA*) which is known to be defective in MPS-I, to enable molecular genetic diagnosis of mutations causing MPS-I for more accurate disease prognosis, and to make possible new therapy protocols such as enzyme replacement therapy and gene replacement therapy in the near future. It was also expected that an understanding of the enzymology of IDUA and the biogenesis of the lysosome may also be obtained.

This introduction is intended to provide an historical, clinical, biochemical and genetic background to MPS-I. The following sections will discuss biochemistry that is applicable to all of the MPS disorders and the characterization of MPS-I, the prototype MPS. Although it is not intended to act as a general review of lysosomal biogenesis

and function or the molecular pathology of the lysosomal storage disorders, much of the detail described for IDUA and other enzymes involved in the degradation of HS and DS is relevant to other lysosomal enzymes.

1.2 THE MUCOPOLYSACCHARIDOSES

The MPS are a group of heritable disorders caused by a deficiency of lysosomal enzymes required for degradation of sulphated GAG or mucopolysaccharides, heparan sulphate (HS), dermatan sulphate (DS) or keratan sulphate. In the MPS, forms of these partially degraded GAG are stored in the lysosomes of all cells (Fig. 1.1) and excreted in the urine. GAG are normally degraded to inorganic sulphate and monosaccharides in the lysosome and the storage of the partially degraded GAG eventually results in cell, tissue and organ dysfunction which presumably gives rise to the clinical symptoms of the MPS. Each of the known MPS specifically involves the deficiency of one of ten enzymes that are required for the step-wise degradation of these GAG. Each enzyme may be required for the degradation of one or more GAG. The MPS disorders show many clinical features in common. In general all MPS are chronic and progressive with central nervous system (CNS) and/or skeletal involvement and often display a wide spectrum of clinical severity within the one enzyme deficiency. Multiple allelism at each enzyme locus has been proposed to explain this spectrum of clinical variability (Neufeld and Muenzer, 1989).

1.2.1 HISTORY OF THE MUCOPOLYSACCHARIDOSES

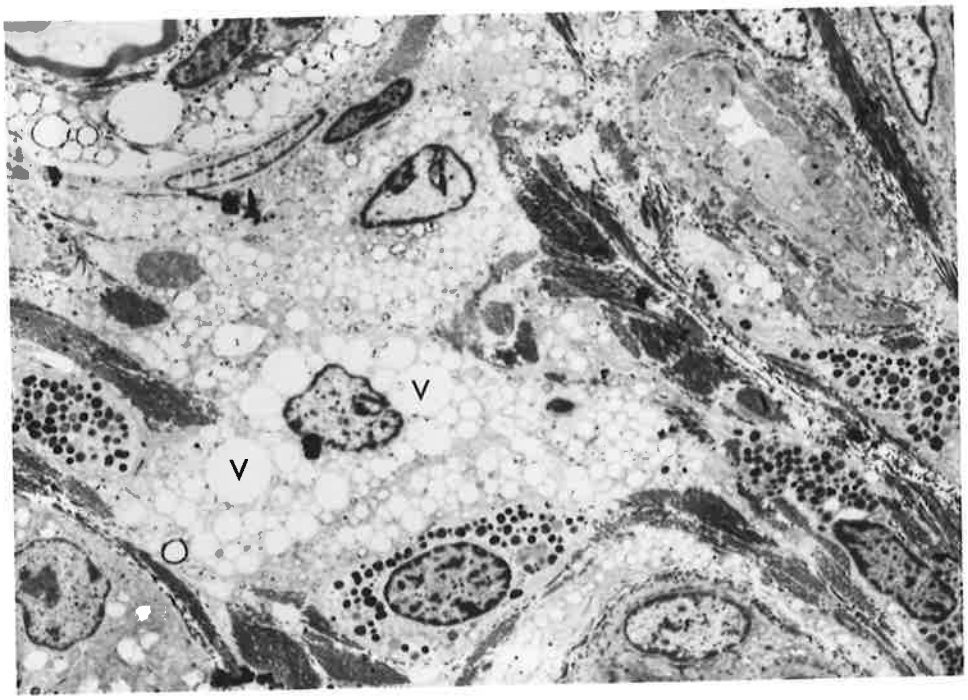
With the benefit of hindsight, the first definitive report of an MPS disorder was a description of the X-linked MPS-II syndrome by Hunter (1917), thus called Hunter syndrome. Prior to this however, there are case reports dating to between 1900 and 1913 of a family in Edinburgh showing that three of nine children in one family had MPS-I or Hurler syndrome (McKusick, 1972). The first definitive report of MPS-I was by a German physician, Gertrude Hurler, in 1919 and thus the MPS-I eponym, Hurler

syndrome. Originally, MPS-I and MPS-II were classified together under the one name, gargoylism, to describe the distinctive facial features associated with the MPS disorders. This unsatisfactory and cruel term began to fade from use with the recognition of two distinct forms of the disease. MPS-II was demonstrated, by extensive pedigree analysis in families with MPS-II patients, to have an X-linked recessive mode of inheritance (Nja, 1946) in contrast with MPS-I which is inherited as an autosomal recessive disease (Halperin and Curtis, 1942). The general term lipochondrodystophy was mistakenly introduced in the 1930s as it was thought that the disorders were ones of lipid metabolism (Washington, 1937). The introduction of misleading terms for genetic diseases is one reason that syndromic names are preferred until the biochemical basis of the disease is established. In 1952, Brante demonstrated that the stored material was acid mucopolysaccharide by isolating dermatan sulphate from the livers of two Hurler syndrome patients, and consequently suggested the term mucopolysaccharidoses. Dorfman and Lorincz (1957) reported the increased urinary excretion of partially degraded mucopolysaccharides (or GAG) in MPS patients, which is still the basis of the diagnostic testing for the MPS disorders. Van Hoof and Hers (1964) used electron microscopy to demonstrate large vacuoles containing granular material in the liver of a Hurler syndrome patient and suggested that the partially degraded mucopolysaccharides were distending the lysosome (Hers, 1965) as shown in Fig. 1.1. At around the same time, Danes and Bearn (1965) demonstrated metachromatic staining of inclusion bodies in the cytoplasm of MPS patient fibroblasts fixed in basic lead acetate.

In 1965, McKusick *et al.* systemized the classification of MPS into six eponymic types, numbered simply I to VI, based on the types of mucopolysaccharides excreted in the urine, the mode of inheritance and the clinical features. Hurler syndrome was considered the model or prototype MPS and thus was labelled MPS-I. The X-linked Hunter syndrome was labelled MPS-II. Sanfilippo syndrome (MPS-III), Maroteaux-Lamy syndrome (MPS-VI) and Scheie syndrome (MPS-V) had previously been considered "Hurler variants". MPS-V was a mild mucopolysaccharidosis similar to Hurler syndrome (MPS-I) first described by Scheie *et al.* (1962). Morquio syndrome

FIGURE 1.1. LYSOSOMAL INCLUSION BODIES.

Skin fibroblasts from a cat affected with Maroteaux-Lamy syndrome (MPS-VI) demonstrate lysosomal inclusion bodies which appear as vacuoles (V). These vacuoles are lysosomes engorged with partially catabolized glycosaminoglycans. (Courtesy of Prof. J.J. Hopwood).


V
V
O

(MPS-IV) has characteristic mucopolysaccharidurias and skeletal features which distinguishes it from the other MPS disorders.

Soon after this classification, a series of historic experiments took place. Fratantoni *et al.* (1968b; 1969) demonstrated that cultured fibroblasts from MPS-I and MPS-II patients accumulated excessive amounts of radioactive sulphated mucopolysaccharides from culture medium compared to normal controls. This storage resulted from defective degradation of the mucopolysaccharides in MPS-I and MPS-II fibroblasts and was corrected by co-cultivation of MPS-I and MPS-II patient fibroblasts (Fratantoni *et al.*, 1968a). Subsequent studies showed that "corrective factors" were excreted into the media of cultured cells, or could be extracted from urine or tissue. It was also shown by co-cultivation of patient fibroblasts that Scheie (MPS-V) and Hurler syndromes (MPS-I), although clinically very different, were in fact the result of the same biochemical deficiency (Wiesmann and Neufeld, 1970) and were reclassified as the clinical subgroups, MPS-IH and MPS-IS (McKusick *et al.*, 1972).

Similar experiments soon led to the delineation of the specific enzyme defects responsible for the MPS disorders. In 1972, it was shown that MPS-I resulted from a deficiency of the lysosomal glycosidase, α-L-iduronidase (IDUA), and thus Hurler and Scheie syndromes were thought to be caused by different mutant alleles at the IDUA locus (Bach *et al.*, 1972; Matalon and Dorfman, 1972). An alternative hypothesis was that the mutations for Hurler and Scheie syndromes resided in different genes, each encoding a different subunit of IDUA (McKusick *et al.*, 1972). Two IDUA isoenzymes specific for HS and DS were possibly thought to exist and mutations causing Hurler syndrome were in a common subunit while Scheie syndrome mutations affected only the DS specific isoenzyme (Matalon and Deanching, 1977). Co-cultivation of patient fibroblasts and studies using GAG derived substrates have led to the identification of ten biochemically distinct MPS disorders, each relating to a specific enzyme defect. From the six MPS disorders originally suggested by McKusick *et al.*, (1965), MPS-V has been reclassified as a subgroup of MPS-I, MPS-III (Sanfilippo syndrome) has been found to contain four biochemically distinct types (MPS-IIIA, B, C, and D), MPS-IV

(Morquio syndrome) has been found to contain two biochemically distinct types (MPS-IVA and B) and a deficiency of β-D-glucuronidase has been classified as MPS-VII (Sly syndrome) to total ten MPS disorders (McKusick and Neufeld, 1983). Once an enzyme deficiency was identified for each MPS disorder, an intensive effort was made to understand the biochemistry and pathology of each disorder, leading to the eventual purification of each MPS enzyme and subsequent cloning of the genes (for reviews, see Hopwood and Morris, 1990; Neufeld, 1991).

1.2.2 BIOCHEMISTRY OF THE MUCOPOLYSACCHARIDOSES

1.2.2a DEGRADATION OF HEPARAN SULPHATE

The understanding of normal GAG catabolism and the MPS is closely linked as many of the enzymes involved were defined as a consequence of their absence. The degradation of HS is a sequential process involving up to nine lysosomal *exo*enzymes. Deficiencies of all but two of these lysosomal enzymes, glucuronate-2-sulphatase and glucosamine-3-sulphatase, have been reported to lead to one of the MPS. All the MPS disorders that result from a defect in HS degradation may have CNS involvement and a correlation between the level of GAG storage and the clinical phenotype has been proposed (Conzelmann and Sandhoff, 1984).

HS is synthesized as a proteoglycan with two or more HS chains covalently linked to a protein core. Some newly synthesized HS proteoglycan is externalized to the cell surface and then endocytosed and catabolized with a half life of 4 to 6 hours (reviewed in Hopwood, 1989). HS proteoglycan is found associated with the extracellular matrix and plasma membrane (see Fig. 1.2; Fransson, 1989). The exact function(s) of HS proteoglycan still remains unknown, but from various experiments it has been implicated to be involved in control of cell adhesion, migration, differentiation, proliferation (for reviews, see Poole, 1982; Stigson and Kjellen, 1991), the binding of growth factors and angiogenesis (for review, see Folkman and Shing,

1992). Fig. 1.2 is a schematic representation of the pathways of synthesis and catabolism of HS.

The first stage of HS degradation is thought to occur in early endosomes, where proteolysis produces single HS chains of approximately 30-kDa (for review, see Freeman and Hopwood, 1992b), which are then hydrolyzed by endo-β-glucuronidase activities to oligosaccharides with a molecular mass of approximately 5-kDa (for reviews, see Hopwood, 1989; Freeman and Hopwood, 1992b). HS oligosaccharides contain alternating uronic acid and α-linked glucosamine residues (for reviews, see Roden, 1980; Fransson, 1989; Hopwood, 1989). The uronic acid residues may be β-D-glucuronic acid or α-L-iduronic acid which may be C-2-sulphated. The glucosamine residues may be *N*-sulphated or *N*-acetylated, C-6- and C-3-sulphated. The 5-kDa HS oligosaccharides are transported to the lysosome to be degraded from their non-reducing-ends by the sequential action of up to three glycosidases, corresponding to the three different monosaccharides in HS, five sulphatases, which remove sulphate residues before the glycosidases act, and an *N*-acetyltransferase to *N*-acetylate terminal glucosamine residues (Fig. 1.3). The reactions of the glycosidases and sulphatases releases monosaccharides and inorganic sulphate respectively. The degradation of the HS to monosaccharides and inorganic sulphate is extremely rapid such that oligosaccharides sized between monosaccharides and 5-kDa are not observed. The monosaccharides and inorganic sulphate generated in the degradative process are thought to be actively transported from the lysosome, possibly to be reused in the cell's biosynthetic processes. Some of the nine *exo*enzymes involved in the degradation of HS must act sequentially due to substrate constraints. For example, the appropriate substrate for IDUA will always be produced by the action of iduronate-2-sulphatase, however the order of action of the other enzymes may depend on the order of the different monosaccharides and their pattern of sulphation in the HS chain. The speed and order of this degradative process, in conjunction with other observations, have led to the hypothesis that the *exo*degradation of heparan sulphate is coupled and handled by a multienzyme complex (Hopwood, 1989; Freeman and Hopwood, 1992a).

FIGURE 1.2. BIOSYNTHESIS AND DEGRADATION OF HEPARAN SULPHATE.

(From Hopwood, 1989)

The diagram represents a simplified scheme detailing the processes involved in the biosynthesis and catabolism of heparan sulphate (HS). On the left, the events involved in the biosynthesis of HS in the Golgi are depicted, followed by secretion from the cell and insertion into the cell membrane. The right side of the diagram represents the steps involved in the endocytosis and catabolism of HS.

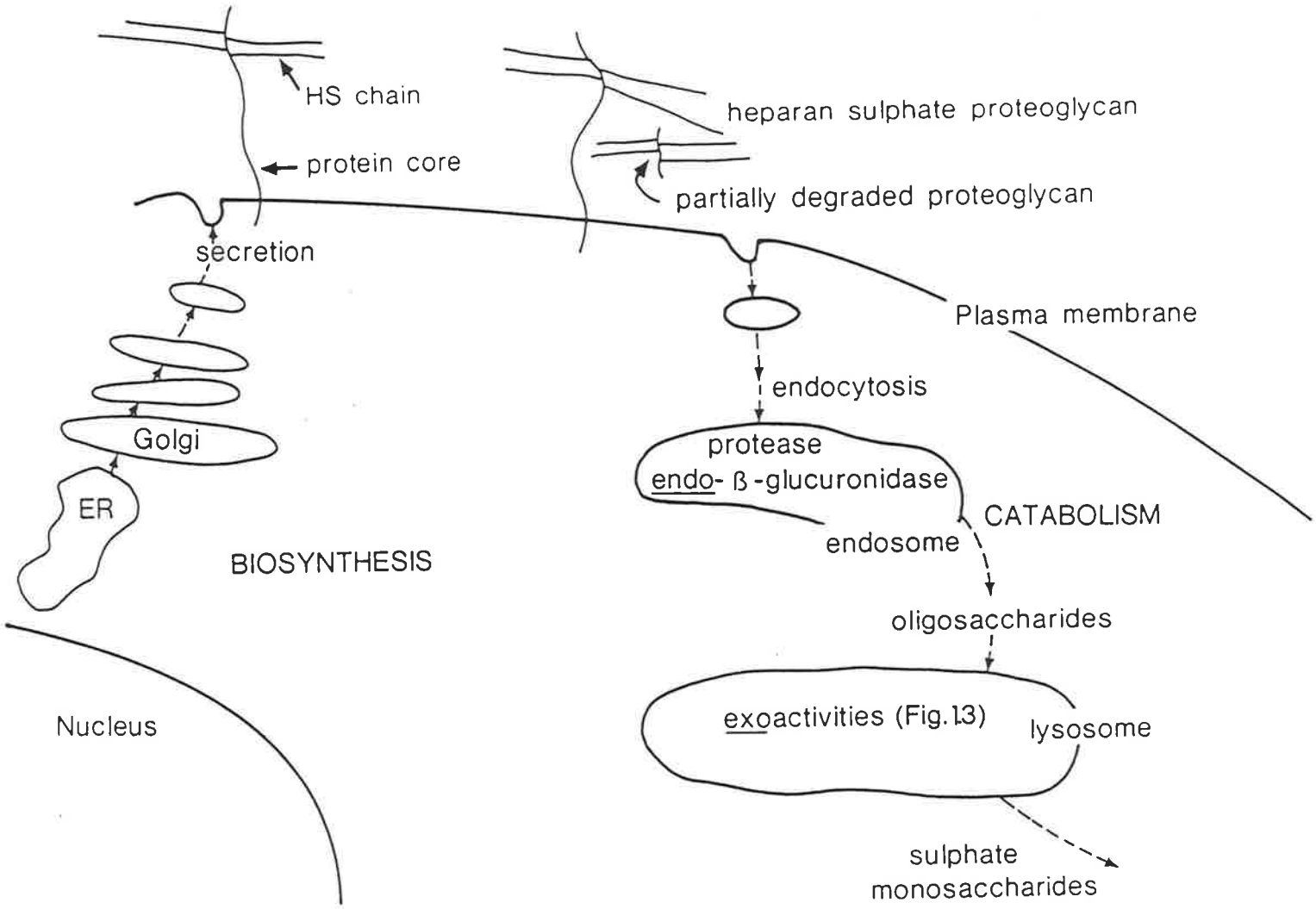

HS chain
protein core
heparan sulphate proteoglycan
partially degraded proteoglycan
secretion
Plasma membrane
endocytosis
Golgi
protease
endo- ß -glucuronidase
ER
CATABOLISM
endosome
BIOSYNTHESIS
oligosaccharides
Nucleus
exoactivities (Fig.13)
lysosome
sulphate
monosaccharides

FIGURE 1.3. DEGRADATION OF HEPARAN SULPHATE.

(From Hopwood and Morris, 1990)

Schematic representation of the step-wise degradation of HS oligosaccharides in the lysosome. The diagram shows the bonds that are modified in HS by the action of the nine *exo*enzymes. The relative positions of the glucuronate-2-sulphate and glucosamine residues are for convenience only. The mucopolysaccharidosis type corresponding to a particular enzyme defect in the degradative pathway is shown alongside the broken arrows.

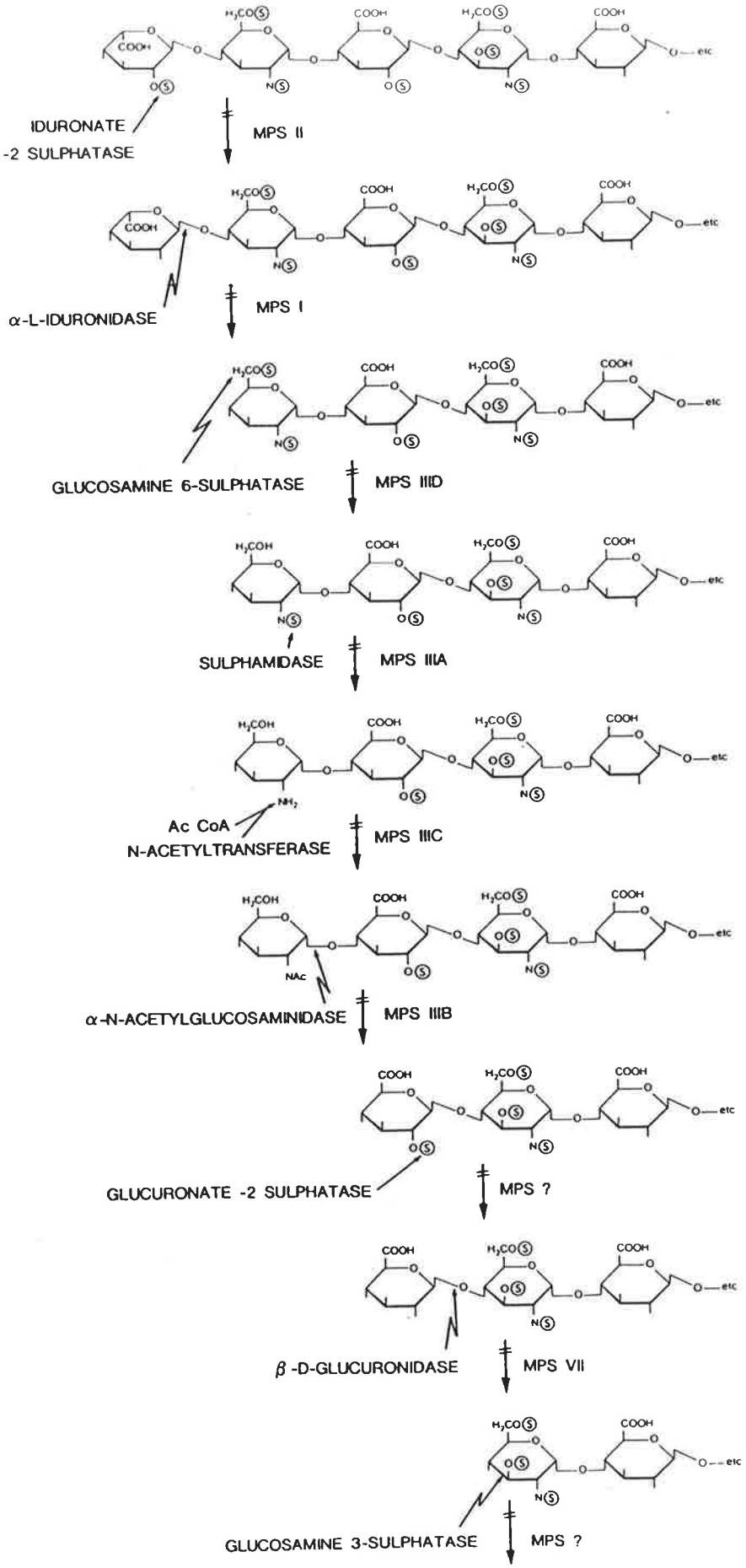

IDURONATE -2 SULPHATASE
MPS II
α-L-IDURONIDASE
MPS I
GLUCOSAMINE 6-SULPHATASE
MPS IIID
SULPHAMIDASE
MPS IIIA
Ac CoA N-ACETYLTRANSFERASE
MPS IIIC
α-N-ACETYLGLUCOSAMINIDASE
MPS IIIB
GLUCURONATE -2 SULPHATASE
MPS ?
β -D-GLUCURONIDASE
MPS VII
GLUCOSAMINE 3-SULPHATASE
MPS ?

Fig. 1.3 illustrates the enzymology of HS degradation and the MPS disorders resulting in the case of each enzyme deficiency. Iduronate-2-sulphatase, the enzyme deficient in MPS-II, hydrolyses terminal iduronate-2-sulphatase esters in HS. IDUA, which is deficient in MPS-I, acts on the oligosaccharide product of iduronate-2-sulphatase to hydrolyse the non-reducing-end α-L-iduronide glycosidic bond in HS to produce iduronic acid and the substrate for glucosamine-6-sulphatase, the next enzyme involved in the sequential degradation of HS (Fig. 1.3). Glucosamine-6-sulphatase, which is deficient in MPS-IIID, hydrolyses the sulphate bond on terminal glucosamine-6-sulphate residues. Sulphamidase (heparan-*N*-sulphatase), the enzyme which is deficient in MPS-IIIA, hydrolyses the *N*-sulphate on the non-reducing-end glucosamine residues of HS. The next step in HS degradation is the only biosynthetic step in the process. At this step, acetyl-CoA:α-glucosaminide-*N*-acetyltransferase or *N*-acetyltransferase (GNAT), the enzyme deficient in MPS-IIIC (Klein *et al.*, 1978), which is an integral membrane protein (Freeman *et al.*, 1983), transfers an acetyl group from acetyl-CoA to *N*-acetylate the terminal glucosamine residue of HS. The reverse of this unusual reaction occurs in the heparin or HS biosynthetic pathway in the Golgi, where the hydrolase, *N*-deacetylase, deacetylates terminal glucosamine residues in HS before further addition of monosaccharides can proceed (Lindahl *et al.*, 1986). The next enzyme in the pathway, α-*N*-acetylglucosaminidase, which is deficient in MPS-IIIB, removes the acetylated terminal glucosamine residue of HS. Glucuronate-2-sulphatase, which may also act on 2-sulphated glucuronic acid residues in chondroitin sulphate, then desulphates the terminal glucuronate residue of HS. Patients deficient for this enzyme have yet to be described, however, they would be expected to be phenotypically similar to mild MPS-IIIA to D patients (Sanfilippo syndrome) as comparatively few 2-sulphated glucuronate residue are present in HS or chondroitin sulphate (Hopwood, 1989; Hopwood and Morris, 1990). β-D-Glucuronidase, which is deficient in MPS-VII, removes terminal glucuronate residues from HS. Unlike the other enzymes in this pathway, this glycosidase is also present in microsomes (Lusis and Paigen, 1977). Glucosamine-3-sulphatase removes sulphate from terminal glucosamine residues. Patients deficient for this enzyme have not been described but, as with glucuronate-2-

sulphatase, they may also be expected be phenotypically similar to mild MPS-IIIA to D patients (Freeman and Hopwood, 1991). At the time the work described in this thesis began, only the genes for glucosamine-6-sulphatase and β-D-glucuronidase from the enzymes described above had been cloned (Robertson *et al.*, 1988 and Oshima *et al.*, 1987 respectively).

1.2.2b DEGRADATION OF DERMATAN SULPHATE

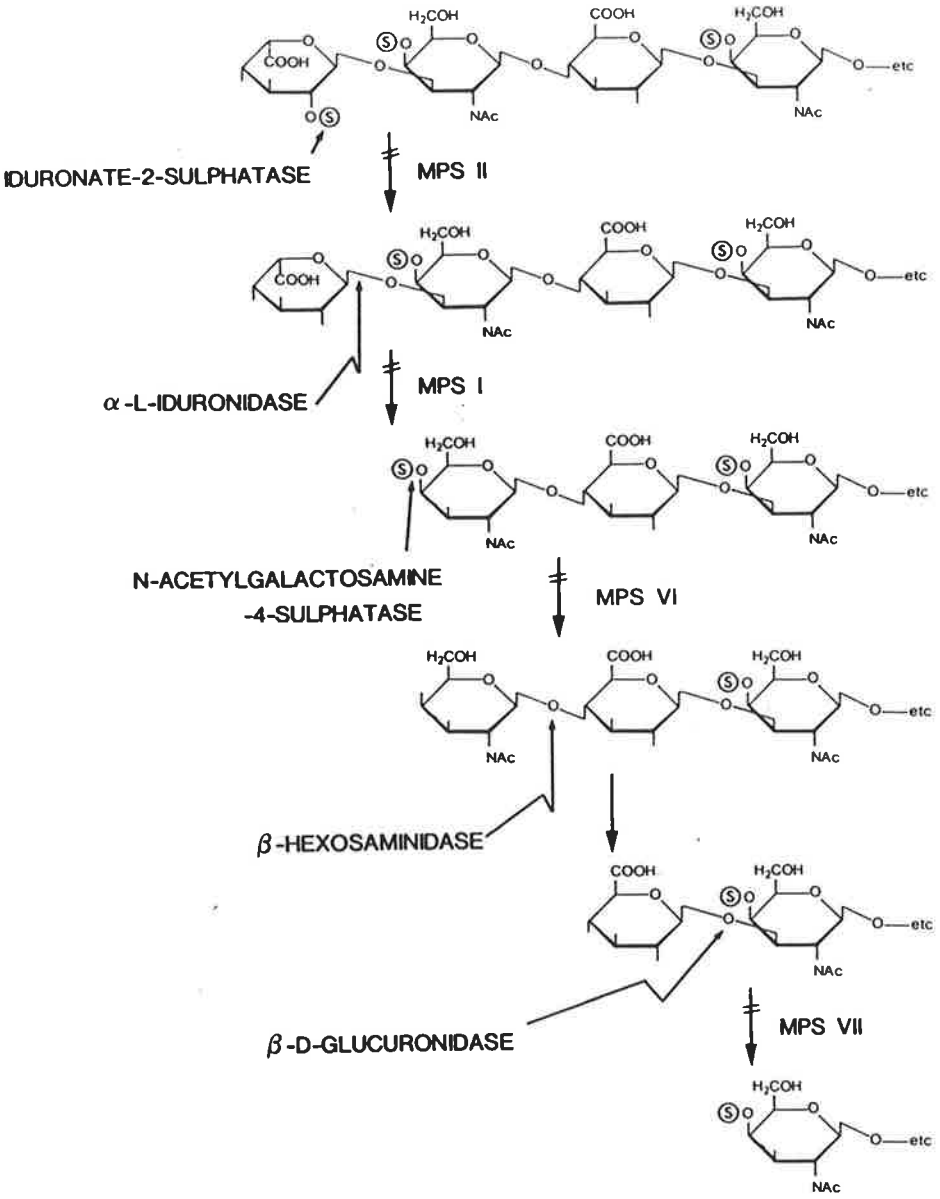
In a similar way to HS, DS is synthesized as a proteoglycan but is mainly found in the matrix of connective tissues. DS is also regularly endocytosed and then degraded. Proteolysis of endocytosed DS proteoglycan produces single DS chains that are further reduced to DS oligosaccharides, presumably by the action of endoglycosidase activities present in the endocytic pathway to the lysosome. In the lysosome, DS oligosaccharides are degraded by the step-wise action of the enzymes shown in Fig. 1.4. Similar to HS, DS consists of iduronic acid and glucuronic acid residues alternating with β-linked N-acetylgalactosamine residues compared with α-linked glucosamine residues in HS. The iduronic residues of DS may be 2-sulphated and the N-acetylgalactosamine residues may be C-4- or C-6-sulphated (for review, see Roden, 1980). Thus, two of the glycosidases, IDUA and β-D-glucuronidase, and the sulphatase, iduronate-2-sulphatase, which are required for the degradation of HS are also required for the degradation of DS, and both HS and DS accumulate in patients who are deficient for these enzymes (MPS-I, MPS-VII and MPS-II respectively). The integral role of DS in connective tissue is reflected by the skeletal deformities seen in MPS patients with a deficiency of the enzymes required to degrade DS.

The degradation of DS oligosaccharides is similar to the degradation of HS. The actions of IDUA, β-D-glucuronidase and iduronate-2-sulphatase are as discussed for HS except that the action of IDUA produces the substrate for *N*-acetylgalactosamine-4-sulphatase. *N*-acetylgalactosamine-4-sulphatase, which is deficient in MPS-VI, desulphates terminal *N*-acetylgalactosamine residues of DS. The removal of *N*-acetylgalactosaminide residues is thought to be by hydrolysis resulting from the action of the β-hexosaminidase isoenzymes A, B or S which also removes *N*-

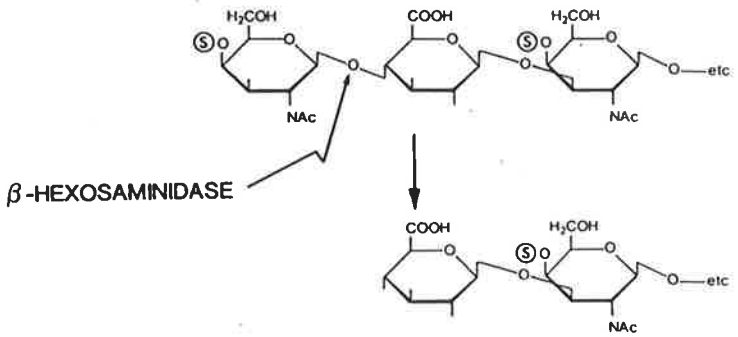
FIGURE 1.4. DEGRADATION OF DERMATAN SULPHATE.

(From Hopwood and Morris, 1990)

Schematic representation of the step-wise degradation of DS oligosaccharides by the action of five *exo*enzymes in the lysosome. The mucopolysaccharidosis type corresponding to a particular enzyme defect in the degradative pathway is shown alongside the broken arrows.



ALTERNATE PATHWAY



acetylgalactosaminide residues from G_{M2}-gangliosides. Unlike the other enzymes in this pathway, β-hexosaminidase is made of subunits encoded for by different genes. Also, a deficiency of β-hexosaminidase does not cause an MPS disorder but rather one of the G_{M2}-gangliosidosis (Tay-Sachs or Sandhoff's disease, Sandhoff *et al.*, 1989). Again, as in HS degradation, β-D-glucuronidase removes terminal glucuronate residues.

Patients deficient in *N*-acetylgalactosamine-4-sulphatase (MPS-VI) accumulate *N*-acetylgalactosamine-4-sulphate and *N*-acetylgalactosamine-4,6-disulphate monosaccharide residues and thus an alternate pathway for DS degradation has been proposed (Hopwood and Elliott, 1985). β-Hexosaminidase A or S could hydrolyse the β-linked *N*-acetylgalactosaminide-4-sulphate residue to produce *N*-acetylgalactosamine-4-sulphate. This monosaccharide sulphate ester substrate can be hydrolysed by *N*-acetylgalactosamine-4-sulphatase to produce inorganic sulphate and *N*-acetylgalactosamine. The role that this alternate pathway may have in the normal degradation of DS is unknown. At the time the work described in this thesis began, from the enzymes described above, only the genes for both subunits of β-Hexosaminidase (Myerowitz *et al.*, 1984; O'Dowd *et al.*, 1985) and β-D-glucuronidase had been cloned (Oshima *et al.*, 1987).

IDUA is not known to be involved in the degradation of any other GAG such as chondroitin sulphate or keratan sulphate. However, in addition to the heparan and dermatan sulphaturia resulting from a deficiency of IDUA, chondroitin sulphaturia has also been reported in some cases of MPS-I (Babarik *et al.*, 1974; Leisti *et al.*, 1976) but a molecular explanation for this is unknown. It has been proposed that analytical error is responsible for these reports, or alternatively, the excretion of chondroitin 4- and 6-sulphates may result from the degradation of hybrid DS-choindroitin sulphate molecules (Neufeld and Muenzer, 1989). Also, in addition to mucopolysaccharide accumulation, the sphingolipid gangliosides G_{M2}, G_{M3} and G_{D3}, have been reported to accumulate in the brains of MPS patients with mental retardation, but are absent from MPS-IS patients without mental retardation (Constantopoulos *et al.*, 1980). It is hypothesized that the accumulated mucopolysaccharides inhibit other lysosomal enzymes, particularly

ganglioside neuraminidase (Baumkotter and Cantz, 1983), and this secondary accumulation of gangliosides may be at least partially responsible for the CNS involvement in the MPS.

1.2.2c BIOSYNTHESIS AND PROCESSING OF LYSOSOMAL ENZYMES

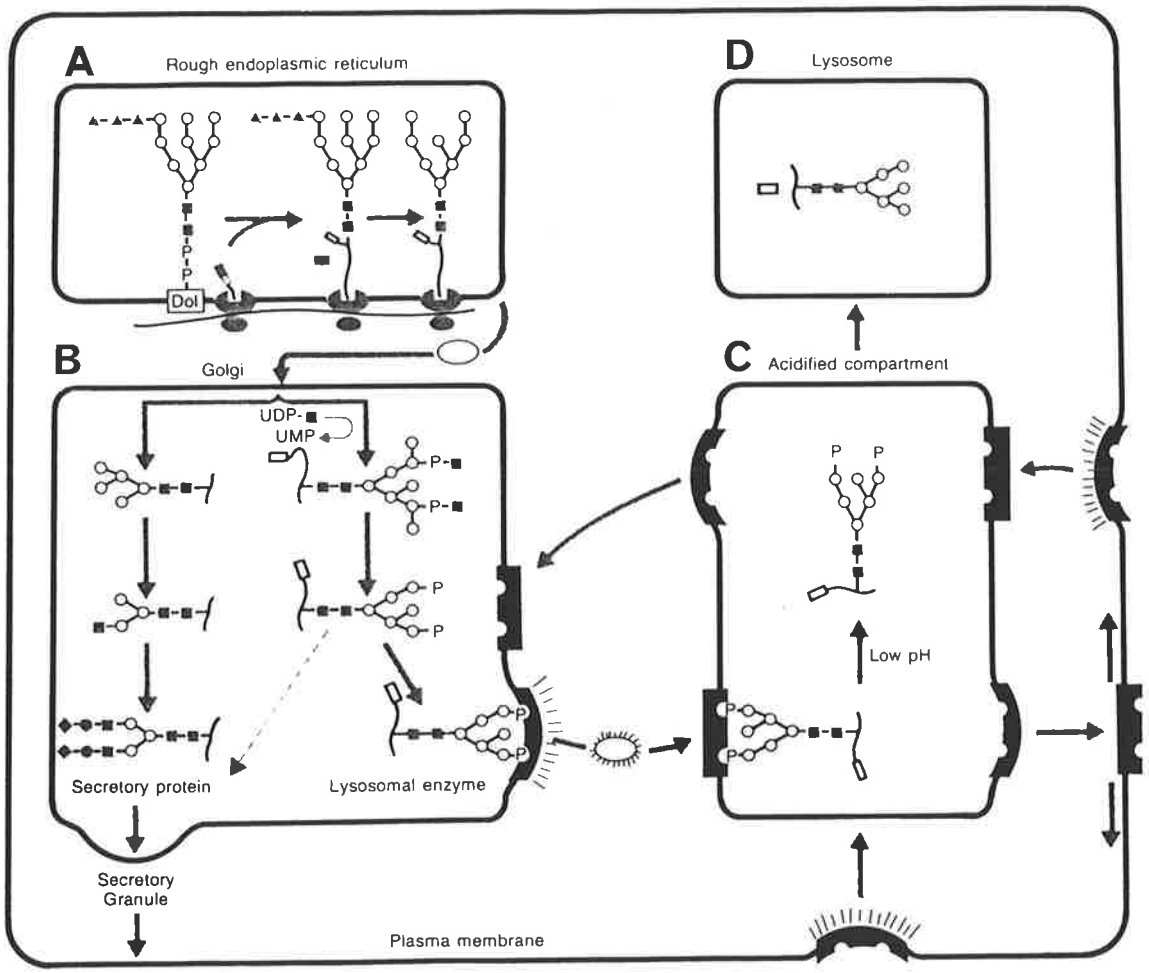
The biosynthetic, receptor-mediated transport and post-translational processing pathways of soluble lysosomal enzymes, such as most of the enzymes involved in GAG degradation discussed above, are well documented (for review, see von Figura and Haslik, 1986; von Figura *et al.*, 1987; Kornfeld and Mellman, 1989). The early events of lysosomal enzyme synthesis and transport are the same as for secretory proteins. Most soluble lysosomal enzymes are synthesized with an N-terminal endoplasmic reticulum (ER) signal sequence that is recognised by a signal recognition particle as it emerges from the ribosome after translation of the mRNA. The newly synthesized protein and signal recognition particle complex is in turn recognised by an ER membrane bound "docking protein". The nascent lysosomal protein is translocated across the membrane into the lumen of the ER followed by cleavage of the signal peptide (Erikson *et al.*, 1983).

Once the polypeptide enters the ER, preformed high mannose oligosaccharides are transferred from dolichol pyrophosphate carriers to some asparagine residues in the protein. The asparagine residues that are glycosylated are found as an Asn-X-Thr/Ser consensus sequence in the protein, although not all of these potential *N*-glycosylation amino acid sequence sites are used (Fig. 1.5, compartment A). These oligosaccharides are subjected to extensive trimming in the ER, and the new glycoprotein is transferred to the Golgi compartment where further trimming and modification occurs, including phosphorylation of some mannose residues in the cis-Golgi by the action of a glycosidase and phosphotransferase (Fig. 1.5, compartment B; Kornfeld and Kornfeld, 1985). This phosphorylation prevents further processing of the phosphorylated mannose residues, but other branches of the same oligosaccharide may be trimmed or

FIGURE 1.5. BIOSYNTHESIS AND TRANSPORT OF LYSOSOMAL ENZYMES.

(Modified from Scriver *et al.*, 1989)

Lysosomal enzymes are synthesized in the rough endoplasmic reticulum (A: RER) before being actively transported to the Golgi apparatus (B) where mannose groups are added to some of the oligosaccharide side chains to form mannose-6-phosphate recognition markers. Enzymes with these markers interact with the mannose-6-phosphate (M6P) receptors and are actively transported to acidified endosomes (C) and then lysosomes (D). M6P receptors on the plasma membrane bind extracellular lysosomal enzymes which are then internalized in coated vesicles for transport to the lysosome.


A
Rough endoplasmic reticulum
Dol
B
Golgi
UDP-
UMP
P
Secretory protein
Lysosomal enzyme
Secretory Granule
D
Lysosome
C
Acidified compartment
Low pH
Plasma membrane

further modified, for example, the addition of galactose and sialic acid residues and/or sulphate (Haslik and von Figura, 1981; Braulke *et al.*, 1987).

The acquisition of mannose-6-phosphate (M6P) residues targets the protein to the lysosome rather than the secretory pathway by its association with M6P receptors late in the Golgi compartment. There are two types of M6P receptors, cation-dependent and cation-independent, that occur throughout the Golgi complex and the trans-Golgi network. Both receptors appear to be involved in the targeting of lysosomal enzymes, although the relative role of each receptor and the role of the cation-dependent M6P receptor is less well understood. The cation-independent M6P receptor is a bifunctional protein with an extracytoplasmic domain consisting of 15 repeating domains homologous to the extracytoplasmic domain of the cation-dependent M6P receptor. These domains recognise phosphorylated lysosomal enzymes, and a separate functional domain acts as the insulin-like growth factor II receptor (for review, see Dahms *et al.*, 1989). The M6P receptor-lysosomal enzyme complexes are packaged into clathrin-coated vesicles and enter the endocytic route to the lysosome via endosomal and other prelysosomal compartments (Lemansky *et al.*, 1987). Prelysosomal compartments are not well characterized but are thought to be endosome-like vesicles, or a vesicle intermediate between the biosynthetic and endocytic pathways to lysosomes. Under the influence of the low pH within these vesicles (Fig. 1.5, compartment C), the lysosomal enzymes and M6P receptors dissociate. The lysosomal enzymes are transported to the lysosome and the receptors are recycled back to the Golgi complex to participate in further rounds of transport (Kornfeld, 1987).

Once the lysosomal enzymes reach the lysosome, the carbohydrate side chains are trimmed through the action of endoglycosidases, and dephosphorylation of M6P residues occurs. As a result of this dephosphorylation, mature forms of lysosomal enzymes isolated from most tissues are generally found to be poor in M6P residues and thus have a low rate of uptake into cells. This trimming of terminal carbohydrate residues does not appear to be necessary for enzyme activity. The acidic environment

of the lysosome provides the optimal conditions for lysosomal enzyme activity and the localization of their specific substrates (Fig. 1.5, compartment D).

The process described above is the only pathway for lysosomal proteins to enter the lysosome. A fraction of the newly synthesized lysosomal protein may follow the pathway for secretory proteins (Fig. 1.5, compartment B). These secreted lysosomal enzymes may then interact with cation-independent M6P receptors found in clathrin-coated pits on the plasma membrane of cells. When the lysosomal enzymes associate with the receptors, the coated pit is endocytosed and the enzymes follow the endocytic pathway described above to the lysosome. The cation-dependent M6P receptor does not appear to be involved in this endocytotic pathway (Kornfeld, 1987)

In I-cell disease, a phosphotransferase, which is one of the two enzymes responsible for producing the M6P signal on lysosomal enzymes, is deficient (for review, see Nolan and Sly, 1989). Many cell types in I-cell disease, such as hepatocytes, have nearly normal levels of intracellular acid hydrolases. This has suggested that alternate pathways exist for lysosomal enzymes to be transported to the lysosome in these cells that do not require the M6P recognition marker. It has been suggested that these enzymes may be internalized via other systems that recognise asparagine-linked carbohydrates such as galactose, *N*-acetylgalactosamine, mannose and *L*-fucose present on these enzymes (Waheed *et al.*, 1982). Also in I-Cell disease, some enzymes are present in lysosomes at nearly normal levels. Integral membrane proteins such as N-acetyltransferase (Freeman *et al.*, 1983) move through the Golgi system and are transported to the lysosome independent of asparagine linked carbohydrate recognition markers (Barriocanal *et al.*, 1986). Membrane associated lysosomal enzymes such as glucocerebrosidase also do not require M6P phosphorylation to reach the lysosome (Aerts *et al.*, 1988). The pathways these integral membrane lysosomal proteins and membrane associated lysosomal proteins may follow to the lysosome are not characterized. Another lysosomal enzyme, acid phosphatase, is synthesized as a transmembrane protein anchored by a C-terminal domain, transported via the cell surface to the lysosome where the C-terminal domain is cleaved to produce

soluble acid phosphatase (Pohlmann *et al.*, 1988; Peters *et al.*, 1990). Internalization of lysosomal acid phosphatase depends on the presence of a tyrosine residue within the C-terminal domain (Peters *et al.*, 1990).

After the removal of the ER signal peptide, the newly synthesized lysosomal enzymes may be subjected to further limited proteolytic processing during transport to, or more often, in the lysosome. This "maturation" may include removal of N- and C-terminal sequences as well as internal cleavages and further proteolysis. The products of these cleavages may remain associated through disulphide bonds or non-covalent bonds. The function of many of these proteolytic steps is not known but they do not appear to be necessary, except in the case of some lysosomal proteases, for the lysosomal enzyme's activity and are probably a result of the degradative nature of the lysosomal environment (for review, see Neufeld, 1991). Lysosomal enzymes that escape the endocytic pathway to the lysosome are secreted as higher molecular mass precursor forms as they do not undergo the proteolysis and oligosaccharide chain modifications in the lysosome as described above. It has been shown that most lysosomal precursors have normal enzyme activity, for example secreted β-*N*-acetylhexosaminidase A (Haslik *et al.*, 1982). Multiple forms of lysosomal enzymes, which differ in their native protein molecular mass and subunit compositions, have been isolated from the different tissue sources used in their purification (e.g. IDUA, see following section). This, in part, may be a reflection of the abundance and distribution of M6P receptors in different cell types (Brown and Farquhar, 1987) which may affect the rate and the pathway by which lysosomal enzymes are transported to the lysosome.

In the case of LSD, the greater part of most mutant lysosomal enzymes do not proceed to the lysosome at all (e.g. Tager *et al.*, 1984). It has been suggested that amino acid substitutions may result in the incorrect folding of the enzyme, faulty recognition by the signal recognition protein at the ER and thus degradation in the ER (Lodish, 1988).

1.3 MUCOPOLYSACCHARIDOSIS TYPE I

1.3.1 ISOLATION AND MATURATION OF α-L-IDURONIDASE

As described in the previous sections, IDUA hydrolyses the non-reducing terminal α-L-iduronide glycosidic bonds in the glycosaminoglycans HS and DS (Neufeld and Muenzer, 1989; Hopwood, 1989). Largely because a deficiency of IDUA results in MPS-I, several groups have attempted the purification of human IDUA protein. Purification attempts have used a variety of tissue sources with varying low degrees of homogeneity obtained, probably due to the low abundance of IDUA. Barton and Neufeld (1971) were the first to attempt to purify "Hurler corrective factor" which was subsequently shown to be IDUA (Bach *et al.*, 1972, Matalon and Dorfman, 1972). IDUA was purified approximately 1000-fold (estimated to be ~10% pure) from urine, and found to be a protein with an molecular mass of 84 to 87-kDa. Subsequent studies showed that two forms of IDUA existed in urine that differed in the rate of endocytosis into cultured human fibroblasts (Shapiro *et al.*, 1976). The high uptake form was identical to the Hurler corrective factor isolated by Barton and Neufeld (1971) and corresponds to IDUA secreted from cells as described in the previous section. This high uptake form was used to characterize the receptor-mediated uptake of IDUA by cultured fibroblasts (Sando and Neufeld, 1977) which is identical to the M6P receptor-mediated endocytosis described in Section 1.2.2c. The low uptake form of IDUA corresponds to IDUA that has passed through the lysosome and undergone extensive carbohydrate sidechain trimming and proteolytic processing as described in Section 1.2.2c. The carbohydrate sidechain trimming removes many of the M6P signals resulting in its less efficient uptake. The findings of high and low uptake forms of IDUA is a consistent factor in most of the reports of IDUA isolation, as described below.

Rome *et al.* (1978) reported the 25,000-fold purification of IDUA which was estimated to be ~80% pure from human kidney. This low uptake non-corrective form of IDUA was found to be a glycoprotein with a native molecular mass of approximately

60-kDa that reduced to subunits of around 31-kDa by denaturing polyacrylamide gel electrophoresis. Antibodies were raised against this purified IDUA for use in maturation and patient studies as described later in this section (Myerowitz and Neufeld, 1981).

Schuchman *et al.* (1984b) purified equal amounts of both high and low uptake forms of IDUA from human lung. Purification was over 175,000-fold to apparent homogeneity to yield a high and low uptake form with the molecular masses of 85-kDa and 68-kDa respectively. Polyclonal antibodies to the low uptake form of IDUA cross-reacted with the high uptake form (Schuchman *et al.*, 1984b; 1984c). Denaturing polyacrylamide gel electrophoresis under reducing conditions showed that both the high and low uptake forms contained three major polypeptides of around 75 to 72-kDa, 67-kDa, 48 to 43-kDa and 40 to 37-kDa. Consistent with the denaturing polyacrylamide gel electrophoresis, high performance liquid chromatography resolved both the high and low uptake forms of IDUA into several components. It was reported that all the polypeptide components of IDUA maintained some degree of catalytic activity and cross-reactivity with the polyclonal antibodies after separation by high performance liquid chromatography. These results are consistent with the results of other groups reported below but, with hindsight, several factors put the absolute purity of this preparation in doubt as discussed in detail in Section 3.5.2b of this thesis.

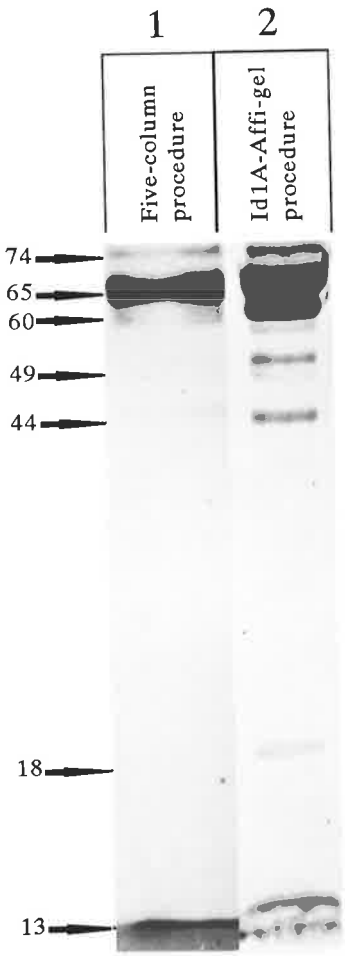
Clements *et al.* (1985a) reported the 20,000-fold purification of human liver IDUA with a molecular mass of 65-kDa by a five-column procedure. This purified IDUA was shown to be a low uptake form. Silver staining and denaturing polyacrylamide gel electrophoresis under reducing conditions revealed the presence of minor species or contaminants of 60-kDa, 54-kDa and other minor bands. Monoclonal antibodies were raised against this preparation of human liver IDUA and used to immunoprecipitate radiolabelled IDUA which revealed bands of 65-kDa and 14-kDa. Both a slight modification of this five-column procedure and the use of a monoclonal antibody immunoaffinity column resulted in the 172,000-fold purification of human liver IDUA (Clements *et al.*, 1989). Seven major bands of 74, 65, 60, 49, 44, 18 and

13-kDa were identified on denaturing and reducing polyacrylamide gel electrophoresis of both preparations (Fig. 1.6). IDUA was also isolated using an immunoaffinity column from human kidney and lung. Two different forms of IDUA from all three tissues were observed with reference to their chromatographic behaviour on the immunoaffinity column. Both forms of IDUA from all three tissues were composed of different proportions of the seven polypeptide species mentioned above. Form II of IDUA from all tissue sources was comprised of a higher proportion of the large polypeptide components of immunopurified human liver IDUA. The lung form II of human IDUA, which had the highest proportion of high molecular mass components, had the highest uptake into cultured MPS-I fibroblasts suggesting that these larger polypeptides may be forms of IDUA that had passed through the secretory pathway and still retained their M6P targeting signals. Both forms of liver IDUA, which was composed mainly of the lower molecular mass polypeptides forms of IDUA, had poor uptake into fibroblasts. Clements *et al.* (1989) speculated that the separation of IDUA into two forms by immunoaffinity chromatography was due to competition between affinity and self-association equilibria. The 172,000-fold purification of human liver IDUA protein with 80-95% yield using an immunoaffinity column provided relatively large amounts of homogeneous IDUA for extensive peptide sequencing (Clements *et al.*, 1989) which was the starting point for the work presented in this thesis. Initial amino acid sequencing revealed that the 65 and 60-kDa polypeptides shared a common N-terminal amino acid sequence and the 49 and 44-kDa polypeptides shared another N-terminal amino acid sequence, which was different to that of the 65/60-kDa N-terminus. The N-terminal amino acid sequence of the 13-kDa band was also determined and thought to show that the 13-kDa band contained two different polypeptide species. Western blotting with two monoclonal antibodies indicated that the two different monoclonal antibody epitopes were present on different polypeptide species of immunopurified IDUA. A complex model was proposed to account for all seven major polypeptides present in immunopurified IDUA being derived from a single gene product (Fig. 1.7).

FIGURE 1.6. PURIFIED IDUA.

(From Clements *et al.*, 1989)

SDS/ polyacrylamide gel showing a comparison of the polypeptides present in human liver α-L-iduronidase purified by a five column procedure (lane 1) and immunopurified using a monoclonal antibody column (lane 2). The molecular mass of the polypeptides is indicated in kDa on the left.


1
2
Five-column procedure
Id1A-Affi-gel procedure
74
65
60
49
44
18
13

As discussed in Section 1.2.1, it had been hypothesized that there were two IDUA isoenzymes specific for HS and DS (Matalon and Deanching, 1977). Reports have suggested that IDUA from human fibroblasts could be separated into two forms; a form that degrades only HS and a form that degrades only DS. HS and DS specific forms of IDUA were reported to be isolated from cultured normal human skin fibroblasts (Minami *et al.*, 1984; Fujibayashi *et al.*, 1984). Clements *et al.* (1985a) speculated that these results may be artefactual due to the experimental conditions used and neither of the two forms of human liver IDUA isolated by Clements *et al.* (1989) showed a difference in activity towards HS or DS derived substrates. Freeman and Hopwood (1992a) attempted to emulate the results of Fujibayashi *et al.* (1984) and also found two forms of IDUA in normal human fibroblasts. However both forms were found to have a similar ability to degrade HS derived substrates.

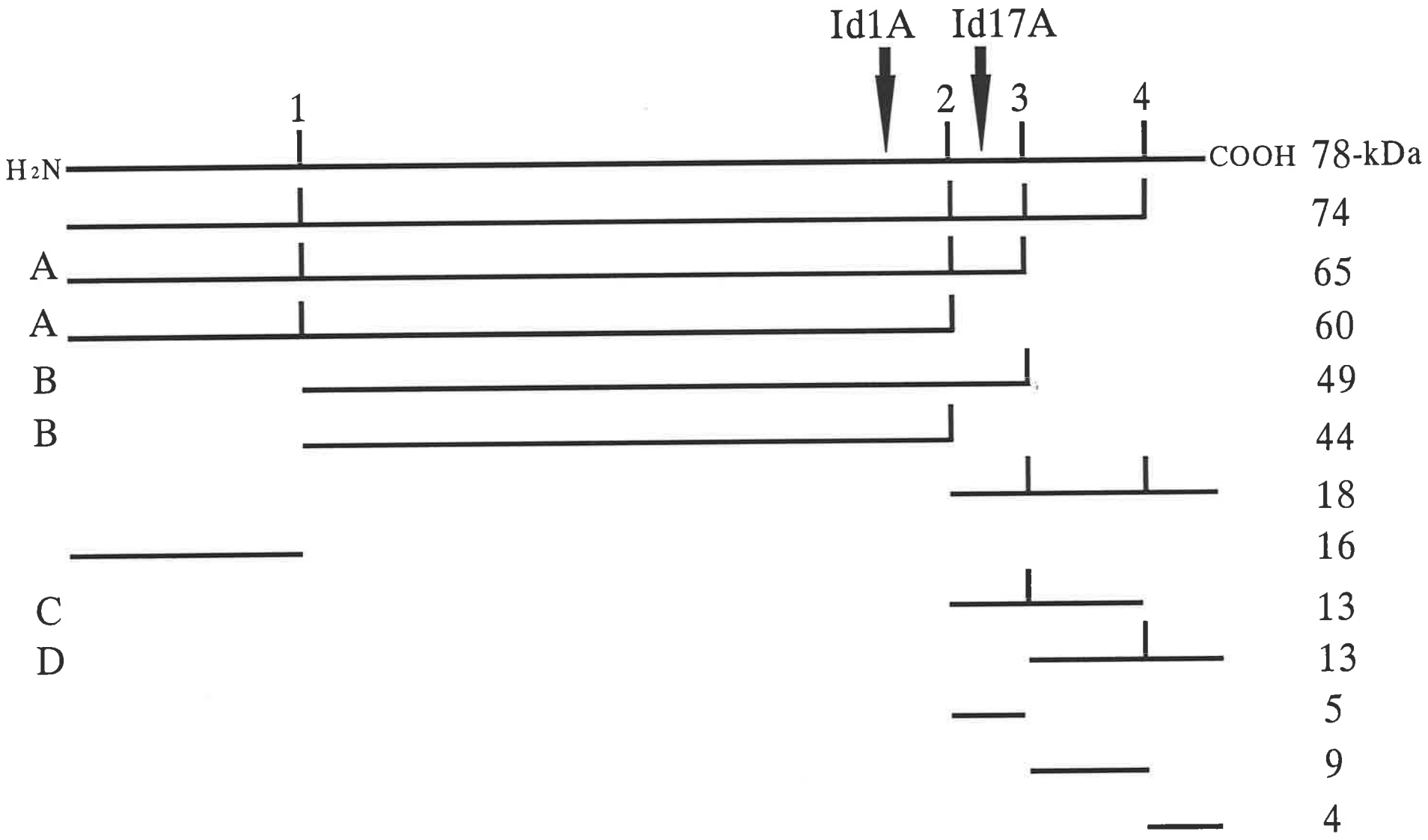
IDUA purified from each source described above has differed both in native protein and subunit molecular mass. In addition, different catalytic properties (pH optima, K_m and V_{max}) have been reported for these purified IDUA preparations and IDUA in leukocyte and cultured skin fibroblast homogenates. These differences are presumably due to many factors such as the tissue source, the method of preparation, the presence of contaminating or modifying proteins and different assay protocols used in each laboratory (Freeman and Hopwood, 1992a). Freeman and Hopwood (1992a) also showed that determination of Michaelis-Menton constants was dependant on the concentration of the IDUA substrates used. Some of the polypeptide species found in the IDUA preparations described above may be due to *in vivo* proteolytic processing and others may be due to *in vitro* degradation and may be specific to the methodology used.

Two studies on the maturation of IDUA in normal cultured fibroblasts have produced consistent results. Myerowitz and Neufeld (1981) used polyclonal antibodies against purified human kidney IDUA (Rome *et al.*, 1978) to immunoprecipitate radiolabelled intracellular IDUA species of 75, 72 and 66-kDa. The identity of these bands as IDUA was confirmed by their absence from MPS-IH patient fibroblasts. The

FIGURE 1.7. PROPOSED MODEL OF IDUA PROTEOLYTIC PROCESSING.

(Adapted from Clements *et al.*, 1989)

The proposed 78-kDa precursor may be processed by different proteolytic events at sites 1, 2, 3 and 4 indicated by the vertical lines. The hypothesized positions of the IdIA and Id17A monoclonal antibody epitopes are arrowed. Polypeptides may be further processed to produce smaller IDUA polypeptides at the proteolytic sites marked. For example, the 18-kDa component can be clipped at site 3 to produce a 13-kDa polypeptide with the Id17A epitope and a 5-kDa component that is either lost from the system or not resolvable by SDS/polyacrylamide gel electrophoresis.


Id1A
Id17A
1
2
3
4
H2N
COOH
78-kDa
74
A
65
A
60
B
49
B
44
18
16
C
13
D
13
5
9
4

75-kDa polypeptide was a precursor that was trimmed to the 72-kDa form within a few hours and then to the 66-kDa form over a period of 4 to 5 days. A 76-kDa form was secreted into the culture media in small amounts and was readily taken up by fibroblasts via M6P receptors. In contrast, the smaller species were low uptake forms of IDUA. Similarly, Taylor *et al.* (1991) used a monoclonal antibody against human liver IDUA to isolate radiolabelled IDUA from cultured fibroblasts by immunoaffinity chromatography. An intracellular 81-kDa IDUA precursor was processed within 24 hours to 69-kDa via intermediates of 76 and 70-kDa. The 69-kDa form appeared to be stable for at least 5 days in culture. Other polypeptides of 59 and 44-kDa were also observed frequently but inconsistently. Again, a small proportion of IDUA (82-kDa) was secreted into the culture medium. The differences in the molecular mass of IDUA species in these two studies are probably due to variation in analytical techniques, as the IDUA species observed and the molecular mass changes involved in processing IDUA from a precursor to a "mature" form via intermediates are directly comparable (Taylor *et al.*, 1991). The major IDUA polypeptide species purified from human liver was 65-kDa (Clements *et al.*, 1985a; 1989) while the major IDUA polypeptide species present in human fibroblasts was 69-kDa (Taylor *et al.*, 1991) and it is thought that these may represent the same polypeptide with minor differences in proteolytic and carbohydrate processing. The high proportion of lower molecular mass species in immunopurified human liver IDUA may reflect continued slow proteolytic maturation, tissue specific proteolytic maturation or proteolysis that has occurred during the purification procedure (Taylor *et al.*, 1991).

1.3.2 CLINICAL DESCRIPTIONS OF MPS-I

As described in Section 1.2.2, a deficiency of IDUA in humans results in the lysosomal storage disorder MPS-I (eponyms; Hurler, Hurler/Scheie and Scheie syndromes) which is inherited as an autosomal recessive disease. MPS-I has acted as the prototype for the clinical classification of the MPS, however most MPS-I patients are more severely affected than patients with other MPS disorders. A deficiency of IDUA can result in a wide spectrum of phenotypes, ranging from Hurler syndrome

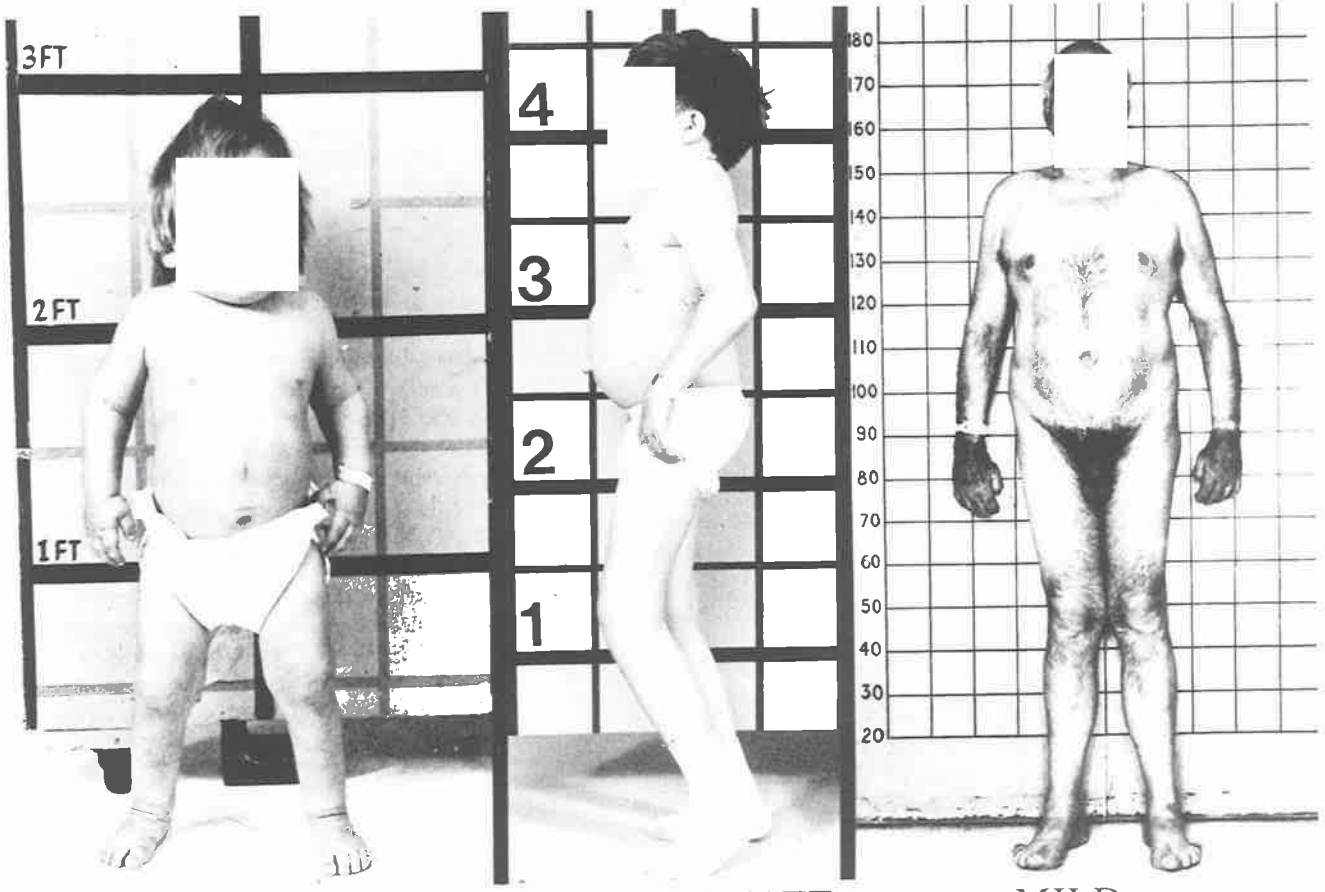
(MPS-IH) at the severe end of the clinical spectrum to Scheie syndrome (MPS-IS) at the mild end, with Hurler/Scheie syndrome (MPS-IH/S) representing intermediate phenotypes (see McKusick, 1972; McKusick and Neufeld, 1983; Neufeld and Muenzer, 1989; Hopwood and Morris, 1990). The clinical phenotypes are not distinguishable biochemically (see Section 1.3.4 for further discussion) as all MPS-I patients have mucopolysaccharidurıa and low IDUA activity and thus any classifications of MPS-I patients into these subgroups is made solely on clinical grounds.

MPS-IH is probably the earliest presenting of the ten MPS and is illustrated by the severe patient in Fig. 1.8. Diagnosis is commonly made between 6 and 10 months of age with the patient normally presenting with a combination of the following: enlarged liver and spleen, skeletal deformities, joint stiffness, coarse facial features, enlarged tongue, prominent forehead, nasal discharge and repeated upper respiratory or ear infection. Infants usually appear normal at birth except perhaps for inguinal or umbilical hernias. The infant may even be considered large in the first year of life but patients will fail to thrive and then progressively deteriorate, both mentally and physically. Delay of growth and mental development may be noticed between 6 to 18 months and 12 to 24 months respectively. Mental retardation is conspicuous in MPS-IH with the maximum functional age obtained by most patients being between 2 and 4 years of age followed by progressive deterioration. Severely affected patients develop limited language skills, probably due to developmental delay, hearing loss and an enlarged tongue. The hearing loss is probably due to a combination of conductive and neurosensory problems. Hydrocephalus normally develops between 2 to 3 years of age but may be seen as early as 6 months, and the intracranial pressure may be relieved by shunting. In contrast to MPS-VI, MPS-I patients that have received a shunt still develop severe neurological symptoms presumably due to GAG accumulation in the CNS. One of the most obvious clinical features that distinguishes MPS-I from MPS-II is the development of corneal clouding from approximately 12 months of age. Glaucoma is also often associated with MPS-I. The liver and spleen are enlarged due to the deposit of mucopolysaccharide in hepatocytes.

FIGURE 1.8. RANGE OF MPS-I CLINICAL PHENOTYPES.

(From Neufeld and Muenzer, 1989 and McKusick and Neufeld, 1983)

Photographs of three patients that are representative of the range of clinical phenotypes seen in MPS-I. The severe patient would previously have been classified as Hurler syndrome and was 4 years old at the time of the picture. The intermediate patient would previously have been classified as Hurler-Scheie syndrome and was 14.5 years old at the time of the picture. He graduated from high school and, at age 19, was planning to attend college. The mild patient would previously have been classified as Scheie syndrome and was 54 years old at the time of the picture and a practising attorney. For a detailed clinical description of the mild patient see McKusick *et al.*, (1965).



SEVERE INTERMEDIATE MILD

The radiological changes seen in MPS-IH are similar to those seen in the other MPS with skeletal involvement such as MPS-VI. These skeletal abnormalities, or dysostosis multiplex, are mainly due to growth plate deformities, presumably caused by the lack of turnover of GAG which is an integral part of the extracellular matrix in the growth plates (Silveri *et al.*, 1991). The diaphyses or shafts of the long bones are enlarged with irregular appearances of the metaphyses and poorly developed epiphyseal centres. Clavicles are short, thickened and irregular. The ribs are narrow at the vertebral ends and broad at the sternal ends resulting in a spatulate or oar-like shape. There is often flaring of the lower ribcage, probably due in part to the hepatosplenomegaly. Anterior hypoplasia of lumbar and/or thoracic vertebrae develop early leading to lumbar or thoracic kyphosis giving the spine an anterior hooklike projection. This may also result in a very short neck, giving the appearance that the head is resting directly on the thorax. Carpal tunnel syndrome and distinctive clawing of the hands is observed due to shortening of the tendons and ligaments surrounding the joints and shortened phalanges with widened shafts and narrow ends. The limited mobility in most joints is thought to be caused by a combination of metaphyseal deformities and thickened joint capsules due to GAG accumulation and fibrosis. In general, these growth plate deformities lead to dwarfism. The skull is large with thickened calvarium, premature closure of sutures, shallow orbits, enlarged J-shaped sella, and widely spaced teeth with dentigerous cysts. These skull abnormalities lead to narrowing of the nasopharyngeal airway and a mass of adenoid tissue may block the remaining airway resulting in noisy breathing. These upper airway abnormalities, together with abnormality of the tracheobronchial cartilages, are responsible for obstructive airway disease and an increased susceptibility to respiratory infection. Deposition of mucopolysaccharide in arteries and in heart valves can often result in heart murmurs and more severe cardiac disease. Obstructive airway disease, respiratory infection and cardiac complications normally result in death before 10 years of age.

MPS-IH/S is used to describe patients with an intermediate clinical phenotype (Fig. 1.8). It is characterized by little or no neurological involvement but progressive somatic disease including dysostosis multiplex. Most of the symptoms described above

for MPS-IH may develop by the early to mid-teens and cause considerable limitation and loss of function, mainly in mobility. Symptoms are first noticed between 3 and 8 years of age and patients often survive to adulthood.

At the mild end of the MPS-I clinical spectrum, MPS-IS patients arc characterized by joint stiffness, aortic valve disease, corneal clouding, mild somatic changes and little or no neurological involvement. Indeed, the mild patient in Fig. 1.8 was class valedictorian in high school and a practising attorney later in life (McKusick, 1972). Symptoms may first be noticed around 5 years of age but diagnosis may not be made until the second decade of life and patients may have a normal stature and life span. The characteristic clawed hands and coarse facial features of MPS-I may be evident. Mild hepatosplenomegaly and hearing loss may develop and the intellect may deteriorate late in life. Psychiatric disturbances have been noted in several MPS-IS patients although its relationship to the metabolic disorder is unknown (McKusick, 1972).

In addition there have been a number of patients described who have unusual clinical phenotypes and thus do not fit into any of these classifications (Roubicek *et al.*, 1985; Colavita *et al.*, 1986; Jellinger *et al.*, 1990). This reflects the fact that the classifications are based on the historical expectations of the genetics of MPS-I as described in the following section.

1.3.3 GENETICS OF MPS-I

The incidence of the LSD has been estimated at 1/2,500 (Scriver *et al.*, 1989) and that of the MPS between 1/10,000 (Hopwood and Morris, 1990) and 1/25,000 births (Scriver *et al.*, 1989). MPS-I seems to be the most common of the MPS in most populations. The frequency of MPS-IH and MPS-IS in British Columbia has been estimated at 1/100,000 and 1/500,000 births respectively (Lowry and Renwick, 1971) but it is thought that these values are an underestimate as complete ascertainment, particularly of mild or variant types of MPS-I, is unlikely to have been achieved in the survey period of 1952 to 1968 (Neufeld and Muenzer, 1989). Jervis (1950) reviewed

families with MPS-IH-affected children where cousin marriages were known to occur in 11 of the 103 families and calculated that 10% cosanguinity corresponded to a frequency of 1/40,000. MPS-I has also been reported in a number of different races such as Negroes, oriental Indians, Chinese and Japanese (for review, see McKusick *et al.*, 1978).

MPS-I is inherited as an autosomal recessive genetic disease (Halperin and Curtis, 1942). Like many rare autosomal recessive genetic diseases, parental cosanguinity and affliction of multiple sibs of both sexes is often observed. Unlike the X-linked MPS-II, there are generally no reports of affected individuals in preceding generations of the families. There is no evidence to suggest partial MPS-I phenotype expression in heterozygotes. Patients with MPS-I, like all other mucopolysaccharidoses, may present within a broad spectrum of clinical phenotype as described above. As described in Section 1.2.1, before the enzyme defect in MPS-I was identified, two separate syndromes of varying clinical severity were classified, i.e. Hurler syndrome (clinically severe) and Scheie syndrome (mild; previously classified as MPS-V) (McKusick and Neufeld, 1983). Following the discovery that Hurler and Scheie syndromes were apparently the result of the same biochemical defect (Wiesmann and Neufeld, 1970) and, following this, that a deficiency of IDUA was responsible for both syndromes (Bach *et al.*, 1972; Matalon and Dorfman, 1972), it was believed that there were two different mutant alleles at the IDUA locus, i.e. Hurler patients were homozygous for the severe allele and Scheie patients were homozygous for the mild allele. By analogy to the haemoglobinopathies such as sickle-cell anaemia, it was suggested that MPS-I patients that were compound heterozygous for the Hurler and Scheie alleles would have an intermediate form of MPS-I (McKusick, 1972; McKusick *et al.*, 1972; McKusick and Neufeld, 1983). The frequencies of the MPS-IH and MPS-IS alleles can be calculated from the population frequencies of the diseases given above to be approximately 1/330 and 1/700 respectively. Thus the frequency of a genetic compound for MPS-IH and MPS-IS should be approximately 1/115,000. Patients with a unique MPS of intermediate severity were sought and, not surprisingly, a number of patients were found who fulfilled the predictions for a genetic compound

and were classified as Hurler/Scheie syndrome (McKusick, 1972; McKusick *et al.*, 1972; McKusick and Neufeld, 1983). Most MPS-I patients have a severe clinical phenotype, thus not fitting the predicted frequency of MPS-IH/S. Also, the appearance of so-called Hurler/Scheie patients as the result of consanguineous marriages (Jensen *et al.*, 1978; Kaibara *et al.*, 1979; Kaibara *et al.*, 1983) and patients with unusual clinical phenotypes which fit into any of the three clinical classifications (Roubicek *et al.*, 1985; Colavita *et al.*, 1986; Jellinger *et al.*, 1990) do not support the hypothesis of only two mutant IDUA alleles. Studies of the molecular genetics of other genetic diseases has generally seen great heterogeneity in the mutations present (Scriver *et al.*, 1989). This has led to the more recent proposal that the broad spectrum of clinical phenotype seen in MPS-I patients is generated by the presence of multiple IDUA mutant alleles (Neufeld and Muenzer, 1989; Hopwood and Morris, 1990).

1.3.4 DIAGNOSIS OF MPS-I

MPS-I is normally diagnosed by clinical features and confirmed by the presence of mucopolysaccharidurıa and the absence of IDUA activity in patient leukocytes and fibroblasts. Screening for mucopolysaccharidurıa has been the basis of diagnostic testing for the MPSs since Dorfman and Lorincz (1957) reported the increased urinary excretion of partially degraded mucopolysaccharides in MPS patients. Urinary GAG analysis can usually discriminate between classes of GAG such as HS, DS, keratan sulphate and chondroitin sulphate. Screening for urinary GAG excretion is an inexpensive, quick and useful strategy for preliminary evaluation, but some methods may have a high percentage of false negatives and positives (for review, see Neufeld and Muenzer, 1989; Hopwood and Morris, 1990). Hopwood and Harrison (1982) developed a simple high-resolution electrophoresis screening test for mucopolysaccharidurıa that, with care, can distinguish between broad groups of GAG and appears not to give false negative or positive results. Many laboratories prefer only to use definitive specific enzyme assays in cultured fibroblasts, leukocytes, serum or plasma for diagnosis that must be used for confirmation of the specific enzyme defects. Specific enzyme assays may use fluorogenic, chromogenic or radiolabelled substrates

derived from the stored GAG. Prenatal diagnosis is possible for all the MPS and the enzyme assay may be performed on homogenates prepared from cells grown from amniotic fluid or chorionic villus (Neufeld and Muenzer, 1989; Hopwood and Morris, 1990). Direct enzyme assay of chorionic villus may also be possible.

As IDUA is required for the degradation of both HS and DS, MPS-I patients accumulate both HS and DS fragments with α-L-iduronide non-reducing termini. Patient pathophysiology has been suggested to be related to type and level of tissue storage of GAG (Conzelmann and Sandhoff, 1984). Levels of GAG excretion appears to correspond to clinical severity although this does not appear to be a reliable diagnostic marker (Hopwood and Morris, 1990). MPS-I patients with mental retardation (MPS-IH) have been shown to excrete larger amounts of urinary HS than patients without mental retardation (MPS-IS and MPS-IH/S; Matalon *et al.*, 1983; Minami *et al.*, 1984). MPS-IS and MPS-IH/S patient fibroblasts were shown to have IDUA activity that could degrade HS but not DS. It was hypothesized that this was the result of IDUA mutations, thus explaining why MPS-IS and MPS-IH/S patients had minimal CNS involvement (Matalon *et al.*, 1983).

IDUA activity is diagnostically assessed using either the fluorogenic substrate 4-methylumbelliferyl-α-L-iduronide (Hopwood *et al.*, 1979b) or radiolabelled disaccharide substrates derived from heparin or HS (Hopwood *et al.*, 1979a; Hopwood and Muller, 1982; Clements *et al.*, 1985b) or DS (Muller and Hopwood, 1984). Attempts to distinguish between the extremes of MPS-I clinical phenotypes by biochemical techniques (Hopwood and Muller, 1979; Ullrich *et al.*, 1981; Matalon *et al.*, 1983; Muller and Hopwood, 1984) and immunological techniques (Myerowitz and Neufeld, 1981; Schuchman and Desnick, 1988; Taylor *et al.*, 1991; Ashton *et al.*, 1992) have only been partially successful. Differences have been observed in the kinetics of residual IDUA in patient fibroblasts from either end of the MPS-I clinical spectrum, but patients of intermediate clinical severity may not be discriminated (Hopwood and Muller, 1979, Ullrich *et al.*, 1981, Muller and Hopwood, 1984). In general, fibroblasts from MPS-IH patients do not have immunodetectable IDUA protein. With some

exceptions, mild or intermediate MPS-I patients have been found to have immunodetectable IDUA protein in fibroblasts (Ashton *et al.*, 1992), which makes the use of immunodetection of IDUA for prediction of the severity of MPS-I patient phenotype dangerous. Carrier detection for MPS-I using enzyme assays is not possible because of the wide range of IDUA activity in normal individuals (Hopwood *et al.*, 1979a; Hopwood *et al.*, 1979b; Hopwood and Muller, 1979; Ullrich *et al.*, 1981; Matalon *et al.*, 1983; Muller and Hopwood, 1984; Gatti *et al.*, 1985; Whitley *et al.*, 1987; Taylor *et al.*, 1991; Ashton *et al.*, 1992). Recombinant DNA methods are considered unlikely to replace the current enzyme assays in the near future except for carrier detection through linkage analysis and direct gene analysis and for prenatal diagnosis (Hopwood and Morris, 1990).

