UNIVERSITY OF ADELAIDE
17.11.94
LIBRARY

THE MOLECULAR GENETICS OF MUCOPOLYSACCHARIDOSIS TYPE VI

by

TOM LITJENS B.Sc.(Hons.)

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
Woman's and Children's Hospital
South Australia

and

Department of Pathology
Faculty of Medicine
University of Adelaide
South Australia

June 1994

Awarded 1995

Whoever loves instruction

loves knowledge

Proverbs 12:1

DEDICATION

To Natasha & Mum

Thank you for your support, patience

and love

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

CHAPTER 2

MATERIALS & METHODS

CHAPTER 3

ISOLATION OF THE 4S GENE

CHAPTER 4

IDENTIFICATION OF 4S MUTATIONS IN MPS-VI PATIENTS

CHAPTER 5

HIGH-LEVEL EXPRESSION AND PRELIMINARY ANALYSIS OF 4S MUTATIONS

CHAPTER 6

APPENDICES

BIBLIOGRAPHY

LIST OF ABBREVIATIONS

4S	*N*-acetylgalactosamine-4-sulphatase
ASO	allele specific oligonucleotide
BMT	bone marrow transplantation
bp	base pair
CHO cells	Chinese hamster ovary cells
cpm	counts per minute
CS	chondroitin sulphate
DMSO	dimethylsulphoxide
DNA	deoxyribonucleic acid
DS	dermatan sulphate
ER	endoplasmic reticulum
ERT	enzyme replacement therapy
GAG	glycosaminoglycans
H	hydroxylamine
hr	hour
HPLC	high performance liquid chromatography
kb	kilobase
kDa	kilo-Dalton
LSD	lysosomal storage disorder
M6P	mannose-6-phosphate
min	minute
MPS	mucopolysaccharidoses
MPS-VI	mucopolysaccharidosis type VI
mRNA	messenger RNA
N-terminal	amino-terminal
O	osmium tetroxide
PBS	phophate buffered saline
PCR	polymerase chain reaction
pfu	plaque forming units
RNA	ribonucleic acid
s	second
SDS-PAGE	sodium dodecyl sulphate - polyacrylamide gel electrophoresis
USP	universal sequencing primer

THESIS ABSTRACT

Maroteaux-Lamy syndrome or mucopolysaccharidosis type VI (MPS-VI) is a lysosomal storage disorder characterized by the defective degradation of dermatan sulphate and chondroitin-4-sulphate, due to the deficiency of *N*-acetylgalactosamine-4-sulphatase (4S). Clinical manifestations of MPS-VI affect a diverse array of somatic tissues; however the clinical severity varies in a continuum from mildly affected to severely affected patients. The primary goal of this work was to identify and analyse both mutant alleles in selected patients, in order to test the hypothesis that the mutant genotype could be correlated with the clinical and biochemical phenotype.

Amino acid sequence was obtained from human liver 4S and used to synthesize oligonucleotide probes for screening cDNA and genomic libraries. A 1.3-kb *Hind*III fragment containing the first coding exon of 4S was isolated from two genomic clones. The fragment was used to refine the chromosomal localization of the 4S gene (*ARSB* locus) to 5q13-5q14. Subsequently, a 4S cDNA clone was obtained from Christoph Peters and Kurt von Figura. Sequence analysis of the disulphide-linked 43-kDa and 8-kDa species of 4S showed that they were derived from the proteolytic processing of the 57-kDa species.

Southern analysis of 18 MPS-VI patients suggested that the majority of 4S mutations were either point mutations or small rearrangements. Nine patients were subjected to detailed molecular analysis using the techniques of cDNA synthesis, polymerase chain reaction (PCR), chemical cleavage, DNA sequencing and allele specific oligonucleotide hybridization. Eight mutations were identified including six single amino acid substitutions (T92M, R95Q, Y210C, H393P, P481L, and L498P), one frameshift mutation (ΔG_{238}), and a partially characterized deletion (Δexon IV). All of the mutations were rare or unique in the MPS-VI patient population screened, the majority of which were compound heterozygotes.

Five amino acid substitution mutants were expressed in CHO cells, the residual 4S activity and protein were determined, and the relative severity predicted. For each patient, the sum of the predicted relative severity of each of the two 4S mutant alleles was in reasonable agreement with the observed clinical phenotype. The demonstration of a genotype-phenotype correlation in MPS-VI will contribute to genetic counselling, clinical prognosis and assessment of therapy in patients.

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

Tom Litjens

Date: June 8, 1994

ACKNOWLEDGEMENTS

I would like to express my sincere appreciation to all those who have supported and patiently endured me during the rather protracted course of this study. I especially wish to thank the following:

My supervisors, Dr. C. Phillip Morris and Prof. John J. Hopwood. Their friendliness, dedication, enthusiasm, and insight have been of great assistance; it has been a privilege to work with them.

Prof. Barrie Vernon-Roberts of the Department of Pathology, University of Adelaide, who provided the opportunity to obtain this degree through his department. Prof. David Thomas, Dr. Evelyn F. Robertson and Prof. Anthony C. Pollard, the current and previous directors respectively of the Department of Chemical Pathology at the Adelaide Children's Hospital, for their interest and support.

I am indebted to Drs. Gary Gibson and Doug Brooks, and also Kerri Beckmann and Paul Nelson, who variously contributed to the purification of *N*-acetylgalactosamine-4-sulphatase (4S), the determination of amino acid sequence data, and the initial gene screening. Their work enabled the project to commence.

The generosity of the following colleagues is particularly appreciated; Dr. Doug Brooks, for guidance with the immunochemical assays described in Chapter 5, and for the provision of a wealth of unpublished data. Dr. Don Anson, for guidance with cell culture, and for the provision of the 4S expression constructs pRSVN.4S.08 and p4SFL. Dr. Gary Gibson, who made available unpublished data concerning the residual 4S activity in patient cultured fibroblasts. Our collaborators, Dr. Christoph Peters and Prof. Kurt von Figura, of the Georg-August-Universität Göttingen, who generously provided a 4S cDNA clone and the sequence prior to publication.

Dr. David Callen and Elizabeth Baker of the Department of Cytogenetics and Molecular Genetics, who determined the chromosomal localization of the 4S gene. 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.

I am grateful to the following doctors for their provision of clinical information on individual MPS-VI patients; Dr. John D. Arnold (Department of Chemical Pathology), Dr. Agnes Bankier (Royal Children's Hospital, Melbourne), Prof. J. Beveridge (Prince of Wales Children's Hospital, Sydney, Australia), Dr. Irène Maire (Centre d'Etude des Maladies Hereditaires Métabolique, Hôpital Debrousse, Lyon), Dr. Alan Parsons (Taranaki Base Hospital, New Plymouth, New Zealand), Dr. Evelyn F. Robertson (Department of Chemical Pathology), Prof. H.J. Weston (Wellington Hospital, Wellington, New Zealand), and Dr. J. Edward Wraith (Willink Biochemical Genetics Unit, Royal Manchester Children's Hospital).

Finally, a special thank you to both past and present colleagues in the Molecular Biology lab, for their camaraderie over the years; Kerri Beckmann, Lianne Blanche, Jill Durrant, Maria Fuller, Doula Gregoratos, Xiao-Hui Guo, John Harvey, Cathy Meaney, Diane Miller, Teresa Occhiodoro, Annette Orsborn, Dan Robertson, Dr. Hamish Scott, Dr. Phillip Thompson, and Dr. Peter Wilson. Drs. Hamish Scott and Peter Wilson also provided generous assistance with the preparation of this thesis.

This work was supported by funds provided from an NH&MRC programme grant to the Lysosomal Diseases Research Unit of the Department of Chemical Pathology.

CHAPTER 1

INTRODUCTION

UNIVERSITY OF ADELAIDE LIBRARY

1.1 GENERAL INTRODUCTION & PRELIMINARY COMMENTS

The heritable or genetic diseases have proven to be a rich source of `natural' experiments which have helped to elucidate complex biochemical processes. To date, hundreds of single gene defects have been described (e.g. Scriver *et al.*, 1989). Although individually rare, genetic diseases considered as a whole are responsible for a significant proportion of human morbidity. The lysosomal storage disorders are a group of more than 36 monogenic diseases in which enzyme deficiencies lead to the accumulation of substrates in the lysosome. Ten of the disorders, the mucopolysaccharidoses (MPS), result from defects in the catabolism of sulphated glycosaminoglycans (GAG). The MPS are differentiated on the basis of their clinical phenotype, the type of GAG stored, and the underlying enzyme deficiency.

The work presented in this thesis was performed with the initial aim of cloning the gene for the lysosomal enzyme *N*-acetylgalactosamine-4-sulphatase (EC 3.1.6.12, 4S), which is deficient in Mucopolysaccharidosis type VI (MPS-VI) or Maroteaux-Lamy syndrome. Subsequently, the mutations present in MPS-VI patients were to be identified and analysed. MPS-VI, in common with the other MPS disorders, exhibits wide clinical heterogeneity, which has hampered the accurate prediction of prognosis. Therefore, the primary goal was to assess the genetic contribution of various 4S mutations toward the clinical and biochemical variation observed in MPS-VI, and determine whether a consistent correlation existed between the genotype and the phenotype. A consistent correlation would contribute to diagnosis, increase the accuracy of predictions of clinical severity, and enable the assessment of therapeutic approaches for this debilitating disease.

This introduction seeks to comprehensively review the normal processes of GAG degradation and lysosomal enzyme trafficking; processes which are defective or perturbed in MPS-VI. In addition, relevant historical, clinical, biochemical, and

genetic aspects of MPS-VI will also be surveyed. Current and future approaches to the therapy of MPS-VI are discussed, and the introduction will conclude with the principles of genotype-phenotype correlation.

1.2 LYSOSOMES & LYSOSOMAL STORAGE DISORDERS

Broadly speaking, lysosomes are acidic intracellular organelles specialized for the terminal degradation of macromolecules. This definition of lysosomal function was originally proposed by de Duve and co-workers (reviewed in de Duve, 1963). Soon after, Hers (1965) introduced the concept of `inborn lysosomal diseases' or lysosomal storage disorders (LSD), based on the observation that cells from a patient with Pompe's disease were deficient in the lysosomal enzyme α-glucosidase, and also accumulated the undegraded glycogen substrate in the lysosome. In each LSD, Hers predicted that the deficiency of a lysosomal enzyme would lead to the accumulation of undigested substrate and subsequent lysosomal engorgement, which was considered to play a significant role in pathogenesis of the disease. Depending on the LSD, the stored substrate can include sphingolipids, glycoproteins, GAG, or other substrates (Neufeld, 1991). Lysosomal structure and biogenesis has been reviewed by Kornfeld & Mellman (1989).

Currently, at least three dozen LSD have been described (Neufeld, 1991), of which the majority are due to deficiencies of individual hydrolytic enzymes. The deficiency of a lysosomal hydrolase can be due to a primary defect in the enzyme itself, as in Pompe's disease which is caused by a deficiency of α-glucosidase (Hers *et al.*, 1989). In contrast, the primary defect may be in an accessory protein of a lysosomal hydrolase, as observed in a form of metachromatic leukodystrophy, where the apparent arylsulphatase A deficiency is due to a fundamental deficiency of the sulphatide activator, saposin (Kolodny, 1989; Neufeld, 1991). LSD and lysosomal enzyme deficiencies can also arise by other mechanisms, for example

mucolipidosis type II (I-cell disease), multiple sulphatase deficiency and galactosialidosis, each of which are deficient in multiple lysosomal enzymes. In particular, Mucolipidosis type II is due to the deficiency of a phosphotransferase which is responsible for the correct targeting of lysosomal enzymes to the lysosome (Nolan & Sly, 1989). Other LSD are due to the deficient transport of metabolites across the lysosomal membrane. In cystinosis, the di-amino acid cystine fails to traverse the lysosomal membrane and consequently accumulates within lysosomes (Gahl *et al.*, 1989). Further investigation of these disorders promises to yield rich insights into the molecular pathology of these diseases, and also lysosomal structure and function.

Ten of the LSD are deficient in enzymes specific for the degradation of sulphated glycosaminoglycans or mucopolysaccharides, and are thus termed mucopolysaccharidoses. The MPS disorders will be briefly reviewed below, followed by a detailed review of MPS-VI.

1.3 THE MUCOPOLYSACCHARIDOSES

The term "mucopolysaccharidosis" was originally devised by Brante (1952) to describe a group of rare genetic disorders of GAG metabolism. The glycosaminoglycans include heparan sulphate, dermatan sulphate, chondroitin-4-sulphate, chondroitin-6-sulphate, and keratan sulphate. Their role, structure, synthesis and degradation are discussed in Section 1.4. Each of the MPS disorders is due to a single enzyme deficiency in the GAG degradative pathways. The study of each disorder has improved the understanding of the disease state and has provided insights into the biochemistry of GAG degradation. An historical appreciation for the progress in MPS research can be obtained from the following reviews; McKusick *et al.* (1965), McKusick (1972), Cantz & Gehler (1976), Kresse *et al.* (1981), Muenzer (1986), Neufeld & Muenzer (1989), and Hopwood & Morris (1990).

The clinical profiles of each of the MPS disorders display shared features, and are characterized by pleiotrophic signs which can include abnormalities of the skeleton, soft tissues and the central nervous system. Generally, the clinical course is chronic and progressively degenerative. In severely affected patients, the prognosis is a reduced life expectancy of considerable morbidity. However, the clinical severity can vary in a wide continuum from mild to severe extremes within each MPS disorder. Consequently, the MPS disorders can be difficult to precisely diagnose on the basis of clinical features alone, due to the broad clinical spectrum and the overlap of clinical features. The pairwise combination of multiple distinct alleles at each enzyme locus has been proposed as an explanation for the clinical heterogeneity (Neufeld & Muenzer, 1989). An exception is the X-linked disorder MPS-II or Hunter syndrome, in which the clinical heterogeneity was proposed to be due to a single copy of distinct mutant alleles. In the majority of families in which more than one person is affected with an MPS, the clinical manifestation is similar. However, intrafamilial variability in clinical features has occasionally been observed in a number of MPS disorders (reviewed by Zlotogora, 1987; e.g. MPS-IH & Sanfilippo: McDowell *et al.*, 1993). As the siblings are assumed to have identical mutant alleles, the role of the `genetic background' and non-genetic factors in the modification of the clinical phenotype needs to be considered.

The MPS disorders are characterized biochemically by the intralysosomal storage of elevated levels of sulphated GAG within the tissues. The storage is progressive and the rate varies depending on the cell type. The lysosomal engorgement with stored GAG leads to organelle proliferation and disruption, which is thought to result in the distortion of cell architecture and function, and hence the development of clinical features. A multiplicity of distended lysosomes are evident in an electron micrograph of skin fibroblasts from a cat afflicted with MPS-VI (Fig. 1.1). In MPS, elevated levels of sulphated GAG are also excreted in the urine, which is exploited

in diagnosis as described in Section 1.8.1. Intralysosomal storage of GAG can also be induced experimentally by drugs (reviewed in Ruben *et al.*, 1993). For example, tilorone induces the storage of dermatan sulphate in the liver, spleen, and cornea of rats (Lüllmann-Rauch, 1989).

The first unambiguous clinical descriptions of patients with MPS were published by Hunter and by Hurler (around the time of World War I), and consequently the disorders were referred to collectively in the literature as the Hurler-Hunter syndrome. The disorders were also referred to as dysostosis multiplex or gargoylism, based on particular clinical features. In the 1950's the biochemical abnormality in Hurler syndrome was first identified as the storage of sulphated GAG in patient tissues (Brante, 1952). The increased excretion of GAG in the urine of MPS patients (mucopolysaccharidurıa) was first described by Dorfman & Lorincz (1957). Van Hoof & Hers (1964) concluded that the lysosome was the intracellular storage site of the sulphated GAG, based on the electron microscopic analysis of liver biopsies from patients with Hurler syndrome. Although at this stage the fundamental metabolic defects underlying the MPS disorders were not known, it was proposed that the storage of GAG within the lysosome was due to the deficiency of lysosomal hydrolases (Hers, 1965). McKusick and co-workers numerically classified the MPS into the six distinct disorders MPS-I to MPS-VI, based on clinical features, the type of mucopolysaccharide excreted in the urine, and the mode of inheritance (McKusick *et al.*, 1965).

The fundamental defect in MPS disorders responsible for the intralysosomal GAG accumulation was first demonstrated in MPS-I and MPS-II, as a reduced rate of GAG degradation (Fratantoni *et al.*, 1968). Relative to normal controls, the patient cultured fibroblasts accumulated excessive amounts of labelled sulphated mucopolysaccharides synthesized from radiolabelled sulphate added to the culture medium. The accumulation was due to a reduction in the rate of turnover of the

FIGURE 1.1 LYSOSOMAL INCLUSION BODIES

Electron micrograph of skin fibroblasts from a cat affected with MPS-VI or Maroteaux-Lamy syndrome show lysosomal inclusion bodies which appear as vacuoles (v). These vacuoles are lysosomes engorged with partially catabolized glycosaminoglycans. (Photo courtesy of John Hopwood.)


V
V

labelled mucopolysaccharides. Subsequently, Fratantoni *et al.* (1969) demonstrated that the aberrant lysosomal storage observed in patient cultured fibroblasts could be corrected by the addition of culture media either from normal cultured fibroblasts or from another MPS type. The purification and characterization of the "corrective factors" present in the culture media enabled the identification of the specific enzyme defects in a number of the MPS. Elucidation of the enzyme defects has enlarged and refined the MPS classification from the original 6 to 10 biochemically distinct disorders. Comprehensive reviews of the clinical and biochemical features of each of the MPS disorders include McKusick *et al.* (1965), McKusick (1972), Cantz & Gehler (1976), Kresse *et al.* (1981), Muenzer (1986), Neufeld & Muenzer (1989), and Hopwood & Morris (1990).

MPS-I is due to the deficiency of α-L-iduronidase. Two forms are distinguished clinically, but are in fact due to the same enzyme deficiency. The clinically severe form is referred to as Hurler syndrome (MPS-IH) and is considered the archetype of the MPS disorders. The clinically milder form, Scheie syndrome, was formerly considered a separate disorder, MPS-V. However, the subsequent demonstration that MPS-IH cells were not corrected by co-cultivation with culture media from MPS-V cells, and ultimately the identification of the enzyme defect in MPS-V, led to the reclassification of MPS-V as MPS-IS. The marked variation in clinical severity observed between Hurler and Scheie syndromes, which are deficient in the same enzyme, is typical of the variation observed within each MPS. MPS-II, also known as Hunter syndrome, is due to the deficiency of iduronate-2-sulphatase. MPS-II is the only X-linked recessively inherited MPS disorder, while the others are inherited as autosomal recessive diseases. MPS-III or Sanfilippo syndrome was originally classified as a single disorder based on clinical and urinary analysis. However, enzyme analysis identified four distinct subtypes A to D based on different enzyme deficiencies. Similarly, MPS-IV or Morquio syndrome is now recognized as two distinct enzyme deficiencies, A and B, with similar clinical

presentation. The reclassification of Scheie disease as MPS-IS has rendered the MPS-V designation vacant. MPS-VI or Maroteaux-Lamy syndrome, was originally considered a Hurler variant based on clinical features. However, MPS-VI is due to the deficiency of *N*-acetylgalactosamine-4-sulphatase. Since the original classification, MPS-VII or Sly syndrome has been added to the list of MPS, and is due to the deficiency of β-D-glucuronidase.

In most of the MPS, the identification of the enzyme defects enabled the purification of the normal enzyme. Subsequently, progress was made in the diagnosis and biochemical analysis of each disorder. In order to obtain insight into the molecular defects responsible for the pathology, attempts were made to clone the genes and identify the patient mutations. The aim of the work presented in this thesis was to ultimately test the hypothesis that a correlation existed in MPS-VI patients between the clinical phenotype, the biochemical phenotype, and the mutant genotype. The rapid progress of research in the MPS over the last few years is exemplified by the fact that at the commencement of this Ph.D. project, the gene for only one of the ten enzymes deficient in the MPS had been cloned (β-D-glucuronidase, Oshima *et al.*, 1987). Currently, only the three genes affected in MPS-IIIA, IIIB, and IIIC remain to be cloned. Prior to a detailed review of MPS-VI, the normal processes of GAG synthesis and degradation, and lysosomal enzyme synthesis and maturation, will be reviewed.

1.4 GLYCOSAMINOGLYCANS & PROTEOGLYCANS

1.4.1 STRUCTURE & FUNCTION

The sulphated GAG are complex and chemically diverse anionic polysaccharides with a wide distribution in animal tissues. The fundamental GAG structure is an alternating copolymer of a hexuronic acid and a hexosamine (except for keratan sulphate, where the hexuronic acid is substituted by galactose). The individual glycosaminoglycans are distinguished based on the hexuronic acid and hexosamine composition, and the configuration of the glycosidic linkages. However, variation in each GAG type is observed, in terms of the constituent saccharides of the core polysaccharide chain, and in the degree and position of sulphation (Cantz & Gehler, 1976). The heterogeneity displayed both between and within GAG chain types has hampered the elucidation of their structure and function.

The glycosaminoglycan chains are usually found as a component of proteoglycans where they are covalently linked to a protein core, however free chains have also been isolated (Hovingh *et al.*, 1993). The core proteins also exhibit diversity. The structure and function of proteoglycans have been reviewed by Rodén (1980), Fransson (1987), Kjéllen & Lindahl (1991), and Hardingham & Fosang (1992). Proteoglycans are essentially ubiquitous in all mammalian tissues, and at the cellular level, have been found in the nucleus, at the cell membrane, and in the extracellular matrix (reviewed in Fransson, 1987; Yanagishita, 1993). The proteoglycans are thought to play a role in such diverse processes as cell-cell recognition and adhesion, control of cell proliferation, matrix assembly and cell-matrix interactions. They are a major component of the ground substance of connective tissue, and are thought to be a determinant of the tissue's viscoelastic and mechanical properties, for example the shock absorbency of cartilage. The vast

chemical and structural diversity of the proteoglycans are thought to reflect their wide range of roles.

The dermatan sulphate proteoglycans are mainly found in connective or supporting tissues, such as cartilage, sclera, skin, tendon and bone. The integral role of dermatan sulphate in these tissues, and the necessity of normal dermatan sulphate catabolism, is reflected by the skeletal and soft tissue abnormalities observed in MPS-VI patients. A number of dermatan sulphate (and the related chondroitin sulphate) proteoglycans have been identified and include the large proteoglycan versican and the relatively smaller proteoglycans decorin, biglycan and proteoglycan-Lb, all of which are located in the extracellular matrix (Yanagishita, 1993). The apparent absence of neurological deterioration in the brain of MPS-VI patients would suggest that either DS is not present or is at a low concentration; the quantity and distribution of dermatan sulphate proteoglycans in the brain is not known.

1.4.2 GLYCOSAMINOGLYCAN SYNTHESIS

The structure and synthesis of proteoglycans has been reviewed by Rodén (1980) and Fransson (1987). The synthesis of the protein core of a proteoglycan precedes that of the GAG chains. The core protein is synthesized on membrane-bound ribosomes, and the nascent polypeptide chain is translocated into the endoplasmic reticulum (ER). Synthesis of the GAG chain occurs during transit of the core protein through the ER and Golgi membranes, and is initiated by the addition of a D-xylosyl moiety to specific serine residues in the core protein. Subsequently, for most GAG types a trisaccharide acceptor sequence composed of D-glucuronyl-D-galactosyl-D-galactose is linked to the xylose. The main GAG chain is then assembled onto the acceptor sequence by the repeated and alternating addition of a hexosamine and a hexuronic acid to the non-reducing end. In the GAG, dermatan sulphate and chondroitin sulphate, the hexosamine is *N*-acetylgalactosamine, while

the hexuronic acid is iduronic and glucuronic acid, or solely glucuronic acid, respectively. A typical dermatan sulphate or chondroitin sulphate chain is comprized of 40-100 disaccharide residues (Yanagishita, 1993).

Dermatan sulphate and chondroitin sulphate are isomeric, as dermatan sulphate is derived from the latter by the epimerization of glucuronic acid residues to iduronic acid, catalysed by uronosyl epimerase acting at the carboxylic acid at the C-5 position. The extent of epimerization can be variable, and contributes to the considerable structural heterogeneity observed in dermatan sulphate chains. For example, dermatan sulphate chains isolated from the adult human meniscus, a fibrocartilagenous disc present in the knee joint, contain about 70% of their uronic acid as iduronate, whereas in human articular cartilage the proportion of iduronate would appear to increase with age from 20 to 50% (Roughley & White, 1992).

Sulphation of the GAG chains can occur at selected *N*- and *O*-positions, and is catalysed by sulphotransferases in the Golgi complex (Rodén, 1980; Graham & Winterbourne, 1988). Usually the C-4 or C-6 position of *N*-acetylgalactosamine is sulphated, and in dermatan sulphate the C-2 position of iduronic residues is also frequently sulphated. The extent and position of sulphation can be variable, and thus contributes to the considerable structural heterogeneity observed within each GAG chain type. Once the elongation and modification of the carbohydrate chains has been completed, the proteoglycans are exported to the cell surface, presumably in vesicle shuttles.

1.4.3 GLYCOSAMINOGLYCAN DEGRADATION

Proteoglycan degradation commences at the cell membrane, during the process of internalization into endosomes (reviewed in Fransson, 1987; Neufeld & Muenzer, 1989; Hopwood & Morris, 1990). In brief, the protein core is degraded by proteases, and *endo*glycosidases present in the endosomes or at the cell surface partially

degrade the GAG chains into shorter oligosaccharide fragments. The GAG fragments are then transported to the lysosome, where degradation is completed by the sequential action of a series of *exo*hydrolases (glycosidases and sulphatases), which degrade the polymer to free monosaccharides and sulphate. The process of GAG degradation within the lysosome, when defective, leads to an MPS disorder. In the MPS, deficiencies in the activity of an individual glycosidase or sulphatase interrupt the sequential degradative process and lead to the intralysosomal storage and urinary excretion of partially degraded GAG. As the rate of dermatan sulphate degradation is reduced in MPS-VI, the normal process of dermatan sulphate degradation within the lysosome will be reviewed below.

The lysosomal degradation of dermatan sulphate oligosaccharides is achieved by the repeated and sequential action of the lysosomal enzymes at the non-reducing end of dermatan sulphate, as shown in Figure 1.2. The first enzyme in the pathway shown is iduronate-2-sulphatase, which desulphates the terminal C-2-sulphated iduronic acid residue; the deficiency of this enzyme results in MPS-II (Hunter syndrome). The second enzyme in the pathway, α-L-iduronidase, removes the iduronate residue, and when deficient leads to MPS-I (Hurler and Scheie syndromes). The desulphation of the resultant terminal *N*-acetylgalactosamine residue at the C-4 position is catalysed by *N*-acetylgalactosamine-4-sulphatase; the deficiency of this enzyme results in MPS-VI (Maroteaux-Lamy syndrome). The subsequent removal of the terminal *N*-acetylgalactosaminide residue is thought to be performed by one or more of the 3 β-hexosaminidase isoenzymes A, B, and S (Hopwood & Elliott, 1985; Neufeld & Muenzer, 1989). Surprisingly, a deficiency of β-hexosaminidase A and B does not cause an MPS disorder but rather the G_{M2}-gangliosidosis, Sandhoff disease. Presumably in these patients, the S isoenzyme is responsible for the removal of the *N*-acetylgalactosaminide residue. Finally, β-D-glucuronidase removes the terminal glucuronate residues, and when deficient leads to MPS-VII (Sly syndrome).

Patients deficient in 4S also accumulate the monosaccharides *N*-acetylgalactosamine-4-sulphate and *N*-acetylgalactosamine-4,6-disulphate, in addition to dermatan sulphate. Therefore, an alternate pathway for dermatan sulphate degradation has been proposed (Hopwood & Elliott, 1985). In this pathway, the β-hexosaminidase isoenzymes act after the α-L-iduronidase and catalyse the hydrolysis of the terminal *N*-acetylgalactosamine-4-sulphate residue of dermatan sulphate to liberate *N*-acetylgalactosamine-4-sulphate (Fig. 1.2). In normal individuals, 4S could then conceivably hydrolyse the sulphated monosaccharide to inorganic sulphate and *N*-acetylgalactosamine, while in MPS-VI patients the deficiency of 4S would lead to accumulation of the sulphated monosaccharide. The degree to which this alternate pathway is utilized in the normal degradation of dermatan sulphate is unknown.

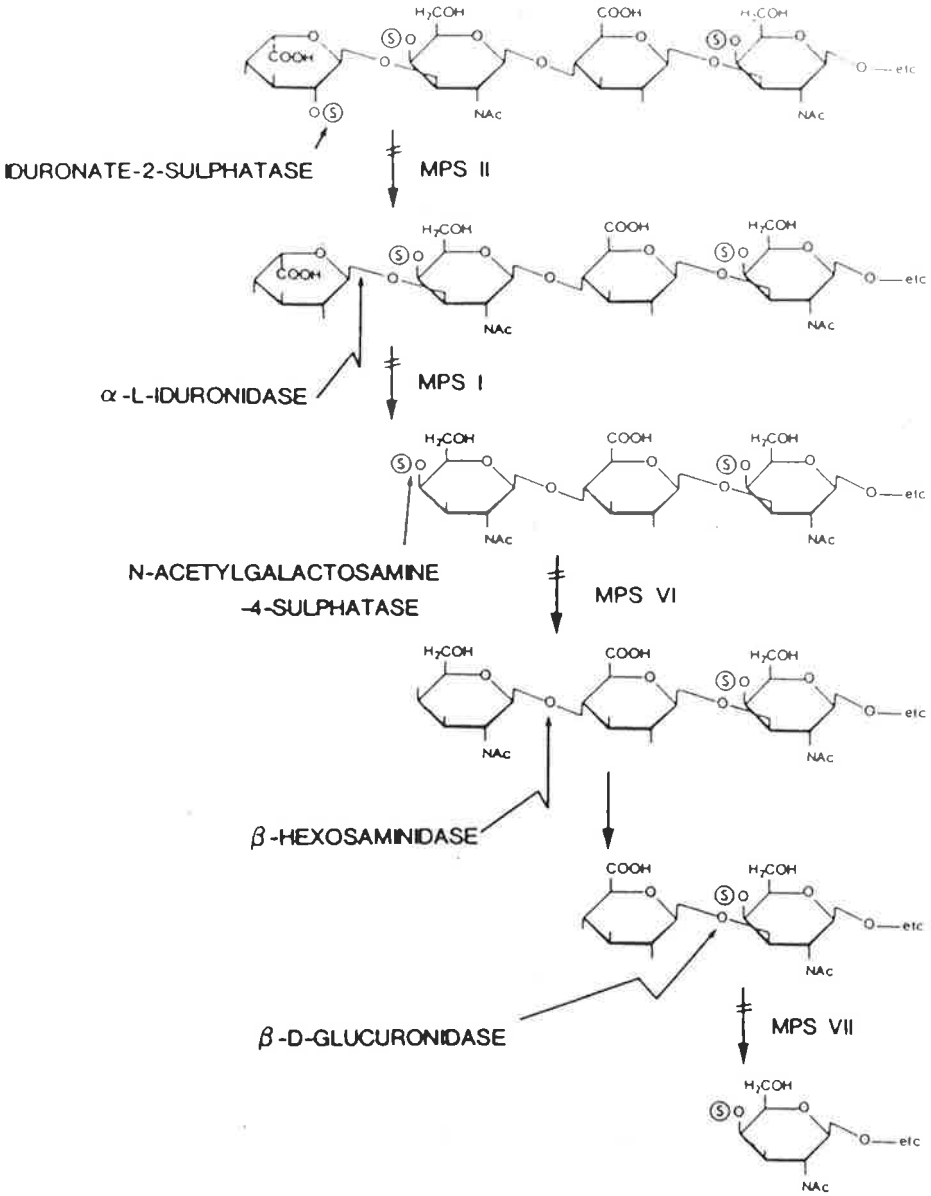
4S is also implicated in the normal catabolic pathway for chondroitin sulphate, as *N*-acetylgalactosamine-4-sulphate residues are present in both dermatan sulphate and chondroitin sulphate. However storage and excretion of chondroitin sulphate is not observed in MPS-VI. The lack of storage has been attributed to the action of lysosomal hyaluronidase, an *endo*-hexosaminidase, which can hydrolyse the terminal *N*-acetylgalactosamine-4-sulphate residue of chondroitin-4-sulphate, and so by-pass the arrest in the usual catabolic pathway due to the 4S deficiency (Neufeld & Muenzer, 1989). The sulphated monosaccharides produced by the action of hyaluronidase could conceivably contribute to the pool of sulphated monosaccharides observed in MPS-VI patients.

In summary, the degradation of the GAG oligosaccharide fragments is accomplished by the sequential action of hydrolases resident in the lysosome. The synthesis, processing, and intracellular trafficking of lysosomal enzymes will now be discussed.

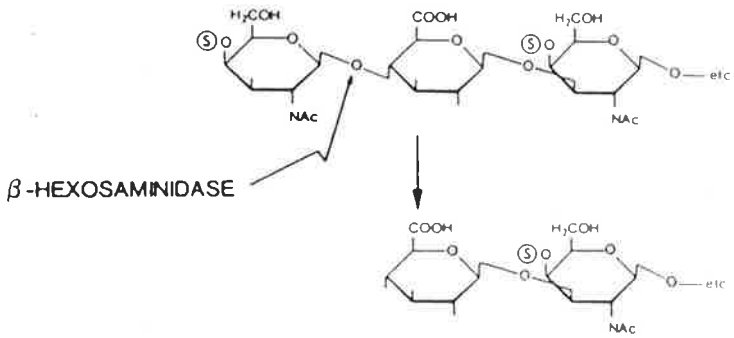
FIGURE 1.2 DEGRADATIVE PATHWAY OF DERMATAN SULPHATE

(from Hopwood & Morris, 1990)

The hypothetical sequential degradation pathway of dermatan sulphate within the lysosome. The participating hydrolases are shown, together with the mucopolysaccharidoses associated with a deficiency of each enzyme in the pathway.



ALTERNATE PATHWAY



1.5 BIOSYNTHESIS & PROCESSING OF LYSOSOMAL ENZYMES

The biosynthesis, post-translational modification, maturation and targeting of proteins in general and lysosomal enzymes in particular is elaborate and of interest. A review of this process is relevant to the MPS disorders and other LSD, as perturbations in the normal pathway are associated with enzyme deficiencies and hence disease. The synthesis and processing of lysosomal enzymes has been reviewed by Hasilik & von Figura (1984), Kornfeld & Mellman (1989), and Hasilik (1992). The trafficking of proteins to lysosomes has been discussed by von Figura & Hasilik (1986), Kornfeld (1986), Pfeffer (1988), and Kornfeld & Mellman (1989). The general process of protein trafficking to organelles has been reviewed by Hong & Tang (1993) and Luzio & Banting (1993).

1.5.1 BIOSYNTHESIS

Proteins destined for residence in the lysosome are synthesized in the rough endoplasmic reticulum (ER) on membrane-bound ribosomes (Fig. 1.3). Subsequently, the nascent polypeptide chain is cotranslationally translocated into the lumen of the ER, and the signal peptide sequence at the amino-terminal end is removed. The process of translocation, the correct folding of the polypeptide chain, and the formation of the appropriate disulphide bonds, are facilitated by a complex of proteins resident in the lumen of the ER, which are collectively known as molecular chaperones. Molecular chaperones have been recently reviewed by Craig *et al.* (1993) and Hendrick & Hartl (1993). The molecular chaperones, which include BiP, peptidyl-prolyl *cis-trans* isomerases and protein disulphide isomerases, transiently associate with the translocated protein and stabilize intermediate folded states and prevent inappropriate interactions. Proteins translocated into the ER are subjected to stringent scrutiny, as mutant or aberrant proteins are rapidly degraded at a site in or near the ER. The proteolytic

degradation of nascent aberrant polypeptides is one of a number of mechanisms which contribute to the enzyme deficient state in MPS disorders, as discussed in Chapter 5, Section 5.8.1.

1.5.2 ADDITION OF OLIGOSACCHARIDES

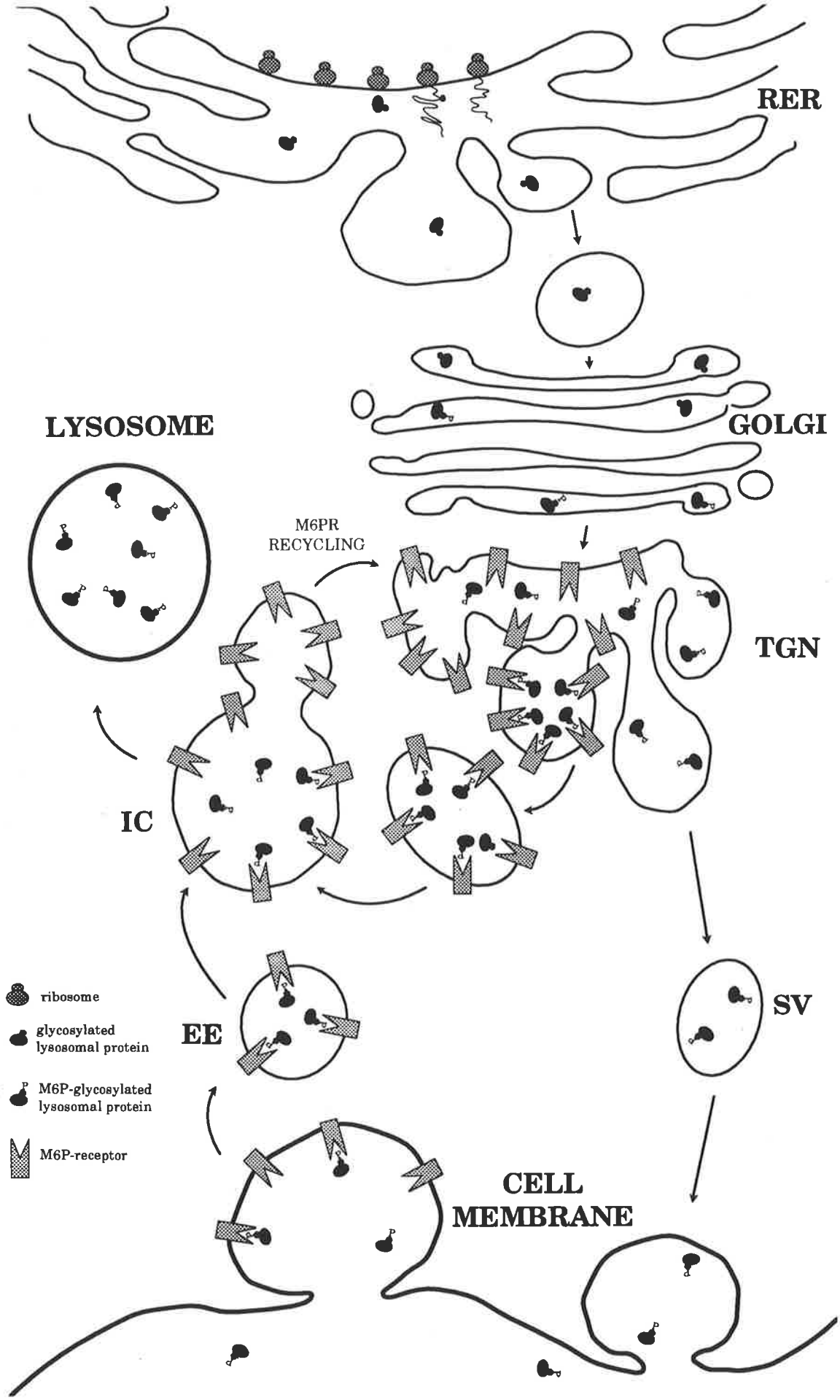
During cotranslation into the ER, the nascent polypeptide chains are modified by the addition of high-mannose oligosaccharide structures (Fig. 1.3). The glycosylation step involves the transfer of a large preformed core oligosaccharide consisting of 2 *N*-acetylglucosamine, 9 mannose and 3 glucose residues, from the activated lipid precursor dolichol pyrophosphate to specific asparagine residues in the protein. The core oligosaccharide is *N*-linked to asparagine residues which conform to the sequence consensus asparagine-x-serine/threonine, but not all potential sites are utilized. The glycosylated proteins are transported through the ER to the Golgi complex by membrane-bound vesicles. Complex processing of the carbohydrate chains occurs during transport, which includes the trimming of a number of saccharides. Further trimming and remodelling of the oligosaccharide side-chains of both lysosomal and non-lysosomal proteins occurs during transit through the compartments of the Golgi complex. The processing is extensive and a detailed description is beyond the scope of this review.

A crucial process for the majority of lysosomal enzymes is the acquisition on the oligosaccharide side-chains of the lysosomal targeting signal, mannose-6-phosphate. The mannose-6-phosphate marker enables lysosomally destined proteins to be selectively removed from the default secretion pathway, in which proteins are directed to the cell membrane, and be targeted to the lysosome. Most but not all lysosomal enzymes require this modification in order to be targeted to the lysosome; targeting independent of the mannose-6-phosphate will be discussed later in Section 1.5.6.

FIGURE 1.3 BIOSYNTHESIS & TRANSPORT OF LYSOSOMAL ENZYMES

A schematic representation of the process of biosynthesis and mannose-6-phosphate-*dependent* trafficking of soluble lysosomal enzymes to the lysosome in a typical cell. Lysosomal enzymes are synthesized and translocated into the rough endoplasmic reticulum (RER), where *N*-linked oligosaccharides are added and subsequently processed. The glycosylated enzymes are then transferred to the Golgi complex via membrane-bound vesicles. Within the Golgi, further processing of the oligosaccharide side chains occurs, including the synthesis of the mannose-6-phosphate marker via phosphorylation of specific mannose residues. Mannose-6-phosphate receptors (M6PR) in the trans-Golgi network (TGN) bind mannose-6-phosphorylated proteins, the complex is then transported to the intermediate compartment (IC), also known as a late endosome, pre lysosome or early lysosome. Acidification of the organelle causes the M6PR to dissociate from the lysosomal protein and recycle back to the Golgi/TGN. The lysosomal protein is then transported from the IC to the lysosome. For simplicity, the M6PR are represented by one symbol, however there are two different M6PR in the cell (see Section 1.5.4).

Lysosomal proteins that are not retrieved by M6PR in the TGN enter the secretory pathway via secretory vesicles (SV), and are released at the cell surface. M6PR located on the cell membrane bind extracellular lysosomal proteins, which are then endocytosed into early endosomes (EE), transported to the IC, and delivered to the lysosome.


RER
GOLGI
LYSOSOME
M6PR
RECYCLING
TGN
IC
EE
SV
CELL
MEMBRANE
ribosome
glycosylated
lysosomal protein
M6P-glycosylated
lysosomal protein
M6P-receptor

It is interesting that both the lysosomal hydrolases involved in GAG degradation and the GAG substrates under synthesis are trafficked through the Golgi, yet the GAG are presumably not degraded. This would suggest that the lysosome is unique in it's provision of a suitable environment for GAG degradation, or that the GAG are kept separate from the lysosomal hydrolases during transit through the Golgi.

The roles of the oligosaccharide side-chains in addition to lysosomal targeting (when modified with mannose-6-phosphate) are not known. However, site-directed mutagenesis studies of the glycosylation sites in a number of enzymes would suggest that the *N*-linked oligosaccharides play a role in correct folding of the polypeptide chain and are necessary for full activity. In β-glucosidase, 1 of the 4 glycosylation sites was found to be essential for the synthesis of a catalytically active enzyme (Berg-Fussman *et al.*, 1993). In β-glucuronidase, *N*-linked glycosylation was also shown to be necessary for expression of activity (Shipley *et al.*, 1993). However, enzymatic deglycosylation of the mature enzyme showed a minimal reduction in activity, which suggested that glycosylation was necessary for the establishment but not the maintenance of enzymatic activity. Studies on the role of glycosylation are hampered by the difficulty in the experimental separation of the effect of the oligosaccharide, and the effect of amino acid substitutions used to abolish the glycosylation site, on enzyme structure and activity.

1.5.3 ADDITION OF MANNOSE-6-PHOSPHATE

The modification of lysosomally-destined enzymes with mannose-6-phosphate is achieved in two steps (reviewed by Kornfeld, 1986; von Figura & Hasilik, 1986; Kornfeld & Mellman, 1989). The first step occurs in the early (*cis*) compartments of the Golgi, where the lysosomal enzyme is selectively recognized by UDP-*N*-acetylglucosamine phosphotransferase. The transferase adds *N*-acetylglucosamine-1-phosphate to the C-6 hydroxyl group of selected mannose residues of the

oligosaccharide side-chain to give the phosphodiester, *N*-acetylglucosamine-1-phospho-6-mannose. In the second step, the *N*-acetylglucosamine is removed by an *N*-acetylglucosaminyl-phosphodiesterase, leaving the mannose-6-phosphate moiety (Fig. 1.3).

The molecular motif/s common to lysosomal enzymes that enable the acquisition of the 6-phosphomannosyl marker have not been determined precisely (Baranski *et al.*, 1992; Cantor *et al.*, 1992). However, the structural determinant on the protein is most likely to display conservation in tertiary structure rather than at the amino acid sequence level, as the deduced amino acid sequences of lysosomal enzymes do not display obvious conserved sequences that are candidates for this motif. Attempts to identify the mannose-6-phosphate recognition determinant are hampered at present, as the tertiary structure for only a few lysosomal enzymes are known, for example cathepsin D (Metcalf & Fusek, 1993), and the phosphotransferase has not been purified to homogeneity. Deficiency of the UDP-*N*-acetylglucosamine phosphotransferase results in the failure of the lysosomal hydrolases to acquire the 6-phosphomannosyl marker. The lysosomal storage disorders mucolipidosis II (I-cell disease) and III (pseudo-Hurler polydystrophy) are deficient in this enzyme activity, and are to various extents consequently deficient in many of the lysosomal enzyme activities. In these disorders, the lysosomal hydrolases are preferentially secreted as precursor species into the extracellular milieu, rather than correctly targeted to the lysosome.

1.5.4 MANNOSE-6-PHOSPHATE RECEPTORS

Mannose-6-phosphate receptors located in the *trans* Golgi network mediate the segregation of lysosomal enzymes bearing the mannose-6-phosphate marker from the default transport pathway destined for the cell surface, and target them towards the lysosome (Fig. 1.3). Two distinct mannose-6-phosphate receptors are currently known; the so-called cation-dependent and the cation-independent

receptor. Both receptors are involved in the recognition and lysosomal targeting of proteins bearing the 6-phosphomannosyl marker. The structure and function of each of the mannose-6-phosphate receptors have been reviewed (Pfeffer, 1988; Kornfeld, 1992).

The cation-independent mannose-6-phosphate receptor (CI-M6PR) is a transmembrane glycoprotein of about 270-300-kDa. The extracytoplasmic portion consists of 15 imperfect repeats of a 147 amino acid unit sequence, which contains a highly conserved region of 13 amino acid residues. The CI-M6PR has been shown to be structurally and functionally identical to the insulin-like growth factor receptor. The CI-M6PR is likely to function as a monomer. The cation-dependent mannose-6-phosphate receptor (CD-M6PR) is also a transmembrane glycoprotein, of about 46-kDa. The extracytoplasmic portion of the CD-M6PR contains a single copy of the 147 amino acid unit sequence that is present in the CI-M6PR. The functional form of the CD-M6PR is likely to be a dimer. Both receptors are primarily located intracellularly, within endosomes, prelysosomes and the Golgi, however a significant fraction (5-10%) are observed at the cell surface, where they appear briefly and are then internalized (Pfeffer, 1988). While both receptors participate in sorting newly synthesized lysosomal enzymes in the Golgi, only the CI-M6PR is efficient at receptor-mediated endocytosis of extracellular enzyme at the cell membrane. The inability of the CD-M6PR to function efficiently in endocytosis reflects its poor binding of ligand at the cell surface rather than it's failure to recycle to the plasma membrane (Stein *et al.*, 1987).

The necessity for two different receptors has not been determined, although the two receptors exhibit differences in such characteristics as subcellular distribution and binding specificities. Interestingly, no patients with a defect in either of the two receptors have yet been identified. The retrieval of lysosomally destined proteins from the secretory pathway is not complete and in all cells examined a

portion (5-20%) of the newly-synthesized enzymes are released extracellularly in precursor form (Kornfeld, 1986). However, the secreted enzyme can still be targeted to the lysosome, via receptor-mediated endocytosis, as described below in Section 1.5.5 (Fig. 1.3).

1.5.5 MANNOSE-6-PHOSPHATE-*DEPENDENT* TRANSPORT

After the mannose-6-phosphate receptors have bound the lysosomal hydrolases in the *trans* Golgi network, the receptor-enzyme complexes are collected into clathrin-coated vesicles, and transported to a destination that is variously termed an intermediate compartment, a late endosome, an acidic endosomal compartment, or a prelysosome (reviewed in Kornfeld, 1986; Pfeffer, 1988; Kornfeld & Mellman, 1989) (Fig. 1.3). Acidification of the intermediate compartment causes the mannose-6-phosphate receptors to release the lysosomal hydrolases, and the receptors are recycled back to the Golgi network for another round of transport (Goda & Pfeffer, 1988). The protein components involved in recycling of the receptors are beginning to be identified (e.g. Rab9, Lombardi *et al.*, 1993). The intermediate compartment has been proposed as the nexus of two lysosomally-destined pathways; the intracellular pathway of *de novo* synthesized lysosomal enzymes (Griffiths *et al.*, 1988), and the endocytotic pathway described below, which imports extracellular material. Lysosomal enzymes delivered to the intermediate compartment by either pathway later appear in mature lysosomes. Whether this delivery involves maturation of the intermediate compartment into a lysosome, or whether there is a transport process from the intermediate compartment to the lysosome, has not been resolved.

As mentioned previously, lysosomally-destined enzymes present in the culture medium or on the plasma membrane can be targeted to the lysosome by receptor-mediated endocytosis. In fibroblasts, about 5-10% of lysosomal enzymes are

lysosomally targeted by this secretion-recapture pathway (cited in Kornfeld, 1986). Mannose-6-phosphate receptors present at the cell surface bind the extracellular lysosomal enzymes and the receptor-enzyme complexes are internalized and transported to endosomes in clathrin-coated vesicles. The receptor-enzyme complexes are transported from the endosomes to the intermediate compartment, where the complexes dissociate. As for the intracellular pathway, the receptor is recycled back to the *trans* Golgi network, and the enzyme is transported to the lysosome. Receptor mediated uptake of lysosomal enzymes at the plasma membrane is not solely due to mannose-6-phosphate receptors, but can also be mediated by receptors specific for galactose, mannose, and *N*-acetylglucosamine residues on the enzymes (Hasilik, 1992).

The extracellular secretion of lysosomal enzyme precursors can be manipulated and enhanced experimentally. Key processes in the delivery of enzymes to the lysosome are the dissociation of the mannose-6-phosphate receptor-enzyme complex in the intermediate compartment, and the recycling of the receptor to the *trans*-Golgi network. The addition of weak bases such as ammonium chloride to the culture medium raises the pH of the acidic organelles and consequently prevents dissociation of the complex. The mannose-6-phosphate receptors are unable to recycle to the *trans*-Golgi network, which leads to the saturation of the available receptor pool. Consequently, newly synthesized mannose-6-phosphorylated enzyme is not retrieved in the *trans*-Golgi network, and follows the default secretion pathway to the cell surface.

The majority of soluble lysosomal enzymes utilize the mannose-6-phosphate-dependent pathway for delivery to the lysosome, however targeting pathways independent of mannose-6-phosphate exist. As described previously, the cells of patients with ML-II and ML-III are unable to synthesize the 6-phosphomannosyl marker on the *N*-linked oligosaccharides of the lysosomal hydrolases, consequently

the lysosomal hydrolases are secreted into the extracellular milieu rather than delivered to the lysosome. However, there is considerable variation in the degree of correctly targeted residual enzyme, which is dependent on both the particular lysosomal enzyme, and the tissue source. For example, fibroblasts derived from ML-II patients are deficient in several lysosomal enzymes; the residual activity of arylsulphatase A is 5% or less of normal, the residual activity of α-L-iduronidase is from 10 to 30% of normal, while β-glucosidase and acid phosphatase display normal activity (Suzuki, 1992). However, the intracellular levels of the majority of these lysosomal enzymes are near normal in liver, spleen, kidney and brain, yet all tissues are deficient in the phosphotransferase activity (Pfeffer, 1988; Nolan & Sly, 1989).

1.5.6 MANNOSE-6-PHOSPHATE-*INDEPENDENT* TRANSPORT

A number of proteins have been identified in which lysosomal targeting occurs independently of mannose-6-phosphate. In addition, the routes followed may vary from the mannose-6-phosphate-dependent pathway. This would suggest that there is at least one and probably more lysosomal targeting mechanisms independent of the mannose-6-phosphate signal, which are not necessary exclusive of each other, or of the mannose-6-phosphate pathway. The peptide sequence determinants of a number of proteins which utilize the independent pathways have begun to be defined. Lysosomal acid phosphatase (LAP) is synthesized and transported from the ER to the Golgi and then to the plasma membrane as a transmembrane precursor. LAP does not contain mannose-6-phosphate moieties, rather, an internalization signal in the cytoplasmic tail of LAP directs transport from the cell surface to endosomes and then to the lysosome. In the lysosome proteolytic cleavage releases the lumenal portion of LAP from the transmembrane portion to produce mature LAP (Gottschalk *et al.*, 1989). The internalization signal was recently defined as a hexapeptide which contained a tyrosine residue that was

conserved in a number of internalized proteins (Lehmann *et al.*, 1992). A tyrosine-containing tetrapeptide internalization motif was also identified at the carboxyl end of the lysosome-associated membrane protein 1 (LAMP1) (Guarnieri *et al.*, 1993). The predominant pathway followed by LAMP1 to the lysosome is the intracellular route from the Golgi, in contrast to LAP. Once within the lysosome, the internalization signal is inactivated by proteolysis at the carboxyl-terminal. The soluble lysosomal enzyme α-glucosidase appears to utilize both mannose-6-phosphate dependent and independent pathways during transport. Wisselaar *et al.* (1993) identified a membrane-bound form of the enzyme that retained the signal peptide and was lysosomally targeted by a mechanism independent of mannose-6-phosphate. The transport of a number of membrane-bound lysosomal enzymes has been shown to be independent of the mannose-6-phosphate pathway (Barriocanal *et al.*, 1986). Although the majority of the LSD are due to deficiencies in soluble lysosomal proteins, clarification of this pathway is relevant to diseases such as MPS-IIIC (Sanfilippo C syndrome), which is due to the deficiency of the lysosomal membrane enzyme, acetyl-CoA:α-glucosaminide *N*-acetyltransferase.

1.5.7 PROCESSING OF LYSOSOMAL PROTEINS

During transit through the prelysosomal compartments and residence in the lysosome, the majority of both soluble and membrane-bound lysosomal proteins undergo limited proteolytic processing and partial trimming of their oligosaccharide side-chains (reviewed by Neufeld, 1991; Hasilik, 1992). Proteolytic processing includes fragmentation and trimming at the amino- and carboxyl-termini of the protein, and is due to the concerted action of *endo*- and *exo*-proteinases present in the prelysosomal and lysosomal compartments. Oligosaccharide processing includes trimming of the carbohydrate residues, and dephosphorylation of the mannose-6-phosphate residues.

Whether these processes are an incidental consequence of the harsh degradative environment of the lysosome, or required for lysosomal enzyme function, is not definitely known for most enzymes. One can envisage the latter possibility as an elegant way to minimise inappropriate enzyme function outside of the lysosome. This is the mechanism utilized in the regulation of the cathepsins, a family of lysosomal proteases. The cathepsins are synthesized as inactive zymogen precursors and only subsequently activated by proteolysis in the appropriate location of the pre-lysosomal and lysosomal compartments (cited in Hasilik, 1992). However the former possibility may be relevant to other lysosomal enzymes, as the precursor species of most lysosomal enzymes observed in cell secretions are catalytically active. For example, the secreted precursor form of 4S was catalytically active toward the artificial substrate *p*-nitrocatechol sulphate (Steckel *et al.*, 1983). Furthermore, Anson *et al.* (1992b) demonstrated that the secreted precursor form of human recombinant 4S expressed in CHO exhibited similar kinetic parameters towards a naturally-derived 4S-specific trisaccharide substrate as the mature form of 4S isolated from human liver.

The following Sections of this Chapter will review the purification of the enzyme deficient in MPS-VI, and the clinical features, diagnosis, and therapy of this disorder. The Chapter will conclude with the attempts to obtain a correlation between a patient's clinical phenotype, biochemical phenotype and mutant genotype in a number of the LSD, including MPS-VI.

1.6 THE PURIFICATION OF 4S

The lysosomal enzyme deficient in MPS-VI is *N*-acetylgalactosamine-4-sulphatase (4S). However, the first reported purification of 4S preceded the identification of the 4S deficiency in MPS-VI, and its role in the metabolism of dermatan sulphate and chondroitin-4-sulphate. In retrospect, 4S was first partially purified in 1953 from ox liver, and was termed arylsulphatase B (ASB). Two arylsulphatases were identified, arylsulphatase A and B, which were distinguished on the basis of differential arylsulphatase activities toward an artificial arylsulphate-ester substrate (Roy, 1953). At that time the role of these enzymes *in vivo* was not known. Two main synthetic substrates were developed for the assay of 4S; *p*-nitrocatechol sulphate, also known as 2-hydoxy-5-nitrophenol sulphate (NCS; Roy, 1953; Baum *et al.*, 1959) and 4-methylumbelliferyl sulphate (4MUS; Christomanou & Sandhoff, 1977). However, a disadvantage of NCS and 4MUS was their relative non-specificity, as the arylsulphatases A and C could also hydrolyze these substrates. Over time, refinements to the assay conditions have increased the sensitivity and specificity of the assay, and have included either a combination of arylsulphatase-specific inhibitors in the assay or electrophoretic separation of the arylsulphatases prior to the assay (Chang *et al.*, 1981; Ratazzi *et al.*, 1973). Roy (1976) provides an informative historical review of the early purification and characteristics of the first members of the sulphatase family, arylsulphatases A, B, and C.

Twenty years after the initial purification of arylsulphatase B a human disorder was linked with it's deficiency. Stumpf *et al.* (1973) were the first to demonstrate that tissues and cultured fibroblasts from MPS-VI patients were deficient in arylsulphatase B. Identification of the *in vivo* substrate of arylsulphatase B was reported soon after by O'Brien *et al.* (1974) and Matalon *et al.* (1974). Their work on cultured fibroblasts of MPS-VI patients showed that the *in vivo* function of ASB was a *N*-acetylgalactosamine-4-sulphatase, which hydrolyzed the C-4 sulphate

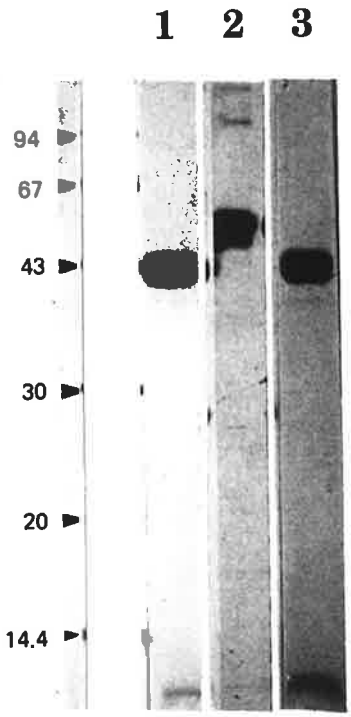
from the non-reducing-end terminal *N*-acetylgalactosamine residue of dermatan sulphate and chondroitin-4-sulphate (O'Brien *et al.*, 1974; Matalon *et al.*, 1974). Subsequent efforts were directed toward purification of the enzyme, which has been achieved by a number of groups from a variety of tissue sources (e.g. McGovern *et al.*, 1982; Steckel *et al.*, 1983; Gibson *et al.*, 1987; Jin *et al.*, 1992; Kobayashi *et al.*, 1992). The very low abundance of 4S in most tissues has hampered purification.

McGovern *et al.* (1982) purified 4S from human and cat liver, however a consistent determination of the molecular weight was not possible as a range of sizes were obtained with different analytical techniques. Steckel *et al.* (1983) achieved a purification of human placental 4S, which was composed of three species; a 47-kDa, 14.5-kDa and a 11.5-kDa species which were disulphide-linked. However, the molecular weight of 4S under non-reducing conditions was 57-kDa, which was 16-kDa smaller than the sum of the individual species. Polyclonal antibodies were raised to this preparation and used to identify two forms of 4S from human fibroblasts. Each form was composed of two species that were disulphide-linked; Form I was composed of a 47-kDa and a 11.5/14.5-kDa species, and Form II was composed of a 40-kDa and 31-kDa species. They also determined that arylsulphatase B activity was associated with Form I and a 64-kDa precursor species which was secreted into the medium, but not Form II. The presence of the mannose-6-phosphate marker on 4S and it's necessity for lysosomal targeting was also shown (Steckel *et al.*, 1983). Gibson *et al.* (1987) used highly specific monoclonal antibodies to purify 4S to homogeneity from liver, and also kidney, lung and urine, by immunoaffinity chromatography. The original purified 4S (obtained by a six-column chromatography procedure) and the immunopurified 4S both possessed a single species of molecular weight 57-kDa under non-reducing conditions, which was composed of two disulphide-linked species of 43- and 13-kDa.

FIGURE 1.4 SDS-PAGE OF PURIFIED 4S

(From Gibson *et al.*, 1987)

SDS-polyacrylamide gel showing human liver 4S purified by a six-column procedure (lanes 1 and 2), and immuno-purified using an anti-4S monoclonal antibody (ASB4.1) affinity column (lane 3). The samples in lanes 1 and 3 were reduced with dithioerythritol prior to electrophoresis. The molecular masses of the 4S polypeptide species are indicated in kDa on the left.


1
2
3
94
67
43
30
20
14.4

The 13-kDa species was later referred to as 8-kDa after a more accurate determination of it's size (Taylor *et al.*, 1990) (Fig. 1.4). Recently, 4S has been shown to contain an additional species of 7-kDa, which is disulphide-linked to the 43-kDa species (Kobayashi *et al.*, 1992). The 7-kDa species was not observed in previous reports as it required modifications of the usual conditions for electrophoresis and protein staining for detection. All three species are derived from the proteolytic processing of a single precursor polypeptide, as discussed in detail in Chapter 3.

The arylsulphatases A, B, and C were originally classified together based on the ability to hydrolyse arylsulphate-ester substrates *in vitro*. *In vivo* however, the arylsulphatases exhibit distinct substrate specificities, intracellular locations, and are associated with different inherited syndromes when deficient. Arylsulphatase A (ASA) is a soluble lysosomal sulphatase involved in the hydrolysis of the glycolipid cerebroside-3-sulphate. A deficiency of ASA leads to metachromatic leukodystrophy, a disorder with predominantly neurological involvement (Kolodny, 1989). Arylsulphatase B or 4S is also a soluble lysosomal enzyme, but the *in vivo* substrate is the terminal *N*-acetylgalactosamine-4-sulphate residues of dermatan sulphate and chondroitin-4-sulphate. 4S is deficient in MPS-VI, in which the main sites of pathology are primarily somatic, as described in Section 1.7.2. Arylsulphatase C or steroid sulphatase (STS) is a non-lysosomal enzyme resident in the microsomal membrane, where it desulphates a variety of 3β-hydroxy steroid sulphates. A deficiency of STS leads to the skin disorder X-linked ichthyosis (Shapiro, 1989). There is some controversy about whether arylsulphatase C and steroid sulphatase are in fact the same enzyme (Shankaran *et al.*, 1991). Interestingly, the three arylsulphatases and all other known sulphatases are deficient in multiple sulphatase deficiency, which is discussed in Section 5.8.1.

1.7 MUCOPOLYSACCHARIDOSIS TYPE VI

1.7.1 HISTORICAL REVIEW

Maroteaux-Lamy syndrome was first recognized as a distinct mucopolysaccharidosis by Maroteaux *et al.* (1963), on the basis of biochemical and clinical features. Prior to their work, Maroteaux-Lamy syndrome was considered a Hurler syndrome variant. In particular, they identified the excessive excretion of dermatan sulphate (chondroitin sulphate B) in the urine, in contrast to Hurler syndrome, in which both dermatan sulphate and heparan sulphate are excreted at elevated levels. Furthermore, the progressive mental retardation observed in Hurler syndrome was not observed in this disorder. Maroteaux & Lamy proposed the term polydystrophic dwarfism for the disorder (Maroteaux & Lamy, 1965), however since then the disease has almost universally been referred to as Maroteaux-Lamy syndrome. As more patients were identified with the disorder, a variation in clinical severity was observed, which confounded precise clinical diagnosis. It was speculated that MPS-VI existed in two forms, a severe MPS-VI A form and a milder MPS-VI B form (Spranger *et al.*, 1970; McKusick, 1972). Whether the two forms were due to different mutations in the same gene or due to mutations in two different genes could not be resolved. At present, a wide spectrum of clinical phenotypes are recognized, forming a continuum between the two forms previously proposed.

Barton & Neufeld (1972) isolated a corrective factor from normal human urine which, when added to cultured MPS-VI patient cells deficient in the factor, reduced the intracellular accumulation of sulphated GAG. Soon after, Stumpf *et al.* (1973) were the first to demonstrate that tissues and cultured fibroblasts from MPS-VI patients were deficient in arylsulphatase B activity. In retrospect, the corrective factor of Barton & Neufeld (1972) was presumed to be arylsulphatase B. Arylsulphatase B was shown to be an *N*-acetylgalactosamine-4-sulphatase,

following identification of the *in vivo* substrates, dermatan sulphate and chondroitin-4-sulphate (O'Brien *et al.*, 1974; Matalon *et al.*, 1974). MPS-VI is inherited as an autosomal recessive disorder; the 4S gene is located on chromosome 5 (Maroteaux & Lamy, 1965; Spranger *et al.*, 1970; Hellkuhl & Grzeschik, 1978; DeLuca *et al.*, 1979).

1.7.2 CLINICAL DESCRIPTION

The clinical features of MPS-VI patients vary from mild through to severe extremes (reviewed in Spranger *et al.*, 1970; McKusick, 1972; Neufeld & Muenzer, 1989). The precise meaning of the terms `severe' or `mild' with reference to the clinical phenotype of MPS-VI is poorly defined in the scientific literature, as these classifications have been made by different observers with different clinical criteria and experience. At present, a uniform and systematized clinical severity scoring system for MPS-VI does not exist. Such a system, if developed, may enable the quantitative assessment of severity, and therefore refine attempts to correlate the patient's genotype with the clinical and biochemical phenotype. The quantitative assessment would be complicated by age dependent factors, and the progressive nature of the disease course. Isbrandt *et al.* (1994) have proposed a severe, intermediate and mild classification based on age of onset. The characteristics of the severe or infantile form are an early onset prior to the age of 2 years and rapid disease progression, while the intermediate or juvenile form is manifested in late childhood, and the mild or adult form is diagnosed after the second decade. Disease progression in the juvenile and adult forms is typically reduced relative to the infantile form.

Clinical diagnosis of MPS-VI is usually based on somatic features; severe MPS-VI is often first noted between the ages of 2 to 3 years because of short stature in the child. In general, growth may or may not proceed normally for the first few years of life, but by the age of 6-8 years has virtually stopped, with the ultimate height of

severely affected patients ranging from 110-140 cm. Other features typical of severely affected MPS-VI patients include a shortened trunk, protrusion of the sternum and/or abdomen, genu valgum (knock-knees), and kyphosis or lumbar lordosis. Craniofacial morphology is abnormal, with an enlarged head and coarse facial features. Progressive corneal clouding is frequent, and often leads to blindness. Four siblings afflicted with the severe form of MPS-VI are shown in Fig. 1.5.

Hydrocephalus is an occasional complication of severely affected patients, and if left unchecked may lead to papilloedema (Goldberg *et al.*, 1970), visual field defects and ultimately mental retardation. Hydrocephalus is probably due to an impaired resorption of cerebrospinal fluid, on account of dermatan sulphate deposition in the meninges (Goldberg *et al.*, 1970). However, the majority of MPS-VI patients exhibit mental development within the normal range, consistent with the suggested absence of GAG deposition within the brain itself. In contrast, profound mental retardation is a characteristic of a number of the other MPS, including MPS-IH, severe MPS-II, and the four subtypes of MPS-III. Mental retardation has been reported in patients with arylsulphatase B deficiency (Taylor *et al.*, 1978), however, as the patients presented clinical and biochemical findings that were atypical of MPS-VI, the diagnosis of MPS-VI requires confirmation.

Another neurological complication is compressive myelopathy of the spinal cord due to thickening of the dura mater, usually in the cervical region, which can lead to paresis. Hypoplasia of the odontoid process of the cervical vertebra can cause spastic paraplegia (McKusick, 1972), due to subluxation of the cranio-vertebral joint and consequent shearing stresses on the spinal cord. Varying severity of deafness is also observed. Both neurosensory and conductive defects are responsible, the latter may be due to recurrent middle ear infections often seen in MPS patients (Shigematsu *et al.*, 1991). The propensity for frequent ear infections

is probably due to the multiple soft tissue deformities observed in the nasopharyngeal region. Other pharyngeal abnormalities, such as an enlarged tongue, enlarged adenoids, congested mucous membranes, and abnormalities of the laryngeal and tracheal cartilage, can cause airway obstruction and predispose patients to respiratory complications and sleep apnoea.

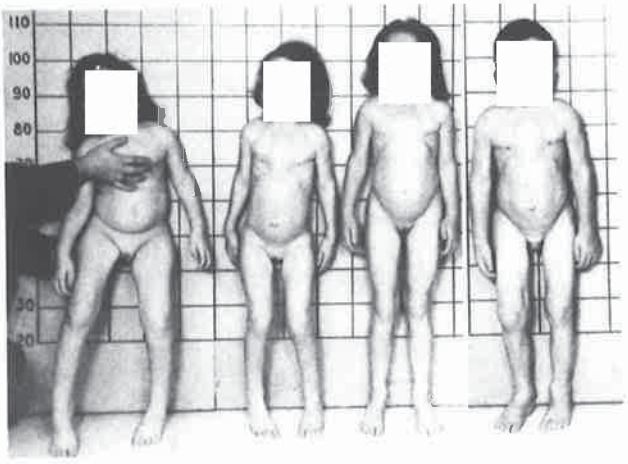
Restriction in the articular motion of knee, hip and elbow joints is noted in the first years of life, and can deteriorate over time. Flexion contractures develop in the fingers due to medial nerve compression (carpal tunnel syndrome), and can lead in later childhood to clawed hands. Dysostosis multiplex is evident upon radiological examination of the skeleton. For example, pelvic abnormalities include small, flared iliac wings, and hypoplasia of the acetabulum, where the pelvis joins the femur. Epiphyseal dysplasia of the proximal femur is also observed. Pathogenesis of the bone abnormalities in MPS-VI may be due to perturbations in the process of bone matrix mineralization. Kazama *et al.* (1992) observed that the process of bone matrix mineralization in rat metaphyseal bone was accompanied by a sharp reduction in proteoglycans, including dermatan sulphate proteoglycans. The proteoglycan reduction is conceivably dependent on normal lysosomal degradation of the GAG. The joint stiffness is probably a combination of both metaphyseal deformities of the bone, and thickening of the tissues that comprise the joint capsules, due to DS deposition (Muenzer, 1986).

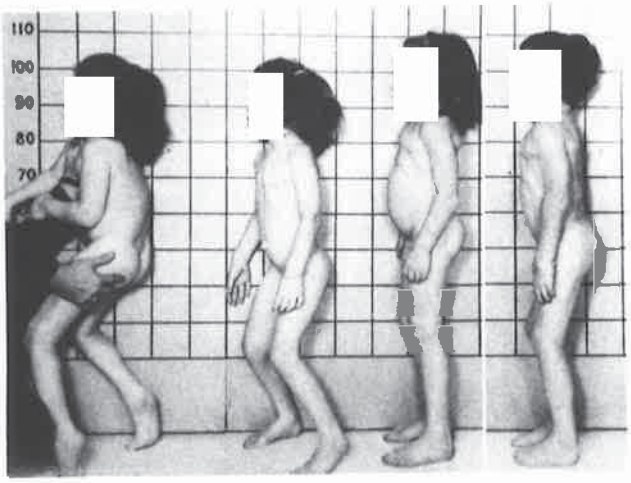
In addition to the skeletal and soft tissue abnormalities, the viscera can be involved. The liver is consistently enlarged in patients over the age of 6 years, and splenomegaly is also observed, although not as frequently (Spranger *et al.*, 1970). Cardiac involvement in MPS-VI is frequent. Adult-onset valvular heart disease, in which the aortic and mitral valves are affected, can occur in mild MPS-VI patients. The valve leaflets become thick and fibrotic, and the normal motion is diminished due to infiltration of dermatan sulphate (Tan *et al.*, 1992).

FIGURE 1.5 MPS-VI: SEVERE CLINICAL PHENOTYPE

(From McKusick, 1972)

Four siblings affected with the severe form of MPS-VI. The patients were aged 14, 9, 13, and 11 years (left to right) at the time this photograph was taken. All four siblings have followed a similar clinical course, the oldest sibling died at 15.5 years (Neufeld & Muenzer, 1989).


110
100
90
80


110
100
90
80
70

An acute infantile cardiomyopathy has also been observed (Hayflick *et al.*, 1992). In the severe form of MPS-VI, heart failure is the most common cause of death, usually before 20 years of age. Clinically mild patients however, may not be diagnosed until 10 to 20 years of age and live to be 50 to 60 years of age or longer (Hopwood & Morris, 1990). There is no evidence to suggest partial MPS-VI phenotype expression in heterozygotes.

The precise frequency of MPS-VI in the population is unknown at present as no studies have appeared in the literature. The disorder is very rare in all populations examined, and thus appears panethnic. The incidence of all the MPS disorders has been estimated at from 1 in 10,000 (Hopwood & Morris, 1990) to 1 in 30,000 births (Muenzer, 1986), but there is uncertainty associated with these figures due to the paucity of epidemiological data for these disorders. A `guesstimate' suggests that MPS-VI is rarer than MPS-IH, which has an incidence of around 1 in 100,000 births (Neufeld & Muenzer, 1989).

1.7.3 ANIMAL MODELS

Animal models of a human disorder can be very useful in the investigation of disease pathology. In addition, animal models enable the evaluation of experimental treatment protocols that would be premature to attempt on human patients. Animal models exist for a number of the MPS disorders including MPS-VI, for which a total of 5 separate models have been found to date, 4 feline and 1 rat.

Jezyk *et al.* (1977) reported a Siamese cat which displayed similar biochemical and clinical features compared to the human MPS-VI syndrome. This initial report was followed by a more complete description of the pathological characteristics and the enzymatic defect (Haskins *et al.*, 1979). Three other groups have subsequently identified MPS-VI in cats (Breton *et al.*, 1983; Gasper *et al.*, 1984; Di Natale *et al.*,

1992). All 4 feline cases of MPS-VI have occurred in the Siamese breed; each of the cases had been obtained independently and were apparently unrelated. There is evidence for genetic heterogeneity in at least two of the MPS-VI cat pedigrees (McGovern *et al.*, 1985). Recently a rat model of MPS-VI has also been described (Yoshida *et al.*, 1993a; 1993b).

In both the feline and rodent MPS-VI models, the disease is inherited as an autosomal recessive disorder, however both models show clinical and biochemical features that are closely similar but not identical compared to the human disease. For example, the hepatomegaly observed in human MPS-VI is not seen in the feline or the rat models. The molecular lesions in the animal models has not yet been identified, however with the cloning of the feline 4S cDNA (Jackson *et al.*, 1992), it is likely that their identification is imminent.

1.8 DIAGNOSIS OF MUCOPOLYSACCHARIDOSIS TYPE VI

Precise diagnosis of a particular MPS disorder based solely on clinical and radiological criteria can be difficult, due to the overlap of features between the different MPS types, and the variable presentation of features in a given MPS. However, diagnostic accuracy has increased with the use of biochemical-based assays. MPS-VI can be diagnosed biochemically in two ways; firstly by detection of the elevated excretion of dermatan sulphate in the urine, and secondly by the assay of the 4S enzyme deficiency. As discussed previously, the elevated lysosomal storage and urinary excretion of dermatan sulphate in MPS-VI is due to the deficiency of 4S. A major impetus for the work presented in this thesis was the development of recombinant DNA-based techniques to aid in diagnosis of MPS-VI, and to gain an insight into the molecular basis of the 4S enzyme deficiency in the MPS-VI disorder.

1.8.1 DERMATANSULPHATURIA

Ever since the original observation of the excessive urinary excretion of mucopolysaccharides in a Hurler patient (Dorfman & Lorincz, 1957), the hypermucopolysaccharidiuria found in MPS disorders has been exploited as the basis for most screening tests. Urinary screening is the simplest biochemical test for MPS, and many variations have been described in the literature. These tests have the advantages of speed and low cost, and thus serve as a useful guide for preliminary evaluation. However, the tests can exhibit a high percentage of false-negative and false-positive results. The specificity of many of the tests is low, and many are unable to distinguish each individual class of mucopolysaccharide in the total excreted pool.

The characteristic increase in urinary excretion of dermatan sulphate in MPS-VI patients was first observed by Maroteaux *et al.* (1963). In general for MPS-VI, the increasing clinical severity correlates with increasing dermatan sulphate excretion. MPS-VI patient urine can also contain elevated concentrations of the monosaccharides *N*-acetylgalactosamine-4-sulphate and *N*-acetylgalactosamine-4,6-disulphate, as discussed previously. The monosaccharide concentration in urine is also generally commensurate with the clinical severity (Hopwood & Elliott, 1985).

An improved assay for elevated mucopolysaccharidiuria was reported by Hopwood & Harrison (1982), in which an electrophoresis technique was used to resolve the different mucopolysaccharide species present in patient urine. Analysis of the relative distribution of individual mucopolysaccharide species allowed a tentative diagnosis of the particular type of MPS to be made. In addition, the technique did not appear to give false-negative or false-positive results. The importance of the separation and analysis of each GAG type was illustrated by Tonnesen *et al.* (1991),

who reported a mildly affected MPS-VI patient with an increased urinary excretion of dermatan sulphate, although the total MPS excretion was within normal limits.

As a result of the dermatan sulphate accumulation in MPS-VI patients, peripheral leukocytes can exhibit metachromatic inclusion bodies. However, this phenomenon is also exhibited in a number of other MPS disorders and therefore is not of precise diagnostic value. The assays based on urinary excretion of dermatan sulphate or sulphated monosaccharides, although useful, are unable to provide a precise diagnosis with indisputable certainty.

1.8.2 4S ACTIVITY

The development of specific assays of enzyme activity has played an indispensable role in the refinement of MPS diagnosis. Enzymatic assays can be performed on patient samples including leukocytes and cultured fibroblasts, using a variety of artificial and naturally-derived substrates. Prenatal diagnosis can also be performed on cultured amniotic cells, and cells obtained from the chorionic villus, which enables detection at an earlier stage.

Early attempts to diagnose the 4S deficiency utilized synthetic artificial substrates, either the chromogen NCS or the fluorophore 4MUS. The synthetic substrates, although convenient to use, had the disadvantage of poor specificity for 4S, as they also functioned as substrates for the other arylsulphatases as mentioned previously. In addition, the artificial substrates bore little resemblance to the natural substrate *in vivo*, and could generate artefactual results. Modified assays were designed in which the competing arylsulphatase activities were either selectively inhibited, or electrophoretically separated (Christomanou & Sandhoff, 1977; Chang *et al.*, 1981; Ratazzi *et al.*, 1973). Hopwood *et al.* (1986) developed an assay for 4S that utilized a radiolabelled oligosaccharide substrate that was derived from chondroitin-4-sulphate. In comparison to the artificial substrates, the

oligosaccharide substrate was much more akin to the natural *in vivo* substrate, and was highly specific for 4S. However, the synthesis of the substrate is cumbersome, and the assays are more laborious than those that utilize the artificial substrates. All three substrates have been utilized to diagnose MPS-VI in patient cultured fibroblasts, however each has limitations. In addition, the very low abundance of 4S in most tissues, especially patient samples, has taxed the sensitivity of the assays.

The development of monoclonal antibody technology and immunochemical techniques has contributed much to the refinement of assays for 4S activity and protein (reviewed in Brooks, 1993). Brooks *et al.* (1991a) developed a 4S activity assay that combined the specificity of an anti-4S monoclonal antibody with the advantages of the fluorogenic substrate 4MUS. The anti-4S monoclonal antibody was used to specifically immunoadsorb 4S from other sulphatases in cell extracts, and the 4MUS substrate was used to detect the captured 4S-specific arylsulphatase activity. In the patient sample examined, the residual activity observed in cultured fibroblasts was about 3% of normal controls. An immuno-quantification assay specific for 4S protein has also been developed (Brooks *et al.*, 1990). The results obtained from the immunochemical investigation of the residual 4S activity and protein in MPS-VI patients will be discussed in Section 1.10.3. All the 4S biochemical assays reviewed above are not suitable for the definitive identification of carriers, as the assays are unable to reliably separate the normal and carrier populations due to the overlap in the range of values.

1.8.3 DETECTION OF 4S GENE DEFECTS

At the commencement of the work contained in this thesis, diagnosis of MPS-VI at the DNA level was not possible, as the gene was not isolated and RFLP analysis was not available. Theoretically, once a mutation is identified, it can be readily used to identify and distinguish normals from heterozygotes and MPS-VI patients,

the latter even before the onset of clinical signs. Therefore, the cloning of the 4S gene was a major initial aim of the work, so as to assess the feasibility of MPS-VI diagnosis at the gene level.

1.9 THERAPY OF MUCOPOLYSACCHARIDOSIS TYPE VI

1.9.1 CURRENT APPROACHES TO MPS-VI PATIENT MANAGEMENT

Current treatment for all the MPS disorders including MPS-VI is restricted to ameliorative care of the patient, due to the absence of specific therapies targeted to the causative enzyme deficiencies. The wide range of organs affected in MPS-VI and the progressive deterioration observed, especially in severe cases, limit the success of the medical intervention to arresting or slowing the rate of disease progression. Specific interventions are discussed in Muenzer (1986), and Neufeld & Muenzer (1989).

Hydrocephalus is observed in severely affected MPS patients, including MPS-VI (Goldberg *et al.*, 1970; Litjens *et al.*, 1992) and can cause mental retardation if not treated. The hydrocephalus is usually resolved by means of a ventriculojugular shunt, which relieves the increased intracranial pressure. Shunting performed for hydrocephalus has also resolved papilloedema (Goldberg *et al.*, 1970). Corneal transplantation, in which the diseased cornea is removed and replaced with a normal donor cornea, has been used as a treatment for clouded corneas. The hearing impairment (Spranger *et al.*, 1970; Shigematsu *et al.*, 1991), may benefit from hearing aids. The obstructive airway disease, due to mucopolysaccharide deposition in and consequent enlargement of the tissues of the airway including tongue, tonsils, and adenoids, can be managed by tonsillectomy, adenoidectomy, or tracheostomy, depending on the degree of severity. The joint stiffness of MPS-VI

can be partially relieved by physical therapy, while surgery is necessary to relieve severe joint contractures due to nerve entrapment, particularly of the carpal tunnel. In MPS-VI, as in other MPS disorders, the cardiovascular system is affected, in particular the aortic and mitral valves. Replacement of the diseased valves with artificial valves has proved beneficial (Tan *et al.*, 1992).

Unfortunately, a number of the clinical features characteristic of MPS disorders can complicate surgical treatment. In particular, induction of anaesthesia is hampered and the risk of mortality can be increased due to the obstructive airway disease, the short and unstable neck, the abnormal rib cage and the compromised cardiac function (Neufeld & Muenzer, 1989; Tan *et al.*, 1992).

1.9.2 FUTURE THERAPEUTIC APPROACHES TO MPS-VI

The main therapeutic approaches under development for the MPS including MPS-VI are attempts to correct the enzymatic deficiencies. These include, enzyme replacement therapy (ERT), bone marrow transplantation (BMT), and gene therapy. All of the experimental therapeutic approaches need to address the problem of targeting sufficient enzyme to the diverse sites of pathology. Studies of the 4S activity in MPS-VI cultured cells (Brooks *et al.*, 1991b) suggest that a residual enzyme activity of approximately 5% or less is associated with GAG storage. Therapeutic approaches which deliver a higher residual activity to the patient cells could conceivably correct the GAG storage responsible for the pathology. However, by the time the clinical features of the disorder are sufficiently manifest to enable diagnosis, the damage wrought by the stored GAG at the cellular, tissue, and organ level may be irreversible. Ideally, therapeutic intervention should take place within the first few months of life, perhaps even prior to the development of clinical symptoms. The importance of sensitive and specific diagnostic techniques is therefore obvious. An advantage of MPS-VI

relative to the other MPS with respect to therapy is the absence of neurological deterioration in the brain. Consequently, it is not considered necessary to achieve transfer of enzyme across the restrictive blood-brain barrier into the brain.

The wide clinical variation in the severity of the MPS hampers the accurate prediction of the rate of disease progression in a patient. However, a reasonably confident prediction of prognosis is necessary for the assessment of the clinical efficacy of a therapeutic procedure. Analysis of the mutant genotype may enable the untreated prognosis to be predicted. For example, a conclusive reduction in disease progression has been reported in BMT-treated MPS-I patients (Hopwood *et al.*, 1993b). The clinical phenotypes of MPS-I patients who had received BMT at an early age were retrospectively compared with those who had not been BMT treated. Both groups were homozygous for a mutation that if untreated, resulted in a consistently severe phenotype and death before the age of 6 years. However, the two treated patients showed relatively mild clinical features and were in their teens. If consistent genotype-phenotype correlations are also observed for other MPS mutant alleles, the clinical efficacy of the various therapeutic approaches will be able to be tested more rigourously than previously possible.

1.9.2a ENZYME REPLACEMENT THERAPY (ERT)

Early attempts at enzyme replacement therapy of MPS disorders involved administration of `corrective factors' purified from normal human tissue to the patient deficient in that factor. This approach exploited the observation of Fratantoni and others (e.g. Fratantoni *et al.*, 1969) that the GAG storage in MPS cells could be reduced by the addition of corrective factors (lysosomal enzymes) to the culture medium. However, clinical benefit to the patient was either transient or nonexistent, and so the early attempts at enzyme replacement therapy were not pursued further. In hindsight, the absence of therapeutic benefit was most likely due to the fact that insufficient enzyme was isolated, or that it was inefficiently

targeted to the patient's tissues. Presently ERT is under re-evaluation, as the original technical difficulties have been substantially overcome by recombinant DNA technology, which can produce large quantities of 4S and other lysosomal enzymes. Furthermore, there is an increased understanding of the processes of lysosomal enzyme targeting, receptor-mediated endocytosis and direct cell-cell transfer (Bou-Gharios *et al.*, 1993).

Recombinant 4S was used to achieve correction of the enzymatic deficiency in cultured MPS-VI fibroblasts (Anson *et al.*, 1992b). The purified recombinant 4S was efficiently endocytosed, processed normally and transported to the lysosome, and initiated degradation of stored dermatan sulphate. A similar degree of correction was achieved by Bielicki *et al.* (1993) in MPS-II fibroblasts using recombinant iduronate-2-sulphatase. Replacement therapy has not yet been attempted in human MPS patients using enzyme synthesized by recombinant DNA techniques. However, the clinical improvement reported in a feline model of MPS-VI after administration of recombinant human 4S (Prof. J. Hopwood, personal communication), are grounds for cautious optimism concerning the clinical efficacy of enzyme replacement therapy.

1.9.2b BONE MARROW THERAPY (BMT)

Further attempts at correction of the enzyme deficiencies in MPS disorders have been via heterologous bone marrow transplantation (BMT), in part due to the failure of the early ERT attempts. For reviews of BMT especially applied to the LSD, see Hobbs (1987), Hoogerbrugge & Vossen (1990), and Krivit *et al.* (1992). Although the bone marrow is not a significant site of pathology in the MPS disorders, the rationale for BMT as a potential therapeutic approach is based on a number of factors. Firstly, the bone marrow derived from the donor supplies enzymatically normal leukocytes to the patient. The donor-derived cells could then transfer normal lysosomal enzymes to the patient's deficient cells by the

mechanisms of direct cell-cell contact, and by secretion-endocytosis. Secondly, the engraftment of the donor pluripotent haematopoietic stem cells would enable the permanent provision of the deficient lysosomal enzymes to the various tissues in the patient. In order to minimize the problems of immunological rejection, BMT is performed using bone marrow obtained from a normal heterologous donor whose HLA antigens are matched (allogeneic) with the patient's. The patient's bone marrow is first destroyed with cytotoxic agents to enable engraftment of the donor BM. Immunosuppresive drugs are also given to the patient to reduce graft versus host disease (GVHD), where the donor marrow rejects the host.

BMT has been performed on MPS-VI patients and also the feline MPS-VI model. Krivit *et al.* (1984) reported biochemical and some clinical improvement in the condition of a 13 year old girl with severe MPS-VI who received an allogeneic BMT. Two years after engraftment an improvement in clinical features was observed and included reduction in hepatosplenomegaly and normalization of cardiopulmonary function. Visual acuity increased, but there was no reduction in corneal clouding. Joint mobility of the hip, knee and elbow were also improved, which presumably reflected changes in the soft tissues. However, there was no improvement in the skeletal abnormalities, nor was there an increase in the patient's height. Biochemical analysis showed a reduction in the excretion of mucopolysaccharide to normal levels, reduction in the intracellular storage of GAG in a number of cells, and increasing levels of arylsulphatase activity in liver biopsy specimens. In summary, the clinical improvement was notable but limited. After consideration of the age of the patient and the time elapsed between diagnosis and BMT, the authors felt that the severity of bone deformities and dysmorphology are likely to be reduced if BMT was performed earlier. BMT has also been performed on an MPS-VI cat, and a beneficial effect was reported including a reduction in facial dysmorphology, improvement in walking ability and increased movement in head, neck and mandible (Gasper *et al.*, 1984). In addition, the corneal clouding was

completely resolved, in contrast to the human patient described by Krivit *et al.* (1984).

The therapeutic benefits of BMT in patients with LSD are variable, and definitive assessment is currently premature, but at best a marked reduction in the rate of disease progression has been observed (Hobbs *et al.*, 1981; Hoogerbrugge & Vossen, 1990; Krivit *et al.*, 1992; Hopwood *et al.*, 1993b). The modest clinical benefits obtained with BMT need to be weighed against the potential problems of graft rejection (or failure of donor engraftment), and the increased risk of mortality associated with the procedure.

Beside donor-derived bone-marrow, other tissues have been proposed for transplantation, for example amnion membrane cells. Amnion cells secrete a variety of lysosomal hydrolases into the medium in culture, and were thought to be non-immunogenic in a heterologous host. However, recent attempts with amnion membrane implantation in 19 patients with MPS-I, II and III were not successful at obtaining stable engraftment, enzyme secretion, or significant clinical improvement (Muenzer *et al.*, 1992).

1.9.2c GENE THERAPY

In the context of the LSD, gene therapy seeks to augment the impaired function of the patient's defective gene by the transfer of a normal gene copy, which enables the patient's cells to manufacture the normal gene product. The experimental design, current applications to genetic and other diseases, and the future prospects of human gene therapy have been reviewed recently (Morgan & Anderson, 1993). Gene therapy trials have commenced in patients with severe combined immunodeficiency (SCID), a deficiency of adenosine deaminase (Blaese, 1993). Although gene therapy is not yet a therapeutic reality for any of the LSD in humans, it is being actively pursued in animal models.

In brief, current gene therapy protocols involve the collection of the recipient cells from the patient, and the introduction of the gene of interest using an appropriate vector. After selection, the genetically-modified cells are reintroduced into the patient. An alternative approach which is also under development is the delivery of the gene of interest directly into the patient. The utilization of autologous or patient-derived cells theoretically circumvents the problems associated with donor tissue incompatibility, as observed in BMT. Bone marrow-derived haematopoietic progenitor cells are considered a prime candidate for the recipient tissue (Krauss, 1992), for reasons previously outlined in the discussion of BMT. Gene transfer methods based on retroviral vectors are currently the most common, but vectors under development include adenovirus, vaccinia virus, adeno-associated virus and herpesvirus, as well as non-viral transfer methods (Morgan & Anderson, 1993).

Gene therapy of murine MPS-VII has been achieved by transgenesis (Kyle *et al.*, 1990), and by retroviral-mediated gene transfer into haematopoietic stem cells (Wolfe *et al.*, 1992). Retroviral gene transfer has also enabled the correction of glucocerebrosidase deficiency in the bone marrow stem cells of Gaucher patients (Fink *et al.*, 1990). Other retroviral-mediated gene transfers to stem cells are in progress for the LSD, canine fucosidosis (Graham Stewart & Don Anson, personal communication). An alternative to gene transfer into bone marrow-derived stem cells was reported by Moullier *et al.* (1993). Skin fibroblasts from MPS-VII mice were removed, and a retroviral vector carrying the β-glucuronidase cDNA was introduced into the cells. The genetically modified cells were then reintroduced into the mice as an autologous transplant. The `neo-organ' supplied enzyme, which corrected storage in the liver and spleen, but did not reverse the skeletal deformities or neurological deficits.

Recent studies in a variety of LSD have shown that retroviral-mediated transfer of functional gene copies into patient cultured cells has successfully restored enzyme activity to levels in excess of normal. For example, gene transfer of arylsulphatase A into metachromatic leukodystrophy fibroblasts (Rommerskirch *et al.*, 1991), α-L-iduronidase into MPS-I fibroblasts (Anson *et al.*, 1992a), and 4S or arylsulphatase B into MPS-VI fibroblasts (Peters *et al.*, 1991). In addition, the first two studies demonstrated that the impaired degradation of stored *in vivo* substrates was normalized subsequent to the transfer of the functional gene. The marked over-expression of the recombinant enzyme in these experimental systems is considered unnecessary for complete correction *in vivo*. A slight enzyme surplus will probably supply sufficient enzyme to the deficient tissues.

Over-expression may actually be counterproductive in some cases. Anson *et al.* (1992a) utilized three retroviral constructs that expressed α-L-iduronidase in MPS-I fibroblasts at levels of about 10, 150 and 250-fold over normal, non-MPS-I cells. A paradoxical inverse relationship was observed between the degree of α-L-iduronidase over-expression and the degree of correction of stored GAG substrate. They proposed that the effect was due to saturation of the mannose-6-phosphate receptor, such that lysosomal targeting of the enzymes was reduced. In analogous studies of retrovirally-introduced 4S in cultured MPS-VI cells, Anson *et al.* (1993) demonstrated that a 200-fold over-expression relative to normal cells corrected the enzymatic defect, but did not correct the dermatan sulphate storage. The 4S over-expression induced the deficiency of the other sulphatases, presumably due to competition for a sulphatase-specific processing mechanism or factor. The relevance of these findings in tissue culture to gene therapy *in vivo* is not known, and awaits gene therapy in animal models.

1.9.3 CONCLUSION

In conclusion, the results obtained so far with ERT, BMT and gene therapy, suggest that a challenging problem will be to achieve the targeting of sufficient enzyme to particular sites of clinically significant pathology, such as the skeleton and cornea, and reverse the GAG storage. Animal models have been and will continue to be indispensable for the development of the therapeutic methodologies. An important prerequisite for the clinical assessment of a therapeutic approach is the prediction of the prognosis in a patient if left untreated. This is not a trivial task in an MPS disorder, as a wide range of clinical severities are observed, which vary in parameters such as age of onset, extent of organs affected, and the rate of disease progression. The analysis of the patient's mutant genotypes may account for the observed clinical (and biochemical) heterogeneity. The concept of a correlation between the patient's genotype and phenotype will be discussed in the next Section.

1.10 GENOTYPE - PHENOTYPE CORRELATION

1.10.1 INTRODUCTION

Genetic heterogeneity had long been suspected as a factor in the observed variation in clinical severity of many genetic disorders. Recent research in many groups has focused on attempts to define the relationship between the clinical features, the biochemical perturbations, and the underlying gene defects. Attempts to correlate the clinical phenotype with the biochemical phenotype and mutant genotype for a large number of genetic disorders is an area of intense interest at present. The work presented in this thesis sought to refine the correlation between the genotype and the clinical and biochemical phenotype in MPS-VI. The correlation of the identified mutation with the clinical and biochemical manifestations of the disease has great utility. For example, a genotype-phenotype correlation may contribute to

the prediction of an accurate disease prognosis. In addition, a genotype-phenotype correlation may also contribute to prenatal diagnosis and genetic counselling of carriers, and the assessment of therapeutic protocols for MPS-VI. This Section will examine the features and limitations of each of the data sources necessary for the correlation, and will review the success or otherwise of genotype-phenotype correlations in a number of genetic diseases, including MPS-VI.

1.10.2 CLINICAL PHENOTYPE

Lysosomal storage disorders including MPS-VI display a characteristic wide continuum of clinical heterogeneity, which ranges from relatively benign to very severe. Clinical heterogeneity is observed in such features as the age of disease onset, the number of organ systems involved, the extent of pathology in the affected organs, and the rate of progression. Attempts to refine relative clinical severity beyond broad categories has been difficult, due to a number of factors. For example, in the MPS, the overall clinical phenotype is composed of the sum of the individual effects on a diverse group of organs, as seen in MPS-VI (Section 1.7.2). Quantification of the total severity is hampered by the uncertainty of the individual contribution of each feature to the disease morbidity, and hence what weighting to assign to each clinical feature. Furthermore, the clinical phenotype may degenerate over time, often at different rates in different tissues. Isolated parameters of clinical severity can be misleading, for example, age at diagnosis may be affected by social or other factors that are independent of the true clinical severity.

1.10.3 BIOCHEMICAL PHENOTYPE

The biochemical phenotype in MPS disorders including MPS-VI can be obtained using a number of techniques. A number of these methods for the determination of the biochemical phenotype were discussed previously in Section 1.8. In review,

these methods include the quantification of the characteristic urinary excretion of elevated levels of sulphated glycosaminoglycans, and quantification of the residual enzyme activity present in cultured cells. More sophisticated analytical techniques using monoclonal antibodies can provide information on the intracellular transport and processing of the deficient enzyme (Taylor *et al.*, 1990), the amount of mutant enzyme (Brooks *et al.*, 1990), and determine the mutant enzyme kinetics (Brooks *et al.*, 1991b). The analysis of MPS-VI cell lines using these techniques will be discussed below. In addition, attempts to correlate the clinical heterogeneity with the biochemical phenotype will be discussed.

In MPS-VI, heterogeneity is observed in the biochemical phenotype. For example, the amount of both residual enzyme activity and protein in a collection of patient cultured fibroblasts was found to be clustered over a narrow range from undetectable to about 5% of normal controls (Brooks *et al.*, 1991b; Gary Gibson, personal communication). Urinary analysis of excreted dermatan sulphate in MPS-VI patients showed a greater range, from barely increased excretion to about 65-fold over normal controls (Brooks *et al.*, 1991b). Attempts have been made to correlate the clinical severity with either the residual 4S activity or the dermatansulphaturia, based on the assumption that the more reduced the residual enzyme concentration or increased the dermatansulphaturia, the more clinically severe the disorder. In general, this hypothesis has been validated by the approximate correlation between biochemical and clinical phenotypes (dermatansulphaturia-clinical phenotype correlation, see Hopwood & Morris, 1990; Brooks *et al.*, 1991b). However, in some MPS-VI patients, apparent deviation from this pattern have been observed in the comparison of the residual enzyme activity with the phenotype. Poor correspondence between the clinical phenotype and the *in vitro* biochemical phenotype has been observed in other disorders. For example, metachromatic leukodystrophy is due to the deficiency of the lysosomal enzyme arylsulphatase A. Biochemical confirmation of this disorder has been complicated

by a clinically normal but enzymatically deficient variant, arylsulphatase A pseudodeficiency, in which the residual enzyme activity is about 10% of the mean normal range (Gieselmann *et al.*, 1989; C. Morris, personal communication).

Theoretical analysis of the lysosomal storage phenomenon has provided a mechanism which can account for the low and narrow range of residual enzyme activity observed in patients with lysosomal storage diseases, and also for the clinical normality of the arylsulphatase A pseudodeficiency homozygote. Conzelmann & Sandhoff (1983/84), developed a model which predicted that low residual activities, down to a critical threshold level, may be sufficient to metabolize substrate and avert lysosomal storage. Below the threshold value, lysosomal storage will occur such that small changes in residual activity may lead to large differences in the rate of substrate storage, and hence clinical severity. Subsequent work by Leinekugel *et al.* (1992) provided experimental support for the model. In studies of cultured fibroblasts from patients with metachromatic leukodystrophy, Tay-Sachs disease or Sandhoff disease, they identified a threshold of 10-15% of normal enzyme activity, above which substrate turnover was essentially normal and lysosomal storage of substrate did not occur. Below the threshold activity, substrate turnover rapidly diminished and lysosomal storage occurred. A reasonable correlation was observed between the residual enzyme activity and the clinical severity for Tay-Sachs and Sandhoff diseases.

In MPS-VI, the critical threshold of activity may be about 5% of normal, based on a MPS-VI patient who was essentially asymptomatic (Brooks *et al.*, 1991b). However, attempts to correlate the residual 4S enzyme activity below this threshold with the clinical severity in MPS-VI are hampered by the low and narrow range of activity observed, and the limitations in sensitivity, specificity, and reproducibility of the assays. In addition, the assay of residual enzyme activity in cultured cells, for example fibroblasts, may not be an accurate indicator of residual

enzyme activity in the clinically significant tissues, as small changes in activity may result in large fluctuations in the rate of storage and hence clinical severity (Conzelmann & Sandhoff, 1983/84; Leinekugel *et al.*, 1992).

A complementary approach to the determination of residual activity is the quantification of the residual 4S protein, using a polyclonal antibody to capture 4S and a panel of 4S-specific monoclonal antibodies. 4S immunoquantification was performed on cultured fibroblasts from 16 MPS-VI patients using a panel of 7 monoclonal antibodies (Brooks *et al.*, 1991b). 4S protein varied from not detectable up to 5% of normal controls, and the degree of reduction from normal showed some correlation to the severity of clinical presentation, but was not sufficiently refined to allow prediction of prognosis. Knowledge of the residual protein and enzyme activity enables the determination of mutant enzyme kinetics, which can provide insights into the nature of the enzymatic defect. However, the low levels of lysosomal enzymes generally, compounded by the fact that they are further reduced in mutants, makes this approach difficult (Brooks *et al.*, 1991b).

A complete biochemical phenotype profile must also take into account the processing of the enzyme and targeting to the lysosome, in addition to residual protein and activity. Taylor *et al.* (1990), investigated the biosynthesis and maturation of 4S in MPS-VI fibroblasts using monoclonal antibodies. They found that all eight patient cell lines synthesized protein at a rate less than 10%, and usually around 3%, of normal; the reduction in 4S protein roughly corresponded with the reduction in 4S activity. The maturation of the residual 4S showed features consistent with delayed intracellular transport and processing, and/or decreased intracellular stability, and/or failure of lysosomal targeting.

In summary, attempts to demonstrate a correlation between the clinical and biochemical phenotypes, such that clinical phenotypes can be clearly differentiated

by biochemical means alone, have not been entirely successful. Identification of the molecular defects in the 4S gene of the patients may enable the determination of a genotype-phenotype correlation.

1.10.4 MOLECULAR GENOTYPE

For a growing number of genetic diseases, the advent of molecular biological techniques has enabled the determination of the gene defects. It is becoming apparent that for many diseases, a large number of mutant alleles exist in the patient population, although a few alleles may predominate. The phenomena of genetic heterogeneity and low incidence has frustrated attempts to perform genetic screening of the mass population for the majority of genetic diseases. For example, in Gaucher disease, over 35 different mutations have been documented in the glucocerebrosidase gene (Beutler, 1993), and 29 point mutations or small deletions have been identified in the X-linked iduronate-2-sulphatase gene (Hopwood *et al.*, 1993a). However, the incidence of a genetic disease in the general population, and the frequency of an individual mutation in the patient population, can vary considerably in different communities. An example of the latter is found in Gaucher disease, in which the mutant allele N370S, an amino acid substitution in the glucocerebrosidase gene, accounts for about 75% of mutations in Jewish patients, while it accounts for only about 33% in European non-Jewish patients, and has not been observed in patients of African or Asian origin (Beutler, 1993). In contrast, the L444P substitution mutation (Tsuji *et al.*, 1987) accounts for approximately 2% and 40% of alleles in Jewish and non-Jewish Gaucher patients respectively (Beutler, 1993).

Genetic screening of the general population may be feasible for diseases which have a high incidence, and in which a few alleles predominate. An example is medium-chain acyl-CoA dehydrogenase deficiency (MCAD), an autosomal recessive disorder of fatty acid oxidation (Yokota *et al.*, 1991). The incidence in the population is not

known with precision, but disorders of fatty acid oxidation as a whole occur at the relatively high frequency of about 1 in 5,000 live births. Mutational analysis of MCAD in 55 patients with a European ethnic background found that 80% of the patients were homozygous and 18% were heterozygous for the same mutant allele A985G, which therefore had an allele frequency of about 90% in the patient population (Yokota *et al.*, 1991).

Once a base change has been identified in the appropriate gene, the putative mutation should be verified. This can be achieved by screening both normal and patient populations for the candidate mutation, and expression and analysis of the candidate mutation in a tissue culture system. Mutation analysis is often complicated by compound heterozygosity, where two different mutant alleles are present within the patient. Both mutant alleles need to be taken into account in an attempted genotype-phenotype correlation.

In a number of diseases, certain alleles or combinations of alleles have been shown to correlate well with the clinical disease pattern, and therefore clinical prognosis is possible. For example, in a recent clinical and genetic analysis of 161 Gaucher disease type I patients, a statistically significant correlation was found between the clinical features and the genotype in patients who were homozygous for the N370S genotype (Sibille *et al.*, 1993). Similarly for Hurler syndrome, which is due to a deficiency of α-L-iduronidase, patients homozygous for the commonly occurring nonsense mutation W402X were uniformly at the extremely severe end of the clinical spectrum (Scott *et al.*, 1992). Although both examples were very useful correlations, they were restricted to a single but frequently occurring mutant allele in each disorder. For the majority of mutant alleles in many genetic disorders, the frequency in patients will be too low to test for a statistically significant genotype-clinical phenotype correlation. Importantly, a correlation does not provide an explanation of the severity of the clinical phenotype.

In addition to verification, the expression of the mutations in tissue culture and *in vitro* analysis of the mutant protein produced can provide insights into the enzyme defect. This approach has the advantage that for two mutant alleles, the effect of each mutation on enzyme structure and function can be independently determined. Furthermore, the expressed protein is generally produced at higher levels than in the patient cells, which facilitates characterization. Expression in tissue culture and subsequent characterization *in vitro* enables an additional biochemical phenotype to be defined, i.e. the predicted residual activity (Trefz *et al.*, 1993) or predicted residual protein, relative to the non-mutant control. These parameters can aid the formation of a genotype-clinical phenotype correlation. For example, Trefz *et al.* (1993) demonstrated a correlation between the predicted residual activity of 6 mutations in the phenylalanine hydroxylase gene, and the degree of brain damage in patients affected with the autosomal recessive disorder, phenylketonuria.

In some cases, genotype-phenotype correlations are not always consistent, for example, Gaucher disease (reviewed by Beutler, 1993). Patients homozygous for the L444P mutation are generally afflicted with the clinically severe forms of Gaucher; the acute neuropathic (infantile onset) or the chronic neuropathic (juvenile onset) form (Tsuji *et al.*, 1987). However, recent work on Japanese patients homozygous for the L444P allele indicated that this allele was associated with a wider spectrum in clinical severity, as it occurred in all three clinical subtypes of Japanese Gaucher disease patients (Kawame *et al.*, 1993), including the milder non-neuropathic form which does not affect the central nervous system. Therefore, other factors beside the presence of a particular glucocerebrosidase mutation are modulating the clinical severity, and can include other currently undefined genetic, environmental, or developmental factors.

1.10.5 GENOTYPE-PHENOTYPE CORRELATION IN MPS-VI

During the course of this Ph.D., the cloning of the 4S gene by other groups has enabled the identification of 4S gene mutations in MPS-VI patients (Table 1.1). The 4S mutations published by other investigators will now be discussed, and where applicable, the genotype-phenotype correlation.

Wicker *et al.* (1991) were the first to report the identification of a 4S mutation, in an MPS-VI patient with an intermediate form of the disease. The patient was homozygous for a G137V amino acid substitution in a region of the 4S polypeptide that is conserved between sulphatases. The G137V substitution was confirmed as a mutation by synthesis *in vitro* and expression in tissue culture. Immunochemical analytical techniques were used to characterize the residual activity, kinetic parameters and protein stability of the mutant protein derived from cultured cells of the patient, and also from LTK$^-$ tissue culture cells transfected with the mutant 4S construct. In addition, the protein processing of the mutant 4S expressed in LTK$^-$ cells was also investigated. The 4S G137V mutant had normal kinetic parameters towards a trisaccharide substrate, but the stability was severely reduced, such that the majority of the precursor enzyme was proteolytically degraded, presumably in the ER or Golgi. The G137V mutation was considered responsible for the intermediate MPS-VI phenotype in the patient, as the behaviour of the 4S mutant isolated from patient cultured cells corresponded with the behaviour of the *in vitro* synthesized G137V mutant expressed in LTK$^-$ cells. Therefore a satisfactory genotype-phenotype correlation was obtained in this patient. However as the mutation was not found in another 14 MPS-VI patients, the reproducibility of the correlation for this mutant allele is not known.

Arlt *et al.* (1994) identified the lesions in the 4S cDNA of a MPS-VI patient with an intermediate form of the disease. The patient was a compound heterozygote

TABLE 1.1 4S MUTATIONS IDENTIFIED IN MPS-VI BY OTHER GROUPS

Patient	Clinical Severity	Allele 1 base change	Allele 1 codon change	Allele 2 base change	Allele 2 codon change	Reference
P.P.	interm.	ΔG_{237}-C_{243}		$C_{454}T$	R152W	Voskoboeva *et al.* (1994)
MS-3	severe	$T_{349}C$	C117R	$T_{349}C$	C117R	Jin *et al.* (1992)
B.K.	interm.	$G_{410}T$	G137V	$G_{410}T$	G137V	Wicker *et al.* (1991)
U.S.	severe	$G_{430}A$	G144R	$G_{430}A$	G144R	Isbrandt *et al.* (1994)
M.T.	interm.	$C_{478}T$	R160X	$G_{479}A$	R160Q	Voskoboeva *et al.* (1994)
A.B.	mild	$T_{574}C$	C192R	$T_{574}C$	C192R	Isbrandt *et al.* (1994)
MS-4	mild	$T_{707}C$	L236P	$G_{1214}A$	C405Y	Jin *et al.* (1992)
B.C.	interm.	Δexon V		$T_{1600}C$	X534Q	Arlt *et al.* (1994)
A.O.	interm.	$T_{962}C$	L321P	$T_{962}C$	L321P	Isbrandt *et al.* (1994)
A.M.	severe	$InsT_{1285}$		$InsT_{1285}$		Isbrandt *et al.* (1994)
M.Ak.	severe	$G_{1562}A$	C521Y	$G_{1562}A$	C521Y	Isbrandt *et al.* (1994)
L.E.	severe	ΔC_{1577}		ΔC_{1577}		Isbrandt *et al.* (1994)

A summary of all the 4S mutations in MPS-VI patients identified by other groups. For each patient, the clinical phenotype, the nucleotide change and the deduced codon change (if applicable) are shown. Intermediate clinical severity is denoted `interm.', a nucleotide insertion is denoted `Ins', and a deletion is denoted `Δ'.

Δexon V and the X534Q alleles. The Δexon V allele was a presumed severe mutation as it was predicted to generate a frameshifted, truncated protein that was devoid of exon V. The X534Q allele was a mutation in the termination codon such that an additional 50 amino acid residues were added at the carboxyl-terminal of the protein. The X534Q mutation was expressed in LTK^- cells, and the biochemical phenotype was determined by analysis of residual 4S activity, processing, and conformation. The X534Q mutant protein exhibited an altered conformation and reduced stability (more than 2,000-fold compared to wild-type expressing cells), the latter due to degradation, presumably in the ER. However, the X534Q mutant protein also exhibited a 9-fold increase in catalytic efficiency. The results obtained from *in vitro* analysis of the mutant protein are in keeping with the intermediate clinical severity. Again, as the mutation was not found in a collection of 27 other MPS-VI patients, the reproducibility of the genotype-phenotype correlation for this mutant allele is not known.

Jin *et al.* (1992) identified three mutations in the 4S cDNA of two MPS-VI patients. A severe patient was homozygous for the C117R mutation, and a mild patient was a compound heterozygote of the L236P and C405Y alleles. The severity of each of the three mutant alleles was predicted based on the biochemical and clinical severity of the patients. However, the severity predictions were not confirmed by expression of the mutations separately in tissue culture.

Recently, Voskoboeva *et al.* (1994) identified four mutations in the 4S gene of two patients with intermediate clinical severity. Both patients were heteroallelic; the first patient for the amino acid substitution R160Q and the truncation mutation R160X, the second patient for the amino acid substitution R152W and the truncation mutation ΔG_{237}-C_{243}. As for *Jin et al.* (1992), the mutations were not characterized by expression studies.

Isbrandt *et al.* (1994) identified six mutant alleles in six MPS-VI patients; each patient was homozygous for a unique mutant allele. Four of the mutations were amino acid substitutions; G144R, C192R, L321P, and C521Y, which were associated with severe, mild, intermediate, and severe clinical phenotypes respectively. One of the mutations was a single base insertion, $insT_{1285}$, which resulted in a 39 amino acid carboxyl-terminal extension. The remaining mutation was a single nucleotide deletion, ΔC_{1577}, which resulted in the loss of 130 amino acids from the carboxyl-terminal of the 4S protein. The insertion and deletion mutations were both associated with severe clinical phenotypes. The four amino acid substitutions and the insertion resulted in the severe reduction of 4S protein and arylsulphatase enzyme activity in comparison to a wild-type control.

The mutations identified by our group in a collection of MPS-VI patients, and their subsequent expression in CHO cells and characterization *in vitro*, will be presented and discussed in Chapters 4 and 5.

1.11 AIMS OF THE PROJECT

The initial aim of the work presented in this thesis was the isolation and characterization of the gene for 4S, in particular a full-length 4S cDNA clone. The provision of a 4S cDNA clone by our collaborators Christoph Peters and Kurt von Figura, enabled the subsequent aims to be pursued.

1. To identify mutations in the 4S gene of MPS-VI patients.

2. To characterize each 4S mutation via expression in tissue culture and analysis of residual activity and protein.

3. To test the hypothesis that the mutant genotype could be correlated with the clinical and biochemical phenotype in MPS-VI patients.

4. To gain insight into residues critical for 4S structure and function, through characterization of MPS-VI mutations.

The successful achievement of these aims may have the following medical significance:

A. The identification of MPS-VI patient mutations could contribute to the diagnosis of MPS-VI, especially carrier detection and prenatal diagnosis in affected families.

B. The successful correlation between genotype and phenotype in MPS-VI may contribute to the prediction of the long-term clinical phenotype (prognosis) for MPS-VI patients.

C. Accurate prognosis would be an especially significant advance in the assessment of the therapeutic efficacy of current and future experimental treatment protocols for MPS-VI.

CHAPTER 2

MATERIALS & METHODS

MATERIALS & METHODS

2.1 MATERIALS

2.1.1 ELECTROPHORESIS

The following reagents were obtained from:

Acrylamide	Bio-Rad, Hercules, 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	BDH Chemicals Ltd., Poole, Dorset, England
Ethidium bromide	Boehringer Mannheim, Mannheim, Germany
Formamide	Ajax, Auburn, NSW, Australia
NuSieve GTG Agarose LMP	FMC Bioproducts, Rockland, Maine, USA
Sequagel sequencing system	National Diagnostics, Manville, New Jersey, USA
SPP-1 *Eco*RI, pUC19 *Hpa*II, λ *Hin*dIII size markers	Bresatec, Adelaide, Australia
N,N,N',N'-tetramethyl--ethylenediamine (TEMED)	Bio-Rad
Urea	Ajax
Xylene cyanol FF	Tokyo Kasei, Tokyo, Japan

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, England
Lysozyme, chicken	Sigma Chemical Co.

DNase I	Boehringer Mannheim
Proteinase K	Sigma Chemical Co.
Reverse transcriptase (Mo-MLV)	Bethesda Research Laboratories, Gaithersburg, Maryland, USA
Pancreatic Ribonuclease A (DNase-free)	5 Prime -> 3 Prime, Boulder, Colorado, USA
RNAsin	Promega, Madison, Wisconsin, 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, 3,000 Ci/mmole	Amersham
γ-^{32}P-ATP, 5,000 Ci/mmole	Amersham
γ-^{32}P-ATP, 5,000 Ci/mmole	Bresatec

2.1.4 BUFFERS & SOLUTIONS

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

Formamide loading buffer	95% (v/v) formamide, 20 mM EDTA, pH 8.0, 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, pH 8.0, 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.
10 X M9 salts	6% (w/v) Na_2HPO_4, 3% (w/v) KH_2PO_4, 1% (w/v) NH_4Cl, 0.5% (w/v) NaCl.
Phosphate buffered saline (PBS)	137 mM NaCl, 2.7 mM KCl, 4.3 mM $Na_2HPO_4.7H_2O$, 1.4 mM KH_2PO_4, pH 7.2.

10 X PCR buffer	500 mM KCl, 100 mM Tris-HCl, pH 8.4, 25 mM $MgCl_2$, 0.01% (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 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, 0.5% (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.
2 X YT	1.6% (w/v) Bacto tryptone, 1% (w/v) Bacto yeast extract, 0.5% (w/v) NaCl.

Bacto tryptone, Bacto yeast extract and Bacto agar were obtained from Difco, Detroit, Michigan, USA.

2.1.5b SOLID MEDIA

L-Agar	L-broth, 1.5% (w/v) Bacto agar
L-Agarose	L-broth, 1.5% (w/v) agarose
L-Amp	L-broth, 1.5% (w/v) Bacto agar, ampicillin (100 μg/ml)
Minimal medium	Minimal medium, 1.5% (w/v) Bacto agar

2.1.6 ANTIBIOTICS

Ampicillin	Sigma Chemical Co.
G418 sulphate (Neomycin) Geneticin	Gibco laboratories, Grand Island, New York, USA
Kanamycin sulphate	Boehringer Mannheim
Penicillin G	Sigma Chemical Co.
Streptomycin sulphate	Boehringer Mannheim

2.1.7 BACTERIAL STRAINS

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

Strain	Genotype and reference
C600	*supE44, hsdR* (r_k^-, m_k^+), *thr-1, thi-1, leuB6, lacY1, tonA21* (Young & Davis, 1983).
DH5αF'	*supE44, hsdR17* (r_k^-, m_k^+), *thi-1, recA1, endA1, gyrA* (Nal^r), *relA1,* Δ(*lacZYA-argF*)*U169, deoR,* F', (*Φ80lacZ*ΔM15) (Raleigh *et al.*, 1993).
ED8799	*hsdS* (r_k^-, m_k^-), *metB7, SupE,* (*glnV*)*44, supF*(*tyr T*)*58,* Δ(*lacZ*)M15.
JM101	*supE, thi* Δ(*lac-proAB*), [F',*traD36, proAB*$^+$, *lacI*^{q}, *lacZ*ΔM15] (Yanisch-Perron *et al.*, 1985).
LE392	*supE44, supF58, hsdR514* (r_k^-, m_k^+), *galK2, galT22, metB1, trpR55, lacY1* (Sambrook *et al.*, 1989).
NM538	*supF, trpR, lacY, hsdR* (r_k^-, m_k^+) (Frischauf *et al.*, 1983).
TG1	*supE, hsdΔ5* (r_k^-, m_k^-, McrB$^-$), *thi,* Δ(*lac-proAB*), [F', *traD36, proAB*$^+$, *lacI*^{q}, *lacZ*ΔM15] (Sambrook *et al.*, 1989).
Y1090	*supF, araD139,* Δ*lon,* Δ*lacU169, rpsL*(str^r), *mcrA, trpC22*::Tn10(tet^r) pMC9 (Huynh *et al.*, 1984)

2.1.8 VECTORS

The vectors used in this study are described below:

Bacteriophage vectors:

Vector	Genotype and reference
λgt10	*λb527, srlλ3°, imm^{434}, srlλ4°, srlλ5°* (Huynh *et al.*, 1984).
λgt11	*λlac5, srlλ3°, cI857, srlλ4°, nin5, srlλ5°, Sam100* (Young & Davis, 1983).
λMBL3A	λ (*A*am32 *B*am1) *sbhlλ1°, b189,* < polylinker, (*Sal*I-*Eco*RI), *int29, ninL44, cI857, trpE,* polylinker (*Eco*RI-*Sal*I) > *KH54, chiC, srlλ4°, nin5, srlλ5°* (Frischauf *et al.*, 1983).

Filamentous phage vectors:

The replicative forms of M13 were purchased from Boehringer Mannheim.

M13mp18	Yanisch-Perron *et al.* (1985)
M13mp19	Yanisch-Perron *et al.* (1985)

Plasmid vectors:

pUC18	Bresatec, Yanisch-Perron *et al.* (1985)
pUC19	Bresatec, Yanisch-Perron *et al.* (1985)
pSP64 poly A	Promega, Melton *et al.* (1984)

2.1.9 RECOMBINANT DNA LIBRARIES

All libraries screened were of human origin and were either purchased from Clontech (Palo Alto, California, USA), or where generously provided by Iain Young or Geoff Howlett.

GENOMIC LIBRARIES

Tissue source	Vector	Source
Leukocyte genomic DNA	λEMBL3A	Iain Young, Australian National University, Canberra, Australia

cDNA LIBRARIES

Tissue source	Vector	Source
Colon cDNA random-primed	λgt10	Promega (Cat. No. HL1034a)
Liver cDNA	λgt10	Geoff Howlett, University of Melbourne, Victoria, Australia
Lung fibroblast cDNA	λgt11	Promega (Cat. No. HL1011)
Placental cDNA	λgt11	Promega (Cat. No. HL1008)
Testis cDNA	λgt11	Promega (Cat. No. HL1010b)

All cDNA libraries were synthesized by oligo-dT priming unless otherwise stated.

2.1.10 TISSUE CULTURE SOLUTIONS

F-12 nutrient media (Ham's)	Gibco
Foetal calf serum	Commonwealth Serum Laboratories, Melbourne, Vic., Australia, or Gibco-BRL
Phosphate buffered saline	Commonwealth Serum Laboratories
Trypsin-Versene solution	Commonwealth Serum Laboratories

2.1.11 MISCELLANEOUS MATERIALS

Centricon-100 microconcentrators	Amicon, Beverly, Massachusetts, USA
Colony/PlaqueScreen™ filters	Dupont-NEN Research Products, Boston, Massachusetts, USA
GF/A glass filter discs	Whatman Ltd., Poole, England
NA-45 DEAE-cellulose paper	Schleicher & Schüll, Dassel, Germany
Nylon filters (GeneScreen*Plus*™)	Dupont-NEN Research Products,
Oligo (dT)-cellulose	Pharmacia P-L Biochemicals
Polyvinyl ELISA plates	Costar, Cambridge, Massachusetts, USA
Positive Land film, type 677	Polaroid Corp., Cambridge, Massachusetts, USA
Sephadex G-50	Pharmacia P-L Biochemicals

Tissue culture flasks and dishes (disposable, polystyrene)	Corning Glass Works, Corning, New York, USA
X-ray film	Fuji Rx X-ray film, Fuji Photo film Co. Ltd.

2.1.12 MISCELLANEOUS FINE CHEMICALS

4-methylumbelliferyl sulphate (4MUS), potassium salt	Sigma Chemical Co.
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	Boehringer Mannheim
Dideoxy DNA sequencing kits	Bresatec or Boehringer Mannheim
Diethylamine	Sigma Chemical Co.
Diethlypyrocarbonate	Sigma Chemical Co.
Dimethylsulphoxide (DMSO)	Sigma Chemical Co.
GENECLEAN™ kit	BIO 101 Inc., La Jolla, California, USA
Guanidinium Isothiocyanate	Bethesda Research Laboratories, Gaithersburg, Maryland, USA
Herring Sperm DNA	Sigma Chemical Co.
Hydroxylamine	Sigma Chemical Co.
Isopropyl-β-D-thiogalactopyranoside (IPTG)	Boehringer Mannheim
Multiprime DNA labelling kit	Amersham
Oligonucleotide-directed *in vitro* mutagenesis kit	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.
Tetramethylammonium chloride (TMACl)	Aldrich

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

2.1.13 SYNTHESIS OF SYNTHETIC DNA OLIGONUCLEOTIDES

The majority of synthetic oligonucleotides used in this study were prepared by Annette Orsborn of the Department of Chemical Pathology at the Adelaide Children's Hospital.

Oligodeoxynucleotides designated ASB-1, 4SS1, 4SS2 and 4SS3 were synthesized by the Department of Haematology, Flinders Medical Centre, Adelaide. Oligodeoxynucleotide 4SS4 was synthesized by Bresatec. Oligodeoxynucleotides designated 4SP1 to 4SP8 were synthesized at the Department of Human Immunology, Institute of Medical and Veterinary Science, Adelaide. All other oligodeoxynucleotides were synthesized using an Applied Biosystems 391 DNA synthesizer in the Department of Chemical Pathology at the Adelaide Children's Hospital.

2.1.14 PREPARATION OF GLASSWARE & 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

Plasmid DNA minipreps were prepared by a modified procedure of Birnboim & Doly (1979). Briefly, single colonies were grown overnight with aeration at 37°C in 2 ml of 2 X YT media with the appropriate antibiotic (eg. ampicillin; 100 μg/ml). 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 on ice 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 10 min followed by addition of 125 μl of an ice-cold solution of 3 M sodium acetate, pH 5.2. The solution was mixed by inversion and incubated on ice for a further 10 min and cellular debris and chromosomal DNA was pelleted by centrifugation at 12,000 g for 5 min at 4°C. The supernatant was transferred to a fresh tube, 1 μl of 10 mg/ml RNase A was added, and the reaction was incubated at 37°C for 60 min. The digest was phenol/chloroform extracted, ethanol precipitated, washed with 70% (v/v) ethanol, vacuum-dried and resuspended in 40 μl of water or TE buffer.

2.2.1b MINIPREPS - M13 REPLICATIVE FORM (RF)

Single plaques were toothpicked into 2 ml of 2 X YT containing a 100-fold dilution of an overnight culture of the host strain, and were grown overnight as for plasmid DNA minipreps. M13 RF DNA minipreps were performed essentially as described for plasmid minipreps in Section 2.2.1a.

2.2.1c LARGE SCALE PLASMID PREPARATION

Recombinant clones, as selected by restriction analysis or dot-blot hybridization of miniprep plasmid DNA, were grown overnight in 100 ml of 2 X YT media supplemented with the appropriate antibiotic. The cells were pelleted by centrifugation at 8,000 g for 10 min at 4°C, and resuspended in 1 ml of TES buffer. Lysozyme was added to a final concentration of 2 mg/ml and the resuspended cells were incubated on ice for 20 min. A solution of ice-cold 0.2 M NaOH, 1% (w/v) SDS was freshly prepared and 2 ml was added to the cell suspension and carefully mixed by inversion. The mixture was incubated on ice for 10 min. Following addition of 1.25 ml 3 M sodium acetate, pH 5.2, the solution was again carefully mixed by inversion and incubated on ice for 10 min. The solution was centrifuged at 15,000 g for 15 min and the supernatant containing supercoiled plasmid DNA was carefully transferred to a fresh tube. RNase A (10 µl of 10 mg/ml) was added to the supernatant and incubated for 60 min at 37°C. The solution was extracted with 1 volume of phenol/chloroform (1:1) saturated with water, vortexed for 1 min, centrifuged at 8,000 g for 10 min and the aqueous phase transferred to a fresh tube. The phenol/chloroform extraction was repeated and the resultant aqueous phase was extracted with 1 volume of chloroform/isoamyl alcohol (24:1); this mixture 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 about 60 min. Centrifugation at 12,000 g for 20 min was performed after which the tube was carefully drained, the DNA pellet rinsed with ethanol and dried under vacuum. The DNA was then resuspended in 100-200 µl of water or TE.

DNA used for cell culture expression studies was further purified by a modification of the procedure of Heilig *et al.* (1993). Specifically, the DNA pellet was resuspended in 2 ml of water, and then 500 µl of 5 M NaCl and 2.5 ml of 13% (w/v) PEG 6000 were added. The solution was mixed and incubated on ice for at least 60 min (more usually overnight). 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 750 µl TE buffer in Eppendorf tubes. Caesium chloride (1.32 g) was dissolved in each DNA solution and 125 µl of a 10 mg/ml ethidium bromide solution was added. The tubes were centrifuged at 12,000 g for 1 min, and the supernatants were gently underlaid beneath 1.9 ml of a 65% (w/v) CsCl solution in polyallomer Quick-Seal™ ultracentrifuge tubes (Beckman, Palo Alto, California, USA). The tubes were centrifuged at 424,000 g for 2.5 hr at room temperature, and the purified DNA bands were collected by syringe and transferred to Eppendorf tubes. An equal volume of TE was added, and the ethidium bromide removed by sequential extractions with water-saturated butanol. The DNA was then precipitated by the addition of 0.6 volumes of isopropanol and left at room temperature for 15 min. The tubes were centrifuged at 12,000 g for 15 min at 4°C; the supernatant was then discarded and the pellet was rinsed with ethanol, drained and dried under vacuum. The final DNA pellet was dissolved in water or TE buffer.

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 5.2 or 3.75 M ammonium acetate, pH 7.5, at -20°C for approximately 1 hr. Following centrifugation, pellets were rinsed with 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. The digests were analysed by agarose gel electrophoresis to ensure complete digestion. Vectors digested with a single restriction endonuclease were 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. Subsequent to the digestion & phosphatase treatment, 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 adjusted to 10-20 ng/µl in water or TE.

2.2.5 LIGATION OF PLASMID VECTORS

The DNA insert and appropriate plasmid vector (20 ng) were usually combined in a molar ratio of 3:1 in a 20 µl reaction volume in a final concentration of 1 x ligation buffer (King & Blakesley, 1986). For ligations of DNA with overhanging ends, 0.1-0.5 units of T4 DNA ligase were used, and for blunt-end ligations, 1.0 unit of enzyme was used. Ligation reactions were incubated at room temperature, generally for 4 hr.

2.2.6 TRANSFORMATION OF *E. coli*.

Strains of *E. coli* were made competent for use in transformations as follows. An aliquot (0.5 ml) of an overnight culture of an *E. coli* strain grown from a single colony was inoculated into a 50 ml culture of 2 X YT. This culture was incubated at 37°C with shaking until an OD_{600nm} of 0.3-0.6 was reached. The cells were pelleted by centrifugation at 3,000 g for 5 min, resuspended in 20 ml of an ice-cold solution of 50 mM $CaCl_2$ and incubated on ice for 30 min. Following centrifugation at 3,000 g for 2 min, the cells were resuspended in 4 ml of 50 mM $CaCl_2$ and left on

ice for at least 2 hr. Aliquots of the DNA ligation reaction (typically 2 and 8 µl) were individually added to 300 µl of the competent cells and incubated on ice for about one hr. A heat-shock reaction was then performed at 42°C for from 45 s to 3 min.

For plasmid vectors, 0.5 ml of 2 X YT was added to the heat-shocked cells and incubated at 37°C with shaking for 30 min. The culture was then plated onto L-plates containing ampicillin at 100 µg/ml. For M13-derived vectors, the heat-shocked cells were returned to ice for 5 min, and then mixed with 3 ml of 0.7% (w/v) L-agar and plated out on L-plates. 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 0.7 to 2.0% (w/v) in TAE buffer. Samples were prepared in loading buffer and were electrophoresed at 5-10 V/cm 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 19:1 (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 vertical gel system at 80 V.

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 formamide 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 DNA was passively eluted from the gel slice by immersion 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 5.2, and re-precipitated with ethanol. 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 from the supernatant using 0.1 volume of 3 M sodium acetate and 2.5 volumes of 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 as described below. The appropriate DNA fragment was excised, placed in 2.5-3 volumes of a NaI solution and incubated at 55°C until the agarose had dissolved. Glassmilk™ solution was added (typically 5 µl for DNA quantities less than 5 µg) to the agarose solution and the tube was left at room temperature for 5 min. Following centrifugation, the pellet was washed three times with 50 volumes of NEW wash (NaCl, EDTA, water) 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 once 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 & Chory (1987), and was performed as described below. After separation of the DNA fragments by agarose gel electrophoresis, the appropriate band was identified, a slit was made in the gel to the side of the fragment, and a piece of NA-45 DEAE paper inserted into the slit. The DNA was electrophoretically migrated onto the paper by rotating the gel 90° from its original migration direction. Subsequently, the paper was 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 water and eight volumes of ethanol were 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 each. 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. 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). Later, oligonucleotide primers designated 4SP22 and over were purified using the method of Sawadogo & 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-RADIOISOTOPE 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-ATP as described by Chaconas & van de Sandé (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 & Vogelstein, 1983) using the Amersham Multiprime DNA labelling 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 using 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. The fractions corresponding to the peak were then pooled.

2.2.12 TRANSFER OF DNA TO NYLON MEMBRANES

2.2.12a PLAQUE LIFTING

Plaques were transferred onto Colony/PlaqueScreen™ filters following the manufacturer's instructions. Briefly, the plates were placed at 4°C for about 30 min. 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 DNA was then denatured by two successive placements of the filters, plaque side up, on pools of 0.4 M NaOH for 2 min each. The filters were removed to blotting paper and then neutralized by two successive placements of the filters, plaque side up, on pools of 1.0 M Tris-HCl, pH 7.5, for 2 min each. Blotted filters were air-dried, and 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) nylon membrane using the alkaline transfer method (Reed & 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

transfer solution (0.6 M NaCl, 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 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 overnight. Following transfer, the filter was neutralized in a solution of 0.5 M Tris-HCl, pH 7.5, 1 M NaCl, air-dried and then prehybridized and hybridized as described in Section 2.2.13.

2.2.13 PREHYBRIDIZATION, HYBRIDIZATION & 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) sodium 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. Library screening filters were generally washed using a tetramethylammonium chloride (TMACl) solution (Wood *et al.*, 1985), as the oligonucleotides used for screening were typically a mixed population differing in base content. Specifically, the filters were rinsed twice in 6 X SSC at room temperature for 5 min. The filters were drained, and then transferred to TMACl-wash-solution (3 M TMACl, 50 mM Tris-HCl, pH 8.0, 2 mM EDTA, 0.1% (w/v) SDS,

0.5% (w/v) sodium pyrophosphate) and gently agitated at room temperature for 5 min. The filters were then transferred to fresh TMACl-wash-solution at a temperature 5-15°C lower than the calculated T_d (Wood *et al.*, 1985), and washed for 10 min. Autoradiography was carried out at room temperature or, for detection of low levels of radioactivity, at -80°C in the presence of calcium tungstate intensifying screens. ASO-probed filters were rinsed twice in 5 X SSC for 10 min at room temperature. The filters were then transferred to wash solution (2 X SSC, 0.1% (w/v) SDS) which was gradually heated from 42°C to 70°C. Filters were removed at the temperatures indicated in Table 4.4. Autoradiography was performed as for the TMACl-washed filters.

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 and the hybridization allowed to incubate overnight at 42°C. Filters were washed in 2 X SSC at room temperature for 5 min with constant agitation, then in 2 X SSC, 1% (w/v) SDS for 30 min with a gradual increase in temperature from room temperature to 65°C, and then the filters were washed in fresh 2 X SSC, 1% (w/v) SDS for 5 min at 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 on 140 mm L-agar plates. 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: LE392 for the EMBL3A genomic library, C600 for the λgt10 cDNA libraries and Y1090 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 (Section 2.2.11). Colony/PlaqueScreen™ 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 and transferred into 1 ml of SM using the large end of a sterile pasteur pipette. 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 were 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 140 mm plate. The confluent plates were covered with 10 ml of 1 X SM solution and incubated at room temperature for at least 1 hr. The solution was then removed and centrifuged at

8,000 g for 15 min at 4°C to pellet the bacterial debris. The supernatants containing the lambda phage were transferred to fresh tubes, DNase I and RNase A were added to final concentrations of 1 μg/ml and 10 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 12,000 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 1 ml of SM buffer per tube and, following addition of 10 μl 0.5 M EDTA and 10 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 10 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 at 4°C, the DNA pellet was washed with 70% (v/v) ethanol, resuspended in TE buffer, ethanol precipitated with sodium acetate, vacuum-dried and resuspended in 100 μl of water.

2.2.16 GROWTH & HARVESTING OF THE (+) 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 and a 1 in 40 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 (ssDNA) for sequencing was prepared as described below.

2.2.16b SINGLE STRANDED DNA PREPARATION

Cells were removed from 1.5 ml of the ssDNA cultures by centrifugation at 12,000 g for 5 min. The centrifugation step was repeated and 1100 µl 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 200 µl of a solution of 50 mM Tris-HCl, pH 8.0, 1 mM EDTA, and extracted with 100 µl of phenol/chloroform (1:1). The mixture was vortexed and then centrifuged at 12,000 g for 5 min and the aqueous phase transferred to a fresh tube. The aqueous phase was re-extracted with phenol/chloroform as above. To the twice-extracted aqueous phase was added 20 µl 3 M sodium acetate, pH 5.2, and 500 µl of ethanol. The DNA was pelleted by centrifugation at 12,000 g for 15 min at 4°C; the supernatant was discarded and the pellet was washed with 1 ml of chilled (-20°C) 70% (v/v) ethanol. The pellet was dried under vacuum and dissolved in 30 µl of water or TE buffer.

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: 6 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 $MgCl_2$) and 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. Generally, 20 μCi (about 7 pmoles) of the radiolabelled nucleotide α-^{32}P-dCTP and 1 unit of the Klenow fragment of DNA polymerase were added to each annealing reaction. The mixture was divided into 4 equal aliquots and added to 4 tubes, each tube contained all four deoxynucleotides and one of each of the four dideoxynucleotides. The final dideoxynucleotide concentrations used in this study varied according to the template to be sequenced, which depended on the ratio of purines to pyrimidines present. Generally, the final concentration in the reaction mixes were 30 μM ddATP, 22 μM ddCTP, 60 μM ddGTP and 150 μM ddTTP. The polymerization reactions were performed at 42°C for 15 min, followed by a chase with all four deoxynucleotides for a further 5 min. 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. Compressed areas of GC-rich sequence were resolved using 7-deaza-dGTP in place of dGTP in the reaction mixes (Mizusawa *et al.*, 1986).

2.2.18 COMPUTER ANALYSIS OF DNA & PEPTIDE SEQUENCES

Nucleotide sequences obtained were screened against the GenBank nucleotide sequence data base (up to Release 80.0, December 1993). 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:

STADEN Programs:

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:

PEPDATA	used to translate a nucleotide sequence into all six peptide reading frames.
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.

Miscellaneous Programs:

CHOFAS	Peptide secondary structure prediction program, using the Chou-Fasman algorithm (Chou & Fasman, 1974; 1978).

CLUSTAL	A multiple sequence alignment program, (Higgins & Sharp, 1988; 1989).
CLUSTALV	An updated version of CLUSTAL (Higgins *et al.*, 1991).
MACAW	A multiple sequence alignment program, (Schuler *et al.*, 1991).
RODENT	For entering DNA sequence data. (Pharmacia).

2.2.19 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, 330 ng of each primer was used (generally 20-mers), and the final concentration of $MgCl_2$ was 2.5 mM unless otherwise indicated, in the buffer system either provided by the manufacturer, or made as described in 2.1.4. 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. In general, 40 cycles were performed, using a Perkin Elmer Cetus DNA thermal cycler. Appropriate positive and negative controls were included in each set of reactions.

CHAPTER 3

ISOLATION OF THE 4S GENE

3.1 INTRODUCTION

Routine biochemical diagnosis of MPS-VI is generally based on the assay of DS, which is excreted at elevated concentrations in the urine of MPS-VI patients. A more precise biochemical diagnosis is based on the measurement of residual 4S enzyme activity in leukocytes or cultured skin fibroblasts of the patient. In general, it is possible to crudely correlate the clinical phenotype of the patient with the biochemical data. However, despite considerable efforts to refine the characterization of the residual levels of 4S activity and protein in patients, the correlation is currently limited in that biochemical analysis can only distinguish between patients at either extreme of the spectrum of clinical severity (Taylor *et al.,* 1990; Brooks *et al.,* 1990; 1991a; 1991b). It was considered that the correlation may be refined by the analysis of the patient's mutant 4S genotype.

The general objective of this Ph.D. project was to test the hypothesis that in MPS-VI patients, the clinical and biochemical phenotype could be correlated with the genotype. The first technical objective was to isolate a 4S cDNA clone, so as to enable the identification and analysis of mutations in the 4S gene of MPS-VI patients.

The two approaches to clone the 4S cDNA that were considered feasible at the commencement of the project (1987-1988) were screening of cDNA libraries with oligonucleotide probes, and screening of cDNA expression libraries with anti-4S antibodies. It was anticipated that the search for a 4S cDNA clone could be a laborious process, necessitating the screening of a large number of clones. 4S is generally present in very low abundance relative to total cellular protein, and turnover is slow (Taylor *et al.,* 1990), so it was considered plausible that the abundance of 4S mRNA and hence the frequency of 4S cDNA in the library could be very low. The approach chosen was library screening with oligonucleotide probes, which were derived from peptide sequences of 4S. Monoclonal and polyclonal

antibodies specific for 4S were also available for use in antibody screening of λgt11 cDNA expression libraries (Young & Davis, 1983). However, the latter approach was considered less likely to be successful as only one in six clones on average would be in the correct orientation and reading frame for successful expression screening, thus exacerbating the anticipated very low frequency of 4S cDNA. Furthermore, not all clones would be full-length and the epitopes on the expressed 4S protein may not be folded correctly to enable detection by the antibody, or the epitope may not be represented in the coding sequence of the clone.

Initially, the oligonucleotide gene probes were designed based on the amino acid sequence obtained from peptide fragments generated by tryptic digestion of purified human liver 4S (Fig. 3.1). 4S has been purified to homogeneity from human liver by members of this laboratory (Gibson *et al.*, 1987), and the mature form was found to be composed of a 57-kDa species, which dissociated under reducing conditions to 43-kDa and 8-kDa subunits on SDS-PAGE. The gene probes were used to screen cDNA libraries derived from lung fibroblast and liver, however, extensive screening of cDNA libraries was unsuccessful. The difficulty experienced in screening for a 4S cDNA clone supported the suspicion that the 4S cDNA was present at a very low frequency in cDNA libraries. Consequently, a genomic library was screened, on the assumption that single-copy genes are present at about the same frequency, independent of the level of gene expression. The oligonucleotide probes used to screen the genomic library were based upon amino acid sequence obtained from direct sequencing of the amino-terminus of the 43-kDa species (Fig. 3.1).

This Chapter describes the isolation of two overlapping genomic clones containing the first exon of 4S within a common 1.3-kb *Hin*dIII fragment. The 1.3-kb *Hin*dIII fragment encoded a 38 amino acid signal peptide, 66 amino acids of the amino-terminal portion of the 43-kDa species of 4S, and 450-bp of 5'-flanking sequence

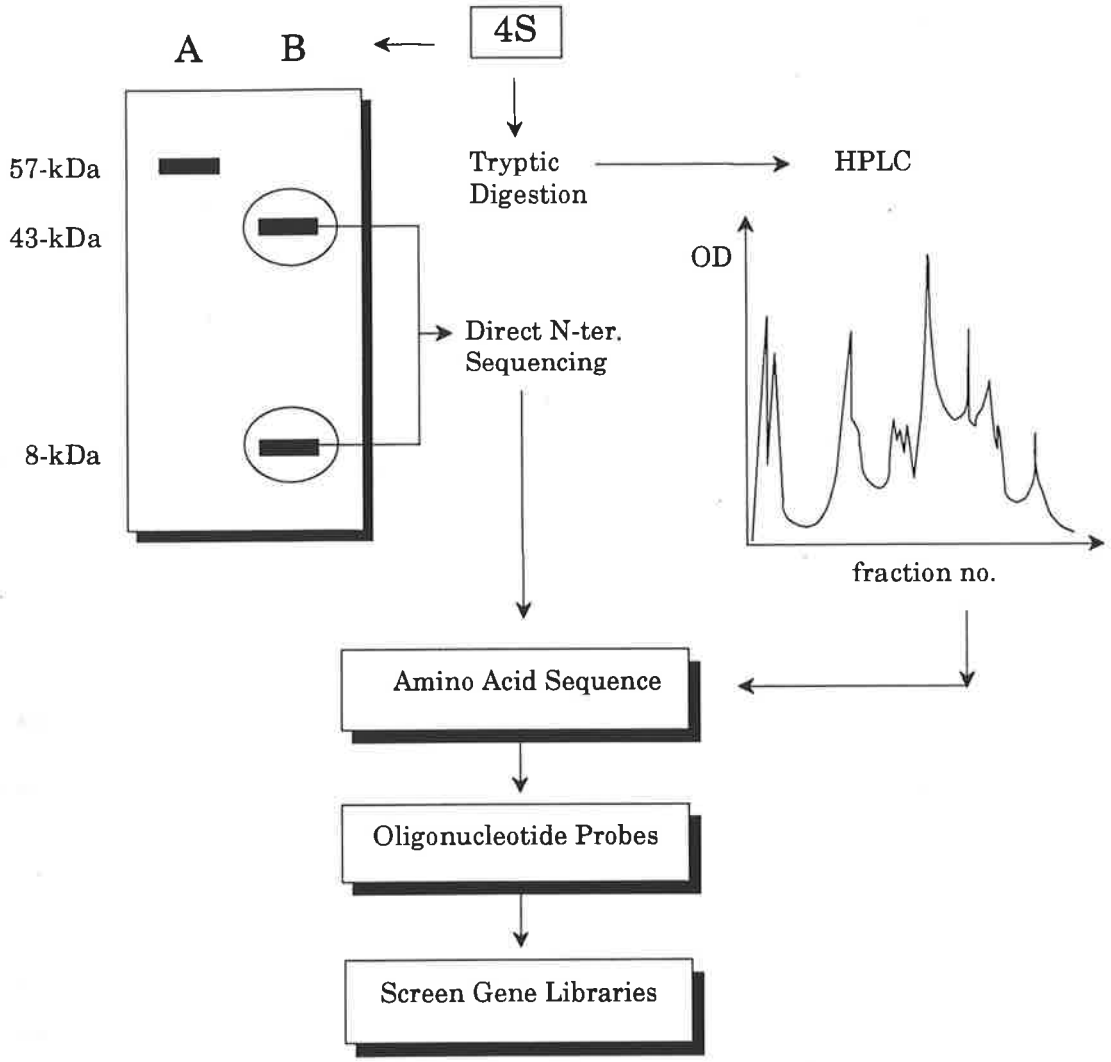


FIGURE 3.1 STRATEGY ADOPTED TO CLONE 4S

A flow diagram of the strategy used to clone the 4S gene. Amino acid sequence of purified 4S was obtained using two approaches; from tryptic fragments which were resolved by HPLC, or directly from the amino-termini of the 43-kDa and 8-kDa species. The 43-kDa and 8-kDa species were obtained by SDS-PAGE under reducing conditions (B) (A = non-reducing conditions). The amino acid sequence data was used to design oligonucleotide probes, which were used to screen cDNA and genomic libraries for a 4S clone.

from the promoter region of the 4S gene. The hypothesis that the fragment encoded 4S was initially supported by the collinearity between the deduced amino acid sequence of the exon and the amino acid sequence obtained from both a tryptic fragment and the amino-terminus of the 43-kDa species. The deduced amino acid sequence was confirmed as a sulphatase as the sequence displayed homology to other recently (1988-1989) cloned sulphatases, including human steroid sulphatase (STS), arylsulphatase A (ASA), glucosamine-6-sulphatase (Gln6S), and a sea urchin arylsulphatase (HpuARS). Chromosomal localization of the 1.3-kb *Hin*dIII fragment to 5q13-5q14, within the *ARSB* (arylsulphatase B) locus, further supported the hypothesis that the fragment encoded a portion of 4S.

Subsequent efforts to isolate a 4S cDNA clone by screening cDNA libraries with the 1.3-kb *Hin*dIII fragment were not successful. A scheme was designed to isolate a 4S cDNA clone using the MOPAC (mixed oligonucleotide primed amplification of cDNA) approach, based on the observation of conserved amino acid sequences between sulphatases. However this approach was not performed, as we became aware that a 4S cDNA had been cloned in April 1989 by Christoph Peters and Kurt von Figura (personal communication), and subsequently published (Peters *et al.*, 1990). Schuchman *et al.* (1990) also published the sequence of a 4S cDNA clone. The DNA sequence of the 1.3-kb *Hin*dIII fragment which contained the first exon and the 5'-flanking region of 4S was found to be identical to the published 4S cDNA sequences, except for a single nucleotide difference in the 5'-untranslated region.

3.2 SPECIFIC MATERIALS & METHODS

3.2.1 PEPTIDE ISOLATION & SEQUENCING

Amino acid sequence from purified 4S was obtained in two ways. Firstly, approximately 150 μg of 4S was digested with trypsin, the digest was applied to an Aquapore RP-300 (Brownlee, Santa Clara, California) cartridge (30 x 4.6 mm), and

the peptides eluted over 180 min with a linear gradient from 0-30% (v/v) acetonitrile in 0.1% (v/v) trifluoroacetic acid at a flowrate of 1 ml/min. Absorbance was followed and fractions were collected manually and stored at -70°C prior to the determination of their amino acid sequence by Edman degradation. Secondly, approximately 20 μg of human liver 4S was subjected to SDS-polyacrylamide gel electrophoresis and transferred to an Immobilon P membrane (Polyvinylidene difluoride; Millipore, Bedford, Massachusetts) essentially by the procedure of Matsudaira (1987), except that the SDS-polyacrylamide gel was pre-elecrophoresed overnight to remove unpolymerized acrylamide and 200 μl of 100 mM-sodium thioglycolate was added to the cathode buffer chamber before electrophoresis. The 43-kDa and the 8-kDa polypeptides were excised and directly sequenced by Edman degradation.

3.2.2 *IN SITU* CHROMOSOMAL HYBRIDIZATION

Techniques of *in situ* hybridization were used as described elsewhere (Robertson *et al.,* 1988a). The 4S probe was labelled with three tritiated nucleotides by nick translation. The chromosomes were denatured in 70% (v/v) deionized formamide, 2 x SSC, pH 7.0, at 70°C for 2 min, and the labelled probe was hybridized *in situ* at 37°C at a concentration of 0.2 μg/ml. The slides were dipped in Kodak NTB-2 nuclear research emulsion diluted 2:1 with water, exposed for 13 days, developed and stained, and the metaphases were scored. Initial localization was achieved by hybridizing to chromosomes prepared from lymphocyte cultures of two normal males. This localization was further investigated by hybridizing the 4S probe to chromosomes from a patient with the karyotype 46,XY,t(5;16)(q14.2;q22.1) (Callen *et al.,* 1988).

RESULTS & DISCUSSION:

ATTEMPTS TOWARD ISOLATION OF THE 4S GENE

3.3 DETERMINATION OF AMINO ACID SEQUENCE

The preparation of 4S from human liver and the transfer to Immobilon P membrane for amino-terminal sequencing was performed by Gary Gibson, and the tryptic digestion and HPLC fractionation of 4S was performed by Paul Nelson, of the Department of Chemical Pathology, The Women's and Children's Hospital, Adelaide. Peptide sequence was determined by Bresatec, Adelaide, South Australia.

The first step towards the isolation of the 4S gene was to determine the amino acid sequence of portions of the 4S protein. Gene probes derived from the peptide sequence could then be utilised to screen for the 4S gene in gene libraries. Gibson *et al.* (1987) had purified 4S to homogeneity from human liver (and also kidney, lung and urine) and found the enzyme had a molecular weight of 57-kDa, composed of disulphide-linked species of 43- and 8-kDa (Taylor *et al.,* 1990).

Amino acid sequence was generated from purified 4S in two ways; sequencing of tryptic fragments, and direct sequencing of the amino-termini of the 43 and 8-kDa species (Fig. 3.1). In the tryptic fragment approach, purified human liver 4S was digested with trypsin and resolved into individual peptide fragments by HPLC and selected fractions were amino acid sequenced. In the direct amino-terminal sequencing approach, purified human liver 4S was subjected to SDS-PAGE under reducing conditions, transferred to an Immobilon membrane, the 43 and 8-kDa species were excised, and the amino acid sequence determined. Tryptic digestion of

4S provided the initial amino acid sequence data while direct amino-terminal sequencing provided sequence data about a year later, enabling the isolation of the genomic clones. A summary of the amino acid sequence obtained from 4S is presented in Table 3.1. Note that the amino-terminus of the 43-kDa species (43-kDa *N*-ter) generated the same amino acid sequence as the tryptic peptide ASB III 52. The tryptic peptides ASB IV 34 and ASB II 61 also generated the same amino acid sequence.

3.4 DESIGN OF OLIGONUCLEOTIDE PROBES AND SCREENING OF GENE LIBRARIES

Kerri Beckmann of the Department of Chemical Pathology at the Women's and Children's Hospital, Adelaide, assisted in the construction of gene probes and the screening of gene libraries.

3.4.1 INTRODUCTION

The preferred approach to isolate the 4S gene at the commencement of the project (1987-1988) was to screen cDNA and genomic libraries with oligonucleotide probes derived from the amino acid sequences of tryptic fragments of 4S. The nucleotide sequence of each probe was based on the reverse complement of the conceptual mRNA sequence obtained from `reverse translation' of the amino acid sequence data. The number of different sequences within the probe population was minimized by designing the sequence from peptide regions with the lowest codon degeneracies. The degeneracy of the probes was further reduced by the substitution of deoxyinosine for all four deoxynucleotides at positions of total codon degeneracy. Deoxyinosine was selected as it had been observed to base-pair with all four deoxynucleotides (dC, dG and dT - Miura *et al.*, 1985 ; dA - Robertson *et al.*, 1988b). The degeneracy of the probe was also reduced by the selection of dG at positions of partial degeneracy that required the inclusion of both dG and dA in the

probe sequence, as dG could form a quasi-stable base-pair with dT in the inferred cDNA sequence, in addition to an orthodox Watson-Crick dG:dC pair.

The probes were used to screen cDNA and genomic libraries. In cDNA libraries, the frequency of a gene was assumed to be directly proportional to the frequency of it's expressed mRNA within the mRNA population of the tissue of origin. In contrast, in genomic libraries the frequency of a single-copy gene was assumed to be about the same as other single-copy genes, independent of the level of it's expressed mRNA. 4S has been observed to be present in low abundance relative to total cellular protein in all tissues examined. Furthermore, 4S was presumed to have a long half-life of the order of 1 to 2 weeks, as observed in normal cultured fibroblasts (Taylor *et al.*, 1990). These observations suggested that the abundance of specific mRNA and hence cDNA could be very low, hence cDNA libraries were screened which were derived from tissue sources where 4S had been or was likely to be isolated.