1.4 TREATMENT OF MPS-I

MPS-I, like the other MPS, is a chronic and progressive disease. No specific treatments are available for the MPS disorders. Management of MPS patients consists of supportive care and treatment of complications (for review, see Neufeld and Muenzer, 1989). A brief discussion of symptomatic treatment and each of the major therapeutic approaches for MPS-I will be discussed below.

1.4.1 CLINICAL MANAGEMENT

As discussed previously, hydrocephalus is common in MPS-I patients and shunting to relieve increased CNS pressure may be palliative (Neufeld and Muenzer, 1989). The corneal clouding seen in MPS-I and several of the other MPS can lead to significant visual disability. Corneal transplants have been performed on a number of patients but the long-term outcome has not always been successful (Orgul *et al.*, 1991). Hearing aids may be of some help with the hearing loss associated with MPS-I (Neufeld and Muenzer, 1989; Ruckenstein *et al.*, 1991). Physiotherapy has often been used to help maintain joint mobility although its benefit may be minimal (Neufeld and Muenzer, 1989). Surgical relief of carpal tunnel syndrome in the clawed hands may be necessary. The obstructive airway disease in MPS-I can lead to sleep apnea which may

be relieved by tracheostomy. Tonsillectomy and adenoidectomy are also frequently performed to decrease airway obstruction (Neufeld and Muenzer, 1989; Ruckenstein *et al.*, 1991). The obstructive airway disease in MPS-I can lead to major risks in general anaesthesia (Sjogren *et al.*, 1987). Valve replacement to overcome valve thickening has been reported in MPS-IS and MPS-IH/S, although experience is limited (Pyeritz, 1983).

1.4.2 ENZYME REPLACEMENT THERAPY

The concept of enzyme replacement therapy for inborn errors of metabolism has been in existence for many years (Baudhiun *et al.*, 1964). Basically, it involves infusion of the missing enzyme into the circulation of a patient with a genetic defect. The finding that the metabolic defect in cultured MPS patient fibroblasts was easily corrected by the addition of "corrective factors", or the missing lysosomal enzymes, stimulated hope for this concept for the MPS (Fratantoni *et al.*, 1969). The recognition that high uptake forms of lysosomal enzymes have the M6P recognition markers (see Section 1.2.2c) meant that enzymes prepared for enzyme replacement therapy not only had to be highly active, pure and nonantigenic, but must bear these M6P recognition markers. As discussed in Section 1.3.1, purification of IDUA, and most other lysosomal enzymes, was difficult to achieve due to their low abundance and the loss of recognition markers either before or during purification.

The first attempts at enzyme therapy by-passed purification. Before the enzyme defects in the MPS were defined, DiFerrante and colleagues infused large amounts of plasma and leukocytes into patients with MPS-I and MPS-II in an attempt to induce degradation of GAG (DiFerrante *et al.*, 1971; Knudson *et al.*, 1971). First results appeared to be encouraging with improved joint mobility. However, these benefits, if any, appeared to be transient and with the advent of specific enzyme assays and the understanding of the M6P recognition markers, it was realised that the comparatively small amount of enzyme being infused did not contain the relevant markers for receptor-mediated endocytosis into cells. Skin grafts in MPS-II led to graft rejection problems and transplants of cultured skin fibroblasts did not alter the clinical course of

the disease. Amnion, a non-immunogenic tissue rich in lysosomal enzymes, was also transplanted with no observable biochemical or clinical improvement (for review, see Neufeld and Muenzer, 1989).

1.4.3 BONE MARROW TRANSPLANTATION

Partly because of the failure of enzyme replacement therapy described above, alternative therapy protocols were pursued for the MPS including allogeneic bone marrow transplantation (BMT). Although the MPS do not specifically involve the hematopoietic system, it was hypothesized that circulating cells, derived from the normal donor hematopoietic progenitor cells (stem cells), could donate the therapeutic enzyme to deficient cells of all tissues through secretion and/or cell to cell interaction. Hobbs *et al.* (1981) were the first to report the apparent success of allogeneic BMT in treating MPS-I. Both biochemical and clinical benefits were observed. Circulating IDUA was measured and the disease progression was reported to be halted, with clearing of the corneas and a decrease in hepatosplenomegaly. To date, a total of at least 38 MPS-I patients have undergone BMT (Hoogerbrugge and Vossen, 1990) and have been followed over several years. Consistent changes include reduction of urinary GAG excretion towards normal levels, a decrease in hepatosplenomegaly with corresponding reduction of lysosomal inclusion bodies in hepatocytes, clearing of corneas and increased joint mobility. It was generally recognised that skeletal abnormalities were not prevented. The lack of a diagnostic test that can be used to give a prognostic outlook for MPS-I patients has lead to controversy about the level of improvement in neurological function in bone marrow transplanted MPS-I patients. The key question was whether those bone marrow transplanted MPS-I patients that maintained a mild phenotype may have had that mild phenotype without BMT (for review, see Neufeld and Muenzer, 1989; Hoogerbrugge and Vossen, 1990; Hopwood and Morris, 1990).

The existence of animal models for some of the MPS has been useful in evaluating allogeneic BMT, as experimental controls that are not possible in human

studies may be performed (canine MPS-I, Shull *et al.*, 1982; Spellacy *et al.*, 1983; feline MPS-VI, Jezyk *et al.*, 1977). Long term BMT in canine MPS-I appears to relieve most of the somatic disease symptoms observed with a corresponding decrease in GAG storage, although development of MPS-I bone pathology and GAG deposit in the cardiovascular system is not completely prevented (Breider *et al.*, 1989; Gompf *et al.*, 1990). Similar alleviation of somatic symptoms has been reported to result from autologous BMT of MPS-VI cats (Gasper *et al.*, 1984). It is difficult to assess whether or not the neurological symptoms of canine MPS-I have been reduced, although IDUA enzyme activity is found in the brain tissue and cerebrospinal fluid of long term BMT MPS-I dogs, resulting in a decrease in the GAG level in those tissues (Shull *et al.*, 1987; 1988). In a canine model of another LSD, fucosidosis (Kelly *et al.*, 1983), it has been demonstrated that allogeneic BMT can be successful in preventing the neurological symptoms of that disease (Taylor *et al.*, 1989; 1992). The therapeutic enzyme is presumably transported across the blood-brain barrier by donor bone marrow-derived macrophages (Hoogerbrugge *et al.*, 1988). However, the BMT must be performed as early as possible to prevent the onset of symptoms (Taylor *et al.*, 1989; 1992). MPS-VII, caused by a deficiency of β-D-glucuronidase which is required for the degradation of both HS and DS, is clinically similar to MPS-I with both skeletal and neurological symptoms. A mouse model for MPS-VII has symptoms similar to those found in severe human MPS-VII patients (Birkenmeier *et al.*, 1989). BMT of an MPS-VII mouse model showed essentially complete correction of spleen, liver, cornea, and glomerular mesangial cells. Storage was partially corrected in meninges and perivascular cells in brain, and in renal tubular epithelial cells. The life span in BMT-treated animals was increased approximately three-fold, approaching that seen in normal mice after BMT (Birkenmeier *et al.*, 1991).

In spite of the apparent success of BMT in treating MPS-I, allogeneic BMT using immunologically-matched donors still carries a high risk of mortality (>30%) and morbidity mainly due to graft-versus-host disease (Neufeld and Muenzer, 1989; Hoogerbrugge and Vossen, 1990) and must still be considered an experimental treatment protocol.

1.4.4 FUTURE PROSPECTS FOR PATIENT THERAPY

The major problem with the past approaches to enzyme replacement therapy for the MPS as discussed above, have been the lack of sufficient enzyme bearing markers for endocytosis. Within the next few years, it is expected that sufficient amounts of therapeutic human enzymes can be produced by recombinant DNA technology for enzyme replacement therapy to become an economically viable option. The complex oligosaccharide side chains of human glucocerebrosidase enzyme have been modified to expose terminal mannose residues allowing mannose receptor-mediated endocytosis for successful enzyme replacement therapy in the LSD, Gaucher's disease (Barton *et al.*, 1990; for review, see Beutler, 1991). Large quantities of recombinant lysosomal enzymes, such as IDUA, may be genetically, chemically or enzymatically modified *in vitro* such that oligosaccharide chains terminate with mannose residues to facilitate enzyme replacement therapy. However, a number of the MPS, including MPS-I, involve CNS deterioration where the presence of the blood-brain barrier is predicted to prevent transport of the therapeutic enzyme to the site of pathology, and thus enzyme replacement therapy is predicted to be of little use in preventing CNS deterioration.

Like enzyme replacement therapy, the concept of gene therapy for genetic diseases has also been considered for the treatment of genetic diseases (Friedmann and Robin, 1972) but technological advances are only just beginning to make it a viable proposition (Friedmann, 1989). The greatest significance of the limited success of BMT in treating MPS-I is that it encourages the prospect of gene replacement therapy through genetic modification of hematopoietic progenitor cells. This is the favoured approach as the genetically engineered stem cells can proliferate indefinitely and colonize the host. The fact that macrophages also appear to cross the blood-brain barrier makes this approach ideal for the treatment of diseases involving CNS pathology (Hoogerbrugge *et al.*, 1988). Insertion of the normal gene into the patient's own stem cells followed by autologous transplantation would be the equivalent of BMT as described above, without the need to completely eradicate the host's original stem cell population, and therefore

considerably reducing the risks involved, particularly of graft-versus-host disease. An in depth discussion of the prospects of BMT, enzyme replacement therapy and gene replacement therapy in treating MPS-I will be given in Chapter 5 of this thesis.

1.5 GENOTYPE TO PHENOTYPE CORRELATIONS

A major objective of mutation analysis for all genetic diseases has been to correlate the nature of mutations with the clinical manifestations of the disease. This will be invaluable for phenotype prediction, genetic counselling of carriers and assessment of experimental treatment protocols. The following section will discuss the theory of genotype to phenotype correlations, some examples of genotype to phenotype correlations and the prospect of genotype to phenotype correlation for MPS-I. This section is deliberately confined to what was known before or at the start the work described in this thesis.

As discussed in the previous sections, all the MPS, and indeed many LSD, show a wide spectrum of clinical phenotypes and it has been proposed multiple allelism at each enzyme locus is responsible for this spectrum of clinical variability (Neufeld and Muenzer, 1989). In general for the LSD, it has been found that the more profound the enzyme deficiency, the more profound the disease symptoms. Conversely, the higher the residual enzyme activity, the greater the protection against disease symptoms. Fibroblasts from the mild MPS-IS patient in Fig. 1.8 have no immunodetectable IDUA protein (cell line number 2474, Ashton *et al.*, 1992) with levels of total IDUA activity similar to severely affected MPS-IH patients, but in contrast to the MPS-IH patients, a near normal K_m (Hopwood and Muller, 1979; Muller and Hopwood, 1984; Ashton *et al.*, 1992). This demonstrates that the level of enzyme activity needed to protect against disease manifestation may only be a small fraction of the normal levels. Obviously, a very small difference in the properties and cellular location of residual enzyme activity may dramatically affect the age of onset, the rate of progression and severity of a disease. Conzelmann and Sandhoff (1984) proposed a kinetic model to explain this

observation for lysosomal storage diseases. Basically, they proposed that a "critical threshold" of activity exists, above which the residual enzyme activity can keep up with substrate influx into the lysosome and below which an accumulation of substrate occurs. A small variation in the residual activity near this threshold can have a large influence on the rate and level of storage of substrate and thus on the development of the disease.

Often, it is possible to tell if accurate genotype to phenotype correlations will be possible for a certain genetic disease before mutations are defined by comparing the clinical phenotypes of patients from the same family, as they may be assumed to have the same genotype. An archetype of poor genotype to phenotype correlations is seen in the peroxisomal disorder adrenoleukodystrophy. Adrenoleukodystrophy and adrenomyeloneuropathy are severe and mild X-linked recessive motor-sensory neuropathies respectively. Repeated occurrences of both these diseases in the same families has led to the understanding that they are clinical variants of the same disease and that these clinical variations occur despite the presence of the same mutation. This disease represents the worst extreme in genotype to phenotype correlation and it has been suggested that the variation between these two disorders could be due to the presence of a modifier gene (Davis *et al.*, 1979).

The LSD Tay-Sachs disease, which has a high incidence in Ashkenazi Jews, is caused by mutations in the α-subunit of β-hexosaminidase resulting in lysosomal storage of G_{M2}-gangliosides (for review, see Sandhoff *et al.*, 1989). A 4-bp insertion that results in a stop codon and early termination of translation accounts for approximately 70% of mutant alleles in Ashkenazi Jews. This mutation causes infantile Tay-Sachs disease in homozygotes (Myerowitz and Costigan, 1988). Another mutation in the α-subunit of β-hexosaminidase, an amino acid substitution of glycine at codon position 269 for a serine ($G_{269}S$), accounts for 3% of mutations in Ashkenazi Jews. The $G_{269}S$ mutation causes either the adult (mild) or chronic (intermediate to mild) form of Tay-Sachs disease in either homozygotes or heterozygotes and thus this mutation must allow enough residual enzyme activity to protect against the

development of the most severe clinical symptoms of the disease (Navon and Proia, 1989). In general, the definition of common mutations for this disorder has allowed good genotype to phenotype correlation, particularly at the severe end of the clinical spectrum.

Gaucher's disease, which is caused by a deficiency of the lysosomal enzyme glucocerebrosidase, also has a high incidence in Ashkenazi Jews (for review, see Barranger and Ginns, 1989). Again, Gaucher's disease has a wide range of clinical phenotypes and has traditionally been grouped into type 1 (adult form, non-neuronopathic, mild clinical severity), type 2 (infantile, neuronopathic, severe) and type 3 (juvenile, neuronopathic, intermediate) forms of the disease. A mutation at codon position 444 of glucocerebrosidase, resulting in the amino acid substitution $L_{444}P$, accounts for approximately 2% and 40% of alleles in Jewish and non-Jewish Gaucher's patients respectively (Tsuji *et al.*, 1987). Homozygotes for the $L_{444}P$ generally have the severe neuronopathic form of Gaucher's disease. However, a homozygote for $L_{444}P$ has been found among Japanese patients with the non-neuronopathic form of Gaucher's disease (Masuno *et al.*, 1990). Another mutation, $A_{370}S$, accounts for approximately 75% and 36% of alleles in Jewish and non-Jewish Gaucher's patients respectively (Tsuji *et al.*, 1988). Patients homozygous or heterozygous for the $A_{370}S$ allele have the non-neuronopathic form of Gaucher's disease. Albeit, as in most adult forms of LSD, there is considerable clinical heterogeneity, even within families and siblings. Genotype to phenotype correlations at the severe end of the clinical spectrum in Gaucher's' disease are reasonable, but are not as simple as in Tay-Sachs disease, perhaps due to the fact that the common mutations defined do not result in nonsense mutations.

The examples described above, and other studies, illustrate the problems that may be encountered in attempting genotype to phenotype correlations. With the definition of mutations in genetic diseases, it is becoming increasingly apparent that for most diseases there is considerable heterogeneity in mutations (Scriver *et al.*, 1989). Although mutation(s) may be common in a particular ethnic population, a large number of rare mutations may account for the remaining alleles. For example, even in diseases

with a common mutation accounting for 50% of disease alleles, only 25% of genotypes will be defined. Thus many patients are likely to be compound heterozygotes. As there are many ways to prevent formation of a functional enzyme, mutations that result in the total loss of enzyme function are common. The clinical phenotypes of patients homozygous for these severe mutations are generally similar. In contrast, the number of mutations that permit some enzymatic function are expected to be limited. Mutations which permit some enzymatic function normally result in a milder disease phenotype and there may be some dose dependence. That is, a patient homozygous for a mild mutation will generally have a milder clinical phenotype than a patient heterozygous for the same mutation. However, as observed by intrafamilial variation in clinical phenotypes of patients with the same genotypes, mutations which permit some enzymatic function may be more susceptible to modulation by variables such as the presence of modifying genes and proteins, timing of treatment received (e.g. surgical procedures and drugs prescribed) and environmental factors (e.g. infection). For example, mild and severe forms of MPS-IIIB have been reported in the same sibship in an isolated case (Andria *et al.*, 1979; Ballabio *et al.*, 1984). The clinical differences observed in mild patients with the same genotype may reflect that, as hypothesized (Conzelmann and Sandhoff, 1984), small differences in residual enzyme activity near a "threshold level" result in minor differences in substrate handling turnover and rate of storage.

Also, identification of mutations at the DNA level can only explain why the respective gene product is defective and not the effect of that mutation on the residual enzyme activity. *In vitro* expression studies may go part of the way to answering this question but in general it will be necessary to understand the three dimensional structure of an enzyme to predict effect of a mutation on enzyme activity, transport and stability in different cell environments during transport to the lysosome. Due to the limitations of mutation analysis discussed above, phenotype prediction with the definition of novel or rare mutations presents a problem. Large deletions or nonsense mutation (introduction of stop codons) generally result in a severe disease phenotype. Caution should always be taken in predicting a patient's clinical outcome from genotype,

especially with the occurrence of "mild" alleles, a novel combination of alleles or the first observation of a described genotype in a different race. Concordance of clinical features in both monozygous and dizygous twins (for review, see McKusick *et al.*, 1972) and sibs (for review, see McKusick *et al.*, 1978) has been reported for MPS-I suggesting that genotype to phenotype correlations should be possible for MPS-I within the limitations discussed above.

1.6 AIMS OF THE PROJECT

The initial aim of the work in this thesis was to isolate and characterize the gene for IDUA, particularly a full-length IDUA cDNA clone. On successful completion of this initial aim, the specific aims of this thesis could be addressed. The specific aims and the significance of the work in this thesis fall into three broad categories that are intimately linked. These aims are:

1) To improve diagnosis of MPS-I by:

   a) Precise chromosomal localization, characterization of the gene for IDUA and definition of polymorphic loci, within or near *IDUA*, which may allow carrier detection and prenatal diagnosis by direct gene analysis or linkage analysis.

   b) Definition of patient mutations could lead to genotype-phenotype correlations, a molecular explanation of the differences between Hurler and Scheie syndromes and improved long-term patient phenotype prediction for:

      i) Genetic counselling, including assessment of whether patients would benefit from specific experimental treatment protocols.

      ii) Assessment of the efficacy of experimental treatment protocols for MPS-I that are currently in use, such as BMT, or assessment of new protocols that may become available, such as enzyme replacement or gene replacement therapy.

c) Definition of mutations could allow carrier detection and prenatal diagnosis by routine detection methods.

2) To improve MPS-I patient treatment by making a full-length IDUA cDNA clone available that may be used for:

a) The *in vitro* production of large amounts of IDUA protein using recombinant DNA technology. The enzyme may be used in enzyme replacement therapy trials in MPS-I animal models and patients.

b) The study of gene replacement therapy *in vitro* in MPS-I patient fibroblasts and *in vivo* in the MPS-I animal models.

3) To increase understanding of lysosomal biogenesis, particularly with respect to transport to the lysosome, maturation, and structure and function of IDUA. This may be achieved though patient studies and/or design of specific experiments once the primary and/or tertiary structure of IDUA becomes known.

CHAPTER 2

MATERIALS AND METHODS

2.1 MATERIALS

2.1.1 ELECTROPHORESIS

The following reagents were obtained from:

Acrylamide	Biorad, Richmond, California, USA
Agarose, type I	Sigma Chemical Co., St. Louis, Missouri, USA
Agarose, low gelling temperature	Sigma Chemical Co.
Ammonium persulphate	Ajax, Auburn, NSW, Australia
Bromophenol blue	B.D.H. Chemicals Ltd., Poole, Dorset, England
Ethidium bromide	Boehringer Mannheim, Mannheim, Germany
Formamide	Ajax, Auburn, NSW, Australia
NuSieve GTG Agarose LMP	FMC Bioproducts, Rockland, USA
Sequagel sequencing system	National Diagnostics, New Jersey, USA
NNN'N'-tetramethylethenediamine (TEMED)	Biorad
Urea	Ajax
Xylene cyanol FF	Tokyo Kasei, Tokyo, Japan
SppI *Eco*RI, pUC19 *Hpa* II, λ *Hin*dIII, λ *Hin*dIII-λ *Hin*dIII/*Eco*RI size markers	Bresatec, Adelaide, Australia

2.1.2 ENZYMES

The enzymes used in this study were obtained from the following companies:

Calf intestinal phosphatase	Boehringer Mannheim
E. coli DNA polymerase I (Klenow fragment)	Amersham, Buckinghamshire, U.K.
Lysozyme, chicken	Sigma Chemical Co.
DNase I	Boehringer Mannheim

Proteinase K	Sigma Chemical Co.
Reverse transcriptase (Mo-MLV)	Bestheda Research Laboratories, Gathersburg, USA
Ribonuclease A	Sigma Chemical Co.
RNAsin	Promega, Maddison, USA
T4 DNA ligase	Boehringer Mannheim
T4 polynucleotide kinase	Pharmacia, Uppsala, Sweden
Taq polymerase	Perkin Elmer Cetus, Norwalk, Connecticut, USA and Biotech International, Perth, Australia

All restriction endonucleases in this study were obtained from Boehringer Mannheim and New England Biolabs (Beverly, Massachusetts, USA) unless specified.

2.1.3 RADIOCHEMICALS

α-^{32}P-dCTP, 3000Ci/mmole	Radiochemical Centre, Amersham
γ-^{32}P-ATP, 5000Ci/mmole	Radiochemical Centre, Amersham
γ-^{32}P-ATP, 5000Ci/mmole	Bresatec, Adelaide, Australia

2.1.4 BUFFERS AND SOLUTIONS

Buffers and solutions routinely used in this study were as follows:

Formamide loading buffer	92.5% (v/v) formamide, 20 mM EDTA, 0.1% (w/v) xylene cyanol, 0.1% (w/v) bromophenol blue.
10 X loading buffer	50% (v/v) glycerol, 1% (w/v) SDS, 100 mM EDTA, 0.1% (w/v) xylene cyanol, 0.1% (w/v) bromophenol blue.
5 X ligation buffer	250 mM Tris-HCl, pH 7.6, 50 mM $MgCl_2$, 25% (w/v) PEG 6000, 5 mM rATP, 5 mM DTT.
M9 salts	1.05% K_2HPO_4, 0.45% (w/v) KH_2PO_4, 0.1% (w/v) $(NH_4)_2SO_4$, 0.05% (w/v) sodium citrate.
Phosphate buffered saline (PBS)	130 mM NaCl, 10 mM $NaHPO_4$, 10 mM NaH_2PO_4, pH 7.2.
10 X PCR buffer	500 mM KCl, Tris-HCl, pH 8.4, 25 mM $MgCl_2$, 1% (v/v) Triton X-100, 0.02% (w/v) gelatin.

20 X SSC	3 M NaCl, 0.3 M tri-sodium citrate.$2H_20$, pH 7.0.
SM buffer	100 mM NaCl, 10 mM $MgSO_4.3H_2O$, 0.01% (w/v) gelatin, 50 mM Tris-HCl, pH 7.5.
TAE	40 mM Tris-acetate, 2 mM EDTA, pH 8.5.
TBE	89 mM Tris-base, 89 mM boric acid, 2.5 mM EDTA, pH 8.3.
TE	10 mM Tris-HCl, pH 7.5, 0.1 mM EDTA.
TES	25 mm Tris-HCl, pH 8.0, 10 mM EDTA, 15% (w/v) sucrose.

2.1.5 BACTERIAL MEDIA

2.1.5a LIQUID MEDIA

All liquid media were prepared using millipore water and were sterilized by autoclaving. The compositions of the various media were as follows:

L-broth	1% (w/v) Bacto tryptone, 0.5% (w/v) Bacto yeast extract, 1% (w/v) NaCl. pH to 7.5 with NaOH.
Minimal medium	1 mM $MgSO_4$, 0.1 mM $CaCl_2$, 1 mM thiamine HCl, 0.2% (w/v) glucose, 1 X M9 salts.
Terrific broth (TB)	3.5% (w/v) Bacto tryptone, 2% (w/v) Bacto yeast extract, 0.5% (w/v) NaCl. pH to 7.5 with NaOH.
2 X YT	1.6% (w/v) Bacto tryptone, 1% (w/v) Bacto yeast extract, 0.5% (w/v) NaCl.

2.1.5b SOLID MEDIA

L-Agar	L-broth, 1% (w/v) Bacto agar
L-Agarose	L-broth, 1% (w/v) agarose
L-Amp	L-broth, 1% (w/v) Bacto agar, ampicillin (100 µg/ml)
Minimal medium plus glucose	Minimal medium, 1% (w/v) Bacto agar, 0.02% (w/v) $MgSO_4$, 0.2% (w/v) glucose, 0.0005% (w/v) thiamine HCl.

2.1.6 ANTIBIOTICS

Ampicillin	Sigma
Chloramphenicol	Sigma
G418 sulphate (Neomycin) Geneticin	Gibco, Glen Waverley, Vic, Australia
Kanamycin	Boehringer Mannheim
Streptomycin	Boehringer Mannheim

2.1.7 BACTERIAL STRAINS

The E. Coli K12 strains used in this study are described below:

C600	*sup*E44, *thi*-1, *leu*B6, *lac*Y1, *ton*A21, λ^-, Restriction:(r_k+, m_k+), *mcr*A(-), *mcr*B(+), (Huynh *et al.*, 1984).
JM101	*sup*E, *thi* Δ(*lac-pro*AB), {F' *tra*D36, *pro*AB, *lac*^{q}ZΔM15}, Restriction: (r_k+, m_k+), mcrA(+), (Yanish-Perron *et al.*, 1985).
Y1088	Δ(*lac*U169), *sup*E, *sup*F, *hsd*R(r-, m_+), *met*B, *trp*R, *ton*A21, *pro*C::Tn5, (pMC9), (Huynh *et al.*, 1984).
NM538	*sup*F, *hsd*R (r_k-, m_k+), (Frischauf *et al.*, 1983).

2.1.8 VECTORS

The vectors used in this study are described below:

Bacteriophage strains:

Vector	Genotype and reference
Lambda-gt10	*λb527, srlλ3°, imm^{434}, srlλ4°, srlλ5°* (Huynh *et al.*, 1984).
Lambda-gt11	*λlac5, srlλ3°, cl857, srlλ4°, nin5, srlλ5°, Sam100* (Young and Davis, 1985).
EMBL3	*λsbhl*λ1°, *b189*, < polylinker, (EcoRI-Sal I), *int*29, *nin*L44, *c*I857, *trp*E, polylinker (Sal I-EcoRI) > *KH*54, *chi*C, *srl*λ4°, *nin*5, *srl*λ5° (Frischauf *et al.*, 1983).

Filamentous phage:

The replicative forms of M13 were purchased from Boehringer Mannheim.

M13mp18	Yanish-Perron *et al.*, 1985
M13mp19	Yanish-Perron *et al.*, 1985

Phagemid vectors:

The following phagemid vectors were purchased from the listed suppliers.

pBLUESCRIPT SK II	Stratagene, La Jolla, California, USA
pTZ19U	Biorad, Richmond, California, USA

Plasmid vectors:

pUC19	Bresatec, Yanish-Perron *et al.*, 1985

2.1.9 RECOMBINANT DNA LIBRARIES

All libraries screened were of human origin and were purchased from Clontech (Palo Alto, California, USA), except for the chromosome 22 specific library which was obtaincd from the American Type Cell Culture Collection.

GENOMIC LIBRARIES

Tissue source	Vector	Catalogue number
Leukocyte genomic DNA	EMBL3	HL1006
Chromosome 22 specific	Charon24A	LL22NSO1

cDNA LIBRARIES

Tissue source	Vector	Catalogue number
Colon cDNA random-primed	λgt10	HL1034a
Lung fibroblast cDNA	λgt11	HL1011
Testis cDNA	λgt11	HL1010b
Placental cDNA	λgt11	HL1008
Umbilical endothelial veins cDNA	λgt11	HL1024b
Umbilical endothelial veins 5'-stretch cDNA	λgt11	HL1070b
T-cell 5'-stretch cDNA random and oligo dT primed	λgt11	HL1068b
Foetal liver 5'-stretch cDNA	λgt10	HL1064a

All cDNA libraries were synthesized by oligo dT priming unless otherwise stated. The 5'-stretch cDNA libraries were reverse transcribed in the presence of methyl-mercury hydroxide to prevent secondary structure formation in mRNA.

2.1.10 TISSUE CULTURE SOLUTIONS

Basal medium eagles (modified)	Flow laboratories
F-12 nutrient media (Ham's)	Gibco
Foetal calf serum	Commonwealth Serum Laboratories or Gibco
Glutamine	Flow Laboratories, Sydney, Australia
Phosphate buffered saline	Commonwealth Serum Laboratories, Melbourne, Vic., Australia

2.1.11 MISCELLANEOUS MATERIALS

Colony/Plaque Screen™ filters	Dupont-NEN Research Products
Hybond N+™ nylon membrane	Amersham
NA45 DEAE-cellulose paper	Schleicher and Schuell, Dassel, Germany
Nylon filters (GeneScreen*Plus*™)	Dupont-NEN Research Products, Boston, MA, USA
Oligo (dT)-cellulose	Pharmacia P-L Biochemicals
Polyvinyl ELISA plates	Costar, Cambridge, MA, USA
Positive Land film, type 677	Polaroid
Sephadex G-50	Pharmacia P-L Biochemicals
X-ray film	Fuji Rx X-ray film, Fuji Photo film Co. Ltd.

2.1.12 MISCELLANEOUS FINE CHEMICALS

5-bromo-4-chloro-3-indolyl -β-D-galactopyranoside (BCIG)	Boehringer Mannheim
Chemicals for oligonucleotide synthesis	Applied Biosystems, Melbourne, Australia
Caesium chloride	Boehringer Mannheim
Deoxynucleotides	Boehringer Mannheim
Dideoxynucleotides	Bochringer Mannheim
Dideoxy sequencing kits	Bresatec or Boehringer Mannheim

Diethlypyrocarbonate	Sigma Chemical Co.
Dimethylsulphoxide (DMSO)	Sigma Chemical Co.
Isopropyl-β-D-thiogalactopyranoside (IPTG)	Boehringer Mannheim
Herring Sperm DNA	Sigma Chemical Co.
Hydroxylamine	Sigma Chemical Co.
4-methylumbelliferyl -α-L-iduronide	Calbiochem
Random priming oligolabelling kits	Amersham
Osmium tetroxide	Sigma Chemical Co.
Phenol	Wako Chemical Co., Osaka, Japan
Piperidine	Sigma Chemical Co.
Pyridine	Aldrich, Milwaukee, Wisconsin, USA
Sarkosyl	Ciba-Geigy, Basle, Switzerland
SDS	Sigma Chemical Co.
Spermidine	Sigma Chemical Co.

All other chemicals used in this study were analytical reagent grade.

2.1.13 DESIGN AND SYNTHESIS OF SYNTHETIC DNA OLIGONUCLEOTIDES

The oligonucleotide probe ID50 and oligonucleotides designated ID1 to ID7 were synthesized by Bresatec (Adelaide, Australia). ID8, ID9, ID47, IDP1 and IDP2 were synthesized at the Department of Haematology of the Flinders Medical Centre. IDP3, IDP4, λgt10 and λgt11 forward (F) and reverse (R) primers were synthesized at the Department of Human Immunology of the Institute of Medical and Veterinary Science. All other oligonucleotides were synthesized using an Applied Biosystems 391 DNA synthesizer in the Department of Chemical Pathology at the Adelaide Children's Hospital.

2.1.13a OLIGONUCLEOTIDE PROBES

Oligonucleotide probes were designed from amino acid sequence data obtained from immunopurified human liver IDUA (Clements *et al.*, 1989). Regions of amino acid sequence were chosen for oligonucleotide design on the basis of low codon degeneracy and homogeneity of peptide sequence. Oligonucleotide probes were synthesized either with incorporation of inosine (I) (Ohtsuka *et al.*, 1985; Martin *et al.*, 1985), all nucleotide options, or choices based on human codon usage at the third base position in a codon (Masuyama *et al.*, 1986). Oligonucleotide probes which contained inosine or all nucleotide options at the third base position were synthesized as 20-mers and, to increase specificity, clones which were positive to these probes were confirmed by hybridization to another 20-mer oligonucleotide probe made to the same, or another, peptide sequence. Oligonucleotide probes incorporating choices based on human codon usage were synthesized as long as possible based on availability of peptide sequence. All probes used in this study are listed in Appendix A and discussed in the relevant parts of the text.

2.1.13b SEQUENCING PRIMERS

Oligonucleotide primers for sequencing were designed to unambiguous previously sequenced regions, priming towards a region of unknown sequence. They were synthesized as 20-mers at spacings of approximately 200 to 400 bp and containing an equal proportion of purine and pyrimidine bases. Later primers used for sequencing were designed according to the parameters for PCR primers described below. The sequence of primers used to sequence the IDUA cDNA and genomic clones are listed in Appendices B and C respectively.

2.1.13c PCR PRIMERS

Oligonucleotides for PCR were designed to contain the same proportion of purine and pyrimidine bases found in the target DNA sequence to be amplified. Efficient priming at high temperatures was maximized by attempting to include one or two G or C residues at the 3' end of the primers. In addition, primer pairs were carefully

checked at their 3' ends to avoid the possibility of primer-dimer formation. The standard length of oligonucleotides used in this study was 24-nt, although 30-mers were used if the target sequence proved difficult to amplify. The sequences of IDUA-specific PCR primers are listed in Appendix D.

2.1.14 PREPARATION OF GLASSWARE AND SOLUTIONS

All solutions were prepared using millipore water (0.04 μSiemens) and sterilized by autoclaving at 125°C and 120 kPa. Glassware was treated in Pyroneg detergent, rinsed in deionized water and dried overnight. Spatulas and centrifuge tubes were rinsed in ethanol and air-dried.

2.2 GENERAL METHODS

2.2.1 PLASMID DNA PREPARATION

2.2.1a MINIPREPS

Plasmid minipreps were prepared by a modified procedure of Birnboim and Doly (1979). Briefly, single colonies were grown overnight in 2 ml of L-broth with the appropriate antibiotic. Cells were pelleted by centrifugation at 12,000 g for 1 min, the pellet resuspended in 100 μl of TES buffer with lysozyme at a final concentration of 2 mg/ml, and the suspension incubated at room temperature for 10 min. Freshly prepared 0.2 M NaOH, 1% (w/v) SDS was added to each tube (200 μl), mixed by inversion and incubated on ice for 5 min followed by addition of 150 μl of an ice-cold solution of 3 M sodium acetate, pH 5.5. The solution was incubated on ice for a further 10 min and cellular debris and chromosomal DNA was pelleted by centrifugation at 12,000 g for 15 min at 4°C. The supernatant was transferred to a fresh tube, phenol/chloroform extracted, ethanol precipitated, washed with 70% (v/v) ethanol, vacuum-dried and resuspended in 40 μl of water or TE buffer. RNase A was added to digestions of the miniprep plasmid DNA to a final concentration of 10 μg/ml.

2.2.1b LARGE SCALE PLASMID PREPARATION

Recombinant clones, as selected by restriction analysis of miniprep plasmid DNA, were grown overnight in L-broth supplemented with the appropriate antibiotic. The cells were pelleted by centrifugation at 8,000 g for 2 min at 4°C, resuspended in 12 ml of TES buffer, lysozyme was added to a final concentration of 2 mg/ml and incubated at room temperature for 10 min. A solution of freshly prepared ice-cold 0.2 M NaOH, 1% (w/v) SDS was added, the solution was carefully mixed by inversion and incubated on ice for 10 min. Following addition of 15 ml 3 M sodium acetate, pH 4.6, the solution was again carefully mixed by inversion and incubated on ice for 20 min. The solution was centrifuged at 12,000 g for 15 min and the supernatant carefully transferred to a fresh tube. RNase A (100 µl of 1 mg/ml) was added to the supernatant and incubated for 20 min at 37°C. The solution was extracted twice with 1 volume of phenol/chloroform (1:1) saturated with TE buffer, and vortexed for 1 min. The solution was centrifuged at 12,000 g for 5 min and the aqueous phase transferred to a fresh tube by addition of 1 volume of chloroform/isoamyl alcohol (24:1). This solution was vortexed and centrifuged as above. The aqueous phase was transferred to a fresh tube, 2 volumes of ethanol were added and the solution left at -20°C for 30 min. Centrifugation at 12,000 g for 20 min was performed after which the tube was carefully drained and the DNA pellet dissolved in 3.2 ml of water. Following resuspension, 1 ml of 4 M NaCl and 4 ml of 13% (w/v) PEG 6000 were added, the solution mixed and incubated on ice for 60 min. The tube was centrifuged at 12,000 g for 10 min, the supernatant removed and the pellet washed once with 70% (v/v) ethanol. The pellet was dried under vacuum and dissolved in an appropriate volume of water or TE buffer. DNA used for cell culture expression studies was further purified by ultracentrifugation on a caesium chloride gradient.

2.2.2 ETHANOL PRECIPITATION OF DNA

In general, all samples were precipitated in ethanol in the presence of 0.3 M sodium acetate, pH 5.5 or 3.75 M NH_4OH at -20°C for approximately 1 hr. Following centrifugation, pellets were rinsed with 70% (v/v) ethanol, dried briefly under vacuum and resuspended in water or TE buffer.

2.2.3 RESTRICTION ENDONUCLEASE DIGESTION OF DNA

Restriction endonuclease digestion of DNA was carried out using the conditions appropriate for each enzyme as detailed by the manufacturer. Spermidine was added at a final concentration of 4 mM in digestions of lambda DNA or where initial digestion yielded poor results.

2.2.4 PLASMID VECTOR PREPARATION

Vector DNA for use in ligations was digested with the appropriate restriction endonuclease(s) for approximately 2 hr. Completely digested DNA, analysed on an agarose gel, was treated with calf intestinal phosphatase (CIP) in the presence of 0.1 mM zinc chloride in order to prevent self-ligation by removal of the 5'-terminal phosphate groups. The reaction was incubated at 37°C for 30 min, after which the reaction volume was increased to 200 µl with 50 mM Tris-HCl, pH 8.0 and a phenol/chloroform extraction followed by a chloroform extraction were carried out. The final concentration of the vector was 10 ng/µl.

2.2.5 LIGATION OF PLASMID VECTORS

The DNA insert and appropriate plasmid vector were combined in a molar ratio of 3:1 in a 20 µl reaction volume in a final concentration of 1 x ligation buffer. For ligations of DNA with overhanging ends, 0.5 units of T4 DNA ligase were used, and for blunt-end ligations, 1.0 unit of enzyme was used. Ligations were performed at room temperature, generally overnight

2.2.6 TRANSFORMATION OF *E. coli.*

Strains of *E. coli* were made competent for use in transformations as follows. A 40 ml culture of L-broth was inoculated with a 2 ml overnight culture grown from a single colony. This culture was incubated at 37°C with shaking until an OD_{600nm} of 0.4 was reached. The cells were pelleted by centrifugation at 2,000 g for 5 min at 4°C, resuspended in 20 ml of an ice-cold solution of 100 mM $MgCl_2$ and incubated on ice for 30 min. Following centrifugation, the cells were resuspended in 2 ml of 100 mM $CaCl_2$ and left on ice for at least 2 hr. The DNA ligation reaction (5µl) was added to 100 µl of the competent cells and incubated on ice for one hr. A heat-shock reaction was then performed at 42°C for 2 to 5 min. For plasmid and phagemid vectors, 1ml of L-broth was added to the transformations and incubated at 37°C with shaking for 1 hr. The cells were pelleted at 6,500 rpm for 2 min, resuspended in 100 µl L-broth and plated onto L-plates containing ampicillin. If appropriate, 20 µl each of BCIG (2% (w/v) in dimethylformamide) and IPTG (25 mg/ml) were added to the cells before plating.

2.2.7 AGAROSE GEL ELECTROPHORESIS OF DNA

Electrophoresis of DNA for analytical purposes was generally carried out in Pharmacia GNA100 tanks for horizontal submerged gel electrophoresis. The electrophoresis of DNA for transfer to nylon membranes was carried out in a variety of tanks according to the number of samples. Agarose was used at 0.8 to 2.0% (w/v) in TAE buffer. Samples were electrophoresed at 120 V until the bromophenol blue marker dye had migrated a sufficient distance to ensure adequate separation of the DNA fragments. DNA was visualized under UV light after brief staining of the gel in 10 µg/ml of ethidium bromide. A positive photograph of the stained gel was then taken using Polaroid land film 667.

2.2.8 POLYACRYLAMIDE GEL ELECTROPHORESIS

2.2.8a NON-DENATURING GELS

For non-denaturing polyacrylamide gels, a prepared solution of 20:1 (acrylamide:bis-acrylamide) was mixed at the required percentage depending on the size of the DNA fragments to be separated. The gels were run in TBE buffer, pH 8.3, on a Hoefer Tall Mighty Small gel system at 80V.

2.2.8b DENATURING GELS

For denaturing sequencing polyacrylamide gels, a commercially prepared solution was mixed and polymerized according to the manufacturer's specifications (usual concentration of 5% (w/v) acrylamide; Sequagel, National Diagnostics). Gels were run in TBE buffer, pH 8.3, at a constant voltage (2,000 V) for the required distance as judged by the migration of the bromophenol blue and xylene cyanol dyes present in the loading buffer. Following electrophoresis, the gel was fixed for 10 min with 10% (v/v) acetic acid, 20% (v/v) ethanol, transferred to a sheet of Whatman 3MM paper, covered with plastic wrap and vacuum-dried. The gel was then autoradiographed overnight at room temperature in the absence of intensifying screens.

2.2.9 ISOLATION OF RESTRICTION FRAGMENTS

2.2.9a FROM POLYACRYLAMIDE GELS

Following excision of the required polyacrylamide band as viewed under a UV transilluminator, the gel slice was immersed in elution buffer (500 mM ammonium-acetate, 10 mM $MgCl_2$, 2 mM EDTA, 20 mM Tris-HCl, pH 8.0, 0.2% (w/v) SDS) and incubated overnight at 65°C. Following ethanol precipitation, the pellet was resuspended in 300 mM sodium acetate, pH 5.5, and ethanol precipitated once again. The pellet was washed with 70% (v/v) ethanol, vacuum-dried and resuspended in the appropriate volume of water or TE buffer.

2.2.9b FROM LOW MELTING POINT AGAROSE GELS

The appropriate DNA fragment was excised from the gel and the DNA extracted by placing the gel slice in 5 volumes of TE buffer (50 mM Tris-HCl, pH 8.0, 1 mM EDTA) and incubating at 65°C until the agarose had completely melted. Phenol extraction was carried out twice and the DNA precipitated using 0.1 volume of 3 M sodium acetate and 2.5 volumes of absolute ethanol.

2.2.9c FROM NORMAL AGAROSE GELS: USING GENECLEAN

DNA fragments were purified from agarose gels in 1 X TAE with a Geneclean™ kit in the following way. The appropriate DNA fragment was excised, placed in a solution of NaI and incubated at 55°C until the agarose had melted. Glass milk solution was added to the melted agarose solution and the tube was placed on ice for 5 min. Following centrifugation, the pellet was washed three times and the final pellet resuspended in a small volume of water or TE buffer. Elution of DNA was allowed to occur at 55°C for 3 min and repeated twice more on the pellet, after which the eluates were pooled.

2.2.9d FROM NORMAL AGAROSE GELS: USING DEAE-CELLULOSE PAPER

Isolation of DNA from agarose gels using DEAE paper was a modification of the method described by Selden and Chory (1987), and was performed as follows. After separation of the DNA fragments by agarose gel electrophoresis, the appropriate band was identified, a slit made in the gel to the side of the fragment, and a piece of DEAE-paper inserted into the slit. The DNA was electrophoretically migrated onto the paper by rotating the gel 90° from its original running position and the paper placed in high salt elution buffer (1 M NaCl, 0.1 mM EDTA, 20 mM Tris-HCl, pH 8.0) and incubated at 65°C for 30 min. The elution buffer was removed to a fresh tube, two volumes of ethanol was added and the DNA was precipitated, washed in 70% (v/v) ethanol, and resuspended in an appropriate volume of water or TE buffer.

2.2.10 PURIFICATION OF OLIGONUCLEOTIDES

Oligonucleotides were cleaved from their synthesis columns by three incubations in 0.4 ml of 10 M NH_4OH at room temperature, for 15 to 30 min. The aliquots of cleaved oligonucleotide were pooled and oligonucleotide protection groups removed by overnight incubation at 55°C.

Initially, oligonucleotides were concentrated by drying in a vacuum centrifuge with addition of a drop of tetra-ethylamine every 30 min. The lyophilized oligonucleotides were resuspended in 100 µl of 100 mM Tris-HCl, pH 8.0, 500 mM NaCl, 5 mM EDTA and the concentration determined by spectrophotometry at 260 nm (A_{260} 1 = 33 µg/ml). Oligonucleotides which did not kinase to a high specific activity were further purified by chromatography on Sep-Pak columns (Millipore) according to the manufacturer's instructions.

Oligonucleotides designated ID70 and over were purified using the method of Sawadogo and Van Dyke (1991). Briefly, 100 µl of the deprotected primer in NH_4OH was added to 1 ml of n-butanol and centrifuged at 12,000 g for 30 s. The pellet was resuspended in 100 µl of water and the procedure repeated. The final pellet was air-dried and resuspended in an appropriate volume of water.

2.2.11 ^{32}P RADIO-ISOTOPE LABELLING OF DNA

2.2.11a 5' END-LABELLING OF OLIGONUCLEOTIDES

Synthetic oligonucleotides were 5' end-labelled using T4 DNA polynucleotide kinase and γ-^{32}P-dATP as described by Chaconas and van de Sande (1980), with the addition of spermidine to a final concentration of 0.1 mM.

2.2.11b PRIMER EXTENSION

Labelling of double stranded DNA was performed by primer extension of random oligonucleotides (Feinberg and Vogelstein, 1983) using the Amersham Multiprime DNA labelling systems kit. Briefly, a small quantity of DNA insert (25-50

ng) was denatured at 100°C for 2 min and added to a solution containing random hexamers, dATP, dGTP, dTTP, α-^{32}P-dCTP, Klenow enzyme and buffer. The reaction mixture was incubated either at room temperature overnight or at 37°C for 2 hr. The radioactive probe was then purified (see below), denatured and used for hybridization.

2.2.11c PROBE PURIFICATION

Unincorporated radionucleotides were separated from the labelled probe through a Sephadex G-50 column. Two drop fractions were collected, resulting in a distinct peak of labelled probe when followed using a mini-monitor β-counter.

2.2.12 TRANSFER OF DNA TO NYLON MEMBRANES

2.2.12a PLAQUE LIFTING

Plaques were transferred onto Colony/Plaque Screen™ filters following the manufacturer's instructions. Briefly, the plates were placed at 4°C for 1 to 2 hr. The filters were gently laid onto the plaques and left at room temperature for 2 min. Subsequent transfers of the same plate were left for an additional 2 min each. The filters were then immersed twice in 0.4 M NaOH for 2 min, and then twice in 1.0 M Tris-HCl, pH 7.5, for 2 min. Blotted filters were either air-dried or heated in a microwave oven for 10-20 s. The dried filters were then prehybridized, hybridized and washed as described in Section 2.2.13.

2.2.12b SOUTHERN BLOTTING

Restriction endonuclease digested DNA was fractionated on agarose slab gels and transferred to GeneScreen*Plus*™ (Dupont, NEN) or Hybond N^{+}™ (Amersham) nylon membrane using the alkaline transfer method (Reed and Mann, 1985). Briefly, if the DNA to be transferred was over 1-kb in size, an acid nicking step was included by soaking twice in 0.25 M HCl for 15 min with gentle shaking. The gel was then immersed in 0.4 M NaOH twice for 15 min with gentle shaking. The nylon membrane was cut to the size of the gel and placed in the transfer solution (0.6 M NaCl, 0.4 M

NaOH) to soak for 15 min. Transfer of the DNA was carried out using a double-layered wick of Whatman 3MM paper onto which the gel was placed. The membrane was placed on top of the gel followed by three to six sheets of 3MM paper and a wad of blotting tissue. Transfer was performed for a minimum of 2 hr but was generally allowed to take place overnight. Following transfer, the filter was neutralized in a solution of 0.1 M Tris-HCl, pH 7.5, air-dried and then prehybridized and hybridized as described in Section 2.2.13.

2.2.13 PREHYBRIDIZATION, HYBRIDIZATION AND WASHING

2.2.13a WITH OLIGONUCLEOTIDE PROBES

Prehybridization was carried out at 42°C in a solution consisting of 20% (v/v) 5 X P (1% (w/v) bovine serum albumin, 1% (w/v) polyvinylpyrrolidone M_r 40,000, 1% (w/v) ficoll M_r 400,000, 250 mM Tris-HCl, pH 7.4, 0.5% (w/v) pyrophosphate), 1 M NaCl, 1% (w/v) SDS, 10% (w/v) dextran sulphate and 100 µg/ml denatured herring sperm DNA for at least 2 hr with shaking. Hybridizations were performed in the same solution and incubated overnight at 42°C with between 1 and 10 ng/ml of 5' end-labelled probe. Filters were washed under conditions determined by the melting temperature of the primer and target sequence, generally with an initial wash in 6 X SSC, 0.1% (w/v) SDS for 15 min at room temperature. This was followed by a 15 min wash in 2 X SSC, 0.1% (w/v) SDS with a gradual increase in washing temperature up to 65°C. Autoradiography was carried out at room temperature or, for detection of low levels of radioactivity, at -80°C in the presence of tungsten intensifying screens.

2.2.13b WITH OLIGOLABELLED PROBES

Conditions used for hybridization using oligolabelled probes were as follows: The prehybridization solution consisted of 50% (v/v) formamide, 1% (w/v) SDS, 1 M NaCl, 10% (w/v) dextran sulphate and 100 µg/ml of denatured herring sperm DNA. Incubation was carried out at 42°C for at least 2 hr. The denatured radioactive probe was added to a final concentration of between 1 and 10 ng/ml (1-4 x 10^5 dpm/ml) and

the hybridization allowed to incubate overnight at 42°C. Filters were washed in 2 X SSC, 0.1% (w/v) SDS at room temperature for 5 min with constant agitation, and then twice in 2 X SSC, 1% (w/v) SDS for 30 min with a gradual increase in temperature from room temperature to 65°C. A further wash in 0.1 X SSC at 65°C for 30 min was performed if background signal was still high. Autoradiography was carried out as described above.

2.2.14 SCREENING OF RECOMBINANT DNA LIBRARIES

All media used with lambda vectors contained 20 mM $MgCl_2$ and 0.2% (w/v) maltose. All libraries were plated using 0.7% (w/v) top agarose, at a density of between 40,000 and 55,000 plaques per 140 mm L-agar plate. Host cells used for plating phage were grown in the presence of 0.2% (w/v) maltose. The host cells used for each library type were: NM538 for the EMBL3 genomic library, C600 for the λgt10 cDNA libraries and Y1088 for the λgt11 cDNA libraries. Plates were grown at 37°C for between 8 and 16 hr. Probes were either 5' end-labelled or labelled by primer extension of random oligonucleotide primers. Colony/Plaque Screen™ filters were prepared as described in Section 2.2.12a and prehybridized, hybridized and washed as described in Section 2.2.13.

Positive plaques observed after autoradiography were aligned to the agar plates and agar plugs were isolated using the large end of a sterile pasteur pipette in 1 ml of SM. The phage solution was replated at a density of between 500 to 2,000 plaques per 140 mm plate. Third round positive plaques were replated at a lower density of between 100 to 500 plaques per 140 mm plate. The procedure was repeated until pure plaques could be isolated which were then plated to confluency on L-agarose plates for lambda DNA preparation as described in Section 2.2.15.

2.2.15 PREPARATION OF LAMBDA DNA FROM PLATE LYSATES

Lambda DNA was prepared by a modification of the method reported in Maniatis *et al.* (1982). Lambda phage were plated on L-agarose plates, containing 20 mM $MgCl_2$ and 0.2% (w/v) maltose, at approximately 1 x 10^5 pfu per 150 mm plate. The confluent plates were covered with 10 ml of 1 X SM solution and incubated at 4°C for at least 2 hr, but the incubation was generally overnight. The solution was then removed and centrifuged at 8,000 g for 20 min to pellet bacteria. The supernatants were transferred to fresh tubes, DNase I and RNase A were added to final concentrations of 0.1 µg/ml and 1 µg/ml respectively, and the tubes incubated at 37°C for 60 min. An equal volume of 2.5 M NaCl, 20% (w/v) PEG was added to the supernatants, mixed, and then incubated on ice for at least 60 min. The tubes were centrifuged at 11,000 g for 20 min to pellet the phage, the supernatants discarded and the pellet drained to ensure the complete removal of PEG. The pellets were resuspended in 0.5 ml of SM buffer per tube and, following addition of 5 µl 0.5 M EDTA and 5 µl 10% (w/v) SDS, they were incubated at 65°C for 20 min. The solutions were extracted twice with water-saturated phenol, once with phenol/chloroform and once with chloroform with centrifugation at 12,000 g for 5 min for each extraction. An equal volume of isopropanol was added to the final supernatants and the tubes incubated at -80°C for at least 20 min. After centrifugation at 12,000 g for 15 min, the DNA pellet was washed with 70% (v/v) ethanol, resuspended in TE buffer, ethanol precipitated with sodium acetate, vacuum-dried and resuspended in an appropriate volume of water.

2.2.16 GROWTH AND HARVESTING (+) STRAND OF RECOMBINANT BACTERIOPHAGE

2.2.16a M13 FILAMENTOUS PHAGE

Recombinant (white) M13 plaques were picked using sterile toothpicks and placed in sterile 10 ml centrifuge tubes containing 2 ml of 2 X YT media with a 1 in 50 dilution of an overnight JM101 culture. The phage were grown at 37°C with vigorous shaking for 5 to 8 hr and single stranded DNA for sequencing was prepared as described in Section 2.2.16c.

2.2.16b PHAGEMIDS

Single recombinant phagemid colonies were grown at 37°C overnight in 2 ml of TB with ampicillin. The overnight culture (100-150 µl) was added to a fresh 2 ml aliquot of TB and ampicillin with between 2.5 x 10^6 and 5 x 10^8 pfu of M13K07 helper phage. In general 5 x 10^8 pfu of helper phage were used per single stranded DNA preparation, however, when difficulties with DNA yields were encountered, a range of helper phage pfu were used. The culture was grown at 37°C for 1 to 2 hr followed by addition of kanamycin to a final concentration of 70 µg/ml and overnight incubation at 37°C. Single stranded DNA for sequencing was then prepared as described in Section 2.2.16c.

2.2.16c SINGLE STRANDED DNA PREPARATION

Cells were removed from 1.5 ml of the ssDNA cultures by centrifugation at 12,000 g for 10 min. This centrifugation step was repeated and 1 ml of the supernatant transferred to a fresh tube. A solution of 20% (w/v) PEG 6000/ 2.5 M NaCl (300 µl) was added to the supernatant, incubated at room temperature for 15 min and centrifuged at 12,000 g for 10 min for precipitation of the phage. After removal of the supernatant, the pellet was resuspended in 180 µl of water and 20 µl of "cracking buffer" (100 mM Tris-HCl, pH.7.5, 1 mM EDTA, 5% (w/v) SDS) and incubated at 80°C for 10 min, as described by Eperon (1986). The solution was cooled to room temperature, followed by

the addition of 80 µl of 7.5 M ammonium acetate and 560 µl of isopropanol. This solution was incubated at room temperature for 15 min and centrifuged at 12,000 g for 10 min. The pellet was washed once in 70% (v/v) ethanol, vacuum-dried and resuspended in 30 µl of water.

2.2.17 DIDEOXY SEQUENCING REACTIONS

2.2.17a ANNEALING

The following were combined in a tube, incubated at 75°C for 5 min and allowed to cool to room temperature: 2 µl of the single-stranded recombinant M13 template, 1 µl of primer (2 ng of Universal sequencing primer or 10 ng of specific primers), 1 µl of 10 X TM buffer (100 mM Tris-HCl, pH 7.5, 100 mM $MgCl_2$), 2 µl of water. The tubes were centrifuged briefly to spin down condensation and used in the polymerization step as described below.

2.2.17b POLYMERIZATION

The chain termination reaction (Sanger *et al.*, 1977) was performed as described by the manufacturers of the sequencing kits. The radiolabelled nucleotide used was α-^{32}P-dCTP and the reactions were carried out at 42°C using the Klenow fragment of DNA polymerase. The final dideoxynucleotide concentrations used in this study varied according to the template to be sequenced, depending on the ratio of purines to pyrimidines present. Generally, the final concentration in the reaction mixes were 33 µM ddATP, 4.4 µM ddCTP, 11 µM ddGTP and 111 µM ddTTP. Compressed areas of GC-rich sequence were resolved using 7-deaza-guanosine in place of dGTP in the reaction mixes (Mizusawa *et al.*, 1986). An aliquot of each reaction was added to formamide loading dye, heat-denatured at 100°C for 3 min and loaded onto a 5% sequencing gel as described in Section 2.2.8b.

2.2.18 COMPUTER ANALYSIS OF DNA AND PEPTIDE SEQUENCES

Nucleotide sequences obtained were screened against the GenBank nucleotide sequence data base (Release 62.0, December 1989) and the encoded protein sequences were screened against the National Biomedical Research Foundation Protein Data Base (Release 23.0, December 1989). The compilation of nucleotide sequence from the multiple subclones and sequencing reactions was performed using various STADEN (Staden, 1980; 1984) and Genetics Computer Group (GCG) programs (Devereux *et al.*, 1984), including:

Miscellaneous Programs:

CLUSTAL	A multiple sequence alignment program, particularly for peptide sequences (Higgins and Sharp, 1988).
DNASCAN	A program used to scan sequences with a matrix file to find consensus sequences such as splice sites and promoter elements (Written by Ian Dodd, Department of Biochemistry at the University of Adelaide).
RODENT	For entering DNA sequence data. (Pharmacia).

STADEN Programs:

ANALYSEQ	Multiple analysis, including search for restriction enzyme sites, translation of DNA sequence into protein sequence plus many other DNA manipulation programs.
MWCALC	For calculating the molecular mass of protein sequences.
SEQFIX	An interactive line editor for keyboard entry and editing of sequences.
SEQH	Searches for local homology between nucleic acid sequences.

SEQHP	Searches for local homology between peptide sequences using Dayhoff similarity (Kanehisa) matrix.
SPCOMP	Used to find initial overlaps in nucleotide sequences.

GCG Programs:

GAP	Uses the algorithm of Needleman and Wunsch to find the alignment of two complete sequences that maximizes the number of matches and minimizes the number of gaps.
FOLD	Finds optimal secondary structures of RNA.
PEPTIDEPLOT	Multiple analyses, including molecular mass and pI calculation, search for epitopes and hydrophilic or hydrophobic domains.

GCG Programs for database searches:

FASTA	Compares whole databases with a nucleotide or peptide sequence, using the Lipman-Pearson algorithm. Sequences in the database with the greatest similarity are noted.
TFASTA	Uses Pearson and Lipman algorithms to search for similarity between a query peptide sequence and any group of nucleotide sequences. TFASTA translates the nucleotide sequences in the databases into all six frames before performing the comparison.

2.2.19 RNA PREPARATION

2.2.19a TOTAL RNA PREPARATION

Total RNA was isolated from organ tissue or cultured human fibroblasts as previously described (Chomczynski and Sacchi, 1987). Briefly, fibroblasts or finely minced tissue was placed in freshly prepared solution D (4 M guanidinium thiocyanate, 25 mM tri-sodium citrate, pH 7.0, 0.5% (w/v) sarkosyl and 0.1 M β-mercaptoethanol) at a ratio of 100 mg of cells or tissue per ml of solution D. Fibroblasts were passed through a syringe and 25 gauge needle and tissues were homogenized at room temperature in a glass teflon homogenizer. Homogenates were transferred to polypropylene centrifuge tubes and the following reagents were added per ml of solution D: 0.1 ml of 2 M sodium acetate, pH 4.0, 1 ml of water-saturated phenol and 0.2 ml of a chloroform/isoamyl alcohol mixture (49:1) with thorough mixing by inversion after the addition of each reagent. The final suspension was vortexed for 10 s and placed on ice for 15 min. Samples were centrifuged at 10,000 g for 20 min at 4°C to separate the RNA, which remained in the aqueous phase, from the DNA and protein, contained in the interface and phenol layer. The aqueous phase was transferred to a fresh tube, an equal volume of isopropanol was added and the solution incubated at -20°C for at least 1 hr. The RNA was precipitated by centrifugation at 10,000 g for 20 min at 4°C, resuspended in 300 µl of solution D and precipitated with 0.3 M sodium acetate and ethanol as described in Section 2.2.2. The RNA was dissolved in 0.5% (w/v) SDS or water by heating at 65°C for 10 min.

2.2.19b POLY(A)$^+$ RNA PREPARATION

Poly(A)$^+$ RNA was obtained from placental RNA as described by Kingston (1987). A siliconized glass pasteur pipette plugged with siliconized glass wool was washed with 10 ml of 5 M NaOH, and a slurry of oligo(dT) (0.5 ml dry oligo(dT) to 1 ml 0.1 M NaOH) poured into the column. The column was washed with 10 ml of DEPC-treated water and equilibrated with 10 to 20 ml of loading buffer (0.5 M LiCl, 10 mM Tris-HCl, pH 7.5, 1 mM EDTA and 0.1% (w/v) SDS). Approximately 2 mg of

total RNA in water was heated at 70°C for 10 min and 10 M LiCl added to a final concentration of 0.5 M. The RNA solution was passed through the column twice, and the column washed with 1 ml of loading buffer. Following this, the column was washed with 2 ml of middle wash buffer (0.15 M LiCl, 10 mM Tris-HCl, pH 7.5, 1 mM EDTA and 0.1% (w/v) SDS) and the RNA eluted into a fresh tube using 2 ml of 2 mM EDTA, 0.1% (w/v) SDS. The RNA was ethanol precipitated with 0.3 M sodium acetate as described in Section 2.2.2 and resuspended in an appropriate volume of TE buffer.

2.2.20 NORTHERN BLOT ANALYSIS

Northern blotting was carried out as described by Selden (1987). Total RNA (40 µg) and poly(A)$^+$ RNA (10 µg and 40 µg) was electrophoresed through a 0.8% (w/v) agarose/ 2.2 M formaldehyde gel. The gel was transferred to GeneScreen*Plus*™ nylon membrane and prehybridization, hybridization and washing performed as previously described (Section 2.2.13).

2.2.21 POLYMERASE CHAIN REACTION

PCR reagents were as described by Saiki *et al.* (1988), with the following modifications. The final concentration of deoxynucleotides was 400 µM, 200 ng of each primer was used (20-mers to 30-mers), and the final concentration of $MgCl_2$ was 2.5 mM in the buffer system provided by the manufacturer. Each PCR reaction contained 2 units of Taq polymerase and 10% (v/v) dimethyl sulfoxide was used in the reaction mixes where indicated. The PCR reaction was performed with denaturation at 94°C for the indicated time, annealing at the designated temperature, dependent on the specific primer pairs being used, for the indicated time, and elongation at 72°C for the appropriate time, depending on the length of PCR product expected (generally based on a synthesis rate of 1 kb per min). In general, 40 cycles were performed. Appropriate positive and negative controls were always included in each set of reaction.

CHAPTER 3

ISOLATION OF THE α-L-IDURONIDASE GENE

3.1 INTRODUCTION

As described in Chapter 1, the lysosomal enzyme α-L-iduronidase (α-L-iduronide iduronohydrolase, EC 3.2.1.76; IDUA) hydrolyses the non-reducing terminal α-L-iduronide glycosidic bonds in the glycosaminoglycans, heparan sulphate and dermatan sulphate (Neufeld and Muenzer, 1989; Hopwood, 1989). IDUA is a key lysosomal enzyme as it has served as a model for processing and maturation events undergone by lysosomal enzymes (Myerowitz and Neufeld, 1981; Sando and Neufeld, 1987; Taylor *et al.*, 1991). A deficiency of IDUA in humans results in the lysosomal storage disorder MPS-I (eponyms; Hurler, Hurler/Scheie and Scheie syndromes) which is inherited as an autosomal recessive disease and shows wide variation of clinical presentation. MPS-I patients are diagnosed by determination of the deficiency of IDUA activity in leucocytes or cultured skin fibroblasts. Multiple different mutant alleles at the *IDUA* locus are thought to be responsible for the spectrum of clinical phenotypes (Neufeld and Muenzer, 1989; Hopwood and Morris, 1990) but extensive biochemical characterization of the residual IDUA activity has only enabled discrimination between the extremes of clinical phenotypes (Hopwood and Muller, 1979; Ullrich *et al.*, 1981; Matalon *et al.*, 1983; Muller and Hopwood, 1984; Myerowitz and Neufeld, 1981; Taylor *et al.*, 1991; Ashton *et al.*, 1992). Consequently, the isolation of the IDUA gene was undertaken in this study in order to provide a DNA probe for molecular analysis of mutations in MPS-I patients and possibly for use in enzyme replacement and gene replacement therapy experiments in the canine or feline models of MPS-I (Shull *et al.*, 1982; Spellacy *et al.*, 1983; Haskins *et al.*, 1979).

Several groups have attempted the purification of human IDUA protein from a variety of different tissue sources with varying degrees of success, probably due to the low abundance of IDUA (Barton and Neufeld, 1971; Rome *et al.*, 1972; Schuchman *et al.*, 1984b; Clements *et al.*, 1985a; Clements *et al.*, 1989). Clements *et al.* (1985a) reported the 20,000-fold purification of human liver IDUA and the subsequent production of monoclonal antibodies to IDUA. Using one of these monoclonal antibodies in an immunoaffinity column allowed 172,000-fold purification to

homogeneity of human liver IDUA protein to provide sufficient IDUA for extensive peptide sequencing (Clements *et al.*, 1989). Immunopurified IDUA from human liver contained seven major bands of 74, 65, 60, 49, 44, 18 and 13-kDa on denaturing acrylamide gel electrophoresis (SDS-PAGE, Fig. 1.6). A model of proteolytic processing was proposed (Clements *et al.*, 1989) to attempt to link all seven polypeptides to IDUA (Fig. 1.7). The main evidence for this model was that the 65 and 60-kDa polypeptides shared a common N-terminal sequence and the 49 and 44-kDa polypeptides shared another N-terminal sequence (different to that of the 65/60-kDa N-terminal sequence).

With the availability of some IDUA N-terminal amino acid sequence data and the prospect of isolating large amounts of IDUA with an immunoaffinity column for further amino acid sequencing, it was decided to attempt to clone the gene for IDUA using the strategy shown in Fig 3.1. Briefly, the strategy involved generation of amino acid sequence data from immunopurified IDUA (Clements *et al.*, 1989) by both amino terminal (N-terminal) sequencing and amino acid sequencing of tryptic peptide fragments. Oligonucleotide probes are then designed to this peptide sequence and used to screen recombinant DNA libraries. Positive clones are sequenced to determine whether or not there is any colinearity between their DNA sequence and the amino acid sequence of the purified IDUA protein. This is a reliable method of gene isolation that has been used successfully to clone a number of other genes for lysosomal enzymes (Robertson *et al.*, 1988; Occhiodoro *et al.*, 1989; Wilson *et al.*, 1990; Litjens *et al.*, 1991). A major problem with this approach at the time this study commenced was that all of the peptides present in the IDUA preparation, and therefore the peptide sequence data, had not definitively been shown to be derived from IDUA. Because immunopurified IDUA from human liver contained seven major bands on SDS-PAGE, it was possible that some of these bands represented polypeptides that co-purified with IDUA. Thus oligonucleotide probes designed to peptide data from immunopurified IDUA may not be specific for the IDUA gene.

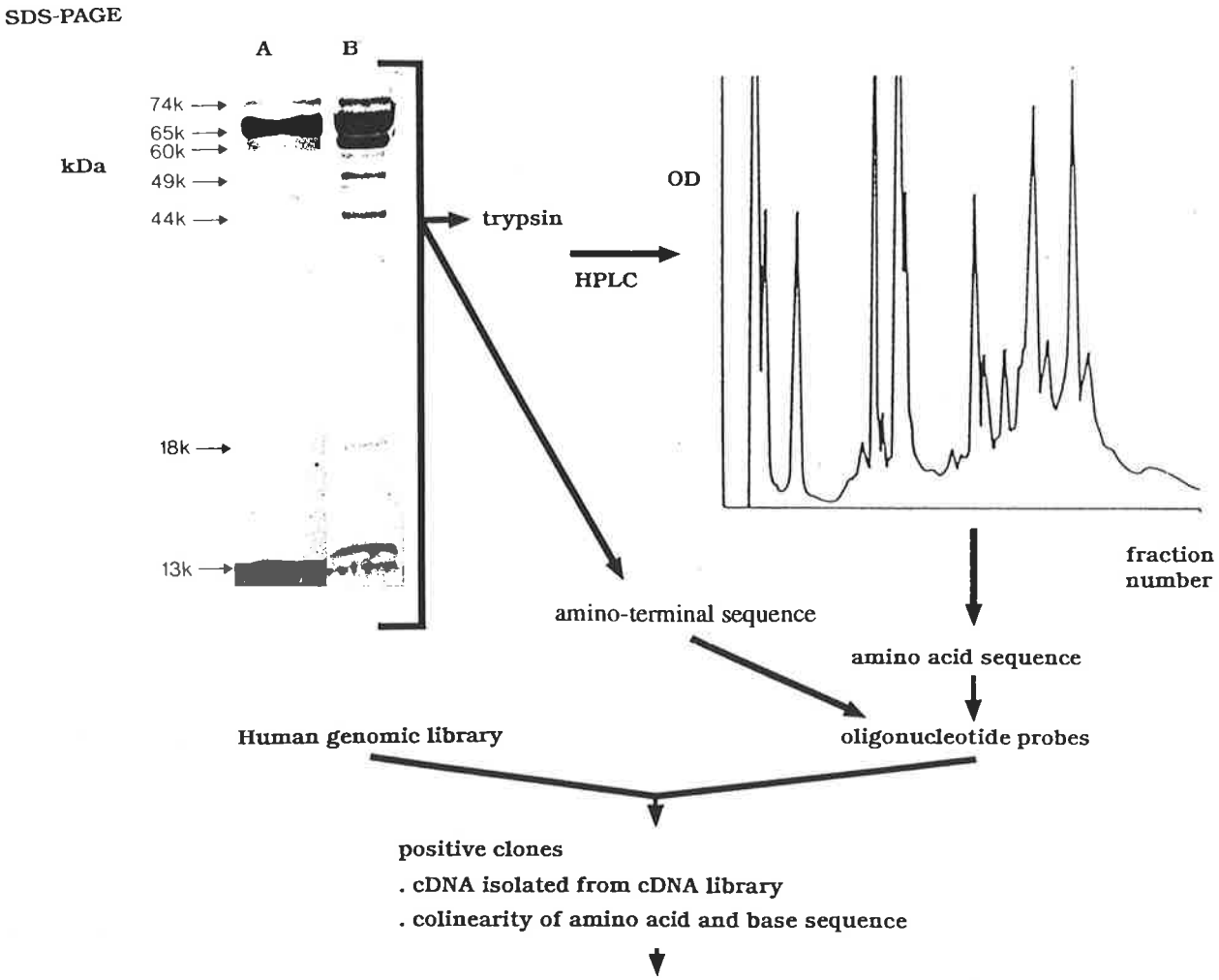


FIGURE 3.1. STRATEGY USED TO CLONE IDUA.

A flow diagram of the strategy used to clone the gene for IDUA.

It would also have been possible to attempt expression screening of λgt11 libraries (Young and Davis, 1985) as IDUA specific monoclonal antibodies were available. However, it was considered that this would be less likely to succeed as the low abundance of the IDUA protein implied the presence of very low levels of the IDUA mRNA and only one in six clones would be in the correct orientation and reading frame for expression screening. In addition, not all clones would be full-length cDNA and thus may miss or alter the antibody epitopes as most epitopes require correctly folded (native) IDUA. Efficient isolation of genes using antibody probes is also more technically demanding than the use of oligonucleotide probes (Young and Davis, 1985).

The polymerase chain reaction (PCR) had recently been modified to use a thermostable polymerase (Saiki *et al.*, 1988) and was being introduced into our laboratory. PCR techniques that could be used for gene isolation such as mixed oligonucleotide priming, amplification and cloning (MOPAC; Lee *et al.*, 1988), PCR amplification from cDNA libraries (Friedman *et al.*, 1988), anchored PCR (Frohman *et al.*, 1988) and ligation mediated PCR (Mueller and Wold, 1989) were attempted as they were described in the literature. The final proof that the full-length cDNA sequence encoding IDUA had been obtained by any of the methods discussed above would be by expression of IDUA enzymatic activity from a full-length cDNA clone.

The results presented in this chapter describe the isolation and characterization of the cDNA and gene for IDUA. Peptide data obtained from purified IDUA was used to design oligonucleotide probes that were used to screen cDNA and genomic libraries. A genomic clone that contained nucleotide sequence colinear to peptide sequence from purified IDUA was isolated. Chromosomal localization studies were performed by *in situ* hybridization to human metaphase chromosomes, Southern blot analysis of human-mouse cell hybrids, and immunocapture of human IDUA from human-mouse cell hybrids. These studies revised the localization of *IDUA* from chromosome 22 to chromosome 4p16.3. Screening of cDNA libraries led to the isolation of a number of IDUA cDNA clones, the longest of which was 1765-bp long. Regional mapping further localized the gene for IDUA to 1100-kb from the telomere of chromosome 4p and

1000-kb from either of the two most likely regions for the Huntington disease gene within a region defined by the cosmid clone A157.1. An oligonucleotide probe to the 74/13-kDa N-terminal peptide sequence of immunopurified IDUA was used to screen the cosmid A157.1 and an exon encoding an initiating methionine, a signal peptide, and the 74/13-kDa N-terminal sequence was identified. PCR amplification of IDUA mRNA sequence from cDNA led to the location of an additional exon and the eventual identification of all the coding sequence for IDUA, as demonstrated by expression studies. Alternative splicing was found to occur, producing two different mRNA species that were missing exons II and IV of the IDUA gene respectively. The gene for IDUA was split into 14 exons. The first 2 exons were separated by an intron of 566-bp followed a large intron of approximately 13-kb (intron 2). The last 12 exons were clustered in a 4.5-kb segment.

3.2 SPECIFIC MATERIALS AND METHODS

3.2.1 *In Situ* HYBRIDIZATION

Metaphase chromosomes were prepared from lymphocyte cultures of two normal males and *in situ* hybridization carried out according to a method previously described (Sutherland *et al.* 1988). Briefly, the IDUA probes were nick translated to a specific activity of approximately 1 x 10^8 cpm/μg with three tritiated nucleotides and hybridized to the metaphases from the two normal males; the cDNA probe at a concentration of 0.5 to 0.1 μg/ml for 14 to 18 days and the genomic probe at 0.01 μg/ml 13 to 20 days. All individual silver grains touching chromosomes were counted to determine the pattern of hybridization.

3.2.2 SOMATIC CELL HYBRID ANALYSIS

3.2.2a HYBRID CELL LINES

Wegroth-D2 was described by Geurts van Kessel *et al.* (1983). WAIVA was described by Deisseroth *et al.* (1977). Construction of the human-mouse cell lines designated CY was described by Callen (1986). CY12, CY13 and CY14 were

described by Callen *et al.* (1989). CY120 and CY123 were made from fibroblast lines containing translocation chromosomes (D. Callen, unpublished observations and Scott *et al.*, 1990). All CY cell lines contain a chromosome background of mouse A9 cells. The human chromosome content of all cell hybrids is shown in Table 3.1.

3.2.2b SOUTHERN BLOTTING

High molecular weight DNA (8-10 µg per track) from A9 mouse cells, normal human cells and the human-mouse hybrid cell lines were digested with *Eco*RI, electrophoresed through 0.8% (w/v) agarose and then transferred to Genescreen*Plus*™ nylon membrane. The IDUA cDNA probe was labelled with α-^{32}P-dCTP by random primer extension as described in Section 2.2.11b. Unincorporated nucleotides were separated from the probe using a 1 ml Sephadex G-50 column (Section 2.2.11c) and hybridization was carried out as described in Section 2.2.13b. The hybridized filter was washed once in 2 X SSC, 0.1% (w/v) SDS for 30 min and then once in 0.2 X SSC, 0.1% (w/v) SDS at 65°C for 20 min.

3.2.2c IMMUNOCAPTURE OF IDUA

Cells were grown to confluent monolayers in 75 cm^2 flasks in Basal Medium Eagles (modified) containing 10% (v/v) FCS. Cells were harvested and washed twice by centrifugation in 20 ml of PBS at 500 g at room temperature and finally resuspended in 100 µl of 0.02 M Tris-HCl, pH 7.0, 0.5 M NaCl. Leucocytes were harvested from 10 ml of whole blood as previously described by Kampine *et al.* (1966) and resuspended in 100 µl of 0.02 M Tris-HCl, pH 7.0, 0.5 M NaCl. All cells were lysed by freeze-thawing the cell preparations five times. The cell debris was removed by centrifugation in a microfuge twice at 12,000 g for 15 min. Aliquots of the suspension were removed for determination of total IDUA activity as described by Clements *et al.* (1985).

The monoclonal antibody designated Id1A has been previously described (Clements *et al.* 1985; Clements *et al.* 1989) and was immunopurified from hybridoma culture supernatants on a sheep anti-mouse immunoglobulin affinity column. Id1A (approximately 10 µg of antibody in 100 µl) was applied to the wells of a polyvinyl

ELISA plate and incubated overnight at 4°C. Residual reactive sites on the plates were blocked by incubation with a blocking solution containing 0.5 M NaCl, 0.02 M Tris-HCl pH 7.0 and 1% (w/v) ovalbumin for 2 hr at 20°C. The blocking solution was then decanted and extracts from the skin fibroblasts, leucocytes and the cell hybrids (30 µl) were applied and incubated overnight at 4°C. The wells were washed three times with 1% (w/v) ovalbumin solution and incubated with 30 µl of 4-methylumbelliferyl-α-L-iduronide as previously described (Clements *et al.* 1985) for 4 hr at 37°C. Relative fluorescence values were converted to pmol/min/mg of cell protein. Protein was determined by the method of Markwell *et al.* (1978).

3.2.3 cDNA SYNTHESIS

Total RNA (3 µg) from normal fibroblasts as prepared in Section 2.2.19a was added to a reaction mix containing 1 X Moloney murine leukemia virus (Mo-MLV) reverse transcriptase buffer (BRL), 40 units RNAsin, 500 ng of random octamers, 0.5 mM deoxynucleotides and 200 units of Mo-MLV reverse transcriptase to a final reaction volume of 50 µl. Incubation at 37°C for 1 hr was followed by hydrolysis of the RNA by the addition of 5 µl of 3 M NaOH and further incubation at 37°C for 30 min. The NaOH was neutralized 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 reaction used 5 µl of cDNA as template.

3.2.4 EXPRESSION OF IDUA

All IDUA expression constructs used the expression vector pRSVN.07 which drives expression of the insert from the Rous Sarcoma Virus LTR (Anson *et al.*, 1992b). CHO cells (strain DK1) were grown in F-12 nutrient media (Ham's), 10% (v/v) FCS, 100 µg/ml penicillin, 100 µg/ml streptomycin sulphate and 120 µg/ml kanamycin sulphate at 37°C in a 5% CO_2 atmosphere. CHO cells (1.2×10^7) were electroporated at 0°C using a BRL Cell-Porator at a pulse of 330 µF and 275 V in the presence 15 µg of the IDUA expression constructs. Cells were grown in non-selective media for 48 hr

TABLE 3.1. HUMAN CHROMOSOME CONTENT OF HUMAN-MOUSE CELL HYBRIDS.

The + and - symbols signify the presence or absence respectively of an entire human chromosome in the cell hybrids. All CY cell hybrids have a translocation chromosome containing the portions of the autosomes indicated. Expression of human IDUA in a cell hybrid (Ex) and the presence of a human specific band in Southern blots (So) is indicated by the + and - signs in the two columns at the extreme right hand side of the table.