The gene libraries were plated out at relatively low plaque densities where each plaque could be clearly discerned, to ensure that the maximum signal intensity was obtained for plaques that were positive to a probe. In order to permit greater control of the washing stringency, a tetramethylammonium chloride (TMACl) solution was used to wash the gene library membranes containing the hybridized oligonucleotide probe (Wood *et al.*, 1985). The 4S oligonucleotide probes used in this project were derived from peptide sequences containing degenerate codons, and so each probe was a mixed population of sequences varying in base content, but of the same length. Usually, the optimum stringency is obtained when washing the hybridized membranes at a few degrees below the dissociation temperature (T_d) of the probe, which is dependent on the length and the GC-content of the probe for a given wash solution. The variation in GC-content within the probe-pool can make the selection of suitably stringent washing conditions for all species within the

Peptide	Sequence	location
ASB III 52	GASRPXHLVFLLADDLGWNDVGF<u>E</u> (24)	40 - 63
ASB II 20	YQIR (4)	103 - 106
ASB II 61	TGLQHQIIWPCQPSCVPLDEK (21)	107 - 127
ASB IV 34	TGLQHQIIWPCQPSCVPLDEK (21)	107 - 127
ASB III 47	(GE)F<u>I</u>(HT)(YI)FGYLLGSEDYYSHE (19)	161 - 179
ASB IV 27	WSLWEGGVR (9)	319 - 327
43-kDa *N*-ter	GASRPXHLVFLLADDLGWNXVG (22)	40 - 61
8-kDa *N*-ter	DPPTKTLWLFDIDRDPEERHDLSREYPHIVTKXL (34)	466 - 499

TABLE 3.1 AMINO ACID SEQUENCES OF PEPTIDES FROM HUMAN LIVER 4S

The amino acid sequences were obtained from tryptic peptides of purified 4S, and from direct amino-terminal sequencing of the 43-kDa and 8-kDa subunits. The amino acid residues are listed in their single letter codes. The length of each reported peptide sequence is in brackets at the end of each sequence. An X residue denotes an unidentified amino acid, underlined residues are different between the peptide sequence and the deduced cDNA, and residues in parentheses denote ambiguity at that position; the two options are listed within parentheses. Amino acid numbering corresponds to Peters *et al.* (1990).

probe population difficult. The use of TMACl in the wash solution was claimed to make the T_d of a probe dependent only on the length of the probe (and not on the base content), thus allowing the selection of washing conditions which would reduce the incidence of false-positive and false-negative clones (Wood *et al.*, 1985). A cautious approach was considered prudent during library screening, especially in the absence of a 4S positive-control. Therefore the washing stringency was generally kept relatively low to ensure that even weakly-positive clones were detected.

Initially, the oligonucleotide probes were derived from tryptic fragments of 4S. However, the peptide data could not be conclusively be assigned to 4S or alternatively to a contaminant protein, at the initial stage of the project. The peptide sequence of ASB IV 34 and ASB II 61 was chosen for the design of the gene probes because the sequence of the two independently isolated peptides was identical, and because the deduced mRNA sequence showed the least codon degeneracy overall compared to the other tryptic peptides, therefore aiding the design of low-degeneracy probes. The 26 nucleotide probe, ASB-1, was synthesized complementary to the mRNA sequence (Fig. 3.2). Reduction in the number of different oligonucleotide sequences present in the probe was achieved by the incorporation of deoxyinosine and dG at the appropriate positions, as previously described.

3.4.2 SCREENING OF cDNA LIBRARIES

The λgt11 human lung fibroblast cDNA library (Clontech, Palo Alto, USA) was selected for the initial screening attempts using the ASB-1 probe, as the average size of the cDNA inserts (1.3-kb) and the number of independent recombinant clones (2.2×10^6), were the largest of the cDNA libraries that were at hand. The library was plated-out on a Y1090 host at a relatively low density of 20,000 pfu (plaque forming units) per 140 mm plate, as described in Section 2.2.14. The

Consensus amino acid sequence and conceptual mRNA sequence for the ASB II 61 / ASB IV 34 tryptic peptide

```
    T   G   L   Q   H   Q   I   I   W   P   C   Q   P   S   C   V   P   L   D   E   K

5′ ACU GGU UUA CAA CAU CAA AUU AUU UGG CCU UGU CAA CCU UCU UGU GUU CCU UUA GAU GAA AAA 3′
     C   C   G   G   C   G   C   C       C   C   G   C   C   C   C   C   G   C   G   G
     A   A CUU               A   A       A           A   A       A   A CUU
     G   G   C                   G           G       G   G       G   G   C
             A                                   AGU                 A
             G                                     C                 G
```

Probe sequence:

```
3′  GTC GTG GTC TAT TAT ACC GGI ACG GT  5′
      T       T   G   G
```

ASB-1 primer

FIGURE 3.2 DESIGN OF OLIGONUCLEOTIDE PROBE ASB-1

The consensus amino acid sequence of the ASB II 61 and ASB IV 34 tryptic fragments of 4S is shown, together with all possible codons of the conceptual mRNA coding sequence. The sequence of the derived oligonucleotide probe, ASB-1, is aligned complementary to the mRNA sequence. The specific activity of the probe was increased by the incorporation of deoxyinosine (I) and the choice of dG, at selected positions of nucleotide degeneracy. The ASB-1 probe is a 26-mer oligonucleotide comprised of 16 species and contains 1 deoxyinosine residue.

plaques were transferred in duplicate to nylon membranes and hybridized overnight to the end-labelled ASB-1 oligonucleotide probe, according to Sections 2.2.12a and 2.2.13a. The filters were removed from the hybridization solution and washed first in 6 x SSC, then twice in TMACl-wash-solution at 55°C. The TMACl washes were at a low stringency to ensure the detection of weakly-positive clones in the first-round screen. Two clones that were positive to the ASB-1 probe, λ5.1 and λ6.1, were identified after autoradiography, from a screen of 1.2 x 10^5 recombinant clones.

Based on washing the probe-clone hybrids at a range of stringencies, the sequence of the clones that was complementary to the ASB-1 probe was estimated to be about 18-bp in length. 18-bp was shorter than the length of the probe (26-bp), and so suggested that the `positives' were probably false. However, it was considered that the inclusion of deoxyinosine and dG at selected positions of degeneracy in the ASB-1 probe may have reduced the expected T_d, as these deoxynucleotides may have formed H-bonds with a reduced strength, depending on the actual sequence present in the authentic 4S cDNA.

In order to confirm the identity of the two ambiguous positive clones, the DNA sequence complementary to the ASB-1 probe within the two clones was determined. λ5.1 and λ6.1 were purified, the DNA was extracted from plate lysates as described (Section 2.2.15), and the cDNA inserts were excised by digestion with the restriction endonuclease *Eco*RI. λ5.1 contained a 1-kb insert and λ6.1 contained a 2-kb insert (data not shown), which were subcloned into the plasmid pUC19 to enable the inserts to be produced in sufficient amounts for further analysis. To facilitate DNA sequencing and swift identification of the clones, the smallest restriction fragments of both clones that still hybridized to the ASB-1 primer were isolated. The pUC subclones of the cDNA inserts, designated pUC5.1 and pUC6.1, were digested with restriction endonucleases that cut frequently

(*Alu*I, *Hae*III, *Rsa*I, *Sau*3AI, and *Taq*I), and the digests were electrophoresed on a acrylamide gel. A 0.35-kb *Sau*3AI fragment common to both clones was identified that was positive to the ASB-1 probe. The full-length cDNA inserts of both clones and the 0.35-kb *Sau*3AI fragment were subcloned into the sequencing vectors M13mp18 and 19 (Yanisch-Perron *et al.*, 1985), to enable the determination of the DNA nucleotide sequence using the dideoxy chain-termination method (Sanger *et al.*, 1977).

DNA sequence was obtained from both ends of the 5.1 and 6.1 cDNA clones and the *Sau*3AI fragment (329-bp) with the USP primer (Fig. 3.3). DNA sequence was also obtained using the ASB-1 probe as a sequencing primer. The assembled DNA sequence showed that λ5.1 contained a 1-kb cDNA insert that was identical in sequence but shorter in length compared to the 2-kb cDNA insert of λ6.1 (Fig. 3.3a). The DNA sequence of the region complementary to the ASB-1 primer showed that 19 of 26 bases were identical with the ASB-1 probe, of which 15 bases were consecutively identical (Fig. 3.3b). Internal *Eco*RI sites were not found in the λ6.1 insert, although it was predicted based on the location of the terminal *Eco*RI sites of the shorter clone λ5.1. It was most likely that these *Eco*RI sites were derived from the linkers used in the construction of the library, and not the mRNA.

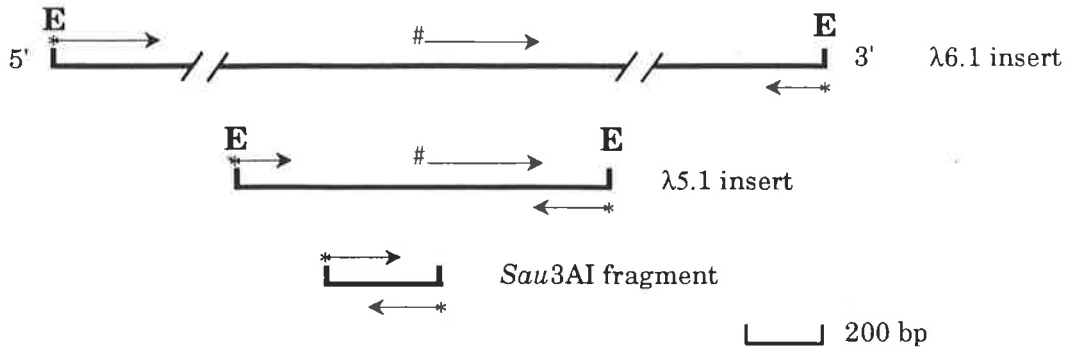
A total of 775-bp of contiguous DNA sequence was obtained around the ASB-1 site in the two cDNA clones. Translation of the DNA sequence into all six possible reading frames and comparison with the 4S peptide data showed no identity, except for four amino acids that coincided with the region of identity between the ASB-1 probe and the clones. Therefore, λ5.1 and λ6.1 were considered false-positives. No significant homology was detected between the 775-bp DNA sequence and the DNA sequences within the GenBank database (up to Release 80.0, December 1993) using the sequence comparison program FASTA.

It was considered technically difficult to distinguish a false-positive clone from a conceivable true-positive clone using the ASB-1 probe at the library screening stage. False-positive clones such as λ5.1 and λ6.1, with 19 of 26 bases identical to the probe, would exhibit roughly similar T_d to that of a conceivable true-positive 4S clone, with from 22 to 26 out of 26 bases identical to the probe. The extent of the identity in the authentic clone is dependent on the actual codon usage in the human cDNA at the four positions where dG was chosen (Fig. 3.2).

Therefore, an `adjacent primer' approach was followed for subsequent library screening, in order to select against false-positive clones. The approach utilised two probes which were adjacent or overlapped each other and were derived from the same peptide sequence. Library screening was performed with both probes in parallel for both the first- and second-round screens, or alternatively, a library screen was performed first with one probe and then any first-round positive clones were screened with the second probe. Clones that were positive to both probes were then selected for further analysis. The `adjacent primer' approach was used to construct the overlapping primers 4SS1 and 4SS2 from the ASB II 61 tryptic peptide (Fig. 3.4).

The adjacent probes 4SS1 and 4SS2 were used to screen the λgt11 lung fibroblast cDNA library for a 4S cDNA clone. The library was plated out on a Y1090 host at 30,000 pfu per plate and the plaques were transferred to nylon membranes and hybridized to the probes, as previously described. Library screening was performed with the two probes in parallel and also sequentially. The membranes were washed at a rather low stringency; 4SS1-probed filters were washed at 45°C in TMACl and 4SS2-probed filters were washed at 42°C, to ensure that weakly-positive clones would be detected. Twelve clones positive to the 4SS1 probe were identified after autoradiography, from first-round screening of a total of 7.2×10^5 recombinants, corresponding to about 0.3 library equivalents. However, none of

a. Sequence relationship of the cDNA inserts of λ5.1 and λ6.1



b. Comparison of the DNA sequence of ASB-1 with λ5.1 and λ6.1

```
5'          TGG CAI GGC CAG ATG ATT TGG TGC TG          ASB-1
              * **  *** *** *** *** ***  *
5'  .. aaa aaa acA GGg CAG ATG ATT TGG TGC aGc ttt ..   λ5.1 & λ6.1
```

FIGURE 3.3 THE RELATIONSHIP OF THE FALSE-POSITIVE CLONES λ5.1 AND λ6.1

a. The *Eco*RI inserts in λ5.1 and λ6.1 were sub-cloned into M13mp18 and M13mp19, and portions of the DNA sequence was determined using the universal sequencing primer (*) and ASB-1 (#). Shown in the Figure is the relationship of the 2-kb insert of λ6.1 and the 1-kb insert of λ5.1. Also indicated is the position of the ASB-1 probe and the 0.33-kb *Sau*3AI restriction fragment, which was positive to ASB-1. The arrows represent the direction and extent of the sequencing reactions. E = *Eco*RI site.

b. Comparison of the DNA sequence of the ASB-1 probe with the corresponding sequence obtained from the *Sau*3AI fragment indicating that 19 of 26 bases were identical, of which 15 bases were consecutively identical. Therefore, λ5.1 and λ6.1 were considered false positive clones which did not code for 4S.

the twelve clones were co-positive with the adjacent probe 4SS2, and so were not investigated further. It was considered that the unsuccessful screening for a 4S cDNA clone in the lung fibroblast cDNA library was due to a presumed low abundance of 4S clones in the library. Consequently, a cDNA library derived from another tissue source was screened, where it was thought that the abundance of 4S cDNA clones may be higher.

A λgt10 human liver cDNA library (kindly provided by G. Howlett, Department of Biochemistry, University of Melbourne), was screened with the 4SS1 probe. The library was plated out on a C600 host at a density of 20,000 pfu per plate and the plaques were transferred to nylon membranes, hybridized to the probe and washed in TMACl at 45°C, as previously described. Six clones positive to 4SS1 were obtained from a first-round screen of 3.2×10^5 recombinant clones; λ36, λ42, λ43, λ44, λ46 and λ47. However, none of the six clones were positive to 4SS2 upon second-round screening. Despite the negative result, the clones were retained for further analysis as there was a suspicion that the adjacent 4SS2 probe was defective. The 4SS2 probe had not been observed to give a single plaque-positive signal greater than the background out of a screen of 3.6×10^5 clones of the λgt11 lung fibroblast cDNA library. At the time it was also difficult to obtain an acceptable and reproducible background signal upon autoradiography of the 4SS2-probed membranes. Due to the technical difficulties experienced with the 4SS2 probe, the negative results obtained with the probe were disregarded and the six 4SS1 clones were considered as potential 4S positive clones.

In order to confirm the identity of the six ambiguous positive clones, the DNA sequence complementary to the 4SS1 probe within the clones was determined. The six clones were plaque-purified, the DNA was extracted from plate lysates as described (Section 2.2.15), and the cDNA inserts were excised by digestion with *Eco*RI. λ43, λ46 and λ47 contained 0.5-kb inserts and λ36, λ42 and λ44 contained 1-

Consensus amino acid sequence and conceptual mRNA sequence for the ASB II 61 / ASB IV 34 tryptic peptide

```
    T   G   L   Q   H   Q   I   I   W   P   C   Q   P   S   C   V   P   L   D   E   K

5' ACU GGU UUA CAA CAU CAA AUU AUU UGG CCU UGU CAA CCU UCU UGU GUU CCU UUA GAU GAA AAA
     C   C   G   G   C   G   C   C       C   C   G   C   C   C   C   C   G   C   G   G
     A   A CUU               A   A       A           A   A       A   A CUU
     G   G   C                   G           G       G   G       G   G   C
             A                                       AGU               A
             G                                         C               G
```

Probe sequences:

```
3'  GTT GTA GTT TAI TA  5'      4SS2 primer
      C   G   C

                3'  TAI TAI ACC GGI ACA GTT GG  5'      4SS1 primer
                                    G   C
```

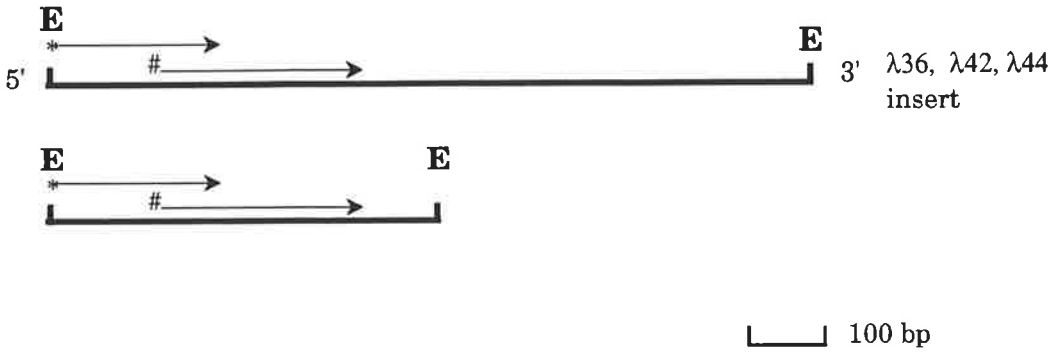
FIGURE 3.4 DESIGN OF OLIGONUCLEOTIDE PROBES 4SS1 & 4SS2

The consensus amino acid sequence of the ASB II 61 and ASB IV 34 tryptic fragments of 4S is shown, together with all possible codons of the conceptual mRNA coding sequence. The sequence of the derived oligonucleotide probes, 4SS1 and 4SS2, are aligned complementary to the mRNA sequence. The order of the probes from the amino-terminal end is 4SS2 and 4SS1 respectively. The specific activity of both probes was increased by the incorporation of deoxyinosine (I) at selected positions of nucleotide degeneracy. The 4SS1 probe is a 20-mer oligonucleotide comprised of four species and contains three deoxyinosine residues, and the 4SS2 probe is a 14-mer oligonucleotide comprised of eight species and contains one deoxyinosine residue.

kb inserts, which were subcloned into M13mp19 for dideoxy DNA sequencing. DNA sequence was obtained from each of the M13 subclones with the USP primer and also with 4SS1. All six clones generated the same DNA sequence with the 4SS1 primer, and the DNA sequence obtained with the USP primer was also identical in all six clones. Comparison of the 4SS1-primed sequence with the USP-primed sequence showed that they overlapped (Fig. 3.5a). The DNA sequence obtained showed that the six clones were derived from the same gene, as the 500-bp inserts of the clones λ43, λ46 and λ47 were contained within and shared a common end with the 1-kb inserts of the clones λ36, λ42 and λ44. DNA sequence of the region complementary to the 4SS1 primer showed that 17 of 20 bases were identical with the 4SS1 probe, of which 16 bases were consecutively identical (Fig. 3.5b). An internal *Eco*RI site was not found in the longer clones λ36, λ42 and λ44, although it was predicted based on the location of the 3'-terminal *Eco*RI site in the shorter clones λ43, λ46 and λ47. It was most likely that the *Eco*RI site present at the 3'-end of the shorter clones was derived from the linkers used in the construction of the library, and not the mRNA.

A total of about 400-bp of contiguous DNA sequence was obtained around the 4SS1 site in the six clones. Translation of the DNA sequence into all six possible reading frames and comparison with the 4S peptide data showed no identity, except for five amino acids that coincided with the region of identity between the 4SS1 probe and the clones. Therefore, λ36, λ42, λ43, λ44, λ46 and λ47 were considered false-positives. The 400-bp of DNA sequence was compared with the DNA sequences within the GenBank database (up to Release 80.0, December 1993) using the sequence comparison program FASTA. The clones are considered to encode a portion of the currently unpublished sequence of the human mitochondrial enzyme, NADH-ubiquinone reductase (also known as the 51-kDa subunit of NADH hydrogenase), based on the similarity between the compiled sequence of the clones and the cDNA sequences of NADH-ubiquinone reductase/NADH hydrogenase from

a. Sequence relationship of the cDNA inserts of λ36-47



b. Comparison of the DNA sequence of 4SS1 with λ36-47:

```
5'          GGT TGA CAI GGC CAI ATI AT          4SS1
            *** *** *** *** *** * *
5'  .. ggt GGT TGA CAT GGC CAG AgA tgt tga ..   λ36-47
```

FIGURE 3.5 THE RELATIONSHIP OF THE FALSE-POSITIVE CLONES λ36, λ42, λ43, λ44, λ46, & λ47

a. The *Eco*RI inserts in λ36, λ42, λ43, λ44, λ46, and λ47 were sub-cloned into M13mp19, and portions of the DNA sequence was determined using the universal sequencing primer (USP) and 4SS1. Shown in the Figure is the relationship of the 1-kb insert of λ36, λ42, and λ44, and the 0.5-kb insert of λ43, λ46, and λ47. Also indicated is the position of the 4SS1 probe. The arrows represent the direction and extent of the sequencing reactions. E = *Eco*RI site.

b. Comparison of the DNA sequence of the 4SS1 probe with the corresponding sequence obtained from the six cDNA clones indicated that 17 of 20 bases were identical, of which 16 bases were consecutively identical. Therefore, λ36, λ42, λ43, λ44, λ46, and λ47 were considered false positive clones which did not code for 4S.

cow (about 84% identity), *N. crassa* (about 68% identity), and *P. denitrificans* (about 64% identity). The results obtained from the screening of the liver cDNA library suggested that it had been exhaustively screened yet did not contain a 4S cDNA clone. The six clones that were positive to 4SS1 were derived from the same gene, of which three identical clones were about 0.5-kb in length and were contained within the other three identical clones of about 1-kb in length.

In review, the extensive screening of cDNA libraries for a 4S clone was not successful. A conservative strategy was followed of characterising even weakly-positive clones, which was thought to ensure the detection of a true 4S cDNA clone if it was present in the cDNA library. The failure to obtain a 4S cDNA clone was considered to be due to the very low abundance of 4S mRNA in tissues generally and those from which the cDNA libraries were derived. Consequently, screening for the 4S gene in genomic libraries was considered more likely to be successful, on the assumption that single-copy genes were represented at about the same frequency in the library, independent of the level at which a gene was expressed. However, screening for genomic clones might not be successful with probes derived from peptide data, as there was a small possibility that the DNA sequence complementary to the probe might be split by an intron. This problem, if it occurred, was to be overcome by designing probes to different portions of the available peptide sequence. The overall priority was still to isolate a 4S cDNA clone, as the definitive determination of the 4S coding sequence was only feasible with a cDNA and not a genomic clone. Therefore, it was envisaged that 4S genomic clones, once identified, would be used to screen cDNA libraries for a complete 4S cDNA clone.

3.4.3 GENOMIC LIBRARY SCREENING

Concurrent with the decision to screen genomic libraries for the 4S gene, amino acid sequence was obtained from the amino-termini of the 43-kDa and 8-kDa

species (Table 3.1). The sequences of the amino-terminus of the 43-kDa species and the tryptic peptide ASB III 52 were shown to be collinear. As the amino-terminus of the 43-kDa species had been definitively identified and the amino acid sequence confirmed by a tryptic peptide, the sequence became a focus of probe design. The oligonucleotide probes 4SS3 and 4SS4 were designed, based on the `adjacent primer' approach (Fig. 3.6).

A λEMBL3A human leukocyte genomic DNA library (kindly provided by Iain Young, Australian National University, Canberra) was screened with both probes. The genomic library was plated out on LE392 at a density of 37,000 pfu per 140 mm plate. The plaques were transferred in duplicate to nylon membranes and each duplicate membrane was hybridized overnight in parallel to either of the two end-labelled probes. The filters were removed from the hybridization solution and washed first in 6 x SSC and then twice in TMACl-wash-solution at 45°C. The TMACl washes were at a low stringency to ensure the detection of even weakly-positive clones in the first-round screen. Five clones positive to both 4SS3 and 4SS4 probes were identified after autoradiography, from a first-round screen of 4 x 10^5 recombinant clones, which corresponds to about 2 genome equivalents. Second-round screening left two clones, λ113 and λ122, that remained positive to both probes. The next Section will describe the characterization of λ113 and λ122.

3.5 CHARACTERIZATION OF THE GENOMIC CLONES

Kerri Beckmann of the Department of Chemical Pathology, The Women's and Children's Hospital, Adelaide, assisted in the characterization of the genomic clones.

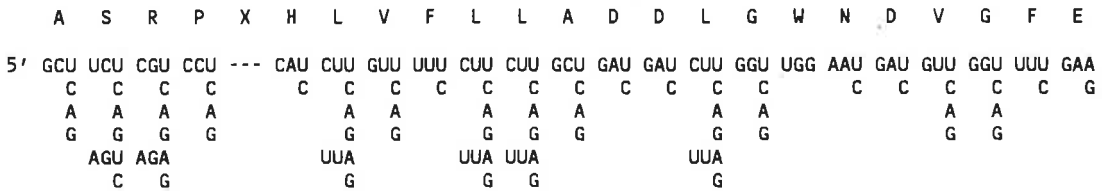
3.5.1 ISOLATION OF A 1.3-kb *Hin*dIII FRAGMENT

This Section reports the characterization of the two genomic clones, λ113 and λ122, and the evidence supporting the hypothesis that they encoded the amino-terminal portion of 4S. The clones were characterized initially by identification of the smallest fragment that hybridized to probes 4SS3 and 4SS4, and then determining the DNA sequences of the fragments that were complementary to the probes. This was achieved by digestion of each λ clone with a range of restriction endonucleases, electrophoresis of the digests on a 0.7% (w/v) agarose gel, transfer of the fragments to a nylon membrane and hybridization of the 4SS4 probe (and later 4SS3) to the membrane. A 1.3-kb *Hin*dIII fragment common to both clones and which hybridized to both oligonucleotide probes was identified (data not shown). The 1.3-kb *Hin*dIII fragment was isolated from both λ clones and the fragments were subcloned into the plasmid vector pUC19 to enable the inserts to be produced in sufficient amounts for further analysis. The fragments were also subcloned into the sequencing vector M13mp19 to enable the nucleotide sequence to be determined.

3.5.2 DEDUCED SEQUENCE COLLINEARITY WITH THE 4S 43-kDa *N*-TERMINUS

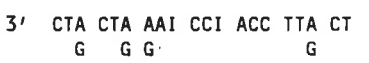
The 1.3-kb *Hin*dIII fragments from λ113 and λ122 were considered identical. DNA sequence was generated from both ends of each M13 subclone with the universal sequencing primer, and internal DNA sequence was obtained with the probe 4SS4 used as a sequencing primer. In each case the DNA sequence obtained from both subclones was identical. The identity of the 1.3-kb *Hin*dIII fragment as encoding for 4S was supported initially by sequencing the M13 subclones with 4SS4. The peptide sequence obtained by the translation of the 4SS4-primed DNA sequence displayed collinearity with the amino-terminal portion of the 43-kDa amino-terminal peptide sequence (data not shown).

Consensus amino acid sequence and conceptual mRNA sequence for the amino-terminal of the 43-kDa subunit / ASB III 52 tryptic peptide



Probe sequences:

4SS3 primer



4SS4 primer

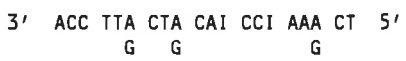


FIGURE 3.6 DESIGN OF OLIGONUCLEOTIDE PROBES 4SS3 & 4SS4

The consensus amino acid sequence of the amino-terminal of the 43-kDa subunit and the ASB III 52 tryptic fragment of 4S is shown, together with all possible codons of the conceptual mRNA coding sequence. The sequence of the derived oligonucleotide probes, 4SS3 and 4SS4, are aligned complementary to the mRNA sequence. The order of the probes from the amino-terminal end is 4SS3 and 4SS4 respectively. The specific activity of both probes was increased by the incorporation of deoxyinosine (I) at selected positions of nucleotide degeneracy. The 4SS3 probe is a 20-mer oligonucleotide comprised of 16 species and contains two deoxyinosine residues, and the 4SS4 probe is a 20-mer oligonucleotide comprised of eight species and contains two deoxyinosine residues.

In order to define the extent of the 4S coding region within the 1.3-kb *Hin*dIII insert, additional DNA sequence was determined from overlapping fragments of the insert. The 1.3-kb *Hin*dIII fragment was prepared from pUC subclones and the purified fragment was digested with several restriction endonucleases including *Alu*I, *Hae*III, *Rsa*I, and *Taq*I, to generate a range of different sized DNA fragments, which were then end-filled by Klenow. These restriction enzymes were selected as they cut frequently on account of their four-base specificity. The fragments were resolved by electrophoresis on a 5% (w/v) polyacrylamide gel and selected fragments were excised and the DNA isolated. Dot-blots of the DNA fragments transferred onto a nylon membrane and probed with 4SS4 indicated that a 270-bp *Alu*I fragment was the smallest convenient sub-fragment of the 1.3-kb *Hin*dIII clone that contained the probe-binding site. The end-filled 270-bp *Alu*I fragment, and other selected *Alu*I and *Rsa*I fragments, were then cloned into M13mp18 *Sma*I sequencing vectors.

The sequencing strategy utilized in the characterization of the nucleotide sequence of the 1.3-kb *Hin*dIII insert is indicated in Figure 3.7. Also shown in the diagram is the position of the oligonucleotide probes, 4SS3 and 4SS4. The sequence is required from both DNA strands to ensure that it is completely accurate, especially in regions that are G+C rich and are therefore especially prone to sequencing artefacts. Sequence was obtained from both strands in the designated areas (Fig. 3.7), for the other areas DNA sequence was obtained from one strand only. For example, the DNA sequence extending 3'-adjacent to the 3'-end of coding exon 1 was obtained from one strand only, and therefore was not shown in Figure 3.8 as the sequence contained ambiguities, most likely due to artefacts generated by the very high G+C content. At the time the sequencing of the 1.3-kb *Hin*dIII was undertaken, the focus was on the isolation of a cDNA clone, and therefore the primary interest was in the precise sequence determination of the coding region in the 1.3-kb *Hin*dIII insert. As the sequence 3'-adjacent to coding exon 1 was

Restriction endonuclease sites in the 1.3-kb *Hin*dIII fragment

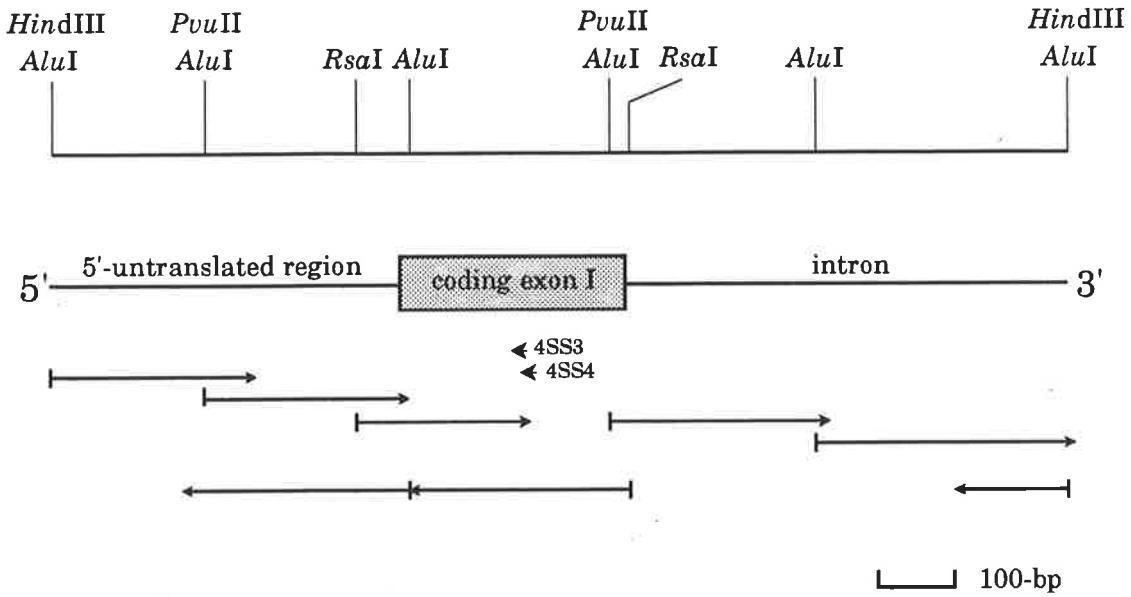
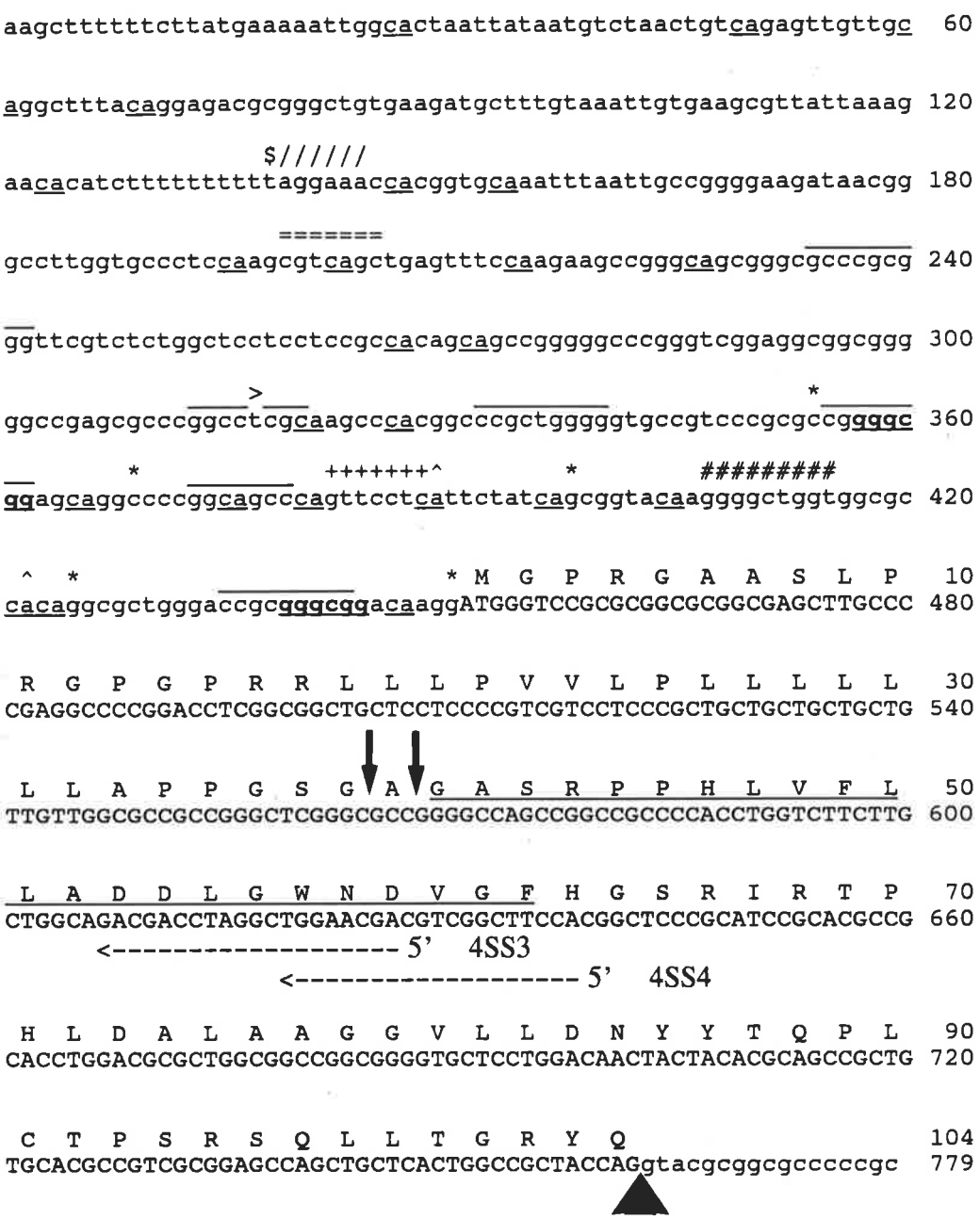


FIGURE 3.7 STRATEGY FOR SEQUENCING THE 1.3-kb *Hin*dIII FRAGMENT

The DNA sequence of the 1.3-kb fragment was generated from restriction endonuclease fragments subcloned into M13mp18 and M13mp19, and sequenced with the universal sequencing primer and 4SS4. The arrows represent the direction and extent of the sequencing reactions. Also shown in this diagram is the position of the oligonucleotide probes, 4SS3 and 4SS4 (arrowheads), the deduced coding region of the exon (shaded box), the 5' untranslated region, and the intron.

FIGURE 3.8 DNA SEQUENCE OF THE 5'-NONCODING REGION AND THE FIRST EXON OF 4S

The sequence shows the first 779-bp from the 5'-end of the 1.3-kb *Hin*dIII fragment. The nucleotide and amino acid numbers are in the right-hand margin. The coding region in the nucleotide sequence is shown in upper-case letters, and the 5'-flanking sequence and the beginning of the first intron are in lower-case letters. The deduced amino acid sequence is shown in upper-case letters in single letter code above the nucleotide sequence. The two vertical arrows represent the sites of proposed cleavage by signal peptidase based on the `weighted matrix' method (von Heijne, 1986). Amino acid residues co-linear with the amino acid sequence of the 43-kDa amino-terminal/ASB III 52 peptide are underlined. The locations of the two probes 4SS3 and 4SS4 are shown below the nucleotide sequence. The arrowhead indicates the position of an intron 5' splice donor site. Highlighted sequences 5'-adjacent to the coding region include fair-scoring consensus transcription initiation or CAP sites (∧), ca dinucleotides (underlined) and two canonical SP1 sites (bold, underlined) (Bucher, 1990). Recently, Modaressi *et al.* (1993) have determined the major CAP site (>) and other CAP sites (*) used in cultured fibroblasts, as well as identifying recognition sites for the following transcription factors; AP-2 (____), NF-S (===), PEA3 (///), Pu.1 (+++) and an additional SP1 site (###). The single nucleotide difference (T_{138}) between the sequence presented, and the sequences of Schuchman *et al.* (1990) and Modaressi *et al.* (1993) is indicated by a $ symbol.


aagctttttttcttatgaaaaattggcactaattataatgtctaactgtcagagttgttgc 60
aggctttacaggagacgcgggctgtgaagatgctttgtaaattgtgaagcgttattaaag 120
$//////
aacacatctttttttttttaggaaaccacggtgcaaatttaattgccggggaagataacgg 180
=======
gccttggtgccctccaagcgtcagctgagtttccaagaagccgggcagcgggcgcccgcg 240
ggttcgtctctggctcctcctccgccacagcagccgggggcccgggtcggaggcggcggg 300
ggccgagcgcccggcctcgcaagccacggcccgctgggggtgccgtcccgcgccgcgggc 360
gqagcaggccccggcagcccagttcctcattctatcagcggtacaaggggctggtggcgc 420
cacaggcgctgggaccgcgggcggacaaggATGGGTCCGCGCGGCGCGGCGAGCTTGCCC 480
CGAGGCCCCGGACCTCGGCGGCTGCTCCTCCCCGTCGTCCTCCCGCTGCTGCTGCTGCTG 540
TTGTTGGCGCCGCCGGGCTCGGGCGCCGGGGCCAGCCGGCCGCCCCACCTGGTCTTCTTG 600
CTGGCAGACGACCTAGGCTGGAACGACGTCGGCTTCCACGGCTCCCGCATCCGCACGCCG 660
<---------- 5' 4SS3
<---------- 5' 4SS4
CACCTGGACGCGCTGGCGGCCGGCGGGGTGCTCCTGGACAACTACTACACGCAGCCGCTG 720
TGCACGCCGTCGCGGAGCCAGCTGCTCACTGGCCGCTACCAGgtacgcggcgcccccgc 779

considered to be the 5'-end of an intron, and therefore non-coding (see below), the precise determination of the DNA sequence in both strands of this and other presumed non-coding regions was not considered a priority.

Conceptual translation of the sequence indicates the presence of an open reading frame which initiates at nucleotide 451 and extends to the position of a 5'-splice site which marks the end of the first exon (Fig. 3.8). Amino acid residues 40 to 63 of the deduced protein are identical to the sequence of the amino-terminus of the 43-kDa species of 4S for 22 of the 23 residues. The discrepancy at residue 63 (histidine in the deduced sequence, glutamic acid in the peptide) is most probably due to a peptide sequencing error resulting from the low signal obtained at the end of the amino acid sequencing run. The otherwise perfect collinearity strengthened the hypothesis that the genomic clones coded for a portion of 4S and included the amino-terminus of the 43-kDa subunit.

The nucleotide sequence 5' to the putative initiating methionine codon has features consistent with the consensus sequence for translation initiation sites in eukaryotes (Kozak, 1987). In particular, the position -3 from the initiating ATG in the consensus sequence is reported as a purine (usually an A residue), and is used in 97% of the initiation sequences examined (Kozak, 1987). The 4S sequence also has an A residue at this position. The consensus sequence also reports a G residue at +4 that is used in 46% of initiation sequences and a C residue at -5 that is used in 39% of initiation sequences. The 4S sequence has the reported residue at each of these positions. The sequence 5'-flanking the initiation of translation is thought to provide the correct context for initiation and to stabilize the binding of the ribosome to the mRNA.

In eukaryotes, the signal peptide is a transient sequence located at the amino-terminal end of proteins destined for export which serves to direct translocation of

the protein into the endoplasmic reticulum. The first 38-39 amino acids of the deduced amino acid sequence immediately preceding the 43-kDa amino-terminal region display features characteristic of a signal peptide, e.g. a basic amino-terminus, hydrophobic core and a more polar carboxyl-terminal region. The first 45 amino acids at the amino-terminus of the deduced protein were analyzed using the weighted matrix method (von Heijne, 1986) to determine the predicted sites of cleavage between the signal peptide and the mature protein. The weighted matrix method was derived from 161 eukaryotic signal sequences and claimed to have a predictive accuracy of 75-80%. Two likely cleavage sites were predicted in different positions of the deduced amino acid sequence of 4S, the first between Gly38 and Ala39 (von Heijne score = 14.1) and the second between Ala39 and Gly40 (von Heijne score = 13.4) (Fig. 3.8).

Direct amino-terminal sequencing of the 43-kDa subunit indicated the presence of glycine (amino acid 40), alanine (amino acid 41) and a small amount of serine (amino acid 42) at the amino-terminus. Subsequent sequencing cycles showed the superposition of a faint preview sequence on the major sequence, indicating that the 43-kDa subunit had a staggered amino-terminus (data not shown). On the basis of this data, it was considered plausible that cleavage of the signal peptide occurred between Gly38-Ala39, and/or between Ala39-Gly40. Subsequently, limited and variable proteolysis occurred at the amino-terminal-end of the 43-kDa subunit, such that up to three amino acids were removed, generating a 43-kDa species with a staggered amino-terminus.

Peters *et al.* (1990) suggested that signal peptidase cleavage of 4S may occur between Ala41-Ser42 in addition to between Ala39-Gly40. Cleavage between Ala41-Ser42 is unlikely, not only because of the poor score (von Heijne score = 7.7) but also because Ala41 was seen as a major component of the amino-terminus of the 43-kDa subunit. Schuchman *et al.* (1990) suggested that a possible site of

signal peptidase cleavage was between Gly40-Ala41, as they determined the amino-terminal amino acid of 4S to be Ala41. Again this is unlikely since Gly40 is seen as a component of the amino-terminus by us, and cleavage between Gly40-Ala41 has a low von Heijne score of 7.3 and not 13.4 as claimed.

The observed removal of up to three amino acids at the amino-terminus of the 43-kDa species is thought to occur *in vivo*, during intracellular transport or more likely when the protein reaches the lysosome. Other lysosomal enzymes have been observed to undergo additional proteolytic processing after the removal of the signal peptide, this can include trimming of amino-terminal and carboxyl-terminal residues, as well as cleavage within the polypeptide (reviewed by Hasilik, 1992). For example, the precursor of the β-subunit of β-hexosaminidase has been reported to undergo trimming of two amino acids at the amino-terminus after the removal of the signal peptide, followed by the removal of a further three amino acids at the amino-terminus to generate the mature form observed within the lysosome (Quon *et al.*, 1989). Limited proteolysis, of the order of one to five amino acids, has been suggested to occur at the amino-terminus of rat liver α-L-fucosidase after cleavage of the signal peptide (Fisher & Aronson, 1989). In addition, the 24-kDa subunit of glycosylasparaginase has been reported as being "frayed", with one-half of the subunit population missing an amino-terminal serine residue (Tollersrud & Aronson, 1989). It is generally thought that this additional trimming is a consequence of the proteolytic environment of the lysosome, and is not generally required for the acquisition of enzyme activity (Neufeld, 1991; Hasilik, 1992).

The 4S open reading frame contained within the 1.3-kb *Hin*dIII fragment (Fig. 3.8) was terminated by a consensus exon-intron splice site between nucleotides 762 and 763 (Ohshima & Gotoh, 1987; Shapiro & Senapathy, 1987). The exon encoded 104 amino acids, which consisted of a 38 amino acid signal peptide and the first 66 amino acids of the 43-kDa component of 4S. Later, the existence of the intron was

confirmed when the deduced amino acid sequence of the reading frame was compared to the 4S cDNA sequence (Peters *et al.,* 1990; Schuchman *et al.,* 1990). The genomic clones λ113 and λ122 did not encode other exons of 4S, as no other restriction fragments of either clone (beside the 1.3-kb *Hin*dIII fragment) hybridized to the 4S cDNA clone provided by Christoph Peters and Kurt von Figura (see Section 3.8.1). This suggested that the first intron was larger than 14-kb in length, based on the restriction map of the genomic clones (data not shown).

3.5.3 SEQUENCE HOMOLOGY TO SULPHATASES

The deduced amino acid sequence of 4S displayed similarities to other (then) recently cloned sulphatases (1988-1989), including human steroid sulphatase, arylsulphatase A, glucosamine-6-sulphatase, and a sea urchin arylsulphatase. The first published amino acid sequence homology between sulphatases was reported by Robertson *et al.* (1988b), who observed regions of similarity between the sequences of glucosamine-6-sulphatase and the previously published steroid sulphatase (Yen *et al.,* 1987). Comparison of the available coding sequence of the 4S 1.3-kb *Hin*dIII fragment with the steroid sulphatase cDNA also identified regions of sequence similarity (data not shown). By early 1989, the complete cDNA sequences for human steroid sulphatase (STS, Yen *et al.,* 1987; Stein *et al.,* 1989b), arylsulphatase A (ASA, Stein *et al.,* 1989a) and a sea urchin arylsulphatase (HpuARS, Sasaki *et al.,* 1988) had been published. The published cDNA sequence for glucosamine-6-sulphatase (Gln6S) was not complete, subsequently however the sequence of a cDNA clone encoding a further 133 amino acids in the 5'-direction relative to the published clone was obtained prior to publication (D. Robertson, personal communication; and later Robertson *et al.*, 1992). The 4S genomic clone encoded only the first 104 amino acids.

Pair-wise comparisons were performed between the deduced amino acid sequences of the sulphatases in order to identify conserved residues, using the programs

SEQHP and SEQA. The programs aligned sequences based on identification of local regions of similarity, and global comparisons of the sequences, respectively. The output was then manually assembled into a multiple alignment, as a multiple sequence alignment program was not available at this stage of the project. The published alignment between ASA and STS was used as a guide (Stein *et al.*, 1989a). Inspection of the multiple sequence alignment identified blocks of amino acids (one to three residues) which were identical between the five sulphatases (Fig. 3.9). The greatest degree of similarity was observed in the amino-terminal portion of the sulphatases. The sequence homology observed between the deduced partial sequence of 4S and the other sulphatases supported the hypothesis that the 1.3-kb *Hin*dIII fragment encoded a portion of a sulphatase. Recently, the publication of the sequences of a number of sulphatases, as well as the complete coding sequence of 4S, has allowed the assembly of a more detailed multiple sequence alignment, which is presented in Section 3.9.1.

3.5.4 ORIGIN OF THE 43-kDa AND 8-kDa SPECIES: A MODEL

The original multiple sequence alignment (Fig. 3.9) also suggested a model accounting for the origin of the 43-kDa and 8-kDa species of 4S. The 4S amino acid data obtained from tryptic peptides and direct amino-terminal sequencing (Table 3.1) were mapped where possible onto the multiple alignment, using the program SEQHP. The following 4S peptide sequences could be aligned; tryptic peptides ASB III 52, ASB III 47, ASB IV 27 and ASB IV 34, and the direct amino-terminal amino acid data, 43-kDa *N*-ter and 8-kDa *N*-ter. The locations of the ASB III 52/43-kDa *N*-ter, ASB IV 27, and 8-kDa *N*-ter sequences could be confidently assigned, based on their modest degree of homology with the other sulphatases. The location of the tryptic peptides ASB III 47 and ASB IV 34 could only be tentatively assigned however, due to the lower degree of homology.

The relative positions of the 43-kDa and 8-kDa amino-terminal peptide sequences assigned on the multiple alignment suggested a model of the subunit structure of 4S, in which a precursor polypeptide was proteolytically cleaved into an amino-terminal 43-kDa species and a carboxyl-terminal 8-kDa species. Later, the relative locations of the 43-kDa and 8-kDa species were exploited in the design of a PCR-based approach to clone the cDNA for 4S, as described in Section 3.7.2.

3.5.5 THE PROMOTER REGION OF THE 4S GENE

The sequence 5'-adjacent to the coding sequence of 4S was compared with other sulphatase sequences, in order to find any nucleic acid sequence features that may be shared between sulphatases. No significant homologies were found between the 5'-adjacent sequence of 4S and the 5'-untranslated regions of human steroid sulphatase (Stein *et al.,* 1989b), human arylsulphatase A (Stein *et al.,* 1989a), and two sea urchin arylsulphatases (Sasaki *et al.,* 1988; Yang *et al.,* 1989), using the program SEQH.

In order to identify potential RNA polymerase II transcriptional promoters or other gene-control elements, the DNA sequence 5'-adjacent to the initiating methionine in the 1.3-kb *Hin*dIII clone was searched using the weighted-consensus sequences for known promoter and control elements (Bucher, 1990). Consensus TATA sequence motifs were absent. The TATA sequence motif is thought to direct the accurate alignment of the RNA polymerase relative to the transcription initiation site (Benoist & Chambon, 1981). Consensus CCAAT motifs were also absent; the function of the CCAAT motif is less well defined. However, potential binding-sites for the transcription factor SP1 were present at nucleotide positions 360 and 442, with sequences identical to the core consensus sequence (GGGCGG) (Fig. 3.8) (Bucher, 1990). The 5'-untranslated sequence is notably G+C rich; 60.2% from nucleotide position 1 to 450, and 77.1% from nucleotide 181 to 450. These sequence features; a high G+C content, absence of the TATA and CCAAT consensus

FIGURE 3.9 ALIGNMENT OF THE DEDUCED 4S CODING SEQUENCE OF THE 1.3-kb *Hin*dIII FRAGMENT WITH OTHER SULPHATASE SEQUENCES

Multiple sequence alignment between the deduced amino acid sequences of the 1.3-kb *Hin*dIII genomic fragment of 4S, and human arylsulphatase A (ASA, Stein *et al.*, 1989a), steroid sulphatase (STS, Stein *et al.*, 1989b), glucosamine-6-sulphatase (Gln6S, Robertson *et al.*, 1988b; D. Robertson, personal communication), and sea urchin arylsulphatase (HpuARS, Sasaki *et al.*, 1988). The amino acid sequences are numbered in the right margin. The residue numbers for Gln6S are in brackets as the sequence is not complete at the amino-terminal end. Underlined sequences are either derived from tryptic peptides of 4S (ASB III 52, ASB III 47, ASB IV 27, ASB IV 34) or direct amino-terminal sequencing of the 43 & 8-kDa species (43-kDa *N*-ter, 8-kDa *N*-ter). Residues totally conserved between the sequences are denoted by an asterisk. Spaces (-) were inserted into the cDNA sequences to maximize the degree of homology between sequences. The multiple alignment was manually assembled from pairwise alignments of the sequences generated using the programs SEQHP and SEQA.


4S PEPTIDE 43-kDa N-ter/ASB III 52:GASRPXHLVF
4S MGPRGAASLPRGPGPRRLLLPVVLPLLLLLLLAPPGSGAGASRPPHLVF 49
ASA MGAPRSLLLALAAGLAVA-RPPNIVL 25
STS MPLRKMKIPFLLLFFLWEAESHEASRPNIIL 31
HpuARS MKSAPFLFLLGLLGLVTAQTQDPALLDLLRENPDLLSLLLQSNEHRAPLV-KPNVVL 56
Gln6S TRRPNVVL (8)
4S PEPTIDE LLADDLGWNDVG
4S LLADDLGWNDVGFHG-SRIRTPHLDALAAGG-VLLDNYYTQPLCTPSRSQLLTGRYQ 104
ASA IFADDLGYGDLGCYGHPSSTTPNLDQLAAGGLRFTDFYVPVSLCTPSRAALLTGRLPVRMGM----YPGV 91
STS VMADDLGIGDPGCYGNKTIRTPNIDRLASGGVKLTQHLAASPLCTPSRAAFMTGRYPVRSGMASWSRTGV 101
HpuARS LVADDMGSGDLTSYGHPTQEAGFIDKMAAEGLRFTNGYVGDAVCTPSRSAIMTGRLPVRIGT--FGETRV 124
Gln6S -LTDD---QDEVL-GGMTPLKKTKALIGEMGMTFSSAYVPSALCCPSRASILTGKYPHNHHVVNNTLEGN (63)
4S PEPTIDE
ASA -LVPSSRGGLPLEEVTVAEVLAARGYLTGMAGKWHLGVGPEG----AFLP-PHQGFHRFLGIPYSHDQGP 155
STS FLFTASSGGLPTDEITFAKLLKDQGYSTALIGKWHLGMSCHSKTDFCHHP-LHHGFNYFYGIS------- 163
HpuARS -FLPWTKTGLPKSELTIAEAMKEAGYATGMVGKWHLGMNENSSTDGAHLPFNH--GFDFVGHNLPFTNSW 191
Gln6S -CSSKSWQKIQEPNTFPAILRSMCGYQTFFAGKYLNEYGAPDAGGLEHVPLGWSYWYALEK--------- (123)
4S PEPTIDE
ASA CQNLTCFPPATPCDGGC--------------------------DQGLVPIPL----------------- 181
STS ---LTNLRDCKPGEGSVFTTGFKRLVFLPLQIVGVTLLTLAALNCLGLLHVPLGVFFSLLFLAALILTLF 230
HpuARS SCDDTGLHKDFPDSQRCY-- 209
Gln6S ----------------------------------NSKYYNYTLSING-KARKHGENYSVDYLTDVLANVS (158)
4S PEPTIDE ASB III 47:G TI EFIHYFGYLLGSEDYY-SH
ASA ----------------L-ANLSVEAQPPWLPGLEARYMAFAHDLMADAQRQDRPFFLYYASHHTHY-PQ 232
STS LGFLHYFRPLNC-FMM---RNYEIIQQPMSYDNLTQRLTVEA--AQFIQRNTETPFLLVLSYLHVHT-AL 293
HpuARS ----------------LYVNATLVSQPYQHKGLTQLFTDDALGFIED-NHAD-PFFLYVAFAHMHT-SL 259
Gln6S LDFLDYKSNFEPFFMMIATPAPHS---PWTAAPQYQKAFQNVFAPRNKNFNIH-GTNKHWLIRQAKTPMT (224)
4S PEPTIDE
ASA FSGQSFAERSGRGPFGDSLMELDAAVGTLMTAIGDLGLLEETLVIFTADNGPETMRMS-----RGGCSGL 297
STS FSSKDFAGKSQHGVYGDAVEEMDWSVGQILNLLDELRLANDTLIYFTSDQGAHVEEVSSKGEIHGGSNGI 363
HpuARS FSSDDFSCTSRRGRYGDNLLEMHDAVQKIVDKLEENNISENTIIFFISDHGPH--R---EYCEEGGDASI 324
Gln6S NSSIQFLDNAFRKRWQTLLSVDDL-VEKLVKRLEFTGELNNTYIFYTSDNGYHTGQFSL---------- (282)
4S PEPTIDE WSLWEGGVR:ASB IV 27
ASA LRCGKGTTYEGGVREPALAFWPGHI-APGVTHELASSLDLLPTLAALAGAPLP-NVTLDGFDLRPPAAGH 365
STS YKGGKANNWEGGIRVPGILRWPRVIQAGQKIDEPTSNMDIFPTVAKLAGAPLPEDRIIDGRDLMPLLEGK 433
HpuARS FRGGKSHSWEGGHRIPYIVYWPGTI-SPGISNEIVTSMDIIATAADLGGTTLPTDRIYDGKSIKDVLLEG 393
Gln6S -PIDKRQLYEFDIKVPLLVRGPGIK-PNQTSKMLVANIDLGPTILDIAGYDLNKTQM-DGMSLLPILRGA (349)
4S PEPTIDE 8 kDa C-ter:DPPTKTLW
ASA RQEPSAVSLL------LPVLPRRGPWGFCCADWKVQ--GSLLHPGSAHSDTTADPACHASSSLTAHEPPL 427
STS SQRSDHEFLFHYCNAYLNAVRWHPQNNTSIWKAFFF--TPNFNPVVPTDCFATHVCFCFGSYVTHHDPPL 501
HpuARS SASPHSSFFYYCKDNLMAVRVGKYKAHFRTQRVRSQDEYGLECAGGFPLEDYFDCNDCEGDCVTEHDPPL 463
Gln6S SNLTWRSDVLVEYQGEGRNVTDPTCPSLSPGVSQCFP-----DCVCEDAYNNTYACVRTMSALWNLQYCE (414)
4S PEPTIDE LFDIDRDPEERHD LSREYPHIVTK ASB IV 34:TGLQHQIIWPCQPSC
ASA LYDLSKDPGENYNLLGGVAGATPEVLQALKQLQLLKAQLDAAVTFGPSQVARGE---DPA-LQICCHPGC 493
STS LFDISKDPRE-RNPL-TPA-SEPRFYEILKVMQEAADRHTQTLPEVPDQFSWNNFLWKPW-LQLCCPSTG 567
HpuARS LFDLHRDPGEAY-PL-EACGHEDVFLTVKSTVEEHKAALVKGTPLLDSFDHSIVPCCNPANGCICNYVHE 531
Gln6S FDDQEVFV-EVYN-LTADPDQITNIAKTIDP--ELLGKMNYRLMMLQ--SCSGPTCRTP-GVFDPGYRFD (477)
4S PEPTIDE VPLDEK
ASA TPRPACCHCPDPHA 507
STS LSCQCDREKQDKRLSR 583
HpuARS PGMPECYQDQVATAARHYRP 551
Gln6S PRLMFSNRGSVRTRRFSKHLL (498)

promoter sequences, and the presence of consensus SP1 binding sites, are typical of the promoter regions of `house-keeping' genes (Kadonaga *et al.,* 1986). House-keeping genes are expressed at a low, constitutive level in a wide number of different tissues. The proposed house-keeping role of 4S is supported by the low level of expression observed in all tissues examined so far.

The 4S promoter region shares these sequence features with the promoters of some other lysosomal enzyme genes, including lysosomal acid phosphatase (Geier *et al.,* 1989), which has two potential SP1 binding sites, and arylsulphatase A (Kreysing *et al.,* 1990) which has four potential SP1 binding sites. In addition, both promoter regions had a high G+C content, and TATA- or CCAAT-elements were absent from the typical positions. However, these features are not rigidly conserved in the promoters of all lysosomal genes, as for example, the β-subunit of β-hexosaminidase has a CCAAT-element (Neote *et al.,* 1988), the α-L-fucosidase has two CCAAT-elements (Kretz *et al.,* 1992) and the α-subunit of β-hexosaminidase has a TATA- and a CCAAT-element (Proia & Soravia, 1987), beside the other sequence features of house-keeping genes.

Further analysis of the 4S putative promoter region was conducted in order to identify the CAP signal/s, which mark the start of the mRNA. The CAP sequence has been described as the nucleotide sequence CA followed by a number of pyrimidines, however the consensus is very loose (Bucher, 1990). CA di-nucleotides are underlined in Figure 3.8, and two fair-scoring potential CAP sites are also shown (at nucleotide positions 389 and 422). However, the CAP consensus sequence has little diagnostic value when other promoter elements such as TATA are absent, which is the case with the 4S promoter region (Bucher, 1990). Primer extension and S1 nuclease protection studies would be required to determine which of these if any, is the true CAP site. The actual CAP sites utilized will be discussed below.

Recently, Modaressi *et al.* (1993) published the isolation of genomic clones containing all the exons of 4S. They found potential binding sites for a number of transcription factors in a 1.6-kb sequence 5'-adjacent to the coding region, and determined the CAP sites utilised in cultured fibroblasts. In particular, they found consensus binding sites for a number of transcription factors in the section of the 1.6-kb sequence that overlaps with the 5'-flanking sequence of the 1.3-kb *Hin*dIII clone, as well as sites 5'-upstream of the *Hin*dIII clone. Consensus binding sites were reported for transcription factor AP-2 at nucleotide positions 234, 313, 332, 355, 373, and 435, for PEA3 at nucleotide 139, for Pu.1 at 382, for NF-S at 199, as well as an additional SP1 consensus (but non-canonical) binding site centred on nucleotide position 411 (Fig. 3.8) (a recent compilation of transcription factor binding sites is found in Faisst & Meyer (1992)).

Modaressi *et al.* observed multiple CAP sites in the 4S mRNA from cultured fibroblasts. They identified the major CAP site, which produced a 135-bp 5'-untranslated leader, and other CAP sites which produced 5'-untranslated leaders from 1 to 97-bp in length (Fig. 3.8). Surprisingly, the actual CAP sites that were used scored poorly when compared to the consensus CAP sequence (Bucher, 1990), and conversely, the high scoring predicted sites were not observed *in vivo*. However, as mentioned previously, the CAP consensus sequence has little diagnostic value when other promoter elements such as the TATA-box are absent, as is the case with the 4S promoter region. Perhaps some of the predicted high scoring sites may be utilized in other tissues, despite not being detected in fibroblasts. The potential binding sites for AP-2, SP1, and Pu.1 are clustered in the vicinity of the observed CAP sites, yet the functional significance of the putative transcription factor binding sites is not known.

Schuchman *et al.* (1990) isolated a full-length cDNA clone for 4S from a human testis library which contained a 5'-untranslated leader of 559-bp. Their cDNA

clone must have been derived from a transcript which initiated upstream of the 5'-boundary of the 1.3-kb *Hin*dIII fragment and the CAP sites found by Modaressi *et al.* (1993), but within the 1.6-kb sequence of Modaressi *et al.* Primer extension and S1 nuclease protection studies by Modaressi *et al.* were unable to detect transcription start sites in fibroblast mRNA corresponding to the vicinity of the start of the cDNA clone of Schuchman *et al.* Furthermore, a 1.8-kb *Eco*RI-*Hin*dIII fragment (which spanned the 5'-start of the Schuchman *et al.* clone) failed to demonstrate promoter activity in a CAT (chloramphenicol acetyl transferase) assay of promoter activity. Modaressi *et al.* resolved the apparent discrepancy by the suggestion that the Schuchman *et al.* cDNA clone was derived from a cryptic promoter, whose activity was below the level of detection of their primer extension and S1 nuclease protection assays of 4S mRNA in cultured fibroblasts. A number of differences exist between the 4S sequences of the 1.3-kb *Hin*dIII fragment, Peters *et al.* (1990), Schuchman *et al.* (1990), and Modaressi *et al.*, (1993); these will be discussed in Section 3.8.2.

Northern blots of poly $(A)^+$ RNA isolated from fibroblasts indicated three 4S mRNA species (Peters *et al.*, 1990); a 4.8, 2.5, and 1.8-kb species of major, significant, and minor relative abundance respectively. The cDNA clone of Peters *et al.* (1990) appears to correspond to the 2.5-kb species, as the length of sequence between the major CAP site in fibroblasts (135-bp 5'-adjacent to the Methionine initiation codon) and the poly (A) addition site of the cDNA clone is 2.36-kb, which is sufficient, after the addition of a poly (A) tail, to account for the 2.5-kb species. The cDNA clone of Schuchman *et al.* (1990) is too large (over 2.8-kb) to account for the presence of the 2.5-kb species; it may be contained within the 4.8-kb species or alternatively, it may have originated from a cryptic promoter, as suggested by Modaressi *et al.* (1993) (see above).

The 4.8-kb mRNA species of 4S may be due to the use of a different polyadenylation signal compared to the 2.5-kb species, or it be due to alternative splicing events in the 3'-untranslated region of the primary transcript. Heterogeneity in the length of mRNA species has been observed in other lysosomal sulphatases, for example ASA mRNA in fibroblasts is composed of three species of length 2.1-kb, 3.7-kb and 4.8-kb (Kreysing *et al.*, 1990). The three species differed in the 3'-untranslated sequences and it was considered that the use of different polyadenylation signals was responsible.

There are a number of possible explanations for the observed 1.8-kb minor 4S mRNA species. The 1.8-kb mRNA could either be a degradation product, a cross-reacting species, or it may represent an alternatively spliced mRNA that may or may not produce a different functional protein. It is not known whether it codes for an active 4S, however based on the size of the 4S protein (533 amino acids), the absolute minimum length for a functional 4S mRNA is of the order of 1.6-kb.

In summary, the observation that the 1.3-kb *Hin*dIII fragment encoded an open reading frame whose sequence was collinear with the 4S peptide data, and which demonstrated sequence homology with other sulphatases, supported the hypothesis that it encoded a portion of 4S. The hypothesis was further strengthened when the chromosomal location of the 1.3-kb *Hin*dIII fragment was obtained by *in situ* hybridization, as discussed below.

3.6 CHROMOSOMAL LOCALIZATION OF 4S TO THE *ARSB* LOCUS

In situ hybridizations using the 1.3-kb HindIII clone were performed by Elizabeth Baker of the Department of Cytogenetics and Molecular Genetics, The Women's and Children's Hospital, Adelaide.

This Section describes the chromosomal localization of the 4S gene to chromosome 5q13-5q14, by *in situ* hybridization of the ^{3}H-labelled 1.3-kb *Hin*dIII genomic fragment to human metaphase chromosomes. This location was consistent with and refined previous chromosomal assignments of the *ARSB* locus based on the expression of human arylsulphatase B in somatic cell hybrids.

The chromosomal location of the 4S gene was defined by *in situ* hybridization to metaphase chromosomes. The distribution of grains on 50 metaphase chromosomes from a normal male is given in Figure 3.10a. Of the total of 172 grains, 30 (17%) were localized over the long arm of chromosome 5. The distribution of grains from 22 of these metaphases, which showed silver grains on the long arm of chromosome 5, together with an additional 28 metaphases with labelling in this region, are depicted in Figure 3.10b and localize the 4S gene probe to 5q13 to 5q14. *In situ* hybridization of this probe to a metaphase spread from another individual showed an identical localization (data not shov As specific-labelling was only observed at bands 5q13-5q14 (Fig. 3.10a), *in situ* hybridization also established that there were no sequences homologous to the 1.3-kb *Hin*dIII genomic clone elsewhere in the human genome.

The localization of 4S was further investigated by *in situ* hybridization to metaphase chromosomes from a patient with a translocation between chromosomes 5 and 16, with the breakpoint on chromosome 5 in the vicinity of the 4S localization (patient karyotype 46,XY,t(5;16)(q14.2;q22.1)). The distribution of grains on 55

metaphases with this translocation are shown in Figure 3.10c. The distribution of grains on the long arms of normal and derived chromosome 16 was consistent with the background grain distribution, whereas there was an excess of grains on the long arms of normal and derived chromosome 5. This result shows that 4S is located proximal to 5q14.2, the breakpoint of the translocation. The *in situ* data suggest the most likely localization is from 5q13.3 to 5q14.1.

O'Brien *et al.* (1974) and Matalon *et al.* (1974) demonstrated that arylsulphatase B and *N*-acetylgalactosamine-4-sulphatase are the same enzyme (Chapter 1, Section 1.6). The *ARSB* locus was localized to chromosome 5 (Hellkuhl & Grzeschik, 1978; DeLuca *et al.,* 1979) based on the analysis of arylsulphatase B expression in human-rodent somatic cell hybrids. Subsequently, the localization of *ARSB* was refined to the region 5p11-5q13, again using somatic cell hybrids (Fox *et al.,* 1984; Fidzianska *et al.,* 1984). The co-localization of the 1.3-kb *Hin*dIII genomic fragment to the *ARSB* locus is positive evidence that the genomic clone encodes 4S.

The *in situ* localization of the 4S gene has refined the location of the *ARSB* locus. The *HEXB* locus, encoding the β-chain of β-hexosaminidase, has been localized to 5q13 using somatic cell hybrids (Dana & Wasmuth 1982); translocation studies indicate that *HEXB* is distal to *ARSB* (Fox *et al.,* 1984). Taken together with the *in situ* data, this suggests that *ARSB* most probably lies in the distal portion of the 5q13 band at 5q13.3.

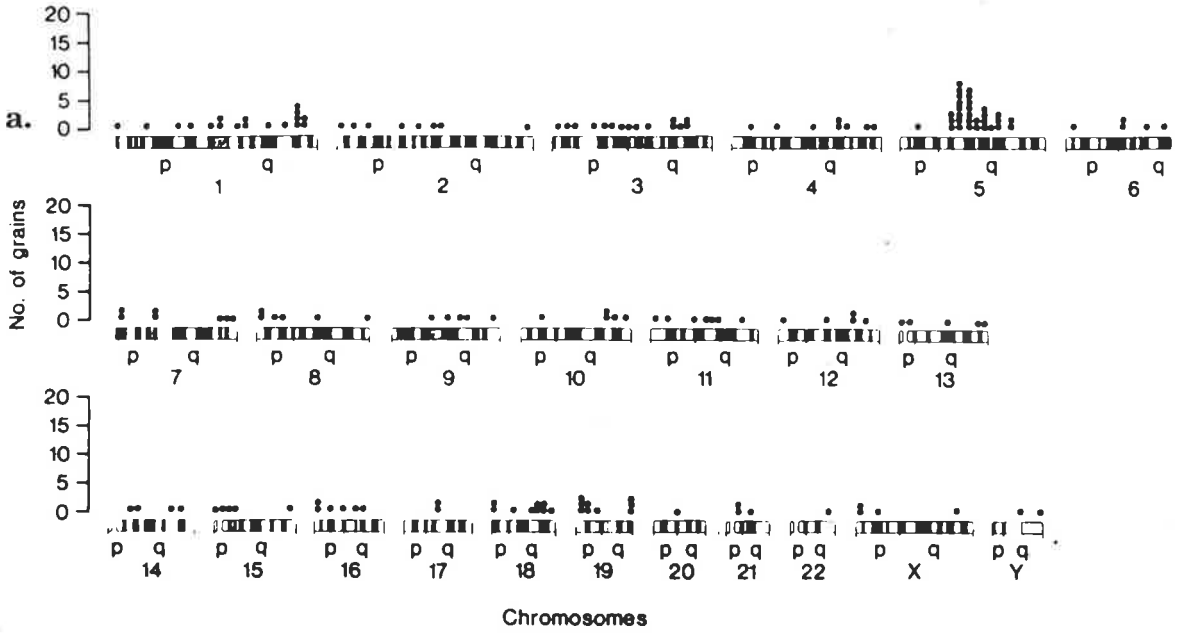
The *in situ* hybridization results were positive evidence supporting the hypothesis that the 1.3-kb *Hin*dIII fragment encoded for 4S. However, the genomic clones had the disadvantage that they did not encode the entire 4S gene. In order to establish the authenticity of the genomic clones the complete 4S protein sequence was required. The next Section will present the attempts to isolate a clone containing the complete coding sequence of 4S.

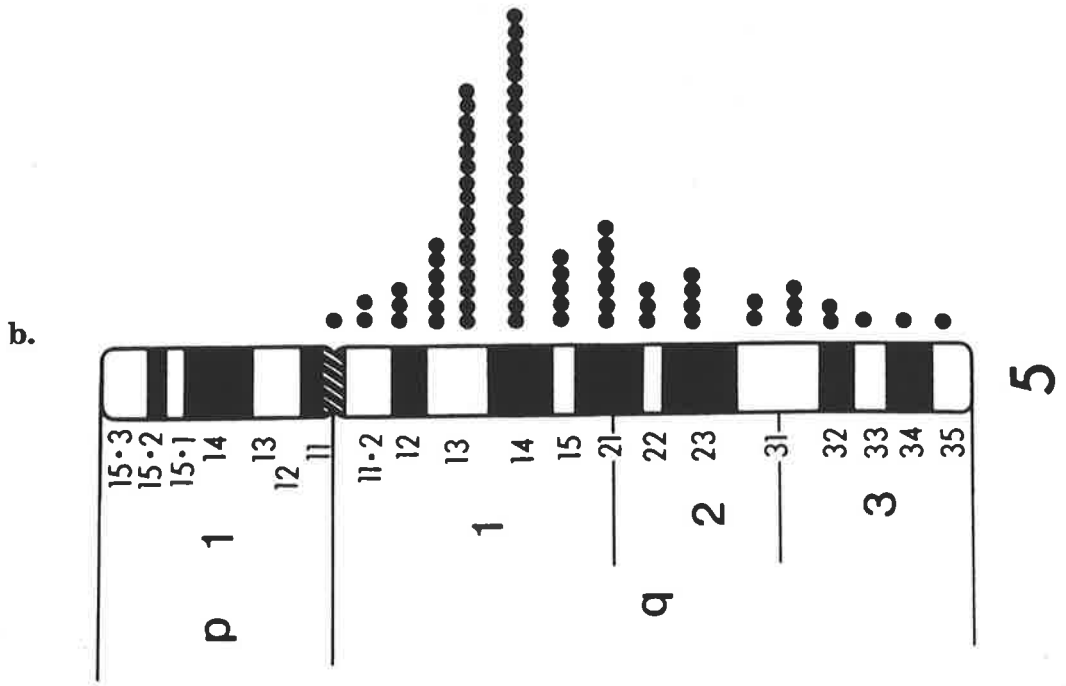
FIGURE 3.10 *In Situ* HYBRIDIZATION OF THE 1.3-kb *Hin*dIII FRAGMENT TO HUMAN METAPHASE CHROMOSOMES

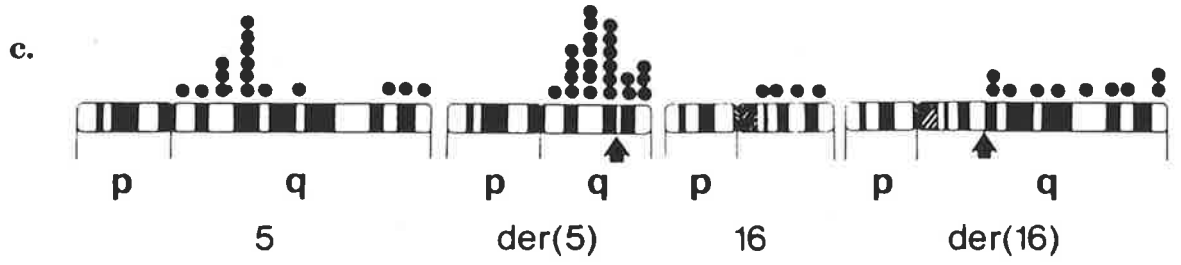
a. Ideograms of G-banded human chromosomes showing the distribution of 172 silver grains from 50 metaphases after hybridization of the ^{3}H-labelled 4S genomic probe (1.3-kb *Hin*dIII fragment) to individual human metaphase chromosomes. Note that chromosome 5q contained 17% of the total grains.

b. Ideogram of G-banded chromosome 5 illustrating distribution of silver grains on 50 metaphases after hybridization of the ^{3}H-labelled 4S genomic probe (1.3-kb *Hin*dIII fragment). Note that the distribution of grains peaks at 5q13 to 5q14.

c. Ideograms of G-banded chromosomes from a patient with the karyotype 46,XY,t(5;16)(q14.2;q22.1) showing the distribution of silver grains from 55 metaphases after hybridization of the ^{3}H-labelled 4S genomic probe (1.3-kb *Hin*dIII fragment). Note that the 4S gene is proximal to the 5q14.2 breakpoint (denoted by the arrow); der = derived.







TR2 t(5;16)(q14.2;q22.1)

3.7 SCREENING cDNA LIBRARIES FOR A 4S cDNA CLONE

Kerri Beckmann of the Department of Chemical Pathology, The Women's and Children's Hospital, Adelaide, assisted in the screening of the cDNA libraries with the 4S genomic clone.

3.7.1 CONVENTIONAL cDNA LIBRARY SCREENING

In order to isolate a cDNA clone containing the complete coding sequence of 4S, the 1.3-kb *Hin*dIII genomic fragment derived from λ113 and λ122 was used as a probe to screen cDNA libraries. It was thought that the genomic fragment would be more likely to detect a 4S clone than the previous oligonucleotide probes, as the longer length of the genomic clone (312-bp of coding sequence) would allow increased specificity and sensitivity, compared to the short (14-26-bp) and degenerate oligonucleotide probes. The 1.3-kb *Hin*dIII genomic fragment was oligolabelled using α-^{32}P-dCTP according to the procedure outlined in Section 2.2.11b, and the efficiency of the labelling was determined semi-quantitatively by fractionation of the reaction on a Sephadex G-50 column as described in Section 2.2.11c.

A λgt10 colon random-primed cDNA library (Clontech, Palo Alto, USA) was selected for screening with the 1.3-kb *Hin*dIII fragment, as a 1.5-kb cDNA clone for iduronate-2-sulphatase, another low abundance lysosomal enzyme, was previously isolated from this library (Wilson *et al.*, 1990). Approximately 1.4 x 10^6 recombinants were plated out on the bacterial hosts LE392 or C600 and screened at a density of 20,000 to 80,000 pfu per plate. The plaques were transferred in duplicate to nylon membranes and hybridized overnight to the oligolabelled 1.3-kb *Hin*dIII probe. The hybridized membranes were washed sequentially in 6 x SSC at room temperature and in 2 x SSC / 1% (w/v) SDS at 50°C, as described in Section 2.2.13b. Washing was not performed with the TMACl solution (as for the

oligonucleotide probes) because the estimated T_d of the 1.3-kb *Hin*dIII probe (about 95°C) was too high for safe and convenient manipulation (Wood *et al.*, 1985). No positive clones were identified after autoradiography. The lack of success in screening for a 4S cDNA clone in the colon cDNA library was considered to be due to either the presumed very low abundance of 4S clones in the colon cDNA library, or alternatively, the probe was not sufficiently sensitive to detect a 4S clone at the plaque density used during library screening. Consequently, an internal positive control consisting of the 4S genomic clone λ122 `seeded' onto one of the library plates, was included in subsequent library screens.

The following cDNA libraries were screened in succession with the 1.3-kb *Hin*dIII probe; a λgt11 lung fibroblast cDNA library, which had been screened previously with oligonucleotide probes (Section 3.4.2), a λgt11 placenta cDNA library, and a λgt11 testis cDNA library (all obtained from Clontech, Palo Alto, USA). The λ122 4S genomic clone was seeded onto one of the lung fibroblast and placenta cDNA library screen plates, as a positive control. Hybridization and membrane washing were as described above. A total of 2.7×10^5 recombinants of the λgt11 lung fibroblast cDNA library were screened, at a density of 30,000 pfu per plate; a total of 9.3×10^5 recombinants of the λgt11 placenta cDNA library were screened, at a density of 93,000 pfu per plate; and a total of 2.0×10^6 recombinants of the λgt11 testis cDNA library were screened, at a density of 100,000 pfu per plate. In summary, no positive cDNA clones were obtained.

It was considered that the failure to isolate cDNA clones with the 1.3-kb *Hin*dIII fragment may have been due to the very low predicted abundance of 4S mRNA in tissues generally and those from which the cDNA libraries were derived. The 1.3-kb *Hin*dIII probe was considered sufficiently sensitive to identify 4S clones at the plaque density used, as the autoradiograph signals obtained on the positive control plates containing the λ122 genomic clone were strong, despite the cDNA library

plaque density, which was between 30,000 and 93,000 per plate. In hindsight, it was thought that insufficient clones were screened. This rationalization was supported anecdotally by the experience of Peters *et al.* (1990), who initially isolated a partial 4S cDNA clone that lacked the 5'-coding end, after screening 1.2 x 10^6 clones of the same testis cDNA library that we had screened unsuccessfully. The missing 13 amino acid codons (38-bp) were later obtained from a genomic clone (Peters *et al.*, 1990). Schuchman *et al.* (1990) isolated a full-length 4S cDNA clone also from the same testis library, using a partial clone isolated from a hepatoma cDNA library as a probe, however they did not report the number of clones screened.

3.7.2 PCR-BASED cDNA LIBRARY SCREENING

Prior to the publication of the 4S cDNA clones in 1990, the failure to isolate a 4S cDNA clone by conventional library screening prompted us to consider a different approach. The fresh approach sought to `amplify' the 4S cDNA from cDNA libraries with oligonucleotide primers, utilizing the (then) newly emergent technique of thermostable-polymerase chain reaction (PCR) (Saiki *et al.*, 1988). The PCR primers were designed based on the known amino acid sequences of 4S peptides. The observation of conserved sequence features between sulphatases enhanced the potential utility of the approach, allowing the locations of the 4S peptide portions within the 4S protein to be estimated, and hence exploited in optimizing the PCR amplification and cloning of the 4S gene.

Four PCR primers derived from selected 4S peptide assignments were designed to enable the PCR amplification of the 4S cDNA in two portions of roughly equal size, on the assumption that the length of the 4S coding region was comparable to that of the other sulphatases (Fig. 3.11). PCR primer 4SP1 was derived from the 5'-end of the coding region of the 4S genomic clone; PCR primers 4SP2 and 3 were derived from the ASB III 47 tryptic peptide, which was tentatively mapped to about

halfway along the presumed length of the 4S mRNA coding region; and PCR primer 4SP4 was derived from the 8-kDa amino-terminus, which was mapped to the 3'-end of the coding region (Fig. 3.11). The tryptic peptide ASB IV 27 was confidently assigned in the sulphatase alignment, however it was not considered suitable for PCR primer design as it was rich in degenerate codons.

This PCR-based approach to recover the 4S cDNA from cDNA libraries using primers derived from the amino acid sequences of 4S peptides, was based on a combination of various published techniques (Lee *et al.,* 1988; Frohman *et al.,* 1988; Friedman *et al.,* 1988). The MOPAC (mixed oligonucleotide primed amplification of cDNA) technique originally described by Lee *et al.* (1988) used two primers derived from the same peptide sequence to PCR-amplify the DNA sequence between them. A limitation of the technique was that the length of the PCR product was restricted by the length of the known peptide sequence. This restriction was by-passed by the RACE technique (Frohman *et al.,* 1988) and by a technique of Friedman *et al.* (1988), enabling the entire coding region of the cDNA of interest, including the extreme 5' and 3' ends, to be cloned by PCR. The RACE (rapid amplification of cDNA ends) technique utilised primers derived from known peptide sequence and adaptor primers designed to prime from the 5'- and 3'-termini of the mRNA, to synthesize the first-strand cDNA and PCR-amplify the cDNA of interest. The approach of Friedman *et al.* (1988) obviated the need to synthesize the cDNA, as the template for the PCR amplification was an aliquot of the cDNA library. This technique used PCR primers designed from the cloning site sequences of the λ vectors in which the cDNA libraries were packaged. PCR amplification was performed between the vector primers and peptide-derived primers to obtain products spanning the entire coding region of the cDNA of interest. The use of aliquots of cDNA libraries as the template for PCR amplification has the advantages of chemical stability, convenience and the capability to be regenerated readily, compared to cDNA synthesized *de novo* from cellular RNA (as in the

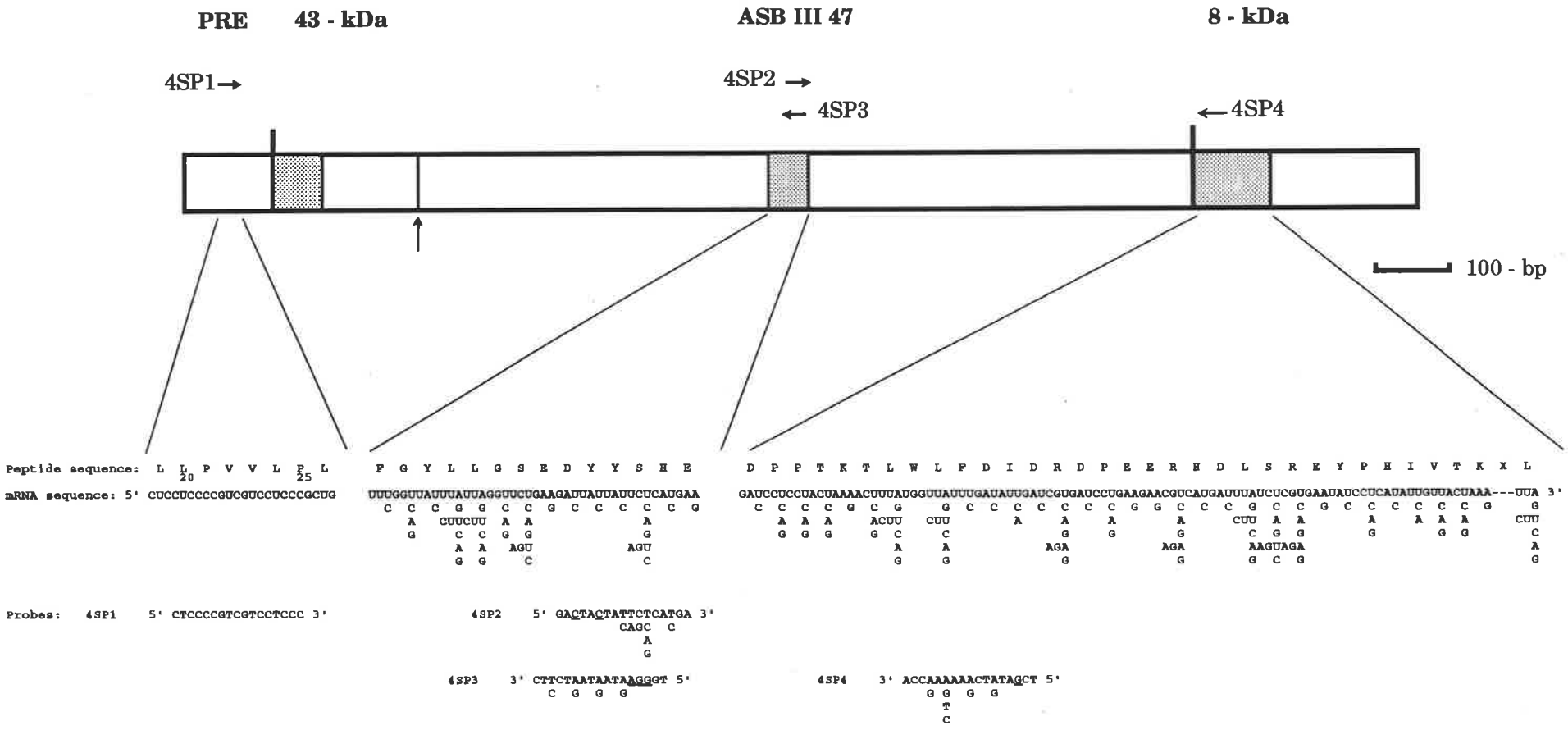
FIGURE 3.11 DESIGN OF PRIMERS FOR PCR AMPLIFICATION OF THE 4S cDNA

Diagrammatic representation of the coding sequence of a hypothetical cDNA for 4S. The coding sequence is about 600 amino acids in length, based on the other sulphatases. The locations of the signal peptide (PRE) and the amino-terminal sequence of the 43-kDa species are indicated, together with the predicted locations of the ASB III 47 tryptic peptide and the 8-kDa species based on sulphatase homology (shaded boxes). The locations of the PCR primers (17-mers) 4SP1, 4SP2, 4SP3, and 4SP4, are shown together with their orientation; sense (right arrow) or antisense (left arrow). The 3'-extent of the coding sequence of the 1.3-kb *Hin*dIII fragment is also indicated (vertical arrow). In the lower half of the figure, the amino acid sequences of the selected peptides are shown, together with the conceptual mRNA coding sequence, and the derived probes.

4SP1 was designed based on the deduced coding sequence of the 1.3-kb *Hin*dIII genomic 4S clone, and was derived from the signal peptide sequence of 4S.

4SP2 and 4SP3 were designed based on the conceptual mRNA coding sequence of the tryptic peptide ASB III 47. The degeneracy of the primers was reduced by the choice of nucleotides (underlined) at selected positions of nucleotide degeneracy, based where possible on codon usage observed in the coding region of the 1.3-kb *Hin*d III genomic fragment and on known codon usage in human genes. Nucleotide choices were restricted to the 5'-end of primers; all nucleotides were incorporated at positions of degeneracy that were within the 3'-half of the primer, so as to maximize the efficiency of the extension reaction during the PCR cycle of amplification. 4SP2 was comprised of 64 species and incorporated two nucleotide choices, while 4SP3 was comprised of 16 species and incorporated three nucleotide choices.

4SP4 was designed based on the conceptual mRNA coding sequence of the amino-terminal of the 4S 8-kDa species. The degeneracy of the primer was reduced as described for 4SP2 and 4SP3. 4SP4 was comprised of 32 species and incorporated one nucleotide choice.


PRE
43 - kDa
ASB III 47
8 - kDa
4SP1
4SP2
4SP3
4SP4
100 - bp
Peptide sequence: L L P V V L P L F G Y L L G S E D Y Y S H E D P P T K T L W L F D I D R D P E E R H D L S R E Y P H I V T K X L
mRNA sequence: 5' CUCCUCCCCGUCGUCCUCCCGCUG UUUGGUUAUUUAUUAGGUUCUGAAGAUUAUUAUUCUCAUGAA GAUCCUCCUACUAAAACUUUAUGGUUAUUUGAUAUUGAUCGUGAUCCUGAAGAACGUCAUGAUUUAUCUCGUGAAUAUCCUCAUAUUGUUACUAAA---UUA 3'
Probes: 4SP1 5' CTCCCCGTCGTCCTCCC 3'
4SP2 5' GACTACTATTCTCATGA 3'
4SP3 3' CTTCTAATAATAGGGT 5'
4SP4 3' ACCAAAAAACTATAGCT 5'

original RACE protocol). In addition, this technique enables the screening of a larger number of clones and at a faster rate than conventional screening and plaque purification procedures.

In order to PCR amplify the entire 4S cDNA, it was envisioned that the 4S-specific PCR primers 4SP1, 4SP2, 4SP3, and 4SP4, would be used in combination with the PCR primers derived from the cDNA library λ vectors. This approach was about to be attempted when in early April 1989, Christoph Peters and Kurt von Figura reported a 4S cDNA clone isolated from a λgt11 human testis cDNA library by conventional screening. Their work was subsequently published (Peters *et al.*, 1990). Soon after, Schuchman *et al.* (1990) also published the sequence of a 4S cDNA clone.

As a matter of interest, an approach similar to that described above was successfully used to clone the murine homologues of human arylsulphatase A and B (Grompe *et al.*, 1992). The MOPAC primers were designed based on the known conservation of residues in human arylsulphatases.

3.8 THE 4S cDNA CLONE OF THE GÖTTINGEN GROUP

The principal motivation for our attempt to isolate a 4S cDNA clone was to enable MPS-VI patient mutations to be determined. The mutations, once identified, would enable the testing of the hypothesis that correlations may be achieved between a patient's clinical phenotype, biochemical phenotype and mutant genotype. The cloning of the 4S cDNA by Peters *et al.* (1990) and Schuchman *et al.* (1990) achieved this initial technical goal, albeit ahead of us, and so a collaboration with Kurt von Figura and Christoph Peters from the Göttingen group was established in order to characterize MPS-VI patient mutations. As part of the collaboration, the Göttingen group provided prior to publication the complete deduced coding

sequence of 4S. In addition, they provided their longest 4S cDNA clone, pASB2-5, a 2.3-kb 4S cDNA cloned in the *Eco*RI site of M13mp18, which was missing 13 amino acids (38-bp) at the 5'-end.

3.8.1 REFINEMENT OF THE 4S STRUCTURE MODEL

The full-length 4S cDNA sequence confirmed and refined the model of 4S subunit structure and processing originally suggested by the homologies observed between sulphatases as described in Section 3.5.4. The amino acid sequences obtained from the amino-termini of the 43-kDa and 8-kDa species (Table 3.1) were found to be essentially collinear with portions of the 4S cDNA sequence. Specifically, residues 1-22 of the 43-kDa species were collinear with the deduced residues 40-61 of the 4S cDNA, and residues 1-34 of the 8-kDa species were collinear with the deduced residues 466-499 (Peters *et al.*, 1990). The complete 4S cDNA encoded 533 amino acids.

Our original model of the relationship between the 43-kDa and 8-kDa species of 4S (Litjens *et al.*, 1991) proposed that proteolytic clipping of the 57-kDa polypeptide generated an amino-terminal 43-kDa species and a carboxyl-terminal 8-kDa species which were disulphide linked. This model was consistent with the then known subunit composition of 4S, the biosynthesis and maturation of 4S (Taylor *et al.*, 1990), and the cDNA and gene sequences. The model envisioned the synthesis of the 4S precursor on membrane bound ribosomes, followed by the translocation of the nascent 4S polypeptide into the lumen of the ER and the cleavage of the signal peptide. The resultant 55.8-kDa 4S species (calculated from the amino acid sequence) then underwent glycosylation, including the addition of *N*-linked oligosaccharides, and acquired the lysosomal targeting signal, mannose 6-phosphate. During transport through the ER and the Golgi apparatus, 4S attained a molecular mass of 66-kDa, but was subsequently processed to a 57-kDa species. The 57-kDa species was cleaved during transport to the lysosome into an amino-

terminal 43-kDa species and a carboxyl-terminal 8-kDa species, linked by a disulphide bond. The linkage between the 43 and 8-kDa polypeptides must be a single disulphide bond involving the sole cysteine present in the deduced 8-kDa polypeptide at residue 521 of the protein sequence (Fig. 3.12). The mannose 6-phosphate residue/s on 4S that are involved in its targeting to the lysosome were thought to reside on the 43-kDa subunit as the 8-kDa subunit had no potential *N*-glycosylation sites.

As mentioned previously, a number of lysosomal enzymes are subjected to proteolytic trimming at the amino- and/or the carboxyl-termini, subsequent to the removal of the signal peptide. Consequently, it was thought that the initial proteolytic cleavage site responsible for the generation of the 8-kDa species from the 57-kDa precursor was localized to at most 15 residues between amino acids 450 and 466 (Fig. 3.12). This site was bounded by the amino-terminal amino acid of the 8-kDa species (Table 3.1) (D466) on the carboxyl side and the carboxyl-terminal amino acid of the (then) nearest 4S tryptic peptide (Peters *et al.*, 1990) on the amino-terminal side (W450). Subsequent proteolytic processing was therefore thought to be constrained to a region of the 57-kDa species encompassing at most 15 residues. The carboxyl-terminus of the 8-kDa species was not proteolytically trimmed, as Peters *et al.* (1990) sequenced a tryptic peptide that extended to the last amino acid of the deduced cDNA sequence (Fig. 3.12).

Recently, 4S has been shown to also contain a 7-kDa species (Kobayashi *et al.*, 1992). The 7-kDa species was not previously observed by us or others as it required modified techniques to detect. The isolation of the 7-kDa species led to a slightly revised model, in which the 57-kDa polypeptide was proteolytically processed to an amino-terminal 43-kDa species, a central 7-kDa species and a carboxyl-terminal 8-kDa species (Fig. 3.12; Kobayashi *et al.*, 1992). The 7 and 8-kDa species were disulphide linked to the 43-kDa species; the 7-kDa species was

presumably linked through the single cysteine at amino acid 447. Analysis of the amino-termini sequences of the 7 and 8-kDa species relative to other peptide sequences (Fig. 3.12), suggested that the initial proteolytic cleavage responsible for the generation of the 8-kDa species was localized to a region of 11 residues, which was smaller than the earlier prediction. The site was bounded on the amino-terminal side by the carboxyl-terminal amino acid of the 7-kDa *N*-terminal sequence (P454) (Kobayashi *et al.,* 1992), and on the carboxyl-terminal side by the amino-terminal amino acid of the 8-kDa species (D466) (Table 3.1). Prior to the discovery of the 7-kDa species, the fact that the sum of the 43 and 8-kDa species (51-kDa) did not equal the molecular weight of the non-denatured species (57-kDa) was considered to be due to uncertainties in the precise determination of the molecular weights on SDS-PAGE.

3.8.2 DIFFERENCES BETWEEN THE PUBLISHED 4S SEQUENCES

A number of differences exist between the 4S sequences encoded in the 1.3-kb *Hin*dIII genomic fragment, the 4S cDNA of Peters *et al.* (1990), the 4S cDNA of Schuchman *et al.* (1990), and the genomic clones of Modaressi *et al.*, (1993). Comparison of the 1.3-kb *Hin*dIII fragment with the other sequences found only a single nucleotide difference (T_{138} [$], Fig. 3.8). T_{138} occurred in the 5'- non-coding region of the 4S gene, and was located in a homopolymeric stretch of 10 T nucleotides, while in the other published reports only 9 T nucleotides were identified. Although the sequence of this region of the 1.3-kb *Hin*dIII fragment was obtained from one strand only, we consider the sequence was clearly 10 T nucleotides. This albeit minor discrepancy awaits clarification.

More significant differences exist between the Schuchman *et al.* sequence on the one hand, and the Peters *et al.* and Modaressi *et al.* sequences on the other. Firstly, the first 21 nucleotides of the 5'-untranslated sequence of the Schuchman *et al.*

FIGURE 3.12 THE DEDUCED AMINO ACID SEQUENCE OF 4S

The deduced amino acid sequence of 4S is from Peters *et al.* (1990). The amino-terminal residues of the 43-kDa and 8-kDa species are indicated, the position and identity of peptides listed in Table 3.1 are double-underlined, and additional sequences described by Peters *et al.* (1990) and Schuchman *et al.* (1990) are single-underlined. Kobayashi *et al.* (1992) described a 7-kDa species of 4S in addition to the 43-kDa and 8-kDa species; the amino-terminal residues of the 7-kDa species are indicated (....). Potential *N*-glycosylation sites are marked (*), and the single cysteine residues present in the 7-kDa and 8-kDa species are denoted (+). Below the deduced amino acid sequence of 4S is a schematic representation of the mature 4S protein, showing the location of the 43-, 7-, and 8-kDa species, the disulphide bridges (S-S), and the potential *N*-linked oligosaccharides (triangles) (modified from Kobayashi *et al.*, 1992).

N-ter 43-kDa

MGPRGAASLPRGPGPRRLLLPVVLPLLLLLLLAPPGSGAGASRPPHLVFLLADDLGWNDV 60

ASB II 20
ASB II 61

GFHGSRIRTPHLDALAAGGVLLDNYYTQPLCTPSRSQLLTGRYQIRTGLQHQIIWPCQPS 120

ASB III 47

CVPLDEKLLPQLLKEAGYTTHMVGKWHLGMYRKECLPTRRGFDTYFGYLLGSEDYYSHER 180

CTLIDALNVTRCALDFRDGEEVATGYKNMYSTNIFTKRAIALITNHPPEKPLFLYLALQS 240
*

VHEPLQVPEEYLKPYDFIQDKNRHHYAGMVSLMDEAVGNVTAALKSSGLWNNTVFIFSTD 300
* *

ASB IV 27

NGGQTLAGGNNWPLRGRKWSLWEGGVRGVGFVASPLLKQKGVKNRELIHISDWLPTLVKL 360

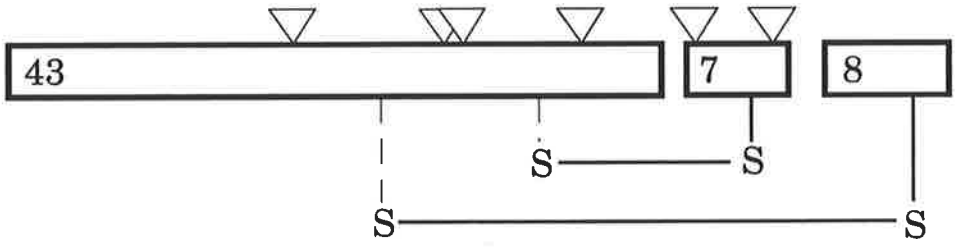
ARGHTNGTKPLDGFDVWKTISEGSPSPRIELLHNIDPNFVDSSPCPRNSMAPAKDDSSLP 420
*

N-ter 7-kDa

. . . . . N-ter 8-kDa

EYSAFNTSVHAAIRHGNWKLLTGYPGCGYWFPPPSQYNVSEIPSSDPPTKTLWLFDIDRD 480
* + *

PEERHDLSREYPHIVTKLLSRLQFYHKHSVPVYFPAQDPRCDPKATGVWGPWM 533
+



clone did not correspond with the equivalent region in the Modaressi *et al.* sequence. Secondly, in the 3'-untranslated region of the Schuchman *et al.* sequence, nucleotides 1794 to 1842 appear to be a complex duplication of the nucleotide sequence 1794 to 1819 in Peters *et al.* Thirdly, the nucleotides C_{2040} and G_{2174} in Peters *et al.* are not present in the Schuchman *et al.* sequence. Whether the above discrepancies are due to natural variation in the 4S gene, or are sequencing artefacts, was not resolved by the authors. Two nucleotide sequence differences were also found in the coding region; at nucleotide position 1126, a G in Peters *et al.*, while an A in Schuchman *et al.*, and at position 1191, an A and G respectively. The $A_{1126}G$ nucleotide substitution resulted in a V376M amino acid change, while the $A_{1191}G$ nucleotide substitution was a silent change in the 3rd nucleotide of the Pro397 codon. Both these changes were later shown to be polymorphisms (see Chapter 4, Section 4.5.2).

3.9 SULPHATASE HOMOLOGY

3.9.1 THE SULPHATASE FAMILY

The sequence homology between 4S and other sulphatases has been published previously (Peters *et al.*, 1990; Schuchman *et al.*, 1990). However, since those reports the cDNA sequences of a number of other sulphatases have been isolated. Comparison of recently published sulphatase sequences with the earlier published data may refine the identification of conserved residues that may play crucial roles in the structure and/or catalytic function of sulphatases.

Consequently, the deduced protein coding sequences of 10 sulphatases from a number of different organisms and with diverse substrate specificities were compared. The sequences included three human arylsulphatases; arylsulphatase A (ASA, Stein *et al.*, 1989a), arylsulphatase B or 4S (ASB, Peters *et al.*, 1990; Schuchman *et al.*, 1990) and arylsulphatase C or steroid sulphatase (STS, Yen *et*

al., 1987; Stein *et al.,* 1989b), a sea urchin arylsulphatase from *Hemicentrotus pulcherrimus* (HpuARS, Sasaki *et al.,* 1988), two arylsulphatases of bacterial origin; *E. coli* (EcoARS, Murphy *et al.*, 1992) and *K. aerogenes* (KaeARS, de Hostos *et al.,* 1989), and a protozoan arylsulphatase from *C. reinhardtii* (CreARS, Murooka *et al.,* 1990). Three human non-arylsulphatases; *N*-acetylgalactosamine-6-sulphatase (Gal6S, Tomatsu *et al.*, 1991), glucosamine-6-sulphatase (Gln6S, Robertson *et al.,* 1992) and iduronate-2-sulphatase (Id2S, Wilson *et al.,* 1990), were also included.

A number of other sulphatase sequences were not included in the alignment. For example, sulphatase sequences from other species that were homologues of the human sulphatases listed above were not included, as they excessively biased preliminary alignments (e.g. feline 4S, Jackson *et al.,* 1992; murine Id2S, Daniele *et al.,* 1993; murine ASA and 4S, Grompe *et al.*, 1992). These homologous sequences showed respectively, about 91%, 85%, 95% and 85% protein sequence identity, relative to the human counterpart. Likewise another arylsulphatase from sea urchin (*Strongylocentrotus purpuratus*) was not included, as it was closely similar to the HpuARS sequence (Yang *et al.,* 1989). Conversely, an arylsulphatase from *Alteromonas carrageenovora* (Barbeyron, 1992) was not included in the alignment as it displayed no significant sequence similarity with the other sulphatases. Prior to the description and analysis of the multiple sequence alignment, the characteristics of each of the sulphatases will be briefly reviewed.

The selected sulphatases show great diversity in their cellular location, their regulation, and the sulphated substrates utilized *in vivo*. For example, ASA is a soluble lysosomal sulphatase involved in the hydrolysis of the glycolipid, cerebroside-3-sulphate (Stein *et al.,* 1989a). A deficiency of ASA leads to metachromatic leukodystrophy. ASB or 4S is also a soluble lysosomal enzyme, and hydrolyzes the sulphate-ester bond of the terminal *N*-acetylgalactosamine-4-

sulphate residues of dermatan sulphate and chondroitin-4-sulphate. As mentioned previously, a deficiency of 4S is responsible for MPS-VI. STS is an arylsulphatase that is apparently resident within the microsomal membrane, and desulphates a number of 3β-hydroxysteroid sulphates, which serve as precursors for oestrogens, androgens and cholesterol (Yen *et al.,* 1987; Stein *et al.,* 1989b). X-linked icthyosis results from STS deficiency.

Gal6S hydrolyzes the sulphate-ester bond of the terminal *N*-acetylgalactosamine-6-sulphate residues of the GAG chondroitin-6-sulphate, and the galactose-6-sulphate residues at the non-reducing end of the GAG keratan sulphate (Tomatsu *et al.*, 1991), and when deficient, leads to MPS-IVA. Gln6S cleaves the sulphate group from the terminal glucosamine-6-sulphate residues of the GAG substrates heparan sulphate and keratan sulphate (Robertson *et al.,* 1992). Gln6S is deficient in MPS-IIID. Id2S desulphates the iduronate-2-sulphate terminal residues of the heparan and dermatan sulphate substrates. A deficiency of Id2S results in MPS-II (Wilson *et al.*, 1990). Gal6S, Gln6S, and Id2S are all lysosomal enzymes, and are classified as non-arylsulphatases (see below).

The precise *in vivo* substrate of the HpuARS arylsulphatase from sea urchin is not known. However, in sea urchin embryos the enzyme is located both in the cytoplasm and in the extracellular matrix, and is thought to play a structural role during morphogenesis (Sasaki *et al.*, 1988; Yamada *et al.*, 1989). The CreARS arylsulphatase is located in the periplasm of *C. reinhardtii*, a soil dwelling, unicellular green alga. CreARS appears to be a scavenger enzyme that releases sulphate from external organic ester substrates for utilization by the host (de Hostos *et al.,* 1989). The KaeARS enzyme perhaps performs a similar role, as enzyme synthesis is regulated depending on the availability of free sulphate and other sulphate-containing compounds in the medium (Murooka *et al.*, 1990). The arylsulphatase-like enzyme *aslA* from *E.coli* (EcoARS) has recently been

serendipitously identified. The gene was isolated on the basis that it suppressed a mutation in the guanosine pentaphosphate phosphatase gene. Subsequent sequence analysis identified it as an arylsulphatase-like protein (Murphy *et al.*, 1992).

Despite the structural diversity of the sulphate substrates hydrolyzed *in vivo* by members of the sulphatase family, sulphatases can be classified into two broad groups; the aryl- and the non-arylsulphatases. The aryl- and non-arylsulphatases are distinguished *in vitro* according to their ability to actively hydrolyze an arylsulphate substrate, such as NCS or 4MUS. The arylsulphatases efficiently desulphate the 4MUS substrate whereas the non-arylsulphatases such as Gln6S, Gal6S, and Id2S, even though they are able to bind to this substrate, are unable to desulphate it efficiently (Robertson *et al.*, 1992; J. Bielicki, unpublished data). For example, Gln6S desulphates 4MUS with a catalytic efficiency about 800-fold lower relative to 4S (Robertson *et al.*, 1992). Analysis of multiple sequence alignments may enable the identification of residues that perform functions specific to either the arylsulphatases or the non-arylsulphatases. For example, the regions responsible for arylsulphatase specificity toward 4MUS may be strongly conserved among the arylsulphatases but not in the non-arylsulphatases.

3.9.2 MULTIPLE SEQUENCE ALIGNMENT

The final alignment was composed from the results of three multiple sequence alignment programs; CLUSTAL, CLUSTALV, and MACAW, which utilized different alignment algorithms. CLUSTAL (Higgins & Sharp, 1988; 1989), and CLUSTALV (Higgins *et al.*, 1991) are automatic multiple sequence alignment programs, while MACAW (Schuler *et al.*, 1991) is an interactive program that allows the user to edit the results of the similarity search. A number of alignment programs were used, as preliminary analysis indicated that no single program was able to detect all of the optimal local alignments. In summary, the final global

alignment was composed by comparing the individual alignments generated by the three programs, and then manually assembling a composite global alignment which maximized regions of local similarity. Of course, the actual biological significance of the alignments was unable to be tested directly, as none of the tertiary structures of the sulphatases is known at present.

The multiple sequence alignment for 10 sulphatases is presented in Figure 3.13. In general, the highest degree of homology is observed in the amino-terminal third of the sulphatase polypeptides. The alignments between the 10 sulphatases indicated that the arylsulphatases from human and sea urchin showed a greater degree of sequence similarity to each other than that displayed between all the sulphatases as a whole. The three other arylsulphatases were divergent from the core arylsulphatase group; EcoARS was closer to the core sequences than KaeARS or CreARS. Of the non-arylsulphatases, Gal6S was more similar to the core arylsulphatases than Id2S or Iln6S.

Blocks of conserved amino acid residues are seen in subsets of the sulphatases. For example, the CTPSRxxxxTGR sequence commencing at residue 91 (numbered according to ASB) is totally conserved between the four core arylsulphatases (ASA, ASB, STS and HpuARS). The sequence GKWHLG commencing at residue 144 is also totally conserved in the core arylsulphatases, and is also found in KaeARS and the non-arylsulphatase Gal6S. Another sequence EGGxR (commencing at residue 323), is totally conserved in the core arylsulphatases, EcoARS, and the non-arylsulphatase Gal6S. Examination of these and other semi-conserved blocks has failed to identify residues that are clearly exclusive to either the arylsulphatase or non-arylsulphatase group.

Residues totally conserved in all 10 sulphatases may be crucial for catalytic function, on the assumption that the catalytic mechanism, and the residues

involved in catalysis, are conserved across the sulphatases. 15 residues were totally conserved in the multiple alignment of sulphatases presented in Figure 3.13; the amino acids were D53, D54, P93, R95, P231, F233, H242, D274, T293, D300, G302, D351, T355, G362, and G373 (numbered according to the ASB sequence). This is a considerable reduction in the total of conserved residues reported in earlier alignments, where the number of cloned arylsulphatase sequences was smaller (Peters *et al.*, 1990; Wilson *et al.*, 1990; Robertson *et al.*, 1992). The reduction in conserved residues has led to alterations in some of the earlier hypotheses. Peters *et al.* (1990) speculated that particular residues conserved in arylsulphatases may contain the catalytic residues required for arylsulphate-ester hydrolysis, namely the histidine residue in GKWHLG, and one or both of the arginine residues in CTPSRxxxxTGR. This hypothesis was supported indirectly by earlier studies reporting that one histidine residue (Lee & van Etten, 1975) and two or more arginine residues (James, 1979) were essential for the catalytic activity of ASA. However, the arginine in the TGR sequence and the histidine residue were not conserved when subsequent sulphatase sequences (both aryl and non-aryl) were added to the alignment (Fig. 3.13). These results cast doubt on their predicted essential role in arylsulphatase catalytic function. In contrast, the arginine in CTPSR is totally conserved in all sulphatases to date. The availability of additional sulphatase sequences has enabled refinement in the prediction of putative catalytic residues. One can speculate that the totally conserved arginine within the CTPSR at position 95, and the totally conserved histidine at position 242, may be directly involved in the sulphatase catalytic mechanism. Further speculation on the catalytic mechanism of sulphatases is presented in Chapter 5, Section 5.8.2.

Alternatively, conserved residues may have crucial roles apart from catalysis, perhaps in the processing and folding of the sulphatase. Multiple sulphatase deficiency in humans is thought to be due to the absence of a post-translational

FIG. 3.13 MULTIPLE SEQUENCE ALIGNMENT OF SULPHATASES

A multiple protein sequence alignment of the following sulphatases; human arylsulphatase A (ASA), human 4S or arylsulphatase B (ASB), human arylsulphatase C or steroid sulphatase (STS), sea urchin arylsulphatase (HpuARS, *Hemicentrotus pulcherrimus*), *E. coli* arylsulphatase-like enzyme (EcoARS), human *N*-acetylgalactosamine-6-sulphatase (Gal6S), human iduronate-2-sulphatase (Id2S), human glucosamine-6-sulphatase (Gln6S), *C. reinhardtii* arylsulphatase (CreARS), and *K. aerogenes* arylsulphatase (KaeARS). The alignment presented was manually assembled from the output of three multiple sequence alignment programs; CLUSTAL, CLUSTALV, and MACAW. The manual alignment was constructed so as to maximize regions of local similarity. The sequences are represented with the amino-terminal on the left and the carboxyl-terminal on the right. The 4S sequence is numbered according to Peters *et al.* (1990). Amino acid residues totally conserved in all 10 sulphatases are marked (■), residues conserved in 9 of 10 sequences are masked (●), and residues marked (▲) are conserved in 8 of 10 sulphatases. Residues conserved in at least 8 of the 10 sulphatases are in bold.


ASA
ASB
STS
HpuARS
EcoARS
Gal6S
Id2S
Gln6S
CreARS
KaeARS

process which is necessary for sulphatase activity (Chapter 5, Section 5.8.1). It is conceivable that a shared determinant on the sulphatases is recognized by the multiple sulphatase factor, perhaps one of the sequence blocks conserved among the human sulphatase sequences may function as the determinant.

Although the three human arylsulphatases display considerable homology at the amino acid level, the intron-exon boundaries and organisation of ASA (Kreysing *et al.*, 1990) ASB (Modaressi *et al.*, 1993) STS (Yen *et al.*, 1988) and also the non-arylsulphatase Id2S (Flomen *et al.*, 1993; Wilson *et al.*, 1993) display no similarity.

In conclusion, the observation of totally conserved residues in the sulphatases suggests that they have important roles to perform. However, as the tertiary structure is not known for any of the sulphatases, and as detailed kinetic and/or structural investigations of site-directed mutants have not yet been performed, predictions concerning specific roles are highly speculative.

CHAPTER 4

IDENTIFICATION OF 4S MUTATIONS IN MPS-VI PATIENTS

4.1 INTRODUCTION

As described in Chapter 1, mucopolysaccharidosis type VI (MPS-VI) or Maroteaux-Lamy syndrome is an autosomally inherited, recessive genetic disorder caused by a deficiency of the lysosomal enzyme *N*-acetylgalactosamine-4-sulphatase (4S). A wide variation in clinical severity is observed in MPS-VI patients, ranging in a spectrum from relatively mild to severe forms, based on the age of onset and the rate of progression of clinical features. The molecular basis for the clinical and biochemical variation is not understood, but is thought to be due to pair-wise combinations of multiple distinct 4S mutant alleles (Neufeld & Muenzer, 1989). In the context of the work presented in this thesis, a 4S mutation was defined as a base change in the gene that had a deleterious result on 4S function and led to clinical pathology, while a 4S polymorphism was defined as a base change that had no deleterious effect.

The cloning of the 4S cDNA (Peters *et al.*, 1990; Schuchman *et al.*, 1990) has enabled the identification and analysis of 4S mutations in MPS-VI patients. Mutational analysis was undertaken in an attempt to provide a molecular explanation for the clinical variation seen in MPS-VI, based on the hypothesis that in MPS-VI patients the clinical and biochemical phenotype can be correlated with the mutant 4S genotype. In addition, the analysis of patient mutations may provide an insight into the structure and catalytic mechanism of 4S, especially mutations that are located in regions that are conserved between sulphatases, and mutations that severely reduce enzyme activity.

MPS-VI patients were screened for 4S mutations in two stages; an initial low-resolution analysis consisting of the Southern blotting technique, followed by precise high-resolution analysis, consisting of the techniques of the PCR amplification of the 4S cDNA, chemical cleavage, direct DNA sequencing of PCR products, and allele specific oligonucleotide hybridization. Southern blot analysis

was performed on genomic DNA isolated from the patient's cultured skin fibroblasts. This technique, which can detect gross changes in the structure of the gene, had already identified rearrangements and deletions within the X-linked iduronate-2-sulphatase gene in MPS-II (Hunter syndrome) patients (Wilson *et al.*, 1990). Subsequent mutational analysis utilized patient RNA rather than genomic DNA, as the sequences of the 4S intron-exon junctions were not then known. The selected patients were analyzed by the technique of PCR amplification of the 4S cDNA (cDNA-PCR), which enabled gross changes in the structure and relative abundance of the 4S mRNA to be detected. Briefly, a complementary DNA (cDNA) copy of the total RNA isolated from the patient's cultured skin fibroblasts was synthesized. Selected portions of the 4S cDNA were then PCR amplified via 4S-specific DNA primers, and the size and yield of the products compared with a normal control. The cDNA-derived PCR products were subsequently used as substrates for the sensitive and high-resolution analysis techniques, in order to detect subtle changes such as small deletions, rearrangements and point mutations.

The techniques and methods available for mutation research has expanded at a rapid rate in recent years (reviewed by Cotton, 1989; 1993). Briefly, the methods available when the mutational analysis of MPS-VI patients was considered included denaturing gradient gel electrophoresis (reviewed in Myers *et al.*, 1989), RNAase cleavage methods (Myers *et al.*, 1988), chemical cleavage of mismatch (Cotton *et al.*, 1988) and single-strand conformation polymorphism analysis (Orita *et al.*, 1989). The incorporation of PCR technology (Saiki *et al.*, 1985) into these techniques has increased the efficiency and speed with which mutations are identified. A summary of the major advantages and disadvantages of these techniques is given in Table 4.1. The chemical cleavage technique was judged to be the best choice for the detection of point mutations, especially as it was claimed to be almost 100% successful in detection of mutations (Cotton, 1989), DNA substrate

could be readily generated by cDNA-PCR amplification, and the technique enabled the site of mismatch to be defined. Curiel *et al.* (1990) reported that the chemical cleavage technique successfully detected 14 out of 14 previously characterized single-base mutations in the p53 gene of lung cancer patients.

The chemical cleavage technique provides a powerful tool to detect mutations at the nucleotide level. This technique is based on the principle that mismatched or unmatched residues in a DNA duplex are more susceptible to chemical modification. A diagrammatic representation of the chemical cleavage technique is shown in Figure 4.1. The reaction involves the formation of a heteroduplex between normal and patient-derived PCR products, either or both of which are radiolabelled. Mutations or polymorphisms in the DNA will result in mismatching of the strands, which renders the mismatched bases more susceptible to chemical modification than the matched bases. Mismatched T and C residues are specifically reactive to osmium tetroxide and hydroxylamine respectively. Subsequent cleavage of the DNA strand at these 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 subsequently determined by DNA sequencing, as described below. To detect mismatched A or G residues, probes labelled on the opposite strand can be employed as the complementary mismatched bases will be T and C respectively.

The DNA sequence at the mismatched sites predicted by chemical cleavage analysis was obtained by the direct DNA sequencing of PCR products. The advantage of this approach compared to conventional DNA sequencing was that it obviated the cloning of the PCR products into a DNA sequencing vector. A further disadvantage with the sequencing of cloned PCR products is that multiple isolates

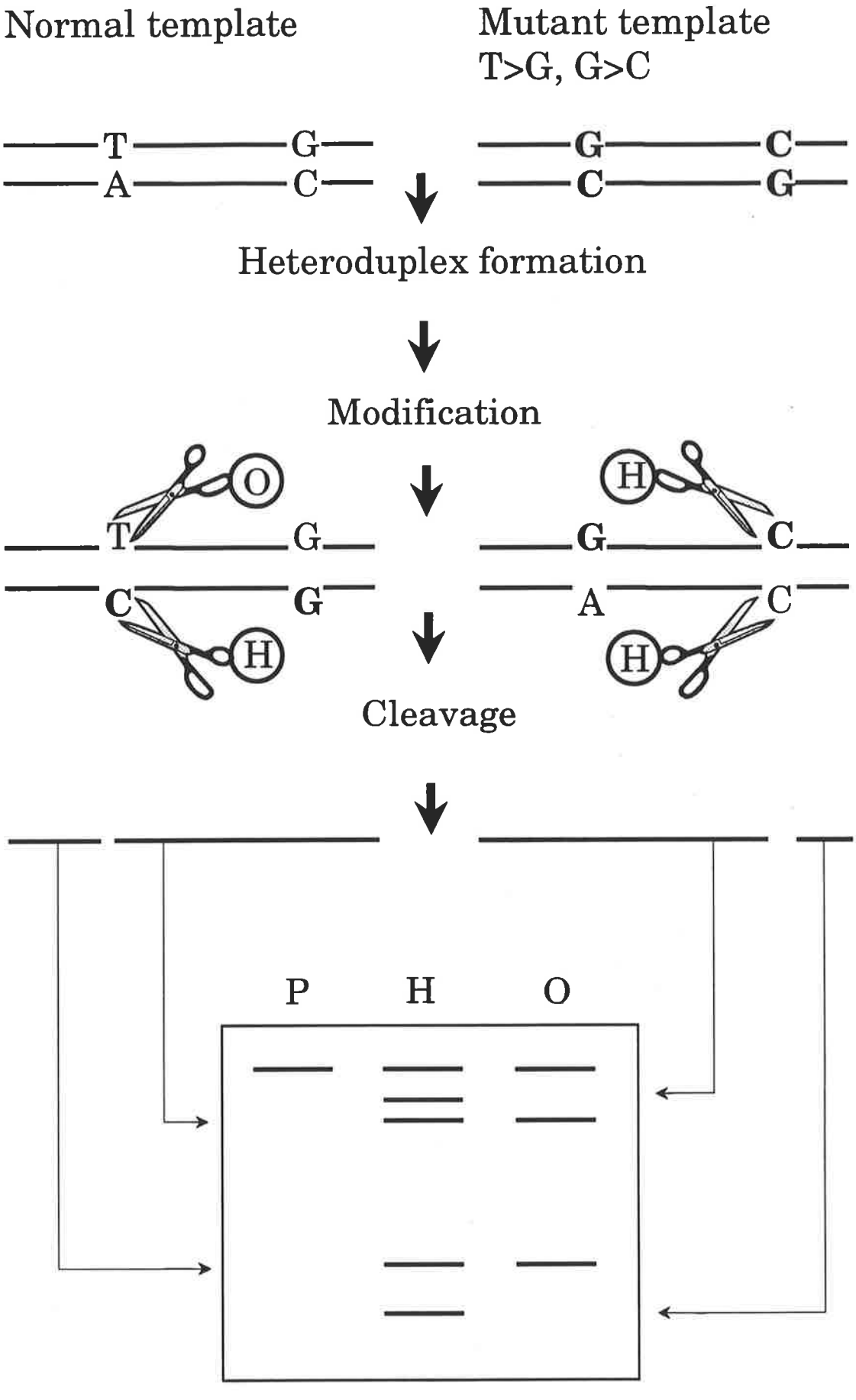
TABLE 4.1 COMPARISON OF MUTATION DETECTION TECHNIQUES

A list of the mutation screening methods considered for detection of MPS-VI mutations is shown with a brief summary of their main advantages and disadvantages (adapted from Cotton, 1989; Cotton, 1993).

Screening Method	Mutation position defined	% Mutations detected	Advantage(s)	Disadvantage(s)
Sequencing	yes	100	Position and change defined in one step.	Time consuming and technically demanding. Only limited coverage
Denaturing Gradient Gel Electrophoresis (DGGE)	no	50-100	Can isolate mutant band. Detects almost all mutations.	Requires special equipment. Needs GC clamp added to each duplex.
RNAase A	yes	~70	Simple. Position defined.	Partial detection. Need to prepare RNA probe.
Single Stranded Conformational Polymorphism (SSCP)	no	~80	Simple.	Limited coverage and partial detection
Chemical cleavage (modified protocol)	yes	~100	May detect and position all mutations.	Labour intensive. Toxic reagents.

FIGURE 4.1 PRINCIPLE OF THE CHEMICAL CLEAVAGE TECHNIQUE

Normal and mutant template DNA samples are heat denatured and annealed to give heteroduplex molecules. Mismatched T residues are modified by osmium tetroxide (O) and mismatched C residues are modified by hydroxylamine (H). The DNA strands at the modified sites are cleaved by piperidine, and the cleavage products are resolved by denaturing gel electrophoresis. The figure shows an uncleaved probe band in lane (P), and the cleavage products after reaction with hydroxylamine (H) and osmium tetroxide (O). The bands actually detectable by autoradiography depend on the manner of labelling the normal and the mutant templates.


Normal template
Mutant template
T>G, G>C
T
G
A
C
G
C
C
G
Heteroduplex formation
Modification
O
H
T
G
G
C
C
G
A
C
H
H
Cleavage
P
H
O

must be sequenced in order to distinguish true mutations from errors introduced by *Taq* DNA polymerase (Dunning *et al.*, 1988; Jansen & Ledley, 1990). Tanaka *et al.* (1990) observed a number of artefactual `mutations' when sequencing M13 subclones of PCR amplified products, even in multiple clones. Their laborious solution was to repeat the re-amplification, cloning and sequencing procedures in order to distinguish the true from the artefactual mutations. Direct sequencing of PCR products eliminates the requirement to sequence multiple isolates, as the DNA sequence obtained is a consensus from a population of PCR products, rather than from a single cloned species. PCR sequencing has been reviewed by Gyllensten (1989).

The frequency of an identified base change can be readily determined in a panel of patient and normal samples with the technique of allele specific oligonucleotide (ASO) hybridization (Conner *et al.*, 1983). The relative frequency of the base change in the patient and normal populations can suggest whether the base change is a mutation or a polymorphism.

The results presented in this Chapter describe the molecular characterization of MPS-VI patients. Southern blot analysis of 18 MPS-VI patients failed to detect major deletions or rearrangements of the 4S gene. Nine patients were selected for detailed mutation analysis. cDNA-PCR analysis of the 4S mRNA revealed only one patient (SF368) that produced an anomalous-sized PCR product (see Section 4.5.1). The initial chemical cleavage protocol was applied to the 4S PCR products amplified from the cDNA of the selected patients. In a number of patients, chemical cleavage detected only one mismatch in the 4S cDNA. Subsequent direct DNA sequencing showed that the patient was heterozygous rather than homozygous for the identified base change. Therefore in these patients the initial chemical cleavage protocol had detected only one of the two mutant alleles in the 4S cDNA; the other mutant allele remained to be determined. It was considered

that the undetected mutant alleles were likely to be T-G mismatches, based on the (then) recent paper by Forrest *et al.* (1991), which showed that some T residues in T-G mismatches were resistant to chemical cleavage detection. Therefore, the modified protocol was introduced and used to re-screen three MPS-VI patients, in order to fully define all mutations and polymorphisms. The remaining patients were not re-screened with the modified chemical cleavage protocol due to time constraints. The differences between the initial and modified chemical cleavage protocols is discussed in Sections 4.2.3 and 4.5.2.

The techniques of chemical cleavage and direct DNA sequencing of PCR products, enabled the identification of a total of eight mutations; six were amino acid substitutions, one a single nucleotide deletion, and one an undefined mutation that caused exon IV to be absent from the 4S mRNA. A total of three polymorphisms were also identified, of which two were amino acid substitutions, and one was a silent change in the third base of an amino acid codon. The mutations identified, and their correlation with the biochemical and clinical phenotype of the MPS-VI patients is the subject of discussion in Section 4.6. Not all the results of chemical cleavage, DNA sequencing or ASO hybridization are reproduced in the thesis, in order to keep the thesis to a manageable size.

4.2 SPECIFIC MATERIALS & METHODS

4.2.1 PATIENT FIBROBLASTS

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

4.2.1a RNA EXTRACTION

Total RNA from normal and patient cultured skin fibroblasts was prepared essentially as previously described (Chomczynski & Sacchi, 1987). Specifically, monolayers of confluent cultured fibroblast cells were harvested using a rubber policeman into 10 ml phosphate buffered saline (PBS) per flask of cells. The suspended cells were transferred to a 50 ml Falcon tube and were pelleted by centrifugation at 300 g for 5 min. The cell pellet was transferred to preweighed Eppendorf centrifuge tubes, rinsed with 1 ml of PBS and spun for 10 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. The pellet was resuspended by vortexing in freshly prepared guanidinium isothiocyanate solution (solution D; 4 M guanidinium isothiocyanate, 25 mM tri-sodium citrate, pH 7.0, 0.5% (w/v) Sarkosyl, 0.1 M β-mercaptoethanol) (Chomczynski & Sacchi, 1987) at a ratio of 1 ml solution D per 100 mg of cells.

Subsequently, 1/10th volume of 2 M sodium acetate, pH 4.0, was added to the suspension and the mixture was vortexed, followed by the addition of an equal volume of phenol saturated with water, and then 1/5th volume of (49:1) chloroform:isoamyl alcohol. The mixture was vortexed and left on ice for 15 min. The suspension was centrifuged at 12,000 g for 10 min at 4°C to separate the RNA, which remained in the aqueous phase, from the DNA and protein, which were 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 placed at -20°C for at least 1 hr to allow the RNA to precipitate. The mixture was centrifuged at 12,000 g for 15 min at 4°C, the supernatant was drained, and the RNA pellet was resuspended in 40 µl of solution D. To the resuspended RNA was added 4 µl of 3 M sodium acetate, pH 5.2, and 100 µl of ethanol, and the mixture was placed at -20°C for at least 1 hr. The sample was centrifuged at 12,000 g for 15 min at 4°C and the RNA pellet was washed once with 70% (v/v) aqueous ethanol, vacuum-dried and dissolved in 20 µl of sterile water. A 2 µl sample was then used to determine the concentration of RNA by absorbance spectroscopy at a wavelength of 260 nm.

4.2.1b cDNA SYNTHESIS

First-strand cDNA was prepared from total cellular RNA by reverse transcription. cDNA synthesis was performed using 3 µg of total RNA (see above) mixed with 2 µg random hexamers in 1 x Moloney murine leukaemia virus reverse transcriptase (MoMLV RT) buffer (50 mM Tris-HCl pH 8.3, 75 mM KCl, 3 mM $MgCl_2$, 10 mM DTT; BRL) in a total volume of 50 µl containing 40 units RNAsin (Promega) and 0.5 mM each of dATP, dCTP, dGTP and dTTP (Boehringer Mannheim). The reaction was placed at 65°C for 3 min then cooled on ice, 600 units of MoMLV reverse transcriptase (BRL) were added and the reaction incubated at 37°C for 60 min. The RNA was hydrolyzed by the addition of 5 µl of 3 M NaOH and incubation at 65°C for 15 min, the reaction mixture was then neutralized by the addition of 1.3 µl 32 % (w/v) HCl. The cDNA was precipitated with $1/10^{th}$ volume (5 µl) 3 M sodium acetate, pH 5.2, and 2.5 volumes (125 µl) of ethanol. The solution was centrifuged at 12,000 g for 15 min at 4°C to pellet the precipitated cDNA. The cDNA pellet was rinsed once with 95% (v/v) aqueous ethanol, dried and resuspended in 50 µl of water. Aliquots of 5 µl were used in typical PCR reactions.

4.2.1c GENOMIC DNA PREPARATION

Monolayers of confluent cultured fibroblast cells were harvested using a rubber policeman into 10 ml PBS per flask of cells. The suspended cells were transferred to a 50 ml Falcon tube and were pelleted by centrifugation at 300 g for 5 min. The cells were resuspended in 1ml of cell lysis buffer (320 mM sucrose, 10 mM Tris-HCl, pH 7.5, 5 mM $MgCl_2$, 1% (w/v) Triton X-100) in a 10 ml polypropylene tube and incubated on ice for 30 min. The suspension was centrifuged at 3,600 g for 5 min and all of the supernatant was removed by suction. The cell pellet was gently resuspended in 100 µl of proteinase-K buffer (10 mM Tris-HCl, pH 7.4, 10 mM EDTA, 10 mM NaCl) and transferred to Eppendorf centrifuge tubes. A further 225 µl of proteinase-K buffer was added, followed by 50 µl of 10% (w/v) SDS and 20 µl of a 10 mg/ml stock of proteinase-K (Boehringer Mannheim) dissolved in proteinase-K buffer. The mixture was then gently mixed by rotation during an overnight incubation at room temperature.

The next day, 300 µl of water was added to the digest, followed by 600 µl of water-saturated phenol, and the sample was mixed gently by inversion. The suspension was centrifuged at 12,000 g for 1 min at room temperature to separate the DNA, which remained in the aqueous phase, from the cell debris. The aqueous phase was transferred to a fresh tube, and was re-extracted with water-saturated phenol, as above. The resultant aqueous phase was then extracted with an equal volume of (24:1) chloroform : isoamyl alcohol and centrifuged at 12,000 g for 15 s at room temperature to resolve the phases. The aqueous phase was collected and divided into two equal volumes in Eppendorf tubes, and the DNA was precipitated by the addition to each aliquot of 1/2 volume 7.5 M ammonium acetate and 3 volumes 95% (v/v) aqueous ethanol. The tubes were mixed gently and the DNA precipitate was pelleted by centrifugation at 12,000 g for 1 min at room temperature. The supernatant was removed and the DNA pellet was rinsed with 1 ml of 70% (v/v) aqueous ethanol, drained, and the DNA was dried under vacuum. The genomic

DNA was gently resuspended in 50 µl of TE buffer (10 mM Tris-HCl, pH 7.5, 0.1 mM EDTA). Aliquots of 1 µl were used in typical PCR reactions.

4.2.2 POLYMERASE CHAIN REACTION

Two main polymerase chain reactions (PCR) were performed in order to cover the entire coding region of the 4S cDNA. The PCR A product which spanned the 5'-end of the 4S cDNA was synthesized with primers 4SP5A and 4SP7, and the PCR B product which spanned the 3'-end of the 4S cDNA was synthesized with primers 4SP6 and 4SP8. Other PCR products were also generated to enable the identification of base changes by the direct DNA sequencing technique, and for ASO hybridization. The locations of the PCR products relative to the 4S cDNA are represented in Figure 4.2. Table 4.2 lists the primers used, the template, the conditions of the PCR reaction and the size of the PCR product, for all relevant 4S PCR reactions. The DNA sequences of the primers used are listed in Appendix A. The templates for the PCR reactions included the normal and patient cDNA synthesis reactions (see Section 4.2.1b). The 1.65-kb *EcoR*I-*Stu*I fragment from the modified 4S cDNA clone p4SFL (described in Section 4.5.1, and Anson *et al.*, 1992b) was used as the normal reference control template. Genomic DNA isolated from normal and MPS-VI patients was also used as templates, depending on the PCR reaction.

Generally, the PCR conditions were essentially the same as described (Saiki *et al.*, 1988), with the exception of the final deoxynucleotide concentration and the need for dimethyl sulphoxide (DMSO) in some reactions. The addition of DMSO was necessary in order to obtain the PCR A product and other 5'-end PCR products. Presumably the DMSO denatured the secondary structures which resulted from the rich G+C content of the sequence at the 5'-end, and which inhibited complete strand synthesis. DMSO was not required to obtain the PCR B product and other 3'-end PCR products. For the PCR A and PCR B reactions, a 5 µl aliquot of the

cDNA synthesis reaction was mixed with 50 pmoles of each PCR primer in a total volume of 100 µl containing PCR buffer [50 mM KCl, 10 mM Tris-HCl pH 8.4, 2.5 mM $MgCl_2$, 0.01 % (w/v) gelatin], 10 % (v/v) DMSO if indicated, 2.5 units Taq polymerase (Cetus or Biotech International) and 0.4 mM each of dATP, dCTP, dGTP and dTTP. Forty cycles using the following conditions were performed: denaturation at 94°C for 45 s, annealing at 58°C for 45 s, and DNA polymerization at 72°C for 2 min, using a Perkin Elmer Cetus DNA thermal cycler, followed by a final extension at 72°C for 3 min. Generally, the thermal cycling parameters were the same for the other PCR reactions; the annealing temperatures were as in Table 4.2. The DNA polymerization step in the PCR reaction was at 72°C for 2 min for products over about 600-bp, while for the shorter products the PCR reaction was at 72°C for 1 min or less. PCR products 22-30 and 31-32 required the final $MgCl_2$ concentration at 4.0 mM and 2.0 mM respectively for optimal results. Control PCR reactions, in which no cDNA template was added were included during each set of PCR reactions. Unless otherwise stated, the nucleotide sequence of 4S is numbered according to Peters *et al.* (1990) for sequence distal to the methionine initiation codon, and numbered according to Schuchman *et al.* (1990) for sequence proximal to the methionine initiation codon.

4.2.3 CHEMICAL CLEAVAGE OF HETERODUPLEXES

The chemical cleavage technique was the method of choice for the detection of mutations at the nucleotide level. Chemical cleavage analysis of MPS-VI patients was performed using the PCR A and PCR B products as substrates, which together spanned the entire coding region of 4S. Two chemical cleavage protocols were utilised for the detection of mutations, which differed primarily in the labelling of the PCR products derived from the patient and the normal control. In the initial protocol, a heteroduplex was formed between a radiolabelled normal PCR product and an unlabelled patient PCR product. The normal PCR product was labelled in both strands using 5'-end-labelled PCR primers. The initial protocol was

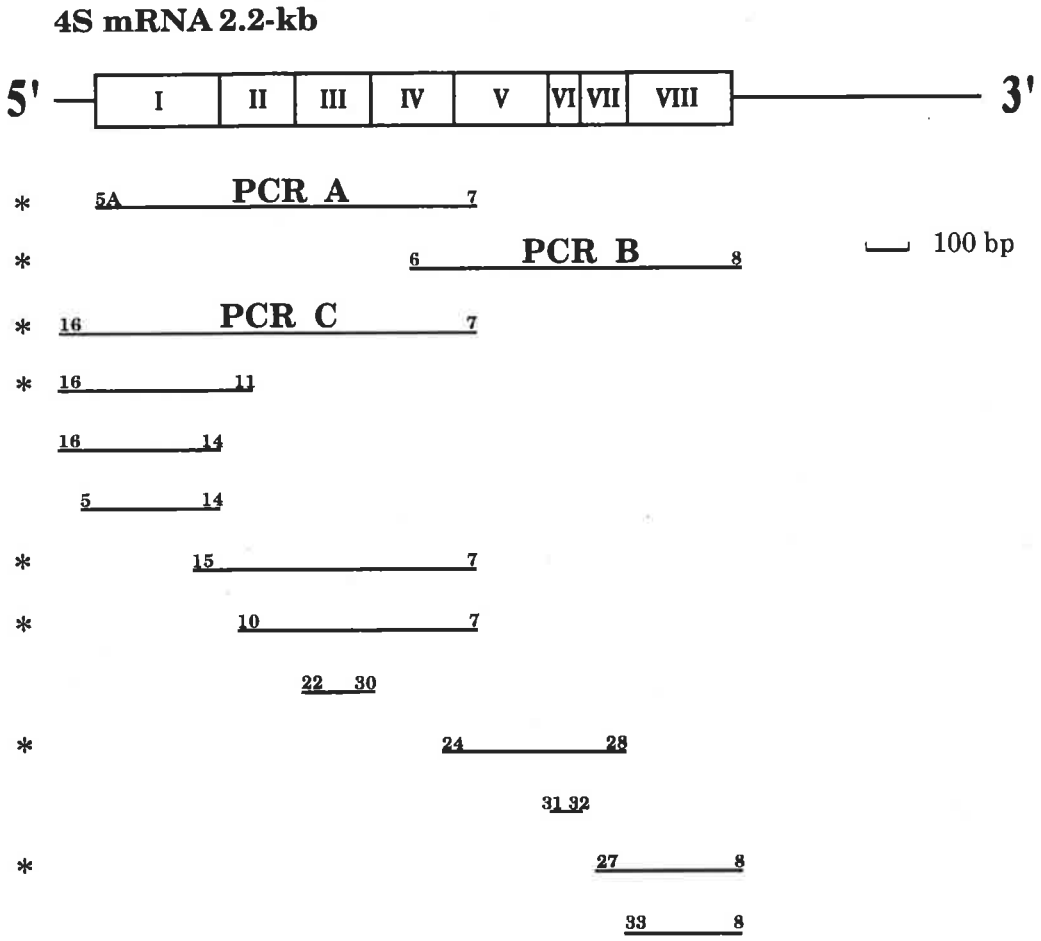


FIGURE 4.2 PCR AMPLIFICATION OF THE ENTIRE 4S CODING REGION

Diagrammatic representation of the 4S coding region, showing the position of the eight exons and the location and length of all the PCR products used in identification of the 4S mutations. PCR products marked (*) can only be amplified from a cDNA template, while the other products can be amplified from both genomic and cDNA templates. The intron-exon boundaries were provided by Christoph Peters and Kurt von Figura (Göttingen). The PCR reactions are described in Section 4.2.2. The numbers represent the oligonucleotide primers used, which are listed in Appendix A.

PCR primers	Template	Size (bp)	Annealing temp. (°C)	$MgCl_2$ (mM)	DMSO (+/-)
5A-7 (A)	cDNA	937	58	2.5	+
6-8 (B)	cDNA	835	58	2.5	-
16-7 (C)	cDNA	1016	58	2.5	+
16-11	cDNA	457	58	2.5	+
16-14	cDNA/genomic	388	58	2.5	+
5-14	cDNA/genomic	328	58	2.5	+
15-7	cDNA	699	58	2.5	-
10-7	cDNA	624	58	2.5	-
22-30	cDNA/genomic	98	60	4.0	-
24-28	cDNA	411	58	2.5	-
31-32	cDNA/genomic	89	60	2.0	-
27-8	cDNA	354	58	2.5	-
33-8	cDNA/genomic	293	58	2.5	-

TABLE 4.2 4S PCR REACTIONS

The oligonucleotide primer pairs used to synthesize the 4S PCR products are shown, together with the template and the size of each product. The annealing temperature, the final $MgCl_2$ concentration, and the presence (+) or absence (-) of 10% (v/v) DMSO in the reaction mixes are also shown. PCR was performed as described in Section 4.2.2. The sequences of the PCR primers are listed in Appendix A.

performed essentially as described in Cotton *et al.* (1988). In the modified protocol, a heteroduplex was formed between a normal and a patient PCR product, both of which were labelled. However, the normal and the patient PCR products were each labelled in a single strand only at a time by performing the PCRs with one of the PCR primers 5'-end-labelled. The modified protocol was performed essentially as described in Forrest *et al.* (1991).

4.2.3a PROBE SYNTHESIS

The PCR A and PCR B products were synthesized from the following templates; the 1.65-kb *EcoR*I-*Stu*I fragment of p4SFL (the normal reference control), and the MPS-VI cDNA samples. Aliquots of these PCR reactions were used as substrates for the probe PCR reactions. In the initial protocol, the radioactive probes were produced using labelled primers in a PCR reaction with 1 µl of normal reference control PCR product as template. PCR conditions were as described in Section 4.2.2 except that 20 ng (3 pmole) of each primer was used and both primers in each primer-pair were 5'-end-labelled with γ-^{32}P-ATP (5,000 Ci/mmol) (Section 2.2.11a). Thus, for each PCR region, 1 probe was produced from the normal reference control DNA. In the modified protocol, the templates for the probe PCR reactions were first gel purified. The normal reference control and patient PCR products were electrophoresed on a 3% (w/v) Nusieve GTG / 1% (w/v) normal agarose gel, the PCR product was excised, and four volumes of water was added and heated at 65°C for 10 min. The radioactive probes were produced using labelled primers in a PCR reaction, with 1 µl of the gel purified sample of either the normal reference control or the MPS-VI patient PCR product as template. PCR conditions were as described in Section 4.2.2 except that 30 ng (4.5 pmole) of each primer was used and only one of the two primers was 5'-end-labelled at a time. Thus, for one PCR region, a total of 4 probes were produced from both normal DNA and MPS-VI patient DNA. In both protocols, the probes were precipitated, analyzed on 2% (w/v) agarose gels,

electrophoresed onto DEAE NA-45 paper (Schleicher & Schüll) and eluted as described below.

The contents of each probe PCR reaction (100 µl) was transferred to an 1.5 ml Eppendorf centrifuge tube, 150 µl of isopropanol was added, and the DNA was allowed to precipitate at -20°C for at least 30 min. The samples were centrifuged at 12,000 g for 30 min at 4°C, the supernatant was carefully removed and the DNA pellet was vacuum-dried. Each probe was resuspended in 10 µl of water and 2 µl of gel loading buffer, and was then loaded onto a 2% (w/v) agarose gel. The probes were resolved by electrophoresis, the bands visualized by staining with ethidium bromide, and the probe products were electrophoresed onto NA-45 paper. The NA-45 paper containing each bound probe was removed from the gel and transferred to a 1.5 ml Eppendorf centrifuge tube containing 150 µl of elution solution (20 mM Tris-HCl, pH 8.0, 0.1 mM EDTA, 1 M NaCl). The probe was eluted from the NA-45 paper by incubation at 65°C for 30 min. The eluates were removed to fresh tubes and 300 µl of water and 1100 µl of 95% (v/v) aqueous ethanol were added. The DNA was allowed to precipitate at -20°C overnight. The probe samples were centrifuged at 12,000 g for 30 min at 4°C, the supernatant was carefully removed and the DNA pellet was vacuum-dried. The probe was resuspended in 50 µl of water. The concentration of the probe was visually estimated by analytical gel electrophoresis against DNA samples of known concentration.

4.2.3b HETERODUPLEX FORMATION

In the initial protocol, the labelled normal control PCR probe (10-20 ng) (synthesized as described in Section 4.2.3a) was mixed with the patient PCR product (100-200 ng). In the modified protocol, approximately 20 to 100 ng of both the patient and normal probes (synthesized as described in Section 4.2.3a) were combined. In both protocols the DNA mixture was in a total volume of 100 µl in a 0.5 ml PCR tube, and the final concentration of 1 X annealing buffer was 0.3 M

NaCl, 3.5 mM $MgCl_2$, 3 mM Tris-HCl, pH 7.5. The normal-patient duplex mixtures were denatured at 100°C for 5 min, cooled to 65°C, and allowed to anneal as the temperature was cooled linearly from 65°C to 42°C over 60 min, and finally cooled to 4°C, using a Perkin Elmer Cetus thermal cycler. Carrier DNA, in the form of 1 µl of 10 mg/ml herring sperm DNA, was added and the DNA duplexes were precipitated with 10 µl of 3 M sodium acetate, pH 5.2, and 250 µl of ethanol. The DNA precipitations were centrifuged at 12,000 g for 10 min at 4°C, the supernatant was drained, and the DNA pellet was vacuum-dried and resuspended in 12 µl of water.

4.2.3c HYDROXYLAMINE MODIFICATION

Hydroxylamine (1.39 g) was dissolved in 1.6 ml of water and the pH was adjusted to about pH 6 by the addition of approximately 1 ml of diethylamine. A modification reaction consisting of 6 µl of the duplex described in Section 4.2.3b, 1 µl of 10 mg/ml herring sperm DNA and 20 µl of the hydroxylamine stock was mixed and incubated at 37°C for 45 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 750 µl of 95% (v/v) aqueous ethanol. The samples were mixed and centrifuged at 12,000 g for 10 min at 4°C, and the supernatant discarded. The initial protocol included a subsequent step which was omitted from the modified protocol, where the DNA pellet was dissolved in 50 µl of water, and then re-precipitated with 5 µl of 3 M sodium acetate, pH 5.2 and 125 µl of 95% (v/v) aqueous ethanol and centrifuged at 12,000 g for 10 min at 4°C. The DNA pellet was rinsed with 70% (v/v) aqueous ethanol and dried under vacuum. The samples were then subjected to piperidine cleavage. All of these reactions were performed in a fume hood due to the hazardous nature of hydroxylamine and diethylamine.

4.2.3d OSMIUM TETROXIDE MODIFICATION

The remainder of the duplex reaction (6 µl, Section 4.2.3b) was combined with 1 µl of 10 mg/ml herring sperm DNA, 2.5 µl of 10 X O 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 an osmium tetroxide solution. In the initial protocol, the osmium tetroxide solution was a 4% (w/v) stock solution in water, whereas in the modified protocol, the 4% osmium tetroxide stock was first diluted 5-fold in water (v/v) prior to use. The modification reactions were mixed and incubated at 37°C for 5 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 and 25 µg/ml tRNA) and 750 µl of 95% (v/v) aqueous ethanol. The samples were mixed and centrifuged at 12,000 g for 10 min at 4°C, and the supernatant discarded. The initial protocol included a subsequent step which was omitted from the modified protocol, as described above for the hydroxylamine modification. The DNA pellet was rinsed with 70% (v/v) aqueous ethanol and dried under vacuum. The samples were then subjected to piperidine cleavage. All of these reactions were performed in a fume hood due to the volatile and highly toxic nature of osmium tetroxide.

4.2.3e PIPERIDINE CLEAVAGE & ANALYSIS

A solution of 10% (v/v) piperidine in water was freshly prepared and 50 µl was added to the hydroxylamine- and osmium tetroxide-modified DNA pellets. The reaction tubes were incubated at 90°C for 30 min. The samples were precipitated with 5 µl of 3 M sodium acetate pH 5.2 and 125 µl of 95% (v/v) aqueous ethanol. The hydroxylamine-treated samples were centrifuged at 12,000 g for 10 min at 4°C to pellet the DNA. The osmium tetroxide-treated samples required incubation in a dry-ice/ethanol bath for 15 min, prior to the centrifugation described for the hydroxylamine-treated samples, in order to maximise the yield of DNA. The supernatants from both samples were discarded, the DNA pellets were carefully

rinsed with 70% (v/v) aqueous ethanol, vacuum-dried, and resuspended in 5 µl of sequencing loading buffer. The samples were denatured at 100°C for 3 min. All of the previous reactions were performed in a fume hood. An aliquot of each sample (1-2 µl) was loaded onto a 5% denaturing polyacrylamide gel, and the gel was processed as described in Section 2.2.8b. The observed fragments were sized using the 5'end-labelled markers pUC19 digested with *Hpa*II and SPP-1 digested with *Eco*RI, in order to estimate the location of the mutations.

4.2.4 DIRECT DNA SEQUENCING OF PCR PRODUCTS

4.2.4a TEMPLATE PREPARATION

PCR products to be used as templates for direct PCR sequencing were purified (for products >200-bp) using Centricon-100 microconcentrator columns (Amicon), according to the manufacturer's instructions. Centricon columns contain integral 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 at 1,000 g in a fixed angle rotor allowing unreacted nucleotides and primers to pass through the filter while retaining the PCR product. 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. About 0.5-1.0 pmoles of a double-stranded PCR product was used in a sequencing reaction; for a 400-bp PCR product, this corresponded to about 200 ng.

4.2.4b DNA SEQUENCING

Direct PCR sequencing was performed as a modification of the linear PCR method (Murray, 1989). DMSO was added to 10% (v/v) in those PCR sequencing reactions for which the production of the template PCR product required DMSO. The reaction conditions used to sequence the PCR products with each primer are given in Appendix C. The double-stranded PCR template (0.5-1.0 pmoles) was mixed with 3 pmoles of ^{32}P-end-labelled sequencing primer (Section 2.2.11a) in PCR buffer (Section 4.2.2), 1.5 units of Taq polymerase (Cetus or Biotech International), and 7.5 μM each of dATP, dCTP, dTTP and 15 μM of 7-deaza-dGTP. The reaction mixture was divided into four termination reactions containing one of each of the ddNTP chain terminators; the final concentrations were 0.25 mM ddATP, 0.1 mM ddCTP, 0.05 mM ddGTP, and 1 mM ddTTP. Fifteen PCR cycles were performed using the following conditions; denaturation at 95°C for 30 s, annealing at 55-58°C for 30 s and DNA polymerization at 72°C for 30 s. The annealing temperature was varied depending on the PCR product and the primer used for DNA sequencing, as shown in Appendix C. A 5 μl aliquot was removed from each termination reaction, 4 μl of formamide loading dye was added, the reactions were heat denatured for 3 min at 100°C and then loaded onto a 6% sequencing gel. Subsequent manipulations were as for a standard sequencing gel (Section 2.2.8b).

4.2.5 ALLELE SPECIFIC OLIGONUCLEOTIDE (ASO) HYBRIDIZATION

The PCR primers used to amplify the 4S genomic sequences of interest for detection of mutations are shown in Figure 4.2, and the PCR reaction conditions used are shown in Table 4.2. The allele specific oligonucleotides used for the detection of individual mutations, and the final washing temperatures are shown in Appendix D. PCR reactions were run in duplicate on a 3% (w/v) Nusieve GTG / 1% (w/v) standard agarose gel, Southern blotted under alkaline conditions onto a

GeneScreen*Plus*™ nylon membrane (Section 2.2.12b), and the duplicate sections of the filter were separated. The filters were prehybridized in 6 x SSC, 0.1% (w/v) SDS, 0.05% (w/v) sodium pyrophosphate, and 100 μg/ml sheared, denatured herring sperm DNA at 42°C for 1 hr. The duplicate filters were individually hybridized for 4 hr with 50 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 10 min. Autoradiography was carried out 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 the PCR primers and allele specific oligonucleotides are listed in Appendices A and B respectively.

RESULTS & DISCUSSION:

4.3 LOW-RESOLUTION ANALYSIS OF THE 4S GENE

4.3.1 SOUTHERN ANALYSIS OF MPS-VI PATIENTS

In order to detect large-scale rearrangements of the 4S gene in MPS-VI patients, the genomic DNA from patients was subjected to Southern analysis (Southern, 1975). Specifically, genomic DNA samples from MPS-VI and normal individuals were digested with restriction endonucleases and the resulting fragments were resolved by gel electrophoresis. The fragments were then transferred to a membrane and hybridized with the radiolabelled 4S cDNA as a probe, as detailed in Sections 2.2.12b and 2.2.13b. The 4S gene probe was the 2.3-kb *Eco*RI 4S cDNA fragment, which was originally provided by our collaborators in Göttingen as an M13mp18 subclone (pASB2-5). To permit large quantities to be produced for subsequent experiments, the 2.3-kb *Eco*RI 4S cDNA fragment was subcloned into

pUC19. The 2.3-kb *Eco*RI 4S cDNA fragment did not encode the complete coding sequence of 4S as it was missing 13 amino acid codons (38-bp) at the 5'-end. Prior to MPS-VI patient analysis, digestion of normal genomic DNA with a series of restriction endonucleases (11) showed that *Eco*RI resulted in the most favourable distribution of band size for Southern analysis.