TABLE 3.1.

hybrids	Human chromosome number 1	2	3	4	5	7	8	11	12	14	16	17	20	21	22	Ex	So
WAIVA	-	-	-	-	-	-	-	-	-	-	+	-	-	-	+	-	-
Wegroth-D2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-
CY12	+	+	-	-	-	+	+	-	+; q24 to qter	-	p11.2 to qter	-	-	+	-	-	-
CY13	q44 to qter	-	+	-	-	-	-	+	-	+	p13.11 to qter	+	+	+	+	-	-
CY123	-	-	-	-	+; q35.2 to qter	-	-	-	-	-	p12.2 to qter	-	-	-	-	-	-
CY120	-	-	-	pter to q25	-	-	-	-	-	-	q22 to qter	-	-	-	-	+	+
CY14	+	-	-	+; q31.1 to qter	+	-	-	-	+	+	p13.3 to qter	-	+	+	-	+	+

and then 1/20 and 1/100 dilutions of the electroporated cells were selected using 750 μg/ml of G418 sulphate. A bulk culture of resistant cells was extracted (Section 3.2.2c and Scott *et al.*, 1990) and assayed for IDUA activity with the fluorogenic substrate 4-methylumbelliferyl-α-L-iduronide (Clements *et al.*, 1985). The Bio-Rad protein assay was used to quantify the amount of protein in each sample according to the manufacturer's instructions. The monoclonal antibody Id1A was used for immunocapture and immunoquantification in conjunction with a polyclonal antibody (Ashton *et al.*, 1992) to assay the specific activity of the expressed IDUA (Clements *et al.*, 1989).

RESULTS AND DISCUSSION

3.3 INITIAL APPROACHES TO CLONING *IDUA*

3.3.1 PEPTIDE SEQUENCE AND OLIGONUCLEOTIDE PROBE DESIGN

Preparation of IDUA from human liver and transfer of IDUA to Imobilon P membrane for amino terminal sequencing was performed by Peter R. Clements of the Department of Chemical Pathology at the Adelaide Children's Hospital. Tryptic digestion and HPLC fractionation of IDUA was performed by Paul V. Nelson and C. Phillip Morris of the Department of Chemical Pathology at the Adelaide Children's Hospital. All peptide sequence was generated by Bresatec (Adelaide, S.A.).

All seven major polypeptides of IDUA (Clements *et al.*, 1989) were directly sequenced from their amino termini as described by Matsudaira (1987) and Le Gendre and Matusaira (1988). Tryptic peptide fragments were generated from 150 μg of immunopurified IDUA as described by Van Driel *et al.*, (1984) and Robertson *et al.*, (1988).

At the commencement of this study, amino-terminal sequence was available for the 65, 60, 49, 44 and 13-kDa IDUA peptide species. This amino acid sequence data showed that the 65 and 60-kDa peptides had a common N-terminal sequence, the 49

and 44-kDa peptides had another common N-terminal sequence, and the 13-kDa peptide another (Table 3.2). A consensus amino acid sequence for the 65 and 60-kDa N-termini was compiled from the 3 available sequences (65 and 60-kDa N-terminal, plus tryptic peptide 4, Table 3.3) and an oligonucleotide probe was designed that incorporated choices based on human codon usage (ID50, Fig. 3.2A).

After tryptic digestion and HPLC separation of immunopurified IDUA, nine major peptide peaks were sequenced (see Table 3.3). Five of the sequenced tryptic peptides had more than one amino acid sequence recorded, indicating that more than one peptide species was present (peptides 1, 2, 3, 5 and 7). The signal recorded during amino acid sequencing from the two peptides present in these single tryptic peptide peaks was normally at different intensities. That is, two amino acids recorded at a single position were present at different concentrations, representing the levels of each peptide species present in the tryptic peptide peak. This was normally consistent for the length of the tryptic peptides allowing at least partial differentiation of the two peptide sequences present. One tryptic peptide (peptide 4) was the same as the 65/60-kDa amino-terminal sequence and one of the two tryptic peptide species present in peak 3 was contained within the 49/44-kDa amino-terminal sequence. Peptide 8 was initially chosen for oligonucleotide probe design due to its length (36-amino acids), the low codon degeneracy of the amino acids and its apparent homogeneity compared to other peptide sequences. Assuming that the undetermined amino acid at position 16 of peptide 8 was a glycosylated Asn, five oligonucleotide probes were designed for library screening (Fig. 3.3). In addition, an oligonucleotide probe was also designed to peptide 9 (ID9, Fig. 3.2B). Design, synthesis and purification of oligonucleotide probes is described in Section 2.1.13.

3.3.2 RECOMBINANT DNA LIBRARY SCREENING

Initial cDNA library screening was performed with the 20-mer oligonucleotide probes ID5 and ID6 which were designed to the amino acid sequence of tryptic peptide 8 (Fig. 3.3). Inosine can bind to all 4 nucleotides present in DNA and thus both

TABLE 3.2. N-TERMINAL AMINO ACID SEQUENCES OF IDUA.

The amino acids are listed in their single letter codes. The length of the reported sequences is in brackets at the end of each sequence. An amino acid followed by a "?" denotes a tentative assignment of a residue. "X" denotes an unassigned amino acid. Amino acids separated by a "/" are options at the same position

TABLE 3.2.

74 kDa N-terminus	V/G/L E Q F F/V V G D/R D/R G/A/R R R A X X P (16)
65 kDa N-terminus	S/G X Y X F T X L D G Y L D D L R V/E N Q I L P G F (24)
60 kDa N-terminus	L S Y X F X T L X G Y L X L L I E N Q A L P G F E (25)
49 kDa N-terminus	A G V R L D Y I L S V? R K G A L? S? (17)
44 kDa N-terminus	A G V R L D Y I S L H R K G A R S H? (18)
18 kDa N-terminus	L P/S Y X F L R L F/D Y Y L V S L R E N Q/K (19)
13 kDa N-terminus	S/E A/R R/P A/L L/P V/G Q V D A A R A L P/G S/C L/S/C R D? V? (20)

TABLE 3.3. AMINO ACID SEQUENCES OF IDUA TRYPTIC PEPTIDES.

As for Table 3.2, the amino acids are listed in their single letter codes. The length of the reported sequences is in brackets at the end of each sequence. Sequences named with an apostrophe were reported as secondary sequences and in some cases the two sequences have been resolved by the cDNA sequence. An amino acid followed by a "?" denotes a tentative assignment of a residue. "X" denotes an unassigned amino acid. Amino acids separated by a "/" are options at the same position.

TABLE 3.3.

Peptide 1	A L P L T Q G Q L V L V W S D E H V G S K (21)
Peptide 1'	K P S T F N L F V F S P D T G A V S G S Y R (22)
Peptide 2	F A D T P I Y N D E A D P L V G X S X P Q Q? X R (24)
Peptide 2'	S T G F C P X L P H S Q X X Q Y X L X X X X Q X N X A Y (28)
Peptide 3	E Q V F E (5)
Peptide 3'	L D Y I N (5)
Peptide 4	L S Y X F T H L D G Y L D L L (15)
Peptide 5	P G P F S D P V P Y X E (12)
Peptide 5'	Y L D N G L C S P D G E (12)
Peptide 6	Y G L A H V S K (8)
Peptide 7	A L D Y W A R X X X (10)
Peptide 7'	X X X X I L E Q E K (10)
Peptide 8	E/W/G N F E T W N E P D H H D F D X V S M T M Q G F L N Y Y D A X X E G (34)
Peptide 9	K P V L T A M G L L A L L D E E Q L W A E V V? Q A G T V X D S G H T X G (36)

FIGURE 3.2. OLIGONUCLEOTIDE PROBES TO THE IDUA 65-60 kDa N-TERMINUS AND PEPTIDE 9.

A. The consensus N-terminal amino acid sequence compiled from the 65 and 60-kDa N-terminal sequence and the sequence of tryptic peptide 4 of IDUA and all possible mRNA coding sequences are shown. The sequence of the 50-mer oligonucleotide probe, ID50, is aligned with the mRNA sequence and is presented below. The asparagine (N) residue at position 4 (underlined) was hypothesized to form a consensus glycosylation site.

B. The amino acid sequence of IDUA tryptic peptide 9 and the sequence of the oligonucleotide probe designed to peptide 9, ID9.

A: Consensus amino acid sequence and possible mRNA sequence for the 65/60-kDa N-terminus:

```
 L   S   Y   N   F   T   H   L   D   G   Y   L   D   L   L   R   E   N   Q   ?   L   P   G   E
CUN UCN UAU AAU UUU ACN CAU CUN GAU GGN UAU CUN GAT CUN CUN CGN GAA AAU CAA ??? CUN GAU GGN GAA
UUA AGU   C   C   C       C UUA   C       C UUA   C UUA UUA AGA   G   C   G     UUA   C       G
  G   C                       G               G       G   G   G                     G
```

Probe sequence:

```
     3'                                                                 5'
ID50 ATG TTG AAG TGG GTG GAC CTG CCG ATG GAC CTG GAT GAC GCC CTC TTG GT
```

B: Amino acid sequence and possible mRNA sequences for peptide 9:

```
 K   P   V   L   T   A   M   G   L   L   A   L   L   D   E   E   Q   L   Y   A   E   V   Q   A   G   T   V
AAA CCN GUN CUN ACN GCN AUG GGN CUN CUN GCN CUN CUN GAU GAA GAA CAA CUN UGG GCN GAA GUN CAA GCN GGN ACN GUN
  G         UUA               UUA UUA     UUA UUA   C   G   G   G UUA                   G
            C                 C   C       C C                   C
```

Probe sequence:

```
       3'                    5'
ID9 CTA CTT CTT GTT AAI ACC CG
      G   C   C   C G
```

FIGURE 3.3. OLIGONUCLEOTIDE PROBES TO IDUA PEPTIDE 8.

The amino acid sequence of IDUA tryptic peptide 8 and all possible mRNA coding sequences are shown. The sequences of all oligonucleotide probes, except for ID7, designed to peptide 8 are aligned with the mRNA sequence and are presented below. The asparagine (N) residue at position 16 (underlined) was hypothesized to form a consensus glycosylation site. The sequence of ID7 is the same as shown for ID8 except that a mixture of all 4 nucleotides replace the inosine residue at position 12 of ID8.

Amino acid sequence and possible mRNA sequences for peptide 8:

```
  ?  N  F  E  T  W  N  E  P  D  H  H  D  F  D  N  V  S  M  T  M  Q  G  F  L  N  Y  Y  D  A
??? AAU UUU GAA ACN UGG AAU GAA CCN GAU CAU CAU GAU UUU GAU AAU GUN UCN AUG ACN AUG CAA GGN UUU CUN AAU UAU UAU GAU GCN
  C  C  G        C  G     C  C  C  C  C  C  C     AGU              G     C UUA  C  C  C  C
                                                  C                      G
```

Probes:

```
3′                    5′                                                3                    5′
TTA AAA CTT TGI ACC TTA CT ID8                                          TAC TGI TAC GTT CCI AAA GA ID6
  G  G  C        G                                                              C     G A

            3′                    5′
            ACC TTA CTT GGI CTA GTA GT ID5
                  G  C     G  G

5′                                                                  3′
AAC TTC GAG ACC TGG AAC GAG CCC GAC CAC CAC GAC TTC GAC AAC GT ID47
```

oligonucleotide probes incorporated inosine residues at wobble base positions where all 4 nucleotides could be present. It was believed that this would help increase the specific activity of the correct probe sequence present in the mixed oligonucleotide by reducing the number of different oligonucleotide sequences present. As discussed in Section 2.1.13a, it was attempted to obtain cDNA clones that were co-positive to ID5 and ID6. cDNA libraries derived from various tissue types (see Section 2.1.9) were screened, in part because of their availability and in part because the low abundance IDUA protein had been found to be present in relatively high amounts in those tissues (e.g. lung fibroblast, Schuchman *et al.*, 1984b; testis, Dr. E. Neufeld, pers. commum.). Initially, 4 different human cDNA libraries (derived from lung fibroblasts, testis, placenta and colon) were screened with the oligonucleotide probes ID5 and ID6, with the 37 positive clones detected in the first round of screening with ID5 not being positive to ID6. Screening of the genomic DNA library with ID5 and ID6 gave a comparable result. It was suspected that ID6 was flawed as this probe only gave an even background signal and no plaque specific signal with low stringency washing when used for primary screening.

Three further 20-mer oligonucleotide probes, two designed to peptide 8 (ID7, ID8, Fig. 3.3) and one to peptide 9 (Fig. 3.2B) were also unsuccessfully used. ID7, was designed to the first 7 residues of peptide 8 and incorporated all nucleotide options at wobble base positions, thus resulting in a mixture of 64 different oligonucleotide sequences (Fig. 3.3). ID8 was made using the same sequence as ID7 except that an inosine residue was incorporated at the wobble base of the threonine (T) residue (residue 5, Fig. 3.3), resulting in a mixture of 16 sequences instead of 64 thus increasing the specific activity of the correct sequence in the probe mixture. ID9 contained a mixture of 32 oligonucleotide sequences (Fig. 3.2B). Use of all the above probes resulted in no primary positive clones, and screening of second round filters, containing positives to ID5, with these probes were also found to be negative.

Although the approach of using 2 20-mer oligonucleotide probes to isolate cDNA clones had been used successfully in this laboratory (Robertson *et al.*, 1988,

Litjens *et al.*, 1991), it was decided to design the longest probes possible by incorporating "best choice" nucleotides at wobble base positions based on human codon usage tables. Thus a 50-mer, ID50, was designed to the consensus N-terminal sequence of the 65 and 60-kDa polypeptides of immunopurified IDUA (Fig. 3.2A). However screening of all filters with ID50 at varying probe concentrations resulted in consistently strong background signal such that plaque specific signals were not able to be observed . A 47-mer was designed using the same approach as ID50 to peptide 8 (ID47, Fig. 3.3). Again the four previously mentioned cDNA libraries were screened using this 47-mer probe without success.

Great care had been taken when preparing filters for primary library screens. Libraries were plated at the relatively low density of 40,000 to 55,000 pfu per 150mm plate and the libraries were transferred to filters only if the plaques were large and clear. All probes used were purified by HPLC and their purity and ability to kinase analysed on denaturing polyacrylamide gels. IDUA is a very low abundance protein (Clements *et al*, 1989) with a long half life (Taylor *et al.*, 1991) and it was therefore assumed that the lack of success in the cDNA library screening described above may have been a reflection of the low levels of IDUA mRNA present in the tissues from which the cDNA libraries were derived. To overcome this problem, it was decided to screen genomic DNA libraries.

In a previous report, Schuchman *et al.* (1984a) had localized the gene for human IDUA to chromosome 22pter to 22q11. Thus, a human chromosome 22 specific genomic DNA library was obtained and screened using the 7 oligonucleotide probes described above, also without success. Following this result, 500,000 clones of the EMBL3 human genomic library were screened using ID47 (Fig. 3.3) resulting in 8 positives clones. Screening of second round filters with ID50, ID5, ID8 and ID9, showed that one of these clones, λID475, was positive to all of these IDUA oligonucleotide probes.

3.3.3 MOPAC CONFIRMATION OF λID475

Partly due to frustration at the lack of success of library screening with oligonucleotide probes, and while work on library screening described in the previous section was in progress, new methods were sought to attempt to isolate the gene for IDUA. A PCR based method called mixed oligonucleotide priming, amplification and cloning (MOPAC) had been recently described (Lee *et al.*, 1988) and was applied to this project. Briefly, mixed sense strand and antisense oligonucleotides derived from the 5' and 3' end of a long peptide sequence (~30 amino acids) are used to PCR amplify the cDNA of interest to produce a long (~100-bp), double stranded probe containing internal sequence derived from the gene of interest. Restriction endonuclease sites designed on the ends of the oligonucleotide primers are then be used to clone the PCR product obtained in order to confirm its identity by nucleotide sequencing. The nucleotide sequence should encode unique amino acid sequence data not represented in the primers. This cloned PCR product may then be used as a probe for cDNA library screening following oligolabelling to high specific activity.

The design of this MOPAC experiment for IDUA is shown in Fig. 3.4. At the time of isolation of λID475, the approach of MOPAC cloning had just produced the expected 102-bp IDUA PCR product using cDNA as the template (data not shown). As the MOPAC experiment was designed to amplify sequences encoding peptide 8 and since λID475 was positive to probes derived from peptide 8, it was assumed that the MOPAC experiment could be used for a rapid confirmation to show that λID475 contained nucleotide sequence colinear with peptide 8. PCR amplification of approximately 1 ng of purified DNA from λID475 using primers IDP1 and IDP2 was performed by manual cycling between 3 heating blocks for 30 cycles. Denaturation was at 94°C for 2 min, annealing was at 42°C for 2 min and extension was at 60°C for 2 min. This PCR reaction using λID475 DNA as template resulted in a unique product of 218-bp instead of the expected 102-bp product (Fig. 3.4). The PCR product was cloned into M13mp19 and sequence data revealed that the product was exactly colinear with the amino acid sequence of peptide 8, but was interrupted by a 115-bp intron (Fig. 3.4).

This confirmed that λID475 contained coding sequence that was representative of peptide sequence in the purified IDUA preparation.

3.4 SUBCLONING FROM λID475 AND ISOLATION OF A cDNA CLONE FOR IDUA

In order to identify a small fragment suitable for subcloning and sequencing, and for subsequent screening of cDNA libraries, lambda DNA was prepared from λID475, digested with multiple restriction endonucleases and combinations of restriction endonucleases and subjected to Southern analysis using ID47 as a probe. A 1.6-kb *Pst* I fragment was identified (Fig. 3.5), isolated and subcloned into both M13mp18 and pUC19. Nucleotide sequencing of the 1.6-kb *Pst* I fragment revealed that the previously identified peptide 8 sequence contained in the MOPAC PCR product was present. This indicated that the correct fragment had been isolated.

The 1.6-kb *Pst* I fragment was isolated from its pUC19 subclone (pID89) and used to screen approximately 10^6 recombinants from each of the lung fibroblast, placental, testis and colon cDNA libraries. A single cDNA clone containing a 729-bp insert (λRPC1) was obtained from the λgt10 random-primed colon cDNA library. The nucleotide sequence of the insert of λRPC1 was determined (Fig. 3.6) and was found to contain sequence that was colinear with six peptide sequences from immunopurified IDUA, including the 49/44-kDa N-terminal sequence. This was further evidence that the cDNA and genomic clones obtained encoded IDUA. However, the λRPC1 insert ended just 5' to the nucleotide sequence encoding tryptic peptide 6 and within the sequence encoding tryptic peptide 9 at the 3' end. It was significant that one of the N-terminal peptide sequences from immunopurified IDUA (49/44-kDa) was identified in the cDNA clone as there was greater confidence that these N-terminal sequences, rather than the tryptic peptide sequences, were truly representative of IDUA. The previous lack of success in screening for IDUA cDNA clones was possibly due to both the low sensitivity of the oligonucleotide probes and the apparent low abundance of IDUA mRNA.

FIGURE 3.4. DESIGN OF THE IDUA MOPAC EXPERIMENT.

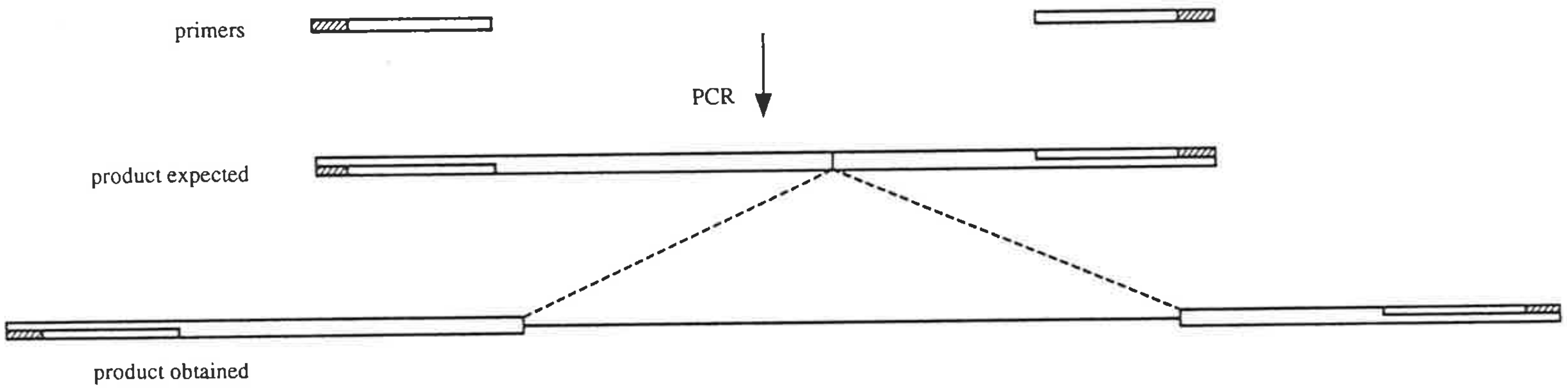
The amino acid sequence of IDUA tryptic peptide 8 and all possible mRNA coding sequences are shown. The sequences of the two oligonucleotide primers designed to peptide 8 with the restriction sites used to clone the PCR product are aligned with the mRNA sequence and are presented below. A diagrammatic representation of the experiment is shown underneath. Note that the 102-bp PCR product was expected from cDNA template but a 218-bp product was obtained when using the EMBL3 genomic clone λID475 as template. This PCR product was cloned and sequenced to show that the clone contained the expected coding sequence (open boxes) separated by a 115-bp intron (line). The nucleotide sequence of the 218-bp MOPAC PCR product is also shown at the bottom of the figure. The amino acid sequence of peptide 8 encoded is shown in single letter code above the nucleotide sequence. The coding regions are shown in uppercase letters and the 115-bp intron is shown in lower case letters. Both the 5' and 3' splice junctions match consensus splice site sequences.

Amino acid sequence and possible mRNA sequences:

```
  N   F   E   T   W   N   E   P   D   H   H   D   F   D   N   V   S   M   T   M   Q   G   F   L   N   Y   Y   D   A
AAU UUU GAA ACN UGG AAU GAA CCN GAU CAU CAU GAU UUU GAU AAU GUN UCN AUG ACN AUG CAA GGN UUU CUN AAU UAU UAU GAU GCN
  C   C   G           C   G       C   C   C   C   C   C   C       AGU             G     C UUA   C   C   C   C
                                                                    C                       G
```

MOPAC Primers:

```
5'                                 3'                                                     3'                         5'
CGGATCCG AAT TTT GAA ACC TGG AA IDP1                                               IDP2 TTA ATA ATA CTA CG CCTTAAGG
 BamHI     C   C   G   G                                                                  G   G   G   G     EcoRI
```



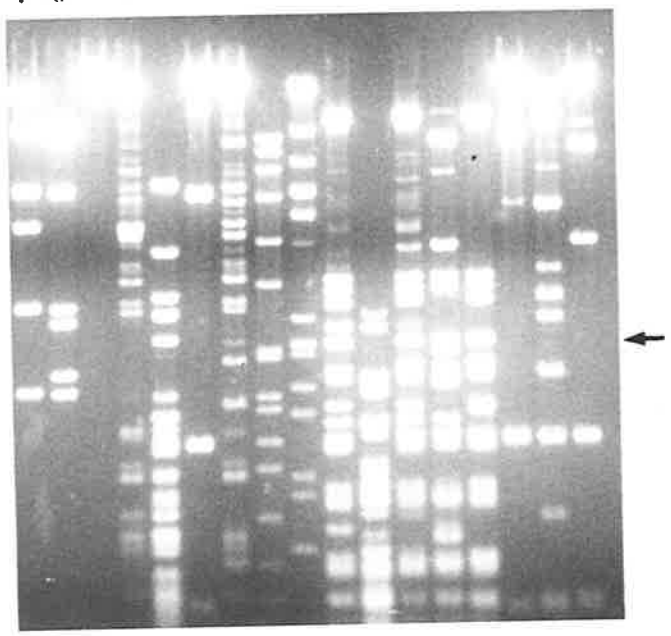

```
                                      peptide 8
 BamHI    N  F  E  T  W  N  E  P  D  H  H  D  F  D  N  V  S  M  T  M  Q
CGGATCCG AACTTCGAGACGTGGAATGAGCCAGACCACCACGACTTTGACAACGTCTCCATGACCATGCAAGgtgtgcaccgcttcctggggtcctgcccggctgaaagg
          10        20        30        40        50        60        70        80        90        100       110

      115-bp intron
                                                                              G  F  L  N  Y  Y  D  A  EcoRI
gggcagaggaaggcaggagcagaggctaagccgctcatccccagggcaggtgtagacgcagtgctcccccggcccagGCTTCCTGAACTACTACGATGC GGTTAACC
     120       130       140       150       160       170       180       190       200       210
```

FIGURE 3.5. RESTRICTION ENDONUCLEASE DIGESTION AND SOUTHERN BLOTTING OF λID475.

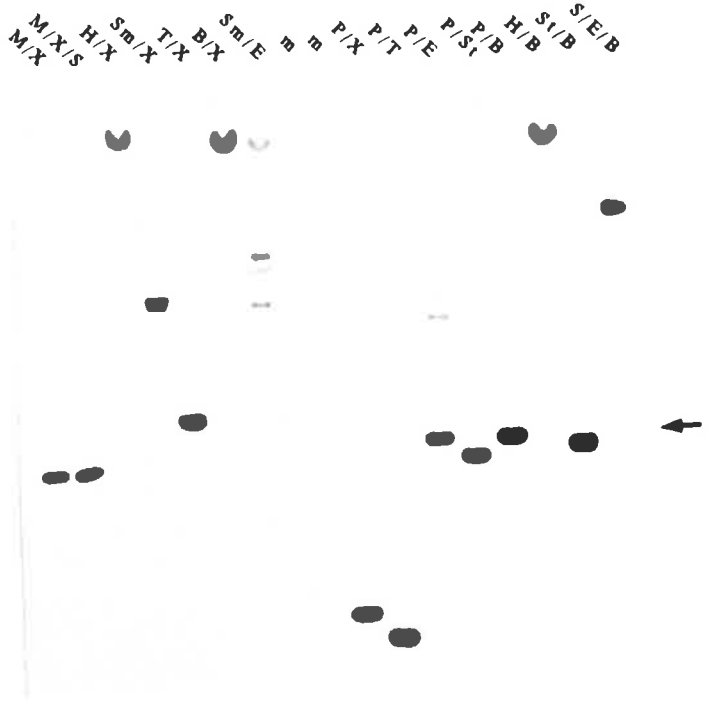
A. TAE agarose gel (0.8%) electrophoresis of λID475 restriction endonuclease digested DNA. Tracks are marked with the restriction enzymes used as follows; M=*Mlu* I, X=*Xho* I, S=*Sal* I, H=*Hin*dIII, Sm=*Sma* I, T=*Taq* I, E=*Eco*RI, P=*Pst* I, St=*Stu* I. The Sm/E reaction was only partially digested. The two marker tracks (m) are λ DNA digested with *Hin*dIII/*Eco*RI and *Hin*dIII and SppI phage digested with *Eco*RI respectively. The position of the 1.6-kb *Pst* I fragment identified in B is marked by an arrow.

B. A Southern blot of the gel in A hybridized to the oligonucleotide probe ID47. The 1.6-kb fragment which hybridizes to ID47 in *Pst* I digests (see tracks P/E and P/B) was isolated for sequence analysis and use as a probe for *in situ* hybridization studies and cDNA library screening (see text).

A.


M/X
M/X/S
H/X
Sm/X
T/X
B/X
Sm/E
m
m
P/X
P/T
P/E
P/St
P/B
H/B
St/B
S/E/B

B.


M/X
M/X/S
H/X
Sm/X
T/X
B/X
Sm/E
m
m
P/X
P/T
P/E
P/St
P/B
H/B
St/B
S/E/B

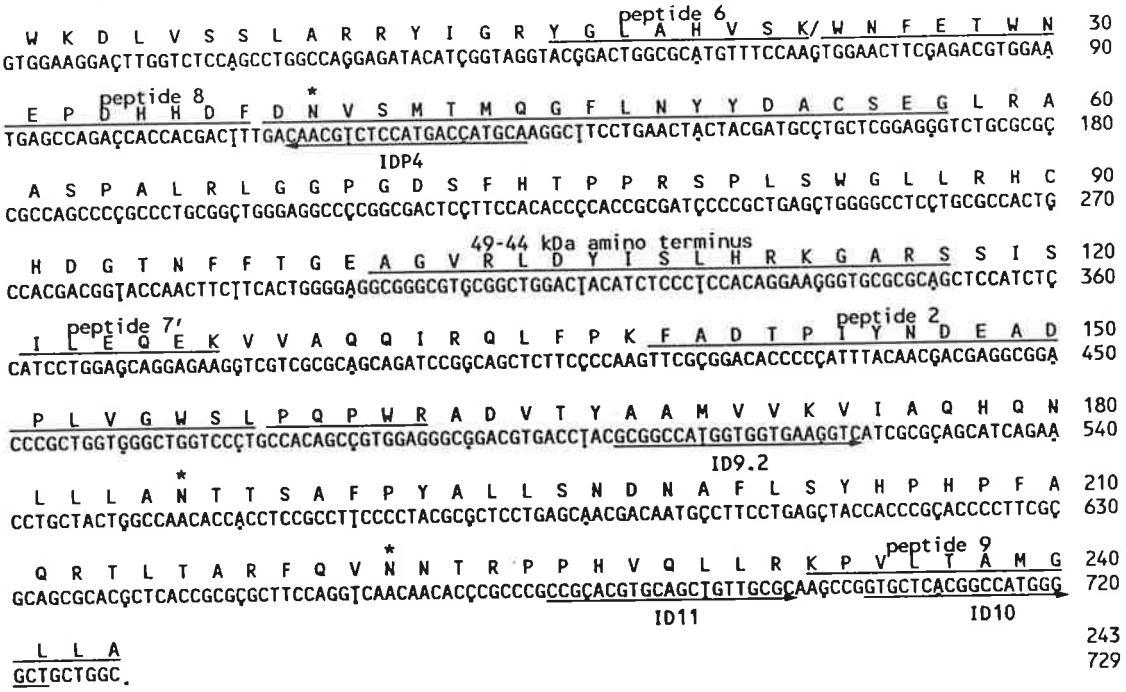


FIGURE 3.6. SEQUENCE OF THE IDUA cDNA CLONE, λRPC1.

Nucleotide sequence for the IDUA cDNA clone λRPC1, and the deduced amino acid sequence of the protein. Amino acid sequence is shown in single letter code above the cDNA sequence. Nucleotide and amino acid numbers are shown in the right margin. Nucleotides are marked at intervals of ten with a dot. Amino acids colinear with either amino-terminal peptide data or tryptic peptides are underlined and named above the sequence. Potential N-glycosylation sites are marked by asterisks. Oligonucleotides used in the studies for PCR amplification of IDUA cDNA from cDNA libraries, by ligation mediated PCR, by anchored PCR and PCR to peptide derived oligonucleotides (see Section 3.6.2) are underlined below the nucleotide sequence with the arrows indicating either sense (→) or antisense (←).

3.5 CHROMOSOMAL LOCALIZATION OF *IDUA*

In a previous report, Schuchman *et al.* (1984a) localized the gene for human IDUA to chromosome 22pter to 22q11 by immunoprecipitation of human IDUA activity from human-mouse cell hybrids using a rabbit anti-human IDUA polyclonal antiserum, and by detection of cross reactive immunological material (CRIM) in human-mouse cell hybrids using a different polyclonal antiserum. Both antisera were against the purified low uptake form of human IDUA (Schuchman *et al.*, 1984b; 1984c). In an attempt to confirm the identity of the 729-bp cDNA (λRPC1) and the genomic clone (λID475), chromosomal localization of *IDUA* was attempted with the expectation that the clones would localize to chromosome 22 if they were derived from IDUA.

3.5.1 *In Situ* HYBRIDIZATION

In Situ hybridizations were performed by Helen J. Eyre and Elizabeth Baker of the Department of Cytogenetics and Molecular Genetics at the Adelaide Children's Hospital.

To define the chromosomal localization of the IDUA gene, the 729-bp cDNA insert of λRPC1 and the genomic 1.6-kb *Pst* I fragment (pID89 insert) subclone were used for *in situ* hybridization to metaphase chromosomes from two normal males. The distribution of grains obtained using the genomic DNA probe on 52 metaphases is shown in Fig. 3.7A. Of 158 grains, 25 (16%) were over the short arm of chromosome 4. A similar distribution was obtained when the cDNA probe was used, with 15% of grains localizing over the short arm of chromosome 4 in 25 metaphases (data not shown).

For the cDNA probe, scoring was then undertaken on 30 additional metaphases, with chromosomes at the 600-1000 band level of resolution, which had grains over the short arm of chromosome 4. The combined data (Fig. 3.7B) demonstrate a localization of the IDUA cDNA probe to the short arm of chromosome 4, with most grains distributed over 4p15.3 to 4pter and a peak at 4p16.3. *In situ* hybridization of the

genomic clone showed an almost identical result from 31 high resolution metaphases with signal on 4p. Hybridization of both probes to metaphase spreads from another unrelated normal male demonstrated an identical localization (data not shown). It is significant to note that there was no specific signal over chromosome 22 (Fig. 3.7A). If these are clones encoding IDUA, this indicates that the most likely chromosomal localization of the human IDUA gene is 4p16.3 and not chromosome 22 as previously reported (Schuchman *et al.*, 1984a).

3.5.2 SOMATIC CELL HYBRID ANALYSIS

The assignment of the structural gene for human IDUA to chromosome 4 was based on the demonstration that both a cDNA and genomic clone for IDUA were localized to chromosome 4p15.3 to 4p16.3 by *in situ* hybridization to human metaphase spreads. As this result conflicted with published data assigning *IDUA* to chromosome 22 (Schuchman *et al.* 1984a), it was possible that these clones were not derived from the IDUA gene. Thus, the localization to chromosome 4 was confirmed by Southern blot analysis of human-mouse hybrids and by direct measurement of expressed IDUA activity in the human-mouse hybrids using an enzyme specific substrate assay. A monoclonal antibody specific to human IDUA was used in order to distinguish human IDUA from mouse IDUA. The panel of somatic cell hybrids (see Table 3.1) was selected to confirm, not to independently localize *IDUA*, the chromosomal localization of *IDUA* to chromosome 4 and exclude localization to chromosome 22.

3.5.2a SOUTHERN ANALYSIS

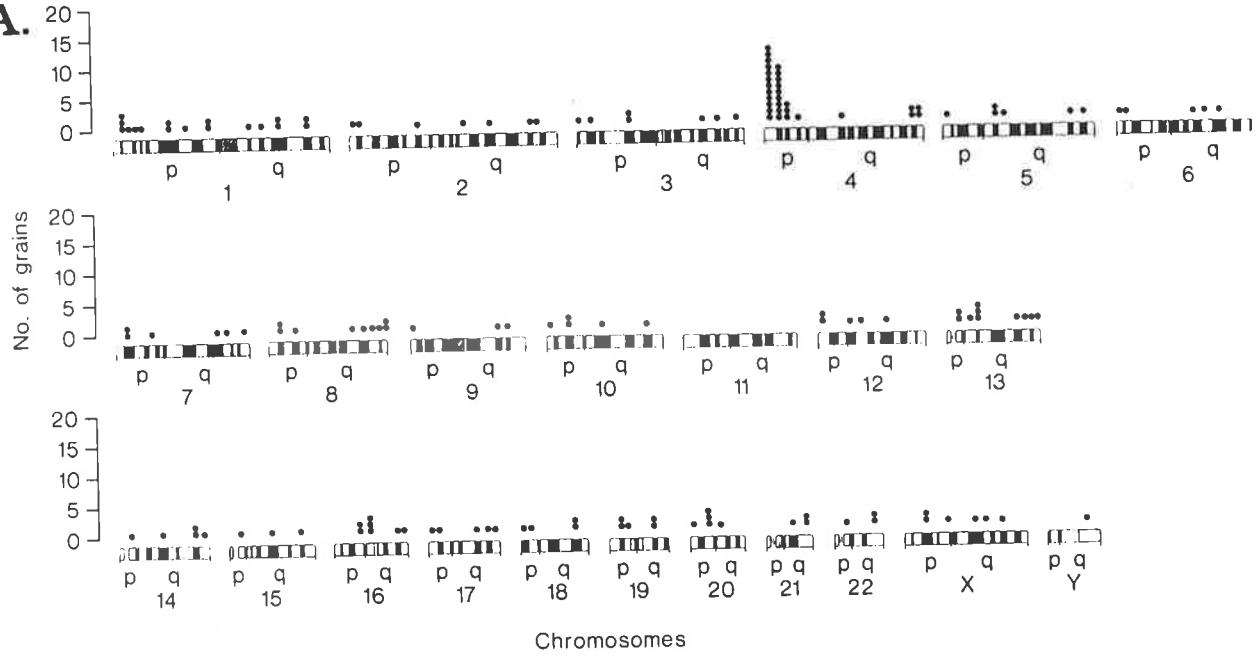
The localization of the IDUA gene to chromosome 4 was confirmed by performing Southern blot analysis on the hybrid cell lines shown in Table 3.1 with the cDNA clone. The gene probe hybridizes to a single 8-kb *Eco*RI band in mouse DNA and to a single 23-kb *Eco*RI band in human DNA. The 23-kb *Eco*RI human specific band is present only in the cell lines CY14 and CY120 (Fig. 3.8). CY14 contains a translocation chromosome, der(16)t(4;16)(q31.1;p13.3), an entire chromosome 4 and five other human chromosomes. CY120 contains the translocation chromosome

FIGURE 3.7. *In situ* HYBRIDIZATION OF IDUA TO METAPHASE CHROMOSOMES AND CHROMOSOME 4p.

A. Ideograms of G-banded human chromosomes showing the distribution of 158 silver grains from 52 metaphases after hybridization to the ^{3}H-labelled IDUA genomic DNA. Note 16% of the silver grains localize over chromosome 4p.

B. An ideogram of G-banded chromosome 4 illustrating the distribution of silver grains in 30 metaphases with chromosomes at the 600-1000 band level of resolution after hybridization to the IDUA cDNA probe. Note that most of the silver grains are distributed over bands 4p15.3 to 4p16.3 with a peak at 4p16.3.

A.


No. of grains
20
15
10
5
0
p q
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
X
Y
Chromosomes

B.

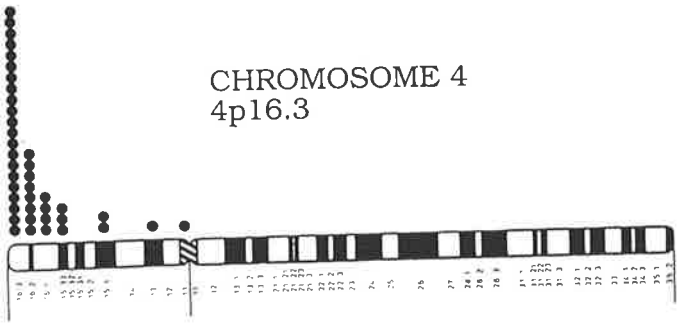
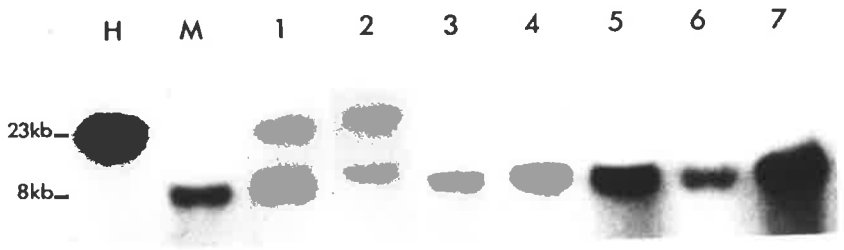

CHROMOSOME 4
4p16.3

FIGURE 3.8. SOUTHERN BLOT ANALYSIS OF HUMAN MOUSE CELL HYBRIDS.

Southern blot analysis of EcoRI digested DNA from a normal human male (lane H), mouse A9 cells (lane M) and the human-mouse cell hybrids; CY14, CY120, CY123, CY12, CY13, Wegroth, and WAVIA in lanes 1 to 7 respectively. The size (in kilobases) of the human and mouse EcoRI fragments is indicated on the left. The presence or absence of human chromosomes 4 and 22 in the cell hybrids is indicated under the Southern. Note that only cell lines containing some or all of human chromosome 4, CY14 and CY120 (Table 4.1), have the human specific band for *IDUA*. For convenience, lane 2 has been photographically inserted from another position on the samc gel.



	H	M	1	2	3	4	5	6	7
CH 4 :	+	-	+	+	-	-	-	-	-
CH 22 :	+	-	-	-	-	-	+	+	+

der(4)t(4;16)(q25;q22) as the only human chromosome present. The other cell lines in the panel contain human chromosomes that allow exclusion of the presence of a human *IDUA* specific band being due to any of the human chromosomes present in CY14, other than chromosome 4 (see Table 3.1). Thus human *IDUA* must be on human chromosome 4 and in particular in the region included in the CY120 translocation chromosome, 4pter to 4q25. Wegroth-D2, WAIVA and CY13 cell lines contain chromosome 22 but do not produce human specific bands in Southern blot analysis with *Eco*RI (Fig. 3.8) or using three other restriction enzymes (data not shown).

3.5.2b IMMUNOCAPTURE

This work was performed by Lesley J. Ashton and Doug A. Brooks of the Department of Chemical Pathology at the Adelaide Children's Hospital.

The localization of *IDUA* to chromosome 4 was directly demonstrated by the immunocapture of human IDUA with the monoclonal antibody Id1A (Clements *et al.*, 1989) on human-mouse cell hybrids (Table 3.1). When used against mixed leucocytes, Id1A captures 44% of the IDUA enzyme activity present. In mouse A9 cells Id1A captures less than 1% of IDUA activity, which is the background capture level, probably due to a small amount of cross reactivity of Id1A with mouse IDUA protein, showing that Id1A is specific for human IDUA. Significant human IDUA activity was detected only in the hybrid cell lines CY14 and CY120 (Table 3.4). Thus, human IDUA activity is being expressed from human chromosome 4, in particular from the region included in CY120, 4pter to 4q25, and no discordant chromosomes were recorded. This is consistent with the *in situ* hybridization and Southern blot results using the IDUA genomic and cDNA clones, and confirms that *IDUA* is located on chromosome 4. Wegroth-D2, WAIVA and CY13 cell lines all contain chromosome 22 and were all negative for expression of human IDUA (Table 3.4). More significantly, it was shown that hybrid cell lines containing human chromosome 22 did not express detectable levels of human IDUA (Table 3.1 and Table 3.4).

The only cell lines that expressed detectable human IDUA activity were CY14 and CY120. While the quantities expressed were ten to twenty fold lower than the

TABLE 3.4. EXPRESSION OF IDUA ACTIVITY IN HUMAN-MOUSE CELL HYBRIDS.

For experimental details see Section 3.2.2c. Cell line A9 is the mouse parent fibroblasts of the CY cell lines. IDUA activity was measured per pmol/min/mg of cell protein. Note that significant IDUA activity was captured using the Id1A monoclonal antibody in cell lines CY120 and CY14, the only hybrid cell lines that contain human chromosome 4 (Table 3.1).

TABLE 3.4.

	α-L-iduronidase Activity (pmol/min/mg protein)									
	WAIVA	Wegroth D-2	CY12	CY13	CY123	CY120	CY14	A9	Human Leucocytes	Human Fibroblasts
Total IDUA activity in extract	504	152	224	191	980	966	823	222	135	2441
Immuno-captured IDUA activity	0.17	0.46	0.43	0.22	0.50	11.3	19.1	.21	58.8	318

values for normal human fibroblasts, they are ten fold higher than the A9 mouse control and are comparable to values obtained from human leucocytes. These lower values probably reflect a lower level of expression of human IDUA in the human-mouse cell hybrids compared with human fibroblasts. It can be seen from the leucocyte IDUA activities recorded in Table 3.4, that the immunocapture assay detected 44% of the human IDUA activity present in the cell extracts. Thus human IDUA accounts for only 2 to 4% of the total IDUA activity present in CY120 and CY14 respectively, which demonstrates the sensitivity of the assay. The immunocapture assay only captured 13% of total IDUA activity present in the human fibroblasts, but this reflects the loading capacity of the assay, and not its sensitivity. That is, the level of IDUA activity detected in a normal human fibroblast control extract (318 pmol/min/mg protein) reflects the maximum binding capacity of the assay which is limited by the number of antigen binding sites present on 10 µg of Id1A monoclonal antibody.

Schuchman *et al.* (1984b) purified two forms of human IDUA, a high and low uptake form, as defined by their receptor-mediated uptake into IDUA deficient (MPS-I) fibroblasts. The polyclonal antibody used in the localization study (Schuchman *et al.*, 1984a) was produced against the low uptake form of IDUA (Schuchman *et al.*, 1984b; 1984c). Schuchman *et al.* (1984a) showed that their antibody against the human IDUA preparation was not species specific, indicating that the human and mouse isozymes shared common antigenic determinants. To distinguish between the human and murine IDUA present in the somatic cell hybrids, Schuchman *et al.* (1984a) mixed known amounts of human and murine IDUA. They demonstrated that 75 units of IDUA activity incubated with the antiserum resulted in >60% of human enzymatic activity detected in the immunoprecipitate, while <2% of the murine enzymatic activity was recovered. Two observations mentioned by Schuchman *et al.* (1984a) indicate that the chromosomal localization was possibly incorrect. Firstly, equal amounts of both the mouse and human enzymes were depleted from the supernatant of the mixtures used in determining the conditions for the species specific assay. It was speculated that the selective detection of human IDUA was due to the inactivation of the murine enzyme in the antigen-antibody complex, or that the murine enzyme was destabilized by the

immunoprecipitation. Secondly, expression of human enzymatic activity in chromosome 22 positive hybrids was noted to be remarkably reduced. Hybrids containing chromosome 22 showed only 15% of immunoprecipitated IDUA activity compared with normal diploid human fibroblasts. It was speculated that this was due to differences in post-translational processing and transport of the human and murine isozymes in the murine based cell hybrids, or because of the absence or suppression of regulatory sequences or proteins. Another antiserum produced during the time of their study was thought to be species specific and was used to detect immunoreactive material from the cell hybrids by both Ouchterlony immunodiffusion and rocket immunoelectrophoresis. Both antisera used in this study may have detected a contaminating polypeptide present in their enzyme preparation, and/or were not species specific. This is supported by the fact that these antisera detected immunologically cross reactive protein, between 38% to 105% of normal levels, in MPS-I patient fibroblasts (Schuchman and Desnick, 1988). This is in contrast to results obtained by other workers (Myerowitz and Neufeld 1981) and ourselves (Ashton *et al.*, 1992). In a study of twenty three different MPS-I cell lines, using two different monoclonal antibodies, all but one cell line contained less than 5% of the IDUA protein found in normal controls. Two of the patient fibroblast lines used in the Schuchman and Desnick (1988) study were shown to contain less than 1% of normal IDUA protein using the monoclonal antibody detection system of Ashton *et al.* (1992). This highlights the potential danger of relying exclusively on polyclonal antibodies to detect a low abundance proteins when the protein preparation that the antibodies are raised against may contain other cross reactive materials.

3.5.3 REGIONAL MAPPING OF *IDUA*

This work was preformed by Marcy E. MacDonald and James F. Gusella of the Neurogenetics Laboratory at the Massachusetts General Hospital of the Huntington Disease Collaborative Research Group.

The gene for Huntington disease (or chorea) is also localized to human chromosome 4p16.3 (Cox *et al.* 1989; Murray and Van Ommen, 1991). Huntington disease (HD) is a late onset autosomal dominant disorder characterized by progressive chorea, psychiatric disturbance and intellectual decline due to loss of striatal neurons. Symptoms typically appear in middle age and death ensues in 10 to 20 years, after an inexorable decline that currently cannot be prevented or delayed by therapy (Hayden, 1981). HD maps in the region distal to the marker locus D4S10, approximately 4,200-kb from the telomere of chromosome 4p, although conflicting data from recombination events in HD kindreds have lead researchers to believe that HD is either within a region 2,500-kb distal to D4S10 or within 1000-kb of the telomere (for a review see Murray and Van Ommen, 1991; or Gusella, 1991). Many cloned genes, including candidates such as the two GABA receptor subunits (*GABRA2*, *GABRB1*; Buckle *et al.*, 1989), and the quinoid dihydropteridine reductase (QDPR; MacDonald *et al.*, 1987), have been mapped on chromosome 4p, but none have been assigned within the immediate vicinity of HD until recently (fibroblast growth factor receptor 3, *FGFR3*; Thompson *et al.*, 1991; and cGMP phosphodiesterase; Weber *et al.* 1991b). An intensive effort has been undertaken to clone the gene for Huntington disease based on its chromosomal location. This has produced detailed genetic and physical maps of the 4p telomeric region. Because of the chromosomal localization of *IDUA* to 4p16.3, studies were performed to determine the precise location of *IDUA* within the 4p16.3 physical map.

To further confirm the chromosomal localization of IDUA to chromosome 4 and provide a finer regional localization, the 1.6-kb *Pst* I insert of pID89 was hybridized to a Southern blot of *Hin*dIII digested DNA from a panel of human x hamster somatic cell hybrids containing chromosomes derived from Wolf-Hirschhorn syndrome (HW) patients (Fig. 3.9). As expected, the probe showed strong hybridization with the

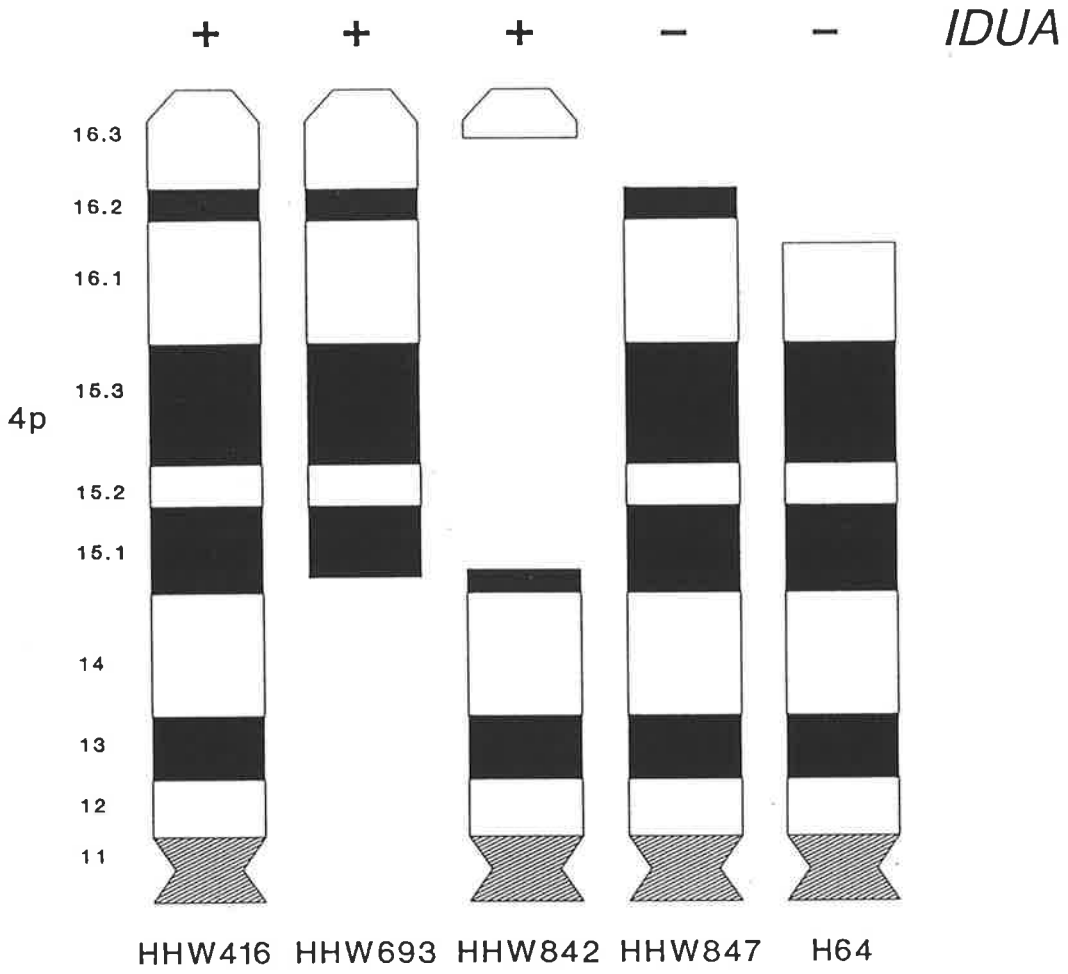


FIGURE 3.9. SCHEMATIC REPRESENTATION OF HW CELL LINE CHROMOSOME 4p CONTENT.

The regions of chromosome 4p present in each cell hybrid of the regional mapping panel containing chromosomes derived from Wolf-Hirschhorn syndrome (HW) patients are shown schematically. Presence or absence of IDUA, as demonstrated by a 20-kb *Hin*dIII fragment positive to the 1.6-kb *Pst* I fragment (see Fig 3.5), is indicated by a + or - respectively.

hamster parent cell line due to evolutionary conservation of *IDUA*. In addition to the 13-kb hamster fragment, a human fragment of approximately 20-kb was observed in some of the hybrid lines, including HHW416 which contains only human chromosome 4 (data not shown). As in the previous analysis (Section 3.5.2a), the human fragment was not present in the hybrid cell lines which contain chromosome 22 as the only intact human chromosome, Eye3Fa6 and GM10888. Rather, IDUA mapped to the most terminal segment of chromosome 4 defined by the regional panel; that is, the distal portion of 4p16.3 (Fig. 3.9). A detailed pulse field map of this distal region, along with the position of numerous cosmid clones and genetic markers had been previously reported by the Huntington Disease Collaborative Research Group (Bucan *et al.*, 1990; Whaley *et al.*, 1991). The region spans approximately 2,200-kb of DNA and contains one unmapped segment.

The 1.6-kb *Pst* I insert of pID89 was then hybridized to the collection of cosmids from 4p16 and specific hybridization to the cosmid A157.1 (Pohl *et al.*, 1988; Whaley *et al.*, 1991) was observed. The cosmid A157.1 contains p157.9, a single copy probe that defined the anonymous locus *D4S111* and detects a highly informative multiallele DNA polymorphism (MacDonald *et al.*, 1989). The placement of p157.9 and pID89 on the restriction map of A157.1 is shown in Fig 3.10. The probes are located approximately 3-kb from each other within the same 6-kb *Not* I fragment. Placement of *IDUA* within the cosmid A157.1 automatically placed IDUA on the physical map of 4p16.3 1,100-kb from the telomere of chromosome 4p, making it the most telomeric cloned gene assigned to this chromosome arm although it is likely that there will be others that are more telomeric. At least 15 CpG islands, indicating the presence of genes, have been reported in the 450-kb area surrounding *IDUA* (Weber *et al.*, 1991a). Transcription of *IDUA* is towards the centromere (Lorne A. Clarke and Bernhard Weber, personal communication). The mouse gene for IDUA (*Idua*) has been mapped to the proximal portion of mouse chromosome 5 near homologues for other loci on human chromosome 4p, however comparison of gene order in mice and humans demonstrated that a chromosomal rearrangement within a conserved synteny has

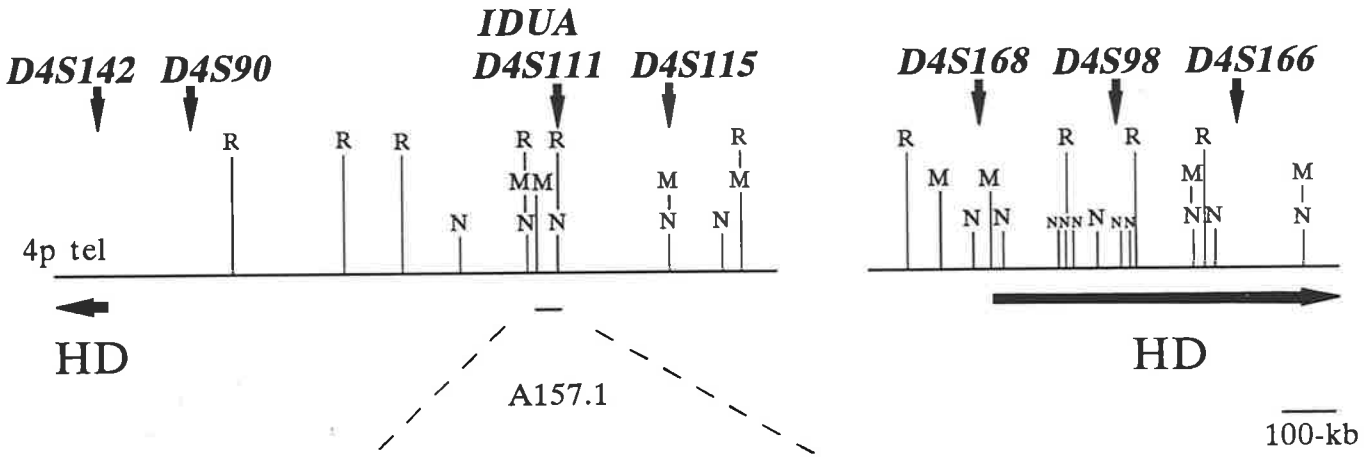
occurred since divergence of lineages leading to mice and humans (Koizumi *et al.*, 1992).

The localization of *IDUA* 1,100-kb from the telomere of 4p16.3, coincident with *D4S111*, places IDUA outside either of the two candidate regions for the HD gene, thus excluding the remote chance that *IDUA* had a role in HD. IDUA is placed in the region hemizygous in Wolf-Hirschhorn syndrome, a congenital abnormality with severe mental retardation involving deletions of portions of chromosome 4p. However, deletion of one of the copies of IDUA is unlikely to contribute to the Wolf-Hirschhorn phenotype given the absence of clinical features in parents and sibs of MPS-I patients (Neufeld and Muenzer, 1989). The chromosome segment of 4p16.3 that contains *IDUA* has been, and is being, extensively analysed for DNA polymorphisms and multiallele markers for use in HD diagnosis and mapping (e.g. Weber *et al.* 1992; Gusella *et al.* 1992). The immediate importance of this study was to make these markers immediately applicable to genetic linkage analysis and carrier prediction for MPS-I, and to the assessment of allelic heterogeneity and identification of potential independent mutation events. However, an abundance of polymorphisms have now been detected within *IDUA* (see Chapter 4) suggesting that they are unlikely to be required. The possibility of finding a patient heterozygous for an MPS-I mutation and a deletion of chromosome 4p, resulting in both MPS-I and Huntington disease symptoms, may lead to finer mapping of the Huntington disease gene. As described in the following sections, this fine mapping of IDUA, and particularly the identification of *IDUA* within the cosmid A157.1, played an important part in the final isolation of a full-length cDNA for IDUA.

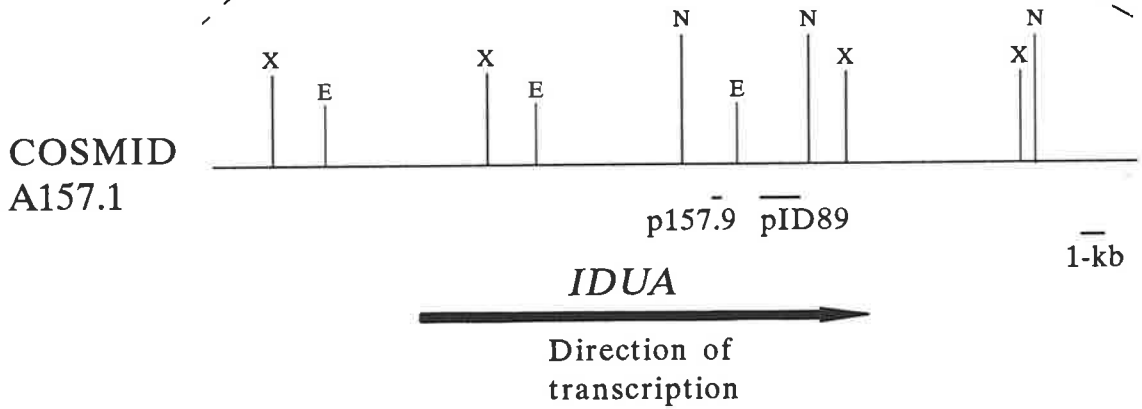
FIGURE 3.10. REGIONAL MAPPING OF IDUA.

A long range restriction map of 4p16.3 with the telomere indicated on the left and polymorphic loci used to create the map shown throughout. Restriction enzymes are *Not* I (N), *Mlu* I (M) and *Nru* I (R). The positions of p157.9 and pID89 in cosmid A157.1 are shown. A restriction map of cosmid A157.1 for the enzymes *Eco*RI (E), *Xho* I (X) and *Not* I (N) is given. The positions of p157.9 and pID89 were determined by hybridization of the probes to restriction digests of A157.1 with these enzymes.

LARGE SCALE RESTRICTION MAP



RESTRICTION MAP



3.6 ISOLATION OF cDNA SEQUENCE FOR IDUA

3.6.1 NORTHERN BLOT ANALYSIS

The 729-bp insert of λRPC1 was used to probe total placental RNA and poly $(A)^+$ RNA. A single 2.3-kb band was detected only when 40 μg of poly $(A)^+$ RNA was loaded in a single track (Fig. 3.11). This indicated that approximately 1.6-kb of IDUA sequence remained to be isolated. The strength of the signal also indicated that, at least in placenta, the mRNA of IDUA was considerably less abundant than the low abundance iduronate-2-sulphatase mRNA from the same preparation (Wilson *et al.*, 1990).

3.6.2 PCR APPROACHES TO CLONING IDUA

Annette Orsborn of the Department of Chemical Pathology at the Adelaide Children's Hospital performed some of the attempts at each of the PCR based approaches described below.

Screening of the four cDNA libraries available at this time (lung fibroblast, testis, placental, and colon) with the insert of λRPC1 and several genomic fragments isolated from λID475 did not result in any positive clones. As a parallel approach for isolation of the rest of the coding sequence of IDUA, a number of PCR techniques were attempted as described below.

3.6.2a PCR FROM cDNA LIBRARIES

PCR amplification from cDNA libraries was described by Friedman *et al.* (1988) and is shown diagrammatically, as attempted for IDUA, in Fig. 3.12. Briefly, an oligonucleotide primer made to the gene of interest is used in a PCR reaction with a primer made to a vector such as λgt11 to PCR amplify an uncloned region of the gene of interest from an appropriate cDNA library. This enables the complete screening of a cDNA library in a single reaction, in contrast to numerous plaque lifts and hybridizations. The PCR products obtained may be cloned and sequenced to confirm

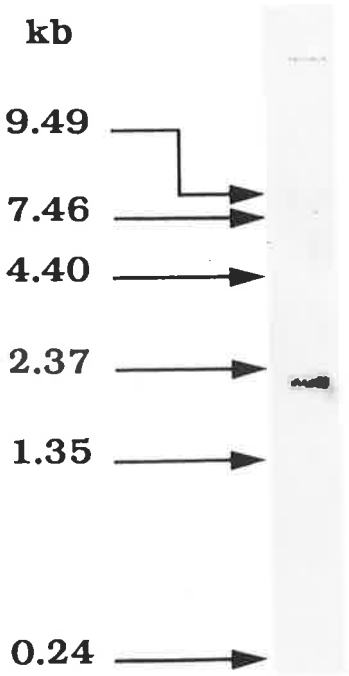
their identity and then oligolabelled to high specific activity for use as probes in cDNA library screening.

This technique was attempted for IDUA using primers within λRPC1 (IDP2, IDP4, ID9.2, ID10 and ID11, Figs. 3.5 and 3.12) in combination with λgt10 or λgt11 forward and reverse primers. DNA was made from the lung fibroblast, testis, placental and colon cDNA libraries to enable a larger number of clones to be screened at one time than could be attempted using a high titre phage stock as template. PCR was as described in Section 2.2.21, except that denaturation was for 1 min at 95°C, annealing was at 55°C for 1 min and extension was at 70°C for 2 min. However, no visible products were obtained in these PCRs, and Southern blotting followed by hybridization with an internal oligonucleotide only detected products from the colon cDNA library that were consistent with the presence of the clone λRPC1.

This lack of success could be for many reasons. It may be that this result was real and that, besides λRPC1, there were no other IDUA cDNA clones in the libraries. However, it is more likely that a combination of the following problems contributed to this lack of success. The PCR reaction conditions were not varied sufficiently to enable amplification of the IDUA target sequences. Experience gained since these attempts with PCR reactions for *IDUA* have shown that the GC-rich IDUA target sequence often requires the use of 10% (v/v) DMSO, high annealing temperatures (>58°C), and longer primers (24 to 30-mers) in the PCR reactions. PCR using cDNA libraries as template also requires great sensitivity to detect rare mRNA transcripts such as IDUA, as the starting amount of target template is very low. This sensitivity can be achieved by using "nested" PCR primers as described some time after our attempts by Gibbons *et al.* (1991). That is, a second pair of primers, located within the first pair are used to PCR amplify a small aliquot of the initial PCR reaction. This decreases the amount of non-specific PCR products amplified and should increase the specific product of interest to enable visualization following ethidium bromide staining.

FIGURE 3.11. NORTHERN BLOT ANALYSIS OF IDUA.

A Northern blot of 40 μg of poly(A)$^+$ mRNA probed with the insert of the IDUA cDNA clone, λRPC1, is shown. A single 2.3-kb band is detected. The size and relative positions of the RNA markerS used is indicated on the left.


kb
9.49
7.46
4.40
2.37
1.35
0.24

3.6.2b ANCHORED PCR

Rapid production of full-length cDNAs from rare transcripts using a single gene specific primer was first described by Frohman *et al.* (1988) and a nested PCR modification of the method by Frohman and Martin (1989). Briefly, either the gene specific primer or the poly(dT) primer is used to generate first strand cDNA. One of the two nested gene specific primers can be used in conjunction with a poly(dT) primer with linker sequences on the 5' end (SEB-T, Fig. 3.12 and Appendix D) to PCR amplify the gene of interest from cDNA template. The poly(dT) primer can prime on the poly(A) tail at the 3' end of an mRNA transcript, or on a poly(dT) tail generated with terminal transferase at the 5' end of a cDNA molecule. A small aliquot of the initial PCR reaction is then reamplified with the second gene specific primer and a primer made to the linker sequences of the poly(dT) primer (SEB-19, Fig. 3.12 and Appendix D). This technique was attempted for IDUA using the oligonucleotides inside λRPC1 mentioned in Section 3.6.2a (Fig. 3.12). Again, this method was unsuccessful, most likely due to the same reasons listed in the above section and also because the IDUA mRNA appears to be at such low levels, making the first strand cDNA synthesis and subsequent addition of poly(dT) very inefficient. The second report of this technique (Frohman and Martin, 1989) stated that, for genes with very low levels of transcription, the technique may only partially purify the cDNA of interest due to non-specific amplification.

3.6.2c LIGATION MEDIATED PCR

Ligation mediated PCR was first described by Mueller and Wold (1989). Briefly, a double stranded cDNA molecule with an overhang at one end is generated using a combination of poly(dT) priming and gene specific priming. A linker with a 5' overhang is then ligated to the cDNA (a combination of SEB-13 and SEB-19, see Appendix D). A primer made to the sequence of one strand of the linker (SEB-19) is then used to PCR amplify the gene of interest in combination with a gene specific primer. The 5' overhang on the linker prevents the linkers from ligating together. The

FIGURE 3.12. PCR APPROACHES TO CLONING IDUA.

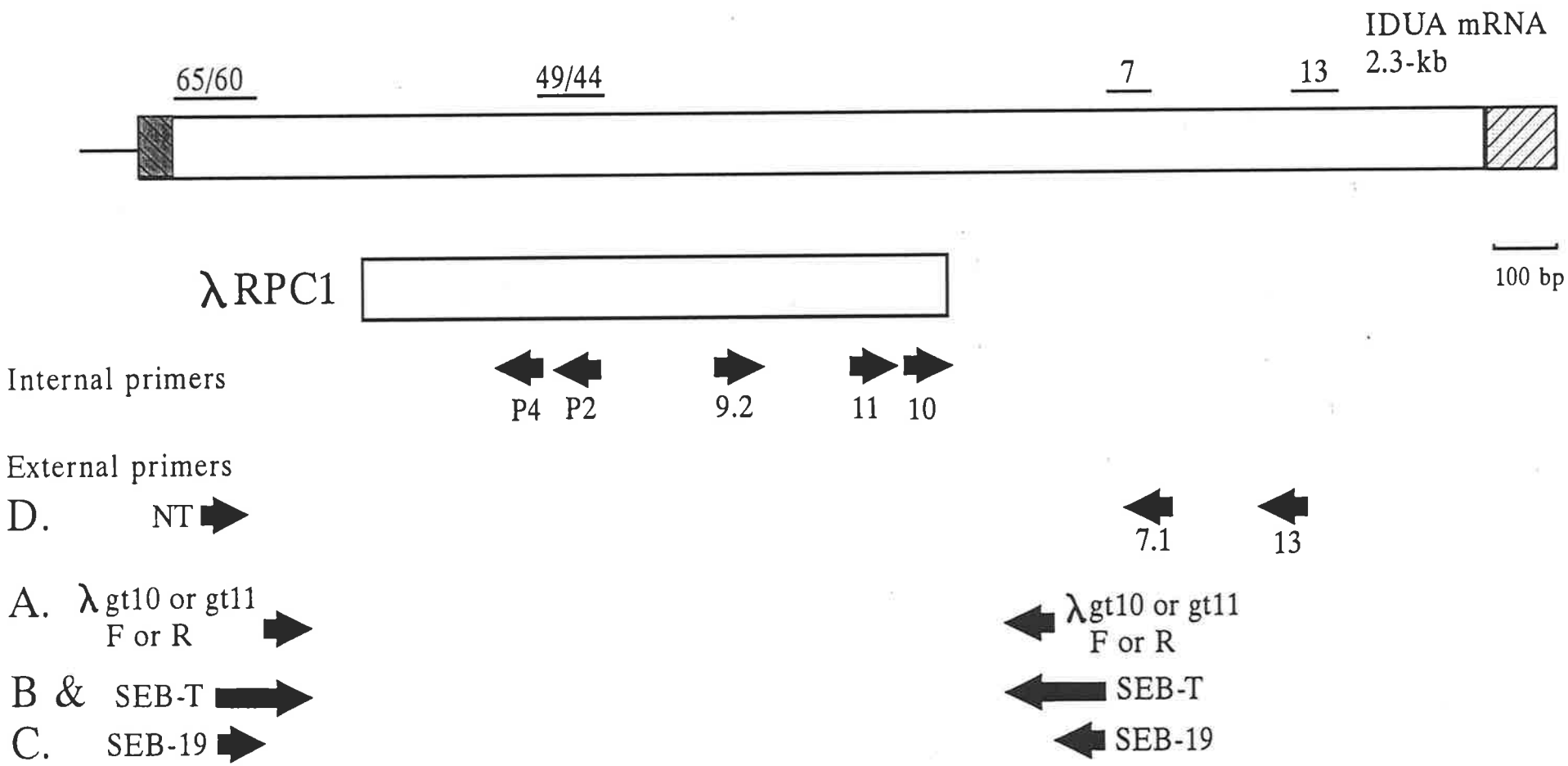
Diagrammatic representation of the PCRs attempted for cloning of IDUA using the primers described in Section 3.6.2. A hypothetical IDUA mRNA is shown with the relative position of λRPC1 indicated below. The striped regions represent the signal peptide and 3' untranslated sequence. The hypothesized position of the N-terminal sequences of IDUA and tryptic peptide 7 are shown above the mRNA. The position and names of primers used are marked by arrows. All PCRs utilised sense (—►) and antisense (◄—) primers, at least one of which was internal to λRPC1 and the other, external. The external primers were;

A. λgt10 or λgt11 forward (F) or reverse (R) primers for PCR using cDNA libraries as template.

B. SEB-T and SEB-19 for anchored PCR.

C. SEB-19 and SEB-13 for ligation mediated PCR.

D. IDNT, ID7.1, ID7.2 and ID13 for PCRs to peptide sequences of IDUA (also see Fig. 3.13).


IDUA mRNA
2.3-kb
65/60
49/44
7
13
100 bp
λRPC1
Internal primers
P4
P2
9.2
11
10
External primers
D.
NT
7.1
13
A.
λgt10 or gt11
F or R
λgt10 or gt11
F or R
B &
SEB-T
SEB-T
C.
SEB-19
SEB-19

production of a double stranded cDNA molecule with an overhang at one end is vital for success with this method to prevent the ligation of the linker to both ends of cDNA molecules, thus enabling amplification of anonymous DNA segments with the linker primer at both ends. This technique was also unsuccessful, once again probably due to the reasons discussed above.

3.6.2d PCR BETWEEN OLIGONUCLEOTIDES MADE TO PEPTIDES OF IDUA

It was felt that another limiting factor present in the three techniques described above was the use of gene specific primers at only one end of the PCR amplification. In an attempt to avoid this limitation, PCRs between the oligonucleotide primers within λRPC1 and mixed oligonucleotide primers designed to peptide sequences from immunopurified IDUA that were not yet represented in cDNA clones were attempted using cDNA as template. Based on the original model for proteolytic processing of IDUA (Fig. 1.7), the 13-kDa IDUA peptide was thought to be at the very 3' end of the IDUA cDNA. Thus, oligonucleotide primers were made to the N-terminal amino acid sequence of the 13-kDa IDUA peptide that had recently been determined (ID13, Fig. 3.13A), to tryptic peptide 7 (ID7.1 and ID7.2, Fig. 3.13B) and also to nucleotide sequence encoding the 65/60-kDa N-terminus that had recently been obtained from genomic clones (IDNT). PCR amplification was attempted between these primers and the primers IDP4, ID9.2, ID10 and ID11 (Fig. 3.12). Again, this technique was not successful. For PCRs utilizing the oligonucleotide primers made to peptide 7 and the 65/60-kDa N-terminus, the reasons for lack of success are probably as described in Section 3.6.2.a. With hindsight, the reason PCRs utilizing the oligonucleotide made to the 13-kDa N-terminal peptide sequence were unsuccessful was obvious, as the 13-kDa peptide is derived from the 5' end of the IDUA cDNA, and not the 3' end. Thus, the oligonucleotides for use in this PCR were designed in the wrong orientation (or sense).

Whilst none of the PCR approaches used for attempts to clone *IDUA* were successful, they were not rigourously attempted. The availability of the cosmid clone, A157.1, and cDNA clone λE8A, as described in the following sections, negated the

need for these methods to be used in the long term. The methods are presented in this thesis as an illustration of the experimental designs that were attempted in order to overcome the lack of success from cDNA library screening for *IDUA*.

3.6.3 ISOLATION OF λE8A

Annette Orsborn of the Department of Chemical Pathology at the Adelaide Children's Hospital performed some of the cDNA library screening, clone isolation and characterization by both PCR and sequencing.

Following the lack of success experienced with exhaustive screening of the four cDNA libraries available at this time (lung fibroblast, testis, placental, and colon), several factors were examined in the process of considering alternative cDNA libraries for screening. These factors were; the number of independent clones in the libraries, the vectors used, the methods of construction, the average insert size and the tissue source of the mRNA. At this time one of the best libraries available regarding the number of independent clones present, was a λgt11 human endothelial cDNA library (see Section 2.1.9) which contained approximately 2.5 x 10^6 independent clones, in contrast to most other libraries containing only 1 x 10^6 independent clones. It was thought that the low abundance of IDUA mRNA was the principal reason for the lack of success in cDNA library screening when using high specific activity probes, such as the cDNA insert from λRPC1. Increasing the number of independent clones available for screening should increase the chance of obtaining IDUA cDNA clones.

Thus, the λRPC1 insert was used to screen approximately 5 x 10^5 recombinant clones from this newly obtained λgt11 human endothelial cDNA library. Twenty clones were isolated and characterized by PCR using IDUA specific oligonucleotides and a primer designed to the arms of the λgt11 vector. The insert of the longest clone, λE8A, was isolated, subcloned into M13mp18 and sequenced using *IDUA* specific oligonucleotides to "walk" through the sequence until the insert was fully sequenced in both directions. The 1765-bp insert (Fig. 3.14) contained an open reading frame that extended from just before the position of the 65/60/18-kDa amino terminus to a stop

FIGURE 3.13. PCR OLIGONUCLEOTIDES TO THE 13-kDa N-TERMINUS AND PEPTIDE 7.

Primers for PCR made to IDUA peptide sequences are shown. The amino acid sequences, all possible mRNA sequences and sequences of the PCR primers made to the 13-kDa N-terminus and tryptic peptide 7 of IDUA are shown in A and B. All primers made to these peptide sequences were designed in the antisense direction, corresponding to their hypothesized positions in the IDUA cDNA relative to λRPC1, and were designed with *Eco*RI restriction sites at their 5' ends to enable the cloning of PCR products obtained.

A. The central 8 residues of the 13-kDa N-terminal sequence were unambiguous (see Table 3.2) and were used to design the primer aligned underneath.

B. Two primers were made to peptide 7 due to the high codon degeneracy of residues 1 and 2 of peptide 7. ID7.1 excluded these residues from the primer design to ensure efficient annealing at the 3' end of the primer during PCR, while ID7.2 included these residues to provide added length and specificity to the primer.

A: **Amino acid sequence and possible mRNA sequences for the 13-kDa N-terminus:**

```
 Q   V   D   A   A   R   A   L
CAA GUN GAU GCN GCN CGN GCN UUA
  G       C           AGA     G
                        G   CUN
```

Primer:

```
          3'                                  5'
ID13  GTT CAC CTA CGG CGG GCC CG CTTAAGG
        C       G                EcoRI
```

B: **Amino acid sequence and possible mRNA sequences for peptide 7:**

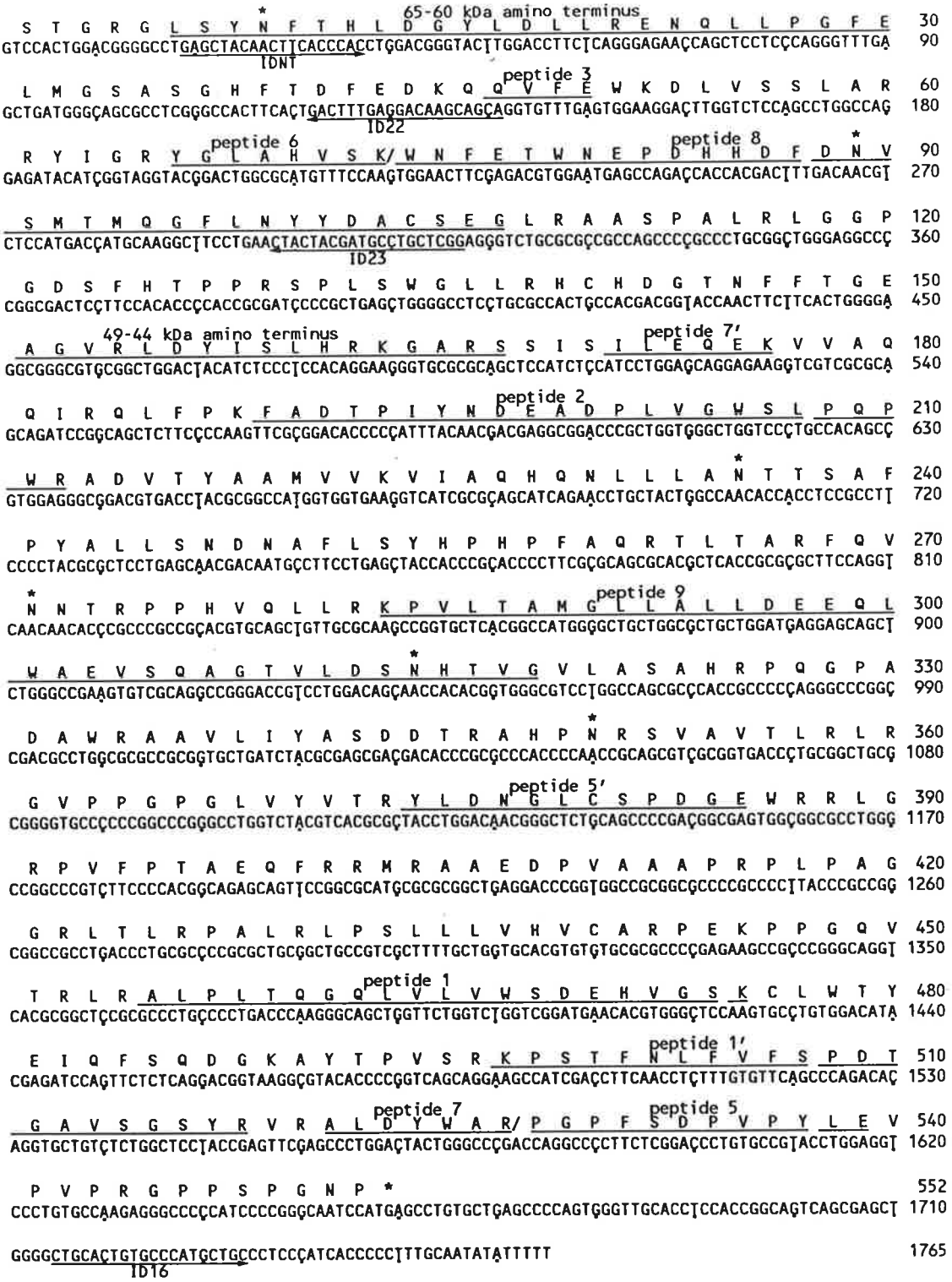
```
 A   L   D   Y   W   A   R
GCN UUA GAU UAU UGG GCN CGN
      G   C   C         AGA
    CUN                   G
```

Primers:

```
              3'                              5'
ID7.1         CTA ATA ACC CGG GCC CCTTAAGG
                G   G              EcoRI

          3'                                  5'
ID7.2 CGC GAC CTA ATA ACC CGG GCC CCTTAAGG
                G   G              EcoRI
```

FIGURE 3.14. SEQUENCE OF THE IDUA cDNA CLONE, λE8A.