A group of 18 MPS-VI patients were selected at random from the MPS-VI cell culture collection; the patients were representative of mild and severe forms of the clinical spectrum, and approximately equal numbers of both sexes were selected. Digestion of the patient genomic DNA with *Eco*RI and subsequent Southern analysis revealed that none showed any discernible change in the Southern band pattern, relative to normal controls (data not shown). Consequently, this suggested that the majority (>95%) of the 4S gene mutations in these MPS-VI patients were point mutations, or very small insertions or deletions.

4.4 PATIENT SELECTION

At the commencement of the attempt to identify 4S mutations in MPS-VI patients, about 70 MPS-VI patient cell lines were available for analysis in the Departmental cell culture collection. It was not considered feasible to perform a detailed analysis for 4S mutations in all 70 patients, given the constraints of time and economy. Consequently, a subset of patients were selected for detailed analysis, based on their clinical severity and the residual 4S enzyme activity (Table 4.3). Residual 4S enzyme activity was measured in extracts of cultured skin fibroblasts using a radiolabelled trisaccharide substrate (G. Gibson, unpublished data). The selected patient subset included patients with clinical phenotypes ranging across the clinical spectrum from mild to severe, so that mutation analysis might provide an insight into the molecular and biochemical basis for the variation in clinical severity. Fibroblast cell lines from patients were also selected that displayed interesting differences in biochemical parameters such as specific activity relative

TABLE 4.3 SELECTION OF MPS-VI PATIENTS FOR MUTATION ANALYSIS

The nine MPS-VI patients selected for detailed mutational analysis are listed, together with the age at diagnosis, clinical phenotype, urinary DS levels, and their known biochemical data. Patient SF912 is also included in the Table, as SF912 and SF913 are siblings . The age at diagnosis, clinical phenotype, and the urinary DS values (grams/mol creatinine) were from Brooks *et al.* (1991b), J. Hopwood (unpublished results), or from the referring doctor. The 4S activity values were obtained from G. Gibson (unpublished results). The 4S activity was determined in normal and patient cultured fibroblast extracts, using a radiolabelled trisaccharide substrate (Hopwood *et al.*, 1986), and is expressed in pmol/min/mg fibroblast protein. The 4S protein was quantified using the anti-4S monoclonal antibody ASB 4.1 in an ELISA assay. The specific activity was defined as the 4S activity per unit of 4S protein (pmol/min/ng 4S). Normal-A refers to the urinary DS ranges determined from over 1000 individuals, which varies depending on age. Normal-B refers to a single normal fibroblast cell line, assayed in duplicate. The criteria used to select the patients is given in detail in Section 4.4. NA = data not available, ND = not detectable.

Patient	Age at diagnosis	Clinical phenotype	Urinary dermatan sulphate	4S activity	4S specific activity
SF1022	7 days	severe	88	NA	NA
SF2424	2 months	severe	127	ND	ND
SF1246	13 months	severe	NA	0.5	0.2
SF2259	2 years	severe	79	0.7	7.5 3.1
SF368	2 years	severe	49	3.6	20
SF912	13 years	mild	8	1.1	0.7
SF913	16 years	mild	9	0.6	0.2
SF2724	17.5 years	mild	4	0.2	0.1
SF2467	35 years	mild	NA	0.5	0.2
SF2357	44 years	asymptomatic	1	3.9 5.8	1.7 2.5
Normal-A	< 2 months	Normal	<4		
	2 mo - 2 years	Normal	<2		
	2 - 6 years	Normal	<1		
	> 6 years	Normal	<0.7		
Normal-B				84 94	1.3 1.5

to normal controls. It was hoped that the selection process would facilitate the identification of 4S mutations that would affect, for example, residues crucial to 4S enzyme catalysis or substrate binding.

For recessively inherited genetic disorders such as MPS-VI, patients are either homozygotes, with two identical mutant alleles, or compound heterozygotes, with two distinct mutant alleles. In order to work up the high-resolution mutational analysis techniques, the initial screen for 4S mutations was performed on a patient that was most likely to possess a homozygous mutation, which was considered to simplify analysis and interpretation of the molecular lesion. Patient SF1022 was the first selected for detailed analysis as the patient was the outcome of a consanguineous union and was therefore likely to be homozygous for a MPS-VI mutation, especially as MPS-VI is a rare inherited disorder. Patient SF1022 displayed a severe clinical and biochemical phenotype as discussed further in Section 4.6.

The patients SF913, SF1246, SF2467 (Repository number GM2849, Human Genetic Mutant Cell Repository, Camden, New Jersey, USA), and SF2724 were selected for detailed mutation analysis as their specific activities (pmol/min/ng 4S) were reduced about 6-13-fold compared to normal controls. Therefore, it was tentatively proposed that in each of these patients, the combination of the two mutant 4S alleles may have affected residues involved in substrate catalysis and/or turnover. In addition, the clinical phenotypes observed in this group of four patients ranged from relatively mild to severe forms, and so mutation analysis in this group was considered likely to provide an insight into the basis for the variation in clinical severity observed in MPS-VI.

Patients SF2259 and SF2357 were chosen because the 4S specific activity values were about normal or elevated, whereas the residual activity was reduced relative

to a normal control; about 100-fold for SF2259 and about 20-fold for SF2357 (Table 4.3). Therefore, the suggestion was made that in each of these patients, the combination of the two mutant 4S alleles may have reduced the steady-state level of 4S protein rather than the specific activity of the enzyme. Patient SF2259 was clinically severe, was diagnosed at 2 years of age, and excreted greatly elevated levels of DS in the urine. Patient SF2357 was considered clinically normal, but was diagnosed serendipitously during a haematological examination for an unrelated minor complaint. Characteristic inclusion bodies were observed in the patient's leukocytes, indicative of a lysosomal storage disorder. The patient also exhibited mildly elevated levels of DS excreted in the urine. Subsequent measurement of lysosomal enzyme activities refined the diagnosis as MPS-VI. This patient was considered especially interesting as biochemical analysis revealed the patient was deficient in 4S, yet was clinically asymptomatic for MPS-VI. Perhaps patient SF2357 defines the boundary of a threshold effect, where the clinical features of MPS-VI only begin to develop once the 4S catalytic capacity of a cell falls below a critical or threshold value.

The severely affected patients SF368 and SF2424 were also chosen for further analysis. SF368 was particularly interesting, as the severe clinical phenotype (diagnosed at 2 years of age), and the elevated levels of DS in the urine, did not correspond directly with the biochemical phenotype *in vitro*. The mutant 4S enzyme in patient SF368 displayed a markedly elevated specific activity relative to a normal control, and the residual activity was similar to that of the asymptomatic patient SF2357 (Table 4.3). For patient SF2424, the kinetic parameters of the mutant 4S were unable to be detected (Table 4.3). The patient was clinically severe, was diagnosed at 2 months of age, and excreted extremely elevated levels of DS in the urine. The biochemical and clinical results are consistent with the suggestion that in this patient, the combination of the two mutant 4S alleles had profoundly deleterious affects on the 4S enzyme.

The residual enzyme activity and specific activity measurements of 4S derived from the patient cultured fibroblasts were useful in the selection of patients for further investigation. However, there were limitations in the activity assay system and the data obtained; firstly, in the reproducibility of the assay. The duplicate determinations of the 4S specific activity in SF2357 and SF2259 varied approximately 2-fold (Table 4.3). In addition, the two patients SF912 and 913 were siblings with presumably the same mutant alleles, yet the 4S specific activities varied >2-fold. Secondly, the residual activity did not exhibit a simple correspondence with the age of diagnosis (a measure of disease severity) for all patients. For example, patients SF1246 and SF2467 both had a similar 4S residual activity, yet the former was diagnosed early at 13 months of age, while the latter was diagnosed at the advanced age of 35 years. A similar phenomenon was observed for patients SF368 and SF2357. These limitations in the residual activity data were also observed in Brooks *et al.* (1991b). The patient activity data in Table 4.3 (and in Brooks *et al.*, 1991b) were obtained before the contribution of arylsulphatase derived from foetal calf serum was appreciated. The patient fibroblasts were cultured in media which contained foetal calf serum that had not been heat inactivated, and hence could conceivably supply 4S activity to the patient fibroblasts. This exogenous 4S activity may be responsible, at least in part, for some of the limitations in the data, and is discussed further Chapter 5, Section 5.6.4.

4.5 HIGH-RESOLUTION MUTATION ANALYSIS OF THE 4S cDNA

This Section describes the use of a range of techniques to determine the precise molecular nature of the 4S mutations in a group of MPS-VI patients. The techniques included cDNA-PCR amplification, chemical cleavage of mismatches, direct DNA sequencing of PCR products, and ASO analysis.

4.5.1 cDNA-PCR ANALYSIS OF MPS-VI PATIENTS

Analysis of the 4S mRNA in MPS-VI patients by the technique of cDNA-PCR amplification enabled the detection of gross changes in the size or relative abundance of the 4S mRNA. The 4S cDNA was derived from the patient RNA by reverse transcription of total RNA, the cDNA was then used as the template for the PCR amplification of portions of the 4S cDNA *in vitro*. Genomic DNA could not be used as a template, as although the exon boundaries were known, the adjacent intronic sequences necessary for PCR amplification from the genome were not provided by our collaborators. If mRNA splicing abnormalities were detected in the 4S cDNA of MPS-VI patients, the identification of the underlying mutation would require genomic DNA sequence. A potential problem with mutational analysis confined to the cDNA is that some mutations have been reported which destabilize the level of RNA expression, and therefore could cause difficulty in detecting both alleles in a patient (e.g. ornithine δ-aminotransferase frameshift mutations in gyrate atrophy, Brody *et al.*, 1992).

The relative abundance of a mRNA can be quantified by densitometric scanning of Northern blot autoradiographs. This technique enables the effects of a mutation on the steady-state level of a mRNA in a patient sample to be determined. The abundance of the 4S mRNA was low in normal and MPS-VI patients, typically 1 μg of patient poly(A)$^+$ RNA from cultured skin fibroblasts was required to detect the three mRNA species of 4S on a Northern blot (Peters *et al.*, 1990). Consequently, 4S mRNA from patients was not subjected to Northern analysis as insufficient mRNA was isolated from patient cultured skin fibroblasts to enable reproducible detection on a Northern.

PCR amplification of cDNA also enables quantification of the relative abundance of a mRNA in normal and patient RNA. However, to be accurately quantitative the

technique requires extensive effort to ensure the linear accumulation of the cDNA-PCR product with cycle number (Delidow *et al.*, 1989). Furthermore, the relationship of the mRNA of interest to it's encoded protein must be reasonably well characterized. Quantitative cDNA-PCR of the 4S mRNA in normal and patient cell lines was not possible, as the structure of each of the three 4S mRNAs and their individual contribution to the synthesis of 4S protein had not been determined. However, cDNA-PCR amplification of the 4S mRNA could still provide a qualitative estimate of the relative 4S mRNA levels in normal and patient cell lines. In addition, gross rearrangements in the size of the 4S mRNA could also be detected, but with the benefit of greater economy as a reduced quantity of mRNA and hence cultured patient fibroblasts would be required compared to Northern analysis.

Total RNA was isolated from cultured skin fibroblasts of MPS-VI patients and normal controls, and the cDNA was generated using random primers. Two PCR products were synthesized that together spanned the entire coding region of 4S (Fig. 4.2). PCR primers 4SP5A and 4SP7 were used to generate the 937-bp product A (PCR A) which spanned the 5'-portion of the 4S cDNA, and PCR primers 4SP6 and 4SP8 were used to generate the 835-bp product B (PCR B) which spanned the 3'-portion of the 4S cDNA. A third PCR product, which overlapped with product A but extended into the 5'-untranslated sequence, was generated between 4SP16 and 4SP7 (PCR C). All three PCR products were generated from both normal and MPS-VI patient cDNA, whereas only PCR A and B could be obtained from the modified 4S cDNA clone, p4SFL. The p4SFL clone contained 5'-untranslated sequence from the rat prepro-insulin gene (Anson *et al.*, 1992), and therefore PCR products incorporating primers designed to sequence 5'-adjacent to the 4S coding sequence could not be generated from this clone. The original 4S cDNA clone, pASB2-5, could not be used to generate either PCR A or PCR C, as it was incomplete.

The PCR C and PCR B products were generated from cDNA from each of the nine selected MPS-VI patients and from normal controls. The PCR products obtained from eight of the patients were indistinguishable in size and yield from normal controls. However, patient SF368 produced a PCR C product which was about 200-bp shorter and in reduced yield compared to the normal controls. In addition, PCR B from patient SF368 was significantly reduced in yield although indistinguishable in size compared to the other patients and the normal control.

In order to determine the location of the 200-bp `deletion' in the 4S cDNA of patient SF368, two PCR products which spanned the 5'-portion and 3'-portion of PCR A were amplified (Fig. 4.2). The patient's 5'-PCR product, which was generated between primers 4SP5 and 4SP14 and spanned nucleotide positions from -20 to 308, was indistinguishable in size and yield from normal controls. However, two PCR products were observed when the patient's cDNA was amplified between 4SP15 and 4SP7, which spanned the 3'-end of PCR A between nucleotide positions 238 and 937. In addition to the predicted 699-bp band which was also observed in a normal control, a novel PCR product of about 480-bp was observed only in SF368 (data not shown).

Taken together, the results suggested that the patient was probably heterozygous for a mutation that caused the 4S mRNA to be missing about 200-bp from a region bounded by nucleotide positions 309 (distal to 4SP14) and 917 (proximal to 4SP7). The precise location and extent of the `deletion' were determined with the subsequent techniques of chemical cleavage and direct DNA sequencing of PCR products, as described below.

As the gross properties of the 4S mRNA were not apparently affected in the other mutant allele of patient SF368, and in the other eight selected MPS-VI patients, the mutations in these alleles were considered to be single-base changes within the

coding region, or modifications in the non-coding regions outside the span of the PCR products. In order to define with precision the molecular nature of the presumed point mutations in the 4S gene, techniques which enabled the detection of single-base changes in the 4S gene were selected. The following Section describes the use of the chemical cleavage technique, direct DNA sequencing of PCR products, and allele specific oligonucleotide hybridization to identify the mutations in MPS-VI patients.

4.5.2 IDENTIFICATION OF 4S MUTATIONS IN MPS-VI PATIENTS

Prior to the presentation of the mutational analysis results for each patient, the chemical cleavage technique is discussed below. Two protocols of the chemical cleavage technique were utilized in the high-resolution analysis of the selected MPS-VI patients, which differed primarily in the labelling of the PCR products derived from the patient and the normal control. The initial protocol generated a heteroduplex between a labelled normal PCR product and an unlabelled patient PCR product, and was essentially based on the method of Cotton *et al.* (1988). The normal PCR product was labelled in both strands by performing the PCR with both primers 5'-end-labelled. Chemical cleavage analysis was performed on normal probe homoduplexes, and normal-patient heteroduplexes. Subsequently, it was shown that some T residues in T-G mismatches were resistant to modification by osmium tetroxide and cleavage by piperidine, and therefore were not detected with the initial protocol (Forrest *et al.*, 1991). The reactivity of a T-G mismatch appears to be strongly influenced by the surrounding DNA sequence context. Fortunately, a heteroduplex containing a T-G mismatch is always accompanied by a complementary heteroduplex containing an A-C mismatch, which can be modified by hydroxylamine. Forrest *et al.* (1991) found that labelling of the patient PCR product enabled the detection of the resistant T-G mismatch as a cleavable C-A mismatch. Consequently, a modified protocol was introduced, in which both the

normal and the patient PCR products were labelled (as suggested by S. Forrest, J. Saleeba & R. Cotton, personal communication). Furthermore, in the modified protocol the normal and the patient PCR products were labelled on one strand only at a time by performing the PCRs with one of the PCR primers 5'-end-labelled. In addition to the normal probe homoduplexes and the normal-patient heteroduplexes, patient-self duplexes were also analyzed. This enabled the determination of whether the patient was homo- or heterozygous for mismatches.

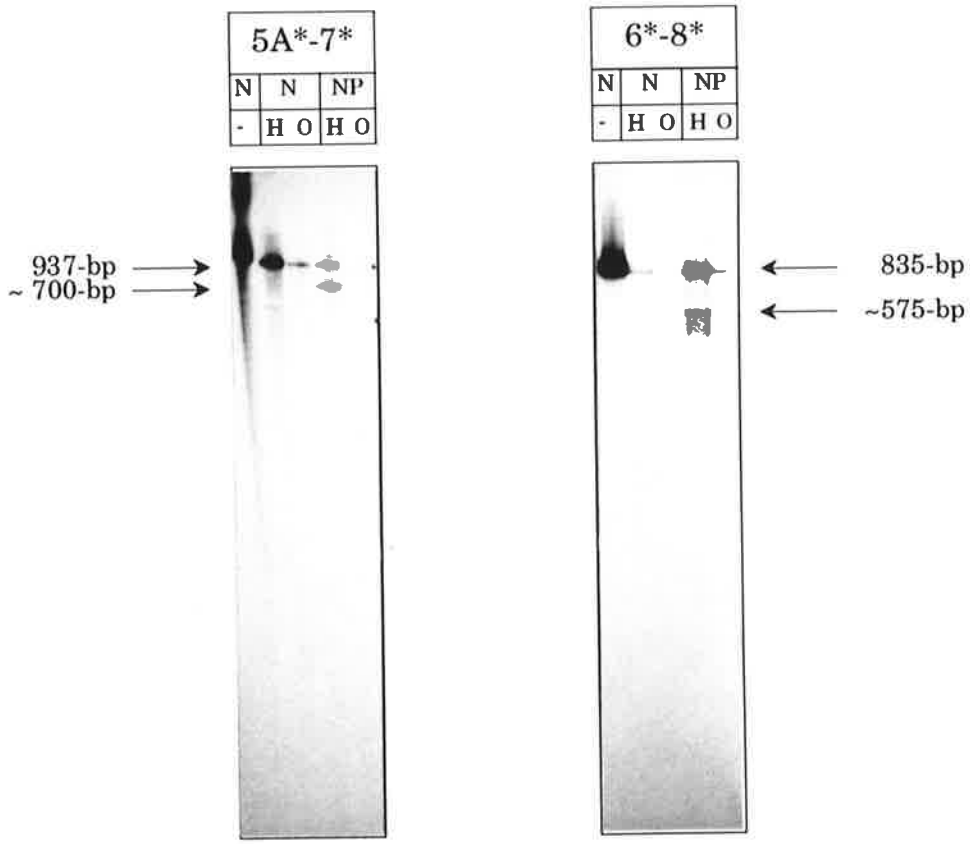
4.5.2a SF1022

The chemical cleavage pattern obtained from patient SF1022 is shown in Figure 4.3. The initial protocol of the chemical cleavage technique was performed on patient SF1022, in which the probe was generated from the modified 4S cDNA clone only, and was ^{32}P-labelled in both strands. Heteroduplexes between the PCR A product of the patient and the probe were specifically cleaved to produce a fragment of about 700-bp in the hydroxylamine reaction, which indicated that the mismatch was in the vicinity of nucleotide position 250 or 700. The location was ambiguous, as the probe was labelled in both strands and therefore it was not possible to determine from which end the labelled fragment was derived from. However, chemical cleavage performed on heteroduplexes between control and patient SF1022 PCR products (4SP16-4SP14) which overlapped the 5'-end of PCR A indicated that the mismatch was at about nucleotide position 250 (data not shown). PCR B heteroduplexes were specifically cleaved to produce a fragment of about 575-bp in the hydroxylamine reaction, which indicated that the mismatch was in the vicinity of nucleotide position 1050 or 1370.

In order to define the mismatch in PCR A of patient SF1022, a PCR product spanning the sequence around nucleotide 250 was generated from both the patient's cDNA and a normal control, with the primers 4SP16 and 4SP11. Direct DNA sequencing of both PCR products using end-labelled 4SP11 showed that the

FIGURE 4.3 CHEMICAL CLEAVAGE ANALYSIS OF PATIENT SF1022

Autoradiograph of the chemical cleavage analysis of the PCR A (5A-7) and PCR B (6-8) products from patient SF1022 on a denaturing polyacrylamide gel. The PCR A and B probes were generated from the modified wild type 4S cDNA clone (N), and were ^{32}P-labelled (*) at the 5'-end in both strands. Unlabelled PCR A and B products were also generated from the patient cDNA. Probe homoduplexes (N) and probe-patient heteroduplexes (NP) were subjected to hydroxylamine (H) treatment for 45 min and osmium tetroxide (O) treatment for 5 min. The homoduplex controls show only the presence of the uncleaved probes (937-bp and 835-bp) while the probe-patient reactions show a cleavage band of about 700-bp in the PCR A product and a cleavage band of about 575-bp in the PCR B product, both in the hydroxylamine-modified reactions. The homoduplex controls not subjected to chemical modification and cleavage are also shown (N-).


5A*-7*
N
N
NP
-
H
O
H
O
937-bp
~ 700-bp
6*-8*
N
N
NP
-
H
O
H
O
835-bp
~575-bp

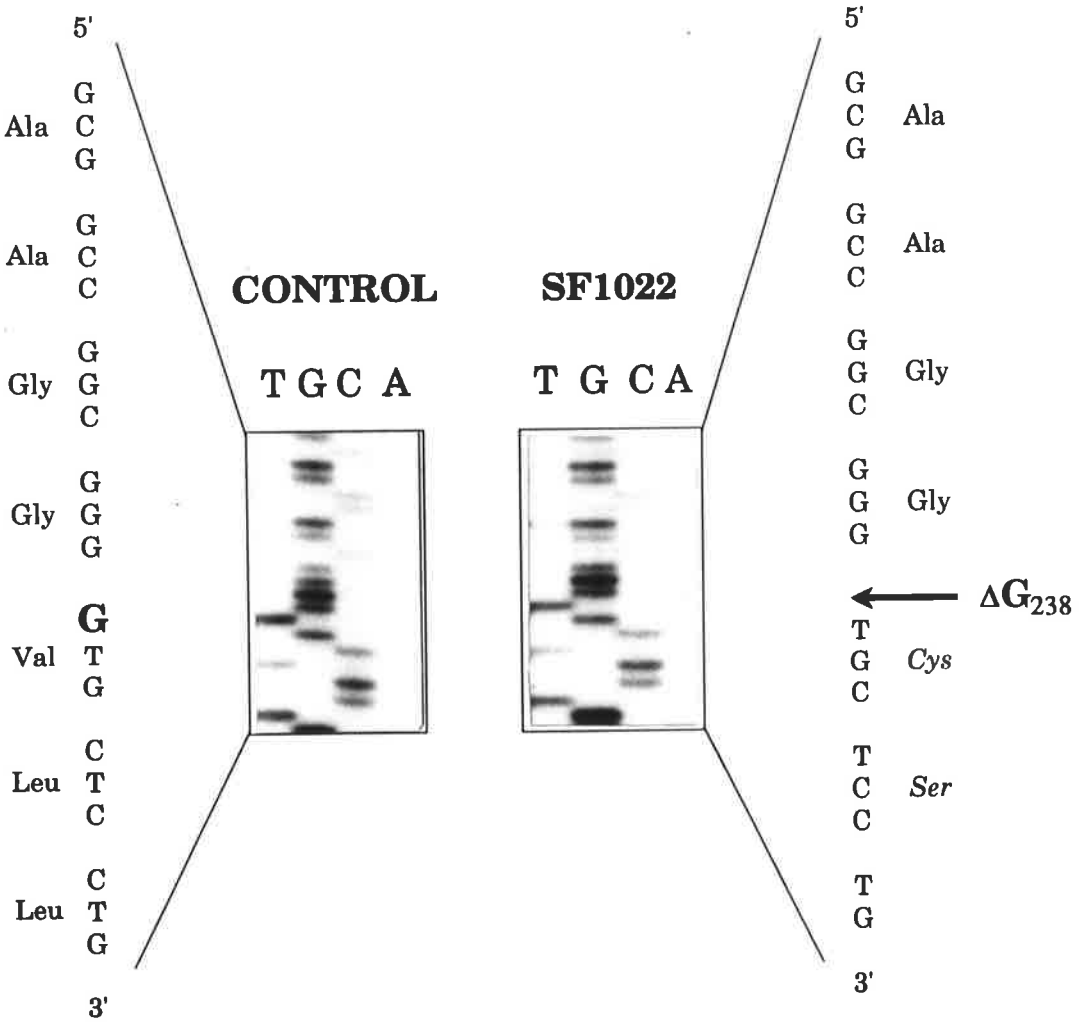
patient was homozygous for the deletion of a G at nucleotide position 238 (ΔG_{238}) (Fig. 4.4). This result was confirmed by determining the sequence in the complementary strand using end-labelled 4SP5A, and by direct DNA sequencing of the PCR product 4SP16-4SP14 synthesized using genomic DNA from both the patient and a normal control (Fig. 4.4). The ΔG_{238} mutation was consistent with and accounted for the observed 700-bp fragment in the hydroxylamine reaction. In addition, the ΔG_{238} mutation was the only nucleotide difference detected after sequencing PCR amplified DNA between 4SP16 and 4SP11 from the control or the patient cDNA.

The ΔG_{238} mutation was predicted to cause a frameshift that results in a truncated 4S polypeptide with an altered amino acid sequence from amino acid 80 (Cys) until an in-frame premature termination codon at codon 113 (TAA), relative to the normal 4-sulfatase reading frame of 533 amino acids (Fig. 4.5). It is most likely that this mutation does not result in the production of enzymically active 4S. The correlation between this mutation and the observed clinical and biochemical phenotype will be discussed in Section 4.6. Nonsense mutations, either direct or as a result of frameshift, have been observed to result in a low mRNA phenotype (e.g. dihydrofolate reductase, Urlaub *et al.*, 1989). Whether the ΔG_{238} mutation reduces the level of 4S mRNA is not known, however cDNA-PCR amplification of the 4S mRNA from patient SF1022 generated products with a yield similar to those derived from normal controls.

ASO analysis was used to independently verify the homozygous ΔG_{238} mutation in SF1022, as well as determine the frequency of this mutation in the MPS-VI population (data not shown). The 4SP5-4SP14 PCR product was generated from the genomic DNA of a normal control, SF1022, and another 25 unrelated MPS-VI patients. The PCR products were subsequently screened with allele-specific

FIGURE 4.4 DNA SEQUENCE OF THE ΔG_{238} MUTATION

Nucleotide and deduced amino acid sequence of the 4S cDNA from the control 4S cDNA clone, and the MPS-VI patient SF1022 who was homozygous for the ΔG_{238} mutation. The arrow shows the deletion of a G nucleotide at position 238 in the patient. The G deletion (in bold) and the corresponding shift in the amino acid reading frame (italics) are shown. The DNA sequence was obtained from the clone and the patient by direct DNA sequencing of the 4SP16-4SP11 PCR product with labelled 4SP11.


5'
Ala
Ala
Gly
Gly
Val
Leu
Leu
3'
CONTROL
T G C A
SF1022
T G C A
5'
Ala
Ala
Gly
Gly
ΔG238
Cys
Ser
3'

oligonucleotide probes. The ΔG_{238} mutation was only found in patient SF1022, indicating that the ΔG_{238} mutation was a rare and possibly private MPS-VI allele.

In order to define the mismatch site in PCR B of patient SF1022, two PCR products were designed which spanned either of the two alternative locations. The 4SP24-4SP28 PCR included the predicted site at nucleotide position 1050, while the 4SP27-4SP8 PCR included the alternative predicted site at nucleotide position 1370. No sequence differences were observed between patient SF1022 and the normal control when the 4SP27-4SP8 PCR products were directly sequenced using end-labelled 4SP27. However, two sequence differences were observed when the 4SP24-4SP28 PCR products were directly sequenced using end-labelled 4SP24 and 4SP28 (data not shown). Patient SF1022 was homozygous for a G to A transition at nucleotide 1072 ($G_{1072}A$), which resulted in a valine to methionine substitution at amino acid 358 (V358M). The $G_{1072}A$ transition was consistent with and accounted for the observed 575-bp fragment in the hydroxylamine reaction. Patient SF1022 was also found to be homozygous for a silent A to G transition at nucleotide 1191 ($A_{1191}G$), in the third base of a proline codon at amino acid 397. This transition was predicted to generate a 439-bp fragment in the osmium tetroxide reaction due to cleavage at the T-G mismatch. However the fragment was not observed, presumably due to the sequence context surrounding the mismatched T which rendered it refractory to modification (Forrest *et al.*, 1991). The $A_{1191}G$ transition was subsequently detected by the modified chemical cleavage protocol as a C-A mismatch, in analyzes of patients SF913, SF1246 and SF2467 (Sections 4.5.2b-d).

The silent $A_{1191}G$ transition was considered a polymorphism as the Pro397 residue was not altered, and as it had been reported in another published 4S cDNA clone (Schuchman *et al.*, 1990). The V358M substitution was also considered a polymorphism as a normal individual was found who was homozygous for this

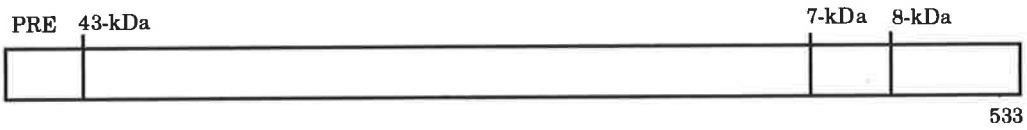
FIGURE 4.5 THE FRAMESHIFT RESULTING FROM ΔG_{238}

a. The cDNA sequence of the ΔG_{238} allele from patient SF1022 showing the nucleotide sequence and the deduced amino acid sequence of the truncated and frameshifted protein. The ΔG_{238} mutation (Δ) is predicted to generate a mutant 4S polypeptide 112 amino acids in length, of which the first 79 amino acids are colinear with the normal 4S sequence, followed by 33 frameshifted amino acids (underlined) at the carboxyl-terminal end. The termination codon is marked by an asterisk.

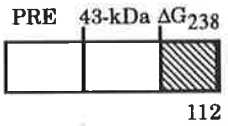
b. Comparison of the 4S cDNA reading frame in the wild type and the mutant (ΔG_{238}). The frameshifted amino acids are denoted by the hatched box.

```
 M  G  P  R  G  A  A  S  L  P  R  G  P  G  P  R  R  L  L  L        20
ATGGGTCCGCGCGGCGCGGCGAGCTTGCCCCGAGGCCCCGGACCTCGGCGGCTGCTCCTC      60

 P  V  V  L  P  L  L  L  L  L  L  A  P  P  G  S  G  A  G         40
CCCGTCGTCCTCCCGCTGCTGCTGCTGCTCTTGTTGGCGCCGCCGGGCTCGGGCGCCGGG      120

 A  S  R  P  P  H  L  V  F  L  L  A  D  D  L  G  W  N  D  V      60
GCCAGCCGGCCGCCCCACCTGGTCTTCTTGCTGGCAGACGACCTAGGCTGGAACGACGTC      180

 G  F  H  G  S  R  I  R  T  P  H  L  D  A  L  A  A  G  G  C      80
GGCTTCCACGGCTCCCGCATCCGCACGCCGCACCTGGACGCGCTGGCGGCCGGCGGGTGC      240
                                                      Δ
 S  W  T  T  T  T  R  S  R  C  A  R  R  R  G  A  S  C  S  L      100
TCCTGGACAACTACTACACGAAGCCGCTGTGCACGCCGTCGCGGAGCCAGCTGCTCACTG      300

 A  A  T  R  S  V  Q  V  Y  S  T  K  *                           112
GCCGCTACCAGATCCGTACAGGTTTACAGCACCAAATAATCTGGCCCTGTCAGCCCAGCT      360
```

4S Normal reading frame:



4S Mutant reading frame:



50 amino acids

substitution (data not shown). In addition, the V358M substitution is a fairly conservative exchange, as both side chains are hydrophobic and uncharged, and neither residue was conserved between sulphatases (Fig. 3.13). Independently, the V358M change was recently described as a polymorphism by Jin *et al.* (1991).

Therefore, the homozygous ΔG_{238} allele was the only candidate mutation identified in patient SF1022. The correlation between the severe clinical and biochemical phenotype observed in patient SF1022 and the predicted phenotype of the ΔG_{238} mutation is presented in Section 4.6.

4.5.2b SF1246

The chemical cleavage pattern obtained from patient SF1246 is shown in Figure 4.6 . The modified protocol of the chemical cleavage technique was performed, in which the PCR probes were generated from both the normal 4S cDNA clone and the patient cDNA, and the probes were individually ^{32}P-labelled in a single strand only. In addition to the normal probe homoduplexes and the normal-patient heteroduplexes, patient-self duplexes were also analyzed, in order to determine whether the patient was homo- or heterozygous for mismatches. As the PCR products in each reaction were labelled at one end only, the position of the mismatched base could be unambiguously defined.

PCR A duplexes, end-labelled in the 4SP7-primed strand, were specifically cleaved to produce a fragment of about 660-bp in the hydroxylamine reaction, indicating that the mismatch was approximately at nucleotide position 280. The 660-bp fragment band was only observed in the patient duplex; it was too weak to be observed in the normal-patient heteroduplexes against the background cleavage signal. The observed self-cleavage indicated that the patient was heterozygous for this mismatch. PCR B duplexes, end-labelled in the 4SP6-primed strand, were specifically cleaved to produce fragments of about 380-bp in both the

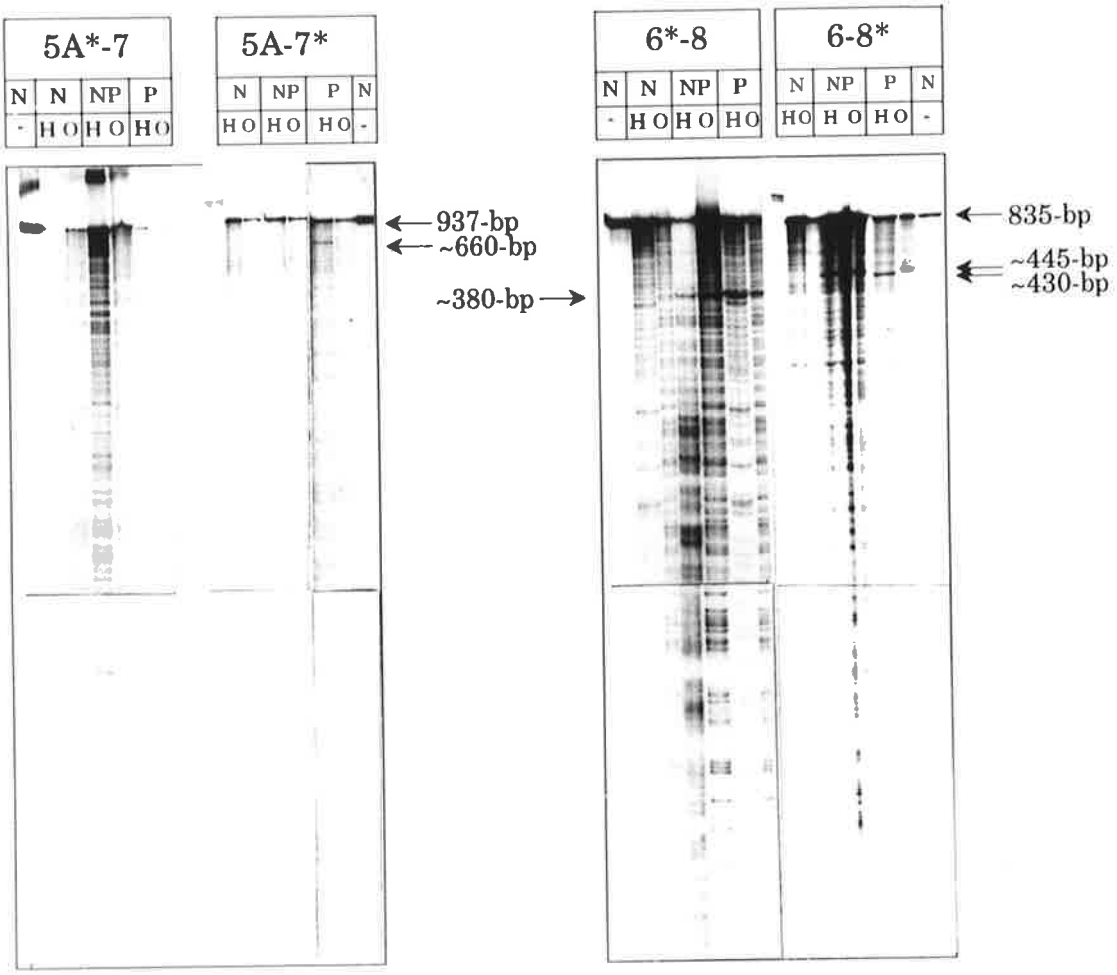
hydroxylamine and osmium tetroxide reactions, indicating that the mismatch was in the vicinity of nucleotide position 1180. The patient was heterozygous for this mismatch as self-cleavage was observed when the PCR B probe derived from the patient was duplexed with itself. PCR B duplexes, end-labelled in the 4SP8-primed strand, were specifically cleaved to produce two fragments. The first fragment was observed in the osmium tetroxide reaction and was about 445-bp in length, indicating that the mismatch was in the vicinity of nucleotide position 1185. The second fragment was observed in both the hydroxylamine and the osmium tetroxide reactions and was about 430-bp in length, indicating that the mismatch was in the vicinity of nucleotide position 1200. Patient 1246 was heterozygous for both mismatches as self-cleavage was observed.

In order to define the mismatch in the vicinity of nucleotide position 280 in PCR A of patient SF1246, the product from the patient and the normal control were directly sequenced with end-labelled 4SP11. Patient SF1246 was found to be heterozygous for a G to A transition at nucleotide 284 ($G_{284}A$) which resulted in a arginine to glutamine substitution at amino acid 95, R95Q (Fig. 4.7). The $G_{284}A$ transition was confirmed in the complementary strand by DNA sequencing with end-labelled 4SP17. The $G_{284}A$ transition was consistent with and accounted for the observed 660-bp fragment in the hydroxylamine reaction. The $G_{284}A$ transition also predicted a 660-bp fragment in the osmium tetroxide reaction due to cleavage at the T-G mismatch, however the fragment was not observed, presumably due to the sequence context surrounding the mismatched T rendering it refractory to modification (Forrest *et al.*, 1991).

ASO analysis was used to independently verify the heterozygous R95Q allele in SF1246, as well as determine the frequency of this allele in the normal and MPS-VI population. The 4SP16-4SP14 PCR product was amplified from the genomic DNA from a panel of 20 normal (non-MPS-VI) and 25 unrelated MPS-VI patients; the

FIGURE 4.6 CHEMICAL CLEAVAGE ANALYSIS OF PATIENT SF1246

Autoradiograph of the chemical cleavage analysis of the PCR A (5A-7) and PCR B (6-8) products from patient SF1246 on a denaturing polyacrylamide gel. The PCR A and B probes were generated from the modified wild type 4S cDNA clone (N) and from patient cDNA (P). Each PCR probe was individually ^{32}P-labelled (*) at the 5'-end in one stand only. Consequently, a total of eight probes were synthesized. Homoduplexes derived from the wild type (N), patient-self duplexes (P), and wild type-patient heteroduplexes (NP), were subjected to hydroxylamine (H) treatment for 45 min and osmium tetroxide (O) treatment for 5 min. The wild type homoduplex controls show only the presence of the uncleaved probes (937-bp and 835-bp). The patient-self duplexes show a cleavage band of about 660-bp in the PCR A product labelled in the 4SP7-primed strand after modification with hydroxylamine. Patient-self duplexes and wild type-patient heteroduplexes also show a cleavage band of about 380-bp in the PCR B product labelled in the 4SP6-primed strand after modification with hydroxylamine and osmium tetroxide. Patient-self duplexes and wild type-patient heteroduplexes of PCR B labelled in the 4SP8-primed strand also show cleavage bands of about 430-bp (hydroxylamine) and 445-bp (hydroxylamine and osmium tetroxide). The homoduplex controls not subjected to chemical cleavage analysis are also shown (N-). Unmarked lanes between the chemical cleavage reactions correspond to end-labelled SPP1 *Eco*RI molecular size markers.


5A*-7
N N NP P
- HO HO HO
5A-7*
N NP P N
HO HO HO -
6*-8
N N NP P
- HO HO HO
6-8*
N NP P N
HO HO HO -
937-bp
~660-bp
~380-bp
835-bp
~445-bp
~430-bp

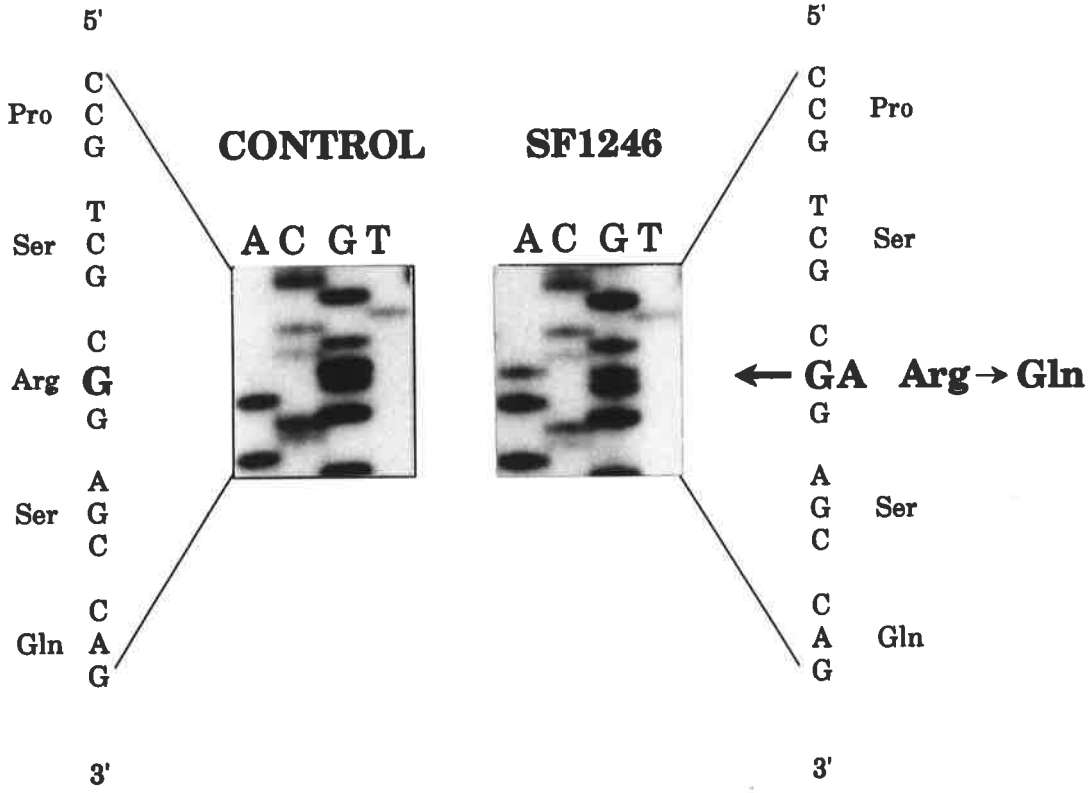
products were screened with allele-specific oligonucleotide probes. Two unrelated MPS-VI patients were heterozygous for the R95Q allele (2/50 alleles), patients SF1246 and SF913 (data not shown). The sibling of patient SF913 (SF912) was also heterozygous for the R95Q allele. The R95Q allele was not found in the 40 normal alleles. Patient SF913 was also independently shown to be heterozygous for the R95Q allele by chemical cleavage and DNA sequence analysis (Section 4.5.2c). The predicted R95Q substitution is most likely to be a mutation as it is a non-conservative change from a basic to a neutral and polar amino acid side-chain at a residue that was totally conserved between 10 different sulphatases (Fig. 3.13).

In order to define the mismatches between nucleotide positions 1180 and 1200 in PCR B of patient SF1246, the 4SP24-4SP28 PCR product was amplified from the cDNA of the patient and the normal control, and the DNA sequenced with end-labelled 4SP28. Patient SF1246 was found to be heterozygous for a A to C transversion at nucleotide 1178 ($A_{1178}C$) which resulted in a histidine to proline substitution at amino acid 393, H393P (Fig. 4.8). The $A_{1178}C$ transversion was consistent with and accounted for the observed 380-bp fragment in the hydroxylamine reaction and the 445-bp fragment in the osmium tetroxide reaction. The 380-bp fragment observed in the osmium tetroxide reaction was not predicted; however as the $A_{1178}C$ transversion was the only base change found, the 380-bp fragment is presumably due to a T nucleotide whose susceptibility to osmium tetroxide modification has been induced by the nearby mismatched C_{1178}. The patient was also found to be heterozygous for the silent $A_{1191}G$ polymorphism. The $A_{1191}G$ polymorphism was responsible for the 430-bp fragments observed in both the hydroxylamine and osmium tetroxide reactions.

The heterozygous H393P allele in SF1246 was confirmed as a mutation and the frequency determined in the normal and MPS-VI population by ASO analysis (data not shown). The 4SP31-4SP32 PCR product was amplified from the genomic DNA

FIGURE 4.7 DNA SEQUENCE OF THE R95Q SUBSTITUTION MUTATION

Nucleotide and deduced amino acid sequence of the 4S cDNA from the control 4S cDNA clone, and the MPS-VI patient SF1246 who was heterozygous for the R95Q mutation. The arrow shows the G to A nucleotide transition at position 284 in the patient. The DNA sequence was obtained from the clone and the patient by direct DNA sequencing of the PCR A product with end-labelled 4SP11.


5'
Pro
C
C
G
Ser
T
C
G
Arg
C
G
G
Ser
A
G
C
Gln
C
A
G
3'
CONTROL
A C G T
SF1246
A C G T
5'
C
C
G
Pro
T
C
Ser
G
C
GA
Arg → Gln
G
A
G
Ser
C
C
A
Gln
G
3'

of a panel of 20 normal and 25 unrelated MPS-VI patients, and the products were screened with allele specific oligonucleotide probes. Four of the unrelated MPS-VI patients were heterozygous for the H393P allele (4/50 alleles); SF1246, SF693, SF839 and SF2984. The H393P allele was not found in the 40 normal alleles. The H393P substitution is a non-conservative change from a basic to a neutral non-polar amino acid side-chain, at a residue that was not conserved between sulphatases (Fig. 3.13). The peptide region containing the H393 residue had a propensity to form a potential α-helical structure, which was abolished when P393 was substituted, according to the Chou-Fasman algorithm for the prediction of protein secondary structure (CHOFAS program: Chapter 2, Section 2.2.18). Taken together, the analyzes predict that the H393P substitution is likely to be a mutation.

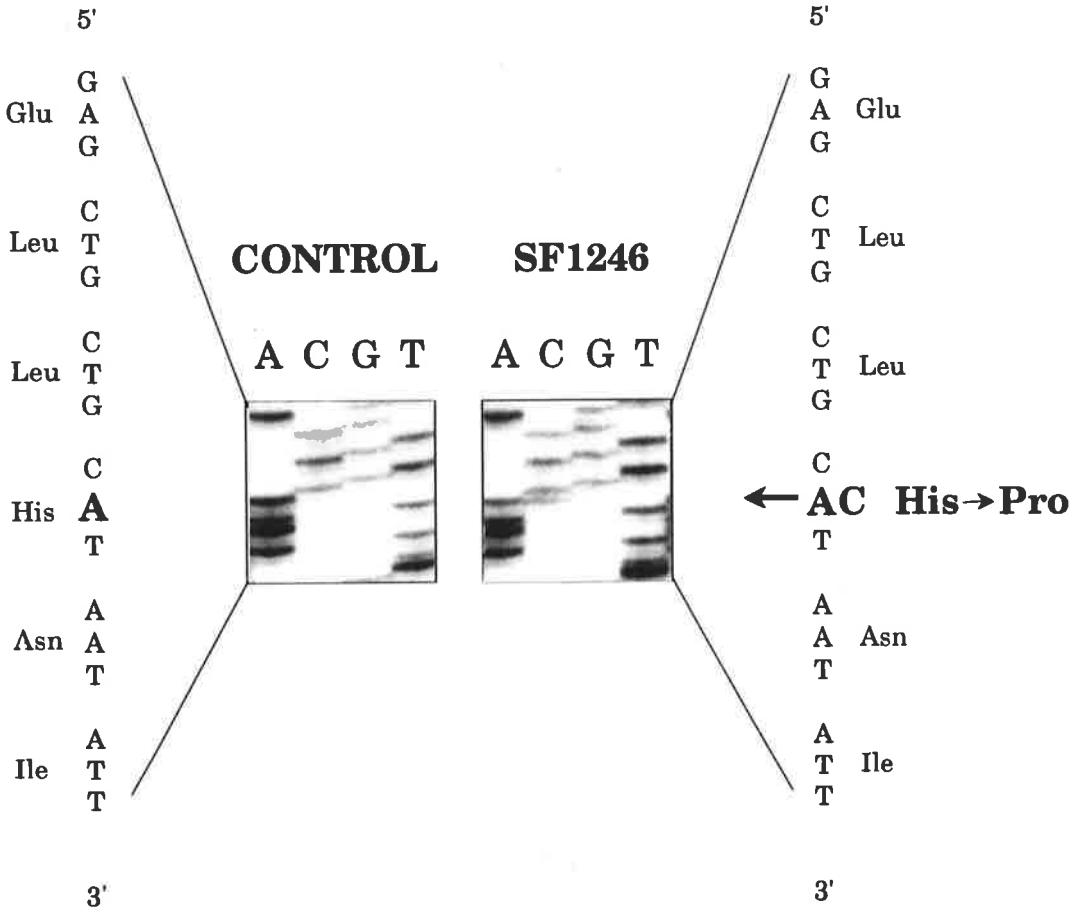
Patient SF1246 is a compound heterozygote of the two candidate mutant 4S alleles R95Q and H393P. The correlation between the severe clinical and biochemical MPS-VI phenotype observed in patient SF1246 and the predicted phenotypes of the R95Q and the H393P mutations is presented in Section 4.6. The effect of the R95Q and H393P alleles on 4S structure and function is discussed in Chapter 5.

4.5.2c SF913

The modified chemical cleavage protocol was used to detect base changes in patient SF913, as shown in Figure 4.9. PCR A duplexes, end-labelled in the 4SP7-primed strand, were specifically cleaved to produce two fragments of about 660-bp and 310-bp in the hydroxylamine reaction. These fragments indicated that the mismatches were approximately at nucleotide positions 280 and 630 respectively. Both fragments were only observed in the patient duplex; they were too weak to be observed in the normal-patient heteroduplexes against the background cleavage signal. The patient was heterozygous for both mismatches as self-cleavage was observed. PCR B duplexes, end-labelled in the 4SP8-primed strand, were

FIGURE 4.8 DNA SEQUENCE OF THE H393P SUBSTITUTION MUTATION

Nucleotide and deduced amino acid sequence of the 4S cDNA from the control 4S cDNA clone, and the MPS-VI patient SF1246 who was heterozygous for the H393P mutation. The arrow shows the A to C nucleotide transversion at position 1178 in the patient. The DNA sequence was obtained from the clone and the patient by direct DNA sequencing of the 4SP24-4SP28 PCR product with end-labelled 4SP28.


5'
Glu
G
A
G
Leu
C
T
G
Leu
C
T
G
His
C
A
T
Asn
A
A
T
Ile
A
T
T
3'
CONTROL
SF1246
A C G T
A C G T
5'
G
A
G
Glu
C
T
G
Leu
C
T
G
Leu
C
AC His→Pro
T
A
A
T
Asn
A
T
T
Ile
3'

specifically cleaved to produce a fragment of about 430-bp in the hydroxylamine reaction, indicating that the mismatch was approximately at nucleotide position 1200. However, the patient was homozygous for this mismatch as self-cleavage was not observed when the PCR B probe derived from the patient was duplexed with itself.

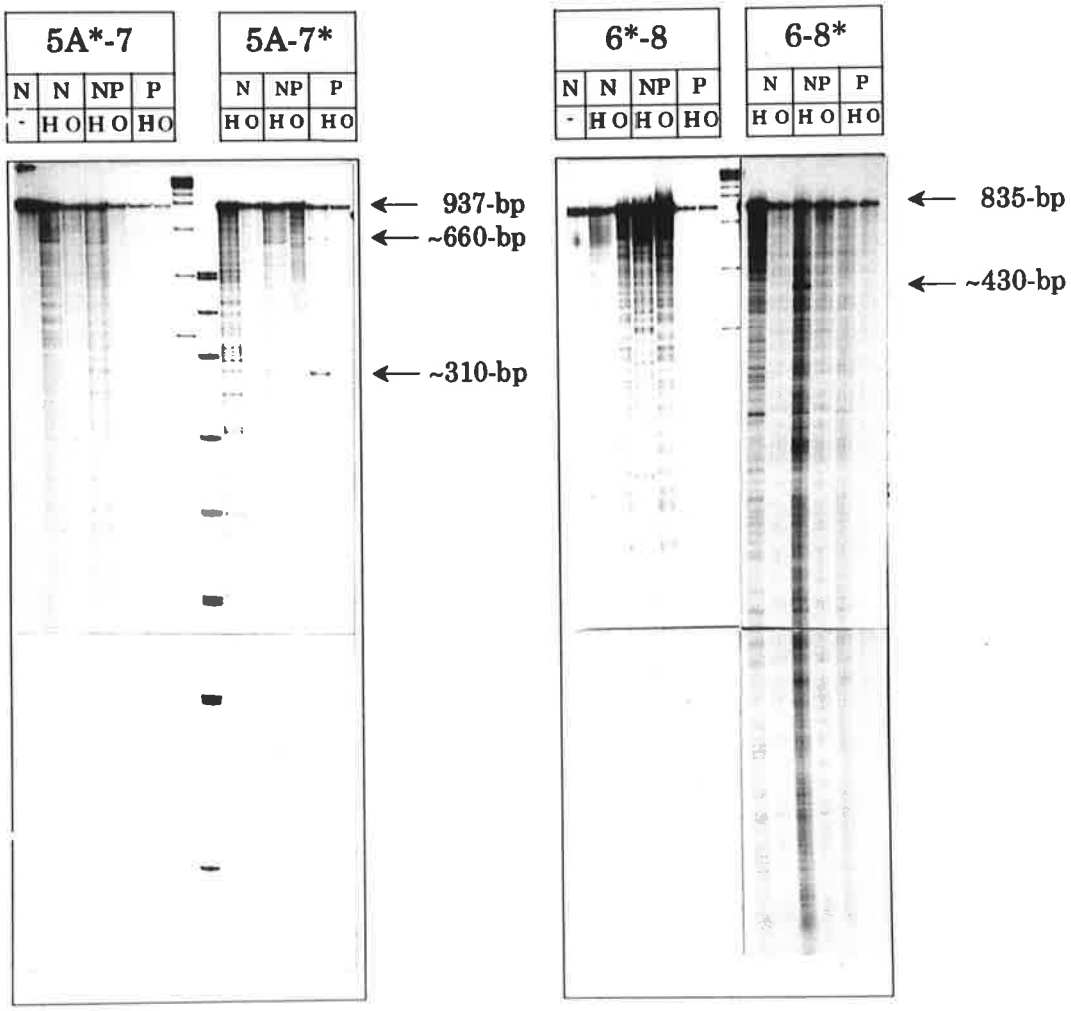
In order to define the mismatch in the vicinity of nucleotide position 280 in PCR A of patient SF913, the product from the patient and the normal control were directly sequenced with end-labelled 4SP17 and 4SP11. Patient SF913 was found to be heterozygous for the $G_{284}A$ transition which resulted in the R95Q substitution, which was also observed in patient SF1246. The R95Q substitution has been discussed previously in Section 4.5.2b.

In order to define the mismatch in the vicinity of nucleotide position 630 in PCR A of patient SF913, the product from the patient and the normal control were directly sequenced using end-labelled 4SP22. Patient SF913 was found to be heterozygous for a A to G transition at nucleotide 629 ($A_{629}G$), which resulted in a tyrosine to cysteine substitution at amino acid 210, Y210C. The $A_{629}G$ transition was confirmed in the complementary strand by DNA sequencing with end-labelled 4SP30 (data not shown). The $A_{629}G$ transition was consistent with and accounted for the observed 310-bp fragment in the hydroxylamine reaction. The $A_{629}G$ transition also predicted a 310-bp fragment in the osmium tetroxide reaction due to cleavage at the T-G mismatch. This fragment was not observed, again presumably due to the unreactive nature of many T-G mismatches.

ASO analysis was used to independently verify the heterozygous Y210C allele in SF913, as well as determine the frequency of this allele in the normal and MPS-VI population. The 4SP22-4SP30 PCR product was amplified from the panel of normal and unrelated MPS-VI genomic DNA samples as described previously, and

FIGURE 4.9 CHEMICAL CLEAVAGE ANALYSIS OF PATIENT SF913

Autoradiograph of the chemical cleavage analysis of the PCR A (5A-7) and PCR B (6-8) products from patient SF913 on a denaturing polyacrylamide gel. The PCR A and B probes were generated from the modified wild type 4S cDNA clone (N) and from patient cDNA (P). Each PCR probe was individually ^{32}P-labelled (*) at the 5'-end in one stand only. Consequently, a total of eight probes were synthesized. Homoduplexes derived from the wild type (N), patient-self duplexes (P), and wild type-patient heteroduplexes (NP), were subjected to hydroxylamine (H) treatment for 45 min and osmium tetroxide (O) treatment for 5 min. The wild type homoduplex controls show only the presence of the uncleaved probes (937-bp and 835-bp). The patient-self duplexes show cleavage bands of about 660-bp and 310-bp in the PCR A product labelled in the 4SP7-primed strand after modification with hydroxylamine. Wild type-patient heteroduplexes show a cleavage band of about 430-bp in the PCR B product labelled in the 4SP8-primed strand after modification with hydroxylamine. The homoduplex controls not subjected to chemical cleavage analysis are also shown (N-). Unmarked lanes between the chemical cleavage reactions correspond to end-labelled SPP1 *Eco*RI and pUC18 *Hpa*II molecular size markers.


5A*-7
N N NP P
- H O H O H O
5A-7*
N NP P
H O H O H O
6*-8
N N NP P
- H O H O H O
6-8*
N NP P
H O H O H O
937-bp
~660-bp
~310-bp
835-bp
~430-bp

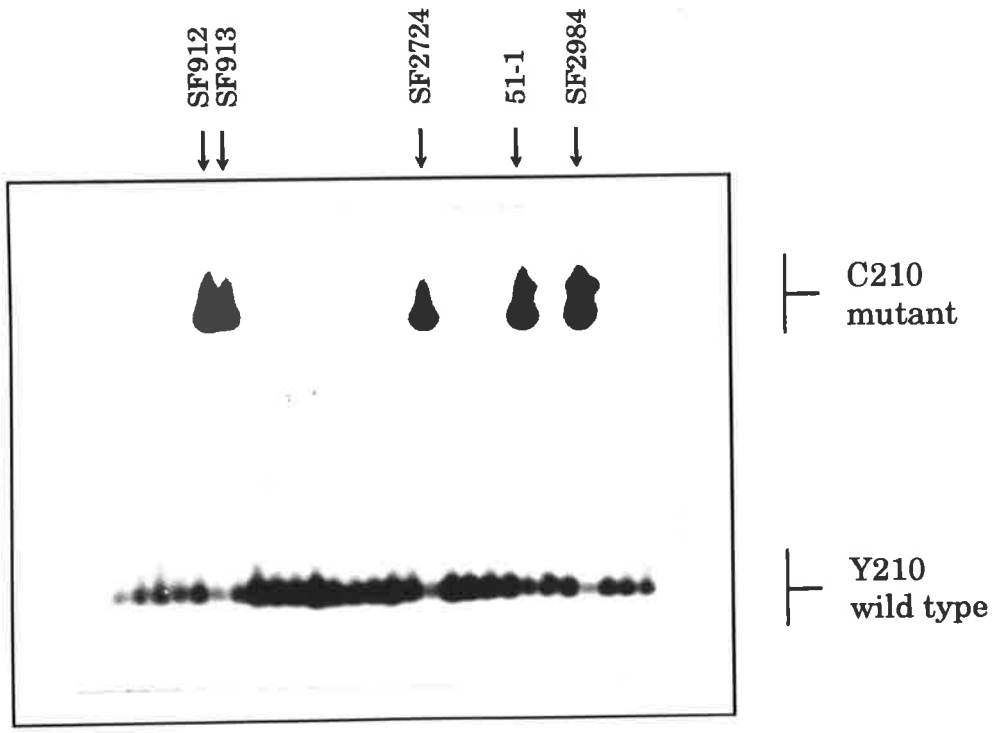
was screened with allele-specific oligonucleotide probes. Four unrelated MPS-VI patients were heterozygous for the Y210C allele (4/50 alleles); SF913, SF2724, SF2984 and 51-1 (Fig. 4.10). SF912 who was a sibling of SF913, was also heterozygous for the Y210C allele. The Y210C allele was not found in the 40 normal alleles. Taken together, these results suggested that the Y210C substitution was likely to be a mutation. The predicted Y210C substitution is a non-conservative change in the amino acid side-chain from a large aromatic species to a smaller sulphydryl group, at a residue that was not conserved between 10 different sulphatases (Fig. 3.13)

The mismatch in the vicinity of nucleotide position 1200 in PCR B of patient SF913 was defined by direct DNA sequencing of PCR products. The 4SP24-4SP28 PCR product was amplified from the cDNA of the patient and the normal control, and DNA sequenced with end-labelled 4SP28. The patient was found to be homozygous for the silent $A_{1191}G$ polymorphism in the third nucleotide of the P397 codon. The $A_{1191}G$ polymorphism has been discussed previously in Section 4.5.2a. The predicted 430-bp osmium tetroxide fragment was not observed in chemical cleavage analysis of SF913 but was observed in SF1246, the observed variable susceptibility of the T-G mismatch was considered to be due to inherent variation in the execution of experimental procedures.

The MPS-VI patient siblings SF913 and SF912 are compound heterozygotes of the two candidate mutant 4S alleles R95Q and Y210C. The correlation between the milder clinical and biochemical MPS-VI phenotype observed in these patients and the predicted phenotypes of the R95Q and the Y210C mutations is presented in Section 4.6. In Chapter 5, the effect of the R95Q and Y210C alleles on 4S structure and function is discussed.

FIGURE 4.10 ASO DETECTION OF THE Y210C MUTATION

The Y210C mutation was detected with the ASOs 4STyr210 (wild type) and 4SCys210 (mutant) as described in Section 4.5.2c. The 4SP22-4SP30 PCR product was amplified from the genomic DNA of 26 MPS-VI patients (25 unrelated) and 20 normal individuals (latter not shown). The PCR products were electrophoresed in duplicate and transferred to nylon membranes, as described in Section 4.2.5. The duplicate filters were hybridized separately with each of the ASOs. The washing conditions were as described in Section 4.2.5 and Appendix D. Patients SF912, SF913, SF2724, SF2984, and 51-1 were heterozygous for the mutant C210 allele.


SF912
SF913
SF2724
51-1
SF2984
C210
mutant
Y210
wild type

4.5.2d SF2467

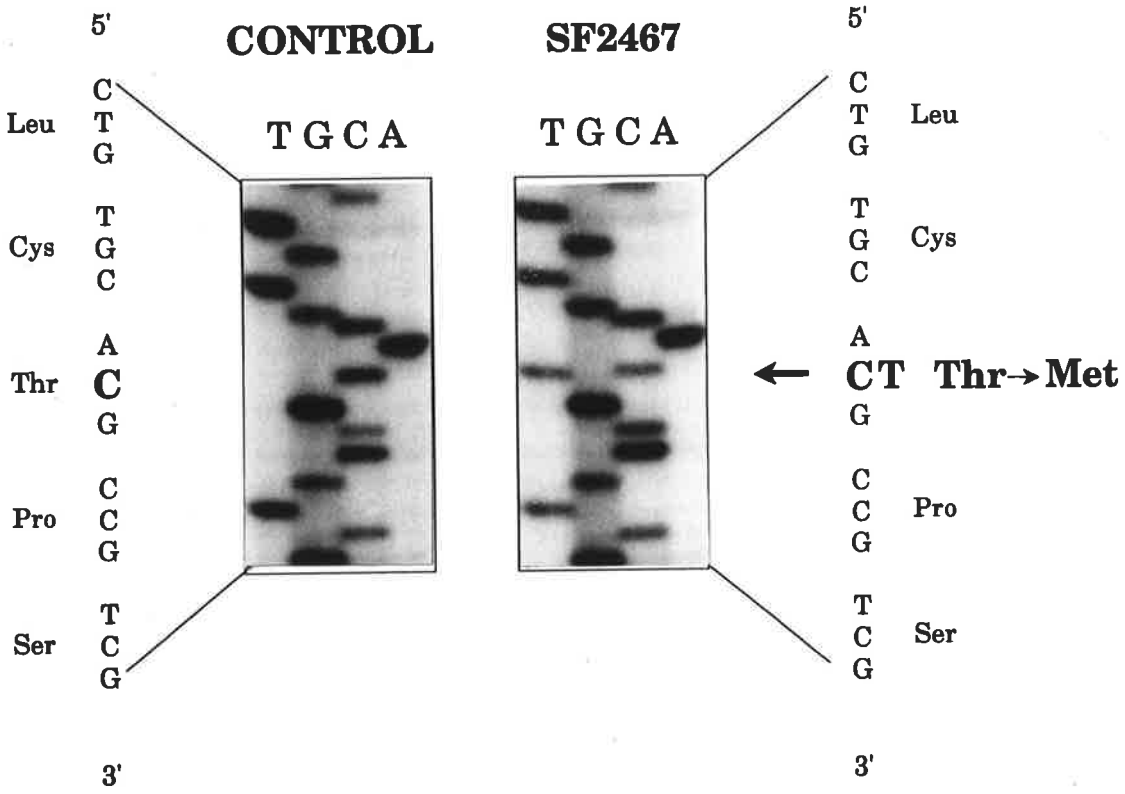
The 4S cDNA obtained from patient SF2467 was analyzed using the modified chemical cleavage procedure (data not shown). PCR A duplexes, end-labelled in the 4SP5A-primed strand, were specifically cleaved to produce a fragment of about 270-bp in the hydroxylamine reaction, which indicated that the mismatch was approximately at nucleotide position 270. The observed self-cleavage indicated that the patient was heterozygous for this mismatch. PCR B duplexes, end-labelled in the 4SP6-primed strand, were specifically cleaved to produce a fragment of about 700-bp in the hydroxylamine reaction, indicating that the mismatch was in the vicinity of nucleotide position 1495. A fragment of about 430-bp was observed in the hydroxylamine reaction of PCR B duplexes end-labelled in the 4SP8-primed strand, which indicated that the mismatch was in the vicinity of nucleotide position 1200. Patient 2467 was heterozygous for both mismatches in the PCR B product.

In order to define the mismatch in the vicinity of nucleotide position 270 in PCR A, the product from SF2467 and the normal control were directly sequenced with end-labelled 4SP17. Patient SF2467 was found to be heterozygous for a C to T transition at nucleotide 275 ($C_{275}T$), which resulted in a threonine to methionine substitution at amino acid 92, T92M (Fig. 4.11). The $C_{275}T$ transition was confirmed in the complementary strand by DNA sequencing with end-labelled 4SP11. The $C_{275}T$ transition was consistent with and accounted for the observed 270-bp fragment in the hydroxylamine reaction. However, a predicted 270-bp fragment in the osmium tetroxide reaction (T-G mismatch) was not observed. T-G mismatches are generally unreactive, as observed in the previous patient analyses.

ASO analysis was used to independently verify the heterozygous T92M allele in SF2467, as well as determine the frequency of this allele in the normal and MPS-VI population. The 4SP16-4SP14 PCR product was amplified from the genomic DNA of a panel of normal and MPS-VI patients, as described previously. The products

FIGURE 4.11 DNA SEQUENCE OF THE T92M SUBSTITUTION MUTATION

Nucleotide and deduced amino acid sequence of the 4S cDNA from the control 4S cDNA clone, and the MPS-VI patient SF2467 who was heterozygous for the T92M mutation. The arrow shows the C to T nucleotide transition at position 275 in the patient. The DNA sequence was obtained from the clone and the patient by direct DNA sequencing of the PCR A product with end-labelled 4SP17.


CONTROL
SF2467
5'
3'
Leu
Cys
Thr
Pro
Ser
C T G
T G C
A C G
C C G
T C G
T G C A
T G C A
CT Thr→Met

were screened with allele-specific oligonucleotide probes. The T92M substitution was only detected in the MPS-VI patient SF2467, as a heterozygous allele (1/50 alleles), and was not found in the 40 normal alleles (data not shown). This result indicated that the T92M substitution was a rare and possibly private MPS-VI allele. The predicted T92M substitution is likely to be a mutation as it is a non-conservative change from a neutral-polar to a neutral-nonpolar amino acid side-chain at a residue that was conserved in the three human arylsulphatases A, B, and C, and a sea urchin arylsulphatase (Fig. 3.13).

In order to define the mismatch in the vicinity of nucleotide position 1495 in PCR B of patient SF2467, the product from the patient and the normal control were directly sequenced with end-labelled 4SP27. Patient SF2467 was found to be heterozygous for a T to C transition at nucleotide 1493 ($T_{1493}C$), which resulted in a leucine to proline substitution at amino acid 498, L498P (data not shown). The $T_{1493}C$ transition was confirmed in the complementary strand by DNA sequencing with end-labelled 4SP8. The $T_{1493}C$ transversion was consistent with and accounted for the observed 700-bp fragment in the hydroxylamine reaction. Again, a predicted fragment in the osmium tetroxide reaction (700-bp) due to a T-G mismatch was not observed.

ASO analysis was used to independently verify the heterozygous L498P allele in SF2467, as well as determine the frequency of this allele in the normal and MPS-VI population. The 4SP33-4SP8 PCR product was amplified from the genomic DNA of a panel of normal and MPS-VI patients, and was screened with allele-specific oligonucleotide probes. The L498P substitution was only detected in the MPS-VI patient SF2467, as a heterozygous allele (1/50 alleles), and was not found in the 40 normal alleles, which indicated that the L498P substitution was a rare and possibly private MPS-VI allele. The L498P substitution is likely to be a mutation as the predicted secondary structure of the region amino-terminal adjacent to the

L498 residue exhibited a propensity to form a potential α-helical structure, which was abolished when P498 was substituted (data not shown). In addition, the L498P substitution induced a predicted β-turn in the polypeptide. The L498 residue was not conserved between 10 different sulphatases (Fig. 3.13).

In order to define the mismatch in the vicinity of nucleotide position 1200 in PCR B of patient SF2467, the 4SP24-4SP28 PCR product was amplified from the cDNA of the patient and the normal control, and DNA sequenced with end-labelled 4SP28. The patient was found to be heterozygous for the silent $A_{1191}G$ polymorphism, which has been discussed previously in Sections 4.5.2a, b, c.

Patient SF2467 is a compound heterozygote of the two candidate mutant 4S alleles T92M and L498P. The correlation between the milder clinical and biochemical MPS-VI phenotype observed in patient SF1246 and the predicted phenotypes of the T92M and the L498P mutations is presented in Section 4.6. The effect of the T92M and L498P alleles on 4S structure and function is discussed in Chapter 5.

4.5.2e SF368

The initial protocol of the chemical cleavage technique was performed on SF368. No fragments were detectable in the osmium tetroxide reactions as the band intensities were extremely faint due to a technical error (data not shown). PCR A duplexes were specifically cleaved to produce a fragment of about 680-bp and also a cluster of fragments of about 40-bp in the hydroxylamine reaction. PCR B duplexes were specifically cleaved to produce a faint fragment of about 750-bp, and strong fragments of about 500-bp, 450-bp, 400-bp, 380-bp, 290-bp and 120-bp. As described in Section 4.5.1, patient SF368 was heterozygous for a `deletion' of about 200-bp from within the 4S cDNA sequence bounded by nucleotides 309 and 917. The deletion was considered partly responsible for the complex pattern observed in the chemical cleavage. The non-deleted sequence present in the normal strand of

the heteroduplex would not be base-paired to the patient strand, and hence would be more susceptible to chemical modification.

In order to identify the missing sequence, the 4SP10-4SP7 PCR product which spanned the putative deletion was generated from both the patient cDNA and a normal control. In SF368, a PCR product of about 400-bp was observed in addition to the anticipated 624-bp product from the normal allele. Direct DNA sequencing of the 400-bp PCR product from SF368 and the 624-bp PCR product from the normal control using end-labelled 4SP22 showed that the patient was heterozygous for the deletion of nucleotides 691-899 (data not shown). This 208-bp deletion was found to coincide with exon IV (C. Peters, personal communication; Modaressi *et al.*, 1993), however the causative mutation was not determined. The causative mutation may be a deletion of genomic DNA spanning exon IV, or a splice-site mutation, or a nonsense mutation causing exon skipping (Dietz *et al.*, 1993). For the sake of convenience and until the underlying mutation is identified, the mutation will be referred to as the Δexon IV allele. Identification of the underlying mutation would require the genomic DNA sequences of exon IV and the boundaries with the adjacent introns 3 and 4 to be determined. The 208-bp deletion in SF368 was consistent with and accounted for the observed 680-bp and 40-bp fragments in PCR A, and the 750-bp and 120-bp fragments in PCR B.

The Δexon IV mutation is predicted to generate a mutant 4S polypeptide of 267 amino acid residues, about 50% of the length of the wild-type 4S reading frame. The first 230 amino acids of the mutant 4S are collinear with normal 4S, while amino acids 231 to 299 encoded in exon IV are absent. The aberrant splicing of exon III to exon V results in the addition of 37 frameshifted amino acids terminated by an in-frame premature stop codon (TGA) (Fig. 4.12). It is most likely that the Δexon IV mutation does not result in the production of enzymically active

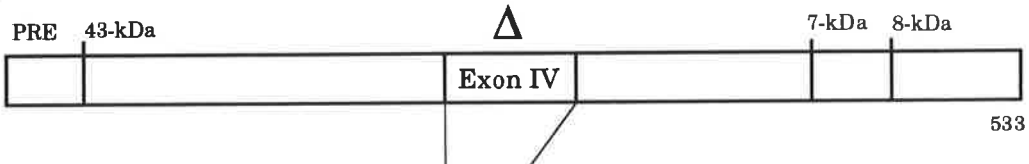
FIGURE 4.12 THE FRAMESHIFT RESULTING FROM ΔEXON IV

a. The cDNA sequence of the Δexon IV allele from patient SF368 showing the nucleotide sequence and the deduced amino acid sequence of the truncated and frameshifted protein. The Δexon IV mutation is predicted to generate an aberrant polypeptide of 267 amino acids in length, compared to the 533 amino acids of the normal 4S reading frame. The amino acid sequence up to and including codon 230 (Lys) of the mutant 4S peptide is collinear with the normal 4S reading frame. The position of the deleted exon IV sequence is indicated (Δ), and the 37 subsequent frameshifted amino acids are underlined. The premature termination codon is marked by an asterisk.

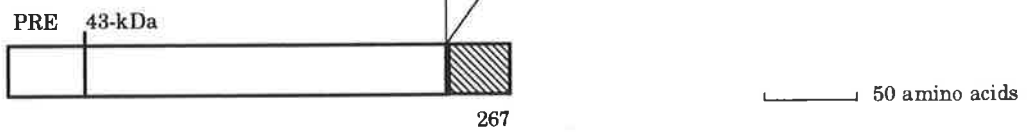
b. Comparison of the 4S cDNA reading frame in the wild type and the mutant (Δexon IV). The frameshifted amino acids are denoted by the hatched box.

```
 M  G  P  R  G  A  A  S  L  P  R  G  P  G  P  R  R  L  L  L        20
ATGGGTCCGCGCGGCGCGGCGAGCTTGCCCCGAGGCCCCGGACCTCGGCGGCTGCTCCTC      60

 P  V  V  L  P  L  L  L  L  L  L  L  A  P  P  G  S  G  A  G        40
CCCGTCGTCCTCCCGCTGCTGCTGCTGCTCTTGTTGGCGCCGCCGGGCTCGGGCGCCGGG     120

 A  S  R  P  P  H  L  V  F  L  L  A  D  D  L  G  W  N  D  V        60
GCCAGCCGGCCGCCCCACCTGGTCTTCTTGCTGGCAGACGACCTAGGCTGGAACGACGTC     180

 G  F  H  G  S  R  I  R  T  P  H  L  D  A  L  A  A  G  G  V        80
GGCTTCCACGGCTCCCGCATCCGCACGCCGCACCTGGACGCGCTGGCGGCCGGCGGGGTG     240

 L  L  D  N  Y  Y  T  Q  P  L  C  T  P  S  R  S  Q  L  L  T       100
CTCCTGGACAACTACTACACGAAGCCGCTGTGCACGCCGTCGCGGAGCCAGCTGCTCACT     300

 G  R  Y  Q  I  R  T  G  L  Q  H  Q  I  I  W  P  C  Q  P  S       120
GGCCGCTACCAGATCCGTACAGGTTTACAGCACCAAATAATCTGGCCCTGTCAGCCCAGC     360

 C  V  P  L  C  E  K  L  L  P  Q  L  L  K  E  A  G  Y  T  T       140
TGTGTTCCTCTGGATGAAAAACTCCTGCCCCAGCTCCTAAAAGAAGCAGGTTATACTACC     420

 H  M  V  G  K  W  H  L  G  M  Y  R  K  E  L  L  P  T  R  R       160
CATATGGTCGGAAAATGGCACCTGGGAATGTACCGGAAAGAATGCCTTCCAACCCGCCGA     480

 G  F  D  T  Y  F  G  Y  L  L  G  S  E  D  Y  Y  S  H  E  R       180
GGATTTGATACCTACTTTGGATATCTCCTGGGTAGTGAAGATTATTATTCCCATGAACGC     540

 C  T  L  I  D  A  L  N  V  T  R  C  A  L  D  F  R  D  G  E       200
TGTACATTAATTGACGCTCTGAATGTCACACGATGTGCTCTTGATTTTCGAGATGGCGAA     600

 E  V  A  T  G  Y  K  N  M  Y  S  T  N  I  F  T  K  R  A  I       220
GAAGTTGCAACAGGATATAAAAATATGTATTCAACAAACATATTCACCAAAAGGGCTATA     660

 A  L  I  T  N  H  P  P  E  K  I  T  E  P  R  L  W  Q  G  V       240
GCCCTCATAACTAACCATCCACCAGAGAAGATAACGGAGGGCAGACTTTGGCAGGGGGTA     720
                              Δ
 I  T  G  P  F  E  E  E  N  G  A  C  G  K  E  A  S  E  G  W       260
ATAACTGGCCCCTTCGAGGAAGAAAATGGAGCCTGTGGGAAGGAGGCGTCCGAGGGGTGG     780

 A  L  W  Q  A  P  C  *                                           267
GCTTTGTGGCAAGCCCCTTGCTGAAGCAGAAGGGCGTGAAGAACCGGGAGCTCATCCACA     840
```

4S Normal reading frame:



4S Mutant reading frame:



4S. The correlation between this heterozygous mutation and the observed clinical and biochemical phenotype in patient SF368 is discussed in Section 4.6.

The five remaining unaccounted fragments in PCR B of patient SF368 were used to estimate the locations of the mismatched sites, which were found to lie between nucleotide positions 1100-1350 of the 4S cDNA. In order to identify the mismatches, the 4SP24-4SP28 PCR product which spanned nucleotide positions 919-1330 was amplified from SF368 cDNA and a normal control. Direct DNA sequencing using end-labelled 4SP28 revealed SF368 was homozygous for a G to A transition at nucleotide 1126 ($G_{1126}A$), which resulted in a valine to methionine substitution at amino acid 376 (V376M). At the time the mutational analysis was performed the V376M substitution was considered a polymorphism, as it had been observed in another published 4S cDNA clone (Schuchman *et al.*, 1990), and also in a normal individual who was heterozygous for this substitution (data not shown). In addition, the V376M substitution is a fairly conservative exchange as both side chains are neutral and non-polar, and neither residue was conserved between sulphatases (Fig. 3.13). Subsequent to this work, Wicker *et al.* (1991) confirmed that the V376M substitution was polymorphic, as they showed that a 4S cDNA clone carrying V376M expressed comparable levels of enzyme activity towards a synthetic substrate, relative to the wild type clone. The $G_{1126}A$ transition was consistent with and accounted for the observed 500-bp fragment in the hydroxylamine reaction. The $G_{1126}A$ transition also predicted a 504-bp fragment in the osmium tetroxide reaction due to cleavage at the T-G mismatch, however it was not observed as the total band intensity of the osmium tetroxide reaction was extremely faint.

Apart from the $G_{1126}A$ transition (V376M), no other definitive DNA sequence differences were observed between SF368 and the normal control using 4SP28 as the sequencing primer. However, as the DNA sequence contained a number of

cross-bands, it was likely that the other base change/s in this patient were obscured by the sequence artefacts. The number of unaccounted fragments (4) is intriguing, and suggests that the patient may have more than two mutations. Alternatively, some of the fragments may be artefactual, due to sites which are susceptible to modification and cleavage although they are not mismatched (Hamish Scott & Peter Wilson, personal communication). It is envisaged that the rescreening of SF368 using the modified chemical cleavage technique will enable the more precise estimation of mismatch positions and therefore facilitate mutation identification. DNA sequencing with other primers in the vicinity of the predicted mismatched sites may overcome the sequence artefacts and enable the determination of the other base changes.

The patient SF368 has not been completely characterized at the molecular level, however the patient is heterozygous for a deletion of exon IV (Δexon IV) from the 4S cDNA. The relationship between the severe clinical and biochemical MPS-VI phenotype observed in patient SF368 and the predicted phenotype of the known mutation is presented in Section 4.6.

4.5.2f SF2357

The initial protocol of the chemical cleavage technique was performed on SF2357 (data not shown). PCR A duplexes displayed no specific cleavages, while PCR B duplexes were specifically cleaved to produce two fragments in the hydroxylamine reaction. The first fragment was about 650-bp in length, which indicated that the mismatch was in the vicinity of nucleotide position 980 or 1445. The second fragment was about 575-bp in length, which indicated that the mismatch was in the vicinity of nucleotide position 1050 or 1370.

The DNA sequences spanning all four of the potential mismatch sites have not been determined. However, preliminary sequence has been obtained from the PCR

B product of SF2357 and the normal control using end-labelled 4SP8. Patient SF2357 was found to be heterozygous for a C to T transition at nucleotide 1442 ($C_{1442}T$), which resulted in a proline to leucine substitution at amino acid 481, P481L (data not shown). The $C_{1442}T$ transition was consistent with and accounted for the observed 650-bp fragment in the hydroxylamine reaction. Although not conclusively shown, the 575-bp fragment in the PCR B product was considered to be most likely due to the V376M polymorphism (as seen in SF1022, Section 4.5.2a). Alternatively, the fragment may be due to the remaining unidentified mutation.

The predicted P481L substitution is most likely to be a mutation as it occurred at a residue that was conserved in 8 out of 10 different sulphatases (Fig. 3.13). Analysis of the predicted secondary structure using the Chou-Fasman algorithm was in support of the proposal that the P481L substitution was a mutation. The peptide region containing the P481 residue was predicted to adopt a β-turn configuration which was disrupted when L481 was substituted.

Patient SF2357 has not been completely characterized at the molecular level, however the patient was heterozygous for the P481L substitution. It is envisaged that rescreening of SF2357 using the modified chemical cleavage protocol will enable the identification of the other mutant allele. The relationship between the asymptomatic clinical and the mild biochemical MPS-VI phenotype observed in patient SF2357 and the predicted phenotype of the known P481L mutation is presented in Section 4.6.

4.5.2g SF2259, SF2424, SF2724

The three MPS-VI patients SF2259, SF2424 and SF2724, were also screened using the initial chemical cleavage protocol. Fragment bands were observed with each patient (data not shown), the sites of mismatch were estimated, and the DNA sequences in the vicinity of each mismatch were determined. However, no

conclusive base changes were observed between the DNA sequences obtained by the direct DNA sequencing of PCR products from each of the patients, and the normal control. As the DNA sequencing was only a preliminary attempt, the failure to identify base changes was considered to be due to the incomplete coverage of the DNA sequences in the vicinity of the predicted mismatch sites, rather than due to a chemical cleavage artefact. In addition, it was difficult to obtain definitive sequence from some of the predicted sites of mismatch as sequencing artefacts were encountered, presumably due to the high G+C content of the regions sequenced.

Further analysis of SF2259, SF2424 and SF2724 was deferred as the 4S mutations were not readily determined, and also to focus on investigation of the fully characterised patients. Subsequently, the Y210C mutant allele, originally identified in patient SF913, was also found in SF2724 by allele specific hybridization. It is envisaged that rescreening these three patients using the modified chemical cleavage protocol will facilitate mutation identification. The modified procedure has demonstrated the ability to detect previously undetectable T-G mismatches (as C-A mismatches) and estimate mismatch positions with greater precision.

4.6 MPS-VI MUTATION SUMMARY

In MPS-VI patients, the entire spectrum of clinical phenotype from asymptomatic to severe is clustered within a low and narrow range of residual enzyme activity between about 0 to 5% of normal (Table 4.3). The mutations in the 4S gene responsible for the low residual enzyme activity could conceivably manifest their effects at any stage in the process from the transcription of the 4S gene through to lysosomal residence of the 4S enzyme. Mutations of particular interest are those that affect the active site of the enzyme, or that result in a reduction in the

stability or intracellular half-life of the enzyme protein, or that result in inefficient intracellular targeting of the enzyme to the lysosome. These mechanisms are not necessarily mutually exclusive. As a first step in the elucidation of the molecular pathology, the 4S mutations in MPS-VI patients were determined.

The initial low-resolution screen for 4S mutations in MPS-VI patients was by Southern blot analysis. All MPS VI patients examined showed normal fragment patterns when probed with the 4S cDNA. Consequently, this suggested that the majority (>95%) of the 4S gene mutations in these MPS-VI patients were point mutations, or very small insertions or deletions. By comparison, Southern analysis of patients with Hunter syndrome (MPS-II), an X-linked lysosomal storage disorder due to the deficiency of iduronate-2-sulphatase, has revealed that of 319 MPS-II patients, about 20% (62 patients) showed gross structural rearrangements in the Id2S gene (Hopwood *et al.*, 1993a). A complete deletion of the Id2S gene was observed in 14 patients and a partial deletion or other gross rearrangement was observed in the remaining 48 patients. In addition, Ballabio *et al.* (1989) reported that 48 of 57 (84%) European patients with steroid sulphatase (STS) deficiency showed a deletion of the X-linked STS gene.

The techniques of chemical cleavage, direct DNA sequencing of PCR products, and ASO hybridization were used to identify the predicted point mutations, and have enabled the identification of 11 of the 18 MPS-VI alleles in the original nine patients selected for analysis (Tables 4.3 and 4.4). Four of the originally selected patients had both mutant alleles identified (SF1022, SF1246, SF913, SF2467), three patients had one of the two mutant alleles identified (SF368, SF2724, SF2357), and for two patients no mutant alleles were identified (SF2424, SF2259). A total of eight mutations distributed throughout the 4S cDNA were found, of which six were amino acid substitutions, one was a single nucleotide deletion, and one an undefined mutation that caused the deletion of exon IV from the 4S mRNA.

ASO hybridization identified five other MPS-VI patients who were heterozygous for three of the known 4S mutations. This enabled the definition of both mutant alleles in two patients (SF2984, SF912), and definition of one allele in each of three other patients (SF693, SF839, 51-1). Patient SF912 was a sibling of the selected patient SF913. A total of eight 4S mutations were found in 11 unrelated patients, which are summarised in Table 4.4, and the position of each mutation is indicated in a schematic diagram of the 4S coding region in Figure 4.13. The number of mutations found provide direct molecular evidence of the genetic heterogeneity underlying the biochemical and clinical heterogeneity observed in MPS-VI. Three polymorphisms were also identified, of which two were amino acid substitutions, V358M and V376M, and one was a silent change in the third base of the codon for Pro397. Patients in which the polymorphisms were identified are shown in Table 4.4.

It has been reported that about 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 (m^5C) residues in the CpG dinucleotide (Cooper & Youssoufian, 1988). Two of the six amino acid substitution mutations reported in this thesis could result from transitions in CpG dinucleotides (T92M and R95Q). Of interest is the observation that the R95Q mutation occurred more than once; in the two siblings SF912 and SF913, and in patient SF1246 who is presumably unrelated to them. The three polymorphisms; V358M, V376M, and the silent $G_{1191}A$ transition in P397, are G->A transitions in CpG dinucleotides. These observations are consistent with the suggestion that transitions in the m^5CpG dinucleotide occur more frequently than elsewhere (Cooper & Youssoufian, 1988). The Y210C and the H393P mutations were the most frequent mutant alleles found in the patient panel, each of which occurred in four heterozygotes out of 25 MPS-VI patients (8%). However, neither mutation occurred in a CpG dinucleotide.

TABLE 4.4 MPS-VI PATIENT MUTATION SUMMARY

The mutant 4S genotypes identified in each patient are given, together with the age at diagnosis, the clinical phenotype, and the technique used to detect the mutation. The techniques used were chemical cleavage and direct DNA sequencing of PCR products (CC/seq), and allele specific oligonucleotide hybridization (ASO). Polymorphisms (where known) are also shown.

Patient	Age at diagnosis	Clinical phenotype	Mutations allele 1	allele 2	Detection	Polymorphisms
SF1022	7 days	severe	ΔG_{238}	ΔG_{238}	CC/seq	M358 V376 G_{1191}
SF1246	13 months	severe	R95Q	H393P	CC/seq	V358 V376 A/G_{1191}
SF368	2 years	severe	Δexon IV	?	CC/seq	V358 M376 A_{1191}
SF693	3.75 years	severe	H393P	?	ASO	
SF2984	6.7 years	intermediate	Y210C	H393P	ASO	
SF912	13 years	mild	R95Q	Y210C	ASO	
51-1	15 years	mild	Y210C	?	ASO	
SF913	16 years	mild	R95Q	Y210C	CC/seq	V358 V376 G_{1191}
SF839	17 years	mild	H393P	?	ASO	
SF2724	17.5 years	mild	Y210C	?	ASO	
SF2467	35 years	mild	T92M	L498P	CC/seq	V358 V376 A/G_{1191}
SF2357	44 years	asymptomatic	P481L	?	CC/seq	

In a recessive genetic disease, the patient phenotype is a function of both individual mutant alleles. Therefore an attempted correlation between the clinical and biochemical phenotype on the one hand and the patient genotype on the other must take into account the contribution of both alleles. Generally, a severe phenotype is due to two severe mutant alleles, while a milder phenotype is due to one of the mutant alleles being mild. In the case of a milder phenotype, the mild allele ameliorates the severity of the severe allele. An important assumption of the genotype-phenotype correlation hypothesis is that the products of each allele do not interact or influence each other directly. The predicted severity of each of the 4S mutations, and the relation to the clinical and biochemical phenotype of the patient, will be discussed below.

The individual polymorphisms identified in the 4S gene are assumed to have at most a minor effect on the structure and function of 4S *in vivo*. However, the cumulative effects of multiple polymorphisms, or of polymorphisms linked to mutations, are not known. It is conceivably possible that the polymorphisms may ameliorate or accentuate the severity of a 4S mutation *in vivo*, and hence influence disease severity. In MPS-I (Hurler syndrome), polymorphisms have been implicated in the modulation of the severity of a mutant allele (Scott *et al.*, 1993). However, the mutation detection techniques as performed did not allow the determination of the individual linkage between each mutation and the polymorphisms.

4.6.1 SF1022

Patient SF1022 was found to be homozygous for a frameshift mutation ΔG_{238}, which was predicted to generate a grossly truncated 4S polypeptide of 112 amino acids, compared to the normal 4S reading frame of 533 amino acids. The aberrant 4S polypeptide was predicted to be enzymically inactive, as only the first 79

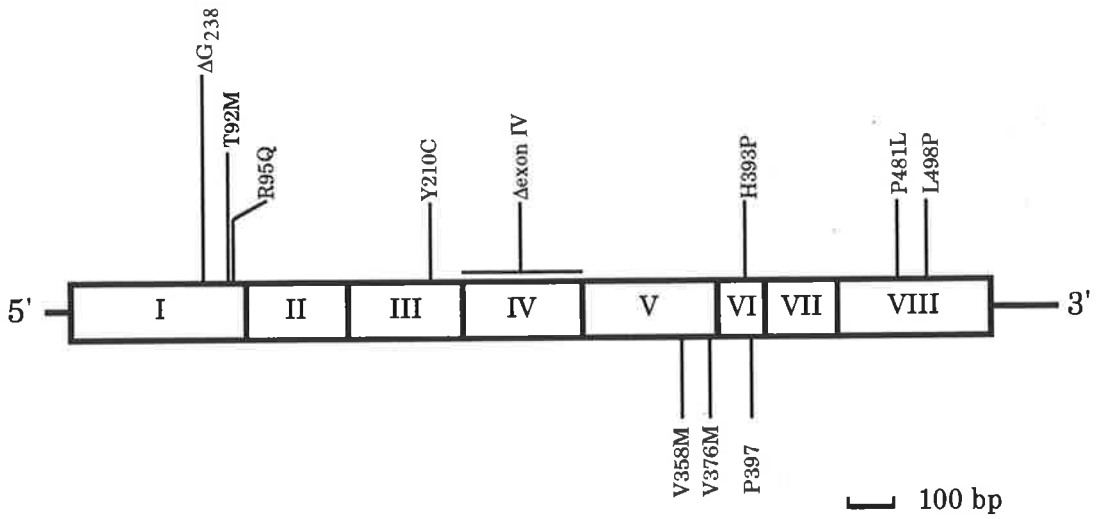


FIGURE 4.13 LOCATION OF THE MPS-VI MUTATIONS IN THE 4S mRNA CODING REGION

Diagrammatic representation of the coding region of the 4S mRNA, the exon boundaries, and the locations of the identified mutations and polymorphisms. The mutations are situated above the 4S coding region, while the polymorphisms are situated below the 4S coding region.

residues were co-linear with normal 4S while the succeeding 33 residues were frameshifted (Fig. 4.5). The predicted severe phenotype of the ΔG_{238} mutation in patient SF1022 correlated closely with the observed severe clinical and biochemical phenotype.

Patient SF1022 (male) presented at 1 day of age, with oedema of hands and feet and respiratory distress. The patient displayed coarse dysmorphic features with a flattened nasal bridge, large open fontanelle, large mouth and tongue, and short neck. Dysostosis multiplex was evident on the skeletal survey at two days, with abnormal vertebrae, small flared iliac wings, diaphyseal distension of tubular bones and large skull as the most prominent features. Haematological screening revealed inclusion bodies in leukocytes. Urine from SF1022 was shown to have a grossly elevated level of dermatan sulphate (88 g/mol creatinine), relative to age-matched normal controls (<4 g/mol creatinine) (John Hopwood, unpublished data). The patient developed all the complications which can occur in MPS-VI at an early age. Corneal clouding was evident by the age of three months, which became so dense as to make him blind by seven years, when a left corneal transplant was done. Hydrocephalus required shunting at four years and cervical chord compression, resulting from instability of the C2 or C3 vertebrae, required vertebral stabilization at six years. Lumbar kyphosis and joint contractures of shoulders, elbows, knees and hips were present by six months of age and gradually worsened. Currently (Dec. 1993) the patient is 13 years of age, and has a large head, short neck, lumbar kyphosis and has had virtually no gain in height since 30 months. The patient's chest shape is abnormal, but heart and lung function are "reasonably good". Patient SF1022 has the clinical features of a child with a severe form of MPS-VI, as typified by the early age of diagnosis and the extent and progression of the clinical symptoms. An exception is the degree of cardiopulmonary involvement, which is usually associated with a mild MPS-VI.

Laboratory analysis also showed that SF1022 had a severe biochemical phenotype, as no consistently detectable 4S activity towards a radiolabelled trisaccharide substrate was found in cultured fibroblasts or in peripheral blood leukocytes, as determined by the assay of Hopwood *et al.* (1986) (John Hopwood, unpublished data). The levels of iduronate-2-sulfatase, α-L-iduronidase and β-D-glucuronidase in both these cell types were normal, confirming the biochemical diagnosis of MPS-VI. Furthermore, out of a panel of seven monoclonal antibodies raised to normal 4S, only one detected protein, albeit at a very low level, in the patient's cultured fibroblasts (Brooks *et al.,* 1991b). In conclusion, a definite correlation was established between the predicted severity of the patient's genotype and the clinical and biochemical phenotype, and that the ΔG_{238} mutation was sufficient to account for the correlation.

4.6.2 SF1246

Patient SF1246 was a compound heterozygote of the two mutant 4S alleles R95Q and H393P. The R95Q mutation was predicted to result in a severe phenotype, as the amino acid substitution was a non-conservative change at a residue that was totally conserved between 10 different sulphatases (Fig. 3.13), and therefore was likely to be crucial in 4S structure and/or function. The predicted phenotype of the H393P mutation was less readily estimated based on homology between sulphatase sequences, as the residue was not conserved. However, the H393P substitution was a non-conservative change with respect to the amino acid side-chain, and the predicted propensity of the polypeptide region in the vicinity of the residue to form a α-helical structure was abolished when P393 was substituted.

Patient SF1246 was classified clinically as a severely affected MPS-VI patient, as she originally presented at the age of 13 months with mild developmental delay and thoracolumbar kyphosis. Hepatosplenomegaly was present, and radiological examination found skeletal features consistent with an MPS. Biochemical analysis

of residual 4S activity in patient cultured fibroblasts could not conclusively show a simple correspondence with a severe phenotype (Table 4.3). However, as discussed in Section 4.4, the assays were performed before the potential contribution of exogenous 4S (calf serum) was appreciated, and hence the significance of the data was reduced. Analysis of residual 4S protein in patient cultured fibroblasts found that of a panel of seven monoclonal antibodies raised to normal 4S, all seven detected protein, albeit at a lower level than the normal controls (Brooks *et al.*, 1991b).

The predicted severity of the H393P allele was supported by the clinical phenotype of patient SF693, who was heterozygous for the H393P mutation. Patient SF693 has the clinical features of a severely affected form of MPS-VI, as typified by the early age of diagnosis and the extent and progression of the clinical symptoms. The patient was diagnosed at the age of 3.75 years and presented with upper respiratory problems and short stature. At diagnosis, the patient had coarse features, corneal clouding, and some restriction of movement in all joints. Height was 96 cm (10^{th} centile), while the head circumference was 51.5 cm (97^{th} centile). Hepatosplenomegaly was slight. Currently at the age of 19 years the patient is "terminally ill", with "extremely short stature, cardiac failure secondary to diffuse airways narrowing, severe sleep apnoea and stiffness of all joints". Therefore, both mutant alleles in this patient were predicted to be severe, the H393P substitution as well as the unidentified mutation.

As MPS-VI is a recessive genetic disorder, and on the assumption that the clinical and biochemical phenotype can be correlated with the genotype, both the R95Q and H393P mutant alleles in SF1246 were predicted to individually result in severe phenotypes. This prediction was subsequently tested and confirmed experimentally. In Chapter 5, the individual effects of the R95Q and H393P alleles on 4S structure and function are presented.

4.6.3 SF912 & SF913

The MPS-VI patients SF912 and SF913 were siblings, and were compound heterozygotes of the two mutant 4S alleles R95Q and Y210C. The R95Q mutation was predicted to result in a severe phenotype, as previously discussed for SF1246. The predicted phenotype of the Y210C mutation was less readily estimated based on homology between sulphatase sequences, as the residue was not conserved. However, the Y210C substitution was a non-conservative exchange with respect to the chemical structure of the amino acid side-chain.

Patients SF912 and SF913 were classified clinically as milder or less severely affected than the severe forms, based on the later age of diagnosis and the reduced extent and progression of the clinical symptoms. Patient SF912 presented at the age of 13 years with short stature; she was 133.5 cm in height (<3rd centile), and 30.3 kg in weight (<3rd centile). The patient had a slight kyphosis, hearing loss in both ears (30 dB), mild mitral valve incompetence, a normal electrocardiogram, and no splenomegaly. Subsequently, an older brother SF913 (16 years) was diagnosed during the clinical investigation of the family. Biochemical analysis of urine from both patients was shown to have an elevated level of dermatan sulphate (8-9 g/mol creatinine) relative to age-matched normal controls (<0.7 g/mol creatinine), but the level was lower in comparison to the severely affected patients (Brooks *et al.*, 1991b). Out of a panel of seven monoclonal antibodies raised to normal 4S, all seven detected protein in cultured fibroblasts from both patients, albeit at a low level relative to the normal controls (Brooks *et al.*, 1991b).

The Y210C mutant allele was predicted to result in a mild phenotype, on the assumptions that the clinical and biochemical phenotype can be correlated with the genotype, and that R95Q was a severe allele. The predicted mild severity of the Y210C allele was supported by the clinical phenotype of the other patients who

were heterozygous for this mutation, SF2724 and 51-1, who were also independently classified as milder MPS-VI patients. Interestingly, all four patients were diagnosed between the relatively late ages of 13-17.5 years. In particular, patient 51-1 presented at 15 years of age because of short stature; she was 142 cm in height (<5th centile), and weighed 45 kg (10th centile). The patient had a normal facial appearance, enlarged liver, "very slight restriction of large joints", "fixed flexion deformity at terminal interphalangeal joints" and a very slight degree of corneal clouding. A soft systolic heart murmur was noted, while echocardiography was normal. The patient is now 21 years of age, she is "fit and well" and the clinical signs have shown no progression. Patient SF2724 was previously diagnosed as a Morquio syndrome at the age of 11 years. At the age of 17.5 years, the patient presented with hearing loss. He was short (10th centile), had coarse facial features, and had restricted movement in the shoulder, hip and other joints. Corneal clouding was slight, and no hepatosplenomegaly was observed. Biochemical investigation showed an increased urinary concentration of dermatan sulphate but not keratan sulphate. Consequently, the diagnosis was corrected to MPS-VI. The clinical history of this patient was an example where factors beside clinical severity alone affected the age of diagnosis.

Patient SF2984 (female) was also heterozygous for the Y210C mutant allele, yet the clinical phenotype was more severe than the other Y210C heterozygotes, and the age of diagnosis was earlier, 6 years and 8 months. In particular, corneal clouding, skeletal abnormalities, and restriction in movement of the hip and shoulder joints were noted at diagnosis. Facial dysmorphism was evident by 8 years. The increased clinical severity of SF2984 was conceivably due to the predicted severity of the other mutant allele, H393P. Comparison of the residual enzyme activities of the patients heterozygous for the Y210C allele did not find similarities (Table 4.3).

The predicted severities of the R95Q and Y210C alleles were subsequently tested and confirmed experimentally. In Chapter 5, the individual effects of these alleles on 4S structure and function are presented.

4.6.4 SF2467

Patient SF2467 was a compound heterozygote of the two mutant 4S alleles T92M and L498P. Assured prediction of the severity of either allele was hindered by a number of factors. The T92M mutation was a non-conservative amino acid change at a residue which lay within the CTPSR motif, a motif which was totally conserved in the three human arylsulphatases A, B, and C, and a sea urchin arylsulphatase. The T residue was not totally conserved in other sulphatases, however the divergent residues were generally conservative exchanges (Fig. 3.13). The predicted phenotype was potentially severe as the mutation was located at a conserved residue. However, the severity does depend on the putative role of the T92 residue, and of the CTPSR motif, in 4S structure and function. The predicted severity of the L498P mutation was also ambiguous, as the mutation was located in the carboxyl-terminal region of the 4S polypeptide which exhibited little conservation of amino acid sequence between sulphatases. The predicted severity was considered to vary from mild to severe, depending on whether the conformation of the polypeptide chain in the vicinity of the L498 residue was crucial or not for 4S structure and function, and the extent to which the L498P mutation perturbed the conformation. The patient's clinical and biochemical phenotype enabled a partial resolution of the ambiguity.

The SF2467 cell line was obtained from the Human Genetic Mutant Cell Repository (GM2849). The patient (male) was considered a mild MPS-VI, as the patient was diagnosed at 35 years of age, however no clinical details were provided. Biochemical analysis indicated that the patient's cultured skin fibroblasts contained residual 4S activity towards a radiolabelled trisaccharide substrate

(Table 4.3). Out of a panel of seven monoclonal antibodies raised to normal 4S, all seven detected protein in the patient's cultured fibroblasts at levels higher than in the other MPS-VI patients examined, but lower than the normal controls (Brooks *et al.*, 1991b). These results suggested that one of the mutant 4S alleles T92M and L498P was a mild mutation.

To clarify the effect of the T92M and L498P alleles on 4S structure and function, and therefore enable the predicted phenotype of each to be refined, each allele was expressed individually in cultured cells. The analysis of the 4S activity and protein produced by the T92M and L498P 4S mutants is presented in Chapter 5.

4.6.5 SF368

The patient SF368 was heterozygous for a deletion of exon IV from the 4S mRNA (Δexon IV). However the patient had not been completely characterized at the molecular level, as the other mutant allele and the underlying mutation responsible for the deletion of exon IV had not been identified. The Δexon IV mutation was predicted to result in a severe phenotype, as it was a frameshift mutation that resulted in a truncated 4S polypeptide of 267 amino acids, relative to the normal 4S reading frame of 533 residues. The putative truncated 4S was collinear with the normal 4S for the first 230 amino acids, however the exon IV-encoded amino acids 231 to 299 were absent, and the aberrant splicing of exon III to exon V resulted in the addition of 37 frameshifted amino acids terminated by an in-frame premature stop codon (TGA) (Fig. 4.12). In addition, 5 residues (Pro231, Phe233, His242, Asp274 and Thr293) of the 15 residues totally conserved between 10 sulphatases are located in the deleted exon IV.

Patient SF368 was classified as a severely affected MPS-VI patient; the diagnosis was made at the early age of two years when the patient presented with coarse features and developmental delays. Biochemical analysis of the urine from SF368

was shown to have a grossly elevated level of dermatan sulphate (49 g/mol creatinine), relative to age-matched normal controls (<2 g/mol creatinine) (Brooks *et al.*, 1991b). Furthermore, out of a panel of seven monoclonal antibodies raised to normal 4S, only two detected protein, albeit at a very low level, in the patient's cultured fibroblasts (Brooks *et al.*, 1991b).

On the assumption that the clinical and biochemical phenotype can be correlated with the genotype, both the Δexon IV allele and the other unidentified mutation were predicted to be severe mutations. The Δexon IV mutant allele was considered unlikely to be able to contribute any 4S activity to the patient, and therefore the effect of the Δexon IV allele on 4S structure and function was not investigated further, nor was the underlying mutation identified. Other MPS-VI patients have been described with deletions of 4S exons; Arlt *et al.* (1994) have reported a patient who was heterozygous for the deletion of exon V from the 4S mRNA. Comparable deletions of entire exons causing disruptions of the reading frame have been described for the dystrophin gene of patients with Duchenne muscular dystrophy, who do not show any residual dystrophin activity (Chelly *et al.*, 1990).

4.6.6 SF2357

The patient SF2357 was heterozygous for the mutant 4S allele P481L. However the patient had not been completely characterized at the molecular level, as the other mutant allele was not identified. The amino acid substitution P481L was predicted to result in a severe phenotype, as it occurred at a conserved residue found in 8 out of 10 sulphatases, which suggested that P481 may be crucial for 4S structure and/or function.

Clinically, the patient was asymptomatic but biochemical analysis identified the patient as a very mild MPS-VI. Urine from SF2357 was shown to have a mildly elevated level of dermatan sulphate (1 g/mol creatinine), relative to age-matched

normal controls (<0.7 g/mol creatinine) (Brooks *et al.,* 1991b). The patient's cultured skin fibroblasts contained low levels of residual 4S activity (Table 4.2), and out of a panel of seven monoclonal antibodies raised to normal 4S, five detected protein in the patient's cultured fibroblasts, albeit at a low level compared to the normal controls (Brooks *et al.,* 1991b).

The P481L allele was predicted to be a severe mutation, and the other unidentified mutant allele was predicted to be a mild mutation, based on the assumption that the clinical and biochemical phenotype can be correlated with the genotype. The P481L mutation was identified too late to be investigated by expression studies, and therefore the effect of the P481L substitution on 4S structure and function was not able to be investigated.

4.7 DISCUSSION

The phenotype severity of each of the identified 4S mutations was predicted based on consideration of the degree of sulphatase sequence conservation, the amino acid side-chain chemistry, the effect on the predicted secondary structure of the 4S polypeptide, and also the clinical and biochemical phenotype of the patient. The mutations, in relative order from the predicted mildest to most severe phenotype were; Y210C, R95Q, H393P, Δexon IV, and ΔG_{238}. The P481L mutation was predicted as severe, but it's severity relative to the above mutations was not known. The predicted individual severity of the T92M and the L498P mutations was also not known, however at least one of the mutants was relatively mild, based on the apparently mild phenotype of the patient.

The predicted phenotypes of the mutations required confirmation if the genotype-phenotype correlation hypothesis was to be supported. In addition, to determine the contribution of each mutant allele to the clinical and biochemical phenotype, both alleles required `dissection' and assessment individually. It therefore became

a goal of this study to analyze further the mutant allele-pairs in a number of MPS-VI patients. In particular, the effect of the 4S mutations on enzyme activity and 4S protein was examined. Chapter 5 reports these protein chemistry results for a number of 4S mutants, and compares the observed behaviour of the 4S mutants *in vitro* with the predicted severities of the mutant alleles. These results were then compared with the clinical and biochemical phenotypes in order to refine the correlations, and test the hypothesis.

Although none of the identified mutations are frequent within the MPS-VI patient population, they may aid in diagnosis, especially prenatal diagnosis in families where the genotype of an affected member is already known. The relevant portion of the 4S gene could be PCR amplified from genomic DNA samples obtained from family members, and the mutant alleles detected by ASO analysis.

CHAPTER 5

HIGH-LEVEL EXPRESSION & PRELIMINARY ANALYSIS OF 4S MUTANTS

5.1 INTRODUCTION

The central aim of the work presented in this thesis was to test the hypothesis that the clinical and biochemical phenotype in MPS-VI patients could be correlated with the mutant genotype. In the previous Chapter, the clinical severity of each of the 4S mutant alleles was predicted by comparison of the mutant genotype with the patient's clinical and biochemical phenotype, and by analysis of the mutated protein sequence. However, unambiguous phenotype prediction of each allele was hindered as two mutant alleles were present in each patient, the majority of whom were compound heterozygotes of rare or unique mutant alleles. In addition, phenotype prediction was imprecise as the putative 4S mutants had not been independently confirmed by the demonstration of reduced activity and/or stability when expressed in cell culture. Therefore, it was thought judicious to consider as tentative the predictions of mutant severity derived solely from patient phenotype, until independently confirmed.

The work described in this Chapter sought to test the central hypothesis via an alternative approach; i.e. given the genotype, do the *predicted* clinical and biochemical phenotypes correlate with the *actual* phenotypes? In summary, selected MPS-VI 4S mutations were individually expressed in cell culture, and the 4S activity and the level of 4S protein were determined. The values obtained comprised the *measured* biochemical phenotype, and were used to derive the *predicted* biochemical phenotype of each mutation in a patient. The *measured* biochemical phenotype values were also used to derive the *predicted* clinical phenotype or severity of each mutant allele. The predictions were then compared with the *actual* clinical severity of the patient, and the *actual* biochemical phenotype parameters measured in patient cultured fibroblasts.

Specifically, each mutant allele was individually produced *in vitro* by oligonucleotide-directed mutagenesis and transferred to an expression vector. To

facilitate the assembly of the mutant 4S expression constructs, small fragments or `cassettes' generated by restriction endonuclease digestion, were used to transfer each of the mutations from the mutagenesis vector. The wild type and mutant 4S expression constructs were introduced individually into Chinese hamster ovary (CHO) cells, and stably-transfected lines were selected for further analysis. CHO was selected for the heterologous expression of 4S, as high-level synthesis of the lysosomal enzymes α-L-iduronidase, iduronate-2-sulphatase and 4S had been previously demonstrated using this system (Scott *et al.*, 1991; Wilson, 1992; Anson *et al.*, 1992b).

The 4S activity and 4S protein were measured for the CHO-expressed wild type and mutant 4S constructs. The 4S activity was determined in an assay which combined the specificity of an anti-4S monoclonal antibody with the ease of use of synthetic arylsulphatase substrates (Brooks *et al.* 1991a). A number of studies had shown that the kinetic properties of the 4S enzyme (pH optimum, K_m and V_{max}) were essentially unchanged upon binding to selected antibodies (Brooks *et al.*, 1991a; Anson *et al.*, 1992b; Doug Brooks, unpublished results). These studies established the validity of the immunocapture technique for the measurement of enzyme activity *in vitro*. The synthetic fluorogenic substrate 4-methylumbelliferyl sulphate (4MUS) was used to determine 4S arylsulphatase activity, rather than the radiolabelled trisaccharide substrate derived from chondroitin-4-sulphate. Each substrate had specific advantages and disadvantages; 4MUS was considered convenient to use but was not specific for 4S, while the oligosaccharide substrate was highly specific but it's synthesis and the assay itself were laborious. As the goal was to achieve a preliminary biochemical characterization of the mutant 4S enzymes, the 4MUS substrate was chosen due to it's convenience; the necessary specificity for 4S was to be conferred by the monoclonal antibody in the immunocapture assay. The selected monoclonal antibodies were also used to

quantify 4S protein in an heterogeneous ELISA assay which was a modification of a previous assay (Brooks *et al.*, 1991b).

Biochemical analysis of the CHO-expressed mutants enabled the residual activity and residual protein to be calculated for each mutant allele, as each was expressed and characterized separately. From the *in vitro* results, the relative biochemical and clinical severity of each mutant allele was predicted. The central hypothesis was tested by comparison of the measured biochemical and predicted clinical phenotypes of the CHO-expressed mutant alleles with the observed biochemical phenotype in patient cultured fibroblasts and the clinical phenotype of the patient. In general for each patient, the predicted severity of the mutant 4S allele pair demonstrated a reasonable correlation with the severity of the observed clinical phenotype, within the intrinsic limitations of all the data. The degree of correspondence between the predicted and actual values, and whether a genotype-phenotype correlation was found in MPS-VI, is discussed in Sections 5.6.5 and 5.7.

In addition to testing the genotype-phenotype correlation, the preliminary biochemical characterization of the CHO-expressed 4S mutants provided insights into the biosynthesis, processing, structure and function of wild type and mutant 4S.

5.2 SPECIFIC MATERIALS & METHODS

5.2.1 POLYCLONAL & MONOCLONAL ANTIBODIES

The monoclonal and polyclonal antibodies were generously provided by Doug Brooks of the Department of Chemical Pathology, The Women's & Children's Hospital, Adelaide.

The monoclonal antibodies ASB 4.1, ASB 56.2, ASB 58.3, and ASB F66, were raised and characterized against immunopurified human liver 4S according to a

method reported previously for the monoclonal antibodies ASB 4.1 and ASB 5.4 (Gibson *et al.*, 1987; Brooks *et al.*, 1990). The rabbit polyclonal antibody was prepared against human 4S which had been immunopurified using ASB 4.1 (Brooks *et al.*, 1991b).

5.2.2 SITE-DIRECTED *IN VITRO* MUTAGENESIS

Site-directed *in vitro* mutagenesis of the 4S cDNA was performed using a commercially available mutagenesis kit (Amersham, UK), essentially according to the manufacturer's instructions. The kit was based on the phosphorothioate approach of Taylor *et al.* (1985) and Nakamaye & Eckstein (1986). The M13 subclone of the 4S cDNA, p4S18.1, was constructed as described in Section 5.3, and was used as the single-stranded DNA template for mutagenesis. The oligonucleotide primers, originally designed for the detection of the 4S mutations by allele-specific hybridization (as described in Chapter 4 and Appendix B), were used to mutagenize the 4S cDNA as detailed below.

5.2.2a 5'-PHOSPHORYLATION OF THE MUTAGENIC OLIGONUCLEOTIDE

The mutagenic oligonucleotide (75 pmoles) was 5'-phosphorylated in a reaction mixture consisting of kinase buffer (100 mM Tris-HCl, pH 7.5, 10 mM $MgCl_2$, 10 mM DTT, 1 mM rATP) and 4 units of T4 polynucleotide kinase, in a reaction volume of 30 µl. The reaction mixture was incubated at 37°C for 30 min, followed by an incubation at 70°C for 10 min to terminate the reaction. An aliquot of the phosphorylated mutagenic oligonucleotide (5-6 pmoles) was then annealed to 5 µg of single-stranded p4S18.1 in a reaction mix consisting of kit annealing buffer. The reaction was denatured at 70°C for 3 min, and then allowed to anneal at 37°C for 30 min. The reaction was then placed on ice.

5.2.2b SYNTHESIS & LIGATION OF THE MUTANT DNA STRAND

The annealed mutagenic oligonucleotide was used to prime second-strand DNA synthesis from the single-stranded DNA template. The following kit-supplied reagents were added to the annealing reaction from above, in a final reaction volume of 50 µl; $MgCl_2$ solution, Nucleotide mix 1 (which contained dNTPs and dCTPαS), 6 units Klenow polymerase and 6 units T4 DNA ligase. The reaction was mixed and then incubated at room temperature overnight.

5.2.2c REMOVAL OF NON-MUTANT SINGLE-STRANDED DNA

Parental (non-mutant) single-stranded DNA which remained after the previous step was removed by the centrifugal filtration of the reaction mix through the kit-supplied columns. The columns contained a nitrocellulose membrane, which bound the single-stranded DNA and allowed the double-stranded DNA to pass through into the collection tube below. Specifically, 170 µl of water and 30 µl of 5 M NaCl was added to the reaction mixture from Section 5.2.2b, which was then applied to a nitrocellulose column and centrifuged at 500 g for 10 min. A 100 µl volume of 500 mM NaCl was added to the column, which was then respun at 500 g for 10 min. The filtrate was collected and the double-stranded DNA was precipitated by the addition of 28 µl 3 M sodium acetate, pH 5.2 and 700 µl of cold (-20°C) ethanol. After incubation at -80°C for 30 min, the DNA was pelleted by centrifugation at 12,000 g for 15 min at 4°C. The DNA pellet was washed with ethanol, vacuum-dried and resuspended in 25 µl of the kit-supplied buffer 2.

5.2.2d NICKING OF THE NON-MUTANT DNA STRAND USING *Nci*I

The non-mutant strand of the double-stranded heteroduplex was then specifically cleaved by digestion with *Nci*I, whereas the mutant strand was resistant to cleavage due to the incorporation of the thionucleotide during synthesis *in vitro*. A 10 µl aliquot of the filtered and resuspended DNA was incubated with 5 units of *Nci*I and 65 µl of the kit-supplied buffer 3 at 37°C for 90 min.

5.2.2e EXONUCLEASE III DIGESTION OF THE NON-MUTANT STRAND

Selective removal of the non-mutant strand from the double-stranded heteroduplex was achieved by incubation of the construct with exonuclease III; the *Nci*I-nicked non-mutant strand was hydrolyzed while the mutant DNA strand remained. To an aliquot of the above digest (66 µl) was added 12 µl of 500 mM NaCl, 10 µl of kit-supplied buffer 4 and 50 units (2 µl) of exonuclease III freshly diluted 10-fold in buffer 4. The reaction was incubated at 37°C for 30 min, and was then terminated by incubation at 70°C for 15 min.

5.2.2f REPOLYMERIZATION & LIGATION OF THE GAPPED DNA

The repolymerization step ensured the resynthesis of double-stranded closed circular DNA carrying the required mutation. To 75 µl of the above reaction was added the following kit-supplied reagents; 13 µl of nucleotide mix 2 (dNTPs), 5 µl of $MgCl_2$ solution, 3 units of DNA polymerase I and 2 units of T4 DNA ligase. The reaction was incubated at room temperature for 3 hr. The final reaction mix was stored at -20°C until required for transformation into competent TG1 bacterial cells.

5.2.2g TRANSFORMATION

Competent TG1 cells were prepared and transformed in duplicate with 20 µl aliquots of the above reaction, according to the procedure in Section 2.2.6. The next day, plaques containing putative mutant 4S cDNA clones were picked, and the RF and ssDNA prepared as described in Sections 2.2.1b and 2.2.16 respectively.

5.2.3 EXPRESSION OF 4S IN CELL CULTURE

5.2.3a ELECTROPORATION

The wild type 4S expression vector pRSVN.4S.08 was constructed and generously provided by Don Anson of the Department of Chemical Pathology, The Women's & Children's Hospital, Adelaide.

CHO cells (strain DK1) were grown in F-12 nutrient media (Ham's), supplemented with 10-20% (v/v) heat-inactivated foetal calf serum (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. The FCS was heat-inactivated by incubation of the FCS stock at 56°C for 60 min, in order to reduce the background arylsulphatase activity. The cells were fed with fresh medium every 2-3 days. Cells were typically cultured in either 75 cm^2 culture flasks or 100 mm diameter culture dishes, with 10-13 ml of nutrient media.

The cells to be transfected by electroporation were harvested with a 10% (v/v) trypsin-versene solution in PBS, and centrifuged at about 500 g for 2 min. The cell pellet was resuspended in 10 ml PBS, and the viable cells were counted by trypan blue staining in a haemocytometer. The cells were centrifuged at about 500 g for 1 min, the supernatant discarded and the cell pellet was resuspended in PBS at 1.2 x 10^7 viable cells/ml. An aliquot of the cell suspension (0.8 ml) was placed in a disposable electroporation chamber (4-mm gap, BRL) and 20 µg of plasmid DNA

was added. The chamber was allowed to equilibrate to 0°C on ice for 10 min. The cells were then electroporated using a pulse of 275 Volts at 330 μF in a BRL Cell-Porator. The pRSVN.4S.08 construct contained the gene for aminoglycoside phosphotransferase, which conferred resistance to the antibiotic neomycin (G418) on transfected cells and therefore enabled selection. For stable expression, the electroporated cells were grown in non-selective medium for 48 hr and subcultured 1:5 and 1:25 into medium containing 375 μg/ml of G418. Cell colonies became visible after about 7-10 days of growth, and were maintained in medium containing 375 μg/ml of G418. Both pooled and individual clones were harvested and subcultured further; the latter were isolated with cloning rings.

5.2.3b PREPARATION OF CELL EXTRACTS & CONDITIONED MEDIUM

The cultured adherent cells were harvested when confluent by trypsinization, the cells were resuspended in extraction buffer (0.5 M NaCl, 20 mM Tris-HCl, pH 7.0), and lysates were prepared by six freeze-thaw cycles alternating between dry ice/ethanol and tepid water. The cell lysates were clarified by centrifugation at 12,000 g for 5 min at 4°C. The total protein content of the extracts was determined using the Bio-Rad protein assay kit according to the manufacturer's instructions, using known quantities of BSA as the standard. Aliquots of conditioned media were retained and were clarified of cells by centrifugation at 500 g for about 2 min. The cell extracts and the culture media were assayed for arylsulphatase activity with the fluorogenic substrate 4-methylumbelliferyl sulphate (4MUS) as described in Section 5.2.4, and the 4S protein was quantified by ELISA as described in Section 5.2.5. Extracts and media aliquots were stored at -80 or -20°C.

5.2.4 DETERMINATION OF 4S ENZYME ACTIVITY

Two approaches were used to determine arylsulphatase activity towards 4MUS. In the first approach, the total arylsulphatase activity (arylsulphatases A, B, and C) of each sample was measured in a direct assay. The second approach utilized immunocapture with various monoclonal antibodies to purify 4S away from other arylsulphatases and therefore permit the detection of 4S-specific arylsulphatase activity.

5.2.4a TOTAL ARYLSULPHATASE ACTIVITY ASSAY

The assay of arylsulphatase activity was performed essentially as described in Gibson *et al.* (1987), which was a modification of the procedures of Christomanou & Sandhoff (1977), and Chang *et al.* (1981). A 5 µl aliquot of either cell extract or conditioned culture media (neat or diluted in cell extraction buffer) was mixed with 95 µl of 4MUS substrate (5 mM 4MUS, 50 mM sodium acetate, pH 5.6, and 0.05 mg/ml bovine serum albumin) in a test-tube and incubated for 20-60 min at 37°C. The reaction was terminated by the addition of 1.5 ml of glycine stop buffer (50 mM glycine-carbonate buffer, pH 10.7). Enzyme activity was based on the hydrolysis of 4MUS to 4-methylumbelliferone (4MU), which was measured spectrophotometrically using a Perkin Elmer spectrofluorometer at the emission wavelength of 446 nm (E_{446nm}) using an excitation wavelength of 366 nm. One Unit of activity was defined as the amount of enzyme required to produce 1 nmole of 4MU per min; 1 nmole of 4MU was equivalent to an E_{446nm} signal of 10. Results were expressed either as activity Units per volume in the case of culture media, or as activity Units per mg of cell protein in the case of cell extracts.

5.2.4b 4S-SPECIFIC ARYLSULPHATASE ACTIVITY ASSAY

The immunocapture assay of 4S-specific arylsulphatase activity was a modification of a previously described method (Brooks *et al.*, 1991a), and is described below. A

sheep anti-mouse polyclonal Ig antibody (Silenus, Hawthorn, Victoria, Australia) was diluted to 20 μg/ml in 0.1 M sodium bicarbonate, pH 8.5, and 100 μl was added to each well of a 96-well microtitre plate. The plate was incubated at 37°C for 2 hr and then 4°C overnight to allow the antibody to bind. Unbound antibody was aspirated from the wells which were then washed three times with 200 μl of buffer A (20 mM Tris-HCl, pH 7.0, 250 mM sodium chloride, 1% (w/v) ovalbumin) to block the remaining binding sites. To each well was then added 100 μl of the appropriate anti-4S monoclonal antibody (obtained from hybridoma culture supernatants), which was allowed to bind to the anti-mouse polyclonal by incubation at room temperature for 4 hr. Unbound monoclonal antibody was aspirated from the wells, which were then washed three times with 200 μl of buffer A to block the remaining binding sites. The blocking solution was removed and replaced with 50 μl of each sample of cell extract or conditioned culture media, either neat or diluted in buffer A. Dilution was required for some samples in order to ensure detection in the linear range of the assay. The microtitre plates containing the samples were incubated overnight at 4°C to allow binding of the 4S protein to the anti-4S monoclonal antibodies.

The next day, the samples were aspirated and the wells were washed three times with 200 μl of buffer A. To each of the wells was added 100 μl of the 4MUS substrate (5 mM 4MUS, 50 mM sodium acetate, pH 5.6, and 0.05 mg/ml bovine serum albumin), which was then incubated at 37°C for 5 hr. The contents of each well were then transferred to 1.5 ml of glycine stop buffer (50 mM glycine-carbonate buffer, pH 10.7) in a polystyrene test tube. As for the total arylsulphatase activity assay, E_{446nm} values were determined spectrophotometrically. However, 4S activities of the samples were determined with reference to a standard curve of known 4S activities, which were immunocaptured in parallel with the samples. Results were expressed either as

activity Units per volume in the case of culture media, or as activity Units per mg of cell protein in the case of cell extracts.

5.2.5 IMMUNOQUANTIFICATION OF 4S PROTEIN

Wild type and mutant 4S protein levels were measured by an heterogeneous immunoassay, as reviewed in Brooks *et al.* (1993). In summary, an immunoaffinity-purified polyclonal antibody was used to bind the 4S protein to an microtitre plate. An appropriate anti-4S monoclonal antibody was then allowed to bind to 4S, and a peroxidase-labelled second antibody directed to the 4S monoclonal antibodies was used to detect the bound protein colourimetrically.

Specifically, an anti-4S polyclonal antibody was diluted 1/32 (v/v) in 100 mM sodium bicarbonate, pH 8.5, and 50 µl was added to each well of a polyvinylchloride microtitre plate and incubated first at 37°C for 2 hr and then at 4°C overnight to allow the antibody to bind. Unbound antibody was aspirated from each well, and the wells were washed three times with 200 µl of buffer A. The final wash was retained on the microtitre plate for about 4 hr at room temperature to block the remaining binding sites. The buffer A solution was aspirated from the wells and 50 µl of the samples to be quantified for 4S protein, either neat or serially diluted in buffer A, were added to each well and incubated at 4°C overnight.

The next day, the samples were aspirated, and the wells washed three times with 200 µl of buffer A. Anti-4S monoclonal antibody (100 µl) was added to each well and the plates were incubated at room temperature for 4 hr. The wells were aspirated and washed three times with 200 µl of buffer A. To each well was then added 50 µl of a peroxidase-labelled anti-mouse Ig second antibody (diluted 1:1000 in buffer A), and the plates were incubated at room temperature for 1 hr. The second antibody was then aspirated, and each well was washed three times with 200 µl of a solution of 20 mM Tris-HCl, pH 7.0, and 250 mM sodium chloride. The

peroxidase chromogenic substrate ABTS (2,2'-azino-di[3-ethyl-benzthiazoline-6-sulphonic acid]) was prepared according to the manufacturer's instructions (Bio-Rad). The wash buffer was aspirated, and 100 µl of the ABTS preparation was added to each well and incubated for 20-30 min at room temperature. The amount of 4S protein bound was directly proportional to the chromophore concentration, which was quantified by measurement of the absorbance at 414 nm on an automated ELISA reader. 4S protein values were determined from a standard curve of known 4S protein concentrations, extracted from pooled normal human fibroblasts, and were expressed as µg of 4S protein per mg of extracted cell protein.

RESULTS & DISCUSSION:

5.3 *IN VITRO* SYNTHESIS OF THE 4S PATIENT MUTATIONS

In order to investigate the effect of each of the 4S mutations on 4S activity and protein, the mutant alleles were first synthesized *in vitro*. Oligonucleotide-directed *in vitro* mutagenesis was considered the method of choice for the synthesis of the 4S patient mutations, due to the efficiency and simplicity of the technique, and the availability of a number of different and well established protocols. A commercially available oligonucleotide-directed *in vitro* mutagenesis kit was chosen, which utilized the phosphorothioate approach (Taylor *et al.*, 1985; Nakamaye & Eckstein, 1986). The phosphorothioate approach was selected based on the advantages of high mutagenic efficiency and simplicity claimed by the manufacturer. In addition, the approach had been used successfully by other members of the group to create point mutations in the α-L-iduronidase cDNA (Phil Thompson, personal communication). The uracil template method (Kunkel *et al.*, 1987) and the gapped duplex method (Kramer & Fritz, 1987) were other variations of the oligonucleotide-directed *in vitro* mutagenesis approach that were also considered. However, as

both these methods required the use of specialized bacterial strains, and as the latter method required specialized vectors and plaque selection techniques, the phosphorothioate method was preferred.

In oligonucleotide-directed mutagenesis, the 4S patient mutations are synthesized *in vitro* using a wild type 4S cDNA clone as the template. An alternative approach to *in vitro* mutagenesis was also considered, in which the required 4S mutation was obtained directly by PCR amplification of the patient's cDNA or genomic DNA. However, during the development and construction of the expression system for the 4S mutations, the PCR-based approach was not thought suitable for the synthesis of the mutants, due to the relatively high error rate of the conventional Taq polymerase in the PCR amplification reaction. For example, sequence artefacts have been observed in cloned PCR products at a frequency of approximately 1:300 bases (Jansen & Ledley, 1990), and 1:360 bases (Dunning *et al.*, 1988), using cDNA and genomic DNA templates respectively. These error rates were unacceptable as they would necessitate the verification of the entire DNA sequence of all cloned PCR products prior to their use in the 4S expression system, which was considered too laborious and time-consuming. In addition, both identified and unidentified polymorphisms present in the patient may be PCR amplified together with the required mutation, and so complicate subsequent interpretation of the effect of the mutation. Therefore, oligonucleotide-directed *in vitro* mutagenesis was used to synthesize the 4S mutations described in this Chapter. Recently, higher fidelity Taq polymerases have become available (e.g. $Vent_R$ DNA polymerase, New England Biolabs). The use of a higher fidelity Taq polymerase, and the PCR amplification of short segments of patient cDNA, may reduce the work-load of the PCR-based approach to a feasible level for future mutation studies. The mutant 4S expression system described in Section 5.4 is sufficiently versatile that the mutant `cassettes' used in the scheme could be obtained from a site-directed mutagenesis

construct or from a patient PCR product. However, the sequence of the patient PCR product would still need to be checked, and any differences removed.

The procedure of oligonucleotide-directed phosporothioate-based *in vitro* mutagenesis is now reviewed. The DNA template to be mutagenized is cloned into a vector capable of single-stranded DNA production, such as a derivative of M13 or pTZ. An oligonucleotide primer which contains the required mutation is hybridized to the complementary single-stranded non-mutant clone. *In vitro* second-strand DNA synthesis and the generation of a covalently closed heteroduplex molecule is achieved by Klenow polymerase and T4 DNA ligase. During the second-strand synthesis reaction, dCTPαS is incorporated into the strand primed by the mutant oligonucleotide. dCTPαS is a nucleotide analogue of dCTP in which an oxygen atom on the α-phosphate is replaced by sulphur, and is the selective `tag' which enables the non-mutant or parental strand to be subsequently selected against. The double-stranded heteroduplex species is then digested with the restriction endonuclease *Nci*I, which results in cleavage of the parental strand, while the phosphorothioate-containing mutant DNA strand remains intact. The parental strand is then extensively hydrolyzed by exonuclease III, which initiates digestion at the nicked sites. A covalently closed double-stranded homoduplex mutant species is then synthesized from the intact mutant strand. Transformation of the homoduplex species into competent host cells enables isolation of single- and double-stranded DNA containing the intended mutation.

The template for the oligonucleotide-directed *in vitro* mutagenesis was the 4S cDNA clone pASB2-5, which consisted of a 2.2-kb 4S cDNA subcloned into the *Eco*RI site of M13mp18 (Peters *et al.*, 1990). The pASB2-5 clone encoded the complete 4S reading frame except for the first 13 amino acid codons (nucleotides 1-38) at the 5'-end (Peters *et al.*, 1990). However, it was still considered a suitable 4S cDNA template for mutagenesis, as none of the patient mutations to be expressed

were situated in this absent region. In retrospect, the plasmid p4SFL may have been superior to pASB2-5 as the source of the 4S cDNA template, as it contained a full-length 4S cDNA reading frame, albeit modified in the 5'-untranslated region (Anson *et al.*, 1992b).

The primers to be utilized for mutagenesis of the 4S cDNA were originally designed for allele specific oligonucleotide hybridization; the majority were derived from the sense strand of the 4S cDNA (Appendix B). However, as the 2.2-kb single-stranded 4S cDNA in pASB2-5 was also in the sense orientation, the mutagenesis primers could not be used. Therefore, in order to obtain the 4S clone in the antisense orientation, the pASB2-5 construct was digested with the restriction endonuclease *Eco*RI, the 2.2-kb 4S cDNA fragment was isolated, and the fragment recloned into an *Eco*RI-cut and phosphatased M13mp18 vector. An M13mp18 subclone, p4S18.1, was identified with the necessary orientation of the 2.2-kb *Eco*RI 4S cDNA fragment for mutagenesis by allele specific hybridization (data not shown). Large scale single-stranded DNA of the p4S18.1 construct was prepared according to a scaled-up version of the procedure in Section 2.2.16b.

Oligonucleotide-directed mutagenesis was performed according to the phosphorothioate approach (Section 5.2.2). Single-stranded DNA of the p4S18.1 clone was used as the wild type 4S cDNA template, and the mutagenic oligonucleotides used to synthesize the 4S mutants are listed in Appendix B. Putative mutant clones were isolated from transformed TG1 hosts, and small scale replicative-form (RF) DNA preps were performed. The RF DNA samples were dot-blotted and mutant clones were identified using allele specific hybridization analysis (data not shown). The procedures for allele specific hybridization and the washing conditions for the oligonucleotides, were as described in Chapter 4, Section 4.2.5. The phosphorothioate-based mutagenesis approach was successfully used to synthesize all five of the selected patient 4S mutations; the average mutagenesis

efficiency for the five mutants was about 65% (19/30 clones). The five mutant 4S cDNA subclones in M13mp18; p4S-T92M, p4S-R95Q, p4S-Y210C, p4S-H393P and p4S-L498P, were used in the construction of the mutant 4S expression vectors, as described below.

5.4 DESIGN & CONSTRUCTION OF THE 4S EXPRESSION VECTOR

The five mutant 4S cDNA subclones in M13mp18 were transferred to a suitable eukaryotic expression vector to enable the transfection and high-level expression of the 4S mutants in cell culture. The vector pRSVN.07 (obtained from Dr. Allan Robins, Department of Biochemistry, University of Adelaide) was selected for the expression of 4S. The pRSVN.07 plasmid contained the Rous sarcoma virus long terminal repeat promoter, which was used to drive the transcription of subcloned cDNA sequences. Anson *et al.* (1992b) utilized the expression construct pRSVN.4S.08, which contained a modified 4S cDNA clone inserted into pRSVN.07, to achieve the high level expression of wild type human 4S in CHO cells. They constructed a modified 4S cDNA clone in order to convert the incomplete coding region of the 4S cDNA fragment within pASB2-5 into a full-length 4S reading frame. In addition, a 30-bp sequence derived from the rat prepro-insulin gene was added to the 5'-untranslated portion of the 4S cDNA, in an attempt to ensure efficient translation of the transcribed 4S mRNA. The 5'-non-coding sequence from the rat prepro-insulin gene was reported to greatly enhance the translation efficiency of interleukin-2 (Cullen, 1988) and iduronate-2-sulphatase mRNAs (Dr. Don Anson, personal communication; Peter Wilson, 1992).

A potential problem of the *in vitro* mutagenesis procedure, albeit of low probability, was the generation of spurious mutations in addition to the designed mutations within the 4S cDNA. The spurious mutations could arise from the aberrant binding of the mutagenic oligonucleotide to secondary sites aside from the designed

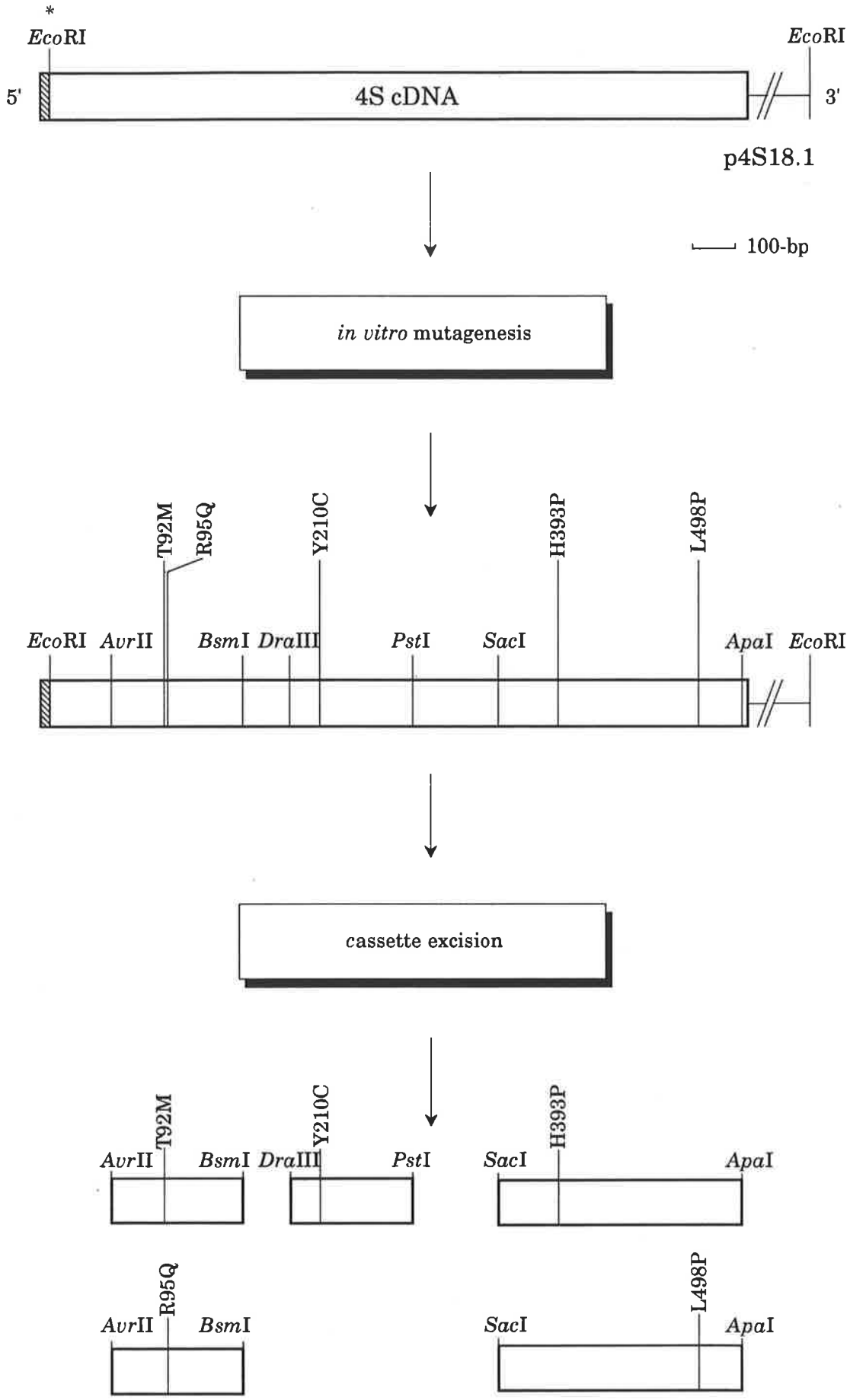
primary site (cited in Kramer & Fritz, 1987). Spurious mutations could also be generated directly by DNA polymerase during DNA synthesis *in vitro*, as the error rate is several orders of magnitude greater than *in vivo*. The spurious mutations pose a serious threat if undetected, as they may confound the interpretation of the effects of the legitimate mutations. A prudent approach would be to determine the entire cDNA sequence of each 4S mutant by DNA sequencing, to ensure that only the intended mutation was present prior to expression in tissue culture. However, this approach was not considered feasible, as the resources to perform the predicted DNA sequencing work-load (about 24-kb) were not available.

5.4.1 CASSETTE SCHEME

Therefore, an alternative strategy was designed, in which small restriction fragments or `cassettes' were utilized to shuttle the legitimate 4S mutations from the *in vitro* mutagenized 4S cDNA clones to the 4S expression vector. It was considered that the use of small mutant cassettes would reduce the probability of transferring illegitimate mutations to the 4S expression vector, compared to the transfer of the entire *in vitro* mutagenized 4S cDNA. The relatively small size of the cassette would also reduce the task of verification of the mutant DNA sequence. The cassette approach was also favoured as it enabled greater flexibility in the construction of the 4S mutant expression vectors. For example, the cassette scheme could utilize mutant cassettes generated either by *in vitro* mutagenesis, or by PCR amplification directly from a patient sample using a high fidelity Taq polymerase, as suggested previously in Section 5.3. The construction of the mutant 4S expression plasmids from the *in vitro* mutagenized 4S cDNA clones is presented below.

FIGURE 5.1 CASSETTE SCHEME FOR THE ISOLATION OF 4S MUTATIONS SYNTHESIZED *IN VITRO*.

A scheme was designed which enabled the removal of an *in vitro* synthesized 4S mutation from the 4S cDNA as a short DNA fragment or `cassette'. The 4S mutations were synthesized as described in the text, using p4S18.1 as the template which contained the 2.2-kb *Eco*RI 4S cDNA insert of pASB2-5. The 5'-end of this 4S cDNA (*) was missing the first 38-bp (shaded) of the complete coding region of the 4S cDNA. The T92M and R95Q mutations were individually isolated by digestion of the mutated 4S cDNA with the restriction endonucleases *Avr*II and *Bsm*I (300-bp fragment), the Y210C mutation was obtained using *Dra*III and *Pst*I (272-bp fragment), and the H393P and L498P mutations were obtained individually by digestion with *Sac*I and *Apa*I (549-bp fragment). The cassettes were then cloned into the appropriate cassette acceptor vector.


*
EcoRI
EcoRI
5'
4S cDNA
3'
p4S18.1
100-bp
in vitro mutagenesis
T92M
R95Q
Y210C
H393P
L498P
EcoRI
AvrII
BsmI
DraIII
PstI
SacI
ApaI
EcoRI
cassette excision
T92M
Y210C
H393P
AvrII
BsmI
DraIII
PstI
SacI
ApaI
R95Q
L498P
AvrII
BsmI
SacI
ApaI

5.4.1a ISOLATION OF THE MUTANT 4S CASSETTES

The cassettes together spanned the majority of the complete 4S cDNA coding sequence, enabling the excision of each intended 4S mutation on one of three small DNA fragments (Fig. 5.1). Based on this design, RF DNA from a single selected M13mp18 subclone of each 4S mutant cDNA (p4S-T92M, p4S-R95Q, p4S-Y210C, p4S-H393P and p4S-L498P) was digested with the appropriate pair of restriction endonucleases, and the cassette which contained the mutation of interest was isolated and purified by agarose gel electrophoresis. The T92M and R95Q mutations were individually isolated on 300-bp *Avr*II-*Bsm*I cassettes, the Y210C mutation was isolated on a 272-bp *Dra*III-*Pst*I cassette, and the H393P and L498P mutations were individually isolated on 549-bp *Sac*I-*Apa*I cassettes (Fig. 5.1). Each mutant cassette was then cloned into the appropriate one of three pUCXB4S `cassette acceptor' vectors. These vectors contained the modified full-length 4S cDNA clone from which the corresponding wild type cassette had been removed by restriction endonuclease digestion. The three pUCXB4S cassette acceptor vectors were constructed from pUC18 as described below.

5.4.1b CONSTRUCTION OF THE CASSETTE ACCEPTOR VECTORS

The standard polylinker in pUC18 contained *Pst*I and *Sac*I sites which would interfere with the successful assembly of the mutant cassettes and the cassette acceptor vectors if unmodified. Therefore the pUC18 polylinker was replaced by a truncated alternative (Fig. 5.2). The standard polylinker was removed by digestion of pUC18 DNA with the restriction endonucleases *Hind*III and *Eco*RI followed by purification and isolation of the large fragment containing the remainder of pUC18 by gel electrophoresis. Two complementary oligonucleotide linkers were designed which together comprised an alternative polylinker containing only four restriction sites; internal sites for *Xba*I and *Bam*HI, and cohesive *Hind*III and *Eco*RI

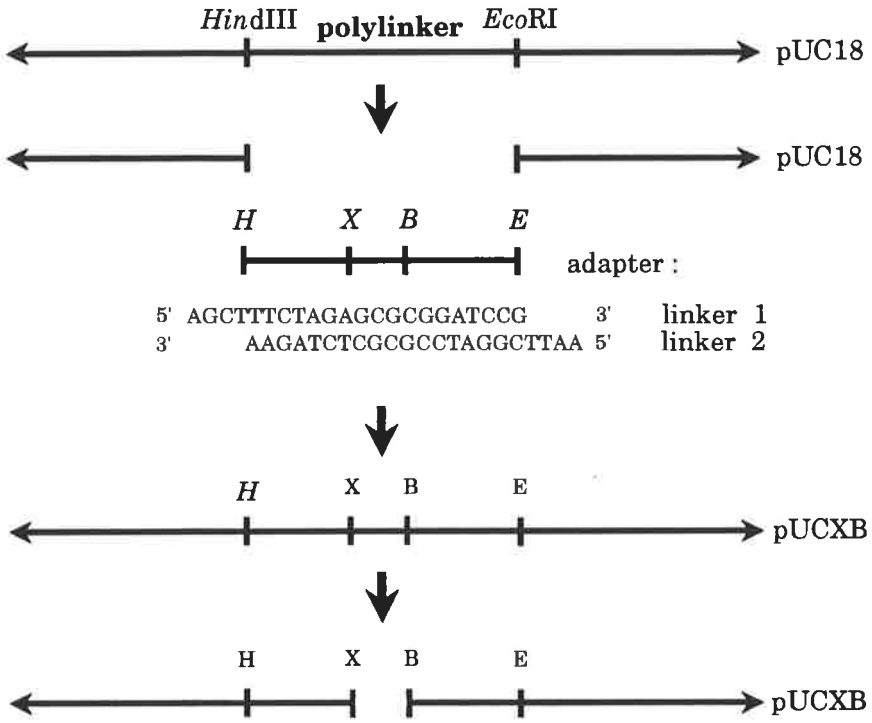


FIGURE 5.2 CONSTRUCTION OF THE VECTOR pUCXB

The pUC18 vector was digested with *Hin*dIII (H) and *Eco*RI (E), the fragments were resolved by agarose gel electrophoresis, and the linearized vector was isolated from the gel. The two linker primers 1 and 2 were heat denatured, annealed, and ligated into the cohesive ends of the linearized pUC vector, to produce pUCXB. DNA of pUCXB was isolated and digested with *Xba*I (X) and *Bam*HI (B). The linearized vector was purified by gel electrophoresis.

overhangs at the ends. The two oligonucleotide linkers were then denatured, annealed and directionally cloned into the *Hind*III-*Eco*RI cut pUC18 vector to create pUCXB. Plasmid pUCXB DNA was prepared from selected transformed cells, and was digested with restriction endonucleases *Xba*I and *Bam*HI, followed by purification and isolation of the linearized vector by gel electrophoresis.

A 1.7-kb 4S cDNA fragment containing the complete coding region (and a modified 5'-untranslated region) was excised from the 4S expression vector pRSVN.4S.08 by digestion with restriction endonucleases *Xba*I and *Bam*HI (Fig. 5.3). The 1.7-kb *Xba*I-*Bam*HI fragment was then purified by agarose gel electrophoresis, and directionally cloned into the pUCXB vector to produce pUCXB4S. pUCXB4S DNA was prepared from selected transformed cells, and the required configuration was confirmed by analytical digestion with restriction endonucleases. The clone pUCXB4S.1 was selected for further manipulations.

pUCXB4S.1 DNA was digested in three separate restriction endonuclease double digests (Fig. 5.3). Each preparative digest contained one of the three pairs of enzymes (*Avr*II-*Bsm*I, *Dra*III-*Pst*I and *Sac*I-*Apa*I) previously used to generate the three mutant cassettes. The products of each digestion were subjected to agarose gel electrophoresis in order to resolve each required cassette acceptor vector from the unwanted wild type 4S cassette fragment. The linearized DNA of each of the three cassette acceptor vectors (pUCXB4S/AB, pUCXB4S/DP, and pUCXB4S/SAp) was then purified.

5.4.1c ASSEMBLY OF THE MUTANT 4S EXPRESSION CONSTRUCTS

The five mutant cassettes containing the 4S mutations were cloned into the appropriate cassette acceptor vector (Fig. 5.4). For each mutant, about six transformed colonies were selected and small-scale DNA preps were performed.

Putative constructs containing mutant cassettes were identified by ASO analysis of plasmid DNA dot-blots (data not shown).

For each of the 4S mutants, the full-length 4S cDNA containing the mutant cassette was then excized from the pUCXB4S vector as a 1.7-kb fragment by digestion with *Xba*I and *Bam*HI. The fragment was purified by gel electrophoresis and directionally recloned into the *Xba*I and *Bam*HI sites of the pRSVN.4S.08 expression construct from which the 4S cDNA had been previously removed by digestion with *Xba*I and *Bam*HI (pRSVN.07 expression vector, Fig. 5.4). For each mutant construct, about four transformed colonies were selected and small-scale DNA preps were performed. Putative clones containing mutant 4S cDNA inserts were verified by ASO analysis of plasmid DNA dot-blots. The integrity of the 4S cDNA insert was checked by analytical digestion of the DNA with *Xba*I and *Bam*HI for the presence of an intact 1.7-kb 4S cDNA fragment (data not shown). Positive clones for each of the 4S mutations were selected, large-scale DNA preps were performed, and the DNA was further purified by caesium chloride gradient ultracentrifugation for subsequent transfection into CHO cells. The five mutant expression vectors derived from the wild type 4S cDNA expression construct pRSVN.4S.08, were designated pRSVN.4S-T92M, pRSVN.4S-R95Q, pRSVN.4S-Y210C, pRSVN.4S-H393P and pRSVN.4S-L498P.

Prior to transfection, the DNA sequence of each of the intended 4S mutations was confirmed by ASO hybridization analysis. After transfection, the mutation in each integrated construct was confirmed by direct DNA sequencing of PCR products derived from the selected cell lines (data not shown).

FIGURE 5.3 CONSTRUCTION OF THE 4S MUTANT CASSETTE ACCEPTOR VECTORS

The wild type 4S expression vector pRSVN.4S.08 was digested with restriction endonucleases *Xba*I and *Bam*HI, and a 1.7-kb 4S cDNA fragment containing the complete coding region (and a modified 5'-untranslated region) was excised. The 1.7-kb *Xba*I-*Bam*HI fragment was then purified by agarose gel electrophoresis, and directionally cloned into the pUCXB vector to produce pUCXB4S.1. pUCXB4S.1 DNA was digested in three separate restriction endonuclease double digests. Each preparative digest contained one of the three pairs of enzymes (*Avr*II-*Bsm*I, *Dra*III-*Pst*I and *Sac*I-*Apa*I) previously used to generate the three mutant cassettes. The products of each digestion were subjected to agarose gel electrophoresis in order to resolve each required cassette acceptor vector from the unwanted wild type 4S cassette fragment. The linearized DNA of each of the three cassette acceptor vectors (pUCXB4S/AB, pUCXB4S/DP, and pUCXB4S/SAp) was then purified.