The sequence of the IDUA cDNA clone λE8A and the deduced amino acid sequence of the IDUA protein is shown in single letter code above the cDNA sequence. Nucleotide and amino acid numbers are in the right margin. Amino acids colinear with either N-terminal peptide data or tryptic peptides are underlined and named above the sequence. Potential N-glycosylation sites and the stop codon are asterisked. The oligonucleotides used to produce the 5' probe for library screening (ID22 to IDNT), to analyse the 5' ends (ID22) and 3' ends (ID16) of cDNA clones by PCR and attempt anchored PCR (ID22 and ID23) are underlined below the nucleotide sequence with the arrows indicating either sense (→) or antisense (←).


65-60 kDa amino terminus
IDNT
peptide 3
ID22
peptide 6
peptide 8
ID23
49-44 kDa amino terminus
peptide 7′
peptide 2
peptide 9
peptide 5′
peptide 1
peptide 1′
peptide 7
peptide 5
ID16

codon 105-bp from the 3' end of the clone. Six further tryptic peptides were matched to the translated DNA sequence (increasing the total to 11) but, significantly, the sequence of the 13-kDa amino terminus, a secondary tryptic peptide (peptide 2'), a signal peptide and an initiating methionine were not present in this clone. PCR analysis performed on the other clones demonstrated that seven of these ended at the same base at the 5' end, and that the remaining clones were shorter (Fig. 3.15).

An 113-bp 5' probe for IDUA was made by PCR between the oligonucleotides ID22 and IDNT (Fig. 3.14) using λE8A as template and then used to screen all 8 cDNA libraries now available (see Section 2.1.9). Positive clones were not obtained from the screening of six of these cDNA libraries. Screening of two 5'-"stretch" cDNA libraries (umbilical endothelial and T-cell) resulted in a further 38 positive clones. PCR analysis showed that all of these clones ended at the same 5' base as λE8A (Fig. 3.15). Kinased PCR products were analysed on a sequencing gel but still failed to show any size variation. The two 5'-stretch cDNA libraries were purchased externally and it is therefore unlikely that they had been contaminated with λE8A. Great care was taken when plating these libraries by using pipettes that had not previously been used for preparation of DNA or high titre stocks of λE8A, and thus the library screen should not have been contaminated at the plating step. Appropriate positive and negative PCR control reactions were performed at all times negating the possibility that this result was due to PCR contamination. It was thought that a major secondary structure in the mRNA for IDUA may be responsible for the premature termination of the clones at their 5' ends and that perhaps the process used to overcome this problem during construction of the 5'-stretch cDNA libraries, that is the addition of methyl mercury hydroxide to the reverse transcription reaction, had failed to fully denature the IDUA mRNA. Nested anchored PCR (see Section 3.6.2b) was again unsuccessfully attempted, using the IDUA gene specific primers ID22 and ID23 (Fig. 3.14), to try to clone the 5' end of the IDUA cDNA.

At this stage it was felt that the remaining IDUA amino acid sequences not represented in λE8A may have derived from contaminating peptides that copurified

with IDUA. Thus, it was felt that the only sequence remaining to obtain a full-length cDNA for IDUA was that encoding a signal peptide. This was supported by the fact that the six amino acids immediately prior to the 65/60-kDa N-terminus appeared to form a reasonable signal peptide cleavage site (von Heijne, 1986). The exon structure of *IDUA* spanned by λE8A had been mapped at this time and was found to contain 12 exons in a 4.2-kb segment. Sequence data from a 2.2-kb *Pst* I fragment immediately 5' to the exon containing the nucleotide sequence encoding the 65/60-kDa N-terminus and the 4.2-kb segment was available and did not appear to encompass an exon encoding a signal peptide (see Section 3.7). Thus, an experiment was designed to express IDUA activity from λE8A. Previously, the signal peptide for lysosomal acid phosphatase (Pohlmann *et al.*, 1988) had been attached to a cDNA clone encoding another lysosomal enzyme, *N*-acetylgalactosamine-4-sulphatase, to successfully express *N*-acetylgalactosamine-4-sulphatase activity *in vitro* (Peters *et al.*, 1990). Similarly, an expression construct was synthesized for IDUA by attaching the signal peptide for arylsulphatase A (Stein *et al.*, 1989) onto the IDUA cDNA clone (Fig. 3.16). Briefly, a 100-mer oligonucleotide primer was made to the sequence of the 18 amino acid arylsulphatase A signal peptide and the first 27-bp of the sequence of λE8A 3' to the proposed signal cleavage site. Using an M13 clone of λE8A as template, PCR was performed between this 100-mer and the reverse sequencing primer of M13 for a total of 30 cycles with duplicate reactions being stopped every 5 cycles. Denaturation was at 94°C for 45 s, annealing was at 55°C for 45 s and elongation was at 72°C for 2 min. The PCR reactions were ethanol precipitated, digested with *Hin*dIII and *Eco*RI, and ligated into the expression vector pRSVN.07 (see Section 3.2.4). Two clones containing the predicted insert of 1868-bp and gave the predicted band pattern upon digestion with *Sma* I were selected for expression. pIDEX1 was cloned from PCR reactions performed for only 5 cycles and pIDEX11 from 15 cycles, thus minimising the chance that errors had been introduced into the clones during amplification due to the low fidelity of Taq polymerase. These expression constructs were electroporated into CHO cells as described in Section 3.2.4 and the selected cells were assayed for

FIGURE 3.15. PCR ANALYSIS OF THE 5' END OF IDUA cDNA CLONES.

PCR between primer ID22, and λgt10 and λgt11 forward or reverse primers to analyse the 5' end of IDUA cDNA clones obtained by library screening are shown. Lanes are pUC19 *Hpa* II (m), positive PCR control (c) and analysis of three IDUA cDNA clones (1 to 3). Electrophoresis was on high resolution 3% (w/v) Nusieve/ 1% (w/v) TAE agarose gels. Note that all three clones end at the same point at the 5' end.

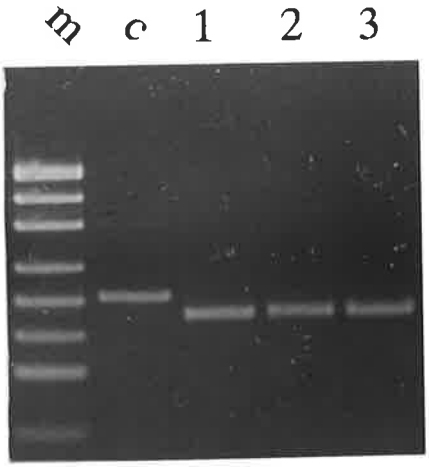

M
c
1
2
3

FIGURE 3.16. AN EXPRESSION CONSTRUCT USING λE8A AND THE SIGNAL PEPTIDE FROM ARYLSULPHATASE A.

Sequence of the oligonucleotide IDEX1 used in PCR with the M13 reverse primer with λE8A as template to produce an IDUA expression construct and its coded amino acids is shown. The first 18 amino acids match the signal peptide of arylsulphatase A. The arrow indicates the position of the expected signal peptide cleavage. The 65/60-kDa N-terminus of IDUA is underlined. The two amino acids between the signal peptide and the N-terminus match the IDUA peptide sequence, however the +1 and +2 positions of arylsulphatase A are R and P respectively (as shown underneath) and thus the weight matrix score for the signal peptide encoded for by this construct is only reduced from 11.4 to 9.7 compared with the native arylsulphatase A signal peptide (von Heijne, 1986). The last 27 nucleotides of the 100-mer could prime at the 5' end of λE8A. A 4-nt GC clamp and a *Hind*III site (underlined) were included at the 5' end of the primer to facilitate cloning after PCR amplification.

IDUA activity. No activity was detected, suggesting that the sequence of λE8A did not encode the complete IDUA protein without a signal peptide as had been hypothesized.

3.7 STRUCTURE AND SEQUENCE OF THE HUMAN IDUA GENE

The isolation of the EMBL3 genomic clone, λID475, has been described in Section 3.3.2. Isolation of, and identification of, the cosmid clone A157.1 containing *IDUA* has also been described (Pohl *et al.*, 1988; MacDonald *et al.*, 1989 and Section 3.5.3 respectively). The isolation and characterization of these genomic clones played an integral part in the eventual description of the entire coding sequence for IDUA. In this section the characterization of these clones and thus the genomic structure of *IDUA* will be described in greater detail.

3.7.1 FULL SEQUENCE OF *IDUA*

Some of the cloning and sequencing of the 1.7-kb Stu I fragment at the 3' end of the 4.5-kb segment was done by Xiao-Hui Guo of the Department of Chemical Pathology at the Adelaide Children's Hospital.

Initially, fragments of the genomic clone were selected for sequence analysis on the basis that they hybridized to the oligonucleotide probes designed to IDUA peptide sequence, that had previously been used for library screening. A 1.6-kb *Pst* I fragment, previously described (Section 3.4; Fig. 3.5), hybridized to the oligonucleotide probes made to peptide 8 (ID5, ID8 and ID47; Fig. 3.3; Section 3.3.1) and peptide 9 (ID9; Fig. 3.2B). An 850-bp *Pst* I fragment hybridized to the 65/60-kDa N-terminal probe (ID50; Fig. 3.2A) that was placed immediately adjacent to the 1.6-kb fragment by restriction mapping. A 1.7-kb *Stu* I fragment, that overlapped the 1.6-kb *Pst* I fragment by approximately 100-bp , hybridized to the oligonucleotide primer made to peptide 7 (ID7.1; Fig. 3.13B). The three fragments described above were isolated from the genomic clone λID475 and subcloned into M13 or pTZ vectors for DNA sequencing with the universal sequencing primer (USP) or IDUA specific primers. Intron and exon

junctions were mapped with reference to λE8A and this part of *IDUA* was found to be split into 12 exons clustered in a 4.2-kb region (Fig. 3.17). The exon at the 5' end of this region ended 11-bp before the 5' end of the cDNA clone λE8A. The exon at the 3' end of this region contained the stop codon and all the 3' untranslated sequence from the cDNA clone λE8A. The expression experiment described in Section 3.6.3 had suggested that λE8A did not contain enough sequence to encode active IDUA. Also, as described in Section 3.6.3, the sequence of the 13-kDa amino terminus, a secondary tryptic peptide (peptide 2'), a signal peptide and an initiating methionine were not present in λE8A.

Immunopurified IDUA from human liver contains seven major polypeptide bands on SDS/PAGE (Clements *et al.*, 1989) but at this time only 5 N-terminal sequences had been determined (65, 60, 49, 44 and 13-kDa). In an attempt to clarify the location of all peptide sequences from purified IDUA, and thus possibly provide more information to locate the rest of the IDUA gene, the N-terminal sequence of the 2 remaining peptides was determined (the 74 and 18-kDa peptides). Some of the N-terminal sequences were not clear, partially due to the very low molar amounts of these polypeptide bands of IDUA (e.g. the 74 and 13-kDa bands, Fig. 1.6). It also appeared that there were other peptide sequences present, perhaps due to the samples being slightly acid labile causing the protein to break down, or due to the presence of contaminating peptides. The resulting sequences were very weak, with mixed signals at some positions, making it difficult to match some of these sequences, particularly the 74 and 13-kDa sequences. Three different amino-terminal sequences were found to be present in immunopurified IDUA (Table 3.2). The 65, 60, and 18-kDa species share a common N-terminal amino acid sequence, the 49 and 44-kDa another and the 74 and 13-kDa species another. A cDNA clone for IDUA, λE8A, had been isolated that contained all amino acid sequence data obtained from immunopurified IDUA except for the 74/13-kDa N-terminus and a tryptic peptide, peptide 2' (Table 3.3). Regional mapping studies for IDUA had just been completed and the cosmid clone, A157.1, became available for study (Section 3.5.3). An oligonucleotide probe made to tryptic peptide 2' (ID2, see Appendix A) did not hybridize to A157.1. Hybridization of the

oligonucleotide ID13 (Fig. 3.13A), previously used in PCR studies, to Southern blots of restriction endonuclease digests of the cosmid A157.1, revealed that a 2.8-kb *Bam*HI fragment, 5' to the rest of the IDUA gene, was positive to this oligonucleotide (Fig. 3.18). Thus it was hypothesized that the 74/13-kDa N-terminal polypeptide sequence was located at the 5' end of the IDUA mRNA. Assuming that all seven IDUA polypeptide bands represent part of a single IDUA polypeptide, a model may be proposed, that differs from that previously proposed (Fig. 1.7 and Clements *et al.*, 1989). This new model demonstrates three sites of proteolytic processing of the 74-kDa polypeptide to produce the seven major bands of IDUA (Fig. 3.19).

The 2.8-kb *Bam*HI fragment was isolated and partially sequenced to reveal the presence of exon I of the IDUA gene. The sequence contained an initiating methionine, a signal peptide, the 74/13-kDa amino terminus and the start of the last unmatched tryptic peptide (peptide 2', Fig. 3.20). To determine if there was any coding sequence between the 5' most exon of the 4.2-kb cluster and exon I of IDUA, oligonucleotides were designed to exon I and to the 5' end of λE8A, corresponding to the 5' most exon of the 4.2-kb cluster (ID56, ID57 respectively, Fig. 3.20 and 3.21). PCR amplification of normal fibroblasts cDNA, annealing at 55°C, revealed a single PCR product of 84-bp, which implies that there was no further coding sequence for IDUA (Fig. 3.21A). However, the interrupted tryptic peptide sequence (Fig. 3.20) was still not accounted for by this result. When other PCRs at the 5' end of the IDUA cDNA were performed with the aim of producing a full-length IDUA cDNA insert, the major PCR product obtained (between ID58 and ID61 or ID60 and ID61) from normal fibroblast cDNA was found to be approximately 140-bp bigger than expected from the previous result (Fig. 3.21B). This PCR was performed as previously described except for the presence of 10% (v/v) DMSO in the reaction mixes. Denaturation was for 45 s, annealing was at 58°C for 45 s and elongation was for 2 min. The major PCR product was directly sequenced (Section 4.2.4) using the oligonucleotides ID56 and ID57 to reveal the presence of an additional 141-bp exon encoding the rest of peptide 2'. The collated cDNA sequence (Fig. 3.22) now encoded a protein containing all amino-terminal and tryptic peptide sequences obtained from purified IDUA and was consistent with the revised model for IDUA (Fig.

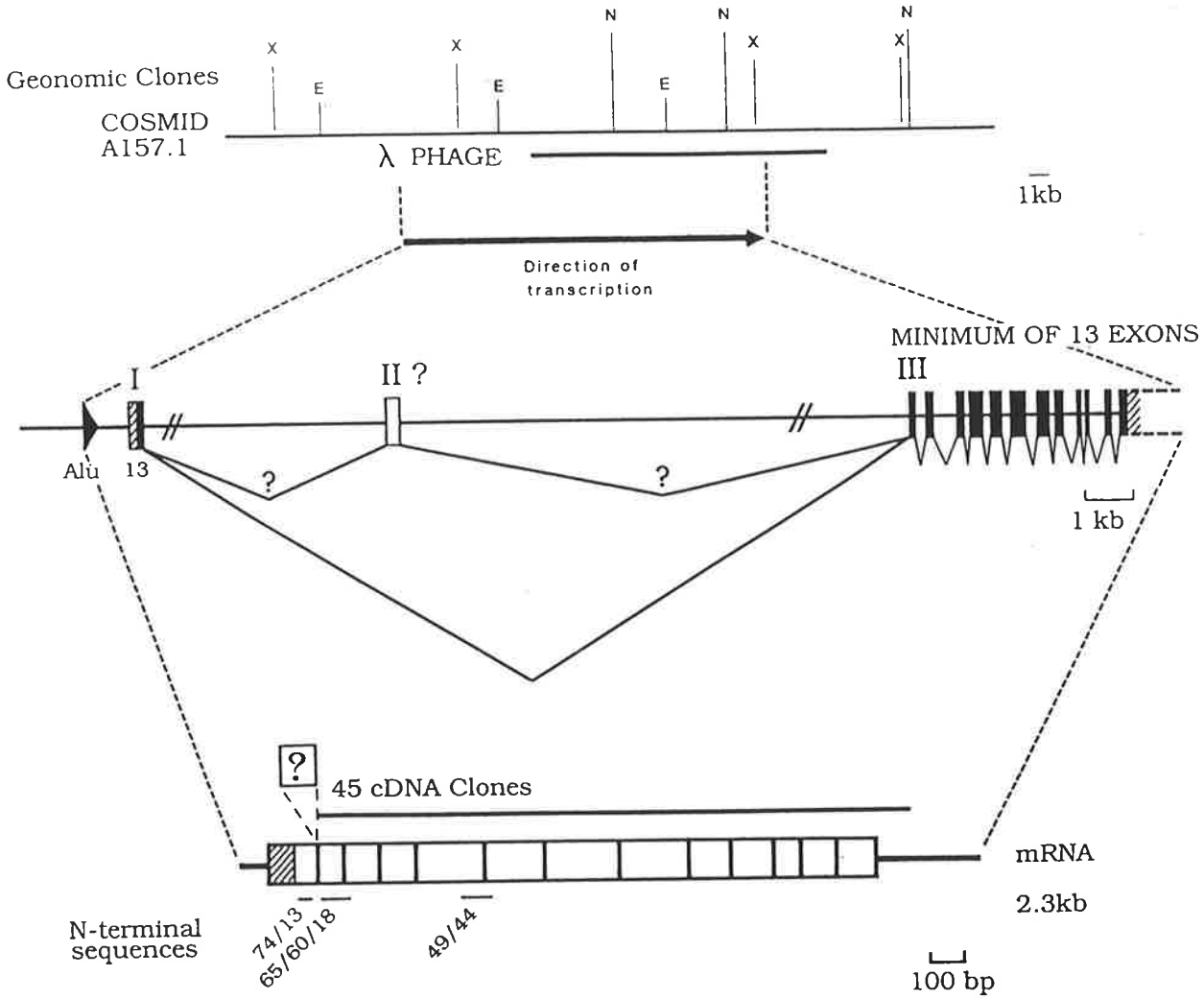


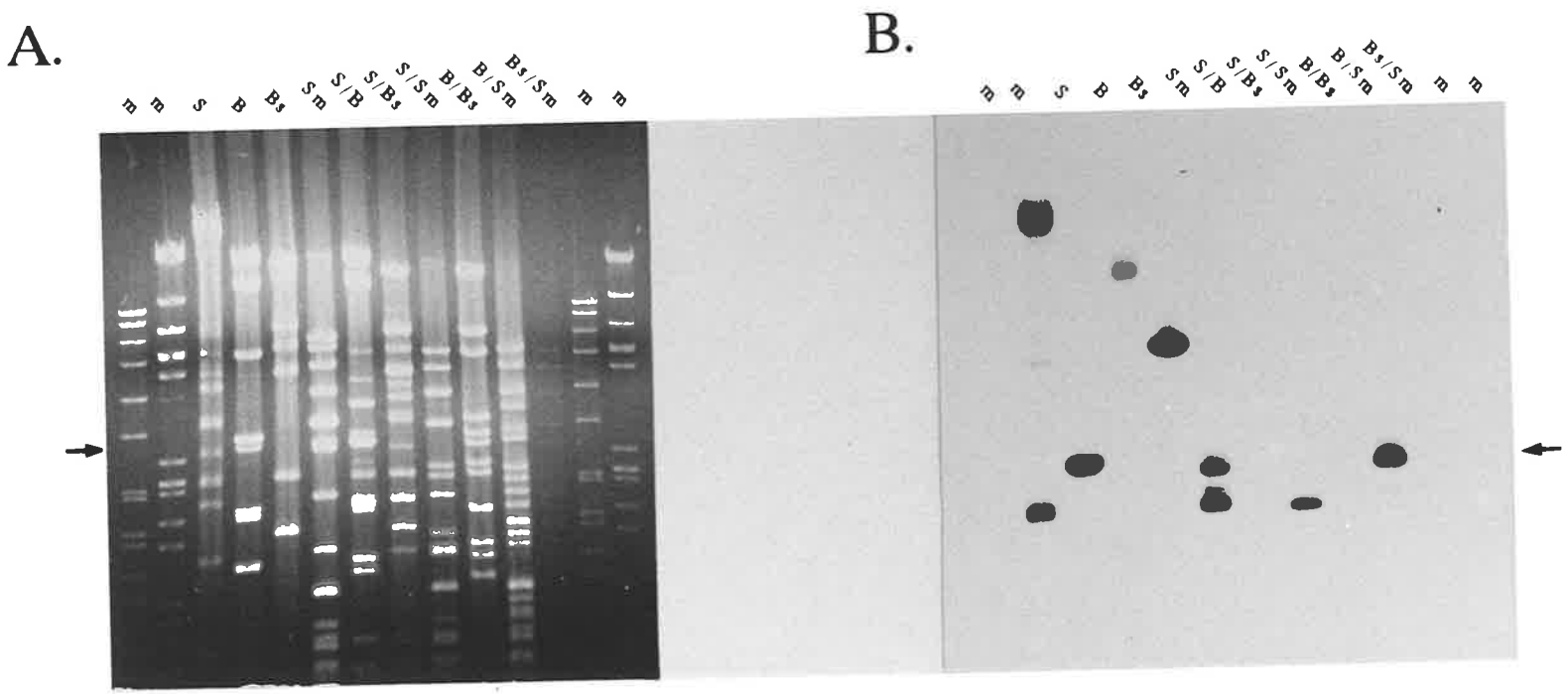
FIGURE 3.17. PARTIAL GENOMIC STRUCTURE OF IDUA.

The restriction map of A157.1 is shown with the structure of the IDUA gene known at this time. Exons are depicted as black boxes. The cluster of twelve exons on the right of the diagram were mapped within λID475 which is indicated under A157.1. Exon I of IDUA was identified using the oligonucleotide ID13 as a probe (see Figs. 3.14 and 3.19), approximately 13-kb from the clustered exons. Further work was required to identify if there was any coding sequence within the 13-kb intron between the exons marked I and III (marked as II? above the open box). Initial PCR results suggested there was no further coding sequence, however the interrupted tryptic peptide (Fig. 3.20), implied that there was in fact further coding sequence to isolate, suggesting the possibility of alternative splicing occurring as indicated by the lines joining the exons.

FIGURE 3.18. IDENTIFICATION OF THE 2.8-kb *Bam*HI GENOMIC FRAGMENT.

A. TAE agarose gel (0.8%) electrophoresis of A157.1 restriction endonuclease digested DNA. Tracks are marked with the restriction enzymes used as follows; S=*Sac* I, B=*Bam*HI, Bs=*Bss*HII, Sm=*Sma* I. The *Sac* I reaction was only partially digested. The marker tracks (m) are SppI phage digested with *Eco*RI and λ DNA digested with *Hin*dIII/*Eco*RI and *Hin*dIII, in that order. The position of the 2.8-kb *Bam*HI fragment identified in B is marked by an arrow.

B. A Southern blot of the gel in A probed with the oligonucleotide probe ID13. The 2.8-kb fragment which hybridizes to ID13 in *Bam*HI digests (see tracks B and B/Sm) was isolated for sequence analysis to identify exon I of *IDUA*.


A.
B.
m m S B Bs Sm S/B S/Bs S/Sm B/Bs B/Sm Bs/Sm m m

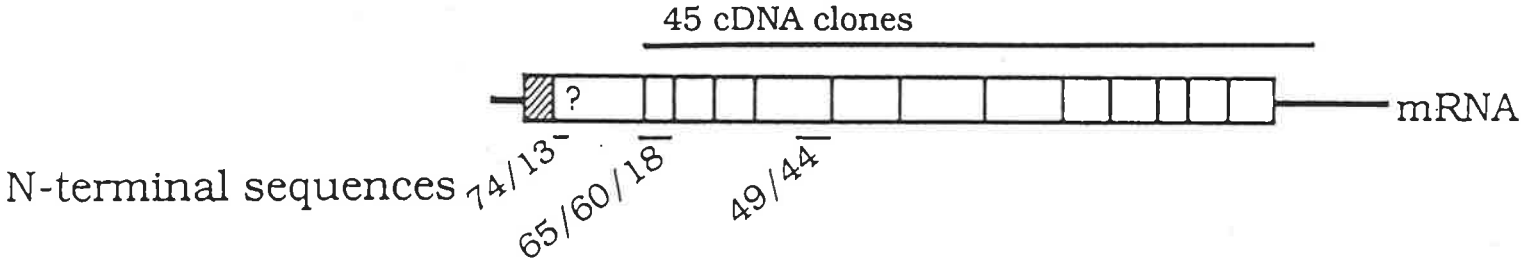
3.19). The oligonucleotide probe, ID2, did not hybridize to A157.1 as the exon I boundary interrupted the sequence of peptide 2'.

As described above, the oligonucleotide primer made to the 74/13-kDa N-terminal sequence (ID13; Fig 3.13A) hybridized to a 2.8-kb *Bam*HI restriction fragment in the cosmid clone, A157.1, approximately 14-kb 5' from the 850-bp *Pst* I fragment. Southern blot analysis of the cosmid A157.1, using the ID58 to ID61 PCR product as a probe, revealed the presence of exon II of *IDUA* in another 2.8-kb *Bam*HI restriction fragment, separated from the first 2.8-kb fragment by a further *Bam*HI fragment of 200-bp. Both 2.8-kb *Bam*HI restriction fragments were isolated and subcloned into either M13 or phagemid vectors and the areas of interest were completely sequenced in both directions using oligonucleotide primers. A PCR product was produced between the oligonucleotides ID80 and ID81 (Appendix C) and directly sequenced to orientate the 200-bp *Bam*HI fragment and link it to the sequence from the two 2.8-kb *Bam*HI fragments at the 5' end of IDUA. The position of subclones used to sequence *IDUA* are shown in Fig 3.23. All subclones used in sequencing regions spanned by λID475 were derived from λID475, except for p157.9 which was derived from the cosmid A157.1. A 550-bp *Eco*RI-*Kpn* I fragment spanning the *Pst* I restriction site between the 2.2-kb and 850-bp *Pst* I fragments, and a 1.15-kb *Kpn* I fragment spanning the *Pst* I restriction site between the 850-bp and 1.6-kb Pst I fragments, were used to link the sequences of these fragments.

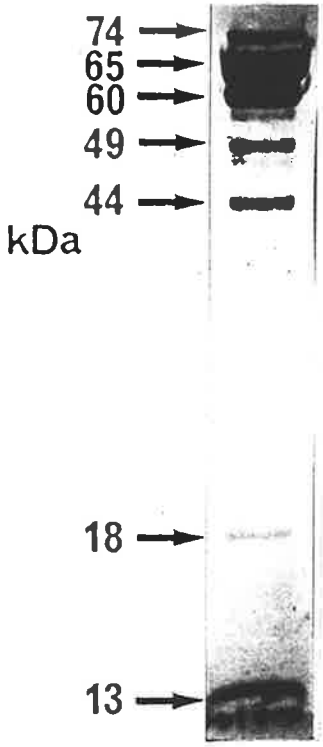
The IDUA cDNA clones, λRPC1 and λE8A, were used to map the positions of exons III to XIV in the lambda and cosmid clones (Fig. 3.23) and the intron exon boundaries (Table 3.5). Mapping of the positions of exons I and II in cosmid clone A157.1 was as described above. The sequence of the human IDUA gene is presented in two segments (Fig. 3.24). The gene for IDUA is split into 14 exons spanning approximately 19-kb. The first two exons are separated by a 566-bp intron and the last twelve exons are clustered in a 4.2-kb region. Intron 11 is one of the smallest recorded introns and has an atypical GC splice donor site (Table 3.5) which does however match the observed exceptions to the GT consensus (Penotti, 1991).

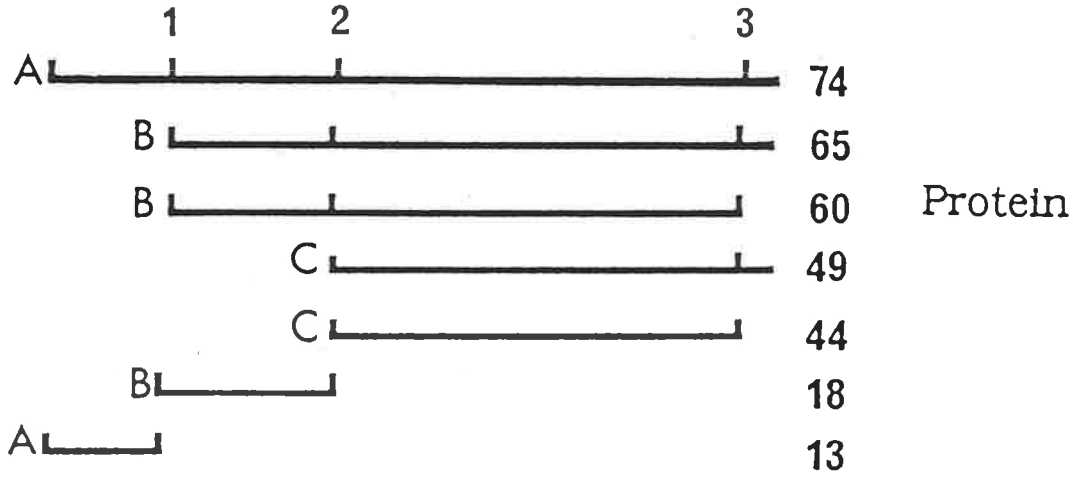
FIGURE 3.19. A POLYPEPTIDE MODEL FOR IDUA.

A model to connect the seven major polypeptides in immunopurified human liver IDUA present on an SDS/PAGE as shown on the left of the figure, with the polypeptide sizes indicated in kDa (Clements *et al.*, 1989). 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 or 3. The IDUA mRNA is shown above the peptide with the final positions of all N-terminal sequences of IDUA marked.



SDS-PAGE OF
PURIFIED IDUA





Northern blot (Section 3.6.1) and PCR analysis of *IDUA* are consistent with a low level of transcription and revealed only a 2.3-kb mRNA transcript, although alternate splicing is known to occur (see Section 3.8.1). No consensus polyadenylation signal was found in the 3' untranslated region of the cDNA clones isolated or the sequence reported here, although two variant polyadenylation signals consistent with the size of the mRNA are proposed (Fig. 3.24B). PCR has shown the mRNA to extend to position 496 (Fig. 3.24A). From the position of the proposed polyadenylation signals, the mRNA's produced would be 2203-bp and 2285-bp, with an additional 20-30-bp prior to the poly(A) tail, matching the size of the observed transcript when a 50 to 100-bp poly(A) tail is taken into account.

The potential promoter for *IDUA* is bounded by an Alu repeat sequence and has only GC box type consensus sequences (Fig. 3.24A). There are multiple transcription start sites (CA) consistent with a GC box promoter of a housekeeping gene with low levels of transcription (Bucher, 1990). From detailed mapping of cosmid clones in 4p16.3 it has been shown that the orientation of A157.1, and thus the *IDUA* transcription unit, is opposite to that reported by MacDonald *et al.* (1991b) (Lorne A. Clarke and Bernhard Weber, personal communication; Weber *et al.*, 1991a). Thus transcription of *IDUA* is towards the centromere.

3.7.2 THE MOLECULAR BASIS OF THE "VNTR-LIKE" POLYMORPHISM AT *D4S111*

IDUA is located on human chromosome 4 at 4p16.3, about 1100-kb from the telomere and flanked by the two most likely candidate locations for the Huntington disease gene (Section 3.5.3). The IDUA gene spans about 19-kb and has been shown to be coincident with the highly polymorphic multi-allele locus *D4S111*, which is used in Huntington disease diagnosis (MacDonald *et al.* 1989). *D4S111* maps 1100-kb from the telomere of 4p and 1000-kb from the two most likely locations of the Huntington disease gene. A "VNTR-like" polymorphism at *D4S111* was detected by Southern analysis with the probe p157.9 (MacDonald *et al.*, 1989). This probe is a plasmid

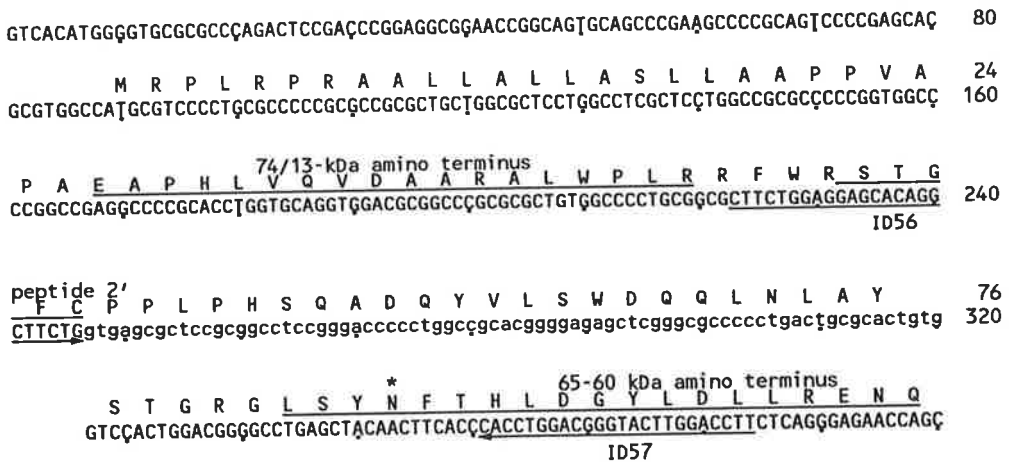


FIGURE 3.20. AN INTERRUPTED TRYPTIC PEPTIDE.

The sequence of exon I and part of intron I of *IDUA* is shown. The 5' end of λE8A is shown as it would join inframe to exon I. All matched peptide sequences are underlined. The 3' boundary of exon I was hypothesized from the genomic DNA sequence using positions of possible 5' splice sites, inframe stop codons and the sequence of tryptic peptide 2'. It can be seen that there are 23 unmatched residues from tryptic peptide 2', strongly implying the presence of further coding sequence for IDUA. The oligonucleotides used to test this hypothesis, ID56 and ID57, are underlined below the nucleotide sequence with the arrows indicating either sense (—►) or antisense (◄—).

FIGURE 3.21. PCRs TO DETERMINE IF THE FULL cDNA SEQUENCE FOR IDUA HAD BEEN OBTAINED.

The position of primers initially used to perform PCR from fibroblast cDNA are shown against exon I and the 45 IDUA cDNA clones that end at the same 5' base. Panel A shows the production of an 84-bp PCR product, implying no coding sequence existed between exon I of IDUA and the start of the cDNA clones. PCRs were performed between the primers ID58 and ID60 to ID61 on the IDUA mRNA as shown, in order to obtain a full-length IDUA cDNA clone. As shown in panel B, the PCRs produced multiple bands varying in intensity depending on the primers and the annealing temperatures used. The 60-61 PCRs were at 58°C and 65°C respectively. Sequencing of the longest of the PCR products revealed the sequence of exon II of IDUA. The smaller and fainter PCR products are the result of alternative splicing of IDUA.

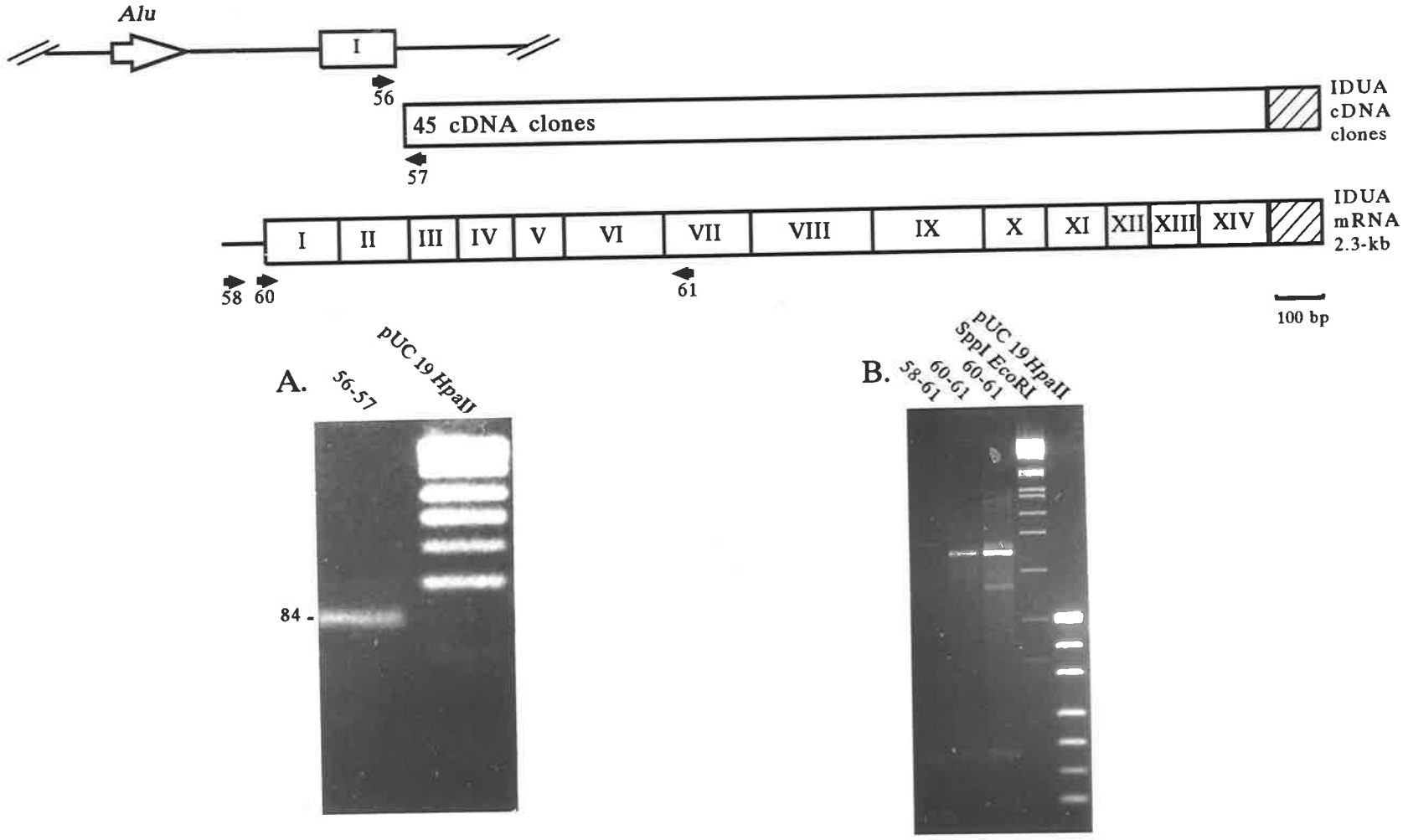

Alu
I
56
45 cDNA clones
57
IDUA
cDNA
clones
I
II
III
IV
V
VI
VII
VIII
IX
X
XI
XII
XIII
XIV
IDUA
mRNA
2.3-kb
58
60
61
100 bp
A.
56-57
pUC 19 HpaII
84 -
B.
58-61
60-61
60-61
SppI EcoRI
pUC 19 HpaII

FIGURE 3.22. THE FULL LENGTH cDNA SEQUENCE OF IDUA.

Compiled nucleotide sequence for IDUA cDNA and the deduced amino acid sequence of the protein. Amino acid sequence is shown in single letter code above the cDNA sequence. Nucleotide and amino acid numbers are shown in the right margin. The probable site of signal 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 marked with an asterisk. Oligonucleotides used to produce the full length cDNA or used in the alternative splicing studies are underlined below the nucleotide sequence with the arrows indicating either sense (→) or antisense (←). The cDNA clone λRPC1 extended from base 541 to 1269 and λE8A from base 391 to the 3' end of the sequence shown.

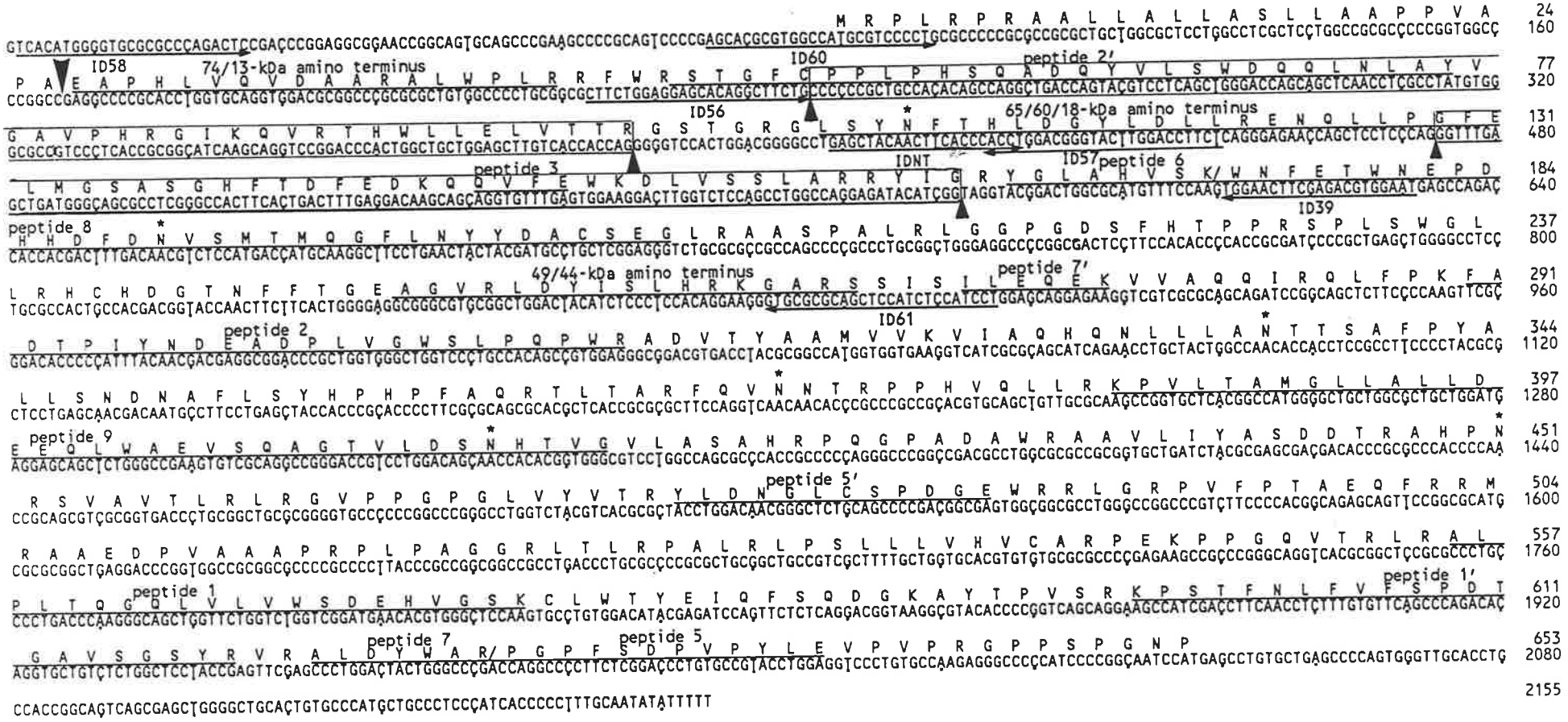

ID58
74/13-kDa amino terminus
ID60
peptide 2′
ID56
65/60/18-kDa amino terminus
peptide 3
IDNT
ID57
peptide 6
ID39
peptide 8
49/44-kDa amino terminus
peptide 7′
ID61
peptide 2
peptide 9
peptide 5′
peptide 1
peptide 1′
peptide 7
peptide 5

FIGURE 3.23. GENOMIC STRUCTURE OF *IDUA*.

A partial restriction map of the cosmid A157.1 as described by MacDonald *et al.*, (1991) with the addition of *Bam*HI sites, showing the position of the 14 exons of the IDUA gene. The position of the λEMBL3 genomic clone originally isolated (λID475), and subclones used for sequencing are shown below. All subclones used in sequencing regions spanned by λID475 were derived from λID475 except for p157.9. The subclone names in general describe the size and restriction enzyme used to isolate the fragment. For example, pIDGP1611 is the M13 subclone of the 1.6-kb *Pst* I fragment originally identified as being positive to the oligonucleotide probe ID47 (Fig. 3.5), and used in the chromosomal localization studies (see Section 3.5).

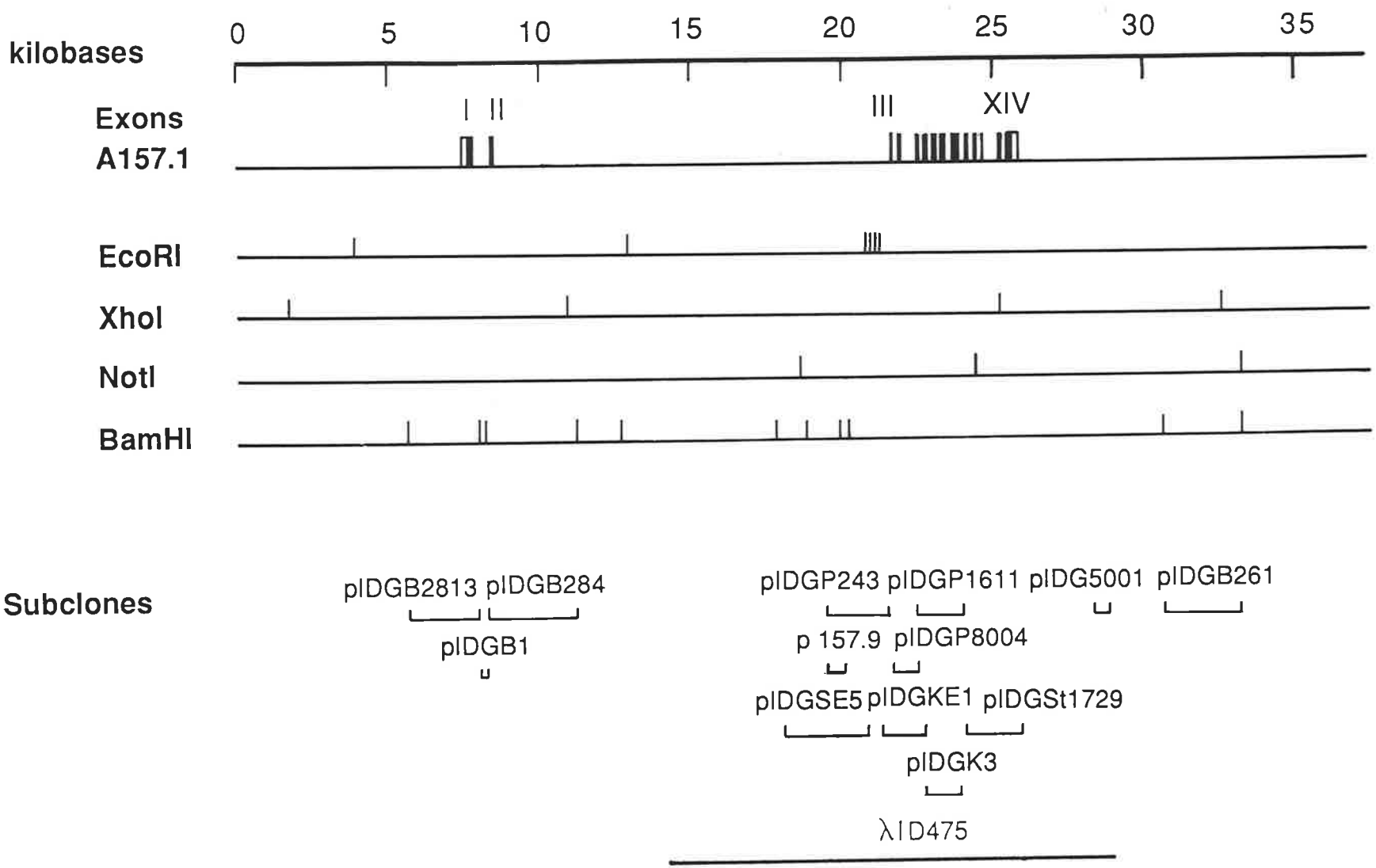

kilobases
0
5
10
15
20
25
30
35
Exons
A157.1
I
II
III
XIV
EcoRI
XhoI
NotI
BamHI
Subclones
pIDGB2813
pIDGB284
pIDGB1
pIDGP243
pIDGP1611
pIDG5001
pIDGB261
p 157.9
pIDGP8004
pIDGSE5
pIDGKE1
pIDGSt1729
pIDGK3
λID475

TABLE 3.5. IDUA GENE STRUCTURE.

Exon and intron sizes and nucleotide sequences of the exon/intron junctions in the human IDUA gene. The numbering of the cDNA position is according to the sequence presented in Fig. 3.22, Section 3.8. Consensus eukaryotic 5' and 3' splice sites according to Penotti (1991) are given underneath. The underlined dinucleotides residues are rigidly conservered in the consensus splice site sequence.

TABLE 3.5.

Exon	size (bp)	cDNA position	5' splice donor	Intron No	size (bp)	3' splice acceptor
I	>245	1->246	GGC TTC TG gtgagc	1	566	tttgttgtccccag C CCC CCG
II	141	247->387	ACC ACC AG gtgggc	2	~13 000	ccgtccttctgcag G GGG TCC
III	86	388->473	CTC CCA G gtgagc	3	184	ctcctgtgttccag GG TTT GAG
IV	108	474->581	TAC ATC G gtgggc	4	478	aggctcctctgcag GT AGG TAC
V	96	582->677	ATG CAA G gtgtgc	5	115	ctcccccggcccag GC TTC CTG
VI	203	678->880	CAC AGG AAG gtgcgc	6	100	ggccccgcccgcag GGT GCG
VII	180	881->1060	GTG GTG AAG gtgggc	7	107	ctggacacccgcag GTC ATC
VIII	217	1061->1267	CTG CTG G gtgagc	8	246	cccccgccccgcag AT GAG GAG
IX	213	1268->1490	GGC CCG G gtaagc	9	91	gtcaggccccgcag GC CTG GTC
X	122	1491->1612	GCG GCT GAG gtaggt	10	187	gccactgcgcccag GAC CCG
XI	126	1613->1738	CCC GGG CAG gcaagt	11	78	cttccctccccag GTC ACG
XII	77	1739->1815	GGC TCC AA gtgcgt	12	386	cctgacctccccag G TGC CTG
XIII	101	1816->1916	AGC CCA G gtgcgc	13	147	ccccttgtctccag AC ACA GGT
XIV	>239	1917->	concensus: $^{A}_{C}$AG gt$^{a}_{g}$aga			$\left(^{c}_{t}\right)_{n}n^{c}_{t}$ag G

A.

FIGURE 3.24. FULL SEQUENCE OF *IDUA*.

The sequence of the human 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 this 1.8-kb segment. The Alu repeat sequence and the four best potential GC 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.

B.

subclone derived from the cosmid, A157.1 containing an approximately 500-bp *Pst* I-*Sau*3A fragment. A157.1 has also been shown to contain the IDUA gene (Section 3.5.3 and MacDonald *et al.*, 1991b). The probe for the D4S111 polymorphism, p157.9, was provided by Marcy E. MacDonald and James F. Gusella of the Neurogenetics Laboratory at the Massachusetts General Hospital.

During the sequencing of λID475, which contains part of the IDUA gene, a 2.2-kb *Pst* I fragment from intron II immediately adjacent to exon III was isolated and sequenced. At this time, exons I and II had not been identified in the cosmid clone, A157.1, and this 2.2-kb *Pst* I fragment was fully sequenced in the hope that it contained another *IDUA* exon. No exons were found, however, the 2.2-kb *Pst* I fragment contained a repetitive element of 86-bp that was shown to be unique in the genome by Southern analysis (data not shown). The sequence of this 2.2-kb fragment is shown in Fig. 3.25A. With the acquisition and subsequent sequencing of p157.9, it was found to extend from a *Pst* I site at nucleotide position 1 to a *Sau*3AI site at position 512 within the 2.2-kb *Pst* I fragment. Analysis of sequences from p157.9 and λID475 revealed that this sequence contained a polymorphism at position 255, i.e. G in λID475 and A in A157.1. An *Alu* repetitive sequence (nucleotide positions 525 to 795, Fig. 3.25A) was found that begins 12-bp 3' to the *Sau*3AI site at the end of p157.9. The sequence shown in Fig. 3.25A contains 7 copies of an imperfect repeat, 5 of which contain 86-bp and 2 of which have an additional base pair inserted at the same relative position. All seven repeats are slightly different as revealed by a comparison of their sequences shown in Fig. 3.25B. When primers ID53 and ID54, which flank these repeats, were used to PCR amplify this region (for conditions see Table 4.3), it was found that there were 3 common alleles in the population which vary in size by increments of 86-bp corresponding to a varying number of repeats. These alleles are numbered 1 to 3 in decreasing order of size, allele 1 having 7 copies of the repeat, allele 2 having 6 copies and allele 3 having 5 copies (Fig. 3.26). In addition to the 3 common alleles, it has been shown that 2 rare alleles exist which contain 8 and 9 copies of this repeat. The frequency of the alleles determined in 136 chromosomes from 68 normal healthy individuals is 0.28 (allele 1), 0.55 (allele 2) and 0.17 (allele 3). These alleles correlate

exactly with the alleles observed at *D4S111*, using the probe p157.9, by MacDonald *et al.* (1989). The variation in repeats creates a unique *Sty* I site in repeat 2 and a unique *Bgl* I site in repeat 6 of allele 1 (see Fig. 3.25B). In order to determine if there is variation in the order of the repeats, which would potentially make the locus more polymorphic, 10 of each of the 3 common alleles and the 2 rare alleles from the *D4S111* locus were digested with these enzymes. This revealed that the relative position of repeats 2 and 6 remains constant and that the observed variation occurs between these two repeat units (data not shown).

3.8 SEQUENCE OF ENTIRE CODING SEQUENCE FOR IDUA

3.8.1 ALTERNATE SPLICING OF IDUA mRNA

As described in Section 3.7.1, 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 (Fig. 3.22) and several minor products that differed in intensity depending on the annealing conditions used (Fig. 3.21B). The minor products also hybridized to an internal oligonucleotide, ID56 (data not shown), indicating that they were representative of alternatively spliced mRNA species from the IDUA gene as has been reported for a number of other genes including human lysosomal hydrolases (β-glucuronidase, Oshima *et al.*, 1987; β-galactosidase, Morreau *et al.*, 1989; acid sphingomyelinase, Quintern *et al.*, 1989). It was thought to be important to identify these other IDUA mRNA species in order to determine which of these species encoded active IDUA protein. The exon structure of *IDUA* had been determined (Fig. 3.23). PCR of normal fibroblast cDNA using the oligonucleotide pairs ID56 to ID57 and IDNT to ID39 (annealing at 60°C and extending for 45 s), produced two distinct products per reaction. The smaller bands were isolated and directly sequenced showing alternative splicing of exons II and IV of the IDUA gene respectively (Fig. 3.27). The polypeptides from these alternatively spliced IDUA mRNA species maintained the translation frame for the IDUA protein (see Fig. 3.27) leaving the primary sequence of the translated peptide identical to the deduced IDUA

FIGURE 3.25. SEQUENCE OF THE 2.2-kb *Pst* I FRAGMENT CONTAINING *D4S111*.

A. Sequence of the 2.2-kb *Pst* I fragment from the IDUA gene. The Alu repeat sequence is underlined and individual 86-bp repeats are boxed. Specific oligonucleotides were made every 200 to 400-bp to fully sequence areas of interest in both directions by dideoxy nucleotide sequencing of M13 subclones. The oligonucleotides used to amplify the VNTR, ID50 and ID51, are underlined below the nucleotide sequence, with the arrows indicating either sense (→) or antisense (←).

B. The sequence of the first (5') 86-bp repeat unit and its comparison to the other 6 repeat units. The presence of sequence differences is indicated with the varying bases aligned under the sequence. The unique *Sty* I and *Bgl* I restriction sites, created by sequence variation, that were used to test if the order of the repeating units remained constant in different alleles are underlined.

A.

CTGCAGATGCAGCACAAAGCCAGCTGCCACAGGAGGTGCACTGCCCCACGGGCAGCTGTGCTCACCTGCGTGGCGTCCGCAGGAAGCCCT 90

CAGACCCTGGCCAGCCTGGTCTCTGCTTCTCACTCCAGGCTCAAGGCTGGCTGCTATTCCATTCACCCCATTCCCTTTCCCTCAGGCCAG 180

AAGCATCACTCACAACTCCAGACCCTTGCACACCAATCTCCTCACTCAGTTCACCCTCAGATGTCCAACTCCACGTGGCCAAAACCGAGA 270

CCAGTCTCACTCCAAGCTCATCCACCCGCCTAGCTGCCCAAGCTGAAATCCACAGTTGCCCTGTACGACCCCCCGACCCCCCCACCTCAC 360

CCCCCAGCTCACCTCAAACCTCATCAGTCCCTCTGCTCACAGCCTGCACAGCCACCACCCAGTCTGGGCACCATACCTCTCTGGGCAACC 450

CCCATGACCCCCAGCCAGTCTTCCAGCTCCCACCCTGACTCCAGTGAGCTGCCAGAGGGATCCTTGAAAAATGCTCTGGAAGCCAAGGCG 540

GGCGGATCACGAGGTCAGGAAATGGAAACCATCCTGGCTAACATGGTGAAACTCCGTCTCTACTAAAAATACAAAAAAATTAGCCGGGCG 630

TGGTGGCGGGCGCCTGTAGTCCCAGCTACTCAGGAGGCTGAGGCAGGAGAATGGTGTGAACCCGGGAGGGGGAGCTTGCAGTGAGCCGAG 720

ATCGCGCCACTGCACCCCAGGCTGGGCAACAGAGTGAGACTCTGTCTCAAAAAAAAAAAAAAGAAAAAAGAAAAATGCCCCTGGCACTGC 810

TCAGCTCAGAACCTTGAAAGGGATCCTTGTTTCGCTCTGGGTCAGGGTCCCATCCTCAAAGCAGTCACTTCCAGAGCCCCATGGAGCCTC 900

ATATCCCATGATATCCACCCTTATTCTGTCTCAGGCCCACCAGCTGCCTCCCTGCTGCCTGATTGTCATGTGTGCATTCCTGTTCCAGAG 990

CCTTTGCACTGGCGATGCCATCTGCTGCCTGGATGTCTTCCTGCTGGCTGGTCCCGTTGCATGGTAGCCTTAGGTGTCTCCTCAGAGAGG 1080

ID53

TCCCTCGCTGACCATCAGGCTCCTCACTCCCTGTCGTATCCCCTTCACCAAAGATTCCCATTCCCTGACGACGCCTGTCCCCTCGGAATG 1170

CAGGCCTCGTGGGAATTCAGCCCATCCGCAAGTGCAGTGGCAGGGCGGCCCCCTTCCCCCTTTGACGACGCCTGTCCCCTTGGAATGCAGG 1260

CCTCGTGGGAATTCAGCCCATCTGCAAGTGCAGTGGCAGGGCGGCCCCCTTCTCCCTTTGAGGACGCCTGTCCCCTCGTAATGCAGGCCTC 1350

GTGGGAATTCAGCCCATCTGCAAGTGCAGTGGCAGGGCGGCCCCCTTCTCCCTTTGAGGACGCCTGTCCCCTCGTAATGCAGGCCTCGTGG 1440

GAATTCAGCCCATCTGCAAGTGCAGTGGCAGGGCGGCCCCCTTCTCCCTTTGAGGACGCCTGTCCCCTCGGAATGCAGGCCTCGTGGGAAT 1530

TCAGCCCCATCTGCAAGTGCAGTGGCAGGGCGGCCCCCTTCTCCCTTTGACGACGCCTGTCCCCTCGGAATGCAGGCCTCGTGGGAATTCA 1620

GCCCCATCGGCAAGTGCAGTGGCAGGGCGGCCCCCTTCACCCTTTGACGACGCCTGTCCCCTCGGAATGCAGGCCTCATGGGAATTCAGCC 1710

CATCTGCAAGTGCAGTGGCAGGGTGGCCCCCTTCTCCCTTTCCTGTGCACTCATGTTGCCTCTTGGGGTGTGGGAGGGGAAATGGGGCAC 1800

TCCTGGGCCTCCAGGAGGTGCAGAGAACCAGGGTGAGGTGTCCACCAGGTCCTGCCTGGCTCCTGACCCCTGGCCCCTGCTGCTCGCGAC 1890

ID54

TGGCCTGCCTCGTGCCACTGAGCCTCAGAGCCATTCCGAACCCCCACCCCAAGTTTTCCATCTCTTGATGGTGTAGGGTTGGGGGGTCTC 1980

CATGTACAGATACTCTAGTTCATACCAGGCCTTCATAGGGTTATTTTCCAAGGGGAAGGGCCCCTCGGGAAGCCGGGATCGGAGTCCTGT 2070

GTGGCACCTTGCAGGCTCCCACATGCTCCGTTGTGGCCACGGTTCCAGCCTGGAGCATGGAGCTGTGTGGGCACCCTGCTTCCTGACGCT 2160

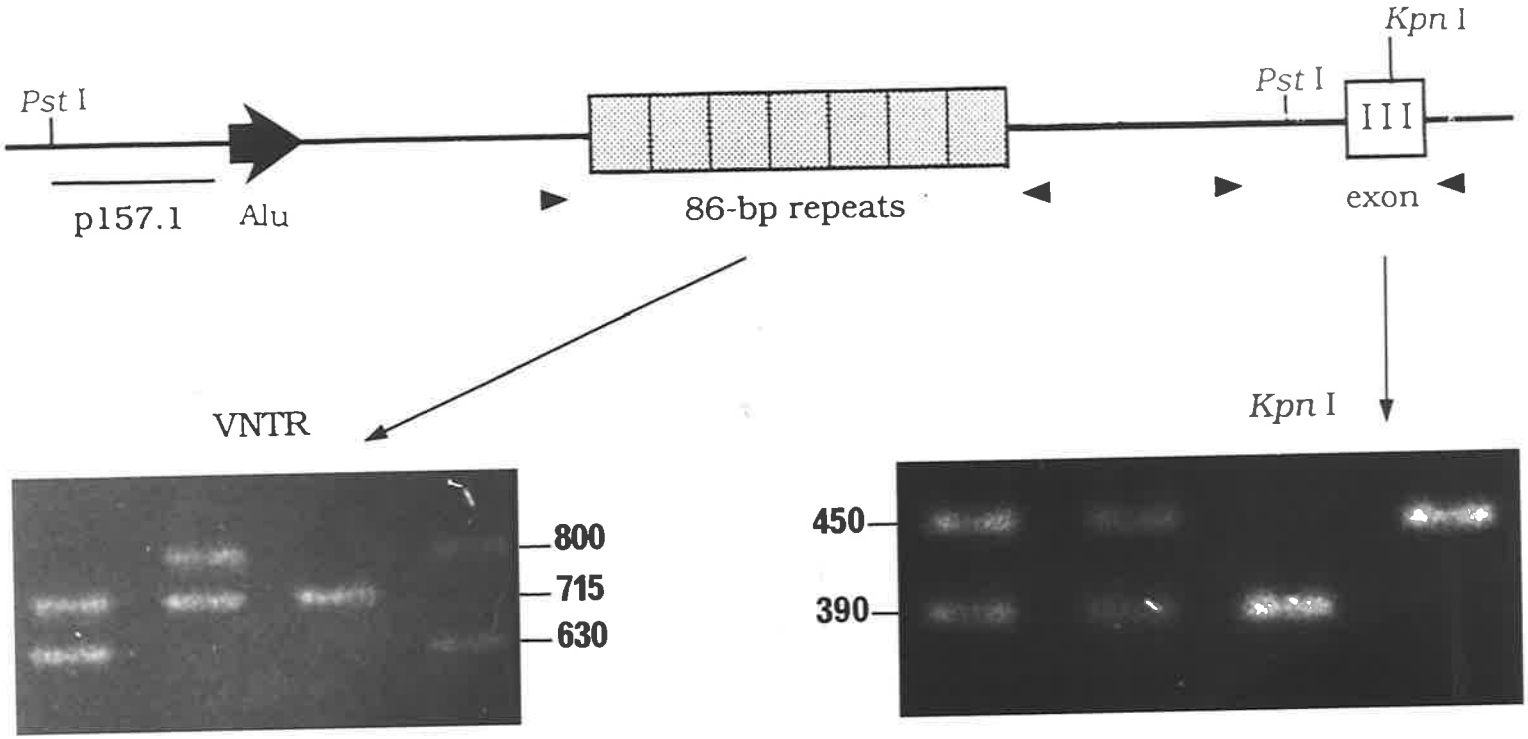
GACCGTCCTTCTGCAG 2176

B.

```
1  TGACGACGCCTGTCCCCTCGGAATGCAGGCCTCGTGGGAATTCAGCCC-ATCCGCAAGTGCAGTGGCAGGGCGGCCCCCTTCCCCCT
2  ..............T..................................T..............................T....
3  ...G.....................................................T..............................T....
4  ...G..............T.....................................T..............................T....
5  ...G.....................................................C...T..........................T....
6  ..........................................................C...G..........................A....
7  ..................................A..................T..................T..........T....
```

FIGURE 3.26. DIAGRAMMATIC REPRESENTATION OF THE 2.2-kb *Pst* I FRAGMENT AND PCR ANALYSIS OF *D4S111*.

Diagrammatic representation of the 2.2-kb *Pst* I fragment containing the VNTR polymorphism in intron 2 of *IDUA*. Exon III of *IDUA* containing the *Kpn* I polymorphism is also shown. The arrows represent the oligonucleotide primers used for PCR detection of these polymorphisms as shown below and described in Section 4.2.5 and Table 4.3. The sizes of the alleles detected is indicated in base pairs on the side of the gels.


Pst I
p157.1
Alu
86-bp repeats
Pst I
Kpn I
III
exon
VNTR
800
715
630
Kpn I
450
390

peptide, except for the omission of 47-amino acids and 36-amino acids respectively. Thus, the alternatively spliced mRNA species individually missing exons II and IV would produce peptide products of 606 amino acids and 617 amino acids respectively.

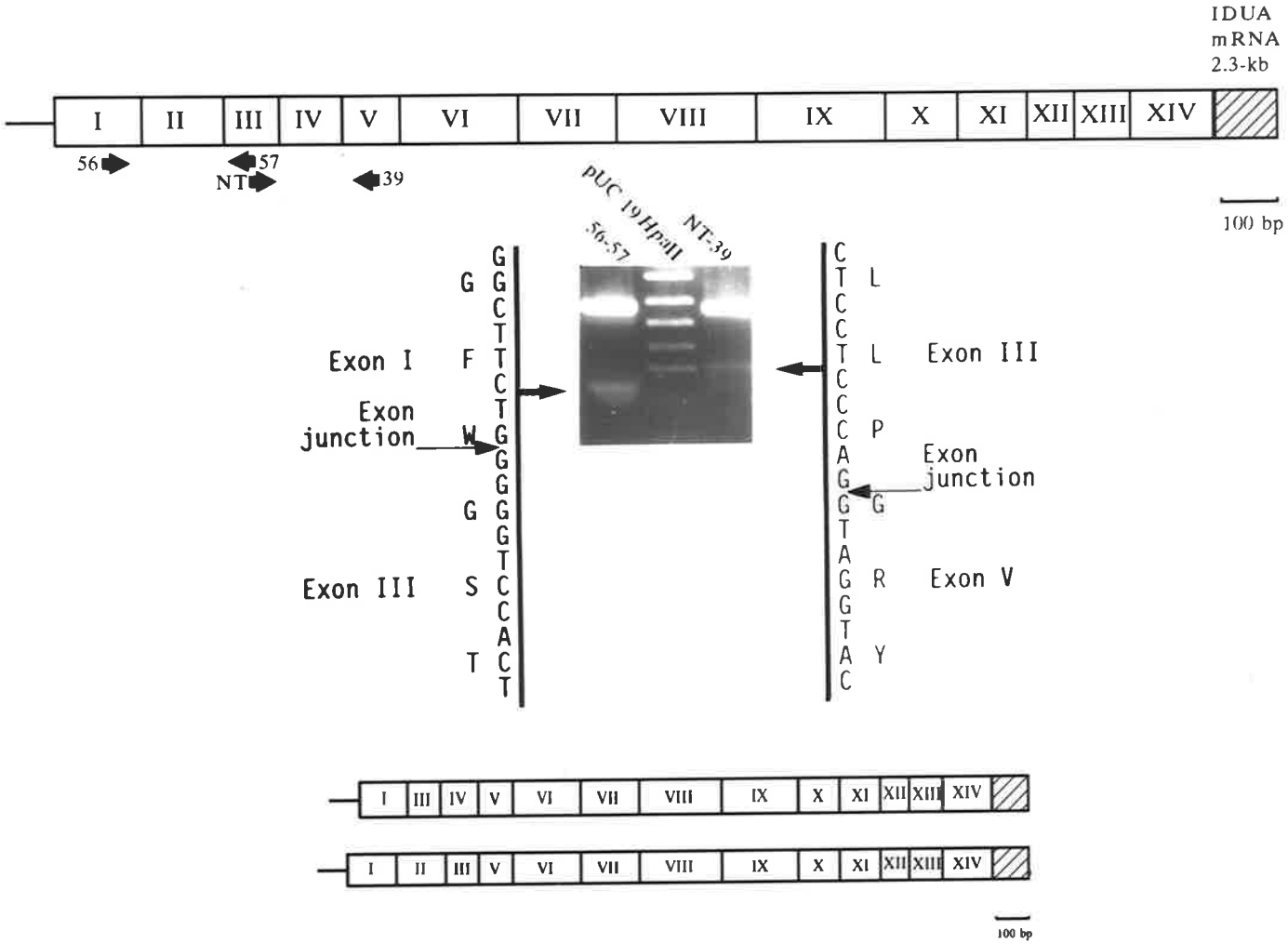
The two distinct PCR products, produced by PCR with the primer pairs ID56 to ID57 and IDNT to ID39 (Fig. 3.27), remain at the same relative intensity when using reverse transcribed liver, kidney or placental RNA as template (data not shown), which indicated that this alternate splicing does not appear to be tissue specific and that they may be minor mRNA species not detectable by Northern blot analysis either due to their low abundance or the resolution of the Northern blot. The alternative splicing of exons II and IV interrupts the observed tryptic peptide sequence for peptide 2' and the 65/60/18-kDa N-terminal sequence of IDUA respectively (see Fig. 3.22). The alternative splicing of exon II also introduces a Trp into the amino acid sequence at the splice junction. As both of the reported splicing events interrupt peptides sequenced from purified IDUA, it was thought that the major PCR product was most likely to represent the full-length mRNA encoding IDUA. Expression of this putative full-length mRNA would establish that the nucleotide sequence presented in Fig. 3.22 encodes enzymatically active IDUA.

3.8.2 CONSTRUCTION OF A FULL LENGTH IDUA cDNA

cDNA from a mixture of normal human fibroblast cell lines was used for PCR as described in Section 2.2.21 using the primers ID60 and ID61. Denaturation was at 94°C for 45 s, annealing was at 58°C for 45 s and extension was at 72°C for 2 min. ID60 spans the initiating ATG codon and has a *Hin*dIII restriction site with a 4-bp GC clamp on the 5' end. ID61 is approximately 100-bp 3' of a unique *Kpn* I restriction site (bases 818 to 823, see Fig. 3.22). Utilising the *Hin*dIII and the *Kpn* I sites, the 840-bp PCR product was directionally cloned into a pTZ19 vector which contained the rest of the IDUA coding sequence from the *Kpn* I site to the *Eco*RI cloning site of the clone λE8A, to produce a full-length IDUA cDNA clone (Fig. 3.28). This full-length IDUA

FIGURE 3.27. ALTERNATIVE SPLICING OF IDUA mRNA.

PCR of reverse transcribed normal fibroblast RNA showing the alternative splicing of exons II and IV. The position of primers on the full length IDUA mRNA are shown above the gel and a diagrammatic representation of the two minor species of IDUA detected are shown below. The PCR between ID56 and ID57 produced a major product of 225-bp and a minor product of 84-bp. The PCR between IDNT and ID39 produced a major product of 222-bp and a minor product of 114-bp. Partial sequences of the two minor products and their encoded amino acids are at the left and right of the figure. The position of the missing exon is indicated by the arrow labelled "Exon junction".

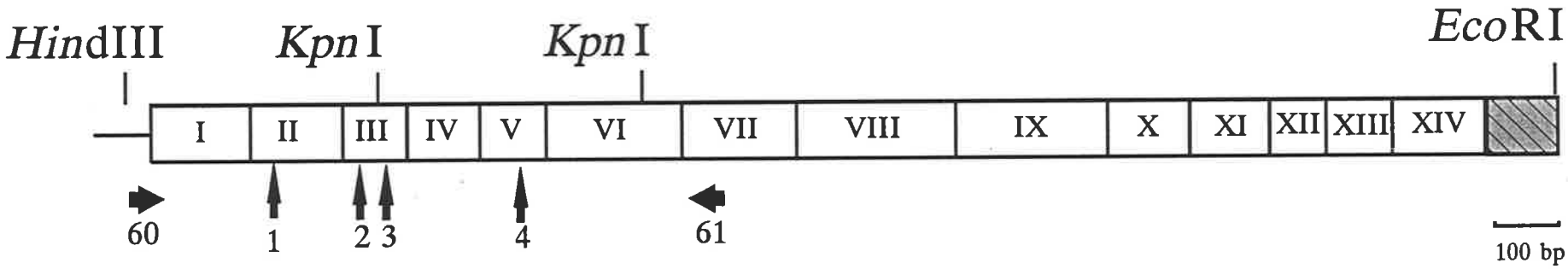

IDUA
mRNA
2.3-kb
I
II
III
IV
V
VI
VII
VIII
IX
X
XI
XII
XIII
XIV
56
57
NT
39
100 bp
pUC 19/HpaII
56-57
NT-39
Exon I
Exon junction
Exon III
G
F
W
G
S
T
GGCTTCTGGGGGGTCCACT
CTCCTCCCCAGGTAGGTAC
L
L
P
G
R
Y
Exon III
Exon junction
Exon V
100 bp

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

In all, 48 clones were analysed and only one, IDX21, was found to be correct (full-length). In addition, sequence analysis of this full-length insert revealed that four nucleotides were different from the previously determined sequence (Figs. 3.22 and 3.28). The differences were A to C (base 276, Fig. 3.22), G to A (base 402), T to C (base 440) and T to C (base 631). The first 2 differences alter the amino acid residues coded for by the cDNA from Gln to Pro (amino acid 63) and Arg to Gln (amino acid 105) respectively. The first amino acid substitution is a non-conservative substitution of a neutral-polar amino acid (Gln, position 63) for a small neutral-non-polar amino acid known for its "turn-inducing" properties (Pro). The second amino acid substitution is a conservative substitution of a basic amino acid (Arg, position 105) for a neutral-polar amino acid (Gln). The A to C base change (base 276, Fig. 3.22), which alters the amino acid residue encoded at position 63 of IDUA from Gln to Pro, created an *Spl* I (Amersham) restriction endonuclease site. This amino acid substitution is not a conservative substitution, thus it was further analysed to demonstrate if this change was a polymorphism. This base change would be occur exon II of IDUA and thus exons I and II of IDUA were PCR amplified (with the primers ID60 and ID66 as described in Section 4.2.2) from the genomic DNA of 20 normal individuals and digested with *Spl* I. No *Spl* I restriction sites were found suggesting that if the Gln to Pro change is a polymorphism, it must be very rare. Since mRNA or fibroblasts were not available at this time from the tissue source the full-length cDNA insert was derivred from, it was not possible to use PCR to directly to detect this base change from its potential source. The T to C base substitution (440) is a silent change which alters a Leu (amino acid 118) codon from TTG to CTG and introduces a second *Kpn* I site into the cDNA. Thus the cloned PCR product presumably resulted from partial digestion with *Kpn* I or the ligation of three fragments. The last change, T to C (base 631), is also a silent change in the third base of an Asn (amino acid 181) codon. The 2 silent wobble base changes have subsequently been shown to be polymorphic but the two amino acid changes are

FIGURE 3.28. CONSTRUCTION OF A FULL LENGTH IDUA cDNA.

Of the IDUA cDNA clones isolated, 45 clones began at the exon II/ III boundary at their 5' end. PCR was performed between the primers ID60 and ID61. The primer ID60 has a *Hin*dIII linker on the 5' end. This PCR product was digested with *Hin*dIII and *Kpn* I and ligated to the 3' *Kpn* I/ *Eco*RI fragment from λE8A to produce a full length IDUA cDNA. The *Eco*RI site at the 3' end of the cDNA is from the linker of λE8A. This full length cDNA insert was sequenced and the four nucleotide changes from the sequence previously observed are arrowed and listed below, together with the amino acid substitutions produced. The 5' most *Kpn* I site was created by change 3.



	1	2	3	4
Nucleotide changes	$A_{276}C$	$G_{402}A$	$T_{440}C$	$T_{631}C$
Amino acid changes	$Q_{63}P$	$R_{105}Q$	–	–

most likely transcription errors introduced by Taq polymerase during PCR in the presence of high concentrations of dNTP's (400 μM) for 40 cycles (Eckert and Kunkel, 1990). However, these conditions were essential to produce enough PCR product to conduct the experiment. It was unsuccessfully attempted to use Vent™ polymerase (NEB), a high fidelity, thermostable polymerase, in place of Taq polymerase to avoid the possibility of errors in the full-length IDUA cDNA insert.

3.8.3 EXPRESSION OF IDUA ACTIVITY FROM THE FULL LENGTH IDUA cDNA

Donald S. Anson of the Department of Chemical Pathology at the Adelaide Children's Hospital performed the transfection of the CHO cells and IDUA enzyme assays.

This full-length cDNA construct (IDX21) was excised using *Hin*dIII and *Eco*RI and directionally cloned into the expression vector pRSVN.07 (which drives expression of the insert from the Rous Sarcoma Virus LTR; Anson *et al.*, 1992b) to produce the construct pRSVNID21. CHO cells were electroporated 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 using an IDUA specific fluorogenic substrate (Clements *et al.*, 1985). CHO cell extracts contained a low level of IDUA activity. Cellular extracts from CHO cells transfected with pRSVNID21 gave a total activity 160-fold greater than the background (CHO cells) and ten-fold greater than the control normal human fibroblast activity (Table 3.6). In order to compare the specific activities of the recombinant and fibroblast IDUA, serial dilutions of the cellular extracts were assayed in parallel using the human IDUA specific Id1A monoclonal antibody based immunocapture (Section 3.2.2c and Scott *et al.*, 1990) and ELISA assays (Ashton *et al.*, 1992). The CHO cell extract gave zero background in both assays. The ELISA result was normalized to the normal fibroblast extract and showed a 12.7-fold higher level of expression of human IDUA in the

TABLE 3.6 EXPRESSION OF IDUA

Expression of IDUA activity in CHO cells, CHO cells transformed with the full length IDUA cDNA expression construct pRSVID21. IDUA activity is in pmol/min/mg of cell protein x 10^{-2}. IDUA activity was measured before (Total) and after capture of IDUA protein in the immunocapture assay (Captured). The amount of human IDUA protein captured in the ELISA assay was normalized against cell protein (per mg) in normal human fibroblasts. Specific activity is expressed as IDUA activity relative to IDUA protein. ND = None detected.