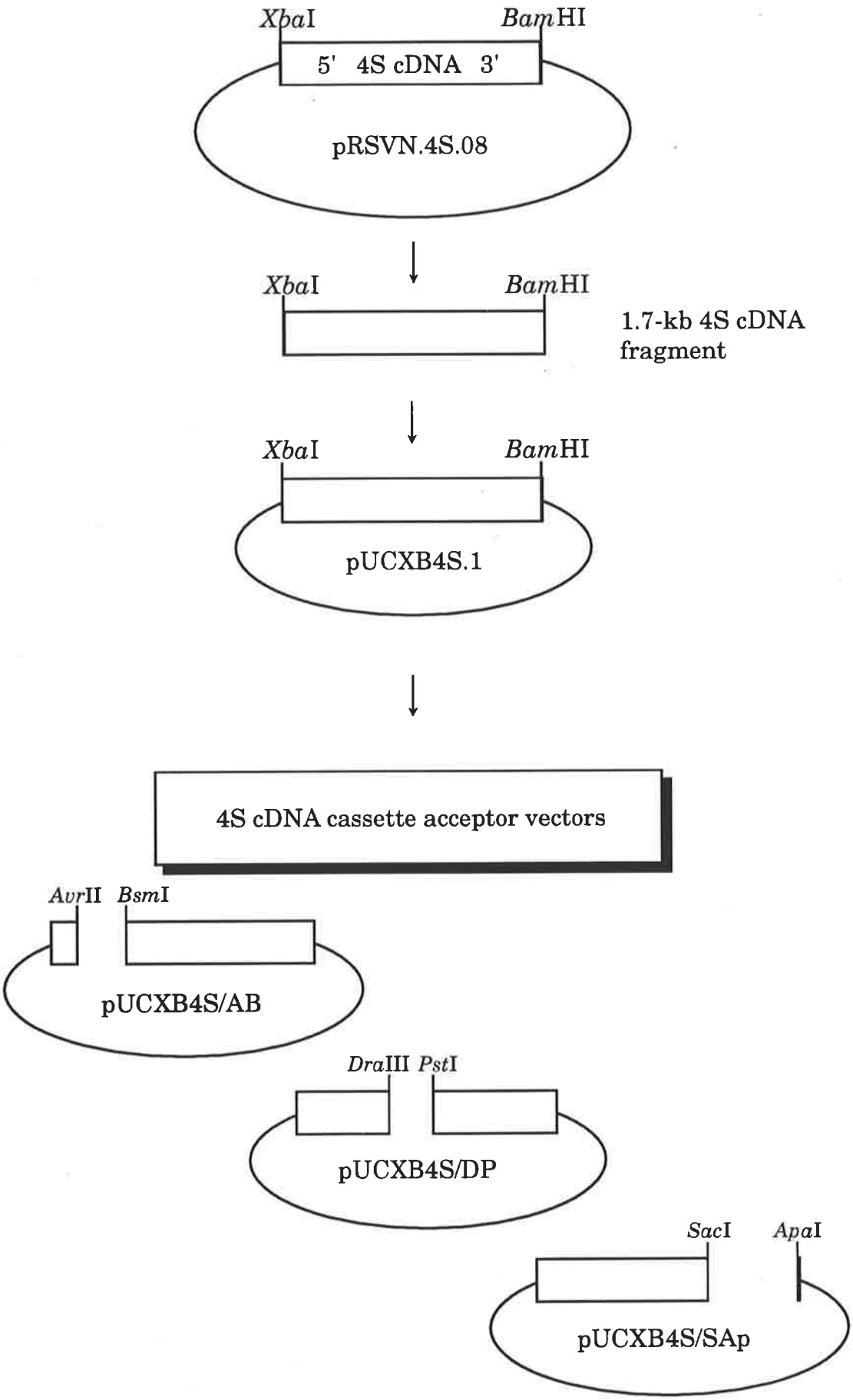
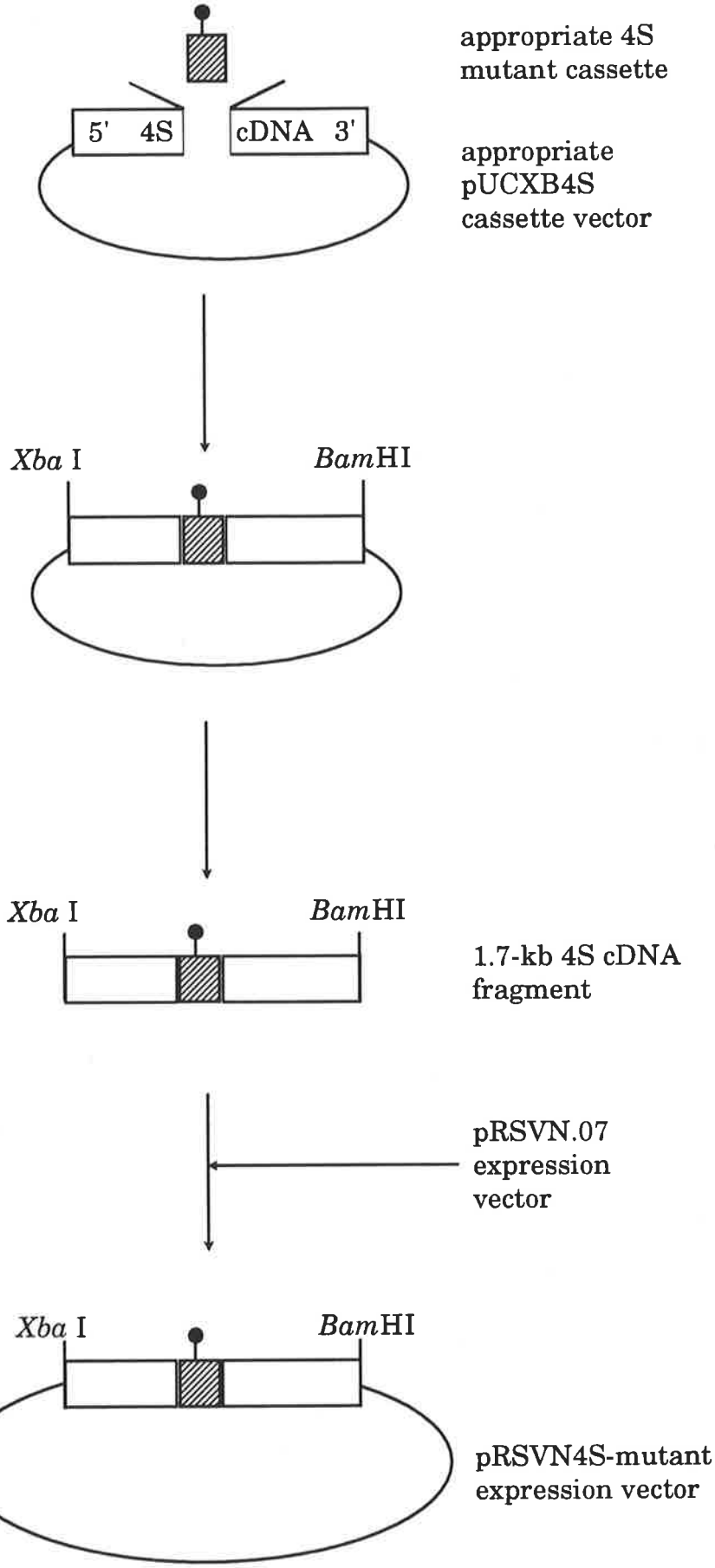

XbaI
BamHI
5' 4S cDNA 3'
pRSVN.4S.08
XbaI
BamHI
1.7-kb 4S cDNA fragment
XbaI
BamHI
pUCXB4S.1
4S cDNA cassette acceptor vectors
AvrII
BsmI
pUCXB4S/AB
DraIII
PstI
pUCXB4S/DP
SacI
ApaI
pUCXB4S/SAp

FIGURE 5.4 ASSEMBLY OF THE 4S MUTANT EXPRESSION CONSTRUCTS

Each 4S mutant cassette (Fig. 5.1) was cloned into the appropriate cassette acceptor vector (Fig. 5.3), and the DNA isolated. For each of the 4S mutants, the 4S cDNA was excized as a 1.7-kb *Xba*I-*Bam*HI fragment, which was then cloned into the pRSVN.4S.08 expression construct from which the 4S cDNA had been previously removed by digestion with *Xba*I and *Bam*HI (pRSVN.07 expression vector) and the DNA purified. Five mutant 4S expression constructs were made; pRSVN.4S-T92M, pRSVN.4S-R95Q, pRSVN.4S-Y210C, pRSVN.4S-H393P and pRSVN.4S-L498P, which were derived from the wild type 4S cDNA expression construct pRSVN.4S.08.


appropriate 4S mutant cassette
5' 4S
cDNA 3'
appropriate pUCXB4S cassette vector
Xba I
BamHI
Xba I
BamHI
1.7-kb 4S cDNA fragment
pRSVN.07 expression vector
Xba I
BamHI
pRSVN4S-mutant expression vector

5.5 EXPRESSION & PRELIMINARY CHARACTERIZATION OF 4S MUTATIONS

5.5.1 CLONE SELECTION

The wild type 4S cDNA expression construct pRSVN.4S.08, and the five mutant derivatives; pRSVN.4S-T92M, pRSVN.4S-R95Q, pRSVN.4S-Y210C, pRSVN.4S-H393P and pRSVN.4S-L498P, were transfected into CHO-DK1 cells by electroporation as described in Section 5.2.3a. The transfected CHO cell lines were designated as 4S-N, 4S-T92M, 4S-R95Q, 4S-Y210C, 4S-H393P and 4S-L498P respectively. For each of the six constructs, 12 stably-transfected G418 resistant colonies were isolated and grown for further analysis. In the case of 4S-N, 4S-H393P and 4S-L498P cells, mass cultures containing the remaining pooled G418 resistant colonies were also set up for further analysis. The mass cultures were performed as they enabled the determination of the residual activity of a 4S mutant averaged over many individual clones. These results could then be compared with the residual activity of single selected colonies. Only 4S-N, 4S-H393P and 4S-L498P lines were mass cultured, as they were used in preliminary optimization of the techniques used in the subsequent analysis of all six cell lines.

For each of the six constructs, the clone line with the highest arylsulphatase activity was selected. Anson *et al.* (1992b) had previously shown that CHO cells transfected with pRSVN.4S.08 secreted active 4S into the culture medium. Therefore, conditioned media from 12 individual 4S-N lines (G418-resistant clones transfected with the pRSVN.4S.08 expression construct) were assayed for total arylsulphatase activity, as described in Section 5.2.4a. A single clone, 4S-N.3, was selected for further analysis as it exhibited the highest arylsulphatase activity secreted in the medium, about 30-fold greater (28.5 Units/ml) than the untransfected CHO-DK1 cell control (1.0 Units/ml). The arylsulphatase activity detected in the untransfected control represents the background activity of the

samples. The background activity is contributed by the foetal calf serum, and was subsequently reduced about 2-3-fold by heat-inactivation of the serum at 56°C for 60 min prior to addition to the F-12 nutrient media.

For the five mutant 4S cell lines 4S-T92M, 4S-R95Q, 4S-Y210C, 4S-H393P, and 4S-L498P, selection of the highest activity clones required a more sensitive and specific assay, as the preliminary analysis of the 4S-H393P and 4S-L498P cell lines was unable to detect total arylsulphatase activity above the untransfected CHO background, in either the conditioned media or the cell extracts. Consequently, an immunocapture technique was used to determine the residual activity of each of the mutants; the principle of the assay is shown in Fig. 5.5. The specificity of the assay is achieved by the purification of recombinant human 4S away from other proteins within the sample (including CHO and foetal bovine arylsulphatases), by anti-human 4S monoclonal antibodies. The specificity of the immunocapture assay resulted in increased sensitivity, as the background was reduced relative to the total arylsulphatase assay.

Cellular extracts were obtained from each of the 12 individual clones of each of the 4S mutant transfected CHO cells, and also from the wild type cell clone 4S-N.3, the mass cultures of 4S-N, 4S-H393P and 4S-L498P, and the un-transfected CHO control. The immunocapture assay was performed using the monoclonal antibody ASB F66, according to the procedure in Section 5.2.4b. The residual activities of the 12 individual clones of each 4S mutant exhibited a quasi-normal distribution (data not shown). The activities obtained from the mass cultures confirmed that a representative population of clones had been selected for each construct, as the activity of each of the 4S-N, 4S-H393P and 4S-L498P mass cultures was close to the mean of the activities of the respective 12 individual clonal lines (data not shown). The two highest activity lines of each 4S mutant construct were selected for further analysis. The variation in 4MUS activity between the different clones of

each construct (wild type or mutant) was presumably due to variation in the transcriptional activity of the expression vector, which integrates into the CHO genome at random and is influenced by the chromosome sequence context. The clonal variation of stably transfected cells and the consequences for the determination of the relative residual activity is discussed at the end of the Chapter.

The selected clone pair of each mutant construct, the 4S-N.3 clone, the 4S-N.MC mass culture, and the untransfected CHO cells, were cultured and the cell extracts obtained as described in Section 5.2.3b. Cellular extracts (batch 1) were obtained from cultured cells that were grown and harvested at confluency, about 2 days after the final media change. For each extract, the total arylsulphatase activity and the ASB F66-immunocaptured 4S activity were determined using procedures described previously. In addition, the 4S protein was quantified immunochemically (Section 5.2.5). The results of each assay will now be presented and discussed in detail.

5.5.2 INITIAL *IN VITRO* ANALYSIS

5.5.2a RESIDUAL ENZYME ACTIVITY

The total arylsulphatase activities of the cell extracts from the 4S mutant cell lines were all markedly reduced relative to the wild type human 4S control (Table 5.1). The activity values were similar for duplicate samples of the 4S-N.3 clone or the two clonal lines of each 4S mutant. The 4S-Y210C mutants had the highest residual activity of the 4S mutants, about 2% of wild type and 200-fold greater than the untransfected CHO control. The 4S-T92M mutants had a very low residual activity of about 0.03% (3X background), while the residual activity of the 4S-R95Q mutants was about 0.02% (2X background) and thus at the resolution limit of the assay to distinguish above background activity. The total arylsulphatase activities

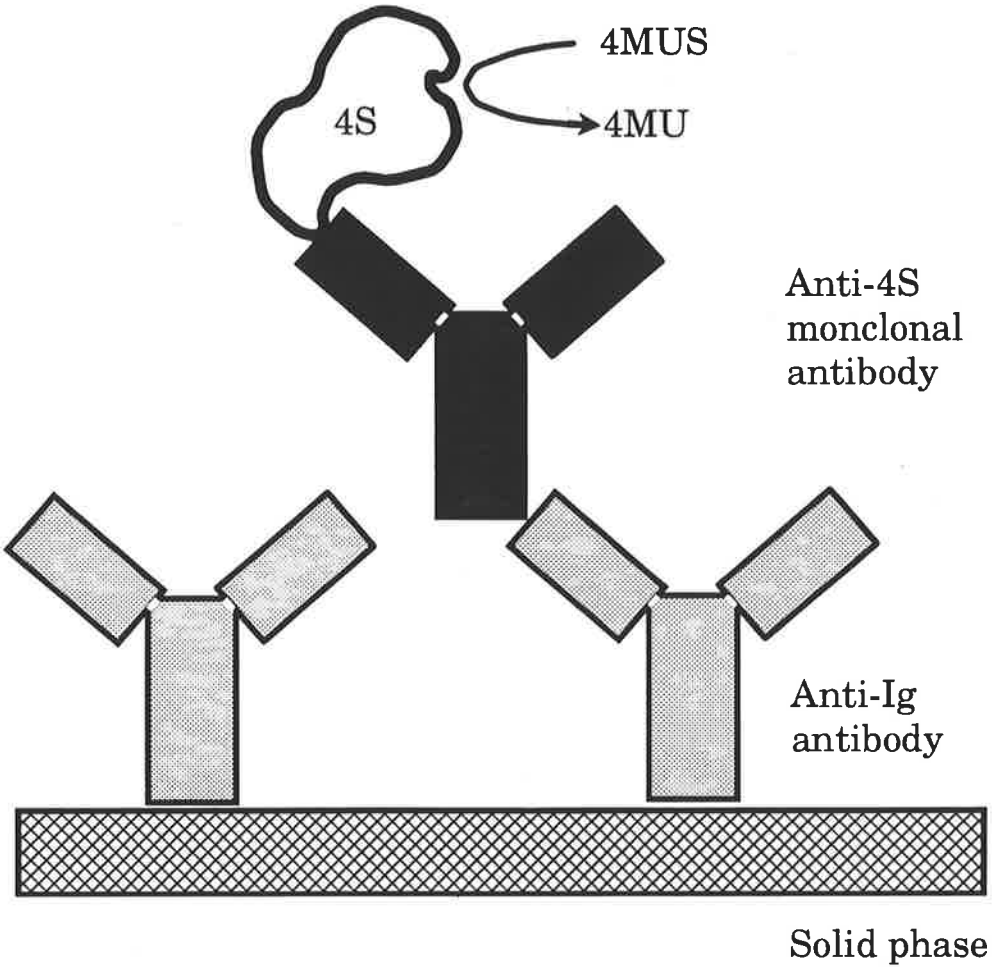


FIGURE 5.5 PRINCIPLE OF THE 4S IMMUNOCAPTURE & 4MUS ACTIVITY ASSAY

The immunocapture assay of 4S-specific arylsulphatase activity was a modification of a previously described method (Brooks *et al.*, 1991a). An anti-Ig antibody bound to a microtitre plate well (solid phase) was used to bind an anti-4S monoclonal antibody (e.g. ASB F66). The monoclonal antibody was then used to `capture' 4S from the sample to be assayed. Bound 4S was assayed using the substrate 4-methylumbelliferyl sulphate (4MUS); the activity was determined by measurement of the hydrolysis of 4MUS to the fluorescent product 4-methylumbelliferone (4MU). The assay is described in detail in Section 5.2.4b.

TABLE 5.1 TOTAL ARYLSULPHATASE ACTIVITY IN TRANSFECTED CHO CELL EXTRACTS

A total arylsulphatase activity assay using the 4MUS substrate was performed on cell extracts of CHO cells transfected with the wild type and mutant 4S expression constructs. The experimental details of the assay are described in Sections 5.2.4a and 5.5.2a. There were two independent cell lines for each of the 4S mutants, and duplicate assays were performed on the CHO (untransfected) and 4S-N.3 (wild type human 4S) samples. MC = mass culture. For each of the extracts (neat and/or serial dilutions where appropriate), 5 µl was mixed with the 4MUS substrate and the reaction was incubated 37°C for 60 min. Enzyme activity was based the hydrolysis of 4MUS to 4MU, which was measured fluorometrically at the emission wavelength of 446nm using an excitation wavelength of 366nm. The E_{446nm} values were converted to units/ml, and then corrected for cell protein in the extract. Units/mg = nmoles/min/mg cell protein. The % relative activity for each extract was calculated relative to the mean activity of the duplicate 4S-N.3 samples, which was assigned a value of 100%. The CHO background activity was not subtracted from the residual activity values. Cell extracts from the H393P and L498P mutants were considered to have no 4MUS activity above the CHO background.

Cell extract	Activity (Units/mg)	% Relative activity
4S-N.3	560	95
4S-N.3	620	105
4S-N.MC	300	50
4S-T92M.4	0.18	0.03
4S-T92M.9	0.21	0.035
4S-R95Q.4	0.10	0.015
4S-R95Q.12	0.11	0.02
4S-Y210C.2	12	2.0
4S-Y210C.5	14	2.4
4S-H393P.2	0.05	0.008
4S-H393P.11	0.06	0.01
4S-L498P.1	0.05	0.008
4S-L498P.7	0.05	0.008
CHO control	0.05	0.008
CHO control	0.04	0.006

of the 4S-H393P and 4S-L498P mutants were indistinguishable from the CHO control.

The activities of the extracts were also assayed using the immunocapture technique, which was specific for 4S and had an increased sensitivity compared to the total arylsulphatase assay above. The cellular extract samples were immunocaptured with ASB F66 and the activity determined by reference to a standard curve derived from duplicate serial dilutions of a 4S preparation with known activity. For each of the 4S mutant constructs, two clonal lines were assayed, while only a single line was assayed for 4S-N.3, 4S-N.MC (mass culture), and the CHO control. The activity of each of the 4S-N.3, 4S-N.MC, 4S-T92M and 4S-Y210C lines was determined over a range of serial dilutions, due to the higher activity of the neat cell extracts. In contrast, the activity of each of the control CHO, 4S-R95Q, 4S-H393P, and 4S-L498P lines was determined in duplicate on neat samples only, due to their lower activity.

The immunocaptured 4S activities of the cell extracts from the 4S mutant cell lines were all markedly reduced relative to the wild type human 4S control (Table 5.2). Broadly speaking, the 4S activities obtained using the immunocapture assay were similar to the results obtained for the total arylsulphatase assay. A possible exception was the wild type clone 4S-N.3, which differed almost 2-fold between the two assays. In retrospect, this difference was considered artefactual. The activity calculated for the 4S-N.3 clone in the total arylsulphatase activity assay was an underestimate, as the E_{446nm} value was later found to be outside the linear response of the spectrofluorometer. The specificity of the immunoassay achieved a 5-10-fold reduction in the background arylsulphatase activity, which enabled the residual activities of the low activity 4S mutants (4S-R95Q, 4S-H393P, and 4S-L498P) to be measured with greater precision. The activity values were similar for

duplicate samples of the 4S-N.3 clone or the two clonal lines of each 4S mutant (Table 5.2).

An unusual and unexpected phenomenon, a `dilution effect', was observed with the 4S-T92M mutant cell extracts. Serial dilution of the 4S-T92M cell extracts resulted in a concomitant increase in the activity (Units/mg cell protein), from 0.1 Units/mg cell protein in the neat extract to 0.4 Units/mg cell protein in an 8-fold dilution (Table 5.2). In contrast, the expected observation of a fairly constant activity over a range of dilutions was observed in the cell lines 4S-N, 4S-Y210C, and CHO cellular extracts. The phenomenon was investigated further in subsequent experiments, where it was also observed in the 4S-R95Q mutant. A possible explanation for the `dilution effect' is that at the higher concentrations of 4S protein present in the neat extract, the mutation promotes the aggregation of the mutant enzyme and the subsequent reduction in the specific activity. Dilution of the extract may reduce aggregation and hence increase enzyme activity. The phenomenon was not an artefact of the fluorimeter, as the activities were observed in the linear portion of the standard curve of the assay.

The reduced *in vitro* activities of the 4S mutants relative to the wild type could be due to a number of factors including reduced specific activity, reduced synthesis or stability, or incorrect cellular trafficking. Alternatively, the reduction in activity may be due to *in vitro* artefacts. The reduced *in vitro* activities of the 4S mutants were unlikely to be an artefact of the ASB F66 antibody used in the immunocapture. The approximate correspondence between the immunocaptured and free enzyme activities for the wild type and mutant 4S (Tables 5.1 and 5.2), suggested that ASB F66 bound the mutant 4S enzymes with about the same affinity as wild type 4S, and did not appear to have either considerably enhanced or retarded the residual enzyme activities. In order to further characterize each mutant, the residual 4S protein was also determined.

TABLE 5.2 IMMUNOCAPTURED (ASBF66) 4S-SPECIFIC ARYLSULPHATASE ACTIVITY IN TRANSFECTED CHO CELL EXTRACTS

An immunocaptured (ASBF66) 4S-specific arylsulphatase activity assay using the 4MUS substrate was performed on cell extracts of CHO cells transfected with the wild type and mutant 4S expression constructs. The experimental details of the assay are described in Sections 5.2.4b and 5.5.2a. As for the total arylsulphatase assay, there were two independent cell lines for each of the 4S mutants, and duplicate assays were performed on the CHO (untransfected) and 4S-N.3 (normal wild type 4S) samples. MC = mass culture. For each of the extracts (neat and/or serial dilutions where appropriate), 50 µl was assayed at 37°C for 5 hr. The progress of the reaction was determined by fluorometric measurement of the liberated 4MU product. The E_{446nm} values were converted to units/ml by reference to a standard curve of known 4S activities, and then corrected for cell protein in the extract. Activity = units/mg = nmoles/min/mg cell protein. The % activities of the mutant 4S samples was calculated as described in Table 5.1. The CHO background was not subtracted from the residual activity values. The residual activity of the 4S-T92M cell line extracts ranged from 0.1 units/mg in neat extracts to 0.4 units/mg at an 8-fold dilution, and therefore was concentration dependent. Cell extracts from the H393P and L498P mutants were considered to have no 4MUS activity above the CHO background. For purposes of comparison, 4S activity immunocaptured using ASB 4.1 from 11 normal cultured fibroblasts is also shown (Doug Brooks, unpublished data).

Cell extract	Activity (Units/mg)	% Relative Activity
4S-N.3	1080	102
4S-N.3	1044	98
4S-N.MC	552	52
4S-T92M.4	0.1->0.4	0.01->0.03
4S-T92M.9	0.1->0.4	0.01->0.03
4S-R95Q.4	0.03	0.003
4S-R95Q.12	0.03	0.003
4S-Y210C.2	14	1.3
4S-Y210C.5	18	1.7
4S-H393P.2	0.02	0.002
4S-H393P.11	0.01	0.001
4S-L498P.1	0.01	0.001
4S-L498P.7	0.02	0.002
CHO background	0.01	0.001
CHO background	0.01	0.001
normal cultured fibroblasts	*0.41 - 0.71*	

5.5.2b RESIDUAL 4S PROTEIN

The 4S protein was quantified in the CHO clones expressing wild type and mutant 4S using a heterogeneous immunoassay or sandwich ELISA technique. The principle of the 4S immunoquantification assay is shown in Fig. 5.6, and was performed as described in Section 5.2.5. In brief, the 4S protein from neat and serial dilutions of cell extracts was first bound to ELISA wells using an anti-4S polyclonal antibody. The bound protein was then quantified using the monoclonal antibody ASB F66, by reference to a standard curve of known quantities of 4S protein.

The residual 4S protein in extracts of CHO cells transfected with the wild type and mutant 4S constructs, and untransfected CHO controls, is shown in Table 5.3. The heterologous expression of human recombinant 4S in cultured CHO cells was at a high level. For example, human recombinant wild type 4S protein and activity were expressed at a level about 500-2,000-fold higher compared to wild type human 4S in cultured fibroblasts (Tables 5.2 and 5.3). At this level, the recombinant 4S comprised about 6% of total cell protein in cultured CHO cells. All of the cell lines transfected with mutant 4S constructs expressed much reduced 4S protein quantities (<5%) relative to the wild type 4S.N construct. In particular, the 4S-H393P and 4S-L498P mutants were not detectable, which suggested that these mutants contained 4S protein below the detection limit of the assay of about 20-50 pg of 4S per 50 µl sample, or about 50-125 pg per mg of cell protein.

The reduced concentration of mutant 4S protein relative to the wild type control was unlikely to be an assay artefact due to a reduced affinity of ASB F66 for the mutant proteins. Firstly, the immunocapture results above suggested that ASB F66 bound all the mutants with approximately equal affinity. Secondly, an ELISA performed with a panel of monoclonal antibodies (ASB 4.1, 15.1, 16.1, 33.1, 58.3

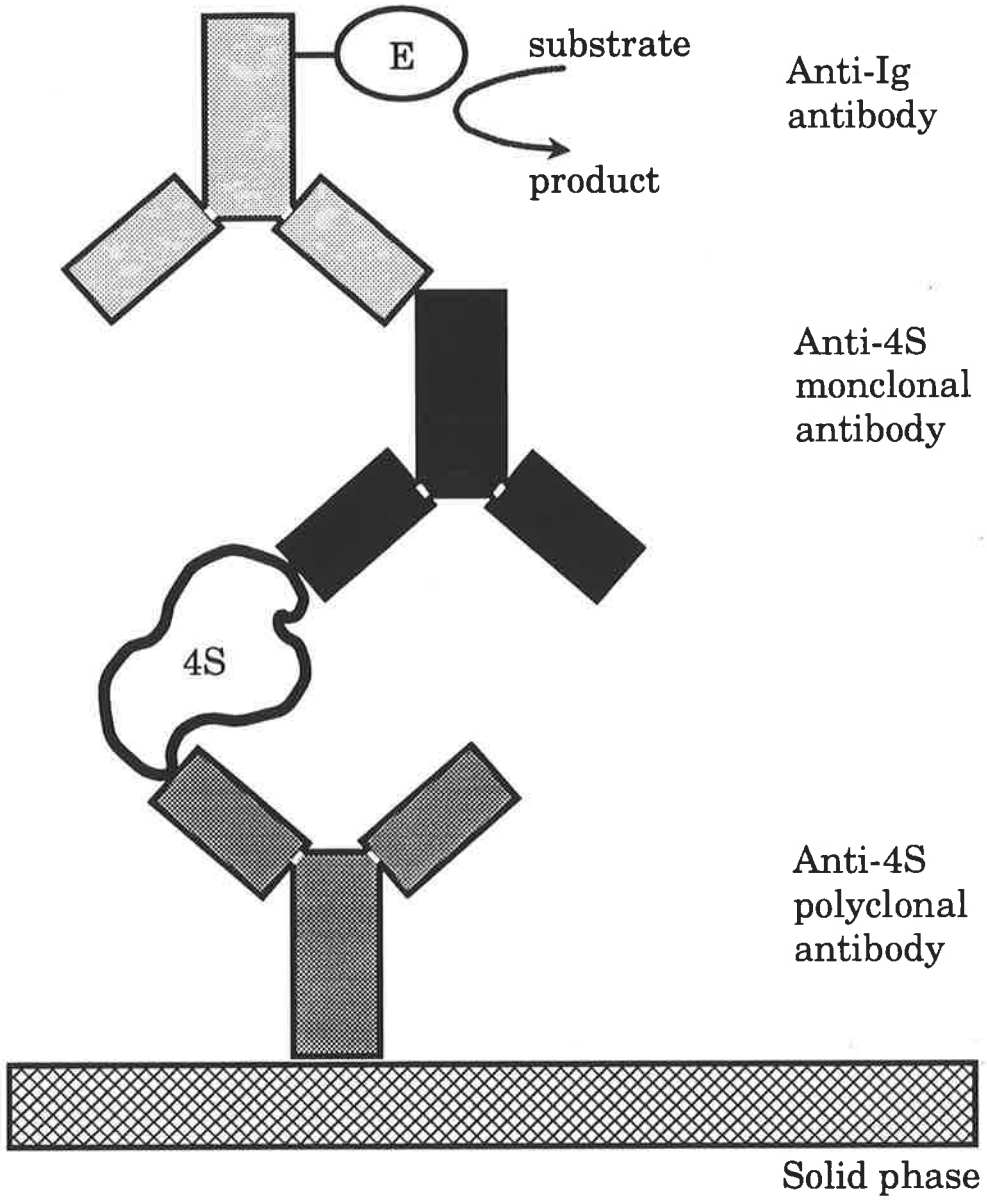


FIGURE 5.6 PRINCIPLE OF THE 4S PROTEIN IMMUNO-QUANTIFICATION ASSAY

Wild type and mutant 4S protein were measured by an heterogeneous immunoassay, as reviewed in Brooks *et al.* (1993). An immunoaffinity-purified anti-4S polyclonal antibody bound to a microtitre plate well (solid phase) was used to bind the 4S protein. An anti-4S monoclonal antibody (e.g. ASB F66) was then allowed to bind to 4S, and a peroxidase-labelled (E) anti-Ig antibody was used to detect the bound protein colourimetrically using the substrate ABTS. The assay is described in detail in Section 5.2.5.

TABLE 5.3 IMMUNOQUANTIFICATION OF 4S PROTEIN IN TRANSFECTED CHO CELL EXTRACTS

Immunoquantification of 4S protein using the monoclonal antibody ASB F66 was performed on cell extracts of CHO cells transfected with the wild type and mutant 4S expression constructs. The experimental details of the assay are described in Sections 5.2.5 and 5.5.2b. As for the activity assays, there were two independent cell lines for each of the 4S mutants, and duplicate assays were performed on the CHO (untransfected) and 4S-N.3 (normal wild type 4S) samples. MC = mass culture. For each of the extracts (neat and/or serial dilutions where appropriate), 50 µl was assayed using the peroxidase chromogenic substrate ABTS (2,2'-azino-di[3-ethyl-benzthiazoline-6-sulphonic acid]), and the 4S protein concentration (µg/mg cell protein) was determined with reference to a standard curve of known concentrations of 4S protein. A value of 0.0001 µg 4S /mg cell protein corresponded to the background of the immunoquantification assay. The % relative 4S protein for each extract was calculated relative to the mean of the protein values for the duplicate 4S-N.3 samples, which was assigned a value of 100%. Cell extracts from the H393P and L498P mutants were considered to have no 4S protein above the CHO background. For purposes of comparison, 4S protein immunoquantified using ASB 4.1 from 10 normal cultured fibroblasts is also shown (Brooks *et al.*, 1991b).

Cell extract	4S protein (μg / mg)	% Relative 4S protein
4S-N.3	57	101
4S-N.3	56	99
4S-N.MC	28	48
4S-T92M.4	0.13	0.24
4S-T92M.9	0.17	0.30
4S-R95Q.4	0.88	1.6
4S-R95Q.12	0.76	1.3
4S-Y210C.2	2.3	4.1
4S-Y210C.5	2.8	5.0
4S-H393P.2	< 0.0001	-
4S-H393P.11	< 0.0001	-
4S-L498P.1	< 0.0001	-
4S-L498P.7	< 0.0001	-
CHO background	< 0.0001	-
CHO background	< 0.0001	-
normal cultured fibroblasts	*0.04 - 0.21*	

and 59.2), detected relative residual protein concentrations in the 4S mutants that were in agreement with the ASB F66 results (data not shown).

5.5.2c SPECIFIC ACTIVITY

The residual 4S protein and immunocaptured 4S activity data were used to calculate the specific activity of the wild type and mutant recombinant human 4S toward the 4MUS substrate, as shown in Table 5.4. All five mutant 4S enzymes exhibit reduced specific activities in comparison to the wild type. The 4S-Y210C mutant was approximately a third of the wild type value, while the 4S-R95Q mutant showed a severe 500-fold reduction. The specific activity of the 4S-T92M mutant varied from about 3 to 15% of normal, depending on the concentration of the enzyme. The specific activities of the 4S-H393P and 4S-L498P mutants, and the CHO background, could not be determined due to insufficient enzyme activity and/or protein. The concordance between the specific activities of the two independent clonal lines of each 4S mutant, and between the 4S-N.3 clone and the 4S-N.MC mass culture, confirmed that the analytical techniques were reproducible and that the clonal cell lines were representative. In addition, the specific activity of the immunocaptured recombinant wild type human 4S (19-20 nmoles/min/μg 4S) compared closely with the value obtained from immunocaptured recombinant enzyme secreted in culture (19 nmoles/min/μg 4S: Anson *et al.*, 1992), or 4S purified from human liver or cultured fibroblasts assayed free in solution (20 nmoles/min/μg 4S: Doug Brooks, unpublished results).

Taken together, these results suggested that firstly, the highly expressed recombinant wild type 4S was enzymatically similar to wild type human 4S purified from tissue. Secondly, that the polyclonal and monoclonal antibodies used in the immunocapture assay had little affect on the enzyme activity, relative to enzyme assayed free in solution. Thirdly, the polyclonal and monoclonal antibodies

used in the ELISA assay bound both wild type and mutant 4S protein with about equal affinity.

5.5.2d DISCUSSION

The specific activity determinations presented in Table 5.4 were preliminary and required refinement. For example, the `dilution effect' observed with the residual activity of the 4S-T92M mutant had not been fully titrated and therefore the specific activity was underestimated. In addition, work presented later showed that these initial experiments were performed under non-optimal conditions of pH and incubation time for a number of the mutants. However, a valid general observation was that the 4S mutations produced a marked reduction in 4S protein. In the 4S-Y210C and 4S-T92M mutants, the reduction in protein was 20 and 370-fold respectively, while the reduction in specific activity of the enzyme was much less severe; 3 and 6-30-fold respectively (Tables 5.3 and 5.4). On the other hand, the R95Q mutation reduced the protein 70-fold, while the reduction in specific activity was more severe; about 500-fold. The 4S-H393P and 4S-L498P mutant proteins were undetectable.

The reduced concentration of mutant 4S protein in the CHO expression system compared to the wild type 4S could be due to a number of mechanisms. For example, it is conceivable that the mutant proteins may be synthesized at a reduced rate, or alternatively, the mutant proteins may be degraded at an increased rate. The latter potential mechanism is discussed in more detail in Section 5.8.1. The consistent and marked reduction in 4S activity and protein observed for all five mutants relative to the wild type made it unlikely that it was solely an artefact of clonal variation in expression of the 4S mutants. The possibility was investigated that gross changes occurred in the integrated pRSVN.4S-H393P and pRSVN.4S-L498P expression vectors, which were responsible for the absence of 4S protein or activity in these mutants. The

TABLE 5.4 SPECIFIC ACTIVITY OF WILD TYPE & MUTANT 4S TOWARDS 4MUS IN TRANSFECTED CHO CELL EXTRACTS

The specific activity (nmoles per min per μg 4S protein) was calculated for the 4S wild type and each of the 4S mutants expressed in transfected CHO cell extracts. The residual 4S activity and protein values shown in Tables 5.2 and 5.3 respectively, were corrected for the CHO background, and the specific activity was calculated by division of the corrected 4S activity by the corrected 4S protein. The 4S activity and protein of the mutants H393P, L498P, and the CHO control were too low to enable a reliable specific activity to be calculated. The % relative specific activities of the mutants was calculated relative to the mean of the specific activities of the two wild type 4S-N.3 samples, which was assigned a value of 100%.

Cell extract	Specific Activity	% Relative Specific Activity
4S-N.3	19	100
4S-N.3	19	100
4S-N.MC	20	105
4S-T92M.4	0.67->2.8	3.6->15
4S-T92M.9	0.54->2.7	2.9->14
4S-R95Q.4	0.026	0.14
4S-R95Q.12	0.030	0.16
4S-Y210C.2	6.1	32
4S-Y210C.5	6.6	35

structural integrity of the integrated expression vectors was checked by PCR analysis using genomic DNA isolated from the transfected CHO lines. The PCR B products generated from the wild type and mutant genomic DNAs were indistinguishable in size or yield (data not shown). This result suggested that the mutant expression vectors had integrated intact and no gross structural rearrangement had occurred.

The DNA sequences of all five of the designed 4S mutations were verified in each of the selected transfected CHO lines (data not shown). Genomic DNA was isolated, and PCR amplification of the integrated 4S cDNA expression construct was performed. The DNA sequence spanning the site of each mutation was then obtained from the appropriate PCR A or PCR B product using a suitable oligonucleotide primer.

The clinical severity of each of the 4S mutations was estimated from the relative residual activity of the CHO-expressed mutant enzyme. The predictions were based on the observation that in cultured patient fibroblasts, the entire spectrum of clinical phenotypes from severe to mild was clustered over a low and narrow range of residual activity from 0 to about 5% of wild type controls. As discussed in Chapter 1, Conzelmann & Sandhoff (1993/1984) predicted that small changes in the residual enzyme activity in the vicinity of a critical threshold level may have a large influence on the rate of lysosomal storage and disease severity. Thus the Y210C mutation was predicted to result in a relatively mild clinical phenotype, while the H393P, L498P and R95Q mutations were predicted to result in relatively severe clinical phenotypes. The clinical severity of the T92M mutation was predicted to lie between that of the Y210C and R95Q mutations, however the `dilution effect' *in vitro* made this assignment somewhat ambiguous.

Further analysis of the CHO-expressed recombinant 4S mutants was required in order to refine the predicted clinical and biochemical phenotypes. In particular, the `dilution effect' observed with the 4S-T92M mutant, and the apparent absence of activity and protein for the 4S-H393P and 4S-L498P mutants. The 4S-T92M `dilution effect' required confirmation using a panel of monoclonal antibodies, to rule out a ASB F66-specific artefact. A similar approach was considered necessary to test the possibility that the 4S-H393P- and 4S-L498P-transfected cell lines were expressing mutant protein that was inactive, and that was not recognized by ASB F66. The general reduction in 4S activity and protein observed in the cell extracts of the 4S mutants, prompted investigation of the culture media for residual 4S activity. This was based on the speculative possibility of incorrect targeting, in which mutant 4S enzyme activity was preferentially secreted to the culture medium rather than remaining in the cell and being targeted to the lysosome.

5.5.3 MULTIPLE ANTIBODY ANALYSIS OF RESIDUAL ENZYME ACTIVITY

In order to confirm and refine the residual activity results obtained with ASB F66, the immunocapture assay was repeated, and included three other anti-4S monoclonal antibodies. As the specific activities of the two clonal cell lines were very similar for each mutant 4S construct, one of the pair was selected for subsequent analysis. The selected clones were amplified in culture, and aliquots of conditioned media and cell extracts were obtained (batch 2). The batch 2 samples were obtained from cultured cells grown and harvested at confluency, about three days after the final media change. The activities of the cell extracts and media were determined using the immunocapture technique described in Section 5.2.4b. However, the assay was performed in parallel, with the four monoclonal antibodies ASB 4.1, ASB 56.2, ASB 58.3, and ASB F66.

To appreciate the results obtained using the ASB F66 immunocapture assay, the data used to calculate activity will be discussed. The data average a number of immunocapture experiments. The linear range of the immunocapture assay was between 0 and 50 fluorometric E_{446nm} units, but was routinely measured between 0 and 15 E_{446nm} units. Typically, the substrate background (4MUS, no extract) was about 0.1 E_{446nm} units, which was unchanged with varying incubation time (20 min to 10 hr). The uncorrected CHO extract control background was about 0.4 E_{446nm} units. The lower limit of sensitivity of the assay corresponded to about 0.3 E_{446nm}, which was about 0.001 to 0.002 4S Units/ml (uncorrected). Neat extracts of 4S-T92M, 4S-R95Q, 4S-H393P and 4S-L498P cells gave signals of around 10, 3, 1.5, and 0.8 E_{446nm}, respectively. The 4S-Y210C conditioned media samples gave signals around 7, while the non-cultured media background was 0.6. Typically, the other samples had to be diluted to ensure detection within the linear portion of the assay. For example, the 4S-N.3 and 4S-Y210C cell extracts had to be diluted about 12,000-fold and about 400-fold to obtain a signal around the middle of the standard curve, while the 4S-N.3 conditioned media sample had to be diluted about 100-fold. In general, the immunocapture assay resulted in less than 2% conversion of the 4MUS substrate to product.

The results of the parallel multiple antibody assays are presented in Tables 5.5 and 5.6. In the cell extract, the immunocaptured residual activities of all five 4S mutants relative to the wild type control were closely similar for the four monoclonal antibodies (Table 5.5). The relative residual activities for the extracts were also in reasonable agreement with the separate experimental results obtained from batch 1 in Table 5.2. In addition, the 4S-T92M dilution effect was observed with each of the four monoclonal antibodies used, and to a similar extent. Unfortunately, 2-fold serial dilutions of neat extract down to 1/16 was not sufficient to fully titrate the dilution effect. Inspection of the absolute values for residual activity indicate that the three antibodies ASB 4.1, ASB 56.2 and ASB F66

captured closely similar activities for the wild type and the mutants. ASB 58.3 however, captured reduced absolute activity relative to the other three for the wild type and the mutants (although % relative activity was about the same), and had a higher background (Table 5.5). In the culture media, no activity was detected for 4S-T92M, 4S-R95Q, 4S-H393P, 4S-L498P, and the CHO control, above the non-cultured media control background. The 4S-Y210C mutant secreted a low level of activity, about 0.5-1.0 units/ml, which was about 1-2% of the 4S-N.3 wild type control (Table 5.6).

The multiple antibody immunocapture experiments confirmed that each of the four monoclonal antibodies (except perhaps for ASB 58.3) bound the wild type and mutant 4S proteins with about equal affinity. In addition, the results suggested that none of the monoclonals had selectively stimulated or retarded the activity of any of the expressed 4S proteins to a significant extent. These results, and the results of the multiple monoclonal ELISA (Section 5.5.2b), strongly supported the proposal that the 4S-H393P and 4S-L498P mutants did not produce activity or protein. In addition, the observation of the 4S-T92M dilution effect of similar magnitude and rate with all four monoclonal antibodies supported the reality of the dilution phenomenon.

The immunocapture results obtained for the media samples showed that no significant 4S activity was secreted into the media for the mutant constructs 4S-T92M, 4S-R95Q, 4S-H393P, and 4S-L498P, and thus the majority of the detected residual activity was intracellularly located. Therefore, subsequent analysis focused on the mutant enzyme in cell extracts.

The polypeptide composition of the 4S wild type and mutants either in the medium or the cell extract was not determined during this work. Presumably, for the wild type 4S at least, the major secreted form was composed of the 66-kDa species

TABLE 5.5 IMMUNOCAPTURE ASSAY OF 4S ACTIVITY (CELL EXTRACT) USING A PANEL OF MONOCLONAL ANTIBODIES

A panel of four anti-4S monoclonal antibodies (ASB 4.1, ASB 56.2, ASB 58.3, and ASB F66) were individually used in an immunocapture assay to determine 4S-specific arylsulphatase activity in cell extracts. Extracts were obtained from 4S-N.3, and from a selected single clone of each of the transfected CHO cells which expressed mutant 4S, and the activity was assayed using the 4MUS substrate. The experimental details of the assay are described in Sections 5.2.4b and 5.5.3. For each of the extracts (neat and/or serial dilutions where appropriate), 50 µl was assayed at 37°C for 5 hr. The progress of the reaction was determined by fluorometric measurement of the liberated 4-methylumbelliferone. The E_{446nm} values were converted to Units/ml by reference to a standard curve of known 4S activities, and then corrected for cell protein in the extract. Activity = Units/mg = nmoles/min/mg cell protein; the activity values are shown in normal type. The % relative activities (italics) of the mutant 4S samples were calculated relative to 4S-N.3, which was assigned a value of 100% for each antibody. The CHO background was not subtracted from the residual activity values. The 4S-T92M extract exhibited the `dilution effect' with each of the antibodies, as discussed in Section 5.5.3. Cell extracts from the H393P and L498P mutants were considered to have no significant 4MUS activity above the CHO background.

Monoclonal	**ASB 4.1**	**ASB 56.2**	**ASB 58.3**	**ASB F66**
Clone:				
4S-N.3	770 *100%*	800 *100%*	260 *100%*	620 *100%*
4S-Y210C.5	26 *3.4%*	29 *3.6%*	8.6 *3.3%*	17 *2.7%*
4S-T92M.9	0.15->.6 *0.02->.08%*	0.18->.6 *0.02->.08%*	0.05->.25 *0.02->.10%*	0.10->.5 *0.02->.1%*
4S-R95Q.4	0.04 *0.005%*	0.06 *0.007%*	0.03 *0.01%*	0.04 *0.006%*
4S-H393P.2	0.02 *0.002%*	0.02 *0.002%*	0.02 *0.007%*	0.02 *0.003%*
4S-L498P.7	0.01 *0.001%*	0.01 *0.001%*	0.02 *0.007%*	0.01 *0.001%*
CHO background	0.004 *0.0005%*	0.004 *0.0005%*	0.01 *0.004%*	0.005 *0.0008%*

TABLE 5.6 IMMUNOCAPTURE ASSAY OF 4S ACTIVITY (SECRETED) USING A PANEL OF MONOCLONAL ANTIBODIES

A panel of four anti-4S monoclonal antibodies (ASB 4.1, ASB 56.2, ASB 58.3, and ASB F66) were individually used in an immunocapture assay to determine 4S-specific arylsulphatase activity secreted into the culture media of CHO cells which expressed wild type and mutant 4S. Culture media samples were obtained from 4S-N.3 and from a selected single clone of each of the transfected CHO cells which expressed mutant 4S, and the activity was assayed using the 4MUS substrate. The experimental details of the assay are described in Sections 5.2.4b and 5.5.3. See Table 5.5 for more information. Activity was defined as Units/ml culture medium. The activity values are shown in normal type, while the % relative activities are in italics. The CHO background was not subtracted from the residual activity values. Of the five 4S mutants, only the Y210C mutant was considered to have significant 4MUS activity above the untransfected CHO cell background. The media background was determined on a sample of media which was not used for culturing cells.

Monoclonal	ASB 4.1	ASB 56.2	ASB 58.3	ASB F66
Clone:				
4S-N.3	48 *100%*	52 *100%*	49 *100%*	45 *100%*
4S-Y210C.5	0.63 *1.3%*	1.0 *2%*	0.53 *1%*	0.65 *1.4%*
4S-T92M.9	0.08 *0.17%*	0.10 *0.19%*	0.30 *0.62%*	0.20 *0.43%*
4S-R95Q.4	0.10 *0.20%*	0.11 *0.21%*	0.10 *0.21%*	0.11 *0.25%*
4S-H393P.2	0.09 *0.18%*	0.09 *0.18*	0.27 *0.55%*	0.09 *0.19%*
4S-L498P.7	0.10 *0.19%*	0.13 *0.25%*	0.30 *0.61%*	0.10 *0.21%*
CHO background	0.10 *0.19%*	0.09 *0.17%*	0.24 *0.48%*	0.11 *0.24%*
media background	0.1 *0.20%*	0.13 *0.25%*	0.24 *0.59%*	0.1 *0.18%*

(Anson *et al.*, 1992b), while the intracellular form was the mature 57-kDa species composed of the disulphide-linked 43-, 8-, and 7-kDa species (Chapter 3.8.1). A formal comparison of the relative affinity of each antibody for the intracellular compared to the secreted species was not performed. Differential affinity affects cannot be ruled out, and therefore values for cell extract and cell media cannot be strictly compared. However, comparison of the total immunocaptured 4S activity in the culture media with that of the cell extract in untransfected CHO cell controls found that the media had about 260-fold more activity than the cell extract. The volume of the culture media was about 130-fold greater, and therefore the concentration of background activity in the CHO control extracts (0.04-0.115 Units/ml) was about a third to half that of the media only control (0.1-0.24 Units/ml). The background activity was presumed to be derived from the CHO and/or bovine homologue of human 4S, presumably due to trace levels of cross-reactivity between the non-human 4S species and the antibodies in the assay.

The previously observed dilution effect for the 4S-T92M cell extract using ASB F66 was also demonstrated using three other monoclonal antibodies separately in the immunocapture assay. However, complete titration was not achieved. In order to further characterize the residual activity of the 4S-T92M mutation, two further immunocapture assays were performed in order to titrate fully the observed dilution affect. A fresh batch (batch 3) of cellular extracts and conditioned media aliquots was obtained from cultured cells that were grown and harvested at confluency, about four days after the final media change. The maximum activity per mg cell protein for the T92M mutant was observed at a 64-fold dilution of cellular extract; about 2 Units/mg or about 0.5% relative activity, which was approximately 7-fold increased over the value of the neat extract (data not shown). During these experiments, a `dilution effect' was also observed for the R95Q mutant, which was not fully titrated. However, activity per mg cell protein of the

most dilute sample (1/64) was about 1 Unit/mg or about 0.2% relative activity, which was increased over 13-fold relative to the neat extract (data not shown).

5.5.4 PRELIMINARY OPTIMIZATION OF THE IMMUNOCAPTURE ASSAY

Obviously, such marked variation of the residual activity, dependent on the concentration of the mutant sample, made prediction of the biochemical and clinical severity of these mutant alleles difficult and ambiguous. Therefore, these results led to the re-examination of the experimental conditions under which the immunocapture assay was performed, in order to determine the optimal conditions for the analysis of the 4S mutations. The original conditions, although considered optimal for the wild type, may have been sub-optimal for the mutants. Optimization of the experimental assay system was considered necessary for the refinement of the predicted clinical and biochemical phenotype of each mutant allele. In order to refine the immunocapture activity assay, time-activity and pH-activity profiles were performed on the wild type 4S-N.3, the 4S mutants 4S-T92M.9, 4S-Y210C.5 and 4S-R95Q.4, and a standard 4S sample obtained from a CHO cell line expressing human 4S.

5.5.4a OPTIMIZATION OF THE TIME-ACTIVITY PROFILE

The immunocapture assay was performed using cell extracts and conditioned media samples (batch 3) as previously described, i.e. ASB F66, 5 mM 4MUS substrate, acetate buffer pH 5.6. Time points were taken after 20 min, 1 hr, 3 hr, 5 hr and 10 hr incubation at 37°C, and the 4S activities determined. The 4S-R95Q.4 extract was only assayed at 1 hr, 5 hr and 10 hr, due to limited quantities of extract. The conditioned media of the 4S-N.3 and 4S-Y210C.5 lines were assayed at a single dilution, while each of the cell extracts and the standard 4S sample were assayed at two dilutions that differed 4-fold. Two dilutions were assayed for a

number of reasons. Firstly, to ensure that the fluorescence signals were in the linear range of the assay. Secondly, to allow for the `dilution effect'. And thirdly, to enable the comparison of roughly equal amounts of bound activity from each of the constructs.

The results of the time course assay are presented in Figure 5.7. For the sake of clarity, only one of the two dilutions of each sample was plotted on the graph. However, the trends observed for the two dilutions, and hence the conclusions reached, were similar. The activities of the standard 4S sample, the 4S-N.3 extract and media, and the 4S-T92M.9 extract, were approximately linear over 10 hr. The activity of the 4S-R95Q.4 extract was linear over 5, but not 10 hr. For the 4S-Y210C.5 mutant, both the media and extract samples were not linear over 5 or even 3 hr.

These results suggested that the 5 hr incubation time for the original immunocapture assay was suitable for all except the 4S-Y210C.5 mutant. The original 5 hr assay would underestimate the residual activity by at least 20%, compared to an assay performed for a shorter time for which the time-activity profile was still linear. However, this was considered unlikely to critically affect the prediction of the biochemical and clinical phenotype of the mutation. Whether the instability of the Y210C mutant *in vitro* was an artefact, or corresponds with the *in vivo* situation is not known.

5.5.4b OPTIMIZATION OF THE pH-ACTIVITY PROFILE

The pH-activity profiles were also determined in order to refine the immunocapture assay. The immunocapture assay was performed using cell extracts and conditioned media samples (batch 3) as previously described, i.e. using ASB F66, 5 mM 4MUS substrate, and a 5 hr incubation. The reactions were performed in

acetate buffer at the following pHs; 3.9, 4.6, 5.3, 5.7, and 6.1. The samples assayed, and the dilutions used, were the same as the time-activity assay.

The results of the pH-activity profiles are presented in Figure 5.8. Again, for the sake of clarity, only one of the two dilutions of each sample was plotted on the graph, although the trends observed and the conclusions reached were similar for both dilutions. The pH-activity profiles of the standard 4S sample, the 4S-N.3 extract, the 4S-N.3 media, the 4S-Y210C.5 extract, and the 4S-Y210C.5 media, were similar to each other, and exhibited a pH optimum around 5.3. The profiles were also in agreement with that of human liver 4S assayed free in solution, which exhibited a pH optimum around 5.5 (Gibson *et al.*, 1987). These results suggested that ASB F66 was not significantly altering the behaviour of 4S towards 4MUS *in vitro*. The 4S-T92M.9 extract however, exhibited a pH optimum which was shifted to a pH around 4.6. In fact, the residual activity of the 4S-T92M extract was about 2 to 3 -fold greater at pH 4.6 than at pH 5.3 and 5.7 respectively. Taking into account the 20 to 30% reduction in activity of the 4S-N.3 extract over the same pH range, the residual activity of the 4S-T92M.9 mutant was about 3-fold greater at pH 4.6 than at pH 5.3-5.7, relative to the wild type 4S-N.3. The 4S-R95Q.4 extract also showed a shifted pH-activity profile that had a optimum greater than pH 6.1.

These results suggested that the pH 5.6 acetate buffer in the standard immunocapture assay was within the known pH range of peak 4S activity, and therefore was suitable for all samples except the 4S-T92M.9 and 4S-R95Q.4 mutants. The 4S-T92M.9 mutant displayed an approximately 3-fold increase in residual activity relative to the wild type 4S, from about pH 5.3-5.7 to pH 4.6. The linearity of the 4S-T92M.9 mutant activity over time at pH 4.6 was not determined. The 4S-R95Q.4 mutant apparently displayed a pH optimum in excess of 6.1, however in order to determine the precise pH optimum, a different assay buffer would have been necessary, as the buffering capacity of acetate is poor above pH

FIGURE 5.7 ACTIVITY vs. TIME ASSAY OF WILD TYPE & MUTANT 4S

The immunocapture assay (ASB F66) was performed using cell extracts and conditioned media samples from CHO cells which expressed wild type and mutant 4S. The substrate was 4MUS, which was used at the standard assay concentration of 5 mM and pH 5.6. Time points were taken after 20 min, 1 hr, 3 hr, 5 hr, and 10 hr incubation at 37°C, and the E_{446nm} values measured. In the legend of the graph, cell extracts are denoted E, while conditioned media samples are denoted M. 4S.STD was a standard 4S sample obtained from a CHO cell line expressing human 4S. 4S.N was 4S-N.3, which expressed 4S wild type; the media sample was diluted 50-fold, and the cell extract sample was diluted 6400-fold, prior to assay. YC was the 4S-Y210C.5 mutant; the media sample was assayed neat, while the cell extract sample was diluted 320-fold. TM was the 4S-T92M.9 mutant and RQ is the 4S-R95Q.4 mutant; the cell extract samples were diluted 16-fold and 8-fold respectively. All samples except the cell extract and media samples for the 4S-Y210C mutant produced linear time-activity profiles over 5 hr, which was the usual assay time of the standard immunocapture assay. Substrate only blanks produced a signal of E_{446nm} <0.1 at all five time points assayed.

Time - Activity (4MUS) Assay

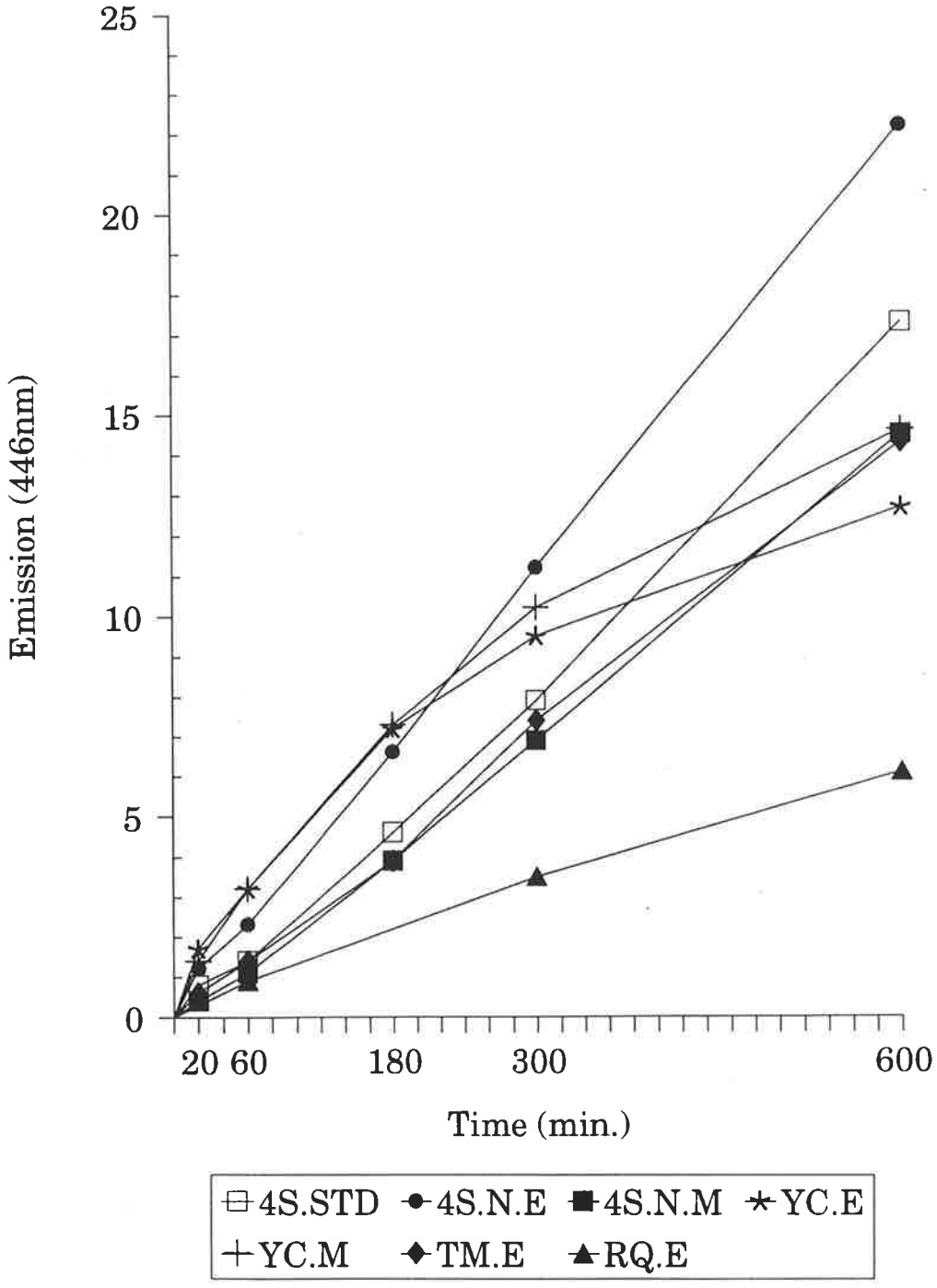
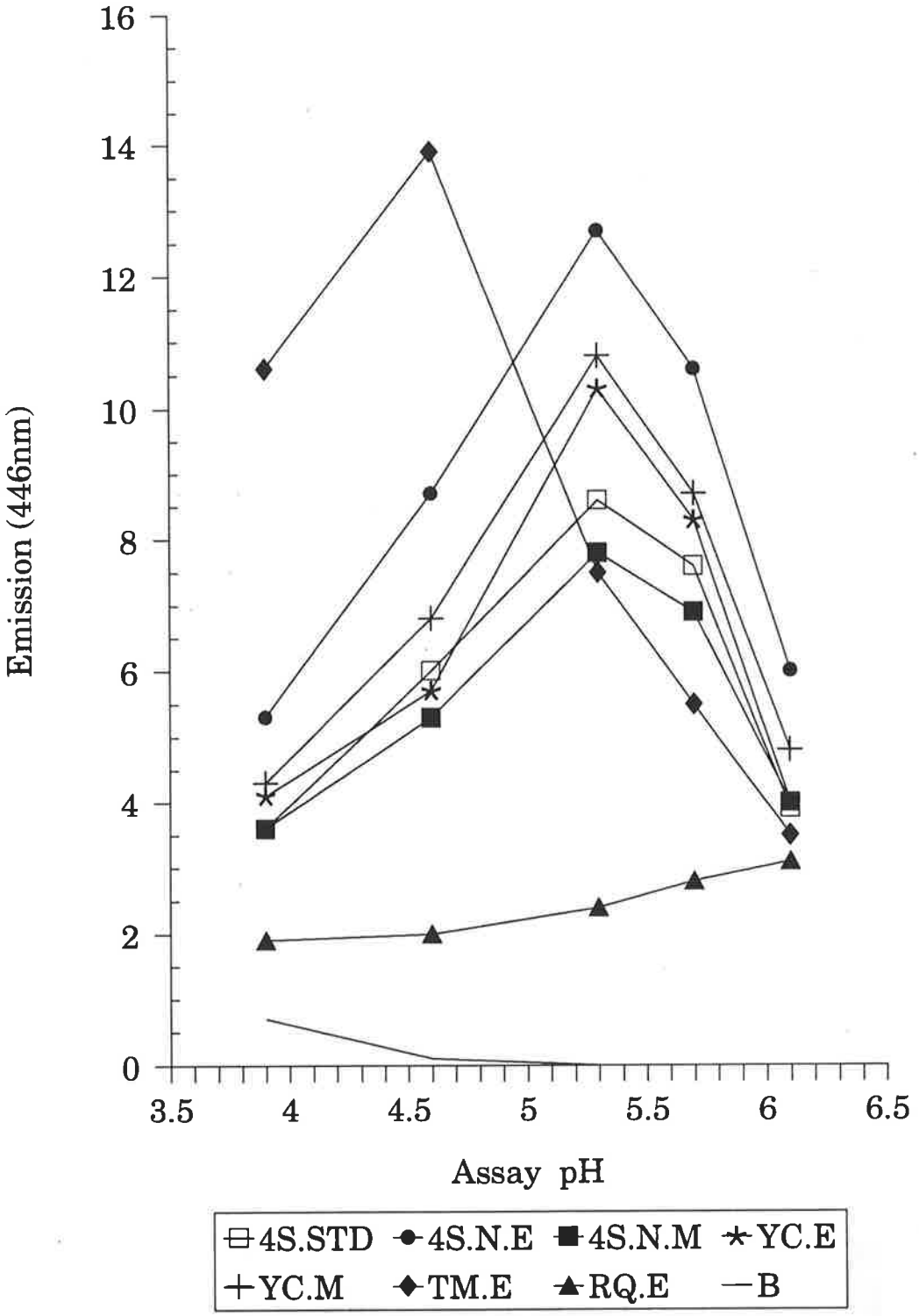


FIGURE 5.8 ACTIVITY vs. pH ASSAY OF WILD TYPE & MUTANT 4S

The immunocapture assay (ASB F66) was performed using cell extracts and conditioned media samples from CHO cells which expressed wild type and mutant 4S. The substrate was 4MUS, which was used at the standard concentration of 5 mM, and the assay was performed for 5 hr. The reactions were performed in acetate buffer at the following pHs; 3.9, 4.6, 5.3, 5.7, and 6.1. The samples assayed, and the dilutions used, were the same as the time-activity assay in Figure 5.7. The pH-activity profiles of the standard 4S sample, the 4S-N.3 extract, the 4S-N.3 media, the 4S-Y210C.5 extract, and the 4S-Y210C.5 media, were similar to each other, and exhibited a pH optimum around 5.3. The 4S-T92M.9 extract however, exhibited a pH optimum which was shifted to a pH around 4.6. The 4S-R95Q.4 extract also showed a shifted pH-activity profile that had a optimum greater than pH 6.1. Substrate only blanks (B) showed that at low pH, the rate of spontaneous hydrolysis of 4MUS was increased.

pH - Activity (4MUS) Assay



6.1. The pH of the lysosomal system is not known with precision, but is thought to be between 4.6 and 5.0 (reviewed in Mellman *et al.*, 1986). Consequently, the possibility exists that the relative activity *in vivo* of the 4S-T92M.9 mutant was underestimated, and the 4S-R95Q.4 mutant overestimated, based on the *in vitro* activity at a standard pH 5.6. Whether the shift in pH optimum of the 4S-T92M.9 and 4S-R95Q.4 mutants was an *in vitro* artefact, or corresponds with the *in vivo* situation is not known.

Preliminary optimization of the activity assay has enabled refinements in the predicted clinical and biochemical phenotype of the mutations to be made. These predictions will be presented below, and then compared with the actual clinical and biochemical phenotype of the patient.

5.6 PREDICTION OF THE CLINICAL & BIOCHEMICAL PHENOTYPE OF 4S MUTATIONS

5.6.1 INTRODUCTION

The work presented in this Chapter attempts to test the validity of the main hypothesis, i.e. given the 4S genotype, do the *predicted* clinical and biochemical phenotypes correlate with the *actual* phenotypes in MPS-VI? In summary, the measurements of the residual 4S activity and protein obtained from the wild type and mutant 4S expressed in CHO cells, although preliminary, were used to predict the biochemical and clinical phenotype of the mutations. These predicted individual biochemical phenotypes were then compared with the actual residual 4S activity and 4S protein measured in cultured MPS-VI fibroblasts, and the clinical phenotype of the MPS-VI patient.

5.6.2 *PREDICTED* BIOCHEMICAL PHENOTYPE

The biochemical severity of the 4S mutant alleles was predicted from the immunocaptured activity of the CHO-expressed 4S enzyme (Tables 5.2 and 5.5, and the results of further titrations of T92M and R95Q activities). The H393P and L498P mutations were predicted to be severe, as no residual 4S activity or protein was detected for either *in vitro*. The R95Q mutation was also predicted to be a severe allele, as the residual activity was detectable although very low; about 0.003% to >0.2% of expressed control 4S depending on the assay and the extent of dilution. The predicted severity of the Y210C mutation was milder than the previous mutations, as it exhibited a greater residual activity, about 3% of expressed control 4S. The T92M mutation was estimated to have a relative severity between that of the R95Q and Y210C mutations, about 0.01 to 0.5% of control activity, although the prediction was also made ambiguous by the `dilution effect'. For the sake of simplicity, the predicted relative activities of the R95Q and T92M mutant alleles were arbitrarily assigned the highest values of the observed ranges. In an autosomal recessive disorder such as MPS-VI, two mutant alleles are required for clinical pathology to occur. Therefore, the predicted relative activity *in vivo* was calculated as the mean of the predicted relative activities of both mutant alleles.

The relative residual 4S protein of the 4S mutants expressed in CHO cells (Table 5.3) was another potentially useful biochemical measurement. Although determination of the relative 4S protein was not likely to be informative of the biochemical or clinical phenotype per se, comparison of the predicted values derived from analysis of transfected CHO cells with the actual values obtained from patient samples may provide an insight into the validity of the expression system. The H393P and L498P mutations were predicted to result in no residual protein in patient samples, as neither was detected in transfected CHO cells. The T92M mutation was estimated to produce a relative residual protein level of about 0.3% of

normal. The R95Q and Y210C mutations were predicted to have a relatively less marked reduction of about 1.5% and 5% protein respectively. Again, the predicted relative 4S protein *in vivo* was calculated as the mean of the predicted relative 4S protein of both mutant alleles.

5.6.3 COMPARISON OF THE *PREDICTED* AND THE ACTUAL BIOCHEMICAL PHENOTYPE

For those MPS-VI patients with at least one mutant 4S allele defined, the predicted relative 4S activity and protein values were compared retrospectively with the actual values obtained from cultured skin fibroblasts a number of years earlier. The relative residual 4S activities of extracts from patient cultured fibroblasts were provided by Doug Brooks (unpublished data). The activity was immunocaptured using ASB 4.1 and assayed with 4MUS, using an earlier protocol (Brooks *et al.*, 1991a) than the one described in Section 5.2.4b. The relative residual 4S protein in patient cultured fibroblasts had previously been determined using an ELISA assay based on ASB 4.1 (Doug Brooks, unpublished data).

The predicted and actual relative activities for each patient were compared (Table 5.7). The actual relative activities of three patient fibroblasts heterozygous for the Y210C allele ranged from 1.1% to 1.7% of normal controls (SF912, SF913, SF2724). In comparison, the predicted value for a patient heterozygous for the Y210C mutation was greater than or equal to 1.5%, depending on the severity of the other allele. Therefore, the predicted and actual activity of the Y210C allele were in reasonable accord. The combination of the H393P and R95Q mutations was predicted to have a relative activity of about 0.1%, based on the low activities of these mutants expressed in CHO (~0% and ~0.2% respectively). However, this prediction was not confirmed when compared to the actual residual activity of patient SF1246 (2.2%), who was heterozygous for both alleles. The biochemical

correlation could not be checked for the T92M and L498P alleles, as they were unique to patient SF2467, and the residual activity had not been determined.

The predicted and actual relative 4S protein for each patient were also compared (Table 5.7). A patient heterozygous for the Y210C allele was predicted to have a relative 4S protein level greater than or equal to 2.5%, depending on the severity of the other allele. This prediction was in reasonable agreement with the four patients heterozygous for the Y210C mutation for which the actual protein values were known (SF912, SF913, SF2724, SF2984). The predicted combined low residual protein of the R95Q and H393P mutants was in reasonable agreement with the actual value in the compound heterozygote SF1246. However, the actual residual protein found in the patient SF2467 fibroblasts (3.6%) did not correlate with the predicted residual protein of a T92M/L498P compound heterozygote (0.15%).

5.6.4 DISCUSSION: *PREDICTED* vs. *ACTUAL* BIOCHEMICAL PHENOTYPE

The attempt to correlate the predicted residual 4S activity and protein of the mutant alleles with the actual values from patient fibroblasts was only partially successful. At present, it is not known whether the anomalies are of biological significance or artefactual. However, the later is more likely on account of a number of limitations in the data obtained from the patient cultured fibroblasts.

As mentioned previously, the immunocapture and ELISA assays of the patient cultured fibroblasts had been performed before the contribution of 4S derived from foetal calf serum was appreciated. Although the bovine 4S was present at a low level, the very low residual 4S activity of the patient fibroblasts may have been variably supplemented by uptake of the exogenous 4S, perhaps by receptor-mediated endocytosis. The anti-4S monoclonal antibodies may have exhibited a

TABLE 5.7 COMPARISON OF THE *PREDICTED & ACTUAL* RELATIVE 4S ACTIVITY & PROTEIN

The relative residual activity and protein of each mutant 4S expressed in transfected CHO cells was measured, and compared with the actual values obtained from patient cultured fibroblasts with the corresponding mutant genotype. The activity of each mutant allele relative to the wild type control was determined by ASB F66-immunocapture and assay with 4MUS substrate. The relative activities of the mutant alleles were; H393P (0%), L498P (0%), R95Q (0.2%), T92M (0.5%), and Y210C (3%) (Section 5.6.2). The predicted relative activity of the patient was calculated as the mean of the predicted relative activity of both mutant alleles. The actual 4S activity had been previously determined in an immunocapture assay using ASB 4.1, and was expressed relative to normal cultured fibroblast samples (Doug Brooks, unpublished data). The relative residual protein of each mutant 4S allele was determined in an ELISA using ASB F66. The relative 4S protein of the 4S mutant alleles were H393P (0%), L498P (0%), R95Q (1.5%), T92M (0.3%), and Y210C (5%) (Section 5.5.2b). The predicted relative 4S protein of the patient was calculated in the same manner as for the predicted relative activity. The actual relative 4S protein had been previously determined in an ELISA assay using ASB 4.1 (Doug Brooks, unpublished data). Note: (ND) not determined, (>) greater than or equal to.

Patient	Age at diagnosis (years)	Clinical phenotype	Mutant genotype	Predicted relative 4S activity	Actual relative 4S activity	Predicted relative 4S protein	Actual relative 4S protein
SF1246	1.1	severe	H393P, R95Q	0.1%	2.2%	0.75%	1.6%
SF693	3.75	severe	H393P, ?	> 0%	ND	> 0%	ND
SF2984	6.7	intermediate	H393P, Y210C	1.5%	ND	2.5%	3%
SF912	13	mild	R95Q, Y210C	1.6%	1.1%	3.25%	3.3%
SF913	16	mild	R95Q, Y210C	1.6%	1.7%	3.25%	1.6%
51-1	15	mild	Y210C, ?	> 1.5%	ND	> 2.5%	ND
SF839	17	mild	H393P, ?	> 0%	3%	> 0%	0.5%
SF2724	17.5	mild	Y210C, ?	> 1.5%	1.7%	> 2.5%	1.4%
SF2467	35	mild	L498P, T92M	0.25%	ND	0.15%	3.6%
Normal				100%	100%	100%	100%

trace of cross-reactivity towards the exogenous 4S, which would have led to an overestimate of the residual 4S activity and protein values in the patient fibroblasts. In addition, the low levels of mutant 4S activity and protein in the patient fibroblasts meant that for some samples the assays were at the limit of sensitivity. Other differences in the experimental procedures over time may have also contributed to the divergent results.

Another limitation was that there were a number of patients for which the mutant genotype had been identified but had not had the residual 4S activity and/or protein determined, and thus were not able to be utilized in the comparison of predicted versus actual biochemical phenotype.

Therefore, the limitations of the biochemical data derived from patient fibroblasts render the discussion of potential mechanisms which account for the discrepancy between the predicted and the actual biochemical phenotype, as premature. Future experiments would need to determine in parallel the residual 4S activity and protein in both the patient cultured fibroblasts and the transfected CHO lines expressing mutant 4S, under optimal culture and assay conditions. This would be a necessary first step in the clarification of the relationship between the predicted biochemical severity and the actual biochemical severity.

5.6.5 COMPARISON OF THE *PREDICTED* BIOCHEMICAL PHENOTYPE & THE *ACTUAL* CLINICAL PHENOTYPE

The predicted biochemical phenotypes of the expressed 4S mutants were also compared with the clinical phenotypes of the MPS-VI patients in which at least one of the mutant 4S alleles had been defined. The only clinical phenotype parameter that was available for all the patients was age at diagnosis, while clinical

descriptions of varying detail were available for eight of the nine patients. The limitations of the clinical phenotype data will be discussed in Section 5.7.3.

The predicted clinical phenotype of each mutation was estimated based on the crude relationship between the patient's clinical severity and the residual 4S activity in cultured fibroblasts, namely that the spectrum of clinical severity from severe to mild extremes was associated with a residual activity range of 0 to about 5% of wild type controls. The CHO-expressed Y210C mutation had the highest residual activity of the five mutant 4S alleles (average about 3%), and was therefore predicted to result in the mildest clinical phenotype. In compound heterozygotes, the milder Y210C allele was proposed to ameliorate the severity of other more severe alleles. This prediction was supported by the age at diagnosis of the five patients who were heterozygous for the Y210C mutation; SF 2984, SF912 & 913, SF2724 and 51-1, which ranged from 6.7-17.5 years (Table 5.8). The milder clinical phenotype was illustrated by patient 51-1, who was diagnosed at the age of 15 years. At diagnosis, the patient had a normal facial appearance, hepatomegaly, very slight corneal clouding, and very slight restriction of movement in the large joints. Currently aged 21, the patient is fit and well, and clinical signs have shown no progression. The other mutant 4S allele in patient 51-1 was not identified. Considerable variation in severity was seen in the Y210C heterozygotes, for example the intermediate patient SF2984. The clinical descriptions of patient 51-1, SF2984, and the other Y210C heterozygotes were presented in detail in Chapter 4, Section 4.6.3.

The R95Q mutation was predicted to be more severe than the Y210C mutation due to a lower residual activity (0.003% to >0.2%), and result in an earlier age of diagnosis. However, a more precise prediction of the severity of the R95Q allele was hampered by the dilution effect observed *in vitro*. The H393P mutation was predicted to be even more severe, as no residual activity was detected when the

TABLE 5.8 COMPARISON OF THE *PREDICTED & ACTUAL* CLINICAL PHENOTYPE IN MPS-VI PATIENTS

The clinical phenotype of MPS-VI patients was predicted based on the residual activity of the corresponding 4S mutants expressed in CHO cells. The predicted severity was compared with the actual clinical severity, which was based on the clinical descriptions presented in Chapter 4. Age of diagnosis is shown, as a measurable parameter of actual disease severity. For patients SF693 and SF839, the phenotype could not be predicted as only one allele (H393P) had been defined, although the H393P allele was predicted to be severe.

Patient	Genotype	Predicted clinical severity	Actual clinical severity	Age at diagnosis
SF1022	ΔG_{238}, ΔG_{238}	severe	severe	7 days
SF1246	H393P, R95Q	severe	severe	13 months
SF693	H393P, ?		severe	3 years
SF2984	H393P, Y210C	mild	intermediate	6.7 years
SF912	R95Q , Y210C	mild	mild	13 years
SF913	R95Q , Y210C	mild	mild	16 years
51-1	Y210C, ?	mild	mild	15 years
SF839	H393P, ?		mild	17 years
SF2724	Y210C, ?	mild	mild	17.5 years
SF2467	L498P, T92M	intermediate	mild	35 years

mutation was expressed in CHO cells. On the speculation that the actual residual activity of the R95Q allele in a patient was closer to 0.2% than 0.003%, patients who were compound heterozygotes of the Y210C and R95Q alleles were predicted to have a slightly higher residual activity than patients who were compound heterozygotes of the Y210C and H393P alleles. Although this may not appear of great significance, the model of Conzelmann & Sandhoff (1983/84) based on the consideration of the lysosomal storage phenomenon predicted that small increases in residual activity in the vicinity of a critical threshold activity may lead to large decreases in the rate substrate accumulation, and hence presumably disease severity. Although the predicted contribution of the R95Q allele to the total residual activity was small, it may affect the rate of lysosomal storage and hence clinical severity, especially if the total residual activity was near the critical threshold. Therefore it was speculated that Y210C/R95Q compound heterozygotes may be less clinically severe and/or diagnosed later than Y210C/H393P compound heterozygotes.