Cell type	IDUA activity		Relative IDUA protein	Relative IDUA specific activity
	Total	Captured		
CHO	1	ND	ND	–
CHO with pRSVNID21	160	152	12.7	12.0
Normal human fibroblasts	16	12.6	1	12.6

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 a similar specific activity (Table 3.6). SDS-PAGE analysis of immunopurified human IDUA protein expressed into CHO cell medium in the presence of NH_4Cl revealed a molecular mass of 80-kDa (D. Brooks, personal communication), corresponding to the predicted molecular mass of the mature 627 amino acid human IDUA protein (Fig. 3.22) of 70,029-Da. This is consistent with the previous estimates of IDUA size after allowing for post translational modifications (Myerowitz and Neufeld, 1981; Clements *et al.*, 1985; Clements *et al.*, 1989; Taylor *et al.*, 1991). 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.

The two amino acid changes that were introduced into the IDUA protein encoded by the full-length IDUA cDNA insert, IDX21, obviously do not affect the activity or stability of the expressed IDUA protein. It is interesting to note with hindsight that neither of the two amino acids altered in IDX21 are conserved between human and canine IDUA (Fig. 3.29) (Stoltzfus *et al.*, 1992). In fact, the second amino acid change, Arg to Gln at position 105, is a substitution between the human and canine IDUA at this postion, implying that this amino acid alteration would have little or no affect on the normal activity of IDUA.

Prior to proving that pRSVNID21 produced IDUA of normal specific activity, it was considered that the two amino acid changes discussed above could prevent the correct expression of IDUA. Also, it was considered that for future use of a full-length IDUA cDNA insert for gene or enzyme replacement therapy, or production of IDUA for X-ray crystallographic studies, the insert should encode the correct IDUA peptide sequence as reported in Fig. 3.22. Thus, an alternative expression experiment was designed. Briefly, oligonucleotides of approximately 50-nt overlapping by around 25-nt, were designed to the IDUA cDNA sequence in Fig. 3.30 to produce a double stranded 394-bp insert from the polymorphic *Kpn* I restriction site in IDX21 to the

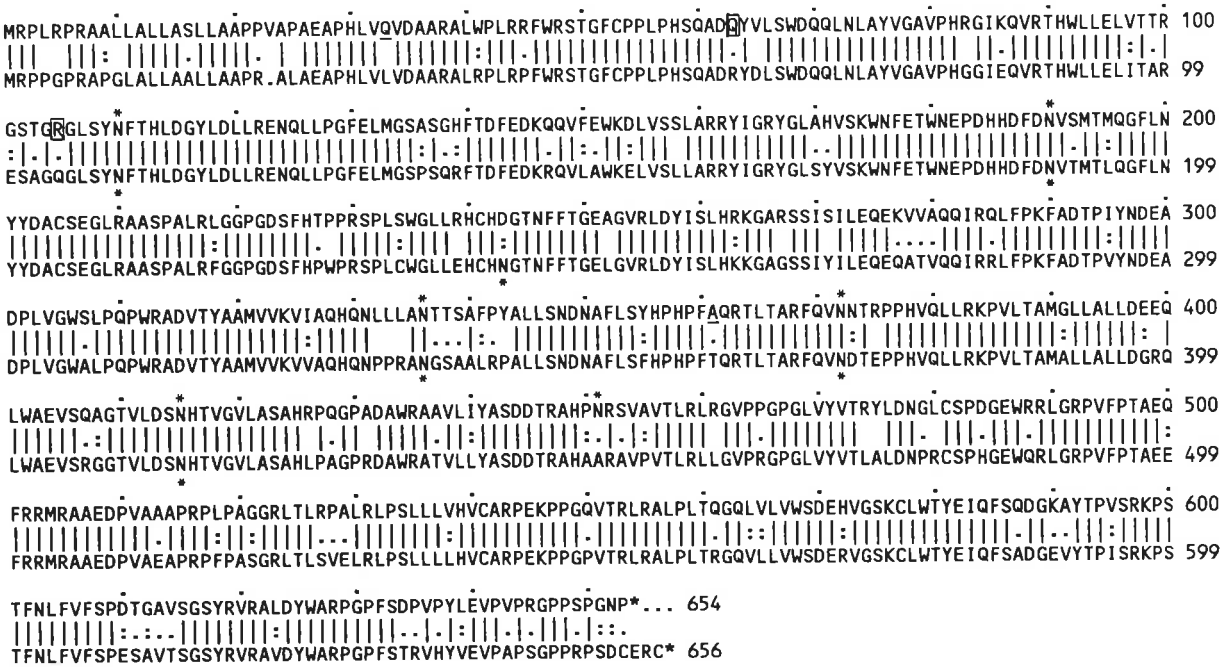


FIGURE 3.29. COMPARISON OF THE HUMAN AND CANINE IDUA PROTEINS.

A GAP sequence alignment of the human (top) and canine (bottom) IDUA proteins. Matches are marked by a line (|), conservative substitutions by a colon (:) and less conservative substitutions by a dot (.). Non-conservative substitutions are unmarked. The proteins are 82% identical, with an additional 6% similarity. Amino acids that have subsequently been found to be polymorphic in human IDUA are underlined, and the two amino acids that are altered in the full length IDUA expression construct are boxed. Potential N-glycosylation sites are marked with an asterisk.

initiating methionine of IDUA. A single nucleotide in a wobble base of amino acid G_{101} was altered from G->A (base position 391 Fig. 3.22 or 346, Fig. 3.30) to introduce a *Bam*HI restriction site into the insert that could be used to "cassette" the IDUA cDNA for production of mutant IDUA cDNA constructs in the future. The rat preproinsulin 5' untranslated sequence was used instead of the 5' untranslated sequence from IDUA as it had previously been shown to increase levels of expression (Cullen, 1988; Anson *et al.*, 1992b). The sequence of these oligonucleotides and the resulting IDUA insert is shown in Fig. 3.30. These 17 oligonucleotides were to be mixed in equimolar proportions, kinased, annealed together to form a double stranded DNA insert with *Hin*dIII and *Kpn* I overhanging ends, and ligated into an appropriate vector to result in a full-length IDUA cDNA encoding the amino acid sequence shown in Fig. 3.22. This experiment has now been performed and expression of this insert in CHO cells has, as expected, resulted in expression of active IDUA protein (Jill Durrant, unpublished results).

3.10 CONCLUDING DISCUSSION

The total coding sequence for IDUA has an open reading frame of 1959-bp encoding a peptide of 653 amino acids. The IDUA cDNA is GC-rich, with 69% of GC over the entire sequence and 80% for the 5' untranslated sequence and the area encoding the signal peptide. A signal peptide of 26 amino acids with a consensus cleavage site (von Heijne, 1986) was present immediately adjacent to the mature amino terminus of the protein (74/13-kDa amino terminus). All major peptide species sequenced (Tables 3.2 and 3.3) are present in the translation of the open reading frame, totalling 234 amino acids (42%) of 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 Fig. 3.22 supports the proposed model of proteolytic processing of the 74-kDa IDUA polypeptide (Fig. 3.19). The protein sequence is rich in arginine residues which are responsible for the basic nature of the protein (pI calculated to be 9.25 using the GCG program PeptidePlot). These

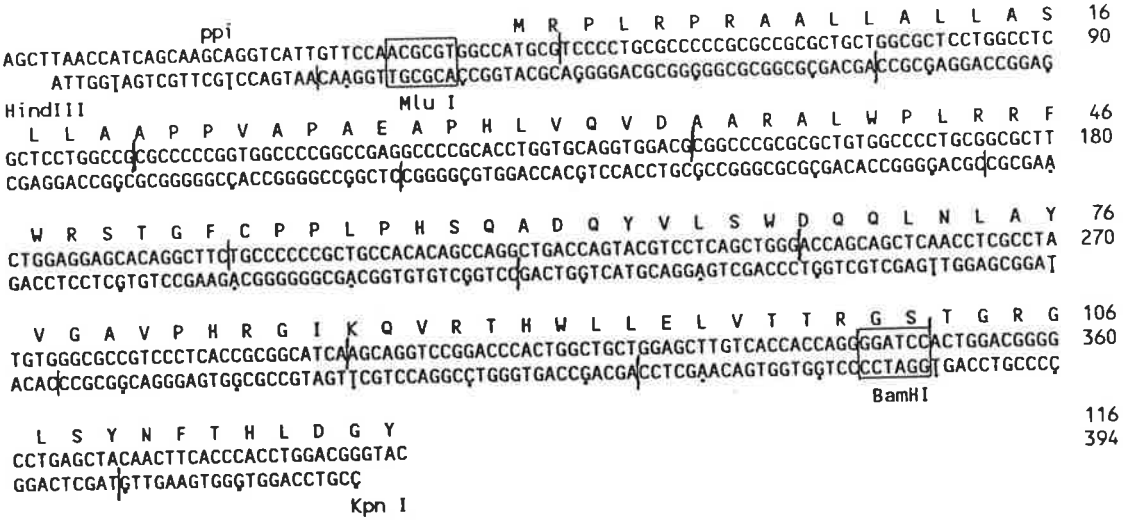


FIGURE 3.30. ALTERNATIVE IDUA EXPRESSION CONSTRUCT.

The sequence of the 394-bp IDUA insert synthesized by the annealing of oligonucleotides is shown. This insert was used to make an IDUA expression construct without the two amino acid errors reported in Fig. 3.28 and spanned from the 5' polymorphic *Kpn* I site to the initiating methionine. The oligonucleotides used to construct the 394-bp insert are separated by vertical lines. The *Mlu* I site that marks the end of the IDUA sequence is boxed. The 5' untranslated sequence is from the rat preproinsulin II gene (ppi). The *Hin*dIII and *Kpn* I restriction sites at the ends of the insert used for cloning are marked. The *Bam*HI site introduced into the sequence by changing a wobble base is boxed. The retention of the *Mlu* I site and the introduction of the *Bam*HI site help to "cassette" the full length IDUA cDNA insert for future expression experiments such as mutagenesis studies.

conclusions are in agreement with data from a published abstract reporting the cloning of a human IDUA cDNA (Moskowitz *et al.*, 1992).

A cDNA for canine IDUA has been isolated and codes for a protein of 655-amino acids with a signal peptide of 25 amino acids and six potential glycosylation sites (Stoltzfus *et al.*, 1992). Purified IDUA from canine testis contains two polypeptides of 68 and 63-kDa and the canine 68-kDa and human 74-kDa amino termini both begin with Glu. The amino terminus of the canine 63-kDa species is Leu 106 corresponding to the human liver 65/60/18-kDa amino terminus of Leu 107. Canine fibroblast IDUA mRNA was reported to be 2.2-kb (Stotzfus *et al.*, 1992). Stoltzfus *et al.*, (1992) reported that the original cDNA clones isolated for canine IDUA were missing a section of 114-bp encoding 38-amino acids. This missing section corresponds to the first half of exon XIII of the human IDUA gene and could be produced by alternative splicing to a cryptic acceptor splice site within exon VIII. PCR analysis of dog IDUA mRNA has shown that <20% of mRNA has this deletion (Stoltzfus *et al.*, 1992).

Of six potential glycosylation sites present (Fig. 3.22), the Asn residue of the potential sites present in the 65/60/18-kDa amino-terminal sequence and peptide 8 did not produce an amino acid signal during automated sequencing and so it is likely that these residues are glycosylated. The potential glycosylation site at the very end of peptide 9 also did not sequence but this may be due to a weak amino acid signal towards the end of the sequencing run rather than a glycosylated residue. Four of the six potential glycosylation sites are conserved between human and canine IDUA. Another potential glycosylation site occurs in a poorly conserved region and has a different consensus sequence between the species (Fig. 3.29). Another human IDUA cDNA clone was reported to encode only 5 potential glycosylation sites (Moskowitz *et al.*, 1992).Thus, there are a maximum of four and at least two glycosylation sites used in IDUA.

Human IDUA protein and cDNA both have 82% homology with the canine IDUA protein and cDNA (Figs. 3.29 and 3.31). With such high levels of homology, no distinct regions of conservation can be determined. However, the 5' and 3' untranslated

sequences of the IDUA cDNAs are very poorly conserved. The polyadenylation signal for the canine IDUA mRNA was reported to be cttaaa, a non-consensus polyadenylation signal similar to the tttaaa polyadenylation signal proposed for human IDUA, although these polyadenylation signals do not align relative to the end of the coding sequence. No significant homology was found between the human IDUA amino acid sequence and proteins in the GenBank, National Biomedical Research Foundation or Swiss-Prot databases (all releases till May, 1991). Stoltzfus *et al.* (1992) reported that a Pearson FASTA search (Pearson and Lipman, 1988) of GenBank (November 2nd release, 1991) showed that the amino acid sequence of IDUA had a 26% identity and a 29% similarity to the 488 amino acid β-xylosidase enzyme encoded by the *xynB* gene of the thermophilic microorganism, *Caldocellum saccharolyticium* (Luthi *et al.*, 1990). The substrates for IDUA and β-xylosidase are both glycosides and are identical, except that idopyranoside uronic acid has a carboxyl group linked to C5 while β-D-xylopyranoside has a hydrogen. An alignment of human IDUA and β-xylosidase is shown in Fig. 3.32. Although the homology is probably significant, it is difficult to draw any conclusions about the position of likely functional domains or residues. It is possible that the two enzymes are derived from an ancestral gene and have undergone divergent evolution on the basis of function.

Alternatively spliced IDUA mRNA species may be due to splicing errors or they may produce IDUA "deletion" products with functions that differ from the normal full-length IDUA. Particularly with the advent of PCR technology, there have recently been many reports of alternative splicing in the literature. In the case of proteins with distinct functional domains, inframe alternative splicing seems to produce functional variants of the same protein where one form may be membrane bound and the other water soluble (e.g. secretin, Kopin *et al.*, 1991). Other reports of alternative splicing resulting in alteration of the reading frame of the mRNA, indicate that the alternative splicing is not of functional significance (e.g. sphingomyelinase Schuchmann *et al.*, 1991). An explanation for the presence of alternatively spliced of IDUA mRNA is that these alternative RNA transcripts are present at low background levels which are detected only due to the sensitivity of PCR. Northern blot analysis of poly(A)$^+$ RNA with the

FIGURE 3.31. COMPARISON OF THE HUMAN AND CANINE IDUA cDNA SEQUENCES.

A GAP sequence alignment of the human (top) and canine (bottom) IDUA cDNAs, demonstrating 82% homology. Only bases that differ from the human sequence are shown for the canine sequence. The ATG initiating methionine codon, stop codons and the canine polyadenylation signals are underlined. As discussed in section 3.10, the tight homology is less obvious in the 5' and 3' untranslated regions.

```
GTCACATGGGGTGCGCGCCCAGACTCCGACCCGGAGGCGGAACCGGCAGTGCAGCCCGAAGCCCCGCAGTCCCCGAGCACGCGTGGCCATGCGTCCCCTG  100
.................. CCG    CG     CTCTC CC     .... C    C GC      GCC G GACC C   CC         G   CC   78

CGCCCCCGCGCCGCGCTGCTGGCGCTCCTGGCCTCGCTCCTGGCCGCGCCCCCGGTGGCCCCGGCCGAGGCCCCGCACCTGGTGCAGGTGGACGCGGCCC  200
G           C CGG        G     G   G     G       ...G    TC  A                    TC                    175

GCGCGCTGTGGCCCCTGCGGCGCTTCTGGAGGAGCACAGGCTTCTGCCCCCCGCTGCCACACAGCCAGGCTGACCAGTACGTCCTCAGCTGGGACCAGCA  300
       C           C           C          C    G                  GC  T A                              275

GCTCAACCTCGCCTATGTGGGCGCCGTCCCTCACCGCGGCATCAAGCAGGTCCGGACCCACTGGCTGCTGGAGCTTGTCACCACCAGGGGGTCCACTGGA  400
        G          T  T        G G     G                                CA    GG      A  AG   G  375

CGGGGCCTGAGCTACAACTTCACCCACCTGGACGGGTACTTGGACCTTCTCAGGGAGAACCAGCTCCTCCCAGGGTTTGAGCTGATGGGCAGCGCCTCGG  500
 AA                          T  C   C   T  C                                T                  C   CC  475

GCCACTTCACTGACTTTGAGGACAAGCAGCAGGTGTTTGAGTGGAAGGACTTGGTCTCCAGCCTGGCCAGGAGATACATCGGTAGGTACGGACTGGCGCA  600
AG G            C         G        G C         GC   G  CT                      G    T    CT AT  575

TGTTTCCAAGTGGAACTTCGAGACGTGGAATGAGCCAGACCACCACGACTTTGACAACGTCTCCATGACCATGCAAGGCTTCCTGAACTACTACGATGCC  700
C                                                 C       GA       C                          T  C    675

TGCTCGGAGGGTCTGCGCGCCGCCAGCCCCGCCCTGCGGCTGGGAGGCCCCGGCGACTCCTTCCACACCCCACCGCGATCCCCGCTGAGCTGGGGCCTCC  800
    C          T  T       G       CT T  C       G    T     C  TGG    C    C  T                        775

TGCGCCACTGCCACGACGGTACCAACTTCTTCACTGGGGAGGCGGGCGTGCGGCTGGACTACATCTCCCTCCACAGGAAGGGTGCGCGCAGCTCCATCTC  900
  GAG  T  T   A   C               C      CT   G    C                              A     C   G G         A  875

CATCCTGGAGCAGGAGAAGGTCGTCGCGCAGCAGATCCGGCAGCTCTTCCCCAAGTTCGCGGACACCCCCATTTACAACGACGAGGCGGACCCGCTGGTG 1000
        A      C   C AC  T           A G                    C       G                                    975

GGCTGGTCCCTGCCACAGCCGTGGAGGGCGGACGTGACCTACGCGGCCATGGTGGTGAAGGTCATCGCGCAGCATCAGAACCTGCTACTGGCCAACACCA 1100
      G       G     C     A  C        G                          G G       C       C  CC G       GG   1075

CCTCCGCCTTCCCCTACGCGCTCCTGAGCAACGACAATGCCTTCCTGAGCTACCACCCGCACCCCTTCGCGCAGCGCACGCTCACCGCGCGCTTCCAGGT 1200
G G G   C G G CC                    C              T           G   A                                    1175

CAACAACACCCGCCCGCCGCACGTGCAGCTGTTGCGCAAGCCGGTGCTCACGGCCATGGGGCTGCTGGCGCTGCTGGATGAGGAGCAGCTCTGGGCCGAA 1300
    G    GGAG                  C                                C         C        C GCCG      G       G 1275

GTGTCGCAGGCCGGGACCGTCCTGGACAGCAACCACACGGTGGGCGTCCTGGCCAGCGCCCACCGCCCCCAGGGCCCGGCCGACGCCTGGCGCGCCGCGG 1400
       G  G      G  G                                                G    TG  GGCC  G  CCGG                A C  1375

TGCTGATCTACGCGAGCGACGACACCCGCGCCCACCCCAACCGCAGCGTCGCGGTGACCCTGCGGCTGCGCGGGGTGCCCCCCGGCCCGGGCCTGGTCTA 1500
     C                   G         G  GC    GC   GC C    G     C    T         G GG     A  G  C        1475

CGTCACGCGCTACCTGGACAACGGGCTCTGCAGCCCCGACGGCGAGTGGCGGCGCCTGGGCCGGCCCGTCTTCCCCACGGCAGAGCAGTTCCGGCGCATG 1600
      C TGGC          CC  G          C           A                                          GG          1575

CGCGCGGCTGAGGACCCGGTGGCCGCGGCGCCCCGCCCCTTACCCGCCGGCGGCCGCCTGACCCTGCGCCCCGCGCTGCGGCTGCCGTCGCTTTTGCTGG 1700
     A  C                 A      G         C      A              G  CA  GTG A              C     GC     C 1675

TGCACGTGTGTGCGCGCCCCGAGAAGCCGCCCGGGCAGGTCACGCGGCTCCGCGCCCTGCCCCTGACCCAAGGGCAGCTGGTTCTGGTCTGGTCGGATGA 1800
          C        G           G  A C   G  C        T          T      GT      G  C  T    G              1775

ACACGTGGGCTCCAAGTGCCTGTGGACATACGAGATCCAGTTCTCTCAGGACGGTAAGGCGTACACCCCGGTCAGCAGGAAGCCATCGACCTTCAACCTC 1900
G G                          C  T              CGC   T  AG A T      G  CA                  C     T     G 1875

TTTGTGTTCAGCCCAGACACAGGTGCTGTCTCTGGCTCCTACCGAGTTCGAGCCCTGGACTACTGGGCCCGACCAGGCCCCTTCTCGGACCCTGTGCCGT 2000
                GT   CC TCAC                G        GG                        G              AC  GC    AC 1975

ACCTGGAGGTCCCTGTGCCAAGAGGGCCCCCATCCCC.....................GGGCAATCCATGAGCCTGTGCTGAGCCCCAGTGGGTTGCACCTCCAC 2084
  G            CA  GTC      G  GCGG  CAGTGACTGTGAGCGGT CTG C TCG      GCG AG CCAGG GC  A A  GG      A 2075

CGGCAGTCAGCG..AGCTGGGGCTGCACTGTGCCCATGCTGCCCTCCCATCACCCCCTTTGCAATATATTTTT.......................... 2155
G AAG C CT  TCC  C    A  G  GCA    G . C TGGG T CG T  TT  G   TG C CAC CCCTTAAAGTGCCTTTGCGCTCACCAAAA 2174
```

full-length IDUA cDNA insert (IDX21) only detected a single 2.3-kb mRNA transcript implying that alternatively spliced mRNAs, if present, are at extremely low levels. Conversely, it is possible that the resolution of this Northern analysis was not sufficient to detect differences of only 100-bp out of 2.3-kb, especially with the added "blurring" effect of varying poly (A) length. However, it is curious that these alternative splicing events do maintain the same reading frame of IDUA, unlike the alternate splicing events described for other human lysosomal hydrolases (Oshima *et al.*, 1987; Morreau *et al.*, 1989; Quintern *et al.*, 1989). It is possible that these IDUA "deletion" products have different substrate preferences compared with the full-length IDUA. Forms of IDUA with different substrate preferences have been reported (Matalon *et al.*, 1983; Minami *et al.*, 1984; Fujibayashi *et al.*, 1984). It will be interesting to determine whether the translated peptides resulting from the alternatively spliced IDUA mRNA are functional, and if so, whether they play any part in the pathology and varying clinical phenotypes of MPS-I.

A summary of what is now known about the IDUA gene is shown in Fig. 3.33. *IDUA* has been mapped to 4p16.3, and regionally mapped within the cosmid A157.1 which is used to define the locus *D4S111*. *D4S111*, and thus *IDUA*, is 1100-kb from the telomere of 4p and 1000-kb from either of the two most likely potential regions for the Huntington disease gene (MacDonald *et al.*, 1991b). The gene for IDUA is split into 14 exons spanning approximately 19-kb. The sequence of the IDUA gene is reported in two segments of 1.8-kb and 4.5-kb and has a 68% GC content. The potential promoter for *IDUA* has only GC box-type consensus sequences consistent with a housekeeping promoter and is bounded by an Alu repeat sequence. The first two exons of *IDUA* are separated by an intron of 566-bp, then there is a large intron of approximately 13-kb and the last 12 exons are clustered in the 4.5-kb segment. No consensus polyadenylation signal was found in the 3' untranslated region although two variant polyadenylation signals are proposed. In addition, the sequence of a 2.2-kb *Pst* I fragment, that overlaps by 426-bp with the 5' end of the 4.5-kb segment and is immediately adjacent to exon III of IDUA, is reported. This fragment contains the probe p157.9, which is used for diagnosis in Huntington disease, and the molecular basis of the *D4S111* polymorphic

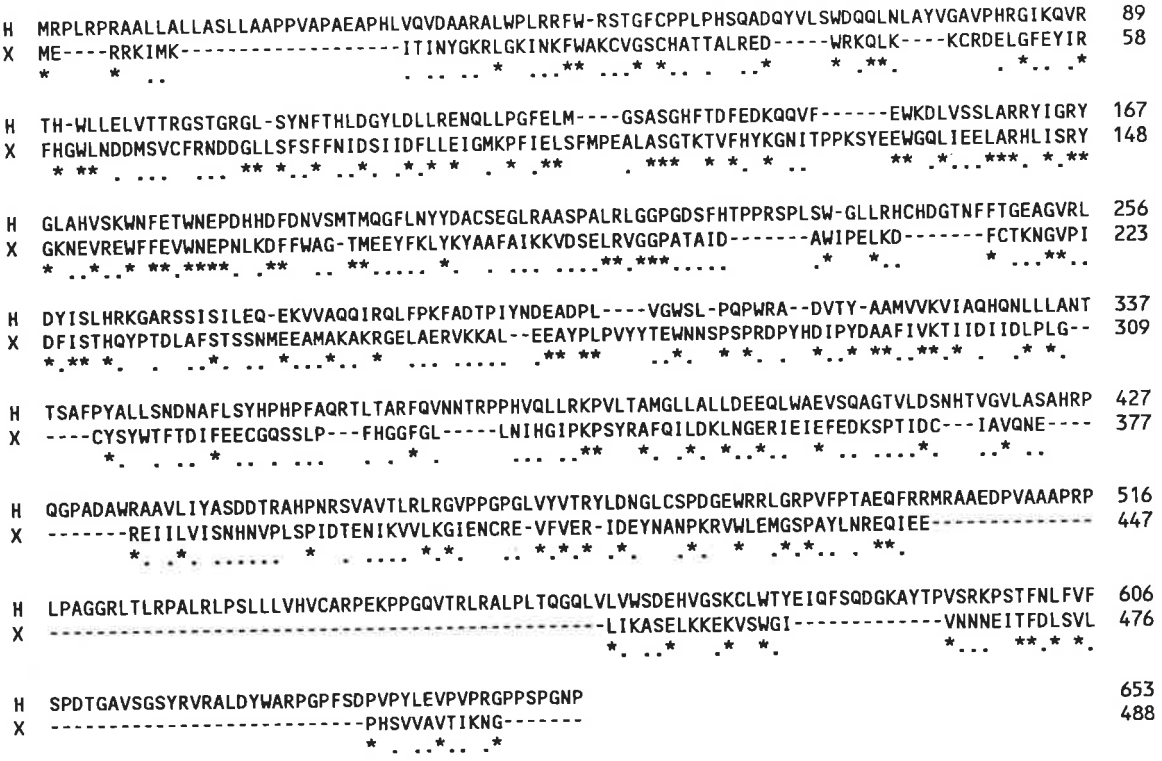


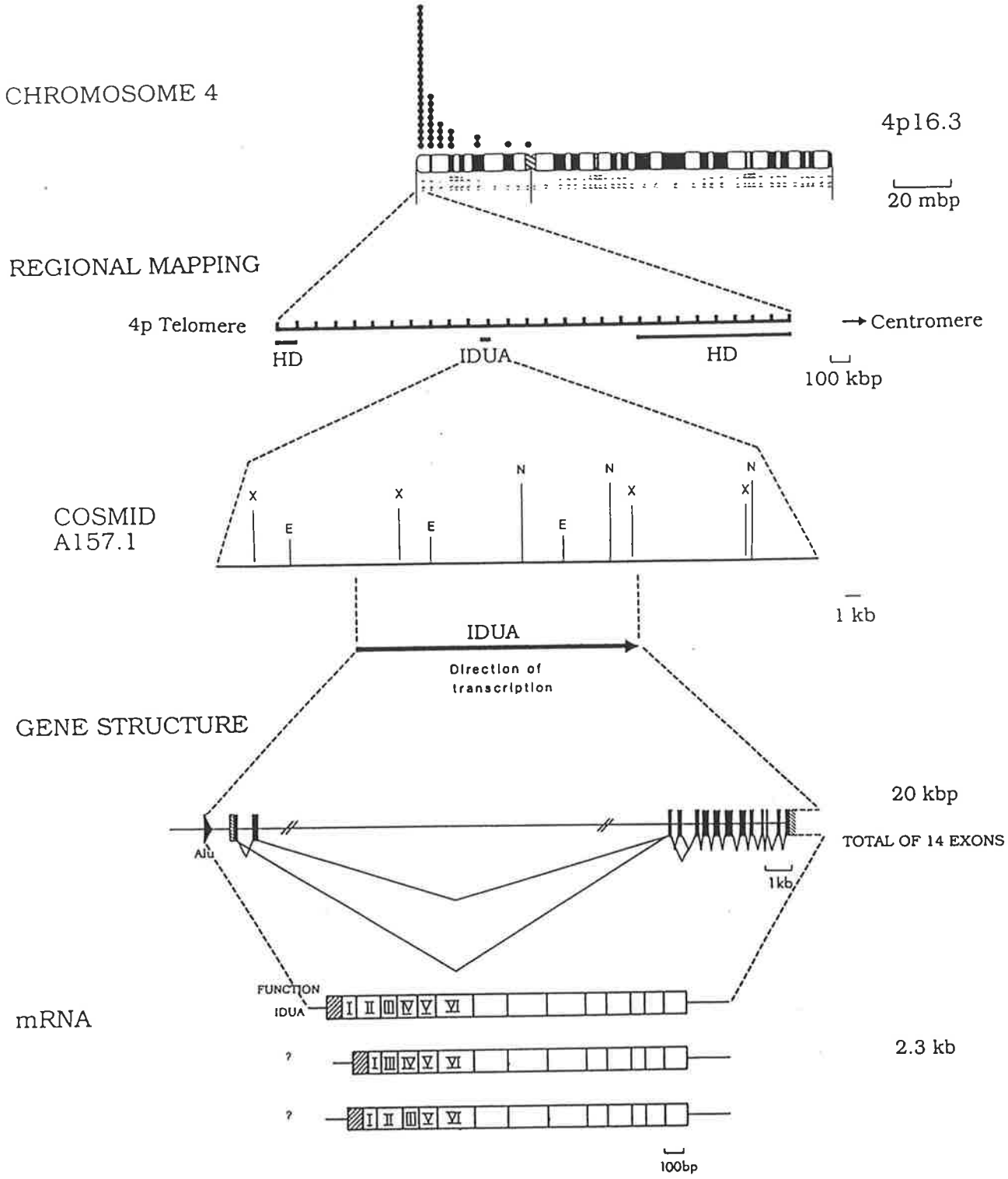
FIGURE 3.32. COMPARISON OF HUMAN IDUA WITH β-XYLOSIDASE.

A Clustal sequence alignment of the human IDUA peptide (H) and the *C. Saccharolyticium* β-Xylosidase protein (X) encoded by the *xynB* gene (Luthi *et al.*, 1990) is presented, showing 20% identity and an additional 32.8% similarity (conservative substitutions) over the length of the proteins. Identical amino acids are marked by an asterisk and conservative substitutions by a dot. Amino acid positions are shown on the right. The similarity was originally picked up by a FASTA scan of the GenBank database (Stoltzfus *et al.*, 1992) and showed 26% identity and an additional 29% similarity from amino acid 30 to 249 of IDUA.

locus that p157.9 detects was shown to be variation in the number of 86-bp repeats contained in this fragment. The *IDUA* genomic sequence will be valuable for defining mutations in MPS-I patients, for defining diagnostically useful polymorphisms for MPS-I and Huntington disease and for refining the genetic and physical map of 4p16.3. The *IDUA* gene produces at least three distinct transcripts, a major 2.3-kb full-length mRNA that is translated into enzymatically active IDUA protein, and two minor transcripts that are products of the alternative splicing of exons II or IV of *IDUA*. These minor transcripts have yet to be characterized in terms of their stability, translational capacity or the activity of the encoded protein products.

FIGURE 3.33. SUMMARY OF THE GENOMIC STRUCTURE OF *IDUA*.

A diagram showing the organization of the human IDUA gene. *IDUA* maps to chromosome 4p16.3, approximately 1100-kb from the telomere, and 1000-kb from either of the two most likely regions for the Huntington disease gene (HD). *IDUA* is split into 14 exons over a 20-kb region within the cosmid A157.9 and transcription is towards the centromere as shown. *IDUA* produces at least 3 mRNA transcripts by alternative splicing. The full-length mRNA produces active IDUA protein and the function of protein produced by the other transcripts is unknown.


CHROMOSOME 4
4p16.3
20 mbp
REGIONAL MAPPING
4p Telomere
Centromere
HD
IDUA
HD
100 kbp
COSMID
A157.1
X
E
X
E
N
E
N
X
N
X
1 kb
IDUA
Direction of
transcription
GENE STRUCTURE
20 kbp
TOTAL OF 14 EXONS
Alu
1kb
FUNCTION
IDUA
I II III IV V VI
?
I III IV V VI
?
I II III V VI
100bp
mRNA
2.3 kb

CHAPTER 4

MOLECULAR CHARACTERIZATION OF MPS-I PATIENTS

4.1 INTRODUCTION

As described in Chapter 1, mucopolysaccharidosis type I (MPS-I, eponyms: Hurler, Hurler/Scheie, and Scheie syndromes) is inherited as an autosomal recessive genetic disease in humans and is caused by a deficiency of the lysosomal glycosidase α-L-iduronidase (IDUA). IDUA is one of a group of *exo*enzymes required to degrade the glycosaminoglycans heparan sulphate and dermatan sulphate. A deficiency of any one of these enzymes leads to lysosomal storage of these partially degraded glycosaminoglycans and a mucopolysaccharidosis-type storage disorder. Patients with MPS-I, like all other mucopolysaccharidoses, may present within a broad spectrum of clinical phenotype that varies from severe mental retardation, skeletal deformations, hepatosplenomegaly, corneal clouding, and a life span of less than 5 years, to mildly affected patients who may have normal intelligence and life span, stiff joints, corneal clouding and mild visceral organ involvement (Neufeld and Muenzer, 1989; Hopwood and Morris, 1990). Before the enzyme defect in MPS-I was identified, two separate syndromes of varying clinical severity were classified, i.e. Hurler syndrome (severe) and Scheie syndrome (mild; previously classified as MPS-V) (McKusick and Neufeld, 1983). Following the discovery that a deficiency of IDUA was responsible for both syndromes (Bach *et al.*, 1972), it was believed that there were two different mutant alleles at the IDUA locus, i.e. Hurler patients were homozygous for the severe allele and Scheie patients were homozygous for the mild allele. MPS-I patients with an intermediate clinical severity were classified Hurler/Scheie to suggest a compound heterozygote of the severe and mild alleles (McKusick *et al.*, 1972; McKusick and Neufeld, 1983). More recently it has been proposed that the broad spectrum of clinical phenotype seen in MPS-I patients is generated by the presence of multiple IDUA mutant alleles (Neufeld and Muenzer, 1989; Hopwood and Morris, 1990).

A deficiency of IDUA activity in patient leucocytes or cultured skin fibroblasts is diagnostic of the MPS-I condition. However, extensive enzymological characterization of the residual IDUA activity in MPS-I fibroblasts has only in part enabled the correlation of biochemical phenotype with the clinical severity of patients

(Hopwood and Muller, 1979; Myerowitz and Neufeld, 1981; Ullrich *et al.*, 1981; Matalon *et al.*, 1983; Muller and Hopwood, 1984; Schuchman and Desnick, 1988; Taylor *et al.*, 1991; Ashton *et al.*, 1992).

Following the isolation of the gene for *IDUA*, it was thought to be important to study the molecular nature of MPS-I and thus mutations within the IDUA gene in the hope that such a study may enable a greater understanding of the disease. In particular, it was hoped that improvement in the following areas would result from these studies.

1) Mutant genotype-phenotype correlations leading to a molecular explanation of the differences between Hurler and Scheie syndromes.

2) Improved long-term patient phenotype prediction for:

   a) Genetic counselling, including assessment of whether patients would benefit from specific experimental treatment protocols.

   b) Assessment of the efficacy of experimental treatment protocols for MPS-I that are currently in use, such as bone marrow transplantation (BMT), or assessment of new protocols that may become available, such as enzyme replacement or gene replacement therapy.

3) An increased understanding of lysosomal biogenesis, particularly with respect to transport to the lysosome, maturation, and structure and function of IDUA.

Identification of mutations in IDUA causing MPS-I was to initially be attempted at the gross DNA level by Southern blotting analysis. It was hoped that several patients with deletions or gene rearrangements would be identified as was observed with the iduronate-2-sulphatase gene in MPS-II (Hunter syndrome) patients (Wilson *et al.*, 1990; 1991). The identification of two polymorphisms within *IDUA* could possibly enable haplotype studies to be performed in order to detect linkage disequilibrium between these polymorphisms and MPS-I. Similar studies have been successful in identifying common mutations for other genetic diseases (e.g. cystic fibrosis, Estivill *et al.*, 1987). Detailed analysis would then be performed in order to detect small deletions or rearrangements and point mutations by more sensitive techniques. Various techniques

were considered for the study of MPS-I patient mutations, including cleavage of RNA/DNA heteroduplexes with RNase A, single-stranded DNA conformational polymorphisms, denaturing gradient gel electrophoresis and chemical cleavage of mismatched DNA heteroduplexes. A summary of the major advantages and disadvantages of these techniques is given in Table 4.1. The chemical cleavage technique (CC) was judged to be the best choice for the detection of point mutations, especially due to the recent refinement of labelling both the mutant and normal strands of the heteroduplex, allowing the detection of all mutations (Forrest *et al.*, 1991). Also, as CC had previously been introduced in our laboratory, MPS-I patient analysis was attempted using this technique.

A diagrammatic representation of CC is shown in Fig. 4.1. CC is based on the principle that mismatched or unmatched residues are more reactive to modification by the chemicals hydroxylamine and osmium tetroxide (OT). Briefly, both mutant and normal PCR products are labelled on one strand at one end by the use of a 5' end-labelled PCR primer. These labelled PCR products are combined, denatured and annealed to form a heteroduplex. The presence of mutations or polymorphisms in the DNA will result in mismatching of the strands, rendering the bases involved susceptible to chemical modification. Mismatched T or C residues are specifically reactive to OT and hydroxylamine respectively. Cleavage of these chemically-modified residues is achieved with piperidine. Cleaved products are resolved on denaturing polyacrylamide gels (Section 2.2.8b) and detected by autoradiography. Cleavage products can be sized to determine the approximate position of the mismatched bases in the PCR product and the exact nature of the mismatch can be determined by direct PCR sequencing. To detect mismatched A or G residues, probes labelled on the opposite strand with the opposite PCR primer can be employed as the complementary mismatched bases will be T and C respectively. It has been shown that some T residues in T-G mismatches are resistant to modification by OT and thus subsequent cleavage by piperidine (Saleeba and Cotton, 1992). The sequence surrounding the T-G mismatch appears to be important in determining whether or not a particular mismatch will be modified.

TABLE 4.1. MUTATION DETECTION TECHNIQUES.

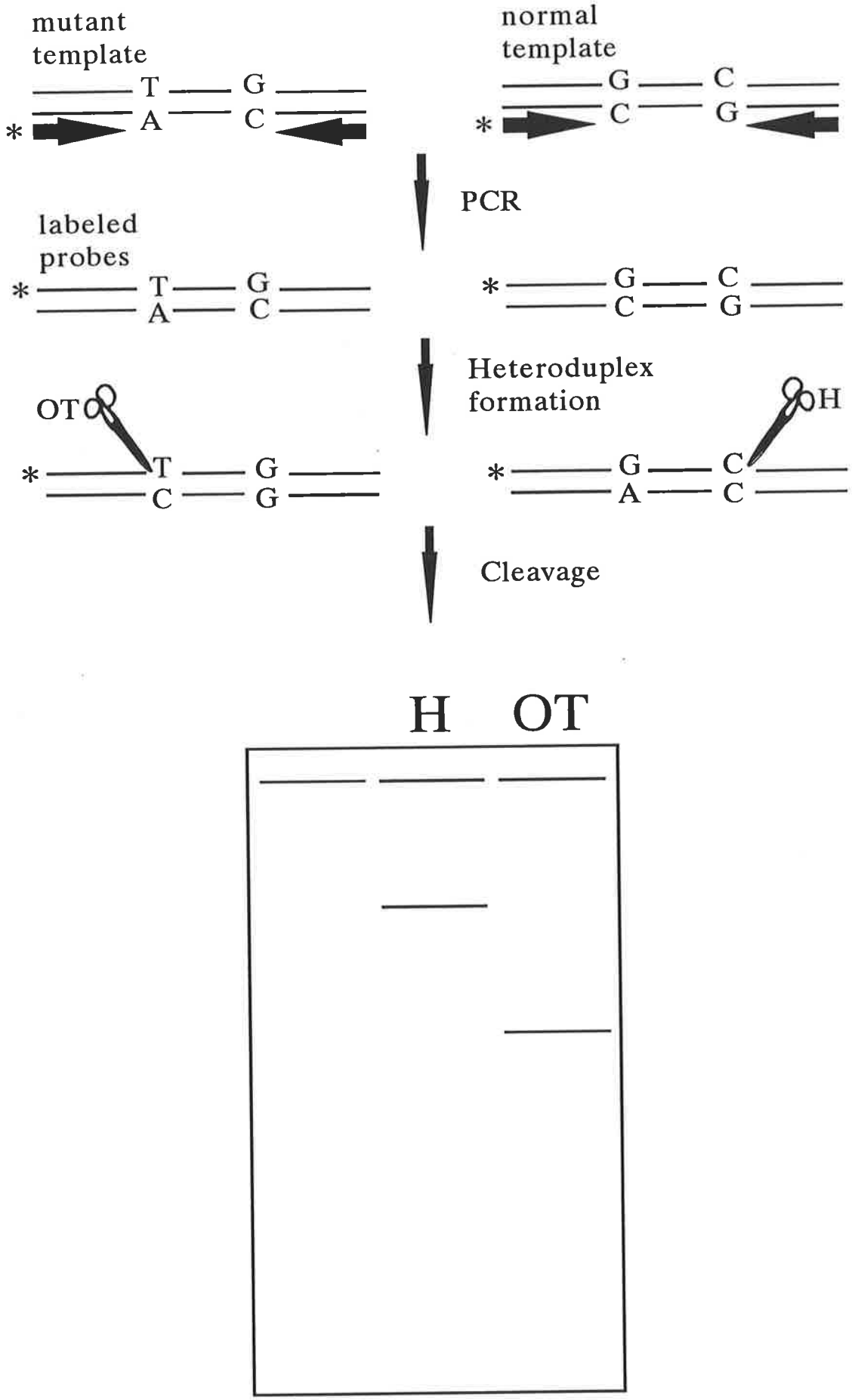
A list of the methods considered for detection of MPS-I mutations is given with a brief summary of their main advantages and disadvantages (adapted from Cotton, 1989). If heteroduplex formation is unnecessary, the heteroduplex type is listed as none.

TABLE 4.1.

Method	Mutation position defined	% Mutations detected	Heteroduplex type	Maximum bases screened/test	Advantage(s)	disadvantage(s)
Sequencing	yes	100	none	200 to 400	Position and change defined in one step.	Time consuming and technically demanding. Only limited coverage.
Denaturing gradient gel electrophoresis	no	50 to 100	none or any combination	1000	Can isolate mutant band. Good coverage.	Requires exspensive special equipment. Does not define position.
RNase A	yes	60 to 70	RNA·DNA or RNA·RNA	1000	Cleavage method; changes positioned.	RNA probe production is difficult and only limited detection.
Single stranded confirmational polymorphism	no	60 to 80	none	300 to 400	Probably the easiest method technically.	Limited coverage and detection.
Chemical cleavage	yes	100?	DNA·DNA or RNA·DNA	1000 to 2000	Cleavage method; may detect and position all mutations.	Many steps handling toxic reagents.

FIGURE 4.1. DIAGRAMMATIC REPRESENTATION OF CHEMICAL CLEAVAGE.

Normal and mutant template DNA strands are shown. A labelled oligonucleotide (arrow with *) is used in PCR to produce both normal and mutant probes which are combined to form heteroduplexes. Hydroxylamine (H) cleaves C mismatches and osmium tetroxide (OT) cleaves T mismatches. Any base changes not detected in this manner can be detected when the opposite PCR primer is labelled


mutant
template
T
G
A
C
*
normal
template
G
C
C
G
*
PCR
labeled
probes
*
T
G
A
C
*
G
C
C
G
Heteroduplex
formation
OT
T
G
C
G
*
H
G
C
A
C
*
Cleavage
H
OT

However, a T-G mismatch will always produce a complementary A-C mismatch which will be detected by hydroxylamine cleavage of the opposite strand.

After the initial haplotype analysis described in Section 4.3.2, a group of MPS-I patients were chosen for detailed analysis on the basis of one or more of the four following criteria:

1) Patients were either homozygous or heterozygous for the 2,2 (VNTR, *Kpn* I) haplotype, implying the possible presence of the hypothesized common mutation for MPS-I.

2) Patients were, or were likely to be, heterozygous for the 2,1 haplotype, implying the possible presence of the hypothesized second common MPS-I mutation. Patients who were homozygous for the 2,1 haplotype were not available at this time.

3) Patients with an interesting biochemical phenotype, implying the presence of a mutation that may be important to improving the understanding of the structure and function of IDUA.

4) Patients exhibiting, as a group, the complete range of clinical phenotypes for MPS-I, so that mutation analysis might provide an understanding of the molecular and biochemical basis of the differences between Hurler, Hurler/Scheie and Scheie syndromes.

The cell line number, phenotype, haplotype and criteria used to select patients for detailed mutation analysis are shown in Table 4.2. Four severely affected patients homozygous for the 2,2 haplotype were selected (538, 728, 985 and 1062). Two patients that were potentially heterozygous for the 2,1 haplotype were selected (296, 2827). Two patients were selected on the basis of interesting biochemical findings and their extremely mild and intermediate phenotypes (2474 and 2662 respectively). Patient 2474 (GM1323, NIGMS, Human Genetic Mutant Cell Repository) is an index Scheie syndrome patient having one of the most mild MPS-I phenotypes described (McKusick *et al.*, 1965; McKusick, 1972). This patient has been extensively studied and

biochemical analysis had detected approximately 0.3% of IDUA activity in fibroblasts from patient 2474 compared to normal controls with a number of different substrates (Hopwood and Muller, 1979; Muller and Hopwood, 1984; Ashton *et al.*, 1992) but IDUA protein was not detected by 2 monoclonal antibodies (Ashton *et al.*, 1992). Patient 283 had a severe to intermediate phenotype with some IDUA enzyme activity and protein detectable (Ashton *et al.*, 1992). Patient 2827 was the only cell line from a group of 23 MPS-I patients studied that contained greater than 7% of the mean level of IDUA protein detected in normal controls. In fact, fibroblasts from patient 2827 had a 6-fold greater amount of IDUA protein than normal controls and low levels of activity towards 4-methylumbelliferyl-α-L-iduronide, and a heparin-derived disaccharide substrate. Maturation studies have shown apparently normal levels of IDUA synthesis with delayed processing to the mature form present in the lysosomes. Brooks *et al.* (1992) hypothesized that the mutation in 2827 affects both the active site and post-translational processing of the enzyme, and as this mutation does not result in reduced stability of the enzyme during maturation or in the lysosome, the mutation is structurally conservative.

The results presented in this chapter describe molecular characterization of MPS-I patients. Southern blot analysis of 40 MPS-I patients failed to detect major deletions or rearrangements of *IDUA*. Linkage disequilibrium was detected between two polymorphic haplotypes and MPS-I suggesting the presence of two common mutations for MPS-I. Nine patients, either displaying a haplotype in disequilibrium with MPS-I, an interesting biochemical phenotype, or an interesting clinical phenotype, were selected for detailed mutation analysis. It was hoped to identify all the mutant (MPS-I) alleles in this group of 9 patients with the initial aim focussing on determination of the hypothesized common mutations. A total of 18 polymorphisms and 8 mutations were detected in the 9 patients, which accounted for all mutant alleles. As hypothesized from the linkage disequilibrium studies, two common mutations, $W_{402}X$ and $Q_{70}X$, associated with the severe Hurler phenotype, were detected and accounted for 33% and 15% of MPS-I alleles respectively. Three amino acid substitutions, a single nucleotide deletion and two splicing mutations were also

TABLE 4.2. SELECTION OF MPS-I PATIENTS FOR MUTATION ANALYSIS.

The MPS-I patients selected for detailed mutation analysis, their VNTR and *Kpn* I polymorphism genotypes, and their known fibroblast biochemical data are listed. The biochemical data in columns marked with a * is from Ashton *et al.*, 1992). UD = undeterminable ND = not detected. A hyphen indicates that the analysis was not done. IDUA protein was detected with the monoclonal antibody Id1A. IDUA activity is expressed as activity detected using the IDUA fluorometric assay (Clements *et al.*, 1985) pmol/min/mg extracted cell protein. Specific activity is expressed as the IDUA activity (fluorometric assay)/pmol/min/ng IDUA protein detected with the monoclonal antibody. K_m was calculated using a heparin derived disaccharide substrate (Hopwood and Muller, 1982). The criteria used to select the patients, as listed in the text of Section 4.1, are given. A more detailed description of selection of patients is given in the text.

TABLE 4.2.

Patient number	Phenotype	VNTR	*Kpn* I	IDUA* protein	IDUA* activity	Specific* activity	K_m* (μM)	Selection criteria
283	severe/ intermediate	2,2	2,2	0.33	7.4	22.4	262	1,3,4
296	severe	1,2	1,2	-	-	-	-	2
538	severe	2,2	2,2	ND	3.7	-	260	1
728	severe	2,2	2,2	ND	1.3	-	227	1
985	severe	2,2	2,2	ND	-	-	-	1
1062	severe	2,2	2,2	ND	-	-	346	1
2474	mild	1,2	2,2	ND	4.2	UD	33	1,3,4
2662	intermediate/ mild	2,3	2,2	2.03	40.3	19.9	18	1,3,4
2827	severe	2,2	1,2	189.30	3.2	0.02	UD	1,2,3,4
Normal	normal	any	any					
Range				17-54	654-2600	30-93	12-111	
Mean				30	1478	58	35	

detected. One amino acid substitution, $D_{349}N$, may be involved in both the active site and post-translational processing of IDUA. Another amino acid substitution, $R_{89}Q$, and a splicing mutation are both associated with a Scheie MPS-I phenotype. A molecular explanation for the range of MPS-I clinical phenotypes is given.

4.2 SPECIFIC MATERIALS AND METHODS

4.2.1 PATIENT FIBROBLASTS AND DNA PREPARATION

Normal and MPS-I patient fibroblasts were stored, grown and maintained by Kathy Nelson and Greta Richardson in the Department of Chemical Pathology at the Adelaide Children's Hospital.

All MPS-I patients had mucopolysaccharidurla and were deficient in IDUA activity in peripheral blood leukocytes and/or cultured skin fibroblasts (Hopwood and Morris, 1990). Normal and MPS-I patient skin fibroblasts were cultured and DNA was prepared as described (Nelson *et al.*, 1989). Fibroblasts were harvested in 10 ml of PBS per flask with a rubber policeman and centrifuged at 1,000 g for 2 min in a 50 ml Falcon tube to pellet the cells. The cell pellet was transferred to a preweighed Eppendorf centrifuge tube, rinsed with 1ml of PBS and spun for 30 s at 12,000 g to repellet the cells. The PBS was carefully removed and the tube reweighed to determine the approximate weight of the cell pellet. Cell pellets were used immediately for DNA or RNA preparations, or stored at -80°C until required.

4.2.2 OLIGONUCLEOTIDE PRIMERS

All coding regions of IDUA were PCR amplified as depicted in Fig 4.2. Briefly, all coding regions of *IDUA* were amplified in five separate PCR reactions. Primers used are listed in Appendix D. The five PCR reactions were as described in Section 2.2.21 except that 10% (v/v) DMSO was added. Forty cycles using the following conditions were performed for the specified reaction: denaturation was at 94°C for 30 s,

annealing was at 65°C for 30 s and extensions were at 72°C for 75 s for ID60 to ID66, ID79 to ID89 and ID90 to ID91, and for 45 s for ID96 to ID97 and ID92 to ID63.

4.2.3 CHEMICAL CLEAVAGE

Samples were subjected to chemical cleavage as described by Saleeba and Cotton (1992), with the modifications outlined below.

4.2.3a PROBE GENERATION

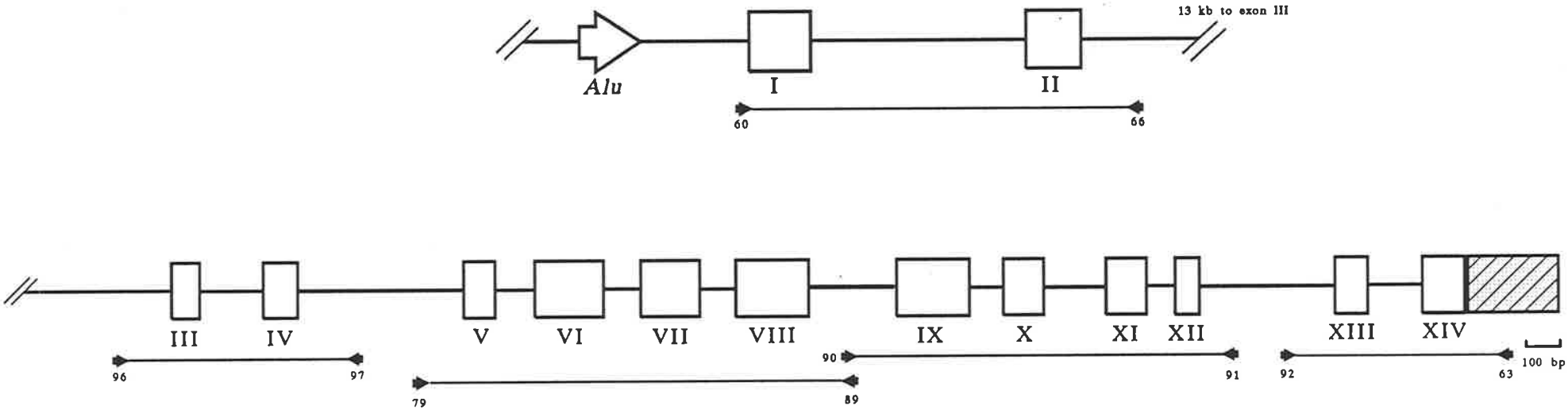
Using the oligonucleotide primers described above, all coding regions of IDUA were PCR amplified in 5 separate products using both MPS-I and normal DNA samples as template. PCR products were run on a 3% (w/v) Nusieve GTG/1% (w/v) normal agarose gel, the PCR product was excised, 3 x the volume of water added and heated at 65°C for 10 min. Radioactive probes were produced using 5'-end-labelled (Section 2.2.11a) primers in a PCR reaction with 1 µl of either normal or MPS-I PCR product from a gel slice as template. PCR conditions were as described above except that 30 ng (3.75 pM) of each primer was used and one of the two primers was 5' end-labelled. Thus, for one PCR region, 2 probes were produced from both normal DNA and MPS-I patient DNA. Probes were precipitated with isopropanol, analysed on a 1% (w/v) agarose gel, electrophoresed onto DEAE NA45 paper and eluted as described (Section 2.2.9d).

4.2.3b HETERODUPLEX FORMATION

Approximately 20 to 100 ng of both the patient and normal probes (generated as described in Section 4.2.3a) were combined in a total volume of 100 µl to result in a final concentration of 1 X annealing buffer (0.3 M NaCl, 3.5 mM $MgCl_2$, 3mM Tris-HCl, pH 7.5). Using a thermal cycler, the mixed probes were heated to 100°C for 5 min, cooled to 65°C, cooled from 65°C to 42°C over 60 min, and finally cooled to 4°C. Carrier DNA, in the form of 1 µl of 10 mg/ml herring sperm DNA, was added and the heteroduplexes were ethanol precipitated.

FIGURE 4.2. PCR AMPLIFICATION OF ALL CODING REGIONS OF *IDUA*.

Diagrammatic representation of *IDUA* and the five PCR reactions designed for use in chemical cleavage analysis. The PCR reactions are described in Section 4.2.2. The arrows represent the oligonucleotide primers used and the numbers refer to the names of the primers.


13 kb to exon III
Alu
I
II
60
66
III
IV
V
VI
VII
VIII
IX
X
XI
XII
XIII
XIV
96
97
79
89
90
91
92
63
100 bp

4.2.3c HYDROXYLAMINE MODIFICATION

Hydroxylamine (1.39 g) was dissolved in 1.6 ml of water and the pH was adjusted to approximately 6.0 by the addition of approximately 1 ml of diethylamine. A solution consisting of 6 µl of heteroduplex, 1 µl of 10 mg/ml herring sperm DNA and 20 µl of the hydroxylamine solution was incubated at 37°C for 10 min. The reaction was halted by the addition of 200 µl of stop solution (0.3 M sodium acetate, pH 5.2, 0.1 mM EDTA, pH 8.0, 25 µg/ml tRNA) and the reactions were ethanol precipitated, washed with 70% (v/v) ethanol and vacuum-dried. The samples were then subjected to piperidine cleavage.

4.2.3d OSMIUM TETROXIDE MODIFICATION

The heteroduplex (6 µl) was combined with 1 µl of 10 mg/ml herring sperm DNA, 2.5 µl of 10 X OT buffer (100 mM Tris-HCl, pH 7.7, 10 mM EDTA, pH 8.0), 0.38 µl of pyridine (to a final concentration of 1.5% (v/v)) and 15 µl of 4% (w/v) OT, followed by incubation at 37°C for 10 min. The reaction was halted by the addition of 200 µl of stop solution followed by ethanol precipitation, a 70% (v/v) ethanol wash and vacuum drying. The samples were then subjected to piperidine cleavage.

4.2.3e PIPERIDINE CLEAVAGE AND ANALYSIS

Piperidine (50 µl of 10% (v/v)) was added to the hydroxylamine or osmium tetroxide modified heteroduplex pellets and incubated at 90°C for 10 to 30 min. The samples were ethanol precipitated with sodium-acetate. Osmium tetroxide treated samples required incubation for 15 min in a dry ice/ethanol bath before centrifugation. Pellets were washed with 70% ethanol, vacuum-dried and resuspended in 5 µl of sequencing loading buffer. The samples were denatured at 100°C for 3 min and 1 to 2 µl was loaded onto a denaturing polyacrylamide gel. The cleaved samples of ID60 to ID66, ID79 to ID89 and ID90 to ID91 (Fig. 4.2) were analysed by both 5% and 3.5% (w/v) denaturing polyacrylamide gel electrophoresis (Section 2.2.8b) to obtain the required resolution up to 1200-bp. Cleaved samples of ID96 to ID97 and ID92 to ID63

were analysed on 5% gels. Observed cleavages were sized using the 5'-end-labelled markers pUC19 *Hpa* II and SppI *Eco*RI for identification of regions to be sequenced.

4.2.4 DIRECT PCR SEQUENCING

4.2.4a TEMPLATE PREPARATION

PCR products to be used as templates for direct PCR sequencing were purified using Centricon 100 (for products >200-bp) or Centricon 30 (for products <200-bp) microconcentrator columns (Amicon) according to the manufacturer's instructions. Briefly, Centricon columns contain built-in filters with varying pore sizes which enable filtration of molecules of specific molecular masses (e.g. Centricon 100 column retain protein molecules of a molecular mass greater than 100-kDa). Thus, PCR products were filtered through Centricon columns together with 2 ml of water and spun for 30 min (Centricon 100 at 1,000 g, 30 at 3,000 g) in a fixed angle rotor allowing unreacted nucleotides and primers to pass through the filter, and the PCR product to be retained. A further 2 ml of water was added and the column spun as before to remove any remaining primers and nucleotides. The column was inverted and spun for 5 min at the same force to pool the remaining fluid containing the PCR product in approximately 60 µl. PCR products were quantified by analytical agarose gel electrophoresis and the DNA was used in PCR sequencing reactions. For a 400-bp PCR product, 200 ng was used in a reaction; the amount of larger or smaller PCR products used as template was varied to maintain a similar molar amount.

4.2.4b SEQUENCING

Direct PCR sequencing was performed as a modification of the linear PCR method (Murray, 1989). When 10% (v/v) DMSO was necessary in production of the template PCR product, it was also added to the PCR sequencing reactions for that template. Each set of 4 reactions for a specific template primer pair contained 20 ng of a 5' end-labelled oligonucleotide primer that would prime towards the area of interest for sequencing. The final concentration of all nucleotides in the reaction mixes was 7.5 µM except that 7-deazaguanosine was used in place of dGTP at a final concentration of

15 µM. Final concentrations of the ddNTPs were ddA, 0.25 mM; ddC, 0.10 mM; ddG, 0.05 mM; and ddT, 0.50 mM. Between 15 and 25 PCR cycles were performed using the following conditions; denaturation at 94°C for 30 s, annealing between 55°C and 65°C for 30 s and extension at 72°C for 30 s. The temperature of annealing was varied depending on the primer used. All base changes were sequenced in both directions using the primers specified and confirmed by restriction endonuclease digestion or ASO analysis.

4.2.5 DETECTION OF POLYMORPHISMS AND MUTATIONS BY RESTRICTION ENDONUCLEASE DIGESTION OF PCR PRODUCTS

PCR primers used to amplify the *IDUA* exon of interest for detection of polymorphisms or mutations, and the reaction conditions used are shown in Fig. 4.3 and in Tables 4.3 and 4.4 respectively. PCR reactions containing DMSO were isopropanol precipitated and the PCR products resuspended in 10 µl of water with 1.5 µl being used for restriction analysis. In the absence of DMSO, 10 µl of the PCR reaction was digested without further purification. PCR products were digested with more than 1 U of the relevant restriction endonuclease in a final volume of 20 µl for 2 hr at the temperature, and in the buffer recommended by the manufacturer. Digests were analysed an a 3% (w/v) Nusieve GTG/ 1% (w/v) normal agarose gel. The sequences of the PCR primers are listed in Appendix D.

4.2.6 DETECTION OF POLYMORPHISMS AND MUTATIONS USING ALLELE SPECIFIC OLIGONUCLEOTIDES

PCR primers used to amplify the *IDUA* exon of interest for detection of polymorphisms or mutations, and the reaction conditions used are shown in Fig. 4.3 and in Tables 4.3 and 4.4 respectively. PCR reactions were run in duplicate on a 1.5% (w/v)

agarose gel, alkaline Southern blotted onto Hybond N^{+TM} nylon membrane (Section 2.2.12b) and the duplicate sections of the filter were separated. The allele specific oligonucleotides for detection of specific polymorphisms or mutations, prehybridization, hybridization and final washing temperatures are shown in Tables 4.3 and 4.4 respectively. Prehybridization was in 6 x SSC, 0.1% (w/v) SDS, 0.05% (w/v) sodium pyrophosphate and 100 µg/µl sheared, denatured herring sperm DNA at the specified temperature for 1 hr. The duplicate filters were individually hybridized for 4 hr with 200 ng of each of the required pair of ASOs. ASOs were 5' end-labelled as described in Section 2.2.11a. The filters were washed twice in 5 x SSC at room temperature for 10 min and then in 2 x SSC, 0.1% (w/v) SDS at the specified temperature for 5 min. Autoradiography was carried out as necessary, either with intensifying screens at -80°C or at room temperature depending on the final level of radioactivity present on the filters. The sequences of PCR primers and allele specific oligonucleotides are listed in Appendices D and E respectively.

RESULTS AND DISCUSSION

4.3 INITIAL ANALYSIS

These analyses were carried out in collaboration with Paul V. Nelson of the Department of Chemical Pathology at the Adelaide Children's Hospital.

4.3.1 SOUTHERN BLOTTING OF MPS-I PATIENT DNA

The isolation of clones that were shown to contain nucleotide sequence colinear with peptide data from purified IDUA enabled Southern blot analysis of MPS-I patient DNA to be performed. The purpose of this analysis was two-fold. Firstly, detection of mutational events such as deletions or major rearrangements in MPS-I patient DNA samples would help confirm that *IDUA* had in fact been isolated. Secondly, Southern blot analysis is a simple and rapid screen for gross mutations which is able to be rapidly performed on a large number of DNA samples. Southern blotting was initially performed on 40 *Eco*RI digested MPS-I patient DNA samples using the 1.6-kb *Pst* I

FIGURE 4.3. EXON SPECIFIC PCR AMPLIFICATION FOR DETECTION OF MUTATIONS AND POLYMORPHISMS.

The arrows represent the primers used to amplify specific *IDUA* exons for ASO detection or restriction endonuclease digestion detection of mutations and polymorphisms within IDUA at the positions indicated. See also Tables 4.3 and 4.4.

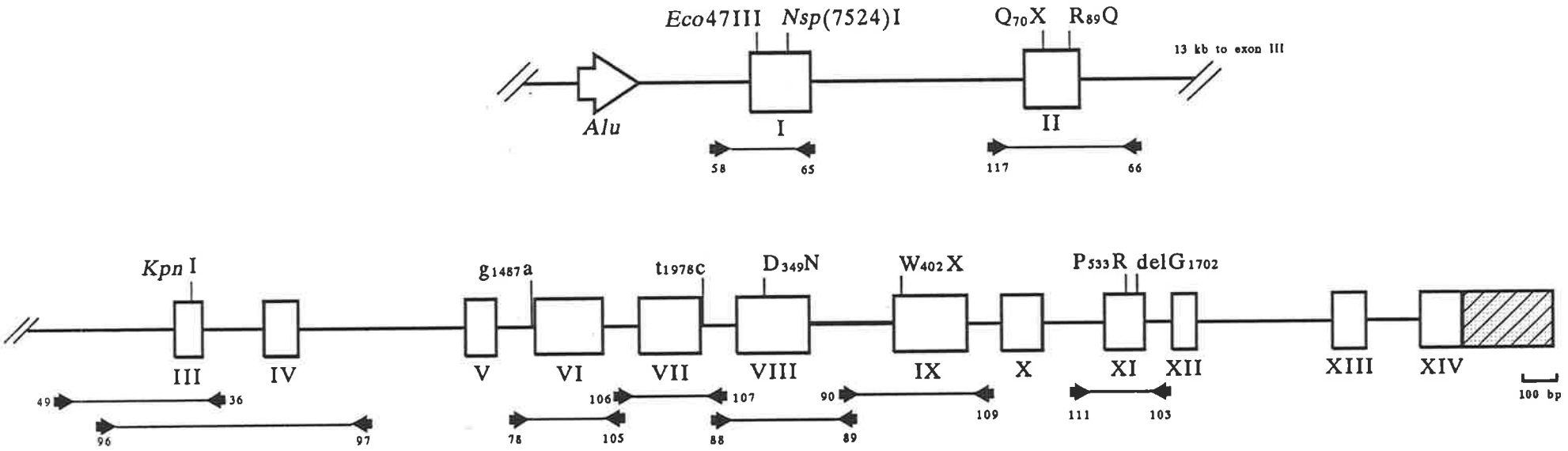

Eco47III
Nsp(7524)I
Q70X
R89Q
13 kb to exon III
Alu
I
II
58
65
117
66
Kpn I
g1487a
t1978c
D349N
W402X
P533R
delG1702
III
IV
V
VI
VII
VIII
IX
X
XI
XII
XIII
XIV
49
36
96
97
106
107
78
105
88
89
90
109
111
103
100 bp

TABLE 4.3. PCR DETECTION OF POLYMORPHISMS IN *IDUA*.

The PCR methods used to detect polymorphisms in *IDUA* are summarized. The nature of the nucleotide substitution and its position in the cDNA (Fig 3.22) and genomic sequence (Fig. 3.24) are given in the first column. The exon or intron of *IDUA* in which the polymorphism occurs (Roman numerals refer to exon numbers) and the amino acid codon positions are given in the second column. Note that two of the polymorphisms change amino acids. The *IDUA* specific PCR primers (Fig. 4.3 and Appendix D) and the annealing temperature used for the PCR reactions are given in the next two columns. DMSO indicates the presence of 10% (v/v) DMSO in the reaction mixes. PCR was performed as described in Section 2.2.21 and relevant sections of the text. The polymorphisms were detected either using the restriction endonucleases or ASOs listed (see Appendix E). The size of the normal PCR product (Nor.) and the size fragments produced following restriction endonuclease digestion are given. For polymorphisms detected by ASOs, the temperature of hybridization (Hyb.) and final washing is shown and was performed as described in Section 4.2.6. A patient cell line number that can act as positive control, because it is heterozygous (het.) for the listed polymorphism, is given in the final column.

TABLE 4.3.

Base changes and position	Exon/ Intron Codon	PCR Primers	PCR Conditions	Detection (Enzyme or ASO's)	Products	ASO detection conditions	Positive control
C->A 112/606	I A_8	58-65	65°C DMSO	*Eco*47III or 143 & 144	245-bp or 132 + 113-bp	Hyb. 50°C Wash 62°C	2662 het.
G->T 187/681	I $Q_{33}H$	58-65	65°C DMSO	*Nsp*(7524)I	245-bp or 187 + 58-bp		2827 het.
86-bp VNTR	intron 2 -	53-54	65°C	size only			
T->C 490/479	III L_{118}	49-36	66°C	*Kpn* I	450-bp or 390 + 60		2827 het.
		or 96-97	65°C DMSO	*Kpn* I	605-bp or 440 & 165-bp		
G->A 1169/2192	VIII $A_{361}T$	88-89	65°C DMSO	ASOs 121 & 122	Nor. 447-bp	Hyb. 53°C Wash 62°C	2827 het.
c->t 1486/-	intron 5 -	78-105	55°C DMSO	ASOs 141 & 140	Nor. 324-bp	Hyb. 60°C Wash 69°C	2474 het.

TABLE 4.4. PCR DETECTION OF MPS-I MUTATIONS.

The PCR methods used to quantify mutations in MPS-I patients and the normal population are summarized. The nature of the nucleotide substitution and position in the cDNA (Fig 3.22) and genomic sequence (Fig. 3.24) is given in the first column. The exon of *IDUA*, the amino acid (aa) codon positions and the mutation names are given in the second column. The remaining columns are as described in the legend of Table 4.3 except that the sizes of the PCR product are given for the normal (Nor.) allele mutant allele (Mut.) for restriction detection are given. A patient cell line number that can act as positive control, because it is heterozygous (het.) or homozygous (homo.) for the listed mutation, is given in the final column.

TABLE 4.4.

Base change and position	Exon/ Intron aa/ name	PCR Primers	PCR Conditions	Detection (Enzyme or ASO's)	PCR and digestion Products	ASO detection conditions	Positive control
G->A 1414/354	II $R_{89}Q$	117-66	65°C DMSO	*Bsp*EI or ASO's 147 & 148	Nor. 183 + 158bp Mut. 341bp	Hyb. 53°C Wash 62°C	2662 het.
C->T 1356/296	II $Q_{70}X$	117-66	65°C DMSO	*Mae* I or ASO's 138 & 139	Nor. 341bp Mut. 239 + 102bp	Hyb. 53°C Wash 60°C	296 het.
g->a 1487	intron 5 $g_{1487}a$	78-105	55°C DMSO	ASO's 140, 141, 142 & 151	Nor. 324bp	Hyb. 60°C Wash 69°C	2662 het. 140 2424 het. 142
t->c 1978	intron 7 $t_{1978}c$	106-107	65°C DMSO	ASO's 129 & 130	Nor. 275bp	Hyb. 48°C Wash 54°C	2662 het.
G->A 2156/1133	VIII $D_{349}N$	88-89	65°C DMSO	ASO's 125 & 126	Nor. 447bp	Hyb. 42°C Wash 47°C	2827 het.
G->A 2562/1293	IX $W_{402}X$	109-90	60°C DMSO	*Mae* I or ASO's 136 & 137	Nor. 371bp Mut. 244 + 127bp	Hyb. 55°C Wash 64°C	728 homo.
delG 3247/1702	XI $delG_{1702}$	111-103	65°C DMSO	ASO's 132 & 133	Nor. 302bp	Hyb. 53°C Wash 63°C	61-1 homo.
C->G 3233/1686	XI $P_{533}R$	111-103	65°C DMSO	ASO's 134 & 150	Nor. 302bp	Hyb. 55°C Wash 64°C	283 het.

genomic fragment from λID475 and the 729-bp fragment from λRPC1 as probes. Both these probes detected a single band of approximately 23-kb that showed no signs of deletions or rearrangements (data not shown). The size of this band suggested that small deletions or rearrangements were unlikely to be detected, as resolution at 23-kb on 0.8% (w/v) agarose gels is poor. Preliminary restriction mapping data of the genomic clone λID475, including the identification of the 65/60-kDa N-terminal polypeptide sequences in an 850-bp *Pst* I fragment and peptide 7 sequences in a 1.7-kb *Pst* I fragment, both adjacent to the 1.6-kb fragment, indicated that restriction endonuclease digestion of the patient DNA samples with *Pst* I would enable the detection of at least 3 bands by Southern blot analysis in the size range of best resolution. Southern analysis of 40 MPS-I *Pst* I digested patient DNA samples with the 1.6-kb, 850-bp, 1.7-kb Pst I genomic fragments, the 729-bp cDNA insert of λRPC1, the 1765-bp insert of λE8A and finally the full-length IDUA cDNA insert , IDX21, failed to detect any major deletions or gene rearrangements (data not shown). This was not surprising as in most autosomal genetic diseases studied to date, major deletions or gene rearrangements account for only a small percentage of mutations (Scriver *et al.*, 1989). Thus more detailed molecular analyses of MPS-I patients were attempted.

4.3.2 HAPLOTYPING OF MPS-I CHROMOSOMES USING TWO POLYMORPHISMS WITHIN *IDUA*

During the cloning of *IDUA*, 2 polymorphisms were detected. The first was a VNTR-like polymorphism caused by variation in the number of 86-bp repeats in intron 2 of *IDUA* (Section 3.7.2). The second was a single base substitution of a T->C at position 490 of the cDNA sequence (Fig. 3.22). This change in a wobble base of the Leu_{118} codon of IDUA, which creates a *Kpn* I restriction site, was originally observed as a difference between the sequence of the IDUA cDNA clone λE8A, and the IDUA genomic clone λID475. Exon III of *IDUA*, which is the location of this polymorphism, was PCR amplified from 114 chromosomes from 57 unrelated healthy individuals and digested with *Kpn* I to detect the presence of the polymorphism. PCR conditions used

TABLE 4.5. LINKAGE DISEQUILIBRIUM STUDIES.

The allele frequency of the Kpn I polymorphism and the VNTR polymorphism within the IDUA gene were studied in both 114 and 136 normal chromosomes respectively and 106 MPS-I chromosomes. The alleles of each polymorphism are listed in decreasing order of size. Note that the frequency of allele 2 of the VNTR is increased by 21% in the MPS-I population compared to the normal population. The probability of this being a random association was tested by $\chi^2 = 19.91$, $P<0.001$, 2 *df*.

	Chromosomes		Alleles		
			1	2	3
Kpn I	Normal	(114)	25 (22%)	89 (78%)	-
	MPS-I	(106)	21 (20%)	85 (80%)	-
VNTR[a]	Normal	(136)	38 (28%)	75 (55%)	23 (17%)
	MPS-I	(106)	18 (17%)	81 (76%)	7 (7%)

[a] $\chi^2 = 19.91$, $P<0.001$, 2 *df*

to detect the polymorphism are shown in Table 4.3. Allele 1 (absence of the *Kpn* I site) and allele 2 (presence of the *Kpn* I site) of the polymorphism were present at frequencies of 0.22 and 0.78 respectively in the normal population (Fig. 3.26).

The genotypes of 114 normal individuals and 106 MPS-I patients were determined for the *Kpn* I polymorphism and the genotypes of 136 normal individuals and 106 MPS-I patients were determined for the VNTR polymorphism (Fig. 3.26). The two *Kpn* I alleles and the three VNTR alleles (numbered 1 to 3 in decreasing order of size) were detected by PCR amplification (Table 4.3). There appeared to be no difference in the frequency of the two *Kpn* I alleles in MPS-I patients compared to normal individuals, but there was a strong bias toward allele 2 of the VNTR polymorphism in MPS-I patients, implying linkage disequilibrium between allele 2 of the VNTR and MPS-I (Table 4.5). Of the 106 MPS-I and 136 normal chromosomes typed using the VNTR polymorphism, 76% of MPS-I chromosomes contained allele 2 compared to 55% of normal chromosomes.

The haplotypes for both polymorphisms were examined in 110 normal and 106 MPS-I chromosomes in an attempt to further delineate the linkage disequilibrium (Table 4.6). Because some individuals were heterozygous for both polymorphisms, it was not possible to define haplotypes in all cases (6 MPS-I patients and 9 normal individuals). There was little, if any, linkage disequilibrium between the *Kpn* I and the VNTR polymorphism. However, there was marked deviation from the expected haplotype frequencies in the MPS-I patients. The haplotype most frequently associated with MPS-I was 2,2 (VNTR, 2; *Kpn* I, 2) which represented 57% of MPS-I haplotypes compared with 37% of normal haplotypes. In addition there also seemed to be an increased proportion of the 2,1 allele in MPS-I patients compared to the expected frequency.

The disequilibrium observed between the 2,2 haplotype and MPS-I is indicative of one or more common mutations that arose in a gene carrying this haplotype. The frequency of a major allele associated with the 2,2 haplotype must be between 57% (the haplotype frequency) and 20% (the increase in frequency). It is not possible to

TABLE 4.6. VNTR, *Kpn* I HAPLOTYPE FREQUENCIES IN MPS-I AND NORMAL CHROMOSOMES.

The alleles are listed in decreasing order of size. The probability of the increase in the haplotype frequencies between MPS-I and normals, as tested by a χ^2 test are shown.

Haplotype		Number of Chromosomes		
VNTR	*Kpn* I	Normal	MPS-I[a]	Expected %
1	1	0 (0%)	0 (0%)	2.7%
1	2	20 (19%)	14 (13%)	18.4%
2	1	13 (12%)	[b]15 (14%)	7.8%
2	2	39 (37%)	[c]60 (57%)	38.8%
3	1	1 (1%)	0 (0%)	1.3%
3	2	15 (14%)	5 (5%)	10.8%
Undefined		18 (17%)	12 (11%)	20.2%

[a] $\chi^2 = 27.49$, $P<0.001$, 6 *df*

[b] $\chi^2 = 5.80$, $P<0.05$, 1 *df*

[c] $\chi^2 = 14.11$, $P<0.001$, 1 *df*

determine the severity of this mutation because it is likely that not all 2,2 alleles are associated with the major allele and many patients are heterozygous for this and another haplotype, however the 2,2 haplotype tends to be associated with patients who have severe symptoms. The presence of an increased frequency of another haplotype (2,1) may indicate the presence of another common, though less abundant, MPS-I allele. The linkage disequilibrium between this 2,1 haplotype and MPS-I may be partially "masked" by the high frequency of the 2,2 haplotype in MPS-I patients. It may be that the discovery of disequilibrium at this locus will assist in the classification of MPS-I patients but, since the IDUA gene had been isolated, the most precise approach would be to use the disequilibrium data to select patients for exact molecular definition of these major MPS-I alleles. This information could then be used to define the frequency of this and eventually other alleles so that it may be possible to define the nature of the clinical variation that exists in MPS-I. It is curious that allele 2 of the *Kpn* I polymorphism, a polymorphism within *IDUA*, is not in disequilibrium with MPS-I (Table 4.5). This may be because the frequency of allele 2 is so high (0.78) in the normal population that an increase is not detectable. It should be noted that the use of linkage disequilibrium to characterize the presence of mutations in genetic diseases is not a precise science.

4.4 DETAILED MUTATION ANALYSIS

Chemical cleavage, direct PCR sequencing and quantification of $W_{402}X$ in MPS-I patients were done in collaboration with Tom Litjens of the Department of Chemical Pathology at the Adelaide Children's Hospital. Quantification of other mutations in MPS-I patients was performed in collaboration with Paul V. Nelson of the Department of Chemical Pathology at the Adelaide Children's Hospital.

Initial attempts at mutation analysis of MPS-I patients were designed to amplify the IDUA mRNA from patient fibroblasts in three sections for chemical cleavage (CC) analysis as outlined in Fig. 4.4. However, this approach failed to work consistently, probably due to the high GC content and low abundance of the IDUA mRNA. A

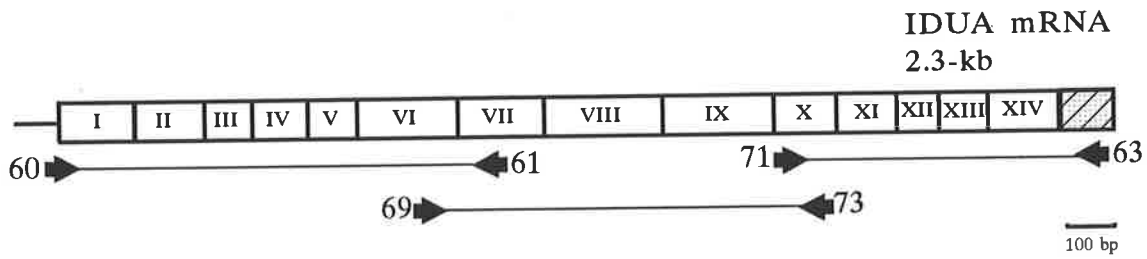


FIGURE 4.4. PCR AMPLIFICATION OF THE IDUA mRNA.

Diagrammatic representation of the PCR reactions attempted for chemical cleavage analysis from IDUA mRNA. The arrows represent the oligonucleotide primers used and the numbers refer to the names of the primers.

number of approaches to overcome this problem were considered. PCR primers of increased length (30-mers compared to the 24-mers currently in use) to allow higher annealing temperatures with efficient priming were considered. An alternative PCR buffer system with different kinetics could be utilized. A thermostable reverse transcriptase (e.g. rTth polymerase, Cetus, Perkin-Elmer or Tet-z polymerase, Amersham) used at high temperatures could possibly eliminate secondary structures in the IDUA mRNA that may cause termination of transcription. However, it was decided that amplification of genomic DNA was the preferred option at this time for the following reasons:

1) Many MPS-I patient DNA samples that had been used in the linkage disequilibrium study were already available, including DNA samples from the patients listed in Table 4.2.

2) RNA preparation is technically more difficult than DNA preparation and fibroblasts for RNA preparation were not available from patients that met the selection criteria discussed in Section 4.1 at this time.

3) The use of genomic DNA templates would avoid the cDNA synthesis step thus avoiding a point of possible low efficiency.

4) Two defined alternative splicing events for IDUA mRNA are described in Section 3.8.1. It is believed that these are not the only alternative splicing events for IDUA mRNA and that the amplification of alternatively spliced mRNA in probes for CC could complicate the interpretation of results and possibly produce cleavage products that could "mask" real cleavage products caused by mutations.

5) If abnormal splicing was detected in MPS-I patients, it would be necessary to determine the nature of these mutations from the genomic sequence of these patients.

6) Some mutations may effect the level of RNA expression (e.g. ornithine delta-aminotransferase mutations in gyrate atrophy, Brody *et al.*, 1992), thus causing difficulty in detecting both alleles in a patient.

7) A relatively large amount of PCR product is preferred for direct PCR sequencing to produce a strong signal. It was felt that this could be more readily and reproducibly produced using genomic DNA as template.

8) For a combination of the above reasons, DNA is the preferred template for PCR amplification for screening large numbers of normal controls and MPS-I patients to determine whether changes are polymorphisms or mutations, and to quantify them in the normal and mutant population.

Oligonucleotide primers were designed from the genomic sequence of the human IDUA gene (Fig. 3.24) and used to PCR amplify all coding regions of *IDUA* (Fig. 4.2) from the selected MPS-I patients (Table 4.2) and a normal control (cosmid A157.1). In general, the GC-rich content of *IDUA* required the use of primers at least 24-nt long and high annealing temperatures in PCR reactions. The oligonucleotide primer, ID60, spans the initiating methionine, thus eliminating 11 coding bases from mutation analysis. It was attempted to amplify exons I and II of *IDUA* with primers in the promoter region of *IDUA* to avoid this problem, however, since these PCR reactions did not work consistently, it was decided to pursue the PCR product produced by ID60 to ID66. A number of 20-mer and 24-mer primer pairs were used to amplify exons III and IV with little success. Eventually, the primers ID96 and ID97 were made as 30-mers and worked consistently well to amplify this GC-rich region. The primer ID63, is 5' to the potential polyadenylation signals in *IDUA* (Fig. 3.24B). The use of primer pairs that would amplify exons XIII and XIV including the potential polyadenylation signals was attempted, however, even 30-mer primers failed to achieve this aim. This may be due to the base composition of the target sequence changeing from GC-rich to AT rich in the 3' untranslated region making it difficult to obey two of the rules for designing PCR primers (Section 2.1.13c), i.e. to match the base composition of primer pairs and to match the base composition of the primers to that of the target sequence.

Thus, the primer pairs shown in Fig. 4.2 were used to amplify all coding sections of IDUA in 5 sections. ID60 to ID66 is a 1035-bp product containing exons I and II. ID96 to ID97 is a 605-bp product containing exons III and IV. ID79 to ID89 is a 1261-bp product containing exons V to XIII. ID90 to ID91 is a 1091-bp product containing exons IX to XII. Finally, ID92 to ID63 is a 579-bp product containing exons XIII and XIV and some of the 3' untranslated region. As described in Section 4.2.3a, the production of probes for CC involved a reamplification step using PCR product as template. The production of probes directly from the genomic DNA of patients by using 100 ng of a kinased primer in PCR reactions was attempted. This was tried for ID60 to ID66 and ID79 to ID89, however it was only successful for ID60 to ID66. It was decided that the reamplification step should be used for the production of all probes for CC to avoid the possible failure of probe production, and the majority of the PCR used as template for this step was unused and could be used as template for direct PCR sequencing if cleavage products were detected. The reamplification step may be necessary to overcome differences in the PCR reactions caused by the use of 5' end-labelled primers and the presence of unincorporated γ-^{32}P-ATP. Once CC changes had been identified, the initial emphasis of this study was to determine whether or not they were the common mutations hypothesized in Section 4.3.2.

4.4.1 COMMON MUTATIONS

4.4.1a DETECTION OF $W_{402}X$

CC analysis of the PCR products described above was used to identify sequence differences between normal and MPS-I patient DNA. The initial CC reactions were performed on the ID60 to ID66 PCR product spanning exons I and II. Two cleavage products common to a number of the patients were immediately observed.

1) The same CC occurred within a region of exon I of the IDUA gene in 5 patients (283, 296, 728, 2474 and 2662). Cleavage products of 46-bp and 989-bp were obtained from primers ID60 and ID66 in the hydroxylamine and osmium tetroxide reactions respectively (data not shown). Direct sequencing

of the PCR products with the primers ID60 and ID65 revealed that the observed cleavages resulted from a single base substitution of C->A (base position 112, Fig. 3.22) in exon I of *IDUA*. The base substitution was in a wobble base of an Ala (A) codon at position 8 (GCC to GCA), thus not affecting the amino acid sequence of the IDUA protein. This observed polymorphism introduces an *Eco*47III (Toyobo) restriction site, and PCR analysis of 32 chromosomes from 16 normal individuals revealed that this polymorphism was present in 31% of normal chromosomes (Table 4.3).

2) The same CC occurred within a region of exon I of the IDUA gene in 3 patients (296, 1062, and 2827). Cleavage products of 121-bp and 914-bp were obtained from primers ID60 and ID66 in the osmium tetroxide and hydroxylamine reactions respectively (data not shown). Direct sequencing of the PCR products with the primers ID116 and ID62 revealed that the observed cleavages resulted from a single base substitution of G->T (base position 187, Fig. 3.22) in exon I of *IDUA* that altered a Gln (Q) codon at position 33 (CAG) to a His (H) codon (CAT). This is a non-conservative amino acid substitution of a neutral-polar amino acid containing an amide side chain (Gln) for a basic amino acid (His) close to the mature N-terminus of IDUA. This base substitution creates a *Nsp*(7524)I (Amersham) restriction site and PCR analysis (Table 4.3) of 32 chromosomes from 16 normal individuals revealed that this polymorphism was present in 9% of normal chromosomes. A normal homozygote for this $Q_{33}H$ substitution was also found, ruling out that this amino acid change is an MPS-I allele. Therefore, this single base substitution is a polymorphism. It is interesting to note that this residue is not conserved between human and canine IDUA (Fig. 3.29).

To further delineate patients for this study, these two polymorphisms in exon I of the IDUA gene were incorporated into the haplotype analysis of the nine patients that had been selected for detailed mutation analysis (Table 4.7). Only patients 538 and 985 were homozygous at all polymorphic loci. The hypothesized common mutation would

presumably be associated with the 1,1,2,2 (*Eco*47III, *Nsp*(7524)I, VNTR, *Kpn* I) haplotype and thus the identification of sequence differences detected by CC in these two patients became a priority as it would lead to the identification of the hypothesized common mutation. Patients that were heterozygous at one or more polymorphic loci were no longer annealed to the normal probe for heteroduplex formation in the CC procedure as two different *IDUA* alleles had been shown to be present, and thus heteroduplexes should automatically form upon denaturation and annealing.

Weak CC occurred within a region of exon IX of the IDUA gene in patients 538 and 985. These were the only cleavages observed in these two patients. The cleavages were 128 bp from primer ID90 in the osmium tetroxide cleavage reactions and 962 bp from primer ID91 in the hydroxylamine cleavage reactions. Direct sequencing of the PCR products with the primers ID27 and ID90 revealed that this CC was the result of a single base substitution of G->A (base position 1293, Fig. 3.22) altering a Trp (W) codon at position 402 (TGG) to a stop (X) codon (TAG), thus resulting in termination of translation approximately two thirds of the way through the 653 amino acid IDUA protein ($W_{402}X$, Fig. 4.5). Cleavage should theoretically not have produced observable bands when using labelled ID90 in the osmium tetroxide reaction, however, a band was observed, presumably due to the partial destabilization of the T residue immediately 3' to the mismatched G/A. When ID91-labelled probe was used in the CC, the osmium tetroxide reaction did not cleave the T/G mismatch, presumably because this mismatch does not destabilize the heteroduplex.

The G->A substitution in $W_{402}X$ introduced a *Mae* I restriction endonuclease site into the IDUA gene. When this was used to test for the presence of $W_{402}X$ in amplified sequences from a total of 76 MPS-I patients, the mutation was found to be present in 34% of all MPS-I alleles (51/152 alleles), including 13 $W_{402}X$ homozygotes, and absent from 20 alleles from normal individuals (Fig. 4.6). It had previously been shown that single base mutations can severely affect the stability of mRNA molecules and thus the site of molecular pathology of the disease (e.g. Brody *et al.*, 1992). To show that the $W_{402}X$ mutation was transcribed into mRNA, total RNA was prepared

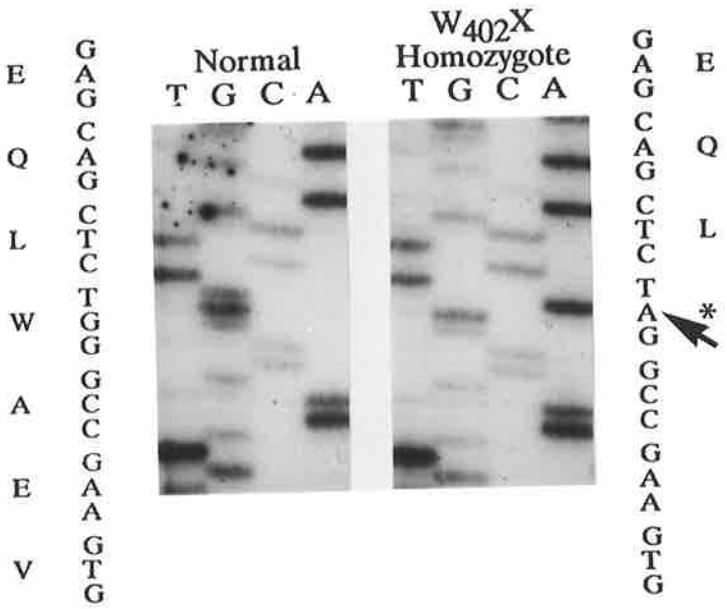
TABLE 4.7. EXTENDED PATIENT HAPLOTYPES.

Extended haplotyping of the nine patients used for detailed mutation analysis of MPS-I. Note that only patients 985 and 538 are still homozygous at all four loci compared to the five patients that were homozygous at the VNTR and *Kpn* I loci (Table 4.2).

Patient	Phenotype	*Eco*47III	*Nsp* I	VNTR	*Kpn* I
283	severe/ intermediate	1 2	1 1	2 2	2 2
296	severe	1 2	2 1	2 1	1 2
538	severe	1 1	1 1	2 2	2 2
728	severe	1 2	1 1	2 2	2 2
985	severe	1 1	1 1	2 2	2 2
1062	severe	1 1	1 2	2 2	2 2
2474	mild	1 2	1 1	2 1	2 2
2662	intermediate/ mild	1 2	1 1	2 3	2 2
2827	severe	1 1	1 2	2 2	2 1

FIGURE 4.5. SEQUENCE OF $W_{402}X$.

Nucleotide and amino acid sequence of a normal individual and a MPS-I patient who is homozygous for the $W_{402}X$ mutation. An arrow indicates the position of the G->A substitution which introduces a stop codon.


E Q L W A E V
GAGCAGCTCTGGGCCGAAGTG
Normal
T G C A
W402X
Homozygote
T G C A
GAGCAGCTCTAGGCCGAAGTG
E Q L *

from a normal individual, an unaffected $W_{402}X$ carrier, an MPS-I $W_{402}X$ heterozygote and an MPS-I $W_{402}X$ homozygote, and cDNA was synthesized from the RNA. A PCR product produced from cDNA using the oligonucleotide primers ID73 and ID127, that spanned $W_{402}X$, was digested with *Mae* I to detect the presence of $W_{402}X$ alleles. The $W_{402}X$ allele was present at approximately the same level as the normal allele when amplified from a normal $W_{402}X$ carrier (Fig. 4.7). This showed that a stable mRNA was transcribed from $W_{402}X$ alleles, and thus should be translated.

Immunochemical assay (Ashton *et al.*, 1992) of IDUA protein in cell lines available from nine of the $W_{402}X$ homozygous patients found no detectable IDUA in eight patients and 0.4% of the average control value in one patient (cell line number 2732, Ashton *et al.*, 1992). Using a disaccharide IDUA substrate, extremely low levels of residual IDUA activity were detected in these cell lines (Hopwood and Muller, 1979; Taylor *et al.*, 1991; Ashton *et al.*, 1992). Four cell lines tested had elevated K_m values compared to normal control fibroblasts (Ashton *et al.*, 1992). These biochemical findings are consistent with the $W_{402}X$ mutation producing a truncated IDUA protein that is unstable and has a reduced ability to bind and turnover substrate.

Thirteen of the 76 MPS-I patients were homozygous for $W_{402}X$, five more than the eight expected on the basis of the frequency of the mutation in MPS-I patients. This may be due to an increased rate of homozygosity in MPS-I patients from a higher than normal rate of consanguinity, however it is not statistically significant ($\chi^2 = 3.49$, $0.1<P<0.05$, 1 *df*).

Patients homozygous for $W_{402}X$ typically presented within their first year of life with classic Hurler-like coarse facial features with a prominent forehead, enlarged tongue, liver and spleen, skeletal defects and stiff joints. Their clinical deterioration was very rapid and death usually occurred before four years of age from cardiac or respiratory failure. These are some of the patients who presented very early and had the most rapid progression to death of all the MPS-I patients. Although all $W_{402}X$ homozygous patients are found at the severe end of the clinical spectrum, there is still some variation in their presentation and rate of disease development. This is probably

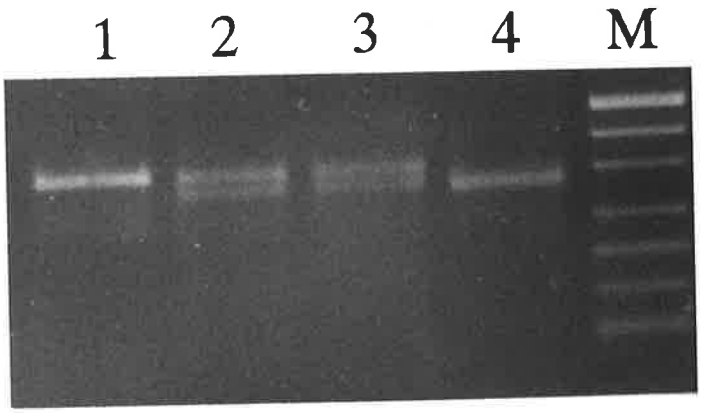
FIGURE 4.6. DETECTION OF $W_{402}X$.

Detection of the $W_{402}X$ mutation with the restriction endonuclease *Mae* I. The sizes of the bands are shown in bp on the right of the gel. Lanes 1 and 6 are MPS-I DNA samples from patients who are homozygous for $W_{402}X$. Lanes 2 and 3 are MPS-I DNA samples which do not have the $W_{402}X$ mutation. Lanes 4, 5 and 7 are heterozygotes for $W_{402}X$.


1
2
3
4
5
6
7
371
244
127

FIGURE 4.7. EXPRESSION OF $W_{402}X$ IN mRNA.

Transcription of the $W_{402}X$ allele. $W_{402}X$ mRNA was detected by PCR from cDNA followed by *Mae* I digestion. The normal PCR product is 333-bp and the $W_{402}X$ mutant band is 301-bp after digestion. The mRNA samples were from: lane 1, a normal individual; lane 2, an unaffected carrier of $W_{402}X$; lane 3, an MPS-I patient who is heterozygous for $W_{402}X$; lane 4, an MPS-I patient who is homozygous for $W_{402}X$. Lane 5 is pUC19 *Hpa* II markers.


1
2
3
4
M

due to heterogeneity in their genetic background and environmental factors such as the type of treatment received, as has been observed for other genetic diseases such as Gaucher's disease (Beutler, 1991).

Heterozygotes for $W_{402}X$ show the full range of clinical presentation of MPS-I. For example, 2474 (GM1323, NIGMS, Human Genetic Mutant Cell Repository) is an index Scheie syndrome case and was a practising attorney in New York (McKusick *et al.*, 1965; Hopwood and Muller, 1979; McKusick and Neufeld, 1983; Muller and Hopwood, 1984; Ashton *et al.*, 1992). This patient is heterozygous for $W_{402}X$. Biochemically, 2474 fibroblasts have no detectable IDUA protein using 2 different IDUA monoclonal antibodies. GM1323 fibroblasts have approximately 0.3% of IDUA activity and a K_m in the normal range (Table 4.2; Ashton *et al.*, 1992). This IDUA activity must result from a mild mutation in the other undefined MPS-I allele present in this patient.

As expected from the linkage disequilibrium studies, $W_{402}X$ seems to be rigidly associated with the 2,2 haplotype (Section 4.3.2). If the 2,2 (VNTR, *Kpn* I) alleles associated with $W_{402}X$ are subtracted from the number of 2,2 MPS-I alleles used in the linkage disequilibrium studies, there is no longer disequilibrium. However, the disequilibrium between the 2,1 haplotype and MPS-I is greatly increased ($\chi^2 = 18.08$, $P<0.0005$ 1 *df*, Table 4.8). This implies that $W_{402}X$ was the mutation responsible for the disequilibrium between the 2,2 haplotype and MPS-I and strongly implied the presence of a second major MPS-I allele. Thus, the patients that were heterozygous for the 2,1 haplotype (296 and 2827) and had been selected for detailed mutation analysis were examined in more detail to determine the nature of this second postulated common mutation.

4.4.1b DETECTION OF $Q_{70}X$

A strong CC was observed within exon IV of IDUA in both patients 296 and 2827 (Fig 4.8). This cleavage was approximately 170-bp from ID97 and occurred in both the osmium tetroxide and hydroxylamine reactions, implying an A/G base mismatch. As this was a strong cleavage in both patients that possibly had the 2,1

haplotype, and thus the second common MPS-I mutation, sequencing of this region was attempted. However, no base changes were observed in either patient from multiple sequencing attempts. Two other patients who were heterozygous for $W_{402}X$ and the 2,1 haplotype (340, and 1805), and one patient who was homozygous for the 2,1 haplotype (3296) were selected and CC was performed on the ID96 to ID97 PCR product at the same time as CC on 296 and 2827. The same CC products were once again observed but sequencing of exon IV from all five patients still failed to detect a base change. It was thought that the two MPS-I alleles were possibly being PCR amplified at different rates and, as no base change was observed in patient 3296, all patients must have been heterozygous. By digesting the ID96 to ID97 PCR product with *Kpn* I, the allele associated with the 2,1 haplotype was isolated from patient 296. It appeared from these digestions that the alleles were evenly amplified. The full-length PCR product remaining after digestion (the allele associated with the 2,1 haplotype) was isolated and directly sequenced, but still no base change was observed. Thus the PCR product was digested with *Msc* I to give a blunt ended fragment that could be cloned. This 433-bp fragment was isolated and subcloned into a *Sma* I M13mp18 vector, and recombinants were obtained and sequenced using the universal sequencing primer (USP). In this way, the region of interest and exon III, could be sequenced to determine if the correct allele had been isolated (Fig. 4.9). Four clones were shown to be derived from the *Kpn* I 1 allele, and four from the 2 allele but base changes that could account for the observed cleavage products were not detected by sequencing in both directions. The observed cleavages were therefore thought to be artifacts associated with the presence of the *Kpn* I 1 allele. The mismatch caused by this polymorphism in the heteroduplex possibly alters the stability of the region in exon IV where this artifactual cleavage occurs. No examples of CC artifacts of this nature have been described in the literature. Thus, it was decided to concentrate on other possible cleavages that were common to patients 296 and 2827 in order to identify the second common MPS-I mutation.

Closer examination of the other CC reactions from 296 and 2827 revealed that the same CC occurred within a region of exon II of *IDUA*. Weak cleavage products of 797-bp were obtained from primer ID60 in both the osmium tetroxide and

TABLE 4.8. HAPLOTYPE FREQUENCIES IN MPS-I CHROMOSOMES MINUS $W_{402}X$ ALLELES.

Haplotype frequencies of the 2,2 haplotype in MPS-I chromosomes using the VNTR and the *Kpn* I polymorphisms before and after the $W_{402}X$ alleles were subtracted. For $W_{402}X$, 2 indicates the presence of the mutation as detected by *Mae* I or ASOs. Note that the linkage disequilibrium of MPS-I with the 2,2 haplotype is negated (a->b) and the disequilibrium with the 2,1 haplotype is increased (c->d).

Haplotype			Number of Chromosomes		
VNTR	*Kpn* I	$W_{402}X$	MPS-I	Expected %	Total
2	2		[a]60 (57%)	38.8%	106
2	2	1	[b]24 (34%)	38.8%	70
2	2	2	36		
2	1		[c]15 (14%)	7.8%	106
2	1		[d]15 (21%)	7.8%	70

[a] $\chi^2 = 14.11$, $P<0.001$, 1 *df*

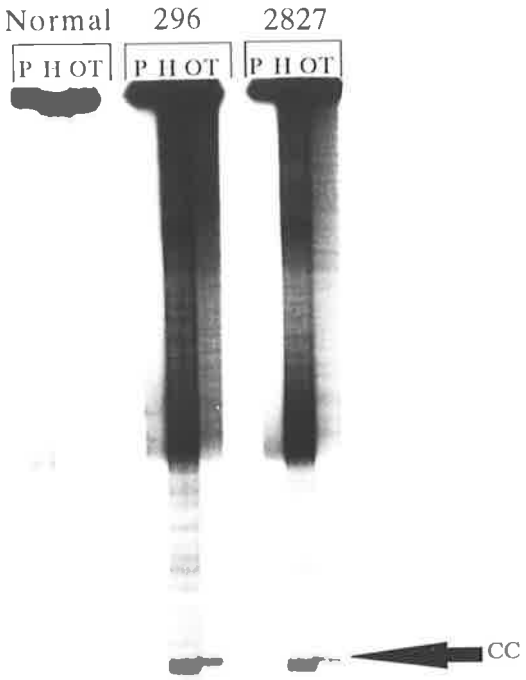
[b] $\chi^2 = 0.60$, $0.4<P<0.5$, 1 *df*

[c] $\chi^2 = 5.80$, $P<0.05$, 1 *df*

[d] $\chi^2 = 18.08$, $P<0.0005$ 1 *df*

FIGURE 4.8. CHEMICAL CLEAVAGE IN EXON IV OF PATIENTS 296 AND 2827.

Chemical cleavage of the ID96 to ID97 PCR products from patients 296 and 2827 are shown on a 5% (w/v) denaturing polyacrylamide gel. The homoduplex control is shown in the first set of lanes and the order of the lanes for each set of reactions is: unreacted probe (P), hydroxylamine treated reactions (H) and osmium tetroxide treated reactions (OT). The arrow indicates the position of the approximately 170-bp cleavage product. As described in Section 4.4.1b, this cleavage is believed to be an artifact.


Normal
296
2827
P H OT
P H OT
P H OT
CC

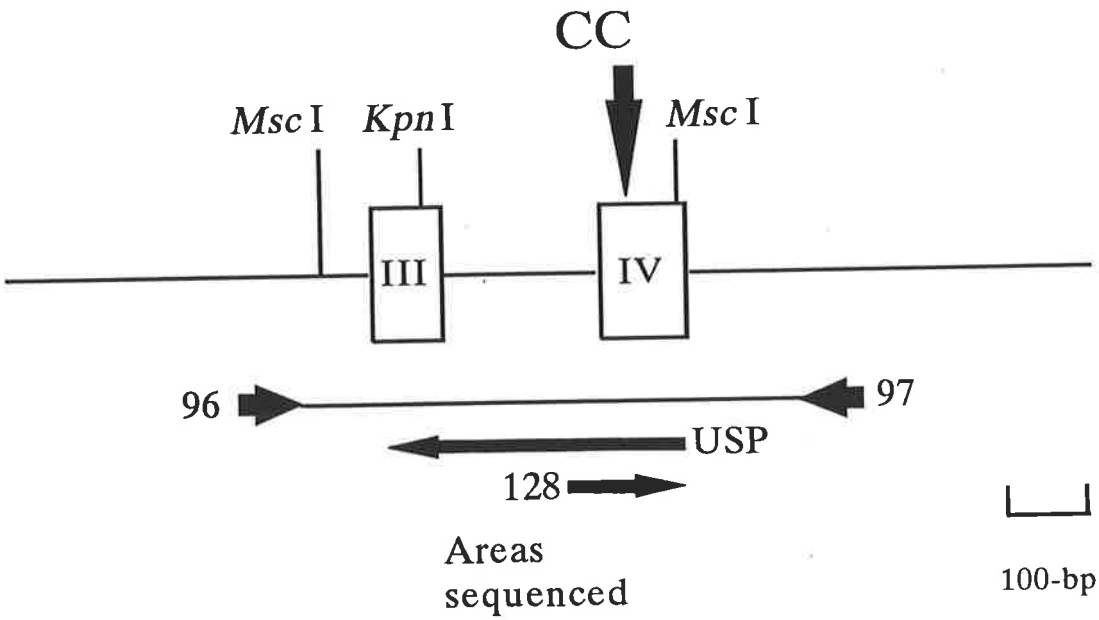


FIGURE 4.9. CLONING OF THE 2,1 ALLELE.

A diagrammatic representation of exons III and IV of IDUA is shown. The position of the PCR primers ID96 and ID97 are shown. The ID96 to ID97 PCR product from patient 296 was digested with the restriction enzyme *Msc* I, cloned, and the area in exon IV of IDUA involved in the cleavage (arrowed and labelled CC) was sequenced in both directions with the primers ID128 and the USP as indicated. Sequence read from the USP spanned the *Kpn* I polymorphic site enabling accurate determination that the 2,1 allele had been isolated.

hydroxylamine reactions (data not shown). Direct sequencing of the PCR products with the primers ID117 and ID118 revealed that the observed cleavages resulted from a single base substitution of C->T (base position 296, Fig. 3.22), altering a Gln (Q) codon at position 70 (CAG) to a stop (X) codon (TAG) and thus resulting in early termination of translation of the 653 amino acid IDUA protein ($Q_{70}X$, Fig. 4.10). PCR analysis with ASOs (Table 4.4) of 40 chromosomes from 20 normal individuals revealed that this substitution was not present in normal chromosomes and was present in 22 of 146 MPS-I chromosomes, thus accounting for 15% of all MPS-I alleles (Fig. 4.11). Two $Q_{70}X$ homozygotes and six $Q_{70}X/W_{402}X$ compound heterozygotes were found in the group of 73 patients that have been completely analysed for $W_{402}X$ and $Q_{70}X$, which correlates with the expected numbers.

$Q_{70}X$ is in effect a null allele. It has been found that mutations leading to early termination of translation can decrease the stability of the mutant mRNA, presumably due to the lack of ribosomes at the 3' end of the message (Brody *et al.*, 1992). The stability of the $Q_{70}X$ mRNA has not been tested, but it is expected that a decrease in mRNA stability would not dramatically alter the path of molecular pathogenesis for the $Q_{70}X$ mutation. It is expected that the $Q_{70}X$ mutation produces a severely truncated IDUA protein that is unstable and has no ability to bind and turnover substrate. Biochemical findings are consistent with this hypothesis. Immunochemical assays to detect the presence of IDUA protein in cell lines available from both $Q_{70}X$ homozygous patients and from three of the six $W_{402}X/Q_{70}X$ compound heterozygotes found no detectable IDUA protein (see cell line number 1805; Ashton *et al.*, 1992). Using a heparin-derived disaccharide IDUA substrate, extremely low levels of residual IDUA activity were detected in two $W_{402}X/Q_{70}X$ cell lines (see cell lines KB, Hopwood and Muller, 1979; Muller and Hopwood, 1984; and 1805, Ashton *et al.*, 1992). Cell line KB had an elevated K_m value compared to normal control fibroblasts. This elevated K_m is probably due to the $W_{402}X$ allele. $Q_{70}X$ is in exon II of *IDUA*. As described in Section 3.8.1, alternative splicing of the IDUA gene produces an mRNA transcript missing exon II of *IDUA*, and thus patients homozygous for the $Q_{70}X$ mutation would still produce an IDUA mRNA missing exon II that can be translated. As the clinical phenotype of

patients homozygous for $Q_{70}X$ is so severe and IDUA activity present in their fibroblasts is negligible, this implies that the translational product from this transcript has no IDUA-like function.

The clinical phenotype of a $Q_{70}X$ homozygote has been described (Pedersen *et al.*, 1979). Briefly, coarse facial features were noticed at birth, the child developed the appearance typical of Hurler syndrome and died of cardiac insufficiency at the age of 2 years and 7 months. $W_{402}X/Q_{70}X$ compound heterozygotes typically presented within their first year of life with classic Hurler-like symptoms similar to, or even more severe than $W_{402}X$ homozygotes (Section 4.4.1a and Scott *et al.*, 1992b). Their clinical deterioration was very rapid and death usually occurred before four years of age from cardiac or respiratory failure. As for $W_{402}X$ homozygotes, all patients who are $Q_{70}X$ homozygotes or $Q_{70}X/W_{402}X$ compound heterozygotes are found at the severe end of the clinical spectrum, but there may still be some variation in their presentation and rate of disease development. Again, this may be due to heterogeneity in their genetic background and environmental factors such as the type of treatment received.

As expected from the linkage disequilibrium studies, $Q_{70}X$ seems to be associated with the 2,1 haplotype (Section 4.3.2). If the 2,1 (VNTR, *Kpn* I) alleles associated with $Q_{70}X$ are subtracted from the number of uncharacterized MPS-I alleles used in the linkage disequilibrium studies, there is no longer disequilibrium between MPS-I and the 2,1 haplotype (Table 4.9). This implies that $Q_{70}X$ was the mutation responsible for the disequilibrium. If all uncharacterized MPS-I alleles used in the linkage disequilibrium study are further analysed, there is no further linkage disequilibrium, although there is an apparent increase in the frequency of 1,2 haplotypes compared to the expected frequency (24% to 18.4%) and a small decrease in the frequency of three other haplotypes (Table 4.10). This increase may become significant if the total number of uncharacterized MPS-I chromosomes is increased. As is described in the following sections, there are other less common MPS-I mutations that do not cause linkage disequilibrium in the group of patients studied, however, it then

FIGURE 4.10. SEQUENCE OF $Q_{70}X$.

Nucleotide and amino acid sequence of a normal individual and an MPS-I patient who is homozygous for the $Q_{70}X$ mutation. An arrow indicates the position of the C->T substitution which introduces a stop codon.

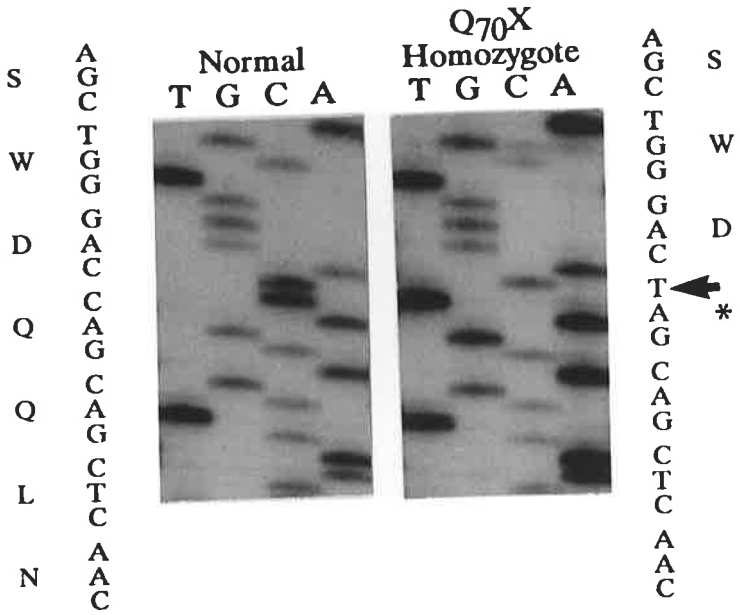

S W D Q Q L N
AGCTGGGACCAGCAGCTCAAC
Normal
T G C A
Q70X
Homozygote
T G C A
AGCTGGGACTAGCAGCTCAAC
S W D *

FIGURE 4.11. ASO DETECTION OF $Q_{70}X$.

Detection of the $Q_{70}X$ mutation with the ASOs ID138 (normal) and ID139 (mutant) as described in Section 4.2.6 and Table 4.4. Lane 1 is an MPS-I sample from a patient homozygous for $Q_{70}X$. Lanes 3, 4, 5, 7, 8 are MPS-I patient samples which do not have the $Q_{70}X$ mutation. Lanes 2, 6 and 9 are samples from patients who are heterozygotes for $Q_{70}X$.

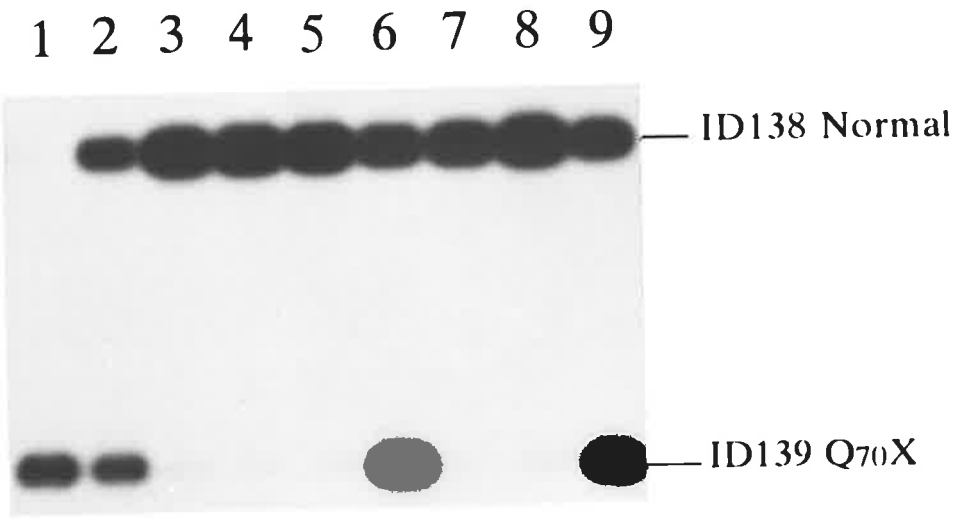

1 2 3 4 5 6 7 8 9
ID138 Normal
ID139 Q70X

TABLE 4.9. HAPLOTYPES FREQUENCIES IN MPS-I CHROMOSOMES MINUS $W_{402}X$ AND THEN $Q_{70}X$

Haplotype frequencies of the 2,1 haplotype in MPS-I chromosomes without a known mutation using the VNTR and the *Kpn* I polymorphisms before and after the $W_{402}X$ and then the $Q_{70}X$ alleles are subtracted. For $Q_{70}X$, 2 indicates the presence of the mutation. Note that the linkage disequilibrium of MPS-I with the 2,1 haplotype is negated (b->c).

Haplotype			Number of Chromosomes		
VNTR	*Kpn* I	$Q_{70}X$	MPS-I	Expected %	Total
2	1		[a]15 (14%)	7.8%	106
2	1		[b]15 (21%)	7.8%	70
2	1	1	[c]3 (5.1%)	7.8%	58
2	1	2	12		

[a] $\chi^2 = 5.80$, $P<0.05$, 1 *df*

[b] $\chi^2 = 18.08$, $P<0.0005$ 1 *df*

[c] $\chi^2 = 0.54$, $0.4<P<0.5$, 1 *df*

seems likely that most of the remaining MPS-I alleles in this group of patients will be rare or unique.

4.4.2 RARE OR UNIQUE MUTATIONS

4.4.2a DETECTION OF $P_{533}R$

During the time of identification and quantification of the two common mutations in the MPS-I population, CC reactions had been performed on all five *IDUA* PCR products from each of the nine patients selected for detailed mutation analysis. As cleavage products were observed and sequenced, the base changes were quantified in the normal population by exon specific PCR and ASOs or restriction endonuclease detection, as a preliminary assessment to determine if they were polymorphisms or mutations. If the base changes were not present in the normal population, MPS-I patient DNA samples were analysed to determine if the changes occurred in MPS-I chromosomes more than once. If a nucleotide substitution causing an amino acid substitution was not detected in the normal population and was detected in more than one MPS-I patient, then it was concluded that the nucleotide substitution was likely to be a mutation. Two nucleotide substitutions of this nature were detected.

Cleavage products of 799 bp and 292 bp from primers ID90 and ID91 respectively, in the hydroxylamine reactions only, were obtained within exon XI of patient 283 (data not shown). Direct sequencing of the PCR products with the primers ID43 and ID103 revealed that the observed cleavages resulted from a single base substitution of C->G (base position 1686, Fig. 3.22), altering a Pro (P) codon at position 533 (CCG) to an Arg (R) codon (CGG) ($P_{533}R$, Fig. 4.12).

$P_{533}R$, a non-conservative substitution of a neutral non-polar amino acid with a basic amino acid, was absent in 82 normal alleles and found in 3% of MPS-I alleles (5/146). A homozygote for $P_{533}R$ with a severe Hurler clinical phenotype was the result of a consanguineous mating. Although immunochemical data is not available on the $P_{533}R$ homozygote, two $P_{533}R$ heterozygous fibroblast cell lines have been tested. One, 3296, had no detectable IDUA protein implying that $P_{533}R$ is a very destabilizing

TABLE 4.10. HAPLOTYPE FREQUENCIES IN UNCHARACTERIZED MPS-I CHROMOSOMES.

The VNTR, *Kpn* I haplotype frequencies in the 58 uncharacterized MPS-I chromosomes after the $W_{402}X$ and $Q_{70}X$ alleles are subtracted from the chromosomes used in the linkage disequilibrium studies (Table 4.3). The alleles are listed in decreasing order of size. Note that there is no longer any significant disequilibrium.

Haplotype		Number of Chromosomes	
VNTR	*Kpn* I	MPS-I[a]	Expected %
1	1	0 (0%)	2.7%
1	2	[b]14 (24%)	18.4%
2	1	[c]3 (5%)	7.8%
2	2	[d]24 (41%)	38.8%
3	1	0 (0%)	1.3%
3	2	5 (9%)	10.8%
Undefined		12 (21%)	20.2%

[a] $\chi^2 = 4.50$, $0.6<P<0.7$, 6 *df*

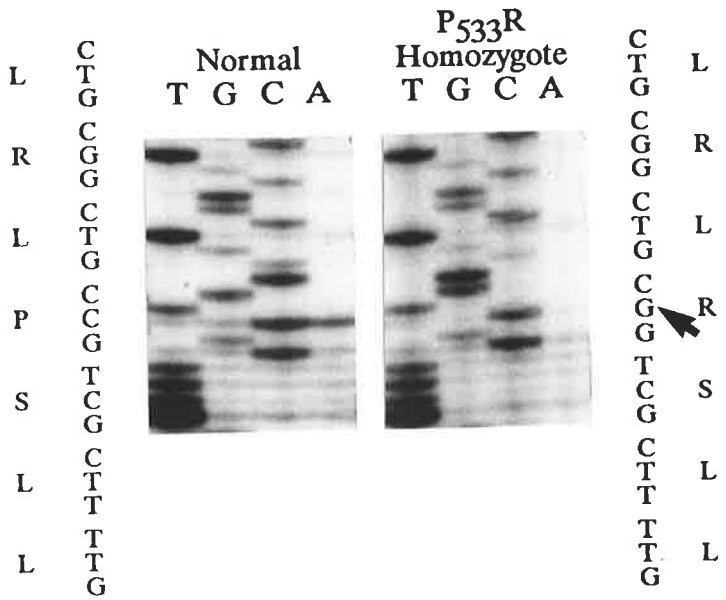
[b] $\chi^2 = 1.27$, $0.2<P<0.3$, 1 *df*

[c] $\chi^2 = 0.55$, $0.4<P<0.5$, 1 *df*

[d] $\chi^2 = 0.10$, $0.7<P<0.8$, 1 *df*

FIGURE 4.12. SEQUENCE OF $P_{533}R$.

Nucleotide and amino acid sequence of a normal individual and an MPS-I patient homozygous for the $P_{533}R$ mutation. An arrow indicates the position of the C->G substitution which alters amino acid 533 of IDUA from Pro (P) to Arg (R).


L R L P S L L
CTG CGG CTG CCG TCG CTT TTG
Normal
T G C A
P533R
Homozygote
T G C A
CTG CGG CTG CGG TCG CTT TTG
L R L R S L L

amino acid substitution or destroys the antibody epitope detected by this monoclonal antibody. The other patient, a $W_{402}X/P_{533}R$ compound heterozygote, had 1% and 7% of the normal mean detectable IDUA protein with the Id1A and Id17A monoclonal antibodies respectively. Using the fluorogenic substrate 4-methylumbelliferyl-α-L-iduronide, these fibroblasts also had 0.5% of the normal mean IDUA activity and a greatly increased K_m (260 compared with 35 μM). It is expected that the $W_{402}X$ allele does not produce any detectable IDUA protein (Section 4.4.1a and Scott *et al.*, 1992b) suggesting that the immunoreactive material must be produced by the $P_{533}R$ allele. Thus the immunochemical data from this patient implies that the $P_{533}R$ allele is producing an unstable IDUA protein with a greatly reduced ability to bind and/or turnover IDUA substrate compared to normal. In a group of 12 patients with detectable monoclonal antibody epitopes, 283 is the only patient cell line that has an appreciable increase in the percentage of the Id17A epitope compared with the Id1A epitope (cell line number 517, Ashton *et al.*, 1992). This may reflect that the instability caused by $P_{533}R$ is greatest around the actual amino acid substitution, and thus that the Id1A epitope is closer to $P_{533}R$ than the Id17A epitope. The conflicting information from these $P_{533}R$ compound heterozygotes may be clarified by immunochemical analysis of the $P_{533}R$ homozygous patient fibroblasts or expression of the $P_{533}R$ mutation. General MPS-I disease progression, including mental deterioration, is less severe in the $W_{402}X/P_{533}R$ compound heterozygote and the $P_{533}R$ homozygous patients compared with $W_{402}X$ or $Q_{70}X$ homozygotes or $W_{402}X/Q_{70}X$ compound heterozygotes. The $W_{402}X/P_{533}R$ compound heterozygote patient died at 12 years of age.

As reported for MPS-I patients who are $W_{402}X$ heterozygotes (Section 4.4.1a and Scott *et al.*, 1992b), heterozygotes for $Q_{70}X$ or $P_{533}R$ may show the full range of clinical presentations of MPS-I. Mild phenotypes may result from a mutation in the patient's other MPS-I allele that produces a mutant IDUA protein with some stability and capacity to bind and turnover substrate. As for $W_{402}X$ and $Q_{70}X$ heterozygotes, this emphasises the need to define both alleles before phenotype prediction is attempted. Biochemical and immunochemical analysis of patient fibroblasts may also be valuable in the prediction of patient phenotype (Ashton *et al.*, 1992).

4.4.2b DETECTION OF $R_{89}Q$

A cleavage product unique to patient 2662 from the nine patients analysed was observed in exon II of *IDUA*. Weak cleavage products of 180-bp were obtained from primer ID66 in both the osmium tetroxide and hydroxylamine reactions. Direct sequencing of the PCR products revealed that the observed cleavages resulted from a single base substitution of G->A (base position 354, Fig. 3.22) in exon II of *IDUA* altering an Arg (R) codon at position 89 (CGG) to a Gln (Q) codon (CAG) ($R_{89}Q$).

This non-conservative substitution of a basic amino acid (Arg) for a neutral-polar amino acid (Gln) occurs at a residue that is conserved between human and canine IDUA and also between IDUA and β-xylosidase (Figs. 3.29 and 3.32 respectively). The base substitution destroys a *Bsp*EI restriction site which was confirmed by the loss of this restriction site in one allele of patient 2662 (Table 4.4). PCR analysis with ASOs of 44 chromosomes from 22 normal individuals revealed that this substitution was not present in normal chromosomes and was present in 3 of 146 MPS-I chromosomes, implying that $R_{89}Q$ is a mutant *IDUA* allele that causes MPS-I.

Immunochemical analysis of fibroblasts available from one of the two $W_{402}X/R_{89}Q$ compound heterozygotes detected 1.5% of the normal levels of IDUA protein with two different monoclonal antibodies, 0.5% of normal IDUA activity with 4-methylumbelliferyl-α-L-iduronide, and an increased K_m with a heparin-derived disaccharide substrate (333 μM compared with 12-111 μM in normal control fibroblasts, cell line number 2730, Ashton *et al.*, 1992). However the specific activity is not greatly decreased (15.3 compared with 30-93 pmol/min/ng of IDUA protein in normal control fibroblasts). This implies that $R_{89}Q$ is a destabilizing amino acid substitution producing IDUA protein that has a reduced ability to bind and/or turnover substrate. Ashton *et al.* (1992) found that, except for one patient with a mild phenotype with no immunodetectable IDUA (cell line number 2474), the presence of some immunodetectable IDUA protein was coincident with mild or intermediate patient phenotype. The fact that the $R_{89}Q$ also appears to be producing a detectable mutant

IDUA protein with only slightly decreased specific activity would probably enable prediction of a mild patient phenotype.

Two patients that are $W_{402}X/R_{89}Q$ compound heterozygotes are both mildly affected. One patient (44-1, Royal Manchester Children's Hospital) is 14 yrs of age, mentally normal, but has corneal clouding and a number of skeletal abnormalities, including badly clawed hands and severe backache although her stature is reasonable for her age (140 cm). She also has mitral valve thickening, but no cardiomyopathy. The other $W_{402}X/R_{89}Q$ compound heterozygous patient, 2730, was diagnosed at 29 years of age and is mentally normal, but has an enlarged tongue, corneal clouding, mitral valve thickening, and short stature. In contrast to $W_{402}X$, $Q_{70}X$ and $P_{533}R$, it is expected that most MPS-I patients with one $R_{89}Q$ allele will be mildly affected with normal intelligence. The small amount of $R_{89}Q$ IDUA protein present in these patients must be enough to produce the relatively mild phenotypes observed.

As previously hypothesized for MPS-I mutations (Matalon *et al.*, 1983), $R_{89}Q$ may be an MPS-I allele that has a greatly reduced ability to degrade DS but a less reduced ability to degrade HS, explaining the minimal CNS involvement in these patients. It will be possible to test this hypothesis, and whether $R_{89}Q$ results in reduced substrate binding and/or turnover, by *in vitro* expression of $R_{89}Q$ IDUA protein and other mutant IDUA proteins with conservative amino acid substitutions at R_{89}. The clinical phenotype of $R_{89}Q$ heterozygotes will be discussed in more detail in Section 4.4.2f.

4.4.2c DETECTION OF $delG_{1702}$

From the group of 9 patients selected for detailed mutation analysis, 4 MPS-I alleles remained to be defined. Thus, the CC reactions performed on these patients were analysed closely. A unique CC occurred within a region of exon XI of *IDUA* in patient 296. The following cleavages, all relating to the same base change, were observed; a strong cleavage product of 276-bp was obtained from primer ID91 in the hydroxylamine reaction, a weak cleavage of 276-bp was obtained from primer ID91 in the osmium tetroxide reaction, weak cleavage products of 795-bp were obtained from

primer ID90 in both the osmium tetroxide and hydroxylamine reactions (data not shown). After sequencing of this region in the sense direction with primer ID43, this base change was originally designated $delG_{1700}$, however ASO hybridization revealed that this was incorrect. The mutation was resequenced in the antisense direction using primer ID103, leading to the correct assignment of the deleted base which was confirmed by PCR analysis and ASO hybridization (Table 4.4). Direct sequencing of the ID90 to ID91 PCR product from patient 296 revealed that the observed cleavages resulted from a deletion of a single G residue ($delG_{1702}$, base position 1702, Fig. 3.22) resulting in a frameshift mutation.

The altered reading frame from codon 538 (Val) would therefore be translated until an inframe stop codon is reached at position 1764 (TGA), thus resulting in a 558 amino acid IDUA protein matching the normal IDUA amino acid sequence for the first 538 amino acids, followed by 20 incorrect amino acids at the C-terminal end (Fig. 4.13). PCR analysis and ASO hybridization of 40 chromosomes from 20 normal individuals revealed that this deletion was not present in normal chromosomes and was unique in the 73 MPS-I patients analysed.

No immunodetectable IDUA protein is present in fibroblasts from patient 296 suggesting that the IDUA protein produced from the $delG_{1702}$ allele is comparable to the $W_{402}X$ mutant IDUA protein, as they are both truncated to approximately the same size. Thus the $delG_{1702}$ protein would be expected to be very unstable and have little ability to bind and/or turnover IDUA substrate thus resulting in a severe phenotype.

Thus the genotype of patient 296 was found to be $Q_{70}X/delG_{1702}$. As previously stated, $Q_{70}X$ is a null allele for *IDUA* and therefore the phenotype of patient 296 can be directly related to the $delG_{1702}$ mutation. Patient 296 was severely affected with a classical Hurler syndrome phenotype. She was diagnosed at 8 months of age and had a large head, coarse facial features with broad saddle nose and wide nostrils, thick lips, an enlarged tongue, an open mouth, corneal clouding, hepatosplenomegaly, and bilateral peribronchial thickening. Patient 296 died at 4 years and 9 months of age from bronchopneumonia.

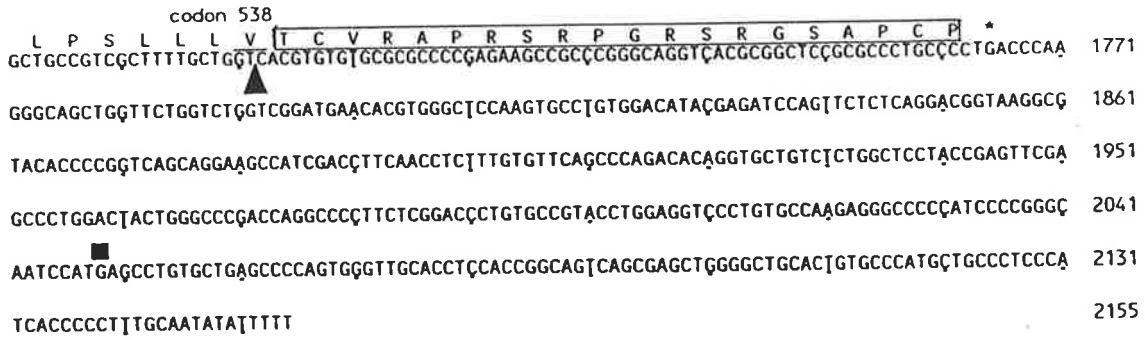


FIGURE 4.13. THE FRAMESHIFT RESULTING FROM $delG_{1702}$.

The cDNA sequence of the $delG_{1702}$ allele from patient 296 showing the deduced sequence of the resulting protein. The position of the deleted base is indicated by an arrow and the sequence up to and including codon 538 (V_{538}, labelled above and underlined) of the IDUA peptide is normal. The 20 incorrect amino acids at the carboxy terminus of this mutant IDUA protein are boxed. The stop codon utilized is marked by an asterisk and the normal stop codon by a square. Base sequence numbers given on the right are the normal cDNA sequence numbers corresponding to Fig. 3.22.

4.4.2d DETECTION OF $D_{349}N$

It proved difficult to define the MPS-I alleles in patients 2662 and 2827 since many cleavages were observed in both patients (Fig. 4.14). All points of cleavage in these patients were eventually sequenced. The same CC occurred within a region of exon VIII of *IDUA* in both patients. Cleavage products of 265-bp were obtained from primer ID89 in both the osmium tetroxide and hydroxylamine reactions (Fig. 4.14). Direct sequencing of the PCR products with the primers ID108 and ID101 revealed that the observed cleavages resulted from a single base substitution of G->A that altered an Ala (A) codon at position 361 (GCG) to a Thr (T) codon (ACG) (base position 1133, Fig. 3.22; $A_{361}T$). This is a structurally conservative amino acid substitution of an neutral-non-polar amino acid (Ala) for a neutral-polar amino acid (Thr). PCR analysis with ASOs (Table 4.4) of 38 chromosomes from 19 normal individuals revealed that this substitution was present in normal chromosomes. In fact, the only $A_{361}T$ alleles detected were from a normal healthy homozygote, precluding this amino acid substitution from causing MPS-I.

At this time the $Q_{70}X$ and $R_{89}Q$ alleles were defined, and thus one MPS-I allele had been detected in both patients 2827 and 2662 respectively. The undefined MPS-I alleles in these patients were expected to be different from each other due to the vastly different clinical phenotypes and the immunochemical data available (Table 4.2). The undefined MPS-I allele in patient 2827 was expected to be the mutation which resulted in the large amounts of inactive IDUA protein detected in fibroblasts, while the undefined allele in patient 2662 was uncharacterized by either clinical or immunochemical techniques. It was thus expected that the cleavage products leading to detection of these undefined alleles would be unique to each patient. A unique CC occurred within a region of exon VIII of *IDUA* in patient 2827. Cleavage products of 301-bp were obtained from primer ID89 in both the osmium tetroxide and hydroxylamine reactions (Fig. 4.14). Direct sequencing of the PCR products with the primers ID108 and ID101 revealed that the observed cleavages resulted from a single base substitution of G->A that altered an Asp (D) codon at position 349 (GAC) to an

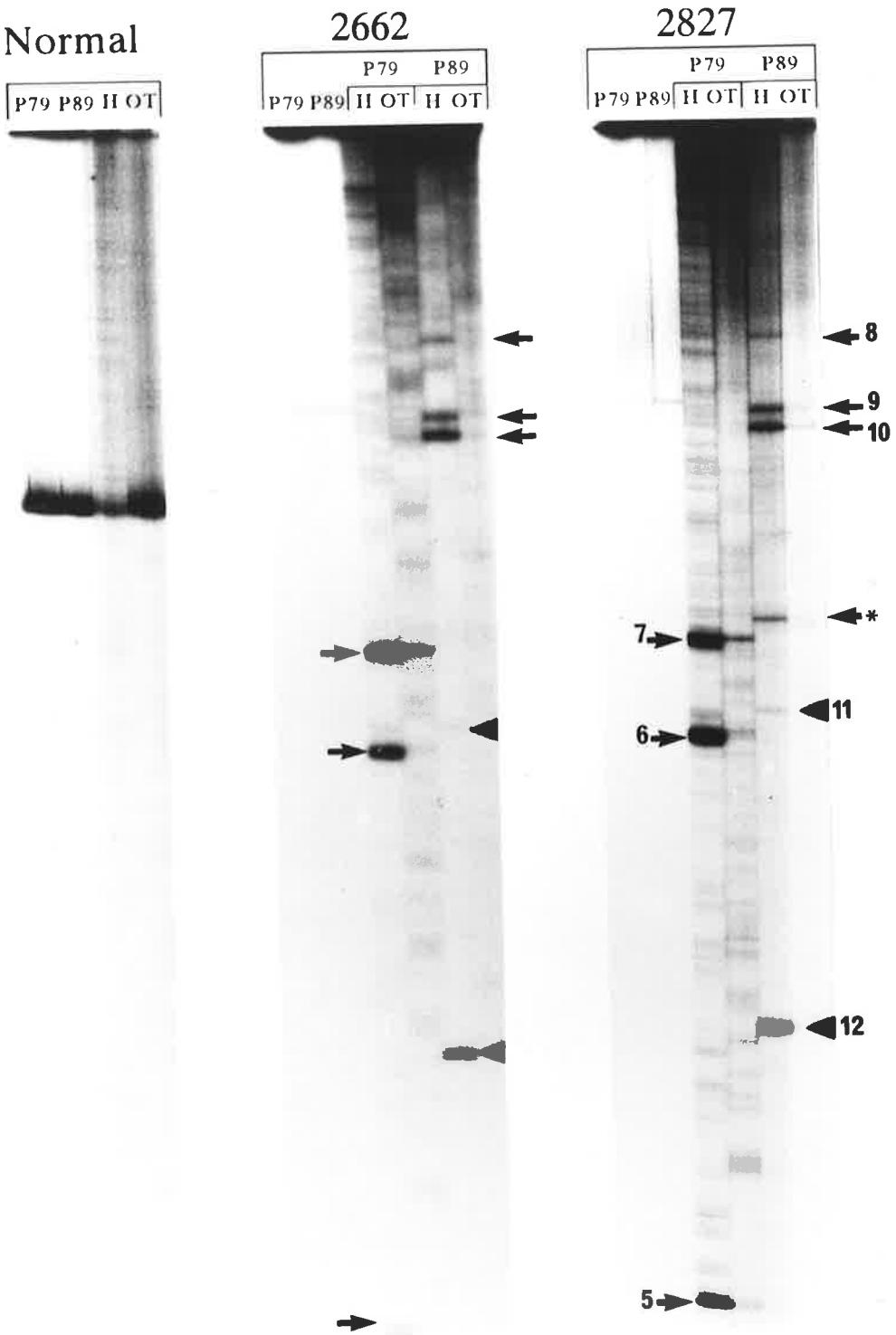
Asn (N) codon (AAC) (base position 1133, Fig. 3.22; $D_{349}N$). PCR analysis with ASOs (Table 4.4) of 40 chromosomes from 20 normal individuals revealed that this substitution was not present in normal chromosomes and, as expected from the immunochemistry, was unique in the 73 MPS-I patients analysed.

The clinical description and immunochemistry of patient 2827 has been described in detail (Brooks *et al.*, 1992). Patient 2827 was 3 yrs old at the time of clinical evaluation. Briefly, the patient was severely affected with coarse facial features, dwarfism, developmental delay, and corneal clouding. Fibroblasts from patient 2827 had a 6-fold greater amount of IDUA protein than normal controls and extremely low or undetectable levels of activity towards 4-methylumbelliferyl-α-L-iduronide or a heparin-derived disaccharide substrate. Maturation studies had shown apparently normal levels of IDUA synthesis with delayed processing to the normal mature 69-kDa form of IDUA present in the lysosomes (Taylor *et al.*, 1991; Brooks *et al.*, 1992). It was hypothesized that patient 2827 has a mutation that affects both the active site and post-translational processing of the enzyme and that the mutation is structurally conservative as it does not result in instability either during trafficking from the ER to the lysosome or maturation in the lysosome (Brooks *et al.*, 1992).

Thus the genotype of patient 2827 is $Q_{70}X/D_{349}N$. It has been shown that $Q_{70}X$, is in effect a null allele for *IDUA*, producing no detectable IDUA activity or protein (Section 4.4.1b) and thus the immunochemical findings in patient 2827 fibroblasts must result from the $D_{349}N$ allele. The D_{349} residue is conserved between human and canine IDUA (Fig. 3.29) and $D_{349}N$ is a structurally conservative amino acid substitution of an acidic amino acid (Asp) for a neutral-polar amino acid with the same structure except for a terminal amide group in place of a carboxyl group (Asn). The negatively charged carboxyl group of aspartic acid (Asp) residues have been implicated in active site catalysis, acting by hydrolysis of the glycosidic bond between sugar residues of other polysaccharide degrading enzymes (e.g. lysozyme, Stryer, 1981; Muraki *et al.*, 1991; α-glucosidase, Hermans *et al.*, 1991; β-glucosidase, Grace *et al.*, 1990). It is possible that the D_{349} residue of human IDUA plays a similar role in the

FIGURE 4.14. CHEMICAL CLEAVAGE OF PATIENTS 2662 AND 2827.

Chemical cleavage of the ID79 to ID89 PCR products from patients 2662 and 2827 are shown on a 5% (w/v) denaturing polyacrylamide gel. The homoduplex control is shown in the first set of lanes and the order of the lanes for each set of reactions is: unreacted probes (P79 and P89), hydroxylamine treated reactions (H) and osmium tetroxide treated reactions (OT). The probes were combined for cleavage in the homoduplex control. The arrows indicate the position of the many cleavage products detected and are numbered on patient 2827 reactions according to Table 4.11. All cleavages shown here are polymorphisms except for the unique cleavage in patient 2827 marked by an asterisk which corresponds to $D_{349}N$. Many larger cleavage products in these heteroduplexes were resolved on a 3.5% (w/v) gel.


Normal
P79 P89 H OT
2662
P79 P89
P79 P89 H OT H OT
2827
P79 P89
P79 P89 H OT H OT
8
9
10
*
7
11
6
12
5

hydrolysis of the glycosidic bond between the uronic acid and glucosamine residues of heparan sulphate, or the iduronic acid and glucuronic acid residues of dermatan sulphate. This hypothesis will be strengthened if it is found that the $D_{349}N$ mutant IDUA protein, or other IDUA proteins mutated at position 349, are still capable of binding substrate. As the other 2827 MPS-I allele is the null allele, $Q_{70}X$, it should be possible to perform experiments to determine substrate binding on 2827 patient fibroblasts or on $D_{349}N$ IDUA produced from *in vitro* expression systems. Thus, the interest in the $D_{349}N$ mutation is not due to its association with a severe clinical phenotype, but in the possibility that the D_{349} residue is involved in the active site and/or post translational processing of IDUA. This will be tested in the future by *in vitro* expression studies.

A total of 18 single base substitution polymorphisms have been detected in *IDUA*, including 14 polymorphisms for which patients 2662 and 2827 were both heterozygous. Patient 2827 is also heterozygous for the *Nsp*(7524)I and *Kpn* I polymorphism while patient 2662 is heterozygous for the *Eco*47III polymorphism. It is possible that the 14 polymorphisms common to patients 2662 and 2827 are on one chromosome in each patient and thus derived from an ancestral chromosome containing these polymorphisms. As patients 2662 and 2827 do not share an MPS-I mutation, two separate mutational events must have occurred to this ancestral chromosome. Only one of these 14 polymorphisms, $g_{3025}t$ in intron 10, was observed in other patients as cleavage products. Although the remaining polymorphisms may not have been detected in the other patients, it seems likely that most of these polymorphisms are rare, as evidenced by the finding of $A_{361}T$ in only 2 of 40 alleles. A list of all polymorphisms detected within *IDUA* is presented in Table 4.11. Surprisingly, nine of the polymorphisms are in the coding region of IDUA with two of these changes altering amino acid residues that are not conserved between the human and canine IDUA proteins (Fig. 3.29). Nonpathogenic IDUA variants causing low levels of IDUA activity have been reported (Gatti *et al.*, 1985; Whitley *et al.*, 1987). It may be interesting to determine whether either of the two polymorphisms that change amino acids, $A_{361}T$ and $Q_{33}H$, alter IDUA activity in fibroblasts from individuals who are

TABLE 4.11. A LIST OF ALL POLYMORPHISMS DETECTED IN *IDUA*.

The frequency and position of polymorphisms in *IDUA* as numbered on the finalized genomic sequence in Figures 3.24 and 3.25 and cDNA sequence as in Fig. 3.22 are given. The patients that the polymorphisms were detected in are also shown. Polymorphism number 3 was detected as a sequence difference between the clones p157.9 and λID475 (see Section 3.7.2). Restriction enzymes that could be used to detect the polymorphisms are indicated. If no convenient restriction sites were available then the polymorphisms were (or would have to be) detected by ASOs. ND = not done.

TABLE 4.11.

	Base Gen/cDNA	Amino acid	Exon/ Intron	Patient(s)	frequency in normals	detection
1	C->A 606/112	A_8	I	283/296/728/ 2474/2662	0.31	*Eco*47III or ASO's 143 & 144
2	G->T 681/187	$Q_{33}H$	I	296/1062/2827	0.09	*Nsp*(7524)I
3	g->a 255	-	intron 2	p157.9	ND	ASO
4	T->C 479/490	L_{118}	III	296/2827	0.22	*Kpn* I
5	T->C 1332/631	N_{181}	V	2662/2827	ND	ASO
6	g->c 1449	-	intron 5	2662/2827	ND	*Fnu*4HI
7	c->t 1486	-	intron 5	2662/2827	0.15	*Nci* I or ASO's 140 & 141.
8	G->C 1946/1030	A_{314}	VII	2662/2827	ND	ASO
9	a->g 2024	-	intron 7	2662/2827	ND	*Sma* I
10	g->c 2039?	-	intron 7	2662/2827	ND	*Cfr*10I
11	G->A 2192/1169	$A_{361}T$	VIII	2827/2662	0.05	ASO 121 & 122
12	G->C 2275/1252	T_{388}	VIII	2662/2827	ND	*Eae* I
13	C->G 2587/1318	T_{410}	IX	2662/2827	ND	ASO
14	t->c 2795	-	intron 9	2662/2827	ND	*Nsp*BII or ASO
15	C->T 2915/1555	R_{489}	X	2662/2827	ND	*Nar* I, *Bbe* I, or *Nun* II
16	g->t 3013	-	intron 10	2662/2827	ND	*Ppu*MI
17	g->t 3025	-	intron 10	728/2474 2662/2827	ND	ASO
18	t->c 3122		intron 10	2662/2827	ND	ASO

Exons XIII and XIV - no mismatches detected by CC.

homozygous for these polymorphisms compared to normals. One of the polymorphisms, T->C (N_{181}), was previously observed in the full-length IDUA cDNA construct, IDX21 (Section 3.8.2). It is interesting to note that neither of the two base changes in the full-length IDUA insert, IDX21, that alter amino acid residues were detected as polymorphisms, suggesting that they are most probably errors introduced by Taq polymerase during replication of the IDUA cDNA in the PCR reaction. It is also noted that no polymorphisms were detected by CC in exons XIII and XIV of *IDUA*.

4.4.2e DETECTION OF $g_{1487}a$

As described in Section 4.1, patient 2474 is an index Scheie syndrome case. One allele for this patient had been designated $W_{402}X$ and, as discussed in Section 4.4.1a, the other MPS-I allele must be responsible for the very mild phenotype observed in this patient. A unique CC product of 971-bp was obtained from primer ID89 in the hydroxylamine reaction, close to the 3' boundary of exon V of *IDUA* in patient 2474. Direct sequencing of the PCR products with the primers ID17 and ID23 revealed that the observed cleavage resulted from a single base substitution of g->a which altered the 3' consensus splice site, or splice-acceptor site of intron 5 of IDUA (base position 1487, Fig. 3.24B; $g_{1487}a$, see Fig. 4.15) and thus could possibly be a mutation affecting the splicing of the IDUA precursor RNA to mRNA.

It was considered important that the molecular pathology of this potential mutation was followed, not for identification of abnormally spliced mRNAs, but to understand why this mutation may produce a mild phenotype. This could help define the lower limit of IDUA activity necessary for a relatively normal phenotype. The following methods were used to test the hypothesis that the $g_{1487}a$ substitution was a splicing mutation.

1) Population studies were undertaken to show that the base substitution was not present in the normal population.

2) The production of mRNAs from both alleles, at an area unaffected by the proposed splicing mutation, was tested by using polymorphisms for which

patient 2474 was heterozygous to distinguish between the two different mRNA molecules.

3) Following this, the production of only one mRNA from the $W_{402}X$ allele was tested by PCR with a primer spanning the affected exon junction, again using polymorphisms to distinguish the alleles.

4) The production of abnormally spliced mRNA was tested by PCR across the affected exon junction, or by PCR at some other region in the IDUA mRNA.

Population studies for the $g_{1487}a$ substitution were complicated by the presence of a c->t substitution immediately 5' to the $g_{1487}a$ substitution at position 1486 of the genomic sequence (Fig. 3.24B). This base substitution was thought to be polymorphic as it was detected in two other patients with defined genotypes (2662 and 2827) and did not disrupt the consensus sequence for a 3' intron splice site (Table 3.5). Thus, a set of four ASOs was required for the detection of the $g_{1487}a$ and $t_{1486}c$ substitutions. PCR analysis with ASOs (Table 4.4) of 116 chromosomes from 62 normal individuals revealed the $g_{1487}a$ substitution was not present in normal chromosomes and was present in 3 of the 146 MPS-I chromosomes analysed. The $t_{1486}c$ substitution was indeed polymorphic and occurred with a frequency of 15% (18/124) and 8% (2/26) in normal and MPS-I chromosomes respectively. This is good preliminary evidence to indicate that $g_{1487}a$ is indeed a mutation. Two patients with a genotype of $W_{402}X/g_{1487}a$ have been detected (2474 and 32-1), and all three patients with the $g_{1487}a$ mutation have a mild phenotype. A detailed clinical description of patient 2474 has been given (Section 1.3.2; McKusick *et al.*, 1965; McKusick, 1972; McKusick and Neufeld, 1983) and is the mild patient shown in Figure 1.8. Patient 2474 was 54 yrs old at the time of the picture and was a practising attorney. Patient 2474 had a sister who was also affected with MPS-I, with a similar or slightly more severe clinical phenotype (McKusick, 1972). Patient 32-1 (from the Willink Biochemical Genetics Unit, Royal Manchester Children's Hospital) is 47 years old with normal intelligence and is employed as an office worker. She is of normal height and has mild joint stiffness, mainly in the large joints. Surgical relief of bilateral carpal tunnel syndrome was

performed at 43 years of age and at 46 years of age the patient had a corneal graft in one eye due to corneal clouding. Echocardiography shows mild thickening of the aortic and mitral valves with no evidence of cardiac failure.

Points 2, 3 and 4 described earlier were tested by experiments shown in Fig. 4.15. Patient 2474 is heterozygous for the *Eco*47III polymorphism in exon I of IDUA. Thus PCR between the primers ID60 and ID68 was performed from cDNA to amplify exons I and II of IDUA to see whether or not normally spliced IDUA mRNA was produced from either allele. The PCR products were analysed by Southern blotting and hybridization of ASOs to the *Eco*47III polymorphism (Table 4.3). This experiment showed that stable mRNA molecules of normal size were produced in approximately equal amounts from both alleles of patient 2474 in the regions of exon I and II of the IDUA mRNA (data not shown).

An oligonucleotide primer, ID123, which spanned the exon V/VI junction that was hypothesized to be affected by the $g_{1487}a$ substitution in patient 2474 was used with ID60 to test whether correctly spliced mRNA is produced from both alleles at this junction (Fig. 4.15). By digesting the PCR product with *Eco*47III, it has been shown that, as hypothesized, only one message that did not contain the *Eco*47III restriction site (allele 1) was amplified to detectable levels (data not shown). Allele 1 of the *Eco*47III polymorphism is normally associated with the $W_{402}X$ mutation as described in Section 4.4.1a. This shows that the $g_{1487}a$ base substitution does seriously affect the IDUA mRNA splicing at the intron 5/ exon VI boundary as no normal mRNA is produced at the exon V/VI junction by the $g_{1487}a$ allele within the detection limits of this assay (estimated at 10%). The $g_{1487}a$ substitution is therefore a mutation.

Fibroblasts from patient 2474 had no immunodetectable IDUA protein, but IDUA activity towards a radiolabelled disaccharide substrate has a normal K_m of 33 μM (Table 4.2). Any IDUA protein that may be present in patient 2474 fibroblasts is below the levels of the immunodetection assay and thus a V_{max} for this IDUA cannot be determined. This patient therefore represents an exception to the finding that MPS-I patients with mild or intermediate phenotypes have detectable IDUA protein (Ashton *et*

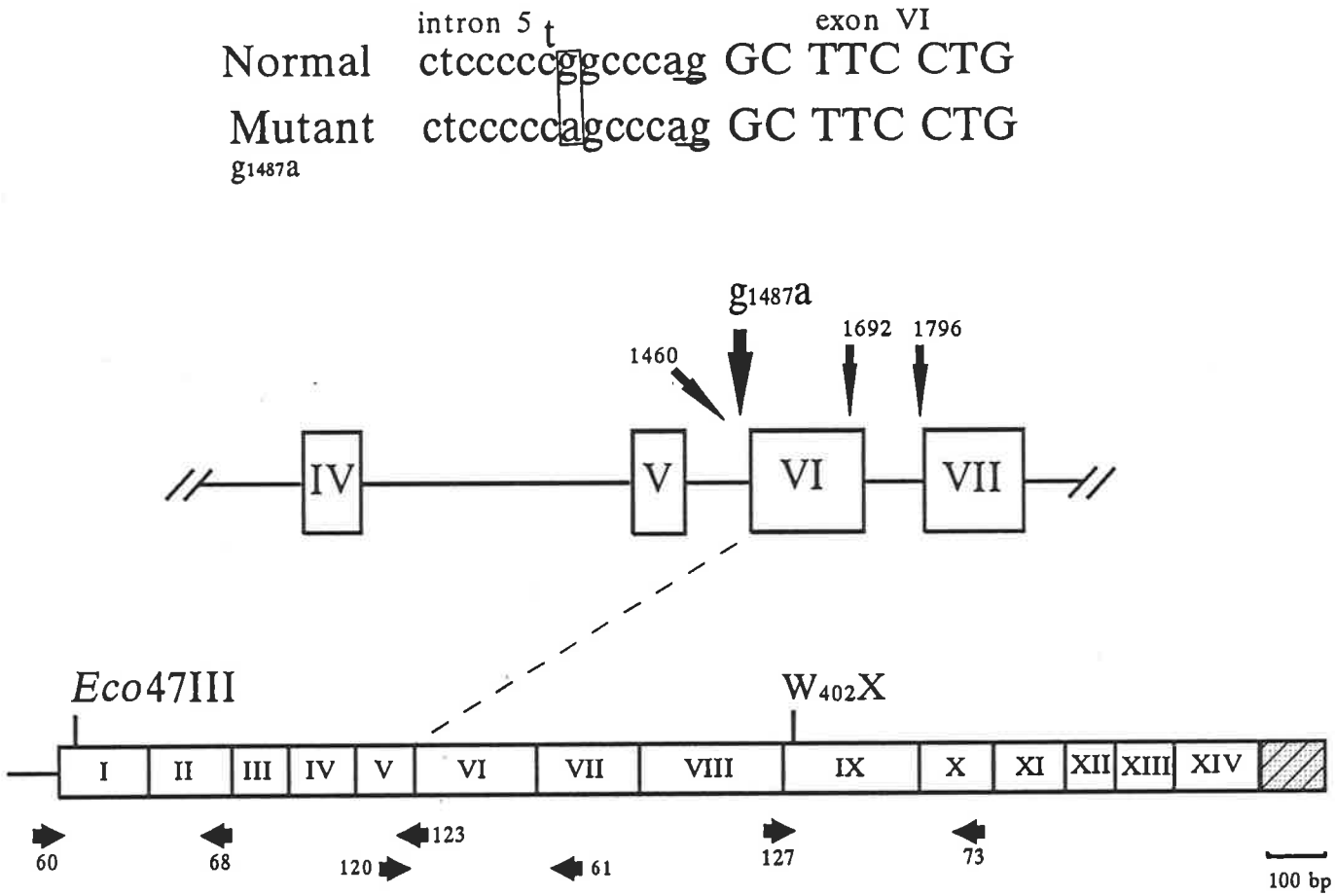
al., 1992). The genotype of patient 2474 is $W_{402}X/g_{1487}a$ and since $W_{402}X$ can be considered a null allele, the $g_{1487}a$ allele must be responsible for the extremely mild phenotype. The $g_{1487}a$ allele must produce enough IDUA, either mutant or normal, with sufficient ability to bind and turnover substrate to confer a mild phenotype to patient 2474. This IDUA protein may be produced from the $g_{1487}a$ allele in several ways as described in the following paragraphs.

It was thought that the possible mode of action of the $g_{1487}a$ mutation at the 3' intron 5 boundary was the creation of an alternative splice site at the base substitution that would be utilized in preference to the correct splice site (Fig. 4.15). If this mutation still allowed a small amount of correct splicing to occur, and thus the production of a small amount of normal IDUA protein, it could account for the extremely mild phenotype observed in this patient. If this hypothesis was correct, an additional 5-bp would be present in the mRNA causing a frameshift in the coding sequence from codon 196, and the addition of two incorrect amino acids on the carboxyl-terminal-end before the premature termination of the IDUA protein. To test this hypothesis, PCR was performed on cDNA from 2474 patient fibroblasts with the oligonucleotide primers ID61 and ID120 which spans the exon V/VI splice junction. The resulting PCR product should contain the correctly spliced mRNA from the $W_{402}X$ allele and an mRNA incorporating the 5 additional nucleotides between the proposed alternative splice site and the correct splice site. This PCR product was Southern blotted and hybridized with ASOs that would detect the two RNA species described above (ID123 and ID124, see Appendix E). Only the normal mRNA, presumably from the $W_{402}X$ allele, was detected indicating that this hypothesis was incorrect and use of this cryptic splice site, if any, must be below the detection limits of ASOs (estimated at 1%).

Mutant, abnormally spliced IDUA mRNA molecules may produce a partially active IDUA protein that is unrecognisable by the two monoclonal antibodies used in the immunochemical assay of Ashton *et al*, (1992). The utilization of cryptic splice sites when mutations affect 3' splice sites has been reported (Steingrimsdottir *et al.*, 1992; Hedley *et al.*, 1989) and it has been shown that the polypyrimidine tract (C/T) 5'

FIGURE 4.15. EXPERIMENTAL DESIGN USED TO TEST IF $g_{1487}a$ IS A SPLICING MUTATION.

The sequence of the normal and mutant intron 5 3' splice sites of *IDUA* are shown. The base substitution is boxed. Arrows indicate the position of the mutation in the gene, the best three cryptic splice sites and the position of primers on the IDUA mRNA used to study the molecular nature of the mutation, as described in Section 4.4.2e are shown. The primer, ID123 spans the exon V/VI exon boundary.


intron 5
t
exon VI
Normal ctcccccggcccag GC TTC CTG
Mutant ctcccccagcccag GC TTC CTG
g1487a
g1487a
1460
1692
1796
IV
V
VI
VII
Eco47III
W402X
I
II
III
IV
V
VI
VII
VIII
IX
X
XI
XII
XIII
XIV
60
68
120
123
61
127
73
100 bp

to the invariant AG dinucleotide of splice-acceptor sites is vital to efficient mRNA splicing (Huang and Van der Ploeg, 1991). The best three cryptic splice sites that may be utilized by the $g_{1487}a$ mRNA as plotted by the DNASCAN program are at base positions 1460, 1692 and 1796 (Fig. 3.24B) and are indicated in Fig. 4.15. The cryptic 3' splice site at position 1692 is very close to the end of exon VI and would therefore not enable the correct splicing the next intron, intron 7, from the IDUA mRNA. Previously, an unusual pattern of IDUA PCR products in patient 2474 had been observed between the primers ID73 and ID127 which had been used to study the expression of $W_{402}X$ in mRNA. ID127 spans the exon VIII/ IX boundary, and thus only RNA, and not genomic DNA derived products, can be amplified. If $g_{1487}a$ is responsible for the generation of abnormally spliced mRNA then these unusual PCR products would all be expected to hybridize to the normal $W_{402}X$ ASO, indicating that they are derived from the $g_{1487}a$ allele, however this experiment is yet to be performed. Utilization of the cryptic splice site at position 1796 could possibly result in the patterns observed in the ID127 to ID73 PCR described above and would result in frameshifts and premature termination of the IDUA protein.

The cryptic splice site at base position 1796 is the intron 7 3' splice site and thus utilization of this site would result in "exon skipping" and the exclusion of exon VI from most IDUA mRNA molecules. It is possible that "exon skipping" could occur as has been reported for a number of other mutations to 3' splice sites or splice-acceptor sites (Steingrimsdottir *et al.*, 1992). If mRNA is produced by skipping exon VI, this would cause a frameshift mutation from codon 197, well before the hypothesized catalytic D_{349} residue, and the addition of 132 incorrect amino acids on the carboxyl terminal end before the premature termination of the IDUA protein. If this hypothesis is correct, the ID120 to ID61 PCR (Fig. 4.15) would produce a 280-bp product from the $W_{402}X$ allele and a 77-bp product that excluded the 203-bp exon VI from the $g_{1487}a$ allele. The expected 280-bp product and a number of larger PCR products were visible, implying that at least some mRNA splicing was incorrect at the exon V and VI boundary (data not shown). Use of either of these two cryptic splice sites presented

above (positions 1692 and 1796) that are 3' to the normal 3' splice site of intron 5, is in conflict with the larger, and not smaller, ID120 to 61 PCR products that were observed.

The ID120 to ID61 PCR products that are larger than the normal product can be explained by the use of cryptic splice sites 5' to the mutation, or by the affected intron remaining in the mRNA. Examples of both of these occurrences have been reported for 3' splice site mutations in other genes (Nelson *et al.*, 1992 and Patterson *et al.*, 1992 respectively). Utilization of the cryptic splice site at position 1460, which is 5' to the normal intron 5 splice site, would result in the inclusion of 33-bp from the 3' end of intron 5 into the IDUA mRNA. This would result in an ID120 to ID61 PCR product 33-bp larger than the normal 280-bp product consistent with the result obtained and with the size of one of the major abnormal PCR products. The translation product resulting from this abnormally spliced mRNA would have the correct IDUA amino acid sequence for 197 amino acids, followed by a valine residue and then the premature termination of the IDUA protein by a stop codon introduced into the IDUA mRNA by the 33-bp insertion.

Translation of an IDUA mRNA with the inclusion of intron 5 would again result in the correct IDUA amino acid sequence for 197 amino acids, and then the addition of 238 amino acids before the premature termination of the IDUA protein. Again, a major ID120 to ID61 PCR product that corresponds to the expected size of an IDUA mRNA including intron 5 was observed. However, it is impossible to determine whether this is the result of amplification of mRNA or contaminating genomic DNA. To determine if an mRNA molecule including intron 5 was produced, it would be possible to perform a PCR using a primer that spans an exon junction in the IDUA mRNA to ensure that only mRNA is amplified.

Thus, from the evidence presented above, it appears that the $g_{1487}a$ base substitution destroys the 3' splice boundary of intron 5 and that splicing for the rest of the IDUA mRNA is disturbed. This is probably due to the inefficient utilization of a number of cryptic splice sites 5' to $g_{1487}a$, which would alter the formation and position of branch sites needed for correct splicing of all exons 3' to $g_{1487}a$. As described above,

none of the abnormal IDUA mRNA molecules that are likely to be produced by the $g_{1487}a$ allele would result in active IDUA protein and thus the mild phenotype observed in patients with this mutation. The total IDUA activity measured with the fluorogenic IDUA substrate in fibroblasts from patient 2474 (0.3% compared to normal) is similar to that seen in some $W_{402}X$ homozygotes and $W_{402}X/Q_{70}X$ compound heterozygotes (see cell line numbers 538 and 1805, Ashton *et al.*, 1992) and thus must be a background level of activity. Although the IDUA activity levels in 2474 fibroblasts is extremely low, it seems to be slightly higher than the background levels in Hurler syndrome patients and has a consistently normal K_m, unlike Hurler syndrome fibroblasts with similar levels of IDUA activity, if measured with radiolabelled disaccharide substrates derived from heparin or dermatan sulphate (Hopwood and Muller, 1979; Muller and Hopwood, 1984; Ashton *et al.*, 1992). This implies that a small amount of IDUA protein with normal activity is present in 2474 fibroblasts. The sensitivity of the immunodetection assay has been calculated at 10pg of IDUA, which would give approximately 0.1% of normal IDUA activity. If normal IDUA protein is present in 2474 fibroblasts, it is at, or below, the lower detection limits of the immunodetection assay.

As all three patients with the $g_{1487}a$ have similarly mild phenotypes, it is considered unlikely that other genetic or environmental factors are responsible for the clinical phenotype observed. Thus, as hypothesized for the HPRT gene 3' splice site mutations causal in Lesch-Nyhan syndrome or gouty arthritis (Rossi *et al.*, 1990) and shown for a 3' splice mutation in *Urd*$^-$A CHO-K1 cells for the multifunctional CAD gene (Patterson *et al.*, 1992), it is hypothesized that a small amount of normal IDUA mRNA is being produced from the $g_{1487}a$ allele. This normal IDUA mRNA must produce a small amount of normal IDUA protein, undetectable by immunochemical assay, conferring on these patients an extremely mild clinical phenotype. If this hypothesis is correct it would show, not surprisingly, that a very small amount of normal IDUA protein in all cells is more effective than a slightly larger amount of mutant IDUA protein such as $R_{89}Q$. Future experiments will include a Southern blot of the ID123 to ID60 PCR products for ASO detection of the *Eco*47III polymorphism (Fig.

4.15). This should enable more sensitive detection of allele 2 of the *Eco*47III polymorphism, which must be on the same chromosome as $g_{1487}a$ (Table 4.12), to show production of normal IDUA mRNA. However, it is possible that the hypothesized normal splicing is below the detection limits of ASOs and S1 nuclease assays, and thus a complete explanation for the molecular pathogenesis of this mutation will remain speculative.

Haplotype analysis of patients with the $g_{1487}a$ allele reveals that it is associated with the 1,2, VNTR, *Kpn* I haplotype. If the two $g_{1487}a$ alleles that were detected in patients included in the linkage disequilibrium studies (Section 4.3.2), are subtracted from the uncharacterized MPS-I chromosomes remaining after subtraction of $W_{402}X$ and $Q_{70}X$ from the haplotype analysis (Table 4.10), there is a decrease in the probability that another common MPS-I mutation exists as discussed in Section 4.4.1b (χ^2=0.28, 0.6<P<0.7, 1 *df*). The haplotype analysis of patients with a $g_{1487}a$ allele also shows a genetic novelty as demonstrated below. One of the MPS-I patients with a $g_{1487}a$ allele is married and has a daughter with MPS-I. One of the daughter's alleles is defined as $P_{533}R$ which she inherited from her mother. The father's and thus the daughter's second MPS-I allele is undefined meaning that accurate phenotype prediction cannot be attempted for the daughter. However, as most MPS-I patients and thus alleles cause a severe phenotype, it is likely that the father's second MPS-I allele is a severe allele, and thus the daughter would have a severe prognosis consistent with her clinical description (Dr D. Wenger, pers. commun.).

	VNTR	*Kpn* I	Mutation
Father, MPS-I.			
	1	2	$g_{1487}a$
	2	1	?
Mother, carrier.			
	2	1	$P_{533}R$
	1	2	normal
Daughter, MPS-I.			
	2	1	$P_{533}R$
	2	1	?

4.4.2f DETECTION OF $t_{1978}c$

As described in Section 4.4.2b, the $R_{89}Q$ MPS-I allele had been detected in patient 2662 and many other CC products common to patients 2662 and 2827 were also observed. However, a unique CC occurred close to the 5' boundary of exon VII of *IDUA* in patient 2662. A cleavage product of 782-bp was obtained from primer ID79 in the hydroxylamine reaction (data not shown). Direct sequencing of the PCR products with the primers ID18 and ID25 revealed that the observed cleavage resulted from a single base substitution of t->c (base position 1978, Fig. 3.24B) that altered the rigidly conserved gt dinucleotide of the 5' splice site or the splice-donor site of intron 7 of *IDUA* to gc ($t_{1978}c$), and thus could possibly be a mutation that affects the splicing of the IDUA precursor RNA to mRNA (for consensus splice site see Table 3.5). It was attempted to show that $t_{1978}c$ was a splicing mutation by the three criteria listed in Section 4.4.2e.

PCR analysis with ASOs (Table 4.4) of 42 chromosomes from 21 normal individuals revealed that the $t_{1978}c$ substitution was not present in normal chromosomes and was unique in the 73 MPS-I patients analysed.

The next experiments were similar to those performed for the $g_{1487}a$ mutation and were designed to test points 2, 3 and 4 as listed in Section 4.4.2e as shown in Fig. 4.16. Patient 2662 is heterozygous for the *Eco*47III polymorphism in exon I of IDUA. Thus, as for patient 2474, PCR between the primers ID60 and ID68 was performed from cDNA to amplify exons I and II of IDUA to observe whether normally spliced IDUA mRNA was produced from either allele. The PCR products were analysed by Southern blotting and hybridization of ASOs to the *Eco*47III polymorphism (Table 4.3). This experiment showed that stable mRNA molecules of normal size were produced in approximately equal amounts from both alleles of patient 2662 in the regions of exon I and II of the IDUA mRNA (data not shown).

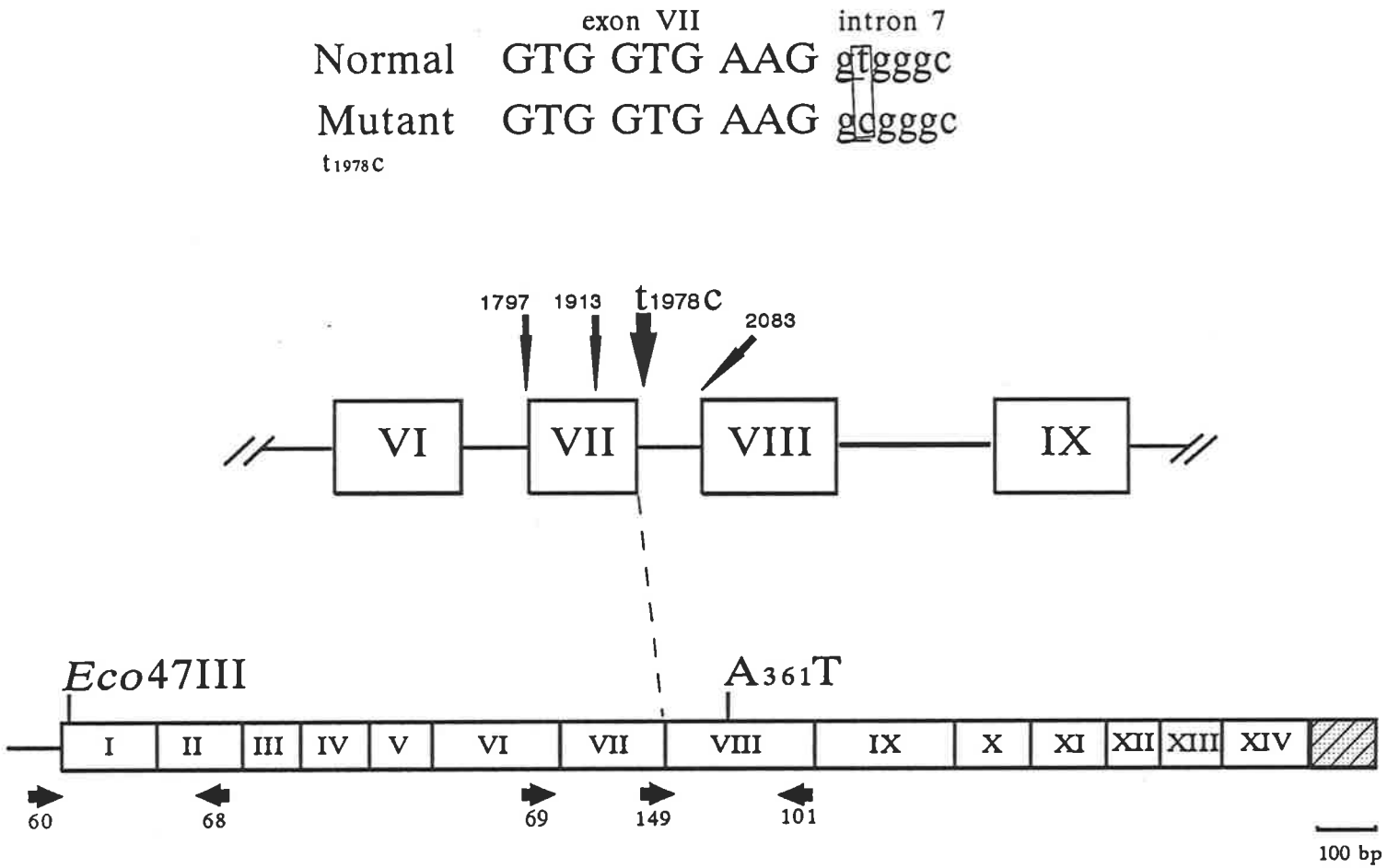
Patient 2662 is also heterozygous for the $A_{361}T$ polymorphism. An oligonucleotide primer, ID149, was synthesized to span the exon VII/ VIII splice

junction that would be affected by the $t_{1978}c$ base substitution at the 5' splice site of intron 7 (Fig. 4.16). As for the ID123 to ID60 PCR for $g_{1487}a$, PCR products from patient 2662 mRNA between ID149 and ID101 are expected to amplify only one mRNA molecule as detected by hybridization with ASOs for the $A_{361}T$ polymorphism. This would show that the $t_{1978}c$ base substitution is a mutation that prevents the correct splicing of the IDUA mRNA at the exon VII/ intron 7 boundary. In order to demonstrate abnormally spliced mRNA in patient 2662, PCR reactions were attempted between the primers ID69 and ID101 (Fig. 4.16), however preferential amplification of genomic DNA contaminating the 2662 RNA preparation prevented the observation of mRNA allele specific amplification or amplification of abnormally spliced mRNAs.

A large number of mutations affecting 5' splice sites leading to abnormal mRNA splicing and thus human genetic diseases have now been described (for review see Sakuraba *et al.*, 1992; Talerico and Berget, 1990; Robberson *et al.*, 1990). Of 5' splice site mutations, 77% result in exon skipping or a combination of exon skipping and use of cryptic splice sites. Inclusion of intron 7 in the IDUA mRNA from the $t_{1978}c$ allele would result in a frameshift in the coding sequence from codon 324, and the addition of 57 incorrect amino acids on the carboxy terminal end, before the premature termination of the IDUA protein. The best three 5' cryptic splice sites that may be utilized by the $t_{1978}c$ mRNA as plotted by the DNASCAN program are at base positions 1797, 1913 and 2083 (Fig. 3.24B). The cryptic 5' splice site at position 1797 is in fact the 3' splice site of intron 6 and is unlikely to be utilized as a 5' splice site as this would also require the use of a cryptic 3' splice site within intron 6. This is a complicated and unlikely scenario. The use of the cryptic splice site at position 1913 would result in an IDUA protein that is identical to normal IDUA except for the exclusion of 21 amino acids from position 304 to 325 of the normal IDUA protein (Fig. 3.22). The cryptic 5' splice site at position 2083 is in fact the 3' splice site of intron 7 and is unlikely to be utilized as a 5' splice site as this would also require the use of a cryptic 3' splice site within exon 7. However, it is likely that this site is still used as a 3' splice site and thus the $t_{1978}c$ mutation may cause skipping of the 180-bp exon VII from the IDUA mRNA. As with most other mutations in the rigidly conserved gt dinucleotide of 5' splice-donor sites, it

FIGURE 4.16. EXPERIMENTAL DESIGN USED TO TEST IF $t_{1978}c$ IS A SPLICING MUTATION.

The sequence of the normal and mutant intron 7 5' splice sites of *IDUA* are shown. The base substitution is boxed. Arrows indicate the position of the mutation in the gene, the best three cryptic splice sites and the position of primers on the IDUA mRNA used to study the molecular nature of the mutation, as described in Section 4.4.2f are shown. The primer, ID149 spans the exon V/VI exon boundary.


exon VII
intron 7
Normal GTG GTG AAG gtgggc
Mutant GTG GTG AAG ggggggc
t1978C
1797
1913
t1978C
2083
VI
VII
VIII
IX
Eco47III
A361T
I
II
III
IV
V
VI
VII
VIII
IX
X
XI
XII
XIII
XIV
60
68
69
149
101
100 bp

is expected that no normally spliced IDUA mRNA would be produced from the $t_{1978}c$ allele. In addition, consistent with another gt to gc 5' splice site mutation that was reported by Weil *et al.* (1988), it is expected that 100% of mRNA from the $t_{1978}c$ allele will result in skipping of exon VII. Again, this hypothesis was tested by the ID69 to ID101 PCR (Fig. 4.16) but failed, again due to the preferential amplification of genomic DNA contaminating the 2662 RNA preparation. An mRNA produced by skipping exon VII would result in an IDUA protein that is identical to normal IDUA except for the exclusion of 60 amino acids from position 265 to 325 of the normal IDUA protein (Fig. 3.22).

Thus the final genotype of patient 2662 is $R_{89}Q/t_{1978}c$. The clinical description of patient 2662 is complicated by the addition of two younger sisters who presumably have the same $R_{89}Q/t_{1978}c$ genotype. The youngest sister presented with failure to thrive at 10 months of age and was given the diagnosis if MPS-III on the basis of mucopolysaccharidurie and X-ray radiology. A shunt was performed to relieve hydrocephalus but the patient died at only 22 months of age from bronch-opneumonia. Patient 2662 was 2 years of age at the time of diagnosis of his younger sister and was noted as being well. However, patient 2662 first presented to his local doctor at 7 months of age with a large head (>97th percentile), mouth breathing and snoring. In all other aspects he was reported as normal. At 3 years of age, he presented again with a persistent runny nose, mouth breathing, snoring and poor speech. At 4 years and 9 months of age his health was recorded as good despite middle ear infections, joint stiffness, enlarged liver and spleen, hearing loss and radiology suggested to be consistent with a diagnosis of MPS-III. No definite abnormal neurological signs were found. Patient 2662 attended special schools until 15 years of age. His motor development has been limited by increasing severe joint contractures. He has had recurrent ear infections and has not attained the ability to read and write. At 18 years of age, he presented with fixed flexion deformities on his fingers and an increasingly poor memory. Measurement of IDUA activity in leukocytes and fibroblasts from patient 2662 corrected the clinical diagnosis from MPS-III to MPS-I. He is now over 20 years old with a short stature (151 cm), obesity (71 kg) and a large head (circumference of 60

cm). He has coarse facial features, hearing and visual problems, and multiple joint contractures that severely limit his movement. His second sister presented at 2 years and 6 months of age with good general health apart from upper respiratory tract infection, coarse facial features, enlarged liver and spleen, and limitation of elbow extension. This sibling is now 20 years of age, has coarse facial features, joint limitations, is obese and has a short stature. She is reported as bright and able to live in her own flat with visiting help.

As discussed in Section 4.4.2b, immunochemical analysis of a $R_{89}Q/W_{402}X$ compound heterozygote detected 1.5% of the normal levels of IDUA protein with two different monoclonal antibodies, 0.5% of normal IDUA activity, and an increased K_m (333 μM compared with 12-111 μM in normal control fibroblasts, cell line number 2730, Ashton *et al.*, 1992). Immunochemical analysis of patient 2662 fibroblasts detected 7% to 10% of the normal levels of IDUA protein with the monoclonal antibodies Id1A and Id17A respectively, 3% of the normal IDUA activity, and a K_m of 18 μM which is at the lower end of the normal range (Table 4.2 and Ashton *et al.*, 1992). In addition, low levels of radiolabel were incorporated into IDUA in fibroblasts from patient 2662 (Taylor *et al.*, 1991). These two sets of biochemical results appear to be significantly different and could lead to the expectation that the clinical phenotype of the $R_{89}Q/t_{1978}c$ compound heterozygotes should be milder than that of the $R_{89}Q/W_{402}X$ compound heterozygotes, in contrast to the clinical data presented in this section and section 4.4.2b respectively. As speculated above, it is expected that IDUA protein from the $t_{1978}c$ allele would miss the 60 amino acids encoded by exon VII. Although no *N*-glycosylation sites are encoded in exon VII, it is possible that the $t_{1978}c$ protein is not transported to the lysosome, and thus does not help degrade substrate which could contribute to a milder clinical phenotype. The only conclusions that can be drawn from the above data are that the IDUA protein being produced from the $t_{1978}c$ allele has a slightly reduced stability and ability to turnover substrate.

From the above descriptions, it would appear that these $R_{89}Q/t_{1978}c$ patients are more clinically severe than the two $W_{402}X/R_{89}Q$ compound heterozygous patients

(Section 4.4.2b). As observed in other LSD, mutations such as $W_{402}X$ or $Q_{70}X$ that result in total loss of function consistently cause a severe clinical phenotype with comparatively little phenotypic variation. However, mutations which permit some enzymatic function, such as $R_{89}Q$ and presumably $t_{1978}c$, normally result in a milder disease phenotype but may be more susceptible to modulation by variables such as novel combinations of mutant alleles, treatment received (e.g. surgical procedures and drugs prescribed) and environmental factors (e.g. infection) (reviewed in Neufeld, 1991). Caution should always be taken in predicting a patients clinical outcome from their genotype, especially with the occurrence of "mild" alleles or a novel combination of alleles. From these results, it is difficult to predict the clinical outcome or the enzymatic function of mutant IDUA protein resulting from the $t_{1978}c$ mutation.

4.4.3 HAPLOTYPING AND MPS-I MUTATIONS

As expected from the linkage disequilibrium studies, $W_{402}X$ and $Q_{70}X$ seem to be rigidly associated with the 2,2 and 2,1 haplotypes respectively (Section 4.3.2). However, expanded haplotyping of ten $W_{402}X$ homozygotes with the *Eco*47III and *Nsp*(7524)I polymorphisms resulted in the finding that $W_{402}X$ was associated with three different haplotypes. Sixteen $W_{402}X$ alleles were associated with the 1,1,2,2 haplotype (*Eco*47III, *Nsp*(7524)I, VNTR, *Kpn* I), 1 with 1,2,2,2 and 3 with 2,1,2,2. The MPS-I patient homozygous for $P_{533}R$ is also homozygous for the 2,2 (VNTR, *Kpn* I) haplotype. A patient heterozygous for $P_{533}R$ is homozygous for the 2,1 haplotype and thus $P_{533}R$ is seen associated with both the 2,2 haplotype and the 2,1 haplotype. $R_{89}Q$ is associated with the 2,2 haplotype and a $W_{402}X/R_{89}Q$ compound heterozygote is heterozygous at the *Kpn* I locus, implying that $R_{89}Q$ is also associated with two haplotypes or that $W_{402}X$ is associated with a fourth haplotype in addition to the three described above. This haplotype data implies that these mutations may have originated more than once, or that intragenic *IDUA* recombination events may have occurred. The $Q_{70}X$ and $g_{1487}a$ mutations are only seen associated with the 2,1 and 1,2 haplotypes respectively. Extended haplotype analysis on patients with a $Q_{70}X$ or $g_{1487}a$ allele may also reveal these mutations in association with more than one haplotype.

It has been reported that 35% of mutations in human genetic disease occur in CpG dinucleotides. Of these mutations, 90% are C->T or G->A transitions that are thought to result from methylation-induced deamination of the 5-methyl-cytosine residues in the CpG dinucleotide (Cooper and Youssoufian, 1988). Three of the mutations reported in this thesis could result from transitions in CpG dinucleotides ($R_{89}Q$, $G_{1414}A$; $g_{1487}a$; $D_{349}N$, $G_{2156}A$; nucleotide numbering according to genomic sequence in Fig. 3.24). $R_{89}Q$ and $g_{1487}a$ were observed more than once and, as discussed above, it is possible that $R_{89}Q$ is associated with two haplotypes. As these mutations result from CpG dinucleotide transitions, there is an increased possibility that they have occurred more than once. $P_{533}R$ is also seen associated with more than one haplotype, and occurs in a CpG dinucleotide but is not a transition (transversion, $C_{3233}G$). Thus 4 of the 8 mutations described in this thesis occur in CpG dinucleotides and 3 of these 4 are transitions. Of the 18 polymorphisms detected within IDUA (Table 4.11), 9 occur in CpG dinucleotides but only 4 of these 9 are transitions ($g_{255}a$; $c_{1486}t$; $A_{361}T$ is $G_{2192}A$; $t_{2795}c$).

Importantly, the common mutations $W_{402}X$ and $Q_{70}X$ do not occur in CpG dinucleotides and are unlikely to have occurred more than once. It is believed that the association of 3 mutations with at least 7 different haplotypes, indicates the occurrence of intragenic recombination in *IDUA*. If recombination has occurred, then the observation of at least 4 recombination events in such a small sample (28 alleles) implies a high frequency of recombination within *IDUA*. *IDUA* maps 1100 kb from the telomere of chromosome 4 at 4p16.3, 1000 kb from either of the two most likely locations for the gene involved in Huntington disease (MacDonald *et al.*, 1991b). Complex patterns of linkage disequilibrium have been observed in the Huntington disease gene region (MacDonald *et al.*, 1991a; Andrew *et al.*, 1992), and recently it has been found that Huntington disease chromosomes exhibit many different haplotypes in the Huntington gene candidate region (MacDonald *et al.*, 1992). This has suggested, in contrast to long-held dogma, that there are a variety of independent Huntington disease mutations. If several regions with a high frequency of recombination, comparable to

the apparent recombination frequency in *IDUA*, are present in the 4p16.3 chromosome region, they may be partially responsible for generating the multiple haplotypes observed on Huntington disease chromosomes, and thus there may still be a major Huntington disease mutation(s). The analysis of recombinant HD chromosomes has already defined a hot spot of recombination within an interval of 300-kb distal to the Huntington disease linked locus *D4S10* (Allitto *et al.*, 1991; Buetow *et al.*, 1991). This may also explain why it has been difficult to map the Huntington disease gene using recombination events in Huntington families and by linkage disequilibrium studies. The presence of many defined polymorphisms within *IDUA* (see following section) and the ascertainment of many more highly informative polymorphisms within the 4p16.3 region (Weber *et al.*, 1992) should enable this hypothesis to be tested, perhaps with the use of CEPH families or the large Huntington disease pedigrees available.

An additional significance of the finding of MPS-I mutations in association with more than one haplotype is an example of where caution should be practiced when using haplotyping in mutation analysis. Extended haplotyping had shown patients 728 and 1062 to be heterozygous at the *Eco*47III and *Nsp*(7524)I polymorphic loci respectively (Table 4.7), and thus it was assumed that normal DNA was not required for heteroduplex formation in CC analysis. In fact, $W_{402}X$ was not detected in these patients by CC analysis in this study although both patients are $W_{402}X$ homozygotes.

4.5 CONCLUDING DISCUSSION

All MPS-I alleles in the group of 9 patients originally selected for patient analysis have been defined by the use of chemical cleavage analysis and direct PCR sequencing. The final genotypes and deduced or hypothesized haplotype associations for the mutations in each patient is given in Table 4.12. Linkage disequilibrium, immunochemical and biochemical analysis have proved to be powerful tools for selecting MPS-I patients with common mutations or mutations of structural and functional interest. Eight mutations have been defined; 2 nonsense mutations, one single nucleotide deletion, 3 amino acid substitutions and 2 splicing mutations. The

position of the mutations in *IDUA* and the PCR reactions used to detect them are shown in Fig. 4.3. Both nonsense mutations, $W_{402}X$ and $Q_{70}X$, are associated with extremely severe MPS-I clinical phenotypes and together account for 48% of all MPS-I alleles in a group of 73 patients studied. The identification of these common mutations and their good genotype to phenotype correlations will allow the rapid assessment of BMT as a therapy protocol for MPS-I, and assessment of other therapy protocols as they become available. The identification of these common mutations will also aid in the selection of patients to undergo therapy and in genetic counselling for MPS-I families, including carrier detection. A unique frameshift mutation, $delG_{1702}$, is also associated with an extremely severe MPS-I clinical phenotype. Two of the three amino acid substitutions, $P_{533}R$ and $R_{89}Q$, were found more than once and are associated with moderately severe and mild to intermediate phenotypes respectively. The third amino acid substitution, $D_{349}N$, has marked the D_{349} residue as being potentially involved in active site catalysis and/or processing of IDUA. Of the two splicing mutations, $g_{1487}a$ was seen in three patients and is thought to be associated with a very mild clinical phenotype. Neither the biochemical or clinical outcome of $t_{1978}c$ splicing mutation is known. The exact molecular pathogenesis of both splicing mutations remains undefined. As expected, it is now apparent that there will be several explanations accounting for the wide range of clinical phenotypes found in MPS-I which is dependent on the combination of mutant alleles. Two types of mutations have already been defined which result in mild or intermediate phenotypes; an amino acid substitution, $R_{89}Q$, and the splicing mutation, $g_{1487}a$. The presence of other mutations such as $P_{533}R$ may result in a moderately severe phenotype. As expected for a recessive genetic disease, only one mild or intermediate allele is required for manifestation of a mild or intermediate phenotype. However, as illustrated in the case of the three $R_{89}Q/t_{1978}c$ sibs (Section 4.4.2f), a mild allele may be more susceptible to environmental influences that may affect the clinical phenotype. Haplotype analysis has been useful in deducing that 3 different MPS-I mutations were derived from up to 7 independent mutation events and/or that intragenic recombination has occurred, implying a high frequency of

TABLE 4.12. FINAL PATIENT HAPLOTYPES AND GENOTYPES.

The final mutant genotypes and deduced or hypothesized haplotype association of the mutations are given for all nine patients. Comments on the nature of the mutation or the haplotype associations are given.

TABLE 4.12.

Patient	Phenotype	*Eco*47III	*Nsp* I	VNTR	*Kpn* I	Mutation(s)	Comments
283	severe/	1	1	2	2	$W_{402}X$	
	intermediate	2	1	2	2	$P_{533}R$	
296	severe	1	2	2	1	$Q_{70}X$	Heterozygous at
		2	1	1	2	$delG_{1702}$	VNTR and *Kpn* I
728	severe	1	1	2	2	$W_{402}X$	Heterozygous
		2	1	2	2	$W_{402}X$	at *Eco*47III
985	severe	1	1	2	2	$W_{402}X$	
		1	1	2	2	$W_{402}X$	
1062	severe	1	1	2	2	$W_{402}X$	Heterozygous
		1	2	2	2	$W_{402}X$	at *Nsp* I
2474	mild	1	1	2	2	$W_{402}X$	
		2	1	1	2	$g_{1487}{}^{a}$	Splice mutation
2662	intermediate/	1	1	2	2	$R_{89}Q$	
	mild	2	1	3	2	$t_{1978}{}^{c}$	Splice mutation
2827	severe	1	1	2	2	$D_{349}N$	Active site mutation?
		1	2	2	1	$Q_{70}X$	

recombination in *IDUA*. A total of 18 polymorphisms were detected within *IDUA*, 14 of which were common to patients 2662 and 2827.

Most MPS-I patients analysed in this study are of European descent. Patient samples were collected during routine diagnosis of MPS-I or because of an interest in MPS-I patients with intermediate and mild phenotypes. Thus it is possible that the group of MPS-I patients used in this study is not representative of MPS-I patients in general, with mild and intermediate patient phenotypes possibly being over-represented. Thus the frequencies of the severe mutations detected in this study, $W_{402}X$, $Q_{70}X$, and $P_{533}R$, may be an underestimate of the frequency of these mutations in MPS-I patients in general. The founder effect may dramatically alter the frequencies of these mutations in relatively small or isolated populations. This may be reflected in the fact that the mutations described account for 91% (33/36) of all MPS-I alleles and 15 complete genotypes from the 18 Australasian patients included in this study. $W_{402}X$ and $Q_{70}X$ account for 50% and 25% of MPS-I alleles respectively in this patient group.

Of the group of 73 patients which have been completely analysed for the five mutations found in more than one patient, 55% of all mutant alleles are defined by $W_{402}X$ (34%), $Q_{70}X$ (15%), $P_{533}R$ (3%), $R_{89}Q$ (2%) and $g_{1487}a$ (2%). The complete genotypes of 26 (36%) of these patients are now fully defined by these mutations, matching the expected 30%. Thus, approximately 30% of MPS-I patients of European background should now have completely definable genotypes such that some clinical phenotype prediction may be offered. With the definition of the 3 unique MPS-I mutations, an additional 3 patients have had their genotypes fully defined bringing the total number of defined genotypes to 29/73 (40%). In addition, 26 (36%) patients have one allele defined. A list of individual patient genotypes detected in this study is shown in Appendix F.

Approximately 20% of the patients in this study had mild to intermediate phenotypes (15/73). Clinical descriptions and patient names were not obtained for all patient samples. Several of the mild patients were not of European origin and did not have any alleles defined by the severe common mutations, $W_{402}X$ and $Q_{70}X$, or by the

mild mutations that were found more than once, $R_{89}Q$ and $g_{1487}a$. As expected, approximately 50% of alleles were defined by $W_{402}X$ and $Q_{70}X$ in the remaining mild patients. Assuming that most patients with mild phenotypes are the result of a combination of a severe and a mild mutant allele, then, in the group of patients studied, 10% of MPS-I alleles will cause a mild to intermediate phenotype and 90% of alleles will cause a severe or severe to intermediate phenotype. Thus, $R_{89}Q$ and $g_{1487}a$ account for 40% of the mild MPS-I mutations and $W_{402}X$, $Q_{70}X$, and $P_{533}R$ account for 60% of severe or severe to intermediate alleles. Many of the patients used in this study that have no alleles defined have mild or intermediate phenotypes (see Appendix F).

CHAPTER 5

CONCLUDING DISCUSSION

5.1 GENERAL DISCUSSION

The work presented in this thesis was primarily aimed at the isolation of the gene for IDUA and its use in characterizing the lysosomal storage disorder (LSD), MPS-I. The following discussion will include a brief synopsis of the information available on the IDUA protein and gene before this study was commenced, that which currently known, and possible future avenues of research. The implications for patient diagnosis, prediction of patient phenotype and assessment of patient treatment are discussed. Also, the future prospects of investigating MPS-I patient therapy by enzyme replacement and gene replacement with the availability of the full-length IDUA cDNA clone are discussed.

At the commencement of the work described in this thesis, the enzyme defect in MPS-I was known to be a deficiency of IDUA (Bach *et al.*, 1972; Matalon and Dorfman, 1972). It was considered possible that the IDUA protein consisted of subunits encoded by different genes which may have been specific for heparan sulphate and dermatan sulphate (McKusick *et al.*, 1972; Matalon and Deanching, 1977). The gene for IDUA had previously been mapped to chromosome 22 by immunoprecipitation of human IDUA activity from human-mouse cell hybrids with an IDUA polyclonal antiserum (Schuchman *et al.*, 1984a). This result has since been found to be incorrect and led to several unsuccessful attempts to clone IDUA. Several groups have attempted the purification of human IDUA protein from a variety of tissue sources with varying degrees of success, probably due to the low abundance of IDUA (Barton and Neufeld, 1971; Rome *et al.*, 1972; Schuchman *et al.*, 1984b; Clements *et al.*, 1985a; Clements *et al.*, 1989). Clements *et al.* (1985a) reported the 20,000-fold purification of human liver IDUA and the subsequent production of monoclonal antibodies to IDUA. The monoclonal antibodies allowed 172,000-fold purification of large amounts of human liver IDUA protein using an immunoaffinity column which provided enough homogeneous IDUA for extensive peptide sequencing (Clements *et al.*, 1989).

Using this amino acid sequence information, it was possible to isolate cDNA and genomic clones that encoded IDUA (Scott *et al.*, 1991a; 1992a; see Chapter 3) for

the first time. Both the cDNA and genomic IDUA clones described in this thesis were used for *in situ* hybridization experiments to human metaphase chromosomes which localized the gene for IDUA to chromosome 4p16.3. Because this result was in conflict with a previous report which mapped *IDUA* to chromosome 22 (Schuchman *et al.*, 1984a), immunocapture of human IDUA activity from human-mouse somatic cell hybrids was used to confirm the localization of *IDUA* to human chromosome 4 (Scott *et al.*, 1990). Regional mapping placed *IDUA* only 1100-kb from the telomere of 4p and 1000-kb from the two most likely candidate regions for the Huntington disease gene (MacDonald *et al.*, 1991b). The isolation of IDUA genomic clones, IDUA cDNA clones, and the chromosomal localization and regional mapping studies for *IDUA* were all an integral part in the eventual characterization of the full-length IDUA cDNA. The characterization of the full-length cDNA showed that the 7 major polypeptides present in immunopurified human liver IDUA were all derived by proteolytic processing of a 74-kDa precursor (Scott *et al.*, 1991a).

The full coding sequence for IDUA was collated from cDNA clones and exons I and II of *IDUA* from genomic DNA sequence. Demonstration that the entire coding sequence for IDUA had been characterized was achieved following construction of a full-length cDNA insert and expression of human IDUA activity in CHO cells (Scott *et al.*, 1991a). The availability of a full-length IDUA cDNA insert (see Section 3.8) will enable mutagenesis studies to be performed for the study of structural and functional aspects of IDUA, including the identification of amino acid residues involved in substrate binding and catalysis. Amino acid residues to be targeted for mutagenesis in IDUA may be identified by protein homology studies or by the study of patient mutations (see Chapters 3 and 4 respectively). Two residues of potential interest, R_{89} and D_{349}, have already been identified through patient studies (see Section 4.4.2b and 4.4.2d). The predicted IDUA protein sequence, perhaps surprisingly, did not show homology to other lysosomal glycosidases that have similar functions and substrates, such as α-L-fucosidase or β-glucuronidase, but did show homology to the β-xylosidase enzyme encoded by the *xynB* gene of the thermophilic microorganism, *Caldocellum saccharolyticium* (Luthi *et al.*, 1990; Section 3.10). The availability of a large number

of IDUA substrates and monoclonal antibodies (Clements *et al.*, 1989; Freeman and Hopwood, 1992a; D. Brooks, unpublished data) will enable the accurate measurement of the properties, such as K_m, V_{max} and specific activity, of various mutant IDUA proteins that are expressed. Information from mutagenesis and MPS-I mutation studies could provide some indication of the areas involved in the binding and catalysis of the IDUA substrates. However, to gain a true understanding of the structure and function of IDUA or other lysosomal enzymes, and thus the effect of mutations on IDUA function and transport, it will be necessary to know the three-dimensional structure of the IDUA protein. Low abundance, partial deglycosylation and proteolytic processing have resulted in IDUA preparations that are too heterogeneous for crystallization, even if sufficient amounts of protein had been purified. The *in vitro* production of large amounts of IDUA has been achieved (J. Durrant and D. Anson, unpublished results) so that studies can begin to produce IDUA crystals for X-ray crystallography which can determine the tertiary structure of IDUA, including identification of the residues involved in the binding and catalysis of the IDUA substrates.

As discussed in Chapter 1, previous attempts to distinguish between the extremes of MPS-I clinical phenotypes by biochemical techniques (Hopwood and Muller, 1979; Ullrich *et al.*, 1981; Matalon *et al.*, 1983; Muller and Hopwood, 1984) and immunological techniques (Myerowitz and Neufeld, 1981; Schuchman and Desnick, 1988; Taylor *et al.*, 1991; Ashton *et al.*, 1992) were only partially successful. The isolation and characterization of the genomic sequence for the IDUA gene enabled the characterization of 8 mutations, which represents all of the MPS-I alleles present in a group of nine patients that were analysed (Scott *et al.*, 1992b; 1992d; see Chapter 4). All except three of these mutations were observed more than once. In fact, two of the mutations accounted for 48% of all MPS-I alleles in a group of 73 patients completely analysed for the 8 mutations. The complete genotypes of 29 MPS-I patients were determined allowing reasonable but limited phenotype to genotype associations to be drawn, thus enabling reliable prognostic predictions to be made for approximately 30% of MPS-I patients in the future. These results have also provided molecular explanations for the differences between Hurler and Scheie syndromes. The hypothesis

that IDUA consists of subunits encoded by different genes which may be specific for heparan sulphate and dermatan sulphate (McKusick *et al.*, 1972; Matalon and Deanching, 1977) is very unlikely considering the identification of mutations in *IDUA* causing mild MPS-I, the identification of all major polypeptides from immunopurified human liver IDUA within the IDUA peptide sequence and the proposed model of proteolytic processing of human liver IDUA described in this thesis. It is expected that the mutations described will also account for a similar number of alleles in other MPS-I patient groups. However, caution should always be used in predicting a patients clinical outcome from their genotype, especially with the occurrence of "mild" alleles, a novel combination of alleles or the first observation of a described genotype in a genetically distinct population (e.g. Japanese Gaucher's $L_{444}P$ homozygote, non-neuronopathic, Masuno *et al.*, 1990).

The work described in this thesis provides PCR based tests for the presence of the MPS-I mutations described, which could be extended for use in prenatal diagnosis or carrier detection after primary diagnosis in a propositus of the IDUA enzyme defect. In fact, a 1933-bp PCR product spanning all the described MPS-I mutations except for $Q_{70}X$ has now been obtained between the primers ID78 and ID103 (Fig 4.3) for use in diagnosis of MPS-I (P. Nelson, unpublished results). It may also be possible to gain an accurate estimate of the frequency of MPS-IH by population screening to determine the carrier frequency of $W_{402}X$ and $Q_{70}X$. The regional mapping of IDUA (MacDonald *et al.*, 1991b) makes possible the use of a large number of polymorphisms, particularly VNTR-type polymorphisms (e.g. Weber *et al.* 1992), for prenatal diagnosis and carrier detection of MPS-I by linkage analysis. The detection of a large number of polymorphisms within *IDUA* (Scott *et al.*, 1991b; 1991c; 1992c; 1992e; see Table 4.11) provides a similar diagnostic tool, however, the possibility of a high rate of intragenic recombination within *IDUA* must be taken into account when performing diagnosis by linkage analysis. With the availability of the IDUA cDNA and genomic sequence and as the technology for analysing patient mutations develops, it may become routine to the determine all MPS-I mutations after primary diagnosis for more accurate MPS-I phenotype predictions.

5.2 PATIENT TREATMENT

5.2.1 THE EFFECTS OF BONE MARROW TRANSPLANTATION ON MPS-I PATIENTS WITH DEFINED GENOTYPES

As described in Chapter 1, the mucopolysaccharidoses, as a group occur with an incidence of around 1/15,000 live births, with MPS-I being the most common of the mucopolysaccharidoses in European populations (Hopwood and Morris, 1990). Currently, bone marrow transplantation (BMT) is the major therapy being evaluated for MPS-I (Neufeld and Muenzer, 1989; Hoogerbrugge and Vossen, 1990; Hopwood and Morris, 1990; Krivit *et al.*, 1990) and other genetic diseases (Weinberg and Parkman, 1990). A lack of accurate diagnostic tests to predict phenotype for MPS-I patients has made it difficult to assess the success of therapies used for this condition. Further, this will continue to be a problem with the advent of new treatments such as enzyme replacement therapy or gene replacement therapy. With some limitations, the elucidation of mutations for genetic diseases such as Gaucher's disease has allowed carrier detection and the accurate prediction of patient phenotype (e.g. $L_{444}P$ neuronopathic Gaucher's disease, see Section 1.5; reviewed in Beutler, 1991). It has also provided an invaluable resource for genetic counselling and the assessment of the efficacy of enzyme replacement therapy now being performed for Gaucher's disease (Beutler, 1991; Beutler *et al.*, 1991).

The results described in this thesis indicate that the identification of the $W_{402}X$, $Q_{70}X$, $P_{533}R$, $R_{89}Q$ and $g_{1487}a$ MPS-I alleles will be similarly useful for MPS-I. To date, a total of at least 38 MPS-I patients have undergone BMT (Hoogerbrugge and Vossen, 1990). The correlation of a severe clinical phenotype with $W_{402}X$, $Q_{70}X$ and $P_{533}R$ homozygotes or compound heterozygotes is currently being retrospectively applied to identify bone marrow transplanted patients with defined genotypes to obtain some measure of the efficacy of BMT. The frequency of these severe mutations, particularly $W_{402}X$, appears to be increased in bone marrow transplanted MPS-I patients. In 33 samples from bone marrow transplanted MPS-I patients (most samples

are distinct from other patient samples in this study), the frequency of $W_{402}X$ is 48.5% (32/66). This suggests that the selection criteria used for choosing MPS-I patients to undergo BMT have been at least partially correct.

Clinical data available on two $W_{402}X$ homozygous patients in the care of Dr J. Edward Wraith (Willink Biochemical Genetics Unit, Royal Manchester Children's Hospital) and Dr Ashok Vellodi (Westminster Children's Hospital, London) indicate that allogeneic BMT markedly decreased progression of disease symptoms in MPS-I. From the results presented in Section 4.4.1a, it is apparent that the clinical deterioration of $W_{402}X$ homozygotes is generally very rapid with profound mental retardation and death usually occurring before four years of age from cardiac or respiratory failure. Patient 3061 (64-1) went to the full term of pregnancy and weighed 9 lbs 8 oz on delivery. She had difficulty feeding at first and had a persistently runny nose. Diagnosis of Hurler syndrome was made at 10 months and the patient was transplanted at 14 months of age in 1982, at which time her development was in the normal range. Grommets were inserted on two occasions post-transplant and an umbilical hernia was repaired at 4 years of age. Some surgical relief of bilateral carpal tunnel syndrome in classical Hurler-like clawed hands was necessary at 7 years and 7 months of age. At 7 years and 10 months of age, the patient developed backache in the area of her worsening lower thoracic kyphosis. At the age of 10 years and 4 months, the patient had a posterior spinal fusion using her own left anterior iliac spine plus donor bone. Echocardiograms, including one performed at 10 years of age, showed no deterioration in left ventricular function or valve lesions. She has hearing loss as a result of mixed conductive and sensory neural problems that has not appeared to have worsened since assessment at 2 years post-transplant. Full scale IQ assessments were as follows: at 6 years one month, 92; 7 years and 7 months, 90; 9 years and 2 months, 87. The patient's mobility is limited due to instability of the right knee and valgus deformity with internal rotation of the knee. There is also a moderate degree of stiffness in both hips. Hepatosplenomegaly has been almost totally relieved and the level of circulating IDUA enzyme activity is within the testing laboratory's normal range. In general, the dwarfism normally associated with severe Hurler syndrome has been partially alleviated

although, as described, skeletal pathology is still apparent, particularly in the spine which has an anterior hooklike projection due to lower thoracic kyphosis. The patient is currently alive at age 11 years and attends a normal school were she is reported to be coping well. The most striking feature must remain that the patient is still alive at 11 years of age and has an intelligence in the normal range.

Patient 3532 was diagnosed with Hurler syndrome at the age of 5 months on the basis of delayed development and clinical features including abnormal facies hepatosplenomegaly, lumbar gibbus, umbilical and inguinal herniae. At 8 months of age on pre-transplant assessment, the patient was also noted to have a persistently runny nose, a large head, corneal clouding and clawed hands. Bone marrow transplantation was performed at 9 months of age using the haploidentical father. The post-BMT recovery has been complicated by the development of acute graft-versus-host disease 8 days post-BMT progressing to chronic graft-versus-host disease and several septicaemic episodes. Increasing pain and stiffness in the hips and knees was first noticed 5 to 6 years post-BMT. Bilateral carpal tunnel syndrome was discovered at 10 years of age. Short form IQ assessments were as follows: at 7 years one month, 89; 8 years and 2 months, 89; 9 years and 3 months, 75; 10 years and 7 months, 71. The patient has very mild corneal clouding, distance vision is poor (3/36) although near vision is very good. The patient has only mild hearing loss on one side. At 13 years of age, the patient now has severe lower limb disabilities and has been wheelchair-bound for the last year. The typical kyphosis of Hurler syndrome has slowly progressed. The patient now attends a special school for physically handicapped children.

These results must be considered preliminary and anecdotal but support the observation that BMT can prevent or delay the development of neurological symptoms in the canine model of the LSD, fucosidosis (Taylor *et al.*, 1989; 1992), and reduce other somatic symptoms and storage of GAG in the CNS of MPS-I dogs (Shull *et al.*, 1988; Breider *et al.*, 1989; Gompf *et al.*, 1990). Studies on these animal models have emphasized that, to have the greatest benefit in arresting or slowing the development of disease symptoms, BMT must be performed at an early stage of life before there has

been a marked clinical development of the disease. The full genotypes of a number of other bone marrow transplanted MPS-I patients is now available and a clearer picture of the effectiveness of BMT may appear as clinical data becomes available and is collated for these patients. The research described in this thesis has, within limits, enabled MPS-I patients to be genotyped and given a clinical prognosis. Depending on the outcome of studies on BMT patients, if a severe clinical outcome is predicted at diagnosis, a difficult ethical question may present, particularly for parents of MPS-I patients. That is, whether or not BMT should be performed as soon as a matching donor is available. The partial success of BMT documented above means that these patients have long term medical problems, and are now intellectually aware of those problems. It is possible that further developments of treatment protocols, as discussed in the next sections, may overcome some of these problems and ensure a much improved quality of life for treated MPS-I patients.

These results and the results from the animal studies also indicate that gene replacement therapy and, to a lesser extent, enzyme replacement therapy may become viable options for treatment of MPS-I in the near future. These therapies, or a combination thereof, should result in high levels of IDUA circulating in the bloodstream. If performed quickly after early diagnosis, these therapies may be more successful than the presently applied allogeneic BMT in treatment of all symptoms, particularly growth plate and skeletal deformities. Gene therapy, as an autologous form of BMT, would also alleviate the major problems presently encountered in allogeneic BMT, namely, the availability of a matched bone marrow donor and the development of graft-versus-host disease. Prenatal diagnosis should be available for families with a propositus using the mutations discussed in Chapter 4. This may enable therapy to begin as soon as possible for many MPS-I patients, perhaps by the collection of stem cells from placenta or cord blood for retroviral infection leading to gene therapy, especially in families were termination of pregnancies is not acceptable. It may also be possible to extrapolate any observed success of BMT for MPS-I patients to the treatment of other MPS disorders. However, the problem with other MPS disorders will

remain the prediction of clinical phenotype and then the assessment of the efficacy of the treatment.

5.2.2 FUTURE PROSPECTS FOR PATIENT TREATMENT

The work described in this thesis also opens up potential for the development of new patient therapy strategies. At present, there is not a generally applicable method for the treatment of LSD, including the MPS disorders. The two main therapeutic approaches under active development are enzyme replacement therapy and gene replacement therapy. The following discussion will include Gaucher's disease as a model, it being the most advanced of the LSD in the trial of these new therapies. The existence of animal models for MPS-I (canine MPS-I, Shull *et al.*, 1982; Spellacy *et al.*, 1983; feline MPS-I, Haskins *et al.*, 1979) and other lysosomal storage diseases (caprine MPS-IIID, Prof. M. Jones, pers. commun.; feline MPS-VI, Jezyk *et al.*, 1977; canine MPS-VII, Haskins *et al.*, 1984; murine MPS-VII, Birkenmieier *et al.*, 1989; canine fucosidosis, Kelly *et al.*, 1983; murine Gaucher's disease, Tybulewicz *et al.*, 1992) will be invaluable in the assessment of these new therapy protocols. It should be possible to perform targeted homologous recombination (Capecchi, 1989) to disrupt the IDUA gene or to introduce specific mutations and use embryonic stem cells for production of a mouse model for MPS-I. This method was used to produce the mouse model for Gaucher's disease, but the affected mice die within 24 hours of birth (Tybulewicz *et al.*, 1992). This emphasizes the fact that, for animal models to be useful, they must match the clinical problems observed in the equivalent human disease. The advantages of mouse models for human genetic diseases are that they are considerably cheaper to maintain, large numbers of animals can be handled and their short life span or quick disease development may give a rapid assessment of the efficacy of various treatment protocols before the results of therapy in the larger animal models, or even humans, is known. However, mice have different anatomy and biochemistry which may affect the correlation of treatment to the human diseases. The main advantage of the canine or other large mammalian models for genetic disease are that the animal's size more

closely resembles that of humans, thus making them suitable for procedures similar to those used on human patients. The fucosidosis dog and caprine MPS-IIID models allows the CNS pathology, similar to that observed in a number of the MPS, to be studied separately from somatic disease pathology since virtually all symptoms of the diseases in these animal models are restricted to the CNS. Similarly, the MPS-VI cat model allows study of the somatic disease pathology, particularly bone pathology, as is found in a number of the other MPS without the complication of severe CNS involvement such as MPS-IVA (Morquio A syndrome; Neufeld and Muenzer, 1989). For all forms of therapy attempted, it is important to remember that BMT in the animal models has demonstrated that treatment performed at an early age will have the greatest effect in alleviating or preventing symptoms of the diseases. In particular, the CNS deterioration or bone growth plate deformities associated with MPS-I are unlikely to be reversible once they have developed.

5.2.2a ENZYME REPLACEMENT THERAPY

The potential for enzyme replacement therapy in treating LSD has been demonstrated by treatment of Gaucher's disease patients with purified human glucocerebrosidase enzyme (Barton *et al.*, 1990; for review, see Beutler, 1991). Oligosaccharide chains on glucocerebrosidase purified from human placenta were modified in order to be terminated with mannose, enabling the efficient uptake of the enzyme by mannose receptors on reticulo-endothelial, macrophage and monocyte cell surfaces. However, several problems still remain. Many cells have mannose receptors resulting in rapid clearance of the enzyme from circulation, partially by cells not involved in the disease pathogenesis, as the enzyme is not specifically targeted to the macrophages, the primary site of pathogenesis in Gaucher's disease (Barranger and Ginns, 1989). Clinically intermediate forms of Gaucher's disease may also involve bone pathology with flaring of the distal femur. Long-term studies of a child with Gaucher's disease indicated that the response to treatment with modified glucocerebrosidase is dose dependent, and that the hematologic response precedes the skeletal response (Barton *et al.*, 1992). The bone pathology associated with some of the

MPS is quite different to that associated with Gaucher's disease as the site of pathology in the MPS is at the growth plates, which have a very limited circulatory system at the region of highest metabolic activity, and therefore GAG turnover. This could make delivery of enzyme to the growth plate to prevent skeletal deformities associated with some of the MPS difficult. A large number of the LSD, including MPS-I and some forms of Gaucher's disease, involve deterioration of the CNS. Although there have been some reports of M6P receptors on isolated brain microvessels (e.g. Frank *et al.*, 1986), enzyme replacement therapy is predicted to be of little use in preventing CNS deterioration due to the presence of the blood-brain barrier which will prevent transport of therapeutic amounts of the enzyme to the site of pathology. Also, enzyme replacement therapy requires the continual administration of the therapeutic agent, leading to prolonged expense and perhaps patient distress. It has also been suggested that the greatest drawback to enzyme replacement therapy for Gaucher's disease is the expense (Beutler, 1991). However, with the continual development and application of recombinant DNA technology in the pharmaceutical industry, recombinant proteins for the treatment of many different genetic diseases should become an economically viable proposition in the near future.

With the production of large quantities of recombinant IDUA with in CHO cells (J. Durrant and D. Anson, unpublished results), it will be possible to begin evaluation of enzyme replacement therapy in MPS-I patients. Recombinant IDUA, expressed into CHO cell media, has been shown to correct the storage of mucopolysaccharides in MPS-I fibroblasts following M6P-dependent endocytosis and transfer of the IDUA to the lysosome (A. Whittle, T. Rozaklis and G. Harper, unpublished results). Similar results have been obtained for iduronate-2-sulphatase using MPS-II fibroblasts (Bielicki *et al.*, 1992) and *N*-acetylgalactosamine-4-sulphatase using MPS-VI fibroblasts (Anson *et al.*, 1992b). Despite the problems discussed above, enzyme replacement therapy for the MPS disorders can be is expected to at least reduce soft tissue storage of GAG, such as hepatosplenomegaly and thickened joint capsules, and thus provide some relief for treated patients. Systems such as macrophage loading which can be used to help target

enzyme to the major sites of pathology in MPS-I, the CNS and growth plates, may also be developed.

5.2.2b GENE REPLACEMENT THERAPY

The apparent success of BMT in treating the neurological symptoms of LSD diseases, including preliminary work presented in this thesis for human MPS-I (Section 5.2.1; for review, see Hoogerbrugge and Vossen, 1990) and canine fucosidosis (Taylor *et al.*, 1989; 1992) implies that enzyme in bone marrow derived cells is able to cross the blood-brain barrier. Presumably, the enzyme is transported across this barrier by bone marrow-derived macrophages (Hoogerbrugge *et al.*, 1988). In the case of canine MPS-I, long term BMT also relieved most of the somatic disease symptoms observed, although development of MPS-I bone pathology is not completely prevented (Shull *et al.*, 1988; Breider *et al.*, 1989). However, allogeneic BMT using immunologically-matched donors still carries a high risk of mortality and morbidity (Hoogerbrugge and Vossen, 1990). The use of recombinant retroviral vectors to transfer genetic sequences into hematopoietic stem cells is the major avenue currently being researched for somatic gene therapy. Genetically engineered pluripotent stems cells can repopulate the hematopoietic cell population and ensure the continual presence of the genetically modified cells in the hematopoietic system. Thus the introduction of genes into hematopoietic cells and their circulation and maturation throughout the body should enable the transfer of enzyme to other somatic cells through cell-to-cell interaction or secretion of the enzyme into extracellular fluid for uptake by other cells. The fact that macrophages also appear to cross the blood-brain barrier makes this approach attractive for the treatment of diseases involving CNS pathology. Two limitations of this approach are the degree of efficiency of transfer of the gene in retroviral vectors into the reconstituting hematopoietic stem cells and the loss of reconstituting ability of hematopoietic stem cells when manipulated *in vitro* during infection and selection. These problems are being addressed by the use of simplified vectors respectively (e.g. Lim *et al.*, 1989), and by prestimulation of the stem cells with growth factors prior to and during infection (Luskey *et al.*, 1990). A major advantage of somatic gene therapy

over enzyme replacement therapy is that, following optimization of the protocols, the corrective procedure should only be required to be performed once during the lifetime of the patient, as retroviral infection of the most primitive hematopoietic stem cells should result in continual production of corrected cells. A number of groups are actively pursuing gene replacement therapy for the LSD, including Gaucher's disease (Correll *et al.*, 1989,; Fink *et al.*, 1990; Kohn *et al.*, 1991; Weinthal *et al.*, 1991) and MPS-VII (Kyle *et al.*, 1990; Wolfe *et al.*, 1990; Birkenmeier, 1991). More preliminary reports are available for MPS-VI (Peters *et al.*, 1991; D.S. Anson pers. commun.), fucosidosis (Occhiodoro *et al.*, 1992) and, as a result of the construction of a full-length IDUA cDNA described in this thesis, MPS-I (Anson *et al.*, 1992a).

All these studies have demonstrated that gene transfer using retroviral vectors can restore normal lysosomal enzyme activity in cells derived from LSD patients. IDUA enzyme activities of 1 to 3% compared with normal, and 7 to 15% compared with normal were found in the brain tissue and cerebrospinal fluid of long term BMT MPS-I dogs (Shull *et al.*, 1988). As is observed in some mild MPS-I patients (e.g. cell line no. 2474, Ashton *et al.*, 1992), if this relatively low amount of enzyme is normally distributed, it is sufficient to protect the CNS from further deterioration and thus is the minimum level of enzyme activity that must be reconstituted by therapy protocols. By using efficient promoters, it may be possible to over-express the therapeutic enzyme leading to the levels of enzyme activity in the CNS of patients treated via autologous retrovirally modified stem cells that approach, or even exceed normal levels. Over-expression may also result in increased secretion of the correcting enzymes for delivery to other sites of disease pathology which may lead to more rapid correction of the pathology compared to allogeneic BMT. However, *in vitro* correction of IDUA deficiency in MPS-I fibroblasts using retroviral constructs containing these viral promoters has led to some complications. The greater the over-expression of IDUA, the lower the level of correction of storage in the MPS-I fibroblasts as measured by the intracellular accumulation of ^{35}S-labelled GAG (Anson *et al.*, 1992a). For example 400-fold over-expression of IDUA resulted in 4-fold inferior correction than normal IDUA expression but 30-fold over-expression resulted in total correction. It is proposed

that this may be due to saturation of the M6P receptors required to target lysosomal enzymes to the lysosome, as evidenced by the increased secretion of other lysosomal enzymes. This would result in reduced amounts of the lysosomal enzymes, that use the M6P pathway, reaching the lysosome and lead to a situation analogous to I-cell disease, where the enzyme responsible for producing the M6P signal on lysosomal enzymes is deficient (Nolan and Sly, 1989). Similar results have been obtained with over-expression of *N*-acetylgalactosamine-4-sulphatase in MPS-VI cells leading to retention of ^{35}S-GAG (D.S. Anson, pers. commun.). This over-expression results in a multiple sulphatase-like deficiency, presumably due to the saturation of a sulphatase-specific processing mechanism. The development of I-cell disease or multiple sulphatase like deficiencies may not be important, as some degree of over-expression may be attained and presumably this effect will be localized to a small number of the genetically modified cells in the hematopoietic system.

It has been observed in animal model experiments that long term expression of enzyme activity following BMT of retrovirus infected cells has been difficult to achieve (e.g. Palmer *et al.*, 1991). It is thought that the use of powerful viral promoters (e.g. retroviral LTRs, the SV40 early promoter and the human cytomegalovirus immediate early promoter) and selectable markers such as neomycin resistance may somehow switch-off or inhibit expression of the gene of interest *in vivo*. Omission of these sequences has enabled the production of retroviral stocks with higher titres, resulting in more efficient infection of stem cells and has also allowed expression of enzyme activity for several months (e.g. Lim *et al.*, 1989).

These results suggest that achievement of gene therapy for the LSD will require careful consideration of vector design, such that inappropriate sequences are excluded, to maintain expression of the correcting enzyme at appropriate levels. These results will also have to be taken into consideration if target tissues other than stem cells are utilized for somatic gene therapy. Unlike allogeneic BMT, somatic gene therapy utilizing stem cells will not require the removal of all affected stem cells in a patient. In fact, the relatively small number of genetically modified stem cells may have a selective

advantage over unmodified stem cells and thus take over the hematopoietic system. If necessary, it may also be possible to freeze patient stem cells for reinfection at a later stage. In the short term somatic gene therapy is considered unlikely to be a cure as it will continue to be difficult to establish grafts which will result in long term expression. Several autologous transplants of genetically modified stem cells may be necessary, or it is possible that a combination of enzyme replacement and gene replacement therapies will be necessary to treat all sites of pathology.

5.3 FINAL STATEMENTS

Many of the aims of this thesis have been achieved. The gene for IDUA has been isolated and characterized. A full-length IDUA cDNA clone is available, and is actively being used to pursue enzyme replacement therapy and gene replacement therapy for MPS-I. Eight mutations that cause MPS-I have been identified, leading to accurate genotype to phenotype correlations and a molecular explanation for the difference between Hurler and Scheie syndromes. Regional mapping of *IDUA*, definition of polymorphisms within *IDUA* and mutation analysis should enable an improvement in the diagnosis of MPS-I, particularly with regard to prenatal diagnosis, carrier detection and prediction of clinical phenotype. Many mutations remain to be defined which, in conjunction with various other approaches, may help in gaining further insight into the structure and function of IDUA. Importantly, encouraging preliminary evidence quantifying the benefit of allogeneic BMT, at least in some genotyped MPS-I patients, has now been obtained (Section 5.2.1). The logical extension of such a finding would be that MPS-I may be treatable by new therapy protocols such as enzyme replacement therapy and gene replacement therapy.

During the course of this thesis, four other genes encoding enzymes deficient in MPS disorders were cloned (iduronate-2-sulphatase, deficient in MPS-II, Wilson *et al.*, 1990: galactose-6-sulphatase, deficient in MPS-IVA, Tomatsu *et al.*, 1991a: β-galactosidase, deficient in MPS-IVB, Oshima *et al.*, 1988: *N*-acetylgalactosamine-4-sulphatase, deficient in MPS-VI, Peters *et al.*, 1990; Schuchman *et al.*, 1990; Litjens *et*

al., 1991). With the cloning of *IDUA*, seven of the ten genes in the defined MPS disorders are now cloned and only the genes for MPS-IIIA to C remain to be isolated. Recently a number of mutations have been defined in MPS-II (Wilson *et al.*, 1990; Sukegawa *et al.*, 1992; Wehnert *et al.*, 1992), MPS-IVB (Oshima *et al.*, 1991), MPS-VI (Wicker *et al.*, 1991; Jin *et al.*, 1992) and MPS-VII (Tomatsu *et al.*, 1990; 1991b; Fukuda *et al.*, 1991). Most of these mutations are rare and have demonstrated that there is considerable genetic heterogeneity in these MPS disorders. Except for the large number of deletions reported in MPS-II (Wilson *et al.*, 1990), these mutations may be of little use in genotype to phenotype correlations and assessment of the efficacy of treatment protocols. The aims of this thesis, as presented in Section 1.6, are applicable to all the enzymes deficient in the MPS. The cloning of *IDUA*, expression of IDUA protein, and definition of common and rare mutations enabling genotype to phenotype correlations should ensure that MPS-I remains at the forefront of MPS research, and thus the prototype MPS, particularly with respect to diagnosis and assessment of patient therapy protocols. Animal models for the MPS will play an important part in assessing these new therapies and also in gaining a greater understanding of the pathogenesis of the MPS. An improved understanding of the function of IDUA and GAG, lysosomal biogenesis, and MPS pathology in growth plates and the CNS may also provide new methods for combating these most debilitating diseases.

APPENDICES

APPENDIX A. OLIGONUCLEOTIDE PROBES FOR IDUA.

27-mer to peptide 2.

ID2 ACCGGCTTCTGCCCCTACCTGCCCCAC

peptide 8 probes.

ID5 TGATGATCIGGTTCATTCCA antisense
G G C G

ID6 AAAAAICCCTGCATIGTCAT antisense
GG T

ID7 TCATTCCANGTTTCAAAATT antisense
G C G G

ID8 TCATTCCAIGTTTCAAAATT antisense
G C G G

ID47 AACTTCGAGACCTGGAACGAGCCCGACCACCACGACTTCGACAACGT
sense

peptide 9 probe.

ID9 GCCCAIAATTGTTCTTCATC antisense
GC C C G

65/60-kDa N-terminal probe.

ID50 TGGTTCTCCCGCAGCAGGTCCAGGTAGCCGTCCAGGTGGGTGAAGTGGTA
antisense

APPENDIX B. OLIGONUCLEOTIDES USED FOR SEQUENCING IDUA cDNA.

The name, sequence, orientation and position of oligonucleotide primers used to sequence the IDUA cDNA clones λRPC1 and λE8A, are given with reference to the sequence in Fig. 3.22. The remaining coding sequence was collated from the *IDUA* genomic DNA sequence.

name	sequence	sense (+) or antisense (-)	position
IDNT	GAGCTACAACTTCACCCACC	+	409-428
IDCS	GCCCGGCGACTCCTTCCACA	+	747-767
ID14	TCCTGGACAGCAACCACACG	+	1320-1339
ID15	CTCGAACTCGGTAGGAGCCA	-	1952-1933
ID16	CTGCACTGTGCCCATGCTGC	+	2105-2124
ID17	GGTACGGACTGGCGCATGTT	+	584-603
ID18	GCGGACACCCCCATTTACAA	+	959-978
ID19	GCGCGCTTCCAGGTCAACAA	+	1187-1206
ID21	CAGCAGGAAGCCATCGACCT	+	1872-1891
ID22	TGCAGCTTGTCCTCAAAGTC	-	531-512
ID23	CCGTGCAGGCATCGTAGTAG	-	707-688
ID24	CGTGCGCGCACCCTTCCTGT	-	892-873
ID25	GGAGGTGGTGTTGGCCAGTA	-	1105-1086
ID26	GGCCCAGAGCTGCTCCTCAT	-	1298-1279
ID27	GTCACCGCGACGCTGCGGTT	-	1458-1439
ID28	TCTTCCCTTGGGTCAGGGGCA	-	1776-1756
ID42	TGGTCTACGTCACGCGCTAC	+	1494-1513

APPENDIX C. OLIGONUCLEOTIDES FOR SEQUENCING IDUA GENOMIC DNA.

The name, sequence, orientation and position of oligonucleotide primers used to obtain the *IDUA* genomic DNA sequence in Fig. 3.24A are given.

name	sequence	sense (+) or antisense (-)	position
ID55	AAGCCTGTGCTCCTCCAGAA	-	737-717
ID56	CTTCTGGAGGAGCACAGGCTTCTG	+	716-740
ID58	TCACATGGGGTGCGCGCCCAGACT	+	496-519
ID59	TGGGAGGGGTTGGGTGGCCT	-	451-432
ID60	GCGCAAGCTTAGCACGCGTGGCCA TGCGTCCCCT	-	569-593
ID62	CCCGCACCTGGTGCAGGTGGACGC	+	666-689
ID66	ACAGCCAGCAAGGACACGCTCTCC	-	1594-1570
ID67	TGGCTGCTGGAGCTTGTCACCACC	+	1422-1446
ID68	GCGCGGATCCCCTGGTGGTGACAA GCTCCAGCAGC	-	1447-1424
ID72	TTCAGTCTGCTTGGTTGAATCCAT	+	40-66
ID80	CGTGTTTGTGGGTGGGTCCTCCAC	-	877-900
ID81	CGCGGCTGACACACACGTTCAAGC	-	1255-1232
ID100	AAGGAAGCGGGGCTCCAAGCCCTG	+	499-522

The name, sequence, orientation and position of oligonucleotide primers used to obtain the *IDUA* genomic DNA sequence given in Fig. 3.25A are given.

name	sequence	sense (+) or antisense (-)	position
ID48	GCACCATACCTCTCTGGCAA	+	427-448
ID49	AGGTCCTGCCTGGCTCCTGA	+	1847-1866
ID50	AAGAGGCAACATGAGTGCAC	-	1774-1755
ID53	TAGGTGTCTCCTCAGAGAGG	+	1061-1080
ID54	AGGACCTGGTGGACACCTCA	-	1853-1834
ID82	CTCACTCAGTTCACCCTCAG	+	221-240
ID83	TGAGACTGGTCTCGGTTTTGG	-	279-259
ID84	TGGGCAACAGAGTGAGACTC	+	743-762
ID85	TTCCTGACCTCGTGATCCGC	-	561-542
ID86	TTCTGAGCTGAGCAGTGCCAG	-	821-801
ID99	ACTCTTCCTGCCCAGGCAGG	-	3693-3674

The name, sequence, orientation and position of oligonucleotides used to obtain the *IDUA* genomic DNA sequence given in Fig. 3.24B are given. Many of the oligonucleotide primers used for sequencing the IDUA cDNA clones were used again and their position relavent to the genomic DNA sequence is given.

name	sequence	sense (+) or antisense (-)	position
IDNT	GAGCTACAACTTCACCCACC	+	448-467
IDCS	GCCCGGCGACTCCTTCCACA	+	1563-1583
ID14	TCCTGGACAGCAACCACACG	+	2589-2608
ID15	CTCGAACTCGGTAGGAGCCA	-	4111-4091
ID16	CTGCACTGTGCCCATGCTGC	+	4263-4282
ID17	GGTACGGACTGGCGCATGTT	+	1286-1305
ID18	GCGGACACCCCCATTTACAA	+	1875-1894
ID19	GCGCGCTTCCAGGTCAACAA	+	2210-2229
ID21	CAGCAGGAAGCCATCGACCT	+	3884-3903
ID22	TGCAGCTTGTCCTCAAAGTC	-	735-754

name	sequence	sense (+) or antisense (-)	position
ID23	CCGTGCAGGCATCGTAGTAG	-	1504-1523
ID25	GGAGGTGGTGTTGGCCAGTA	-	2109-2128
ID26	GGCCCAGAGCTGCTCCTCAT	-	2566-2546
ID27	GTCACCGCGACGCTGCGGTT	-	2727-2708
ID28	TCTTCCCTTGGGTCAGGGGCA	-	3364-3383
ID30	GAGAAGGCCTGGCAGAGCAT	+	632-650
ID31	GCCTGGCCAGGAGATACATC	+	784-803
ID32	TACTCCTGGGCTTGGTGGGT	+	995-1014
ID33	GGAGATGGGGTTCATCTTGA	+	1169-1188
ID34	TGGGGACTCCTTCACCAAGG	+	2435-2454
ID36	GGCTGGGAGCAGAGCCCACA	-	537-518
ID37	ACCACCAACCTAGTCCCGTT	-	962-943
ID38	TGGCTCAAGGGGGTCCAAGA	-	1132-1113
ID39	ATTCCACGTCTCGAAGTTCC	-	1332-1313
ID40	ACCTTCTCCTGCTCCAGGAT	-	1837-1818
ID41	GGAGGAAGTGCGCTTCCGCA	-	2392-2373
ID42	TGGTCTACGTCACGCGCTAC	+	2854-2873
ID43	AGAACCCTGAGGACCGGCCA	+	3129-3148
ID44	GCCTTACCGTCCTGAGAGAA	-	3871-3852
ID45	TTCTGGTCTGGTCGGATGAA	+	3407-3426
ID46	CTAGGGGACATGAGATGGAC	+	3636-3655
ID47	AGTGTGTGGGTGAAGGGCAG	-	3573-3592

APPENDIX D. OLIGONUCLEOTIDE PRIMERS USED FOR IDUA PCRS.

The following appendix lists all oligonucleotide primers cited in the text or figures for use in PCR or in or direct PCR sequencing. Their position is given relevant to the template used, that is, cDNA (Fig. 3.22) or genomic DNA (Fig. 3.24A* or B), unless the primers named have already been listed in Appendices C or D.

The following primers were used for PCR approaches to cloning IDUA as described in secion 3.6.2

ID13 GGAATTC GCCCGGGCGGCATCCACTTC
*Eco*RI G C

ID7.1 GGAATTC CCGGGCCCAATAATC
*Eco*RI G G

ID7.2 GGAATTC CCGGGCCCAATAATCCAGCGC
*Eco*RI G G

λgt11F CTCCTGGAGCCCGTCAGTAT

λgt11R GTAGCGACCGGCGCTCAGCT

λgt10F GCAAGTTCAGCCTGGTTAAG

λgt10R TTATGAGTATTTCTTCCAGG

SEB-T TCGACGAATTCGGATCC $(dT)_{17}$
Sal I , *Eco*RI, *Bam*HI, dT for anchored PCR.

SEB-13 GGATCCGAATTCG To clone on 3'end and PCR with SEB-19

SEB-19 GGTCGACGAATTCGGATCC For nested anchored PCR.

IDP1 CGGATCCG AATTTTGAAACCTGGAA
*Bam*HI C C G G

IDP2 GGAATTCC GCATCATAATAATT
*Eco*RI G G G G

IDP3 TGAGCCAGACCACCACGACT

IDP4 TGCATGGTCATGGAGACGTT

ID9.1 ATGGCCGTGAGCACCGGCTT

ID9.2 GCGGCCATGGTGGTGAAGGT

ID10 GTGCTCACGGCCATGGGGCT

ID11 GGTCGACCCGCACGTGCAGCTGTTGCG

The following primers were used for PCR using IDUA cDNA as template.

name	sequence	sense (+) or antisense (-)	position
ID56	CTTCTGGAGGAGCACAGGCTTCTG	+	223-246
ID57	AAGGTCCAAGTACCCGTCCAGGTG	-	448-425
ID58	TCACATGGGGTGCGCGCCCAGACT	+	1-25
ID60	GCGCAAGCTTAGCACGCGTGGC CATGCGTCCCCT	+	76-89
ID61	AGGATGGAGATGGAGCTGCGCGCA	-	906-883
ID63	ATATTGCAAAGGGGGTGATGGGAG	-	2149-2126
ID68	GCGCGGATCCCCTGGTGGTGACAA GCTCCAGCAGC	-	386-363
ID69	GACTACATCTCCCTCCACAGGAAG	+	857-880
ID71	TGGTCTACGTCACGCGCTACCTGG	+	1494-1517
ID73	AACTGCTCTGCCGTGGGGAAGACG	-	1590-1567
ID101	GGCTTGCGCAACAGCTGCACGTGC	-	1242-1219
ID120	TGGAATGAGCCAGACCACCACGAC	+	626-649
ID123	AGTTCAGGAAGCCTTGCATGGTCA	-	689-666
ID127	GGCTGCTGGCGCTGCTGGATGAGG	+	1260-1283
ID149	TACGCGGCCATGGTGGTGAAGGTC	+	1040-1063

The folowing primers were used for IDUA specific PCR from genomic DNA templates.

name	sequence	sense (+) or antisense (-)	position
ID58*	TCACATGGGGTGCGCGCCCAGACT	+	496-519
ID63	ATATTGCAAAGGGGGTGATGGGAG	-	4307-4284
ID65*	CCCCGGATCCCCAGAAGCCTGTGCT CCTCCAGAAG	-	739-717
ID66*	ACAGCCAGCAAGGACACGCTCTCC	-	1594-1571
ID78	GGAAGGCAGGAGCAGAGGCTAAGC	+	1424-1447
ID79	GCCCTTCATCACCTTGCACCCTCC	+	1196-1219
ID87	CACCGCTCCACCTCCCACCCACAC		

name	sequence	sense (+) or antisense (-)	position
ID88	CACCTTCCTCCCGAGACGGGACAG	+	2011-2034
ID89	TCCCCTTGGTGAAGGAGTCCCCAG	-	2457-2434
ID90	TGGGGACTCCTTCACCAAGGGGAG	+	2434-2457
ID91	CCCGAGTGACCGCATGGGTGAAGG	-	3525-3502
ID92	AGGGAATGAGGCTGTGGGTCCACG	+	3729-3752
ID96	GATCGGAGTCCTGTGTGGCACCTT GCAGGC	+	307-336
ID97	GCCCACCAATGTATCTCCTGACCA GGCAGC	-	911-882
ID101	GGCTTGCGCAACAGCTGCACGTGC	-	2265-2242
ID103	AGGGAAGGGCTGTGATGGCGTCGG	-	3356-3333
ID105	ACAGCGGCTGAGGGCGCAGAACAC	-	1747-1724
ID106	GCTGACCCTGGTGGTGCTGAGGCG	+	1759-1783
ID107	CTGTCCCGTCTCGGGAGGAAGGTG	-	2434-2457
ID108	TGGAGGAAGTGCGCTCCCCGCAGC	-	2394-2371
ID109	CAGAGCCCCAGCGGGGCCAGAGAC	-	2802-2779
ID111	GTGTGGGTGGGAGGTGGAGCGGTG	+	3055-3078
ID116*	AGCTCTCACAGTGCGCAGTCAGGG	-	820-797
ID117*	CGCTGCCAGCCATGCTGAGGCTCG	+	1254-1277
ID118*	CCTCCCATCTGTGCCTCTGTAAGG	-	1506-1483
ID128	GGAGCACAGGCCTGGCAGAGCATG GGTGTG		628-657

APPENDIX E. ALLELE SPECIFIC OLIGONUCLEOTIDES FOR DETECTION OF MUTATIONS OR POLYMORPHISMS IN *IDUA*.

For the polymorphism $A_{361}T$;

ID121 ACC CCT TCG CGC AGC GC common A_{361}.

ID122 ACC CCT TCA CGC AGC GC polymorphic T_{361}.

For $D_{349}N$;

ID125 CTG AGA ACG ACA ATG CC normal D_{349}.

ID126 CTG AGA ACA ACA ATG CC mutant N_{349}.

For $t_{1978}g$;

ID129 GGT GAA GGT GGG CCG normal t_{1978}.

ID130 GGT GAA GGC GGG CCG mutant g_{1978}.

For $P_{533}R$;

ID132 GCG GCT GCC GTC GCT TT normal P_{533}.

ID133 GCG GCT GCG GTC GCT TT mutant R_{533}.

For $delG_{1702}$;

ID134 GCT TTT GCT GGT GCA CGT G normal $delG_{1702}$.

ID135 GCT TTT GCT GTG CAC GTG T mutant $delG_{1700}$ - incorrect.

ID150 GCT TTT GCT GGT CAC GTG T mutant $delG_{1702}$.

For $W_{402}X$;

ID136 GCA GCT CTG GGC CGA AG normal W_{402}.

ID137 GCA GCT CTA GGC CGA AG mutant X_{402}.

For $Q_{70}X$;

ID138 GCT GGG ACC AGC AGC TC normal Q_{70}.

ID139 GCT GGG ACT AGC AGC TC mutant X_{70}.

For $g_{1487}a$;

ID140 CTC CCC TGG CCC AGG CT polymorphic normal $c_{1486}t$.

ID141 CTC CCC CGG CCC AGG CT normal g_{1487}.

ID142 CTC CCC CAG CCC AGG CT mutant a_{1487}.

ID151 CTC CCC TAG CCC AGG CT $c_{1486}t$ polymorphism and $g_{1487}a$ mutation combined.

To see if 5-bp included in cDNA from $g_{1487}a$ allele;

ID123 AGT TCA GGA AGC CTT GCA TGG TCA normal IDUA mRNA at exon V/VI junction.

ID124 CCA TGC AAG CCC AGG CTT 5-bp from intron 5 included in mRNA.

For *Eco*47III polymorphism;

ID143 CCG CGC CGC GCT GCT normal polymorphism *Eco*47III absent.

ID144 CCG CGC AGC GCT GCT normal polymorphism *Eco*47III present.

For $R_{89}Q$;

ID147 AGC AGG TCC GGA CCC ACT G normal R_{89}.

ID148 AGC AGG TCC AGA CCC ACT G mutant Q_{89}.

APPENDIX F. PATIENT GENOTYPES.

Skin fibroblast cell line numbers of patient samples held in the Department of Chemical Pathology at the Adelaide Children's Hospital or DNA sample numbers of patient samples from the Willink Biochemical Genetics Unit, Royal Manchester Children's Hospital are given for patients used in this study. Mutations that are known are given for each patient with a phenotype description for patients with only one allele defined or a micellaneous complete genotype.

MPS-I COMPLETE GENOTYPES.

MISCELLANEOUS COMPLETE GENOTYPES.

SF or DNA number	Phenotype	Genotype
283	Severe/intermediate	$P_{533}R$, $W_{402}X$
296	Severe	$Q_{70}X$, $delG_{1702}$
2474/GM1323	Mild	$W_{402}X$, $g_{1487}a$
32-1	Mild	$W_{402}X$, $g_{1487}a$
2662	Intermediate/mild	$R_{89}Q$, $t_{1978}g$
44-1	Mild	$R_{89}Q$, $W_{402}X$
2730	Mild	$R_{89}Q$, $W_{402}X$
2827	Severe	$Q_{70}X$, $D_{349}N$

$W_{402}X$ homozygotes.

SF or DNA number
538
728
985
1062
2732
3125
3061, 64-1
3461
3532
3568
35-1
48-1
63-1

$Q_{70}X$ homozygotes.

SF or DNA number
3296
3481

$Q_{70}X$, $W_{402}X$ compound heterozygotes.

340
1805
2822
3391
49-1
59.1

$P_{533}R$ homozygote.

SF or DNA number	Phenotype
61-1	Severe/ intermediate

HETEROZYGOTES.

$W_{402}X$ heterozygotes.

1917	?
2475/GM1256	Mild
3054/43-1	Severe
3057/55-2	Severe
3153	Severe
3258	Severe
3460	Severe
3480	?
31-1	Severe
33-1	Severe
36-1	Mild
54.1	Severe
60-1	Severe

$Q_{70}X$ Heterozygotes.

686/GM1391	Severe
1250	Severe
1873	Severe
3014	Severe
3414	?
3482	?
3519	Severe
34-1	Severe
52-1	Severe
57-1	Severe

$P_{533}R$ heterozygotes.

3256	Severe/intermediate
CM	Severe/intermediate

$g_{1487}a$ heterozygote.

PM	Mild

TESTED BUT NO DEFINED MUTATIONS.

SF or DNA number	Phenotype
684/GM963	Intermediate
686/GM91391	Severe
1017	Mild to intermediate
1644	Intermediate
2897	Intermediate
3059	Mild
3357	Mild
3413	Intermediate
3478	?
3479	?
3483	?
3561	?
3573	?
45-1	Severe
47-1	Severe
50-1	Intermediate
53-1	Severe
62-2	Mild

BIBLIOGRAPHY

Aerts, J.M.F.G., Schram, A.W., Strijland, A., van Weely, S., Jonsson, L.M.V., Tager, J.M., Sorrell, S.H., Ginns, E.I., Barranger, J.A. and Murray, G.J. (1988) Glucocerebrosidase, a lysosomal enzyme that does not undergo oligosaccharide phosphorylation. *Biochim. Biophys. Acta.* **964:** 303-308.

Allitto, B.A., MacDonald, M.E., Bucan, M. Richards, J., Romano, D., Whaley, W.L., Falcone, B. Ianazzi, J., Wexler, N.S., Wasmuth, J.J., Collins, F.S., Lehrach, H., Haines, J.L. and Gusella, J.F. (1991) Increased recombination adjacent to the Huntington disease-linked *D4S10* marker. *Genomics* **9:** 104-112.

Andrew, S., Theilmann, J., Hedrick, A., Mah, D., Weber, B., and Hayden, M.R. (1992) Nonrandom association between Huntington disease and two loci separated by about 3 Mb on 4p16.3. *Genomics* **13:** 301-311.

Andria, G., Di Natale, P., Del Giudice, E., Strisciuglio, P. and Murino, P. (1979) Sanfilippo B syndrome (MPS III B): mild and severe forms within the same sibship. *Clin. Genet.* **15:** 500-504.

Anson, D.S., Bielicki, J., and Hopwood, J.J. (1992a) Correction of human MPS type I fibroblasts by retroviral mediated gene transfer of the human α-L-iduronidase gene. *Human Gene Therapy* **3:** 371-379.

Anson, D.S., Taylor, J.A., Bielicki, J., Harper, G.S., Peters, C., Gibson, G.J and Hopwood, J.J. (1992b) Correction of human MPS type VI fibroblasts with recombinant *N*-acetylgalatosamine-4-sulphatase. *Biochem. J.* **284:** 789-794.

Ashton, L.J., Brooks, D.A., McCourt, P.A.G., Muller, V.J., Clements, P.R. and Hopwood, J.J. (1992) Immunoquantification and enzyme kinetics of α-L-iduronidase in cultured fibroblasts from normal controls and mucopolysaccharidosis type I patients. *Am. J. Hum. Genet.* **50:** 787-794.

Babarik, A., Benson, P.F., Dean, M.F. and Muir, H. (1974) Chondroitinsulphaturia with α-L-iduronidase deficiency. *Lancet* **2:** 464-465.

Bach, G., Friedman, R., Weissmann, B. and Neufeld, E.F. (1972) The defect in the Hurler and Scheie syndromes: deficiency of α-L-iduronidase. *Proc. Natl. Acad. Sci. USA* **69:** 2048-2051.

Ballabio, A., Pallini, R. and Di Natale, P. (1984) Mucopolysaccharidosis III B: hybridization studies on fibroblasts from a mild case and fibroblasts from severe patients. *Clin. Genet.* **25:** 191-195.

Barnett, R.W. and Erfle, H. (1990) Rapid generation of DNA fragments by PCR amplification of crude, synthetic oligonucleotides. *Nucl. Acids Res.* **18:** 3094.

Barranger, J.A. and Ginns, E.I. (1989) Galactosylceramide lipidoses: Gaucher disease. In Scriver, C.R., Beaudet, A.L., Sly, W.S., and Valle, D. (eds): *The Metabolic Basis of Inherited Disease.* 6th edn., McGraw-Hill, New York. pp. 1677-1697.

Barriocanal, J.G., Bonifacino, J.S., Yuan, L. and Sandoval, I.V. (1986) Biosynthesis, glycosylation, movement through the Golgi system, and transport to lysosomes by an N-linked carbohydrate-independent mechanism of three lysosomal integral membrane proteins. *J. Biol. Chem.* **261:** 16755-16763.

Barton, N.W., Brady, R.O., Dambrosia, J.M. Doppelt, S.H., Hill, S.C., Holder, C.A., Mankin, H.J., Murray, G.J., Zirzow, G.C. and Parker, R.I. (1992) Dose-dependent responses to macrophage-targeted glucocerebrosidase in a child with Gaucher disease. *J. Pediatr.* **120:** 277-280.

Barton, N.W., Furbish, F.S., Murray, G.J., Garfield, M. and Brady, R.O. (1990) Therapeutic response to intravenous infusions of glucocerebrosidase in a patient with Gaucher disease. *Proc. Natl. Acad. Sci. USA* **87:** 1913-1916.

Barton, R.W. and Neufeld, E.F. (1971) The Hurler corrective factor: purification and some properties. *J. Biol. Chem.* **246:** 7773-7779.

Baudhiun, P., Hers, H.G. and Loeb, H. (1964) An electron microscopic and biochemical diagnosis of type II glycogenesis. *Lab. Invest.* **13:** 1139-1143.

Baumkotter, J. and Cantz, M. (1983) Decreased ganglioside neuraminidase activity in fibroblasts from mucopolysaccharidosis patients. Inhibition of the activity in vitro by

sulfated glycosaminoglycans and other compounds. *Biochim. Biophys. Acta.* **761:**163-170.

Beutler, E. (1991) Gaucher's disease. *N. Engl. J. Med.* **325:** 1354-1360.

Beutler, E., Gelbart, T., Kuhl, W., Sorge, J. and West, C. (1991) Identification of the second common Jewish Gaucher disease mutation makes possible population-based screening for the heterozygous state. *Proc. Natl. Acad. Sci. USA* **88:** 10544-10547.

Bielicki, J., Hopwood, J.J., Wilson, P.J. and Anson, D.S. (1992) Recombinant human iduronate-2-sulphatase: correction of MPS-II fibroblasts and characterisation of purified enzyme. *Biochem. J.*, in press.

Birkenmeier, E.H. (1991) Correction of murine mucopolysaccharidosis type VII (MPS VII) by bone marrow transplantation and gene transfer therapy. *Hum. Gene. Ther.* **2:** 113.

Birkenmeier, E.H., Barker, J.E., Vogler, C.A., Kyle, J.W., Sly, W.S., Gwynn, B., Levy, B. and Pegors, C. (1991) Increased life span and correction of metabolic defects in murine mucopolysaccharidosis type VII after syngeneic bone marrow transplantation. *Blood* **78:** 3081-3092.

Birkenmeier, E.H., Davisson, M.T., Beamer, W.G., Ganschow, R.E., Vogler, C.A., Gwynn, B., Lyford, K.A., Maltais, L.M. and Wawrzyniak, C.J. (1989) Murine mucopolysaccharidosis type VII. Characterization of a mouse with β-glucuronidase deficiency. *J. Clin. Invest.* **83:** 1258-1266.

Birnboim, H.C. and Doly, J. (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res.* **7:** 1513-1523.

Brante, G. (1952) Gargolyism: A mucopolysaccharidosis. *Scand. J. Clin. Lab. Invest.* **4:** 43-46.

Braulke, T., Hille., A., Huttner, W.B., Hasilik, A. and von Figura, K. (1987) Sulphated oligosaccharides in human lysosomal enzymes. *Biochem. Biophys. Res. Commun.* **143:** 178-185.

Breider, M.A., Shull, R.M. and Constantopoulos, G.A. (1989) Long-term effects of bone marrow transplantation in dogs with mucopolysaccharidosis I. *Am. J. Pathol.* **134:** 677-692.

Brody, L.C., Mitchell, G.A., Obie, C., Michaud, J., Steel, G., Fontaine, G., Robert, M.F., Sipila, I., Kaiser-Kupfer M. and Valle D (1992) Ornithine delta-aminotransferase mutations in gyrate atrophy. Allelic heterogeneity and functional consequences. *J. Biol. Chem.* **267:** 3302-3307.

Brooks, D.A., Harper, G.S., Gibson, G.J., Ashton, L.J., Taylor, J.A., McCourt, P.A.G., Freeman, C., Clements, P.R, Hoffman, J.W. and Hopwood, J.J. (1992) Hurler syndrome: A patient with abnormally high levels of α-L-iduronidase protein. *Biochem. Med. Metab. Biol.* **47:** 211-220.

Brown, W.J. and Farquhar, M.G. (1987) The distribution of 215-kilodalton mannose 6-phosphate receptors within cis (heavy) and trans (light) Golgi subfractions varies in different cell types. *Proc. Natl. Acad. Sci. USA* **84:** 9001-9005.

Bucan, M., Zimmer, M., Whaley, W.L., Poutska, A., Youngman, S., Allitto, B., Ormondroyd, E., Smith, B., Pohl, T.L., MacDonald, M.E., Bates, G.P., Richards, J., Volinia, S., Gilliam, C., Sedlacek, Z., Collins, F.S., Wasmuth, J.J., Shaw, D., Gusella, J.F., Frischauf, A.-M. and Lehrach, H. (1990) Physical maps of 4p16.3, the area expected to contain the Huntington disease mutation. *Genomics* **6:** 1-16.

Bucher, P. (1990) Weight matrix descriptions of four eukaryotic RNA polymerase II promoter elements derived from 502 unrelated promoter sequences. *J. Mol. Biol.* **212:** 563-578

Buckle, V.J., Fujita, N., Bateson, A.N., Darlinson, M.G. and Barnard, E.A. (1989) Localization of human GABA-A receptor subunit genes to chromosomes 4 and 5. *Cytogenet. Cell Genet.* **51:** 972 (A2205).

Buetow, K.H., Shiang, R., Yang, P., Nakamura, Y., Lathrop, G.M., White, R., Wasmuth, J.J., Wood, S., Berdahl, L.D., Leysens, N.J., Ritty, T.M., Wise, M.W. and Murray, J.C. (1991) A detailed multipoint map of human chromosome 4 provides

evidence for linkage heterogeneity and position-specific recombination rates. *Am. J. Hum. Genet.* **48:** 911-925.

Callen, D.F. (1986) A mouse-human hybrid cell panel for mapping human chromosome 16. *Ann. Genet. (Paris)* **29:** 235-239

Callen, D.F., Hyland, V.L., Baker, E.G., Fratini, A., Gedeon, A.K., Mulley, J.C., Fernandez, K.E.W., Breuning, M.H. and Sutherland, G.R. (1989) Mapping the short arm of human chromosome 16. *Genomics* **4:** 348-354

Capecchi, M.R. (1989) Altering the genome by homologous recombination. *Science* **244:** 1288-1292.

Chaconas, G. and van de Sande, J.H. (1980) 5'-^{32}P labeling of RNA and DNA restriction fragments. *Methods Enzymol.* **65:** 75-88.

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

Clements, P.R., Brooks, D.A., McCourt, P.A.G. and Hopwood, J.J. (1989) Immunopurification and characterization of human α-L-iduronidase with the use of monoclonal antibodies. *Biochem. J.* **259:** 199-208.

Clements, P.R., Brooks, D.A., Saccone, G.T.P. and Hopwood, J.J. (1985a) Human α-L-iduronidase. 1. Purification, monoclonal antibody production, native and subunit molecular mass. *Eur. J. Biochem.* **152:** 21-28.

Clements, P.R., Muller, V. and Hopwood, J.J. (1985b) Human α-L-iduronidase. 2. Catalytic properties. *Eur. J. Biochem.* **152:** 29-34.

Colavita, N., Orazi, C., Fileni, A., Leone, P.C., Ricci, R. and Segni, G. (1986) A further contribution to the knowledge of mucopolysaccharidosis IH/S compound. Presentation of two cases and review of the literature. *Australas. Radiol.* **30:** 142-149.

Constantopoulos, G., Iqbal, K. and Dekaban, A.S. (1980) Mucopolysaccharidosis types IH, IS, II, and IIIA: glycosaminoglycans and lipids of isolated brain cells and other fractions from autopsied tissues. *J. Neurochem.* **34:** 1399-1411.

Conzelmann, E. and Sandhoff, K. (1984) Partial enzyme deficiencies: residual activities and the development of neurological disorders. *Dev. Neurosci.* **6:** 58-71.

Cooper, D.N. and Youssoufian, H. (1988) The CpG dinucleotide and human genetic disease. *Hum. Genet.* **78:** 151-155.

Correll, P.H., Fink, J.K., Brady, R.O., Perry, L.K. and Karlsson, S. (1989) Production of human glucocerebrosidase in mice after retroviral gene transfer into multipotential hematopoietic progenitor cells. *Proc. Natl. Acad. Sci. USA* **86:** 8912-8916.

Cotton, R.G.H. (1989) Detection of single base changes in nucleic acids. *Biochem. J.* **263:** 1-10.

Cox, D.R., Murray, J.C. and Buetow, K.H. (1989) Report of the committee on the genetic constitution of chromosome 4. *Cytogenet. Cell Genet.* **51:** 121-136.

Cullen, B.R. (1988) Expression of a cloned human interlukin-2 cDNA is enhanced by the substitution of a heterologous mRNA leader sequence. *DNA* **7:** 645-650.

Dahms, N.M., Lobel, P. and Kornfeld, S. (1989) Mannose 6-phosphate receptors and lysosomal enzyme targeting. *J. Biol. Chem.* **264:** 12115-12118.

Danes, B.S. and Bearn, A.G. (1965) Hurler's syndrome: demonstration of an inherited disorder of connective tissue in cell culture. *Science* **149:** 987-989.

Davis, L.E., Snyder, R.D., Orth, D.N., Nicholson, W.E., Kornfeld, M. and Seelinger, D.F. (1979) Adrenoleukodystrophy and adrenomyeloneuropathy associated with partial adrenal insufficiency in three generations of a kindred. *Am. J. Med.* **66:** 342-347.

Deisseroth, A., Nienhuis, A., Turner, P., Velez, R., Anderson, W.F., Ruddle, F. , Lawrence, J., Creagan, R. and Kucherlapati, R. (1977) Localization of the human α-globin structural gene to chromosome 16 in somatic cell hybrids by molecular hybridization assay. *Cell* **12:** 205-218.

Devereux, J., Haeberli, P. and Smithies O. (1984). A Comprehensive Set of Sequence Analysis Programs for the VAX. *Nucl. Acids Res.* **12:** 387-395.

DiFerrante, N., Nichols, B.L., Donnelly, P.V., Neri, G., Hrgovcic, R. and Berglund, R.K. (1971) Induced degradation of glycosaminoglycans in Hurler's and Hunter's syndromes by plasma infusion. *Proc. Natl. Acad. Sci. USA* **68:** 303-307.

Dorfman, A. and Lorincz, A.E. (1957) Occurrence of urinary acid mucopolysaccharides in the Hurler syndrome. *Proc. Natl. Acad. Sci. USA* **43:** 443-446.

Eckert, K.A. and Kunkel, T.A. (1990) High fidelity DNA synthesis by the Thermus aquaticus DNA polymerase. *Nucl. Acid Res.* **18:** 3739-3744.

Eperon, I.C. (1986) Rapid preparation of bacteriophage DNA for sequence analysis in sets of 96 clones, using filtration. *Anal. Biochem.* **156:** 406-412.

Erickson, A.H., Walter, P. and Blobel, G. (1983) Translocation of a lysosomal enzyme across the microsomal membrane requires signal recognition particle. *Biochem. Biophys. Res. Commun.* **115:** 275-280.

Estivill, X., Scambler, P.J., Wainwright, B.J., Hawley, K., Frederick, P., Schwartz, M., Baiget, M., Kere, J., Williamson, R. and Farrall, M. (1987) Patterns of polymorphism and linkage disequilibrium for cystic fibrosis. *Genomics* **1:** 257-263.

Feinberg, A.P. and Vogelstein, B (1983) A technique for radiolabeling DNA restriction fragments to high specific activity. *Analyt. Biochem.* **132:** 6-13.

Fink, J.K., Correll, P.H., Perry, L.K., Brady, R.O. and Karlsson, S.T.I. (1990) Correction of glucocerebrosidase deficiency after retroviral-mediated gene transfer into hematopoietic progenitor cells from patients with Gaucher disease. *Proc. Natl. Acad. Sci. USA* **87:** 2334-2338.

Folkman, J. and Shing, Y. (1992) Angiogenesis. *J. Biol. Chem.* **267:** 10931-10934.

Forrest, S.M., Dahl, H.H., Howells, D.W., Dianzani, I. and Cotton, R.G. (1991) Mutation detection in phenylketonuria by using chemical cleavage of mismatch: importance of using probes from both normal and patient samples. *Am. J. Hum. Genet.* **49:** 175-83. Published erratum appears in *Am. J. Hum. Genet.* **50:** 659 (1992).

Frank, H.J.L., Partridge, W.M., Morris, W.L., Rosenfeld, R.J. and Choi, T.B. (1986) Binding and internalization of insulin and insulin-like growth factors by isolated brain microvessels. *Diabetes* **35:** 654-661.

Fransson, L-A. (1989) Heparan sulphate proteoglycans: structure and properties. In Lane, D.W. and Lindahl, U. (eds): *Heparin: Chemical and Biological Properties, Clinical Applications.* Edward Arnold, London. pp. 115-133.

Fratantoni, J.C., Hall, W.C. and Neufeld, E.F. (1968a) Hurler and Hunter syndromes: mutual correction of the defect in cultured fibroblasts. *Science* **162:** 570-572.

Fratantoni, J.C., Hall, W.C. and Neufeld, E.F. (1968b) The defect in Hurler's and Hunter's syndromes: faulty degradation of mucopolysaccharide. *Proc. Natl. Acad. Sci. USA* **60:** 699-706.

Fratantoni, J.C., Hall, W.C. and Neufeld, E.F. (1969) The defect in Hurler's and Hunter's syndromes, II: deficiency of specific factors involved in mucopolysaccharide degradation. *Proc. Natl. Acad. Sci. USA* **64:** 360-366.

Freeman, C., Clements, P.R. and Hopwood, J.J. (1983) Acetyl-CoA:α-glucosaminide *N*-acetyltransferase: partial purification from human liver. *Biochem. Internat.* **6:** 663-671.

Freeman, C., and Hopwood, J.J. (1992a) Catalytic properties and an integrated role in the lysosomal degradation of heparan sulphate. *Biochem. J.* **282:** 899-908.

Freeman, C., and Hopwood, J.J. (1991) Glucuronate-2-sulphatase activity in cultured human skin fibroblast homogenates. *Biochem. J.* **279:** 399-405.

Freeman, C. and Hopwood, J.J. (1992b) Lysosomal degradation of heparin and heparan sulphate. In Lane, D. and Lindahl, U. (eds): *Heparin and related polysaccharides.* 5th edn., Plenum press, U.K. in press.

Friedman, K.D., Rosen, N.L., Newman, P.J. and Montgomery, R.R. (1988) Enzymatic amplification of specific cDNA inserts from lambda gt11 libraries. *Nucl. Acids Res.* **16:** 8718.

Friedmann, T. (1989) Progress towards human gene therapy. *Science* **244:** 1275-1281.

Friedmann, T. and Robin, R. (1972) Gene therapy for human genetic diseases. *Science* **175:** 949-955.

Frischauf, A.-M., Lehrach, H., Poustka, A. and Murray, N (1983) Lambda replacement vectors carrying polylinker sequences. *J. Mol. Biol.* **170:** 827-842.

Frohman, M.A., Dush, M.K. and Martin, G.R. (1988) Rapid production of full-length cDNAs from rare transcripts: Amplification using a single gene-specific oligonucleotide primer. *Proc. Natl. Acad. Sci. USA* **85:** 8998-9002.

Fujibayashi, S., Minami, R., Ishikawa, Y., Wagatsuma, K., Nakao, T. and Tsugawa, S. (1984) Properties of α-L-iduronidase in cultured skin fibroblasts from α-L-iduronidase-deficient patients. *Hum. Genet.* **65:** 268-272.

Fukuda, S., Tomatsu, S., Sukegawa, K., Sasaki, T., Yamada, Y., Kuwahara, T., Okamoto, H., Ikedo, Y., Yamaguchi, S. and Orii, T. (1991) Molecular analysis of mucopolysaccharidosis type VII. *J. Inherit. Metab. Dis.* **14:** 800-804.

Garrod, A.E. (1909) Inborn errors in metabolism. Oxford University press. U.K. (reprinted in 1963 with a supplement by H. Harris).

Gasper, P.W., Thrall, M.A., Wenger, D.A., Macy, D.W., Ham, L., Dornsife, R.E., McBiles, K., Quackenbush, S.L., Kesel, M.L., Gillette, E.L. and Hoover, E.A. (1984) Correction of feline arylsulphate B deficiency (mucopolysaccharidosis VI) by bone marrow transplantation. *Nature* **312:** 467-469.

Gatti, R., Borrone, C., Filocamo, M., Pannone, N. and Di Natale P (1985) Prenatal diagnosis of mucopolysaccharidosis I: A special difficulty arising from an unusually low enzyme activity in mother's cells. *Prenat. Diagn.* **5:** 149-154.

Geurts van Kessel, A.H.M., Tetteroo, P.A.T., von dem Bourne, A.E.G.Kr., Hajemeijer, A. and Bootsma, D. (1983) Expression of human myeloid-associated surface antigens in human-mouse myeloid cell hybrids. *Proc. Natl. Acad. Sci. USA* **80:** 3748-3752

Gibbons, I.R., Asai, D.J., Ching, N.S., Dolecki, G.J., Mocz, G., Phillipson, C.A., Ren, H., Tang, W-J.Y. and Gibbons, B.H. (1991) A PCR procedure to determine the

sequence of large polypeptides by rapid walking through a cDNA library. *Proc. Natl. Acad. Sci. USA* **88:** 8563-8567.

Gompf, R.E., Shull, R.M., Breider, M.A., Scott, J.A. and Constantopoulos, G.C. (1990) Cardiovascular changes after bone marrow transplantation in dogs with mucopolysaccharidosis I. *Am. J. Vet. Res.* **51:** 2054-2060.

Grace, M.E., Graves, P.N., Smith, F.I. and Grabowski, G.A. (1990) Analyses of catalytic activity and inhibitor binding of human acid β-glucosidase by site-directed mutagenesis. Identification of residues critical to catalysis and evidence for causality of two Ashkenazi Jewish Gaucher disease type 1 mutations. *J. Biol. Chem.* **265:** 6827-6835.

Gusella, J.F. (1991) Huntington's disease. *Adv. Hum. Genet.* **20:** 125-150.

Gusella, J.F., Altherr, M.R., McClatchey, A.I., Doucette-Stamm, L.A., Tagle, D., Plummer, S., Groot, N., Barnes, G., Hummerich, H., Collins, F.S., Housman, D.E., lehrach, H., MacDonald, M.E., Bates, G. and Wasmuth, J.J. (1992) Sequence-tagged sites (STSs) spanning 4p16.3 and the Huntington disease candidate region. *Genomics* **13:** 75-80.

Halperin, S.L. and Curtis, G.M. (1942) The genetics of gargoylism. *Amer. J. Ment. Def.* **46:** 298-301.

Haskins, M.E., Desnick, R.J., DiFerrante, N., Jezyk, P.F. and Patterson, D.F. (1979) α-L-iduronidase deficiency in a cat: a model of mucopolysaccharidosis type I. *Ped. Res.* **13:** 1294-1297.

Haskins, M.E., Desnick, R.J., DiFerrante, N., Jezyk, P.F. and Patterson, D.F. (1984) β-Glucuronidase deficiency in a dog: a model of human mucopolysaccharidosis type VII. *Ped. Res.* **18:** 980-984.

Hasilik, A. and von Figura, K. (1981) Oligosaccharides in lysosomal enzymes: distribution of high-mannose and complex oligosaccharides in cathepsin D and β-hexosaminidase. *Eur. J. Biochem.* **121:** 125-129.

Hasilik, A. and von Figura, K. (1984) Processing of lysosomal enzymes in fibroblasts. In Tingle, J.T., Dean, R.S., and Sly, W.S. (eds): *Lysosomes in biology and pathology.* Elsevier, Amsterdamm. **7:** 3-16.

Hasilik, A. and von Figura, K., Conzelmann, E., Nehrkorn, H. and Sandhoff, K. (1982) Lysosomal enzyme precursors in human fibroblasts. Activation of cathepsin D precursor *in vitro* and activity of a β-hexosaminidase A precursor towards ganglioside G_{M2}. *Eur. J. Biochem.* **125:** 317-321.

Hayden, M.R. (1981) Huntington's chorea. Springer, New York.

Hedley, M.L., Forman, J. and Tucker, P.W.(1989) Mutation of 3' splice sites in two different class I genes results in different usage of cryptic splice sites. *J. Immunol.* **143:** 1018-1025.

Hermans, M.M., Kroos, M.A., van Beeumen, J., Oostra, B.A. and Reuser, A.J. (1991) Human lysosomal α-glucosidase. Characterization of the catalytic site. *J. Biol. Chem.* **266:** 13507-13512.

Hers, H.G. (1965) Inborn lysosomal diseases. *Gastroenterology* **48:** 625-633.

Higgins, D.G. and Sharp, P.M. (1988) CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. *Gene* **73:** 237-244.

Hobbs, J.R., Hugh-Jones, K., Barrett, A.J., Byrom, N., Chambers, D., Henry, K., James, D.C.O., Lucas, C.F., Rogers, T.R., Benson, P.F., Tansley, L.R., Patrick, A.D., Mossman, J. and Young, E.P. (1981) Reversal of clinical features of Hurler's disease and biochemical improvement after treatment by bone marrow transplantation. *Lancet* **2:** 709-712.

Hoogerbrugge, P.M., Suzuki, K., Suzuki, K., Poorthuis, B.J., Kobayashi, T., Wagemaker, G. and van Bekkum, D.W. (1988) Donor-derived cells in the central nervous system of twitcher mice after bone marrow transplantation. *Science* **239:** 1035-1038.

Hoogerbrugge, P.M. and Vossen, J.M.J.J. (1990) Bone marrow transplantation in the treatment of lysosomal storage disorders. In Fernandes, J., Saudubray, J-M. and Tada,

K. (eds): *Inborn Metabolic Diseases; Diagnosis and Treatment.* Springer-Verlag, Berlin. pp. 659-670.

Hopwood, J.J. (1989) Enzymes that degrade heparin and heparan sulphate. In Lane, D.W. and Lindahl, U. (eds): *Heparin: Chemical and Biological Properties, Clinical Applications.* Edward Arnold, London. pp. 190-229.

Hopwood, J.J. and Elliott, H. (1985) Urinary excretion of sulphated *N*-acetylhexosamines in patients with various mucopolysaccharidoses. *Biochem. J.* **229:** 579-586.

Hopwood, J.J. and Harrison, J.R. (1982) High-resolution electrophoresis of urinary glycosaminoglycans: an improved screening test for the mucopolysaccharidoses.. *Anal. Biochem.* **119:** 120-127.

Hopwood, J.J. and Morris, C.P. (1990) The mucopolysaccharidoses: Diagnosis, molecular genetics and treatment. *Molec. Biol. Med.* **7:** 381-404.

Hopwood, J.J. and Muller, V. (1979) Biochemical discrimination of Hurler and Scheie syndromes. *Clin. Sci.* **57:** 265-272.

Hopwood, J.J. and Muller, V. (1982) Diagnostic enzymology of α-L-iduronidase with special reference to a sulphated disaccharide derived from heparin. *Clin. Sci.* **62:** 193-201.

Hopwood, J.J., Muller, V. and Pollard, A.C. (1979a) Post- and pre-natal assessment of α-L-iduronidase deficiency with a radiolabelled natural substrate. *Clin. Sci.* **56:** 591-599.

Hopwood, J.J., Muller, V., Smithson, A. and Baggett, N. (1979b) A fluorometric assay using 4-methylumbelliferyl α-L-iduronide for the estimation of α-L-iduronidase activity and the detection of Hurler and Scheie syndromes. *Clin. Chim. Acta.* **92:** 257-265.

Huang, J. and Van der Ploeg, L.H. (1991) Requirement of a polypyrimidine tract for trans-splicing in trypanosomes: discriminating the PARP promoter from the immediately adjacent 3' splice-acceptor site. *EMBO J.* **10:** 3877-3885.

Hunter, C. (1917) A rare disease in two brothers. *Proc. R. Soc. Med.* **10:** 104-115.

Hurler, G. (1919) Uber eninen type multipler abatungen, vorwiegend am skelettsystem. *Kinderheilkd* **24:** 220-234.

Huynh, T.V., Young, R.A. and Davis, R.W. (1984) Construction and screening of DNA libraries in λgt10 and λgt11. In D. Glover (ed.): *DNA cloning techniques: A practical approach.* IRL press, Oxford. pp. 49-78.

Jellinger, K., Paulus, W., Grisold, W. and Paschke, E. (1990) New phenotype of adult α–L-iduronidase deficiency (mucopolysaccharidosis I) masquerading as Friedreich's ataxia with cardiopathy. *Clin. Neuropathol.* **9:** 163-169.

Jensen, O.A., Pedersen, C., Schwartz, M., Vestermark, S. and Warburg, M. (1978) Hurler/Scheie phenotype. Report of an inbred sibship with tapeto-retinal degeneration and electron-microscopie examination of the conjuctiva. *Ophthalmologica* **176:** 194-204.

Jervis, G.A. (1950) Gargoylism: study of ten cases with emphasis on the formes fruste. *Arch. Neur. Psychiat.* **63:** 681-712.

Jezyk, P.F., Haskins, M.E., Patterson, D.F., Mellman, W.J. and Greenstein, M. (1977) Mucopolysaccharidosis in a cat with arylsulphatase B deficiency: a model of Maroteaux-Lamy syndrome. *Science* **198:** 834-836.

Jin, W.D., Jackson, C.E., Desnick, R.J. and Schuchman, E.H. (1992) Mucopolysaccharidosis type VI: identification of three mutations in the arylsulfatase B gene of patients with the severe and mild phenotypes provides molecular evidence for genetic heterogeneity. *Am. J. Hum. Genet.* **50:** 795-800.

Kaibara, N., Eguchi, M., Shibata, K. and Takagishi, K. (1979) Hurler-Scheie phenotype: a report of two pairs of inbred sibs. *Hum. Genet.* **53:** 37-41.

Kaibara, N., Katsuki, I., Hotokebuchi, T., Takagishi, K. and Kure, T. (1983) Hurler-Scheie phenotype with parental consanguinity. Report of an additional case supporting the concept of genetic heterogeneity. *Clin. Orthop.* **175:** 233-236.

Kampine, J.P., Brady, R.O., Kanfer, J.N., Feld, M. and Shapiro D (1966) Diagnosis of Gaucher`s disease and Niemann-Pick disease with small samples of venous blood. *Science* **155:** 86-88.

Kelly, W,R., Clague, A.E., Barns, R.J., Bate, M.J. and MacKay, B.M. (1983) Canine α-L-fucosidosis: A storage disease of springer spaniels. *Acta. Neuropathol.* **60:** 9-13.

Kingston, R.E. (1987) Preparation of Poly(A)$^+$ RNA. In Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Smith, J.A., Seidman, J.G. and Struhl, K. (eds): *Current Protocols in Molecular Biology*. John Wiley and Sons, New York. pp. 4.5.1-4.5.3.

Klein, U., Kresse, H. and von Figura, K. (1978) Sanfilippo syndrome type C: deficiency of acetyl-CoA: α-glucosaminide *N*-acetyltransferase in skin fibroblasts. *Proc Natl Acad Sci USA* **75:** 5185-5189.

Knudson, A.GJr., DiFerrante, N. and Curtis, J.E. (1971) Effect of leukocyte transfusion in a child with type II mucopolysaccharidosis. *Proc Natl Acad Sci USA* **68:** 1738-1741.

Kohn, D.B., Nolta, J.A., Weinthal, J., Bahner, I., Yu, X.J., Lilley, J. and Crooks, G.M. (1991) Toward gene therapy for Gaucher disease. *Hum. Gene. Ther.* **2:** 101-105.

Koizumi, T., MacDonald, M.E., Bucan, M., Hopwood, J.J., Morris, C.P., Scott, H.S., Gusella, J.F., and Nadeau, J.H. (1992) Linkage, but not gene order, of homologous loci, including α-L-iduronidase (*Idua*), is conserved in the Huntington disease region of the mouse and human genomes. *Mammalian Genome*. **3:** 23-27.

Kopin, A.S., Wheeler, M.B., Nishitani, J., McBride, E.W., Chang, T-M., Chey, W.Y., and Leiter, A.B. (1991) The secretin gene: Evolutionary history, alternative splicing, and developmental regulation. *Proc Natl Acad Sci USA* **88:** 5335-5339.

Kornfeld, R. and Kornfeld, S. (1985) Assembly of asparagine-linked oligosaccharides. *Annu Rev. Biochem.* **54:** 631-664.

Kornfeld, S. (1987) Trafficking of lysosomal enzymes. *FASEB. J.* **1:** 462-468.

Kornfeld, S. and Mellman, I. (1989) The biogenesis of lysosomes. *Annu Rev. Cell. Biol.* **5:** 483-525.

Krivit, W., Whitley, C.B., Chang, P-N., Shapiro, E., Belani, K.G., Snover, D., Summers, C.G. and Blazar, B. (1990) Lysosomal storage diseases treated by bone

marrow transplantation: Review of 21 patients. In Johnson, F.L. and Pochedly, C. (eds): *Bone Marrow Transplantation in Children.* Raven Press, New York. pp. 261-287.

Kyle, J.W., Birkenmeier, E.H., Gwynn, B., Volger, C., Hoppe, P.C., Hoffman, J.W. and Sly, W.S. (1990) Correction of murine mucopolysaccharidosis type VII by a human β-glucuronidase transgene. *Proc. Natl. Acad. Sci. USA* **87:** 3914-3918.

Lee, C.C., Wu, X., Gibbs, R.A., Cook, R.G., Muzny, D.M. and Caskey, C.T. (1988) Generation of cDNA probes directed by amino acid sequence: cloning of urate oxidase. *Science* **239:** 1288-1291.

Le Gendre, N. and Matsudaira, P. (1988) Direct protein microsequencing from Immobilon-P Transfer Membrane. *Biotechniques* **6:** 154-159.

Leisti, J., Rimoin, D.L., Kaback, M., Shapiro, L.J. and Matalon, R. (1976) Allelic mutations in the mucopolysaccharidoses. *Birth Defects* **12:** 81-91.

Lemansky, P., Hasilik, A., von Figura, K., Helmy, S., Fishman, J., Fine, R.E., Kedersha, N.L. and Rome, L.R. (1987) Lysosomal enzyme precursors in coated vesicles derived from the exocytic and endocytic pathways. *J. Cell. Biol.* **104:** 1743-1748.

Lim, B., Apperley, J.F., Orkin, S.H. and Williams, D.A. (1989) Long-term expression of human adenosine deaminase in mice transplanted with retrovirus-infected hematopoietic stem cells. *Proc. Natl. Acad. Sci. USA* **86:** 8892-8896.

Lindahl, U., Feingold, D.S., and Roden, L. (1986) Biosynthesis of heparin. *Trends Biochem. Sci.* **11:** 221-225.

Litjens, T., Morris, C.P., Gibson, G.J., Beckman, K.R. and Hopwood, J.J. (1991) Human *N*-acetylgalatosamine-4-sulphatase: Protein maturation and isolation of genomic clones. *Biochem. Internat.* **24:** 209-215.

Lodish, H.F. (1988) Transport of secretory and membrane glycoproteins from the rough endoplasmic reticulum to the Golgi. A rate-limiting step in protein maturation and secretion. *J. Biol. Chem.* **263:** 2107-2110.

Lowry, K.B. and Renwick, D.H.G. (1971) Relative frequencies of the Hurler and Hunter syndromes. (letter) *N. Engl. J. Med.* **284:** 221.

Lusis, A.J. and Paigen, K. (1977) Relationships between levels of membrane-bound glucuronidase and the associated protein egasyn in mouse tissues. *J. Cell. Biol.* **73:** 728-735.

Luskey, B.D., Lim, B., Apperley, J.F., Orkin, S.H. and Williams, D.A. (1990) Gene transfer into murine hematopoietic stem cells and bone marrow stromal cells. *Ann. N. Y. Acad. Sci.* **612:**398-406.

Luthi, E., Love, D.R., McAnulty, J., Wallace, C., Caughey, P.A., Saul, D. and Bergquist, P.L. (1990) Cloning, sequence analysis and expression of genes encoding xylan-degrading enzymes from the thermophile *Caldocellum saccharolyticium. Appl. Env. Microbiol.* **56:** 1017-1024.

MacDonald, M.E., Anderson, M.A., Lockyer, J.A., Milstein, S., Hobbs, W.J., Faryniarz, A.G., Kaufman, S., Ledley, F.D., Woo, S.L.C. and Gusella, J.F. (1987) Physical and genetic localization of quinoid dihydropteridine reductase gene (QDPR) on short arm of chromosome 4. *Somat. Cell Mol. Genet.* **13:** 569-574.

MacDonald, M.E, Cheng, S.V., Zimmer, M., Haines, J.L., Poustka, A., Allitto, B., Smith B., Whaley, W.L., Romano, D.M., Jagadeesh, J., Myers, R.H., Lehrach, H., Wasmuth, J.J., Frischauf, A.-M. and Gusella, J.F. (1989) Clustering of multiallele DNA markers near the Huntington disease gene. *J. Clin. Invest.* **84:** 1013-1016.

MacDonald, M.E., Lin, C., Srinidhi, L., Bates, G., Altherr, M., Whaley, W.L., Lehrach, H., Wasmuth, J. and Gusella, J.F. (1991a) Complex patterns of linkage disequilibrium in the Huntington disease region. *Am. J. Hum. Genet.* **49:** 723-734.

MacDonald, M.E., Novelletto, A., Lin, C., Tagle, D., Barnes, G., Bates, G., Taylor, S., Allitto, B., Altherr, M., Myers, R., Lehrach, H., Collins, F.S., Wasmuth, J., Frontali, M. and Gusella, J.F. (1992) The Huntington's disease candidate region exhibits many different haplotypes. *Nature Genet.* **1:** 99-103.

MacDonald, M.E., Scott, H.S., Whaley, W.L., Pohl, T., Wasmuth, J.J., Lehrach, H., Morris, C.P., Frischauf, A.-M., Hopwood, J.J. and Gusella, J.F. (1991b) Huntington

disease-linked locus *D4S111* exposed as the α-L-iduronidase gene. *Somatic Cell Mol. Genet.* **17:** 421-425.

Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular cloning: A laboratory manual. Cold Spring Harbour Laboratory. pp. 170.

Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular cloning: A laboratory manual. Cold Spring Harbour Laboratory. pp. 371-373.

Markwell, M.K., Haas, S.M., Bilber, L.L. and Tolbert, N.E. (1978) A modification of the Lowry procedure to simplify protein determination in membrane and lipoprotein samples. *Anal. Biochem.* **87:** 211-222.

Martin, F.H., Castro, M.M., Aboul-ela, F. and Tinoco, I. Jr (1985) Base pairing involving deoxyinosine: implications for probe design. *Nucl. Acids Res.* **13:** 8927-8938.

Masuno, M., Tomatsu, S., Sukegawa, K. and Orii, T. (1990) Non-existence of a tight association between a 444leucine to proline mutation and phenotypes of Gaucher disease: high frequency of a NciI polymorphism in the non-neuronopathic form. *Hum. Genet.* **84:** 203-206.

Masuyama, T., Gojobori, T., Aota, S. and Ikemura, T. (1986). Codon usage tabulated from the Genbank genetic sequence data. *Nucl. Acids. Res.* **14:** r151-r197.

Matalon, R. and Deanching, M. (1977) The enzymic basis for the phenotypic variation of Hurler and Scheie syndromes. *Pediatr. Res.* **11:** 519.

Matalon, R., Deanching, M. and Omura, K. (1983) Hurler, Scheie, and Hurler-Scheie `Compound': residual activity of α-L-iduronidase towards natural substrates suggesting allelic mutations. *J. Inher. Metab. Dis.* **6[suppl]:** 133-134.

Matalon, R. and Dorfman, A. (1972) Hurler's syndrome, an α-L-iduronidase deficiency. *Biochem. Biophys. Res. Commun.* **47:** 959-964.

Matsudaira, P. (1987) Sequence from picomole quantities of proteins electroblotted onto polyvinylidene difluoride membranes. *J. Biol. Chem.* **262:** 10035-10038.

McKusick, V.A. (1972) The mucopolysaccharidoses. In *Heritable disorders of connective tissue.* 4th edn., C.V. Mosby, Saint Louis. pp. 521-686.

McKusick, V.A., Howell, R.R., Hussels, I.E., Neufeld, E.F. and Stevenson, R.E. (1972) Allelism, non-allelism, and genetic compounds among the mucopolysaccharidoses. *Lancet* **1:** 993-996.

McKusick, V.A., Kaplan, D., Wise, D., Hanley, W.B., Suddarth, S.B., Sevick, M.E. and Maumenee, A.E. (1965) The genetic mucopolysaccharidoses. *Medicine* **44:** 445-483.

McKusick, V.A. and Neufeld, E.F. (1983) The mucopolysaccharide storage diseases. In Stanbury, J.B., Wyngaarden, J.B., Fredrickson, D.S., Goldstein, J.L. and Brown, M.S. (eds): *The Metabolic Basis of Inherited Disease.* 5th edn., McGraw-Hill, New York. pp. 751-777.

McKusick, V.A., Neufeld, E.F. and Kelly, T.E. (1978) The mucopolysaccharide storage diseases. In Stanbury, J.B., Wyngaarden, J.B. and Fredrickson, D.S. (eds): *The Metabolic Basis of Inherited Disease.* 4th edn., McGraw-Hill, New York. pp. 1282-1307.

Mendel, G. (1866) Experiments in plant hybridization. Reprinted in Peter, J.A. (ed.): *Classic papers in genetics.* Prentice-Hall, Englewood Cliffs, N.J. (1959). pp. 1-20.

Minami, R., Fujibayashi, S., Igarashi, C., Ishikawa, Y., Wagatsuma, K., Nakao, T. and Tsugawa, S. (1984) Heparan sulfate and dermatan sulfate from the liver of a patient with Hurler syndrome: high performance liquid chromatoghraphy of their degradation products after incubation with α-L-iduronidase deficient fibroblasts. *Clin. Chim. Acta.* **137:** 179-187.

Mizusawa, S., Nishimura, S. and Seela, F. (1986) Improvement of the dideoxy chain termination method of DNA sequencing by the use of deoxy-7-deazaguanosine triphosphate in place of dGTP. *Nucleic Acids Res.* **14:** 1319-1324.

Morreau, H., Galjart, N.J., Gillemans, N., Willemsen, R., van Horst, T.J. and d'Azzo, A. (1989) Alternative splicing of β-galactosidase mRNA generates the classic lysosomal enzyme and a β-galactosidase-related protein. *J. Biol. Chem.* **264,** 20655-20663.

Moskowitz, S.M., Dlott, B., Chuang, P.D. and Neufeld, E. F. (1992) Cloning and expression of cDNA encoding the human lysosomal enzyme, α-L-iduronidase. *FASEB J.* **6:** A77.

Mueller, P.R. and Wold, B. (1989) *In vivo* footprinting of a muscle specific enhancer by ligation mediated PCR. *Science* **246:** 780-786.

Muller, V.J. and Hopwood, J.J. (1984) α-L-Iduronidase deficiency in mucopolysaccharidosis type I against a radiolabelled sulfated disaccharide substrate derived from dermatan sulfate. *Clin. Genet.* **26:** 414-421.

Muraki, M., Harata, K., Hayashi, Y., Machida, M., Jigami, Y. (1991) The importance of precise positioning of negatively charged carboxylate in the catalytic action of human lysozyme. *Biochim. Biophys. Acta.* **1079:** 229-37.

Murray, J.C. and Van Ommen, G.J.B. (1991) Report of the committee on the genetic constitution of chromosome 4. *Cytogenet. Cell Genet.* **58:** 231-260.

Murray, V. (1989) Improved double stranded DNA sequencing using the linear polymerase chain reaction. *Nucleic Acids Res.* **17:** 8889.

Myerowitz, R. and Costigan, F.C. (1988) The major defect in Ashkenazi Jews with Tay-Sachs disease is an insertion in the gene for the alpha-chain of β-hexosaminidase. *J. Biol. Chem.* **263:** 18587-18589.

Myerowitz, R. and Neufeld, E.F. (1981) Maturation of α-L-iduronidase in cultured human fibroblasts. *J. Biol. Chem.* **256:** 3044-3048.

Myerowitz, R. and Proia, R.L. (1984) cDNA clone for the alpha-chain of human β-hexosaminidase: deficiency of alpha-chain mRNA in Ashkenazi Tay-Sachs fibroblasts. *Proc. Natl. Acad. Sci. USA* **81:** 5394-5398.

Navon, R. and Proia, R.L. (1989) The mutations in Ashkenazi Jews with adult G_{M2} gangliosidosis, the adult form of Tay-Sachs disease. *Science* **243:** 1471-1474.

Nelson, C., Rabb, H. and Arnaout, M.A. (1992) Genetic cause of leukocyte adhesion molecule deficiency. Abnormal splicing and a missense mutation in a conserved region

of CD18 impair cell surface expression of beta 2 integrins. *J. Biol. Chem.* **267:** 3351-3357.

Nelson, P.V., Carey, W.F., Morris, C.P. and Pollard, A.C. (1989) Cystic fibrosis: prenatal diagnosis and carrier detection by DNA analysis. *Med. J. Aust.* **151:** 126-131.

Neufeld, E.F. (1991) Lysosomal storage diseases. *Annu. Rev. Biochem.* **60:** 257-280.

Neufeld, E.F. and Muenzer, J. (1989) The mucopolysaccharidoses. In Scriver, C.R., Beaudet, A.L., Sly, W.S., and Valle, D. (eds): *The Metabolic Basis of Inherited Disease.* 6th edn., McGraw-Hill, New York. pp. 1565-1587.

Nja, A. (1946) Sex linked type of gargoylism. *Acta. Paediatr.* **33:** 267-286.

Nolan, C.M and Sly, W.S. (1989) I-cell disease and psuedo-Hurler polydystrophy: Disorders of lysosomal enzyme phosphorylation and localization. In Scriver, C.R., Beaudet, A.L., Sly, W.S., and Valle, D. (eds): *The Metabolic Basis of Inherited Disease*. 6th edn., McGraw-Hill, New York. pp. 1589-1601.

Occhiodoro, T., Beckmann, K.R., Morris, C.P. and Hopwood, J.J. (1989) Human α-L-fucosidase: complete coding sequence from cDNA clones. *Biochem. Biophys. Res. Commun.* **164:** 439-446.

Occhiodoro, T., Morris, C.P., Hopwood, J.J. and Anson, D.S. (1992) Correction of α-L-fucosidase deficiency in fucosidosis fibroblasts by retroviral vector-mediated gene transfer. *Human Gene Therapy* **3:** 365-369.

O'Dowd, B.F., Quan, F., Willard, H.F., Lamhonwah, A.M., Korneluk, R.G., Lowden, J.A., Gravel, R.A. and Mahuran, D.J. (1985) Isolation of cDNA clones coding for the beta subunit of human β-hexosaminidase. *Proc. Natl. Acad. Sci. USA* **82:** 1184-1188.

Ohtsuka, E., Matsuki, S., Ikehara, M., Takahashi, Y. and Matsubara, K (1985) An alternative approach to deoxyoligonucleotides as hybridization probes by insertion of deoxyinosine at ambiguous codon positions. *J. Biol. Chem.* **260:** 2605-2608.

Orgul, S., Daicker, B. and Kain, H.L. (1991) Simultaneous corneal transplantation in mucopolysaccharidosis. *Klin. Monatsbl. Augenheilkd.* **198:** 430-432.

Oshima, A., Kyle, J.W., Miller, R.D., Hoffman, J.W., Powell, P., Grubb, J.H., Sly, W.S., Tropak, M., Guise, S. and Gravel, R.A. (1987) Cloning, sequencing, and expression of cDNA for human β-glucuronidase. *Proc. Natl. Acad. Sci. USA* **84,** 685-689.

Oshima, A., Tsuji, A., Nagao, Y., Sakuraba, H. and Suzuki, Y. (1988) Cloning, sequencing, and expression of cDNA for human β-galactosidase. *Biochem. Biophys. Res. Commun.* **157:** 238-244.

Oshima, A., Yoshida, K., Shimmoto, M., Fukuhara, Y., Sakuraba, H. and Suzuki, Y. (1991) Human β-galactosidase gene mutations in morquio B disease. *Am. J. Hum. Genet.* **49:** 1091-1093.

Palmer, T.D., Rosman, G.J., Osborne, W.R. and Miller, A.D. (1991) Genetically modified skin fibroblasts persist long after transplantation but gradually inactivate introduced genes. *Proc. Natl. Acad. Sci. USA* **88:** 1330-1334.

Patterson, D., Berger, R., Bleskan, J., Vannais, D. and Davidson, J. (1992) A single base change at a splice-acceptor site leads to a truncated CAD protein in Urd-A mutant Chinese hamster ovary cells. *Somat. Cell. Mol. Genet.* **18:** 65-75.

Pedersen, C., Schwartz, M., Guttler, F. and Hobolth, N (1979) Prenatal diagnosis of the Hurler syndrome. *Danish Medical Bulletin* **26:** 357-359.

Penotti, F.E. (1991) Human pre-mRNA splicing signals. *J. Theor. Biol.* **150:** 385-420.

Peters, C., Braun, M., Weber, B., Wendland, M., Schmidt, B., Pohlmann, R., Waheed, A. and von Figura, K. (1990) Targeting of a lysosomal membrane protein: a tyrosine-containing endocytosis signal in the cytoplasmic tail of lysosomal acid phosphatase is necessary and sufficient for targeting to lysosomes. *EMBO J.* **9:** 3497-3506.

Peters, C., Rommerskirch, W., Modaressi, S. and von Figura, K. (1991) Restoration arylsulphatase B activity in human mucopolysaccharidosis type VI fibroblasts by retroviral mediated gene transfer. *Biochem. J.* **276:** 499-504.

Peters, C., Schmidt, B., Rommerskirch, W., Rupp, K., Zuhlsdorf, M., Vingron, M., Meyer, H.E., Pohlmann, R. and von Figura, K. (1990) Phylogenetic conservation of the

arylsulphatases: cDNA cloning and expression of human arylsulphatase B. *J. Biol. Chem.* **265:** 3374-3381.

Pohl, T.M., Zimmer, M., MacDonald, M.E., Smith, B., Bucan, M., Poutska, A., Volinia, S., Searle, S., Zehetner, G. and Wasmuth, J.J. (1988) Construction of a *Not* I linking library and isolation of new markers close to the Huntington's disease gene. *Nucl. Acids Res.* **16:** 9185-9198.

Pohlmann, R., Krentler, C., Schmidt, B., Schroder, W., Lorkowski, G., Culley, J., Mersmann, G., Geier, C., Waheed, A., Gottschalk, S., Grzeschik, K-.H., Hasilik, A. and von Figura, K. (1988) Human lysosomal acid phosphatase: cloning, expression and chromosomal assignment. *EMBO J.* **7:** 2343-2350.

Poole, B.P. (1982) Developmental role of hyaluronate. *Conn. Tiss. Res.* **10:** 93-100.

Pyeritz, R.E. (1983) Cardiovascular manifestations of heritable disorders of connective tissue. *Prog. Med. Genet.* **5:** 191-302.

Quintern, L.E., Schuchman, E.H., Levran, O., Suchi, M., Ferlinz, K., Reinke, H., Sandhoff, K. and Desnick, R.J. (1989) Isolation of cDNA clones encoding human acid sphingomyelinase: occurrence of alternatively processed transcripts. *EMBO J.* **8,** 2469-2473.

Reed, K.C. and Mann, D.A. (1985) Rapid transfer of DNA from agarose gels to nylon membranes. *Nucl. Acids Res.* **13:** 7207-7221.

Robberson, B.L., Cote, G.J. and Berget, S.M. (1990) Exon definition may facilitate splice site selection in RNAs with multiple exons. *Mol. Cell. Biol.* **10:** 84-94.

Robertson, D.A., Freeman, C., Nelson, P.V., Morris, C.P. and Hopwood, J.J. (1988) Human glucosamine 6-sulphatase cDNA reveals homology with steroid sulphatase. *Biochem. Biophys. Res. Commun.* **157:** 218-224.

Roden, L. (1980) Structure and metabolism of connective tissue proteoglycans. In Lennarz, W.J. (ed.): *The biochemistry of glycoproteins and proteoglycans.* Plenum press, New York, pp. 267-371.

Rome, L.H., Garvin, A.J. and Neufeld, E.F. (1978) Human kidney α-L-iduronidase: purification and characterization. *Arch. Biochem. Biophys.* **189:** 344-353.

Rossi, A.M., Thijssen, J.C., Tates, A.D., Vrieling, H., Natarajan, A.T., Lohman, P.H. and van Zeeland, A.A. (1990) Mutations affecting RNA splicing in man are detected more frequently in somatic than in germ cells. *Mutat. Res.* **244:** 353-357.

Roubicek, M., Gehler, J. and Spranger, J. (1985) The clinical spectrum of α-L-iduronidase deficiency. *Am. J. Med. Genet.* **20:** 471-481.

Ruckenstein, M.J., Macdonald, R.E., Clarke, J.T. and Forte, V. (1991) The management of otolaryngological problems in the mucopolysaccharidoses: a retrospective review. *J. Otolaryngol.* **20:** 177-183.

Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B. and Erlich, H.A. (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239:** 487-491.

Sakuraba, H. Eng, C.M., Desnick, R.J. and Bishop, D.F. (1992) Invariant exon skipping in the human α-galactosidase A pre-mRNA: A g^{+1} to t substitution in a 5'-splice site causing Fabry disease. *Genomics* **12:** 643-650.

Salleeba, J.A. and Cotton, R.G.H (1992) Chemical cleavage of mismatch to detect mutations. *Methods Enzymol.*, in press.

Sandhoff, K., Conzelmann, E., Neufeld, E.F., Kaback, M.M. and Suzuki, K. (1989) The G_{M2} gangliosidosis. In Scriver, C.R., Beaudet, A.L., Sly, W.S., and Valle, D. (eds): *The Metabolic Basis of Inherited Disease.* 6th edn., McGraw-Hill, New York. pp. 1565-1587.

Sando, G.N. and Neufeld, E.F. (1977) Recognition and receptor mediated uptake of a lysosomal enzyme, α-L-iduronidase, by cultured human fibroblasts. *Cell* **12:** 612-627.

Sanger, F., Nicklen, S. and Coulson, A.R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* **74:** 5463-5467.

Sawadogo M. and Van Dyke, M.W. (1991) A rapid method for the purification of deprotected oligodeoxynucleotides. *Nucl. Acids Res.* **19:** 674.

Scheie, H.G., Hambrick, G.W. and Barness, L.A. (1962) A newly recognised forme fruste of Hurler's disease (gargoylism). *Amer. J. Ophthal.* **53:** 753-769.

Schuchman, E.H. and Desnick, R.J. (1988) Mucopolysaccharidosis type I subtypes: Presence of immunologically cross reactive material and *in vitro* enhancement of the residual α-L-iduronidase activities. *J. Clin. Invest.* **81:** 98-105.

Schuchman, E.H., Astrin, K.H., Aula, P. and Desnick, R.J. (1984a) Regional assignment of the structural gene for human α-L-iduronidase. *Proc. Natl. Acad. Sci. USA* **81:** 1169-1173.

Schuchman, E.H., Guzman, N.A. and Desnick, R.J. (1984b) Human α-L-iduronidase. I. Purification and characterization of the high uptake (higher molecular weight) and low uptake (processed) forms. *J. Biol. Chem.* **259:** 3132-3140.

Schuchman, E.H., Guzman, N.A., Takada, G. and Desnick, R.J. (1984c) Human α-L-iduronidase. II. Comparative biochemical and immunologic properties of the purified low and high uptake forms. *Enzyme* **31:** 166-175.

Schuchman, E.H., Jackson, C.E., and Desnick, R.J. (1990) Human arylsulphatase B: MOPAC cloning, nucleotide sequence of a full-length cDNA, and regions of amino acid identity with arylsulphatase A and C. *Genomics* **6:** 149-158.

Schuchman, E.H., Suchi, M., Takahashi, T., Sandhoff, K. and Desnick, R.J. (1991) Human acid sphingomyelinase. Isolation, nucleotide sequence and expression of the full-length and alternatively spliced cDNAs. *J. Biol. Chem.* **266:** 8531-8539.

Scott, H.S., Anson, D.A., Orsborn, A.M., Nelson, P.V., Clements, P.R., Morris, C.P. and Hopwood, J.J (1991a) Human α-L-iduronidase: cDNA isolation and expression. *Proc. Natl. Acad. Sci. USA* **88:** 9695-9699.

Scott, H.S., Ashton, L.J., Eyre, H.J., Baker, E., Brooks, D.A., Callen, D.F., Sutherland, G.R., Morris, C.P. and Hopwood, J.J. (1990) Chromosomal localization of the human α-L-iduronidase gene (*IDUA*) to 4p16.3. *Am. J. Hum. Genet.* **47:** 802-807.

Scott, H.S., Guo, X-H., Hopwood, J.J. and Morris, C.P. (1992a) Structure and sequence of the human α-L-iduronidase gene. *Genomics* **13:** 1311-1313.

Scott, H.S., Litjens, T., Hopwood, J.J., and Morris, C.P. (1992b) A common mutation for mucopolysaccharidosis type I associated with a severe Hurler syndrome phenotype. *Human Mutation*. **1:** 103-108.

Scott, H.S., Litjens, T., Hopwood, J.J. and Morris, C.P. (1992c) PCR detection of two RFLP's in exon I of the α-L-iduronidase (IDUA) gene. *Human Genet.*, in press.

Scott, H.S., Litjens, T., Nelson, P.V., Brooks, D.A., Hopwood, J.J., and Morris, C.P. (1992d) α-L-Iduronidase mutations ($Q_{70}X$ and $P_{533}R$) associate with a severe Hurler syndrome phenotype. *Human Mutation*, in press.

Scott, H.S., Nelson, P.V., Hopwood, J.J. and Morris, C.P. (1991b) PCR of a Kpn I RFLP in the α-L-iduronidase gene. *Nucl. Acids Res.* **19:** 5796.

Scott, H.S., Nelson, P.V., Hopwood, J.J. and Morris, C.P. (1991c) PCR of a VNTR linked to mucopolysaccharidoses type one and Huntington disease. *Nucl. Acids Res.* **19:** 6348.

Scott, H.S., Nelson, P.V., MacDonald, M.E., Gusella, J.F., Hopwood, J.J. and Morris, C.P. (1992e) An 86 bp VNTR within IDUA is the basis of the *D4S111* polymorphic locus. *Genomics*, in press.

Scriver, C.R., Beaudet, A.L., Sly, W.S., and Valle, D. (eds) (1989) *The Metabolic Basis of Inherited Disease*. 6th edn., McGraw-Hill, New York.

Selden, R.F. (1987) Analysis of RNA by Northern hybridization. In Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Smith, J.A., Seidman, J.G. and Struhl, K. (eds): *Current Protocols in Molecular Biology*. Wiley, New York. pp. 4.9.1-4.9.8.

Selden, R.F. and Chory, J. (1987) Isolation and purification of large DNA restriction fragments from agarose gels. In Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Smith, J.A., Seidman, J.G. and Struhl, K. (eds): *Current Protocols in Molecular Biology*. Wiley, New York. pp. 2.6.1-2.6.8.

Shapiro, L.J., Hall, C.W., Leder, I.J. and Neufeld, E.F. (1976) The relationship of α-L-iduronidase and Hurler corrective factor. *Arch. Biochem. Biophys.* **172:** 156-161.

Shull, R.M., Breider, M.A. and Constantopoulos, G.A. (1988) Long-term neurological effects of bone marrow transplantation in a canine lysosomal storage disease. *Pediatr. Res.* **24:** 347-352.

Shull. R.M., Hastings, N.E., Selcer, R.R., Jones, J.B., Smith, J.R., Cullen, W.C. and Constantopoulous, G. (1987) Bone marrow transplantation in canine mucopolysaccharidosis type I. Effects within the central nervous system. *J. Clin. Invest.* **79:** 435-443.

Shull, R.M., Munger, R.J., Spellacy, E., Hall, C.W., Constantopoulous, G. and Neufeld E. (1982) Canine α-L-iduronidase deficiency: A model of mucopolysaccharidosis type I. *Am J. Path.* **109:** 244-248.

Silveri, C.P., Kaplan, F.S., Fallon, M.D., Bayever, E. and August, C.S. (1991) Hurler syndrome with special reference to histologic abnormalities of the growth plate. *Clin. Orthop.* **269:** 305-311.

Sjogren, P., Pedersen, T. and Steinmetz, H. (1987) Mucopolysaccharidoses and anaesthetic risks. *Acta. Anaesthesiol. Scand.* **31:** 214-218.

Spellacy, E., Shull, R.M, Constantopoulos, G. and Neufeld, E.F. (1983) A canine model of human α-L-iduronidase deficiency. *Proc. Natl. Acad. Sci. USA* **80:** 6091-6095.

Staden, R. (1980) A new computer method for the storage and manipulation of DNA gel reading data. *Nucl. Acids. Res.* **8:** 3673-3694.

Staden, R. (1984) Computer methods to aid the determination and analysis of DNA sequences. *Biochem. Soc. Trans.* **12:** 1005-1008.

Stein, C., Gieselmann, V., Kreysing, J., Schmidt, B., Pohlmann, R., Waheed, A., Meyer, H.E., O'Brien, J.S. and von Figura, K. (1989) Cloning and expression of human arylsulfatase A. *J. Biol. Chem.* **264:** 1252-1259.

Steingrimsdottir, H., Rowley, G., Dorado, G., Cole, J. and Lehmann, A.R. (1992) Mutations which alter splicing in the human hypoxanthine-guanine phosphoribosyltransferase gene. *Nucl. Acids Res.* **20:** 1201-1208.

Stigson, M. and Kjellen, L. (1991) Large disulfide-stabilized proteoglycan complexes are synthesized by the epidermis of Axolotyl embryos. *Arch. Biochem. Biophys* **290:** 391-396.

Stoltzfus, L.J., Sosa-Pineda, B., Moscowitz, S.M., Menon, K.P., Dlott, B., Hooper, L., Teplow, D.B., Shull, R.M. and Neufeld, E.F. (1992) Cloning and characterization of cDNA encoding canine α-L-iduronidase: mRNA deficiency in mucopolysaccharidosis I dog. *J. Biol. Chem.* **267:** 6570-6575.

Stryer, L. (1981) Mechanisms of enzyme action: lysozyme and carboxypeptidase. In *Biochemistry*. W.H. Freeman and Co., San Francisco, pp 135-148.

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

Sutherland, G.R., Baker, E., Callen, D.F., Hyland, V.J., May, B.K., Bawden, M.J., Healy, H.M. and Borthwick, I.A. (1988) 5-Aminolevulinate synthase is at 3p21 and thus not the primary defect in X-linked sideroblastic anemia. *Am. J. Hum. Genet.* **43:** 331-335.

Tager, J.M., Jonsson, L.V.M., Aerts, J.M.F.G. Elferink, R.P.J.O., Schram, A.W., Erickson, A.H. and Barranger, J.A. (1984) Metabolic consequences of genetic defects in lysosomes. *Biochem. Soc. Trans.* **12:** 902-905.

Talerico, M. and Berget, S.M. (1990) Effect of 5' splice site mutations on splicing of the preceding exon. *Mol. Cell. Biol.* **10:** 6299-6305.

Taylor, J.A., Gibson, G.J., Brooks, D.A. and Hopwood, J.J. (1991) α-L-Iduronidase in normal and mucopolysaccharidosis type I human skin fibroblasts. *Biochem. J.* **274:** 263-268.

Taylor, R.M., Farrow, B.R.H. and Stewart, G.J. (1992) Amelioration of clinical disease following bone marrow transplantation in fucosidase-deficient dogs. *Am. J. Med. Genet.* **42:** 628-632.

Taylor, R.M., Stewart, G.J., Farrow, B.R.H., Byrne, J. and Healy, P.J. (1989) Histological improvement and enzyme replacement in the brains of fucosidosis dogs after bone marrow engraftment. *Transplant. Proc.* **21:** 3074-3075.

Thompson, L.M., Plummer, S., Schalling, M., Altherr, M.R., Gusella, J.F., Housman, D.E. and Wasmuth, J.J. (1991) A gene encoding a fibroblast growth factor receptor isolated from the Huntington disease gene region of human chromosome 4. *Genomics* **11:** 1133-1142.

Tomatsu, S., Fukuda, S., Masue, M., Sukegawa, K., Fukao, T., Yamagishi, A., Hori, T., Iwata, H., Ogawa, T., Nakashima, Y., Hanyu, Y., Hashimoto, T., Titani, K., Oyama, R., Suzuki, M., Yagi, K., Hayashi, Y. and Orii, T. (1991a) Morquio disease: isolation, characterization and expression of full-length cDNA for human N-acetylgalactosamine-6-sulfate sulfatase. *Biochem. Biophys. Res. Commun.* **181:** 677-683.

Tomatsu, S., Fukuda, S., Sukegawa, K., Ikedo, Y., Yamada, S., Yamada, Y., Sasaki, T., Okamoto, H., Kuwahara, T., Yamaguchi, S., Kiman, T., Shintaku, T., Isshiki, G and Orii, T. (1991b) Mucopolysaccharidosis type VII: characterization of mutations and molecular heterogeneity. *Am. J. Hum. Genet.* **48:** 89-96.

Tomatsu, S., Sukegawa, K., Ikedo, Y., Fukuda, S., Yamada, Y., Sasaki, T., Okamoto, H., Kuwabara, T. and Orii, T. (1990) Molecular basis of mucopolysaccharidosis type VII: replacement of Ala619 in β-glucuronidase with Val. *Gene* **89:** 283-287.

Tsuji, S., Choudary, P.V., Martin, B.M., Stubblefield, B.K., Mayor, J.A., Barranger, J.A. and Ginns, E.I. (1987) A mutation in the human glucocerebrosidase gene in neuronopathic Gaucher's disease. *N. Engl. J. Med.* **316:** 570-575.

Tsuji, S., Martin, B.M., Barranger, J.A., Stubblefield, B.K., LaMarca, M.E. and Ginns, E.I. (1988) Genetic heterogeneity in type 1 Gaucher disease: multiple genotypes in Ashkenazic and non-Ashkenazic individuals. *Proc. Natl. Acad. Sci. USA* **85:** 2349-2352. Published erratum appears in *Proc. Natl. Acad. Sci. USA* **85:** 5708.

Tybulewicz, V.L., Tremblay, M.L., LaMarca, M.E., Willemsen, R., Stubblefield, B.K., Winfield, S., Zablocka, B., Sidransky, E., Martin, B.M., Huang, S.P., Mintzer, K.A., Westphal, H., Mulligan, R.C. and Ginns, E.I. (1992) Animal model of Gaucher's disease from targeted disruption of the mouse glucocerebrosidase gene. *Nature* **357:** 407-410.

Ullrich, K., Grobe, H., von Figura, K., Hopwood, J.J. and Muller, V. (1981) Late onset form of mucopolysaccharidosis type I: clinical aspect and biochemical characterization of residual α-L-iduronidase activity. *J. Inher. Metab. Dis.* **4:** 171-172.

van Driel, I.R., Stearne, P.A., Grego, B., Simpson, R.J. and Goding, J.W. (1984) The receptor for transferrin on murine myeloma cells: one-step purification based on its physiology, and partial amino acid sequence. *J. Immunol.* **133:** 3220-3224.

Van Hoof, F. and Hers. H.G. (1964) The ultrastructure of hepatic cells from Hurler's disease (gargoylism). *C. R. Hedb. Sceances Acad. Sci. ser. D* **259:** 1281-1283.

von Figura, K. and Hasilik, A. (1986) Lysosomal enzymes and their receptors. *Annu. Rev. Biochem.* **55:** 167-193.

von Figura, K., Hasilik, A., Pohlmann, R., Braulke, T., Lemansky, P. and Stein, M. (1987) Mutations affecting transport and stability of lysosomal enzymes. *Enzyme* **38:** 144-153.

von Heijne, G. (1986) A new method for predicting signal sequence cleavage sites. *Nucleic Acids Res.* **14:** 4683-4690.

Waheed, A., Pohlmann, R., Hasilik, A., von Figura, K., van Elsen, A. and Leroy, J.G. (1982) Deficiency of UDP-*N*-acetylglucosamine: Lysosomal enzyme *N*-acetylglucosamine-1-phosphotransferase in organs of I-cell patients. *Biochem. Biophys. Chem. Commun.* **105:** 1052-1058.

Washington, J.A. (1937) Lipochondrodystrophy. In *Brennemanns's practice of paediatrics.* W.F. Prior Co., Inc., Hagerstown, Md., Chap 30.

Weber, B., Collins, C., Kowbel, D., Riess, O. and Hayden, M.R. (1991a) Identification of multiple CpG islands and associated conserved sequences in a candidate region for the Huntington disease gene. *Genomics* **11:** 1133-1142.

Weber, B., Hedrick, A., Andrew, S., Riess, O., Collins, C., Kowbel, D. and Hayden, M.R. (1992) Isolation and characterization of new highly polymorphic DNA markers from the Huntington disease region. *Am. J, Hum. Genet.* **50:** 382-393.

Weber, B., Riess, O., Hutchinson, G., Collins, C., Lin, B., Kowbel, D., Andrews, S., Shapppert, K. and Hayden, M.R. (1991b) Genomic organization and complete

sequence of the human gene encoding the β-subunit of the cGMP phosphodiesterase and its location to 4p16.3. *Nucl. Acids Res.* **19:** 6263-6268.

Wehnert, M., Hopwood, J.J., Schroder, W. and Herrmann, F.H. (1992) Structural gene aberrations in mucopolysaccharidosis II (Hunter). *Hum. Genet.* **89:** 430-422.

Weil, D., Bernard, M., Combates, N., Wirtz, M.K., Hollister, D.W., Steinmann, B. and Ramirez, F. (1988) Identification of a mutation that causes exon skipping during collagen pre-mRNA splicing in a Ehlers-Danlos syndrome variant. *J. Biol. Chem.* **263:** 8561-8564.

Weinberg, K.I. and Parkman, R. (1990) Bone marrow transplantation for genetic diseases. In Johnson, F.L. and Pochedly, C. (eds): *Bone Marrow Transplantation in Children*. Raven Press, New York. pp. 243-260.

Weinthal, J., Nolta, J.A., Yu, X.J., Lilley, J., Uribe, L. and Kohn, D.B. (1991) Expression of human glucocerebrosidase following retroviral vector-mediated transduction of murine hematopoietic stem cells. *Bone Marrow Transplant.* **8:** 403-412.

Whaley, W.L., Bates, G.P., Noveletto, A., Sedlacek, Z., Cheng, S., Romano, D., Ormondroyd, E., Allitto, B., Lin, C., Youngman, S., Baxendale, S., Bucan, M., Altherr, M., Wasmuth, J.J., Wexler, N.S., Frontali, M., Frischauf, A.-M., Lehrach, H., MacDonald, M.E. and Gusella, J.F. (1991) Mapping of cosmid clones in the Huntington's disease region of chromosome 4. *Somat. Cell Mol. Genet.* **17:** 83-91.

Whitley, C.B., Gorlin, R.J. and Krivit, W.A (1987) A nonpathologic allele (IW) for low α-L-iduronidase enzyme activity vis-a-vis prenatal diagnosis of Hurler syndrome. *Am. J. Med. Genet.* **28:** 233-243.

Whitley, C.B., Ramsay, N.K.C., Kersey,J.H. and Krivit, W. (1986) Bone marrow transplantation for Hurler syndrome: Assessment of metabolic correction. *Birth Defects* **22:** 7-24.

Wicker, G., Prill, V., Brooks, D., Gibson, G., Hopwood, J.J., von Figura, K. and Peters, C. (1991) Mucopolysaccharidosis VI (Maroteaux-Lamy syndrome). An intermediate

clinical phenotype caused by substitution of valine for glycine at position 137 of arylsulfatase B. *J. Biol. Chem.* **266:** 21386-21391.

Wiesmann, U. and Neufeld, E.F. (1970) Scheie and Hurler syndromes: apparent identity of the biochemical defect. *Science* **169:** 72-74.

Wilson, P.J., Morris, C.P., Anson, D.S., Occhiodoro, T., Bielicki, J., Clements, P.R. and Hopwood, J.J. (1990) Hunter Syndrome: Isolation of a cDNA clone encoding human iduronate-2-sulphatase and analysis of patient DNA. *Proc. Natl. Acad. Sci. USA* **87:** 8531-8535.

Wilson, P.J., Suthers, G.K., Callen, D.F., Baker, E., Nelson, P.V., Cooper, A., Wraith, J.E., Sutherland, G.R., Morris, C.P. and Hopwood, J.J. (1991) Frequent deletions at Xq28 indicate genetic heterogeneity in Hunter syndrome. *Hum. Genet.* **87:** 505-508.

Wolfe, J.H., Schuchman, E.H., Stramm, L.E., Concaugh, E.A., Haskins, M.E., Aguirre, G.D., Patterson, D.F., Desnick, R.J. and Giloboa, E. (1990) Restoration of normal lysosomal function in mucopolysaccharidosis type VII cells by retroviral mediated gene transfer. *Proc. Natl. Acad. Sci. USA* **87:** 2877-2881.

Yanisch-Perron, C., Vieira, J., and Messing, J. (1985) Improved M13 phage cloning vectors and host strains: nucleotide sequence of the M13mp18 and pUC19 vectors. *Gene* **33:** 103-119.

Young, R.A. and Davis, R.W. (1985) Efficient isolation of genes using antibody probes. *Proc Natl Acad Sci USA* **80:** 1194-1188.

PUBLICATIONS RESULTING FROM THIS THESIS

1. Scott, H.S., Ashton, L.J., Eyre, H.J., Baker, E., Brooks, D.A., Callen, D.F., Sutherland, G.R., Morris, C.P. and Hopwood, J.J. (1990) Chromosomal localization of the human α-L-iduronidase gene (*IDUA*) to 4p16.3. *Am. J. Hum. Genet.* **47:** 802-807.

2. MacDonald, M.E., Scott, H.S., Whaley, W.L., Pohl, T., Wasmuth, J.J., Lehrach, H., Morris, C.P., Frischauf, A.-M., Hopwood, J.J. and Gusella, J.F. (1991) Huntington disease-linked locus *D4S111* exposed as the α-L-iduronidase gene. *Somatic Cell Mol. Genet.* **17:** 421-425.

3. Scott, H.S., Anson, D.A., Orsborn, A.M., Nelson, P.V., Clements, P.R., Morris, C.P. and Hopwood, J.J (1991) Human α-L-iduronidase: cDNA isolation and expression. *Proc. Natl. Acad. Sci. USA* **88:** 9695-9699.

4. Scott, H.S., Nelson, P.V., Hopwood, J.J. and Morris, C.P. (1991) PCR of a Kpn I RFLP in the α-L-iduronidase gene. *Nucl. Acids Res.* **19:** 5796.

5. Scott, H.S., Nelson, P.V., Hopwood, J.J. and Morris, C.P. (1991) PCR of a VNTR linked to mucopolysaccharidoses type one and Huntington disease. *Nucl. Acids Res.* **19:** 6348.

6. Scott, H.S., Nelson, P.V., Cooper, A., Wraith E., Hopwood, J.J. and Morris, C.P. (1992) Mucopolysaccharidosis type I (Hurler syndrome) : linkage disequilibrium indicates the presence of a major allele. *Human Genet.* **88:** 701-702.

7. Scott, H.S., Guo, X-H., Hopwood, J.J. and Morris, C.P. (1992) Structure and sequence of the human α-L-iduronidase gene. *Genomics* **13:** 1311-1313.

8. Scott, H.S., Litjens, T., Hopwood, J.J. and Morris, C.P. (1992) PCR detection of two RFLP's in exon I of the α-L-iduronidase (IDUA) gene. *Human Genet.*, in press.

9. Scott, H.S., Nelson, P.V., MacDonald, M.E., Gusella, J.F., Hopwood, J.J. and Morris, C.P. (1992) An 86 bp VNTR within IDUA is the basis of the *D4S111* polymorphic locus. *Genomics*, in press.

10. Koizumi, T., MacDonald, M.E., Bucan, M., Hopwood, J.J., Morris, C.P., Scott, H.S., Gusella, J.F., and Nadeau, J.H. (1992) Linkage, but not gene order, of homologous loci, including α-L-iduronidase (*Idua*), is conserved in the Huntington disease region of the mouse and human genomes. *Mammalian Genome*. **3:** 23-27.

11. Scott, H.S., Litjens, T., Hopwood, J.J., and Morris, C.P. (1992) A common mutation for mucopolysaccharidosis type I associated with a severe Hurler syndrome phenotype. *Human Mutation*. **1:** 103-108.

12. Scott, H.S., Litjens, T., Nelson, P.V., Brooks, D.A., Hopwood, J.J., and Morris, C.P. (1992) α-L-Iduronidase mutations ($Q_{70}X$ and $P_{533}R$) associate with a severe Hurler syndrome phenotype. *Human Mutation*, in press.