This prediction was borne out by the observation that patient SF912 (Y210C and R95Q alleles) was diagnosed at the ages of 13 years, compared to patient SF2984 (Y210C and H393P alleles), who was diagnosed at 6 years 8 months of age. Patient SF2984 was the earliest diagnosed of the Y210C heterozygotes; the clinical features and the age at diagnosis would suggest an intermediate phenotype. At diagnosis the patient displayed restriction in the movement of shoulders and hips, corneal clouding, dysostosis multiplex and other skeletal signs typical of MPS-VI. The growth curve was within the normal range for the first 7 years of life, thereafter it was retarded. Dysmorphology was noted at around 8 years, progressive deterioration in skeletal features was observed over the next few years, and at the age of 17, the patient was 143.5 cm ($<5^{th}$ centile) in height , exhibited joint stiffness and aortic insufficiency. Patient SF2984 had a younger brother who was also

affected with MPS-VI, and was considered by the referring doctor "as following the same clinical course".

The predicted severity of the R95Q and H393P alleles was independently confirmed in patient SF1246 who was heterozygous for both alleles. The patient was considered clinically severe, and was diagnosed at the early age of 13 months. The predicted severity of the H393P allele was also in agreement with the clinical phenotype of patient SF693, who was heterozygous for the H393P allele and an unidentified allele. Patient SF693 was diagnosed at the age of 3.75 years and currently is terminally ill at the age of 19 years. The clinical descriptions of patients SF1246 and SF693 were presented in detail in Chapter 4, Section 4.6.2. Therefore, as the phenotype of patient SF693 is severe, both mutant alleles are considered severe, one of which is the H393P allele. Patient SF839 was also heterozygous for the H393P allele, however the patient was mildly affected and was diagnosed at the age of 17 years. Presumably, the other (unidentified) allele was mild, and ameliorated the severity of the H393P allele.

The predicted severities of the T92M and L498P alleles were difficult to reconcile with the actual clinical severity of patient SF2467. The L498P allele was predicted to be a severe mutation comparable to the H393P mutation, as no residual 4S activity or protein was detected in the CHO expression system. Prediction of the severity of the T92M allele was complicated by the dilution effect observed *in vitro*, but was likely to be between the milder Y210C and more severe R95Q alleles, based on the predicted residual activity (about 0.5%). If the altered pH optimum of the T92M mutant observed *in vitro* (pH 4.6) was relevant *in vivo*, and if the pH in the lysosome was around this value as suggested previously, then the predicted residual activity of the T92M mutant could be up to 1% of normal. Although this is a very small increase in activity in absolute terms, this may result in a proportionally greater decrease in clinical severity, based on the model of

Conzelmann & Sandhoff (1983/84) discussed previously. Therefore, the T92M allele was predicted to have an intermediate clinical severity. Cognizant of the limitations of the CHO expression data, an intermediate clinical severity was proposed for the patient, based on the predicted intermediate and severe phenotypes of the T92M and L498P alleles respectively. However, the patient was diagnosed at the late age of 35 years, which suggested a mild phenotype. Unfortunately, the actual clinical severity was not known, as no clinical details apart from the age of diagnosis were provided by the cell repository from which the patient cell line was obtained. In addition, the T92M and L498P alleles were unique to SF2467 in the patient collection screened, and therefore the discrepancy between the predicted and actual clinical severities of these alleles could not be resolved by comparison with the phenotypes of other patients.

A study of the residual 4S protein in a collection of cultured fibroblasts derived from MPS-VI patients consistently identified levels of residual 4S protein in SF2467 higher relative to other MPS-VI patients, using a panel of 7 anti-4S monoclonal antibodies (Brooks *et al.*, 1991b). In addition, residual activity towards the radiolabelled trisaccharide substrate was also observed. If the mild biochemical phenotype was indeed a valid observation, and not an artefact due to the exogenous bovine 4S, this would suggest that the T92M and/or L498P mutations are milder *in vivo* than the CHO expression data had predicted. If this interpretation is valid, then significant limitations exist in at least one of the components of the experimental system utilized for the prediction of the mutant allele phenotype.

The existence or otherwise of a genotype-phenotype correlation in MPS-VI, will be discussed below, followed by a discussion concerning the potential limitations of the CHO expression system, the ELISA and immunocapture assays, the 4MUS substrate, and the patient clinical data.

5.7 GENOTYPE & PHENOTYPE IN MPS-VI: A CORRELATION?

In review, preliminary biochemical analysis of human 4S mutants expressed in CHO cells enabled the prediction of clinical and biochemical phenotypes, which were subsequently compared with the actual patient phenotypes. The predicted clinical severity of the 4S mutant alleles; Y210C, R95Q, and H393P, showed a fair correspondence with the actual clinical phenotype of the patient, in particular the age of diagnosis. The correspondence was strengthened for each of these mutant alleles as each was observed in more than one patient. However, the predicted clinical severity of the T92M and L498P alleles was apparently discordant with the actual clinical severity of the single patient, to whom these alleles were unique.

The validity or otherwise of the genotype-phenotype correlation hypothesis in MPS-VI is crucially dependent on the data marshalled to test it. However, limitations existed in all the data used to test the hypothesis; the genotype, the biochemical phenotype, and the clinical phenotype. Each of the experimental methods used to obtain the data is appraised below, and refinements are suggested.

5.7.1 LIMITATIONS OF THE PATIENT GENOTYPE DATA

The techniques used to determine the mutant 4S genotypes in MPS-VI patients were described in Chapter 4. These techniques enabled the successful identification of both mutant alleles in the majority of patients, especially when the modified chemical cleavage technique was used. These techniques were designed to detect mutations in the coding region of 4S, using the 4S cDNA as the substrate for analysis. However, a limitation of this approach was that 4S mutations outside of the coding region would not be detectable. A case in point was the Δexon IV

mutation. Although the causal mutation was not identified, the mutation was likely to be located outside the coding region, presumably in the vicinity of an intron-exon splice junction. The recent determination of the genomic structure of the 4S gene (Modaressi *et al.*, 1993), coupled with minor modification of the mutation detection scheme described in Chapter 4, will enable mutations outside the exon regions to be determined.

Three different polymorphisms were identified in the 4S cDNA of a number of patients in addition to the MPS-VI mutations. The polymorphisms were individually considered to have no deleterious affect on normal 4S activity, as they are frequent in the normal population (V358M: Jin *et al.*, 1991), and as the V376M allele was shown to exhibit comparable levels of enzyme activity *in vitro* as the wild type (Wicker *et al.*, 1991). However, as discussed in Chapter 4, Section 4.7, it is not known whether the polymorphisms ameliorate or accentuate the severity of a 4S mutation *in vivo*. This question was not addressed experimentally, as the mutation detection techniques described in Chapter 4 did not allow the determination of the individual linkage between each mutation and the polymorphisms.

5.7.2 LIMITATIONS OF THE BIOCHEMICAL PHENOTYPE DATA

5.7.2a RESIDUAL 4S ACTIVITY ASSAY OF FIBROBLAST EXTRACTS

In this thesis, the residual 4S activity in MPS-VI cultured fibroblasts was the main element of the actual biochemical phenotype used in the test of the genotype-phenotype correlation in MPS-VI. Studies had shown that the residual activity in patients from the entire clinical spectrum was clustered over a low and narrow range (0-5%) (Brooks *et al.*, 1991b; Doug Brooks, unpublished results). Although the immunocapture assay was sensitive, the residual 4S activity in most patients

was clustered around very low values and some values were near the sensitivity limit of the assay. Assay-to-assay variation was observed with either the 4MUS or radiolabelled trisaccharide substrates. The fibroblast culture conditions, and/or the level of exogenous 4S activity derived from foetal calf serum (as discussed in Section 5.6.4), could be responsible at least in part for the variation. All these factors hindered the accurate and reproducible measurement of the residual activity. Consequently, the resolution of the gradations of clinical phenotype based on analysis of the residual enzyme data alone was also difficult.

Another limitation of the residual 4S activity assay was the use of a whole cell extract. Consequently, the assay was not able to distinguish between activity correctly targeted to the lysosome and activity incorrectly targeted to alternative cellular compartments. Immunochemical methods combined with cell organelle fractionation procedures can be used to characterize the biosynthesis, processing, and intracellular trafficking of the 4S protein, and therefore contribute to the refinement of the biochemical phenotype.

The immunocapture assay of 4S activity was performed using the synthetic substrate 4MUS, which had little structural resemblance to the GAG substrates, and therefore may not be a sensitive probe of the mutation-induced alterations in 4S activity *in vivo*. Until sophisticated techniques for measuring the activity of 4S towards GAG substrates in the lysosome are developed, the *in vitro* assays are the only way of directly measuring residual activity. The broadly similar residual activity results obtained when performing the immunocapture assay using either the 4MUS substrate or the trisaccharide substrate derived from chondroitin-4-sulphate (Doug Brooks, unpublished results), suggest that for the majority of patients, the *in vitro* assay is a reasonable approximation of the *in vivo* situation.

In addition to the deficiency of the 4S enzyme, MPS-VI is characterized biochemically by the intracellular storage and urinary excretion of elevated levels of GAG substrates (Chapter 1.8.1). In general, the degree of dermatansulphaturia correlates with the severity of the disease (Brooks *et al.*, 1991b). Accurate quantification of the intracellularly stored GAG substrate may provide a valuable indicator of the biochemical phenotype, as the stored material is directly responsible for the development of disease pathology.

5.7.2b HETEROLOGOUS EXPRESSION OF 4S MUTATIONS IN CHO CELLS

The residual activity of the CHO-expressed 4S mutants was also determined using the immunocapture assay, and therefore shared similar potential limitations as described in the assay of patient fibroblasts. However, as wild type human 4S activity and protein were expressed in CHO cells about 1000-fold greater than in normal human fibroblasts, measurement was not hindered by the sensitivity limits of the assays. Even for CHO-expressed mutants which resulted in low residual protein (eg. 4S-T92M, about 150 ng/mg cell protein, Table 5.3), this was comparable with the wild type 4S protein obtained from normal cultured fibroblasts using a number of different monoclonal antibodies (32-206 ng/mg cell protein, Brooks *et al.*, 1991b). Thus, the expression of the 4S mutants in the CHO system facilitated analysis of the mutant proteins. The immunochemical analysis of 4S activity, although sensitive, did show assay-to-assay variation in the absolute activity values obtained for each of the cell lines. Variation in the culture conditions of the different batches of cells may be responsible, at least in part, for the observed inter-assay variation in 4S activity.

It may be speculated that the high-level expression of human 4S in a heterologous host would yield aberrant results, at variance with the actual behaviour of 4S *in vivo* in humans, perhaps due to saturation of the biosynthetic and processing

pathways. However, the specific activity and pH optimum of recombinant wild type human 4S expressed in CHO cells were comparable to 4S from human liver or fibroblast assayed free in solution, as discussed previously. This suggested that the recombinant enzyme was normally synthesized and processed, even though expressed at a high level. In addition, the reduction in 4S mutant protein observed in patient fibroblasts was paralleled by reductions of the same relative magnitude in the 4S mutants expressed in CHO cells. These observations would suggest that the CHO expression system was suitable for the high-level production of 4S mutant proteins, and a valid model for investigation of the molecular pathology in MPS-VI cells *in vivo*. A potential mechanism accounting for the reduced mutant 4S protein is discussed in Section 5.8.1.

In the work described in this thesis, stable expression in CHO cells was chosen rather than transient expression. Stable expression allowed the selection of clonal lines, which would enable the production of large quantities of mutant 4S for subsequent analysis. However, the clonal variation between transfected CHO clones was quite marked, and in retrospect, the relative residual activity and protein may have been more accurately determined using a massed culture system for each mutant, or transient expression.

5.7.3 LIMITATIONS OF THE PATIENT CLINICAL PHENOTYPE DATA

As discussed in Chapter 1, Section 1.10.2, a wide range in clinical heterogeneity is observed in MPS-VI. The limitations of current clinical phenotype classifications are illustrated by the clinical data used in this thesis. Firstly, the data was incomplete in that clinical descriptions were not obtained for all the patients. Secondly, for those patients where clinical data was available, the information provided differed in the detail of the signs and symptoms reported. This suggested that the thoroughness of the clinical examination varied from clinician to clinician.

Therefore, a standardized clinical assessment protocol is necessary if the patient phenotype is to be well defined, an important element in the refinement of the genotype-phenotype correlation in MPS-VI. Such a standardized system would conceivably take into account the presence or absence of the clinical features of MPS-VI discussed in Chapter 1.7.2. Ideally, the protocol would attempt to quantify a number of the signs and symptoms in MPS-VI. For example, the degree of hepatomegaly and splenomegaly could be quantified by computed tomography (CT) scan, as done for Gaucher disease (Sibille *et al.*, 1993). CT and other radiographic techniques could also be used to quantify or at least rank the severity of a number of the skeletal features in the MPS-VI patient. Quantitative measurement of corneal clouding, the degree of deafness, joint mobility, and joint capsule thickness, may also be useful. Echocardiography and other techniques could be used to quantify heart function, in particular the degree of valve pathology.

The protocol would also need to take into account the previous medical interventions performed on the patient, as treatment can modify subsequent disease progression and hence phenotype. Finally, as MPS-VI is a progressive degenerative disease, the clinical evaluation protocol should ideally monitor the clinical features regularly over a number of years.

5.7.4 SUMMARY

In summary, comparison of the clinical and biochemical phenotype of the patient with the preliminary analysis of the predicted clinical and observed biochemical phenotype of the mutant 4S alleles expressed in cell culture support a reasonable genotype-phenotype correlation in MPS-VI, despite limitations in each of the data sources used to test the hypothesis. Refinements in the determination of the clinical phenotype, biochemical phenotype and the genotype, and a larger patient base, will enable a more rigourous test of the hypothesis. The utility of a successful genotype-phenotype correlation in MPS-VI will be discussed in Chapter 6.

5.8 INSIGHTS INTO THE MOLECULAR MECHANISMS OF 4S DEFICIENCY

Biochemical characterization of mutant 4S can provide insights into the biosynthesis, processing, structure, and function of normal 4S. However, analysis of 4S derived from patient cell lines was complicated as the majority of MPS-VI patients in this study were compound heterozygotes. In these patients, the observed 4S biochemical phenotype is based on the contribution of two different mutant 4S species, which can be difficult to distinguish biochemically. In contrast, the individual expression of each mutant 4S allele in a heterologous system can facilitate the resolution of the effect of each mutation.

Conceptually, a 4S mutation which results in MPS-VI can effect any of the processes from transcription of the 4S gene through to lysosomal residence and catalytic function of the 4S protein. The deleterious effects of the CHO-expressed 4S point mutations described in this Chapter were considered to be manifested post-translationally, as the processes of transcription and translation were assumed to be near normal for the 4S mutants, although this had not been rigourously demonstrated (see Chapter 4, Section 4.5.1). The mutations under study all exhibited reductions in 4S-specific arylsulphatase activity and 4S protein. Preliminary characterization of the 4S mutants has enabled generation of hypotheses discussed below concerning the sites of action of the individual mutations.

5.8.1 PROTEIN REDUCTION

As shown in Table 5.3 and discussed in Section 5.5.2b, the yield of CHO-expressed mutant 4S protein was markedly reduced relative to wild type human 4S for all five 4S point mutations. Comparable relative reductions in 4S protein were observed in patient cultured fibroblasts (Table 5.7, Brooks *et al.*, 1991b). The

reduced mutant protein concentration was thought to be most likely due to an increased rate of intracellular degradation in either the endoplasmic reticulum and/or the lysosome. A number of studies have proposed that mutant polypeptides are retained and degraded in a compartment in or near the ER (Lodish, 1988; Lippincott-Schwartz *et al.*, 1988; Lau & Neufeld, 1989; Morreau *et al.*, 1992; Hermans *et al.*, 1993). Perhaps the mutant 4S proteins are recognized as aberrant and degraded during the processes of translocation and folding of the nascent polypeptide by ER-resident molecular chaperones. Another possibility is that the mutant 4S may enter the ER but be missorted to a non-lysosomal destination, where in cannot fulfil its physiological role. Yet another possibility is that the mutant 4S enzymes may also conceivably be correctly targeted to the lysosome, but exhibit markedly reduced stability in the hydrolytic environment.

Attempts to elucidate the defect in multiple sulphatase deficiency (MSD) may also provide insight into the deficiency of 4S in MPS-VI patients. MSD is an autosomally inherited recessive disorder in which 4S is deficient together with all known lysosomal and microsomal sulphatases (reviewed in Kolodny, 1989; Guerra *et al.*, 1990). The primary defect in MSD has not been identified, but Rommerskirch & von Figura (1992) have concluded that the it affects a co- or post-translational process or factor that activates sulphatases or prevents their inactivation. The process or factor may be involved in folding or covalent modifications of the sulphatases. It is conceivable that an MPS-VI mutation may alter the conformation of the 4S polypeptide such that the interaction with the normal factor deficient in MSD is reduced, and consequently contribute to the 4S deficiency in MPS-VI.

The precise delineation of the mechanisms responsible for the reduced protein levels in the 4S mutants would require extensive experimental investigation, which was not pursued due to time constraints. Conceivably, degradation of the mutant

4S polypeptides in the ER or the lysosome may be resolved by the use of protease inhibitors, weak bases such as ammonium chloride, or the temperature at which the cells are incubated (Lippincott-Schwartz *et al.*, 1988; Hasilik, 1992). Pulse-chase experiments and determination of the residual 4S activity and protein in samples obtained by density gradient fractionation of cellular organelles may also refine the mechanism.

The H393P and L498P mutations in particular merit further analysis, as no detectable 4S protein or activity were found in cell extracts or conditioned media for these mutants. The severity of these mutations is intriguing, as both are located in the carboxyl-terminal portion of 4S, in regions of divergent protein sequence between sulphatases as discussed in Chapter 4.6. Both mutations involve the replacement of wild type residues with proline, a residue known to disrupt protein secondary structures. Determination of the mechanisms responsible for the reduction in mutant 4S protein in both the CHO expression system and the patient cell lines, may provide insights into the structural constraints in operation during the biosynthesis and maturation of normal 4S.

The phenomenon of markedly reduced 4S protein in both MPS-VI patient samples and in heterologous expression systems expressing patient mutations appears to be quite general. Wicker *et al.* (1991) characterized a 4S mutation (G137V) that affected a residue conserved in 9 out of 10 arylsulphatases. The mutation severely reduced the stability of the precursor, but did not affect the enzymes kinetic parameters. Isbrandt *et al.* (1994) identified six 4S mutations and expressed them in culture. All six mutant proteins were below the level of detection. The reduction in 4S protein appears to be the major contributor to the deficiency of 4S enzyme activity in MPS-VI patients. Whatever the mechanisms responsible for the reduced 4S protein, they exhibit a high degree of sensitivity and are particularly stringent,

as the levels of mutant proteins are markedly reduced even for those in which the amino acid substitution is predicted to have little or no effect on protein structure.

5.8.2 ACTIVITY REDUCTION

The yield of CHO-expressed mutant 4S activity was markedly reduced relative to wild type human 4S for all five 4S point mutations, as discussed in Section 5.5.2a and shown in Tables 5.1, 5.2 and 5.5. Comparable relative reductions in 4S activity toward either the 4MUS substrate (Doug Brooks, unpublished results, Table 5.7) or a radiolabelled trisaccharide (Brooks *et al.*, 1991b), were observed in patient cultured fibroblasts.

After correction for the reduction in protein, the CHO-expressed 4S mutants all showed a reduction in the specific activity towards 4MUS (Sections 5.5.2c, 5.5.2d and Table 5.4). These specific activity measurements were performed before the immunocapture assay had been optimized with regards to the concentration of the enzyme (dilution effect), and the incubation time and pH of the assay. Consequently, the optimized activity values described in Sections 5.5.3 and 5.5.4 would suggest that the specific activity of the T92M mutant is perhaps greater than 50% of wild type (especially at the pH optimum of 4.6), while the specific activity of the R95Q mutant may exceed 1% of normal. As the residual 4S protein was not determined in parallel with the optimized residual activity measurements, these tentative predictions have not been tested. In summary, the Y210C and T92M mutations did not appear to have had a major effect on the specific activity, and therefore are not apparently crucial to the structure and/or function of 4S.

The severe reduction in the specific activity of the R95Q mutant suggested that this residue was important to the structure and/or function of 4S, in contrast to Y210C and T92M. This hypothesis was supported by the total conservation of the arginine residue in the CTPSR sequence of 10 different sulphatases, as discussed

in Chapter 3, Section 3.3.8 and represented in Figure 3.13. However, the preliminary characterization of the R95Q mutation did not enable the precise mechanism of the reduced specific activity to be determined. The mutation may directly perturb critical residues involved in catalysis and/or substrate binding, and/or residues which are crucial to the molecular architecture of the enzyme.

In a speculative catalytic mechanism, the uniquely conserved arginine may bind to the sulphate moiety common to all the diverse sulphatase substrates, both natural and synthetic. In particular, the cationic terminal guanidino group of arginine may reduce the electron density over the anionic sulphate moiety by means of electrostatic and H-bonding. In consequence, the sulphate would be activated to enable nucleophilic attack at the central sulphur. The nucleophilic group is more difficult to predict, but is likely to be one of the totally conserved residues. Peters *et al.* (1990), suggested that the conserved histidine at position 147 in 4S may be involved in the active site of sulphatases. As more sulphatase sequences were cloned, this histidine did not remain conserved, but there is a totally conserved histidine at position 242. However, such speculation is premature, as a crystal structure for a sulphatase has not yet been reported, nor have detailed kinetic and/or structural investigations been performed. Such information, when available, may provide insight into the catalytic mechanism of sulphatases in general.

Recently, Schimmel (1993) has cautioned that such a view may be too simplistic. Thymidylate synthase was considered a model system for structure-function analysis; the reaction mechanism was known, the tertiary structure of the enzyme-substrate complex was defined, and the sequences of 17 other thymidylate synthases were available. A totally conserved asparagine residue was proposed to stabilize the dUMP substrate, based on the apparently self-evident structural data. In order to test the hypothesis, a complete replacement set of mutant enzymes at

the conserved asparagine were synthesized, and a thorough kinetic analysis was performed. Unexpectedly, the results were more consistent with the role of the conserved asparagine as a discriminator against the competitive substrate dCMP, rather than a selector for dUMP.

A contrasting caveat was made by Shi *et al.* (1993), who investigated the effects of a mutation in a totally conserved threonine residue of chicken muscle adenylate kinase. Kinetic data were consistent with a role that excluded major structural or catalytic functions. However, detailed NMR analysis indicated that the residue was involved in catalysis. The conclusion was that mutation of a functionally important residue may not lead to a large perturbation in kinetic constants, and that rigourous structure-function studies are necessary.

CHAPTER 6

SUMMARY & CONCLUDING DISCUSSION

The initial aim of the work presented in this thesis was to isolate and characterize a complete 4S cDNA clone. The subsequent aim was to identify and characterize mutations in the 4S gene of MPS-VI patients. The prior completion of these aims was necessary in order to achieve the principle aim, which was to test the hypothesized link between the mutant genotype, the biochemical phenotype, and the clinical phenotype. Additional aims were to develop molecular genetic approaches for the diagnosis and treatment of MPS-VI, and to gain insight into residues critical for 4S structure and function. In order to provide a context for discussion of the work presented in this thesis, the `state of the art' of MPS-VI research at the time the project was commenced will first be briefly reviewed. The results presented in the thesis will then be summarized, and the discussion will conclude with the implications for patient diagnosis, prediction of patient phenotype, assessment of patient therapy, and THE determination of residues critical for 4S structure and function.

Prior to the work described in this thesis, the enzyme deficient in MPS-VI was identified as *N*-acetylgalactosamine-4-sulphatase (4S) or arylsulphatase B (Stumpf *et al.*, 1973). Dermatan sulphate and chondroitin-4-sulphate were shown to be the *in vivo* substrates (O'Brien *et al.*, 1974; Matalon *et al.*, 1974), and a variety of *in vitro* substrates were developed (Baum *et al.*, 1959; Christomanou & Sandhoff, 1977; Hopwood *et al.*, 1986). The gene for 4S (*ARSB* locus) was mapped to chromosome 5p11-5q13 (Hellkuhl & Grzeschik, 1978; DeLuca *et al.*, 1979; Fox *et al.*, 1984; Fidzianska *et al.*, 1984). Several groups had purified human 4S protein from a variety of tissue sources, however the low abundance of 4S and the non-homogeneity of the purified protein caused ambiguity and difficulties with analysis (Roy, 1953; McGovern *et al.*, 1982; Steckel *et al.*, 1983). Gibson *et al.* (1987) reported the 10,000-fold purification of 4S from human liver, and the subsequent production of monoclonal antibodies to 4S, which were used to purify 4S to homogeneity by immunoaffinity chromatography. 4S purified from human liver

consisted of a 57-kDa species, which was composed of disulphide-linked 43-kDa and 8-kDa species.

Attempts to distinguish between the relative severities of MPS-VI clinical phenotypes based on the analysis of the biochemical phenotype of patient fibroblasts, including residual 4S enzyme activity, residual 4S protein, or 4S synthesis and processing, were only partially successful (Brooks *et al.*, 1990; Brooks *et al.*, 1991b; Taylor *et al.*, 1990). In MPS-VI patients, the entire spectrum of clinical phenotype from asymptomatic to severe is clustered around a low and narrow range of residual enzyme activity (and protein) between 0-5% of values from normal control fibroblasts. The low and narrow range, combined with the inherent variability of biological samples and the limitations of the assays, only enabled patients at either extreme of the clinical severity spectrum to be distinguished.

The work presented in this thesis commenced when oligonucleotide probes were designed based on amino acid sequence obtained from immunoaffinity-purified 4S (Gibson *et al.*, 1987). The probes were used to screen gene libraries for the 4S gene. Considerable yet unsuccessful efforts were made to isolate 4S cDNA clones from a number of cDNA libraries. However, screening of genomic DNA libraries resulted in the isolation of the first coding exon of 4S, within a 1.3-kb *Hin*dIII fragment. The 1.3-kb *Hin*dIII fragment encoded 104 amino acids of the 4S protein, namely a 38 amino acid signal peptide and 64 amino acids of the amino-terminal portion of the 43-kDa species (Litjens *et al.*, 1991). *In situ* hybridization was used to localize the putative first exon of 4S to chromosome 5q13 - 5q14 (Litjens *et al.*, 1989). This localization was coincident with the previous chromosomal assignments of the arylsulphatase B (*ARSB*) locus, and therefore confirmed the 1.3-kb *Hin*dIII fragment as encoding 4S, and refined the chromosomal localization of *ARSB*.

The deduced amino acid sequence of the first exon displayed considerable similarity to other (then) recently (1988-1989) cloned sulphatases, including steroid sulphatase (Yen *et al.,* 1987; Stein *et al.,* 1989b), arylsulphatase A (Stein *et al.,* 1989a), a sea urchin arylsulphatase (Sasaki *et al.,* 1988), and glucosamine-6-sulphatase (Robertson *et al.*, 1988). Multiple sequence alignment suggested a model accounting for the origin of the 4S subunits. The model proposed the proteolytic cleavage of the 57-kDa species into two disulphide-linked fragments; an amino-terminal 43-kDa species and a carboxyl-terminal 8-kDa species.

Subsequent efforts to isolate a 4S cDNA clone by screening cDNA libraries with the 1.3-kb *Hin*dIII fragment were not successful, despite much effort. The cloning of the 4S cDNA by Peters *et al.* (1990) and Schuchman *et al.* (1990) achieved this initial technical goal, albeit ahead of us, and illustrate the internationally competitive nature of the work attempted in this thesis. A collaboration was established with Prof. Kurt von Figura and Dr. Christoph Peters (Universität Göttingen), who provided the 4S cDNA clone.

The published full-length 4S cDNA sequence confirmed and refined the original model of 4S subunit structure described above. Recently, 4S has been shown to also contain a 7-kDa species, which led to a revised model, in which the 57-kDa polypeptide was proteolytically processed to an amino-terminal 43-kDa species, a central 7-kDa species and a carboxyl-terminal 8-kDa species, with the 7 and 8-kDa species disulphide linked to the 43-kDa species (Kobayashi *et al.,* 1992).

Mutational analysis was undertaken to provide a molecular explanation for the clinical heterogeneity seen in MPS-VI. Southern blot analysis of 18 MPS-VI patients failed to detect major deletions or rearrangements of the 4S gene, which suggested that the majority of MPS-VI mutations were point mutations or small rearrangements. Nine MPS-VI patients selected from across the spectrum of

clinical and biochemical phenotypes were subjected to detailed molecular analysis using the techniques of cDNA synthesis, polymerase chain reaction (PCR), chemical cleavage of mismatch, direct DNA sequencing of PCR products, and allele specific oligonucleotide hybridization. Eight mutations were identified including six single amino acid substitutions (T92M, R95Q, Y210C, H393P, P481L and L498P), one frameshift mutation (ΔG_{238}), and a partially characterized deletion mutation (Δexon IV).

Screening of a larger population of MPS-VI patients demonstrated that the identified mutations were rare or unique. Specifically, three of the mutations were found in more than one unrelated patient; Y210C in four patients (8% of mutant alleles), H393P in four patients (8% of mutant alleles) and R95Q in two patients (4% of mutant alleles). The three mutant alleles T92M, L498P, and ΔG_{238}, were each unique to individual patients, while the frequencies of the Δexon IV and P481L mutations are yet to be determined. Both mutant alleles were determined in six patients (two of whom were siblings), while in six patients one of the two alleles were identified. The majority of unrelated MPS-VI patients in which at least one mutant allele was defined (10/11) were compound heterozygotes. This is in contrast to the results of Isbrandt *et al.* (1994) who identified 4S mutations present in six patients, all of whom were homozygous for unique mutant alleles. Three polymorphisms were also identified during mutation screening; the amino acid substitutions V358M and V376M, and a silent change in the third base of the P397 codon. Whether the polymorphisms augment or attenuate the severity of the mutant alleles *in vivo* is not known. The number of heterogeneous mutations in the 4S gene of MPS-VI patients was found to be consistent with the broad clinical phenotype seen in this disorder.

Identification of the 4S mutant alleles in MPS-VI patients enabled the preliminary testing of the central hypothesis; the existence or otherwise of a genotype-

phenotype correlation in MPS-VI. A testable formulation of the hypothesis was that, given the genotype, did the *predicted* clinical and biochemical phenotypes correlate with the *actual* phenotypes? For the ΔG_{238} allele, the correlation was unambiguously verified. Sequence analysis of this allele predicted a frameshifted and markedly truncated 4S polypeptide, and therefore a severe phenotype. The predicted severity corresponded with the actual severe clinical presentation and biochemical phenotype of a patient who was homozygous for the ΔG_{238} mutation (Litjens *et al.*, 1992). The predicted severity of another gross mutation, Δexon IV, was also severe which correlated with the severe clinical phenotype of the affected patient. This patient was a compound heterozygote and although the other mutant allele was not known, the actual clinical severity was consistent with both mutant alleles severe.

In contrast to the gross mutations ΔG_{238} and ΔexonIV, the predicted phenotypic severity of the six substitution mutations was generally not self-evident from sequence analysis alone. Consequently, five of the mutant alleles were expressed in cell culture and confirmed as mutations by biochemical analysis. Each mutation was individually synthesized *in vitro* by oligonucleotide-directed mutagenesis of the normal 4S cDNA, and then transferred into the high-level expression vector pRSVN.07. The wild type and five mutant 4S expression constructs were then transfected into CHO cells, and the enzyme activity and the 4S protein concentration were determined. Compared to patient cells, the advantage of high-level expression in a heterologous system was the production of much greater quantities of 4S for analysis. In addition, the individual expression of each mutant allele enabled the effect of each mutation to be dissected, in contrast with patient cell lines, which often contained two different mutant 4S species.

Normal and mutant 4S was isolated from transfected cell extracts by monoclonal antibody immunocapture, and the 4S-specific arylsulphatase activity was assayed

in vitro using 4MUS. The expressed 4S protein was quantified using an ELISA technique. The expression and *in vitro* analysis of the 4S mutations enabled the prediction of the relative biochemical and clinical severity that results from each mutant 4S allele. The H393P and L498P mutations were predicted to be severe or null alleles, as no residual activity or protein were detected in CHO cells transfected with these mutant constructs. The R95Q mutation was also predicted to be a severe allele, as the residual activity was detectable although very low (0.003% to >0.2% of expressed control 4S, depending on the assay conditions and the concentration of the enzyme). The predicted severity of the Y210C mutation was milder than the above, as it exhibited a greater residual activity (about 3%). The T92M mutation was estimated to have a relative severity between that of the R95Q and Y210C mutations. The residual activities of the R95Q and T92M mutants were concentration dependent; both mutants exhibited a `dilution effect', where the specific activity of the mutant increased with dilution. This phenomenon hampered precise determination of the residual activity, and hence prediction of severity.

The contribution of each of the mutant alleles towards the patient's phenotype was estimated from the predicted severity of each of the mutant alleles, and then compared with the actual clinical and biochemical phenotype. The validity of the hypothesized genotype-phenotype correlation was confirmed for the majority of MPS-VI patients examined, given the limitations in each of the data components. For the eight patients heterozygous for one or more of the R95Q, Y210C, and H393P alleles, the predicted relative severity of each of the 4S mutant alleles was generally in reasonable agreement with the observed clinical phenotype, and to a lesser degree the biochemical phenotype measured in cultured fibroblasts. The T92M and L498P mutations were exceptions, in which the predicted relative severity did not correspond with the actual biochemical and clinical phenotype of the single compound heterozygous patient.

It must be stressed that the data used to test the MPS-VI genotype-phenotype hypothesis were preliminary. A more rigourous test of the hypothesis would require improvements in the determination of the individual data components, especially the clinical and biochemical phenotype of the patient, and the biochemical phenotype of the 4S mutants expressed in CHO culture. For example, an improved biochemical characterization of the 4S mutants could include determination of the rates of mutant protein synthesis and degradation *in vivo*, examination of the intracellular processing and transport, and kinetic measurements of the mutant enzymes towards the more specific trisaccharide substrate or even the dermatan sulphate substrate. Such information is likely to provide insight into the effect of a mutation, and the role of the non-mutated residue in enzyme structure and function. Mutational analysis of a larger number of MPS-VI patients and the use of statistical methods, may also clarify the validity of the hypothesis. The implications of the results presented in this thesis will now be discussed.

The techniques developed for the identification and subsequent screening of 4S mutations in MPS-VI patients, will contribute to the molecular genetic diagnosis of MPS-VI. As MPS-VI mutations are identified using the mutation identification techniques, at-risk families in which the mutant alleles are identified can be offered prenatal diagnosis. In such cases, molecular genetic diagnosis can be performed faster, and is technically easier to perform, than conventional enzyme assay. However, application of these techniques to the molecular genetic diagnosis of the majority of MPS-VI patients is currently premature. At present, identification of the mutant 4S alleles is a labour-intensive process. The observed molecular heterogeneity and the absence of predominant mutant alleles in the MPS-VI patient population would increase the workload of mutation identification

and screening to an unsustainable/uneconomic level, if applied as a general diagnostic technique.

Although the techniques are not yet suitable for the identification of undefined mutations in large numbers of patients, the continual development of the techniques is likely to lead to increased sensitivity and reduced workload in future. For example, the recent determination of the 4S genomic structure would permit the multiplex PCR amplification of the 4S gene direct from genomic DNA, without the separate step of cDNA synthesis. In addition, developments in automated DNA sequencing may enable the rapid identification of the mutation directly from the PCR products, by-passing the chemical cleavage technique. In future, it is envisaged that molecular genetic techniques will contribute to the diagnosis of the majority of MPS-VI cases.

The preliminary verification of a genotype-phenotype correlation in MPS-VI may, with further refinement, contribute to the prediction of clinical prognosis. For example, in patients carrying mutant alleles which have previously been identified in other patients and characterized *in vitro*, the prognosis may be predicted cautiously from the clinical course of the original patients. In those patients with unique mutant alleles, the prediction of prognosis is more difficult, and would require detailed biochemical investigation of the patient cultured cells, and of the mutant alleles expressed individually in a heterologous system. The preliminary genotype-phenotype correlation may also contribute to the assessment of various therapeutic approaches in MPS-VI patients. Therapeutic approaches for MPS-VI include bone marrow transplantation, while enzyme replacement and gene therapies for MPS-VI are under development (Anson *et al.*, 1992b; John Hopwood, personal communication). An assessment of each therapeutic approach would require the comparison of the clinical course after intervention with the predicted

prognosis if untreated. Such a prediction would be based on analysis of the mutant alleles.

Exceptions to the general genotype-phenotype correlation in MPS-VI may be detected, for example the T92M and/or L498P alleles in patient SF2467 are potential candidates. Unfortunately, genotype-phenotype anomalies will confound the prediction of clinical prognosis and the assessment of various therapeutic approaches in these MPS-VI patients. However, investigation of such discrepancies between the predicted and actual biochemical and clinical phenotypes may reveal deeper insights into the molecular genetic mechanisms and/or environmental factors which modulate disease severity.

In addition to testing the genotype-phenotype correlation, the heterologous expression and biochemical characterization of 4S mutants may also provide an insight into amino acid residues critical for the structure and catalytic mechanism of 4S. Critical residues are likely to be identified through the investigation of mutations that are located in regions that are conserved between sulphatases, mutations that severely reduce enzyme activity or stability, and mutations that cause inefficient intracellular targeting of the enzyme to the lysosome. A mutation could conceivably mediate its effects through more than one mechanism, which are not necessarily mutually exclusive.

The 4S mutations characterized in this study all led to a reduction in 4S activity and protein to various extents. Of particular interest was the R95Q mutation, which was a non-conservative change of a residue totally conserved between 10 sulphatases, and which led to a marked reduction in specific activity of the enzyme. However, detailed speculation concerning the molecular consequences of the mutations on 4S structure and function is premature. Of crucial importance is knowledge of the 4S tertiary structure, which is yet to be determined. Knowledge

of the tertiary structure, in conjunction with site-directed mutagenesis of selected residues and subsequent analysis (kinetics, intracellular transport, epitope mapping) of the mutant protein may identify residues that are crucial for catalysis, substrate specificity, and structure.

APPENDICES

APPENDIX A OLIGONUCLEOTIDES USED IN THE PCR AMPLIFICATION & DIRECT DNA SEQUENCING OF 4S

The designation, sequence, position, and orientation of synthetic oligonucleotide primers used for PCR amplification and direct DNA sequencing of 4S PCR products. The positions of the primers are numbered according to Peters *et al.* (1990), except for primers 4SP16, 4SP5, and 4SP9, which are proximal to the methionine initiation codon and are numbered according to Schuchman *et al.* (1990).

Primer	DNA sequence (5'->3')	Position	sense (s) or antisense (a)
4SP16	CGG CAG CCC AGT TCC TCA TT	-79 - -60	s
4SP5	GGG ACC GCG GGC CGG ACA AGG	-20 - -1	s
4SP9	GGG CGG ACA AGG ATG GGT CC	-12 - 8	s
4SP5A	ATG GGT CCG CGC GGC GCG GCG	1 - 21	s
4SP1	CTC CCC GTC GTC CTC CC	58 - 74	s
4SP15	GTG CTC CTG GAC AAC TAC TA	238 - 257	s
4SP14	TAG CGG CCA GTG AGC AGC TG	289 - 308	a
4SP10	ATC CGT ACA GGT TTA CAG CA	313 - 332	s
4SP11	TCA TCC AGA GGA ACA CAG CT	358 - 377	a
4SP21	GGA AGG CAT TCT TTC CGG TA	451 - 470	a
4SP12	TAA TCT TCA CTA CCC AGG AG	505 - 524	a
4SP22	GAG ATG GCG AAG AAG TTG CA	590 - 609	s
4SP30	TCT CTG GTG GAT GGT TAG TT	669 - 688	a
4SP23	CTG CAT AGT GAT GCC TGT TC	783 - 802	a
4SP6	CTA TGC AGG AAT GGT GTC CC	795 - 814	s

Primer	DNA sequence (5'->3')	Position	sense (s) or antisense (a)
4SP7	GCC AGT TAT TAC CCC CTG CC	918 - 937	a
4SP24	GCA GGG GGT AAT AAC TGG CC	919 - 938	s
4SP26	TCC TTC ACT GAT GGT TTT CC	1130 - 1149	a
4SP31	TGA AGG AAG CCC ATC CCC CA	1143 - 1162	s
4SP32	CGG TGA AGA GTC CAC GAA GT	1193 - 1232	a
4SP27	AAC ACA TCT GTC CAT GCT GC	1276 - 1295	s
4SP28	AGC CCG TGA GGA GTT TCC AA	1311 - 1330	a
4SP33	GCT GTG GTT ACT GGT TCC CT	1337 - 1356	s
4SP29	GAC AGG AGC TTT GTG ACG AT	1480 - 1499	a
4SP8	TGA AAG GTT TTC TAG CCT CC	1611 - 1630	a

APPENDIX B ALLELE SPECIFIC OLIGONUCLEOTIDES FOR THE DETECTION OF 4S MUTATIONS & FOR SITE-DIRECTED MUTAGENESIS.

The designation of each synthetic oligonucleotide primer and the encoded 4S allele, whether wild-type (wt) or mutant (m), is shown. The DNA sequence, and the position and orientation of the primers are also shown. The positions of the primers are numbered according to Peters *et al.* (1990). The allele-specific nucleotides are underlined.

Primer	Allele	DNA sequence (5'->3')	Position	sense (s) or antisense (a)
4SP17	G_{238} (wt)	GCC GGC GGG GTG CTC CTG G	229 - 247	s
4SP18	ΔG_{238} (m)	GCC GGC GGG TGC TCC TGG	229 - 247	s
4SThr92	T92 (wt)	GCT GTG CAC GCC GTC G	267 - 282	s
4SMet92	M92 (m)	GCT GTG CAT GCC GTC G	267 - 282	s
4SP34	M92 (m)	CGA CGG CAT GCA CAG C	267 - 282	a
4SArg95	R95 (wt)	CGT CGC GGA GCC AGC TG	278 - 294	s
4SGln95	Q95 (m)	CGT CGC AGA GCC AGC TG	278 - 294	s
4STyr210	Y210 (wt)	TAA AAA TAT GTA TTC AAC AA	618 - 637	s
4SCys210	C210 (m)	TAA AAA TAT GTG TTC AAC AA	618 - 637	s
4SHis393	H393 (wt)	CTG CTG CAT AAT ATT GAC CC	1171 - 1190	s
4SPro393	P393 (m)	CTG CTG CCT AAT ATT GAC CC	1171 - 1190	s
4SLeu498	L498 (wt)	GTC ACA AAG CTC CTG TCC C	1483 - 1501	s
4SPro498	P498 (m)	GTC ACA AAG CCC CTG TCC C	1493 - 1501	s

APPENDIX C REACTION CONDITIONS FOR THE DIRECT DNA SEQUENCING OF PCR PRODUCTS

For each PCR product, the template and the oligonucleotide primer pair used to synthesize the product are shown. Direct DNA sequencing of PCR products was performed as described in Section 4.2.4b, using the PCR sequencing primers at the annealing temperatures shown, either in the presence (+) or absence (-) of 10% (v/v) DMSO. The sequences of the primers are listed in Appendix B.

PCR product	Template	Sequencing PCR primer	Annealing temp. (°C)	DMSO (+/-)
5A-7 (A)	cDNA	4SP5A	55	+
		4SP11	55	+
		4SP17	55	+
		4SP22	55	+
		4SP23	55	+
		4SP30	58	+
6-8 (B)	cDNA	4SP6	58	-
		4SP28	58	-
		4SP27	58	-
		4SP8	58	-
16-11	cDNA	4SP16	55	+
		4SP5A	55	+
		4SP11	55	+
10-7	cDNA	4SP22	55	-
		4SP7	55	-

PCR product	Template	Sequencing PCR primer	Annealing temp. (°C)	DMSO (+/-)
16-14	genomic	4SP5A	55	+
		4SP17	55	+
		4SP14	55	+
24-28	cDNA	4SP24	55	-
		4SP28	55	-
27-8	cDNA	4SP27	55	-
		4SP29	55	-
		4SP8	58	-

APPENDIX D WASHING TEMPERATURES FOR THE ASO DETECTION OF 4S MUTATIONS

For each 4S mutation, the primers used to generate the PCR product which spanned the mutation are shown, together with the allele specific oligonucleotides which recognize either the wild type (wt) or mutant (m) allele. Genomic DNA was used as the template for the PCR amplification. The sequences of the primers are listed in Appendix B. Southern transfers of the PCR products were hybridized as described in Section 4.2.5. The final washing of each ASO-probed filter was performed in 2 X SSC, 0.1% (w/v) SDS, at the temperature shown.

Mutation	PCR product	ASO primer	Allele	Washing temp. (°C)
ΔG_{238}	5-14	4SP17	wt	67
		4SP18	m	67
T92M	16-14	4SThr92	wt	62
		4SP34	m	56
R95Q	16-14	4SArg95	wt	57
		4SGln95	m	60
Y210C	22-30	4STyr210	wt	47
		4SCys210	m	48
H393P	31-32	4SHis393	wt	58
		4SPro393	m	60
L498P	33-8	4SLeu498	wt	65
		4SPro498	m	65

BIBLIOGRAPHY

Anson, D.S., Bielicki, J., & Hopwood, J.J. (1992a) Correction of human mucopolysaccharidosis type I fibroblasts by retroviral-mediated transfer of the human α-L-iduronidase gene. *Human Gene Therapy* **3:** 371-379.

Anson, D.S., Muller, V., Bielicki, J., Harper, G.S. & Hopwood, J.J. (1993) Overexpression of *N*-acetylgalactosamine-4-sulphatase induces a multiple sulphatase deficiency in mucopolysaccharidosis-type-VI fibroblasts. *Biochem. J.* **294:** 657-662.

Anson, D.S., Taylor, J.A., Bielicki, J., Harper, G.S., Peters, C., Gibson, G.J. & Hopwood, J.J. (1992b) Correction of human mucopolysaccharidosis type-VI fibroblasts with recombinant *N*-acetylgalatosamine-4-sulphatase. *Biochem. J.* **284:** 789-794.

Arlt, G., Brooks, D.A., Isbrandt, D., Hopwood, J.J., Bielicki, J., Bradford, T.M., Bindloss-Petherbridge, C.A., von Figura, K. & Peters, C. (1994) Juvenile form of mucopolysaccharidosis VI (Maroteaux-Lamy syndrome): a C-terminal extension causes instability but increases catalytic efficiency of arylsulfatase B. *J. Biol. Chem.* **269**: 9638-9643.

Ballabio, A., Carrozzo, R., Parenti, G., Gil, A., Zollo, M., Persico, M.G., Gillard, E., Affara, N., Yates, J., Ferguson-Smith, M.A, Frants, R.R., Eriksson, A.W. & Andria, G. (1989) Molecular heterogeneity of steroid sulfatase deficiency: a multicenter study on 57 unrelated patients, at DNA and protein levels. *Genomics* **4**: 36-40.

Baranski, T.J., Cantor, A.B. & Kornfeld, S. (1992) Lysosomal enzyme phosphorylation. I. Protein recognition determinants in both lobes of procathepsin D mediate its interaction with UDP-GlcNAc:lysosomal enzyme *N*-acetylglucosamine-1-phosphotransferase. *J. Biol. Chem.* **267**: 23342-23348.

Barbeyron, T. (1992) *Alteromonas carrageenovora* gene for arylsulphatase. (unpublished GenBank entry ACARYL).

Barriocanal, J.G., Bonifacino, J.S., Yuan, L. & 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, R.W. & Neufeld, E.F. (1972) A distinct biochemical deficit in the Maroteaux-Lamy syndrome (mucopolysaccharidosis VI). *J. Pediat.* **80**: 114-116.

Baum, H., Dodgson, K.S. & Spencer, B. (1959) The assay of arylsulphatases A and B in human urine. *Clin. Chim. Acta* **4**: 453-455.

Benoist, C. & Chambon, P. (1981) *In vivo* sequence requirements of the SV40 early promoter region. *Nature* **290**: 304-310.

Berg-Fussman, A., Grace, M.E., Ioannou, Y. & Grabowski, G.A. (1993) Human acid β-glucosidase: *N*-glycosylation site occupancy and the effect of glycosylation on enzymatic activity. *J. Biol. Chem.* **268**: 14861-14866.

Beutler, E. (1993) Gaucher disease as a paradigm of current issues regarding single gene mutations of humans. *Proc. Natl. Acad. Sci. USA* **90**: 5384-5390.

Bielicki, J., Hopwood, J.J., Wilson, P.J. & Anson, D.S. (1993) Recombinant human iduronate-2-sulphatase: correction of mucopolysaccharidosis-type II fibroblasts and characterization of the purified enzyme. *Biochem. J.* **289**: 241-246.

Birnboim, H.C. & Doly, J. (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res.* **7**: 1513-1523.

Blaese, R.M. (1993) Development of gene therapy for immunodeficiency: adenosine deaminase deficiency. *Pediatr. Res.* **33** (Suppl.): S49-S55.

Bou-Gharios, G., Abraham, D. & Olsen, I. (1993) Lysosomal storage diseases: mechanisms of enzyme replacement therapy. *Histochem. J.* **25**: 593-605.

Brante, G. (1952) Gargoylism - a mucopolysaccharidosis. *Scand. J. Clin. Lab. Invest.* **4:** 43-46.

Breton, L., Guerin, P. & Morin, M. (1983) A case of mucopolysaccharidosis in a cat. *J. Am. Anim. Hosp. Assoc.* **19**: 891-896.

Brody, L.C., Mitchell, G.A., Obie, C., Michaud, J., Steel, G., Fontaine, G., Robert, M.-F., Sipila, I., Kaiser-Kupfer, M. & Valle, D. (1992) Ornithine δ-aminotransferase mutations in gyrate atrophy: allelic heterogeneity and functional consequences. *J. Biol. Chem.* **267**: 3302-3307.

Brooks, D.A. (1993) The immunochemical analysis of enzyme from mucopolysaccharidoses patients. *J. Inher. Metab. Dis.* **16**: 3-15.

Brooks, D.A., Gibson, G.J., McCourt, P.A.G. & Hopwood, J.J. (1991a) A specific fluorogenic assay for N-acetylgalactosamine-4-sulphatase activity using immunoadsorption. *J. Inher. Metab. Dis.* **14**: 5-12.

Brooks, D.A., McCourt, P.A.G., Gibson, G.J., Ashton, L.J., Shutter, M. & Hopwood, J.J. (1991b) Analysis of N-acetylgalactosamine-4-sulfatase protein and kinetics in mucopolysaccharidosis type VI patients. *Am. J. Hum. Genet.* **48**: 710-719.

Brooks, D.A., McCourt, P.A.G., Gibson, G.J. & Hopwood, J.J. (1990) Immunoquantification of the low abundance lysosomal enzyme N-acetylgalactosamine 4-sulphatase. *J. Inher. Metab. Dis.* **13**: 108-120.

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.

Callen, D.F., Hyland, V.J., Baker, E.G., Fratini, A., Simmers, R.N., Mulley, J.C. & Sutherland, G.R. (1988) Fine mapping of gene probes and anonymous DNA fragments to the long arm of chromosome 16. *Genomics* **2**: 144-153.

Cantor, A.B., Baranski, T.J. & Kornfeld, S. (1992) Lysosomal enzyme phoshorylation. II. Protein recognition determinants in either lobe of procathepsin D are sufficient for phosphorylation of both the amino and carboxyl lobe oligosaccharides. *J. Biol. Chem.* **267**: 23349-23356.

Cantz, M. & Gehler, J. (1976) The mucopolysaccharidoses: inborn errors of glycosaminoglycan catabolism. *Hum. Genet.* **32**: 233-255.

Chaconas, G. & van de Sande, J.H. (1980) 5'-^{32}P labelling of RNA and DNA restriction fragments. *Methods Enzymol.* **65**: 75-85.

Chang, P.L., Rosa, N.E. & Davidson, R.G. (1981) Differential assay of arylsulphatase A and B activities: a sensitive method for cultured human cells. *Anal. Biochem.* **117**: 382-389.

Chelly, J., Gilgenkrantz, H., Lambert, M., Hamard, G., Chafey, P., Récan, D., Katz, P., de la Chapelle, A., Koenig, M., Ginjaar, I.B., Fardeau, M., Tomé, F., Kahn, A., Kaplan, J.-C. (1990) Effect of dystrophin gene deletions on mRNA levels and processing in Duchenne and Becker muscular dystrophies. *Cell* **63**: 1239-1248.

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

Chou, P.Y. & Fasman, G.D. (1974) Conformational parameters for amino acids in helical, β-sheet, and random coil regions calculated from proteins. *Biochemistry* **13**: 211-222.

Chou, P.Y. & Fasman, G.D. (1978) Prediction of secondary structure of proteins from their amino acid sequence. *Adv. Enzymol.* **47**: 45-148.

Christomanou, H. & Sandhoff, K. (1977) A sensitive fluorescence assay for the simultaneous and separate determination of arylsulphatases A and B. *Clin. Chim. Acta* **79**: 527-531.

Conner, B.J., Reyes, A.A., Morin, C., Itakura, K., Teplitz, R.L. & Wallace, R.B. (1983) Detection of sickle cell β^s-globin allele by hybridization with synthetic oligonucleotides. *Proc. Natl. Acad. Sci. USA* **80**: 278-282.

Conzelmann, E. & Sandhoff, K. (1983/1984) Partial enzyme deficiencies: residual activities and the development of neurological disorders. *Dev. Neurosci.* **6:** 58-71.

Cooper, D.N. & Youssoufian, H. (1988) The CpG dinucleotide and human genetic disease. *Hum. Genet.* **78**: 151-155.

Cotton, R.G.H. (1989) Detection of single base changes in nucleic acids. *Biochem. J.* **263**: 1-10.

Cotton, R.G.H. (1993) Current methods of mutation detection. *Mutation Res.* **285**: 125-144.

Cotton, R.G.H., Rodrigues, N.R. & Campbell, R.D. (1988) Reactivity of cytosine and thymine in single-base-pair mismatches with hydroxylamine and osmium tetroxide and its application to the study of mutations. *Proc. Natl. Acad. Sci. USA* **85**: 4397-4401.

Craig, E.A., Gambill, B.D. & Nelson, R.J. (1993) Heat shock proteins: molecular chaperones of protein biogenesis. *Microbiol. Rev.* **57**: 402-414.

Cullen, B.R. (1988) Expression of a cloned human interleukin-2 cDNA is enhanced by the substitution of a heterologous mRNA leader region. *DNA* **7**: 645-650.

Curiel, D.T., Buchhagen, D.L., Chiba, I., D'Amico, D., Takahashi, T. & Minna, J.D. (1990) A chemical mismatch cleavage method useful for the detection of point mutations in the p53 gene in lung cancer. *Am. J. Respir. Cell Mol. Biol.* **3**: 405-411.

Dana, S. & Wasmuth, J.J. (1982) Selective linkage disruption in human-Chinese hamster cell hybrids: deletion mapping of the *leuS*, *hexB*, *emtB*, and *chr* genes on human chromosome 5. *Mol. Cell. Biol.* **2**: 1220-1228.

Daniele, A., Faust, C.J., Herman, G.E., Di Natale, P. & Ballabio, A. (1993) Cloning and characterization of the cDNA for the murine iduronate sulfatase gene. *Genomics* **16**: 755-757.

de Duve, C. (1963) The lysosome concept. In de Reuck, A.V.S. & Cameron, M.P. (eds.); *Lysosomes*. Ciba Foundation symposium on lysosomes. Churchill, London, pp. 1-35.

de Hostos, E.L., Schilling, J. & Grossman, A.R. (1989) Structure and expression of the gene encoding the periplasmic arylsulfatase of *Chlamydomonas reinhardtii*. *Mol. Gen. Genet.* **218**: 229-239.

Delidow, B.C., Peluso, J.J. & White, B.A. (1989) Quantitative measurement of mRNAs by polymerase chain reaction. *Gene Anal. Techn.* **6**: 120-124.

DeLuca, C., Brown, J.A. & Shows, T.B. (1979) Lysosomal arylsulfatase deficiencies in humans: chromosome assignments for arylsulfatase A and B. *Proc. Natl. Acad. Sci. USA* **76**: 1957-1961.

Devereux, J., Haeberli, P. & Smithies, O. (1984) A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.* **12**: 387-395.

Dietz, H.C., Valle, D., Francomano, C.A., Kendzior, R.J., Jr., Pyeritz, R.E. & Cutting, G.R. (1993) The skipping of constitutive exons in vivo induced by nonsense mutations. *Science* **259**: 680-682.

Di Natale, P., Annella, T., Daniele, A., Spagnuolo, G., Cerundolo, R., de Caprariis, D. & Gravino, A.E. (1992) Animal models for lysosomal diseases: a new case of feline mucopolysaccharidosis VI. *J. Inher. Metab. Dis.* **15**: 17-24.

Dorfman, A. & Lorincz, A.E. (1957) Occurrence of urinary acid mucopolysaccharides in the Hurler syndrome. *Proc. Natl. Acad. Sci. USA* **43:** 443-446.

Dunning, A.M., Talmud, P. & Humphries, S.E. (1988) Errors in the polymerase chain reaction. *Nucleic Acids Res.* **16**: 10393.

Faisst, S. & Meyer, S. (1992) Compilation of vertebrate-encoded transcription factors. *Nucleic Acids Res.* **20**: 3-26.

Feinberg, A.P. & Vogelstein, B. (1983) A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.* **132**: 6-13.

Fidzianska, E., Abramowicz, T., Czartoryska, B., Glogowska, I., Gorska, D. & Rodo, M. (1984) Assignment of the gene for human arylsulfatase B, *ARSB*, to chromosome region 5p11-5qter. *Cytogenet. Cell. Genet.* **38**: 150-151.

Fink, J.K., Correll, P.H., Perry, L.K., Brady, R.O. & Karlsson, S. (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.

Fisher, K.J. & Aronson, N.N. (1989) Isolation and sequence analysis of a cDNA encoding rat liver α-L-fucosidase. *Biochem. J.* **264**: 695-701.

Flomen, R.H., Green, E.P., Green, P.M., Bentley, D.R. & Giannelli, F. (1993) Determination of the organisation of coding sequences within the iduronate sulphate sulphatase (IDS) gene. *Hum. Mol. Genet.* **2**: 5-10.

Forrest, S.M., Dahl, H.H., Howells, D.W., Dianzani, I. & Cotton, R.G.H. (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**: 175-183.

Fox, M.F., DuToit, D.L., Warnich, L. & Retief, A.E. (1984) Regional localization of α-galactosidase (*GLA*) to Xpter-q22, hexosaminidase B (*HEXB*) to 5q13-qter, and arylsulfatase B (*ARSB*) to 5pter-q13. *Cytogenet. Cell. Genet.* **38**: 45-49.

Fransson, L.-Å. (1987) Structure and function of cell-associated proteoglycans. *Trends Biochem. Sci.* **12**: 406-411.

Fratantoni, J.C., Hall, C. W. & Neufeld, E.F. (1968) 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, C.W. & Neufeld, E.F. (1969) The defect in Hurler and Hunter syndromes; II. deficiency of specific factors involved in mucopolysaccharide degradation. *Proc. Natl. Acad. Sci. USA* **64:** 360-366.

Friedman, K.D., Rosen, N.L., Newman, P.J., & Montgomery, R.R. (1988) Enzymatic amplification of specific cDNA inserts from λgt11 libraries. *Nucleic Acids Res.* **16**: 8718.

Frischauf, A.-M., Lehrach, H., Poustka, A. & Murray, N. (1983) Lambda replacement vectors carrying polylinker sequences. *J. Mol. Biol.* **170**: 827-842.

Frohman, M.A., Dush, M.K. & Martin, G.R. (1988) Rapid amplification of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. *Proc. Natl. Acad. Sci. USA* **85**: 8998-9002.

Gahl, W.A., Renlund, M. & Thoene, J.G. (1989) Lysosomal transport disorders: cystinosis and sialic acid storage disorders. In Scriver, C.R., Beaudet, A.L., Sly, W.S. & Valle, D. (eds.); *The Metabolic Basis of Inherited Disease.* McGraw-Hill, New York, pp. 2619-2467.

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. & Hoover, E.A. (1984) Correction of feline arylsulphatase B deficiency (mucopolysaccharidosis VI) by bone marrow transplantation. *Nature* **312:** 467-469.

Geier, C., von Figura, K. & Pohlmann, R. (1989) Structure of the human lysosomal acid phosphatase gene. *Eur. J. Biochem.* **183**: 611-616.

Gibson, G.J., Saccone, G.T.P., Brooks, D.A., Clements, P.R. & Hopwood, J.J. (1987) Human N-acetylgalactosamine-4-sulphate sulphatase. Purification, monoclonal antibody production and native and subunit M_r values. *Biochem. J.* **248**: 755-764.

Gieselmann, V., Polten, A., Kreysing, J., Kappler, J., Fluharty, A., Bohne, W. & von Figura, K. (1991) Mutations in arylsulfatase A alleles causing metachromatic leukodystrophy. *Brain Dysfunct.* **4**: 235-243.

Gieselmann, V., Polten, A., Kreysing, J. & von Figura, K. (1989) Arylsulfatase A pseudodeficiency: loss of a polyadenylation signal and *N*-glycosylation site. *Proc. Natl. Acad. Sci. USA* **86**: 9436-9440.

Goda, Y. & Pfeffer, S.R. (1988) Selective recycling of the mannose 6-phosphate/IGF-II receptor to the *trans* Golgi network *in vitro*. *Cell* **55**: 309-320.

Goldberg, M.F., Scott, C.I. & McKusick, V.A. (1970) Hydrocephalus and papilledema in the Maroteaux-Lamy syndrome (mucopolysaccharidosis type VI). *Am. J. Opthal.* **69**: 969-975.

Gottschalk, S., Waheed, A., Schmidt, B., Laidler, P. & von Figura, K. (1989) Sequential processing of lysosomal acid phosphatase by a cytoplasmic thiol proteinase and a lysosomal aspartyl proteinase. *EMBO. J.* **8**: 3215-3219.

Graham, J.M. & Winterbourne, D.J. (1988) Subcellular localization of the sulphation reaction of heparan sulphate synthesis and transport of the proteoglycan to the cell surface in rat liver. *Biochem. J.* **252**: 437-445.

Griffiths, G., Hoflack, B., Simons, K., Mellman, I. & Kornfeld, S. (1988) The mannose 6-phosphate receptor and the biogenesis of lysosomes. *Cell* **52**: 329-341.

Grompe, M., Pieretti, M., Caskey, C.T. & Ballabio, A. (1992) The sulfatase gene family: cross-species PCR cloning using the MOPAC technique. *Genomics* **12**: 755-760.

Guarnieri, F.G., Arterburn, L.M., Penno, M.B., Cha, Y. & August, J.T. (1993) The motif Tyr-X-X-hydrophobic residue mediates lysosomal membrane targeting of lysosome-associated membrane protein 1. *J. Biol. Chem.* **268**: 1941-1946.

Guerra, W.F., Verity, M.A., Fluharty, A.L., Nguyen, H.T. & Philippart, M. (1990) Multiple sulfatase deficiency: clinical, neuropathological, ultrastructural and biochemical studies. *J. Neuropathol. Exp. Neurol.* **49**: 406-423.

Gyllensten, U.B. (1989) PCR and DNA sequencing. *Biotechniques* **7**: 700-708.

Hardingham, T.E. & Fosang, A.J. (1992) Proteoglycans: many forms and many functions. *FASEB J.* **6**: 861-870.

Hasilik, A. (1992) The early and late processing of lysosomal enzymes: proteolysis and compartmentation. *Experientia (Basel)* **48**: 130-151.

Hasilik, A. & von Figura, K. (1984) Processing of lysosomal enzymes in fibroblasts. In Dingle, J.T., Dean, R.T. & Sly, W. (eds.); *Lysosomes in biology and pathology*. Elsevier, Amsterdam, **7:** 3-16.

Haskins, M.E., Jezyk, P.F. & Patterson, D.F. (1979) Mucopolysaccharide storage disease in three families of cats with arylsulphatase B deficiency: leukocyte studies and carrier identification. *Pediat. Res.* **13**: 1203-1210.

Hayflick, S., Rowe, S., Kavanaugh-McHugh, A., Olson, J.L. & Valle, D. (1992) Acute infantile cardiomyopathy as a presenting feature of mucopolysaccharidosis VI. *J. Pediatr.* **120**: 269-272.

Heilig, J.S., Lech, K. & Brent, R. (1993) CsCl/Ethidium bromide equilibrium centrifugation. In Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A. & Struhl, K. (eds.); *Current Protocols in Molecular Biology*. Wiley, New York, pp. 1.7.5-1.7.7.

Hellkuhl, B. & Grzeschik, K.-H. (1978) Assignment of a gene for arylsulfatase B to human chromosome 5 using human-mouse somatic cell hybrids. *Cytogenet. Cell Genet.* **22**: 203-206.

Hendrick, J.P. & Hartl, F.-U. (1993) Molecular chaperone functions of heat-shock proteins. *Annu. Rev. Biochem.* **62**: 349-384.

Hermans, M.M.P., de Graaff, E., Kroos, M.A., Wisselaar, H.A., Willemsen, R., Oostra, B.A. & Reuser, A.J.J. (1993) The conservative substitution Asp-645->Glu in lysosomal α-glucosidase affects transport and phosphorylation of the enzyme in an adult patient with glycogen-storage disease type II. *Biochem. J.* **289**: 687-693.

Hers, H.-G. (1965) Inborn lysosomal diseases. *Gastroenterology* **48:** 625-633.

Hers, H.-G., Van Hoof, F. & de Barsy, T. (1989) Glycogen storage diseases. In Scriver, C.R., Beaudet, A.L., Sly, W.S. & Valle, D. (eds.); *The Metabolic Basis of Inherited Disease*. McGraw-Hill, New York, pp. 425-452.

Higgins, D.G., Bleasby, A.J. & Fuchs, R. (1991) CLUSTAL V: improved software for multiple sequence alignment. *Comput. Appl. Biosci.*, submitted.

Higgins, D.G. & Sharp, P.M. (1988) CLUSTAL: a package for performing multiple sequence alignments on a microcomputer. *Gene* **73**: 237-244.

Higgins, D.G. & Sharp, P.M. (1989) Fast and sensitive multiple sequence alignments on a microcomputer. *Comput. Appl. Biosci.* **5**: 151-153.

Hobbs, J.R. (1987) Experience with bone marrow transplantation for inborn errors of metabolism. *Enzyme* **38**: 194-206.

Hong, W. & Tang, B.L. (1993) Protein trafficking along the exocytotic pathway. *Bioessays* **15**: 231-238.

Hoogerbrugge, P.M. & Vossen, J.M.J.J. (1990) Bone marrow transplantation in the treatment of lysosomal storage disorders. In Fernandes, J., Saudubray, J.-M. & Tada, K. (eds.); *Inborn Metabolic Diseases: Diagnosis and Treatment.* Springer-Verlag, Berlin, pp. 659-670.

Hopwood, J.J., Bunge, S., Morris, C.P., Wilson, P.J., Steglich, C., Beck, M., Schwinger, E. & Gal, A. (1993a) Molecular basis of mucopolysaccharidosis type II: mutations in the iduronate-2-sulphatase gene. *Human Mutation* **2**: 435-442.

Hopwood, J.J. & Elliott, H. (1985) Urinary excretion of sulphated *N*-acetylhexosamines in patients with various mucopolysaccharidoses. *Biochem. J.* **229:** 579-586.

Hopwood, J.J., Elliott, H., Muller, V.J. & Saccone, G.T.P. (1986) Diagnosis of Maroteaux-Lamy syndrome by the use of radiolabelled oligosaccharides as substrates for the determination of arylsulphatase B activity. *Biochem. J.* **234**: 507-514.

Hopwood, J.J. & 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. & Morris, C.P. (1990) The mucopolysaccharidoses: diagnosis, molecular genetics and treatment. *Mol. Biol. Med.* **7:** 381-404.

Hopwood, J.J., Vellodi, A., Scott, H.S., Morris, C.P., Litjens, T., Clements, P.R., Brooks, D.A., Cooper, A. & Wraith, J.E. (1993b) Long-term clinical progress in bone marrow transplanted mucopolysaccharidosis type I patients with a defined genotype. *J. Inher. Metab. Dis.* **16**: 1024-1033.

Hovingh, P., Piepkorn, M. & Linker, A. (1993) Differentially expressed patterns of glycosaminoglycan structure in heparan sulphate proteoglycans and free chains. *Eur. J. Biochem.* **211**: 771-779.

Huynh, T.V., Young, R.A. & 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.

Isbrandt, D., Arlt, G., Brooks, D.A., Hopwood, J.J., von Figura, K. & Peters, C. (1994) Mucopolysaccharidosis VI (Maroteaux-Lamy syndrome): six unique arylsulfatase B gene alleles causing variable disease phenotypes. *Am. J. Hum. Genet.* **54**: 454-463.

Jackson, C.E., Yuhki, N., Desnick, R.J., Haskins, M.E, O'Brien, S.J. & Schuchman, E.H. (1992) Feline arylsulfatase B (ARSB): isolation and expression of the cDNA, comparison with human ARSB, and gene localization to feline chromosome A1. *Genomics* **14**: 403-411.

James, G.T. (1979) Essential arginine residues in human liver arylsulfatase A. *Arch. Biochem. Biophys.* **197**: 57-62.

Jansen, R. & Ledley, F.D. (1990) Heterozygous mutations at the *mut* locus in fibroblasts with *mut*^{0} methylmalonic acidemia identified by polymerase-chain-reaction cDNA cloning. *Am. J. Hum. Genet.* **47**: 808-814.

Jezyk, P.F., Haskins, M.E., Patterson, D.F., Mellman, W.J. & Greenstein, M. (1977) Mucopolysaccharidosis in a cat with arylsulfatase B deficiency: a model of Maroteaux-Lamy syndrome. *Science* **198:** 834-836.

Jin, T., Kobayashi, T., Honke, K., Gasa, S. & Makita, A. (1992) Use of immunoadsorbent column chromatography for improved purification of arylsulfatase B from human placenta. *Biochem. Int.* **26**: 1025-1033.

Jin, W.-D., Desnick, R.J. & Schuchman, E.H. (1991) A common polymorphism in the human arylsulfatase B (ARSB) gene at 5q13-q14. *Nucleic Acids Res.* **19**: 4305.

Jin, W.-D., Jackson, C.E., Desnick, R.J. & 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.

Kadonaga, J.T., Jones, K.A. & Tjian, R. (1986) Promoter-specific activation of RNA polymerase II transcription by Sp1. *Trends Biochem. Sci.* **11**: 20-23.

Kawame, H., Maekawa, K. & Eto, Y. (1993) Molecular screening of Japanese patients with Gaucher disease - phenotypic variability in the same genotypes. *Human Mutation* **2**: 362-367.

Kazama, T., Takagi, M., Ishii, T. & Toda, Y. (1992) Immunoelectron microscopic studies of glycosaminoglycans in the metaphyseal bone trabeculae of growing rats. *Histochem. J.* **24**: 747-755.

King, P.V. & Blakesley, R. (1986) Optimizing DNA ligations for transformation. *FOCUS* **8(1)**: 1-3.

Kjellén, L. & Lindahl, U. (1991) Proteoglycans: structures and interactions. *Annu. Rev. Biochem.* **60**: 443-475.

Kobayashi, T., Honke, K., Jin, T., Gasa, S., Miyazaki, T. & Makita, A. (1992) Components and proteolytic processing sites of arylsulfatase B from human placenta. *Biochim. Biophys. Acta.* **1159**: 243-247.

Kolodny, E.H. (1989) Metachromatic leukodystrophy and multiple sulfatase deficiency: sulfatide lipidosis. In Scriver, C.R., Beaudet, A.L., Sly, W.S. & Valle, D. (eds.); *The Metabolic Basis of Inherited Disease.* McGraw-Hill, New York, pp. 1721-1750.

Kornfeld, S. (1986) Trafficking of lysosomal enzymes in normal and disease states. *J. Clin. Invest.* **77:** 1-6.

Kornfeld, S. (1992) Structure and function of the mannose 6-phosphate/insulin like growth factor II receptors. *Annu. Rev. Biochem.* **61**: 307-330.

Kornfeld, S. & Mellman, I. (1989) The biogenesis of lysosomes. *Annu. Rev. Cell. Biol.* **5:** 483-525.

Kozak, M. (1987) An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res.* **15**: 8125-8148.

Kramer, W. & Fritz, H.-J. (1987) Oligonucleotide-directed construction of mutations via gapped duplex DNA. *Methods Enzymol.* **154**: 350-367.

Krauss, J.C. (1992) Hematopoietic stem cell gene replacement therapy. *Biochim. Biophys. Acta* **1114**: 193-207.

Kresse, H., Cantz, M., von Figura, K., Glössl, J. & Paschke, E. (1981) The mucoploysaccharidoses: biochemistry and clinical symptoms. *Klin. Wochenschr.* **59**: 867-876.

Kretz, K.A., Cripe, D., Carson, G.S., Fukushima, H. & O'Brien, J.S. (1992) Structure and sequence of the human α-L-fucosidase gene and pseudogene. *Genomics* **12**: 276-280.

Kreysing, J., von Figura, K. & Gieselmann, V. (1990) Structure of the arylsulfatase A gene. *Eur. J. Biochem.* **191**: 627-631.

Krivit, W., Pierpont, M.E., Ayaz, K., Tsai, M., Ramsay, N.K.C., Kersey, J.H., Weisdorf, S., Sibley, R., Snover, D., McGovern, M.M., Schwartz, M.F. & Desnick, R.J. (1984) Bone-marrow transplantation in the Maroteaux-Lamy syndrome (mucopolysaccharidosis type VI): biochemical and clinical status 24 months after transplantation. *N. Engl. J. Med.* **311**: 1606-1611.

Krivit, W., Shapiro, E., Hoogerbrugge, P.M. & Moser, H.W. (1992) State of the art review: bone marrow transplantation treatment for storage diseases. *Bone Marrow Transpl.* **10**: 87-96.

Kunkel, T.A, Roberts, J.D. & Zakour, R.A. (1987) Rapid and efficient site-specific mutagenesis without phenotypic selection. *Methods Enzymol.* **154**: 367-382.

Kyle, J.W., Birkenmeier, E.H., Gwynn, B., Volger, C., Hoppe, P.C., Hoffmann, J.W. & Sly, W.S. (1990) Correction of murine mucopolysaccharidosis type VII by

a human β-glucuronidase transgene. *Proc. Natl. Acad. Sci. USA* **87:** 3914-3918.

Lau, M.M.H. & Neufeld, E.F. (1989) A frameshift mutation in a patient with Tay-Sachs disease causes premature termination and defective intracellular transport of the α-subunit of β-hexosaminidase. *J. Biol. Chem.* **264**: 21376-21380.

Lee, C.C., Wu, X., Gibbs, R.A., Cook, R.G., Muzny, D.M. & Caskey, C.T. (1988) Generation of cDNA probes directed by amino acid sequence: cloning of urate oxidase. *Science* **239**: 1288-1291.

Lee, G.D. & van Etten, R.L. (1975) Evidence for an essential histidine residue in rabbit liver aryl sulfatase A. *Arch. Biochem. Biophys.* **171**: 424-434.

Lehmann, L.E., Eberle, W., Krull, S., Prill, V., Schmidt, B., Sander, C., von Figura, K. & Peters, C. (1992) The internalization signal in the cytoplasmic tail of lysosomal acid phosphatase consists of the hexapeptide PGYRHV. *EMBO J.* **11**: 4391-4399.

Leinekugel, P., Michel, S., Conzelmann, E. & Sandhoff, K. (1992) Quantitative correlation between the residual activity of β-hexosaminidase A and arylsulfatase A and the severity of the resulting lysosomal storage disease. *Hum. Genet.* **88**: 513-523.

Lippincott-Schwartz, J., Bonifacino, J.S., Yuan, L.C. & Klausner, R.D. (1988) Degradation from the endoplasmic reticulum: disposing of newly synthesized proteins. *Cell* **54**: 209-220.

Litjens, T., Baker, E.G., Beckmann, K.R., Morris, C.P., Hopwood, J.J. & Callen, D.F. (1989) Chromosomal localization of *ARSB*, the gene for human N-acetylgalactosamine-4-sulphatase. *Hum. Genet.* **82**: 67-68.

Litjens, T., Morris, C.P., Gibson, G.J., Beckmann, K.R. & Hopwood, J.J. (1991) Human N-acetylgalactosamine-4-sulphatase: protein maturation and isolation of genomic clones. *Biochem. Int.* **24**: 209-215.

Litjens, T., Morris, C.P., Robertson, E.R., Peters, C., von Figura, K. & Hopwood, J.J. (1992) An N-acetylgalactosamine-4-sulfatase mutation (ΔG_{238}) results in a severe Maroteaux-Lamy phenotype. *Human mutation* **1**: 397-402.

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.

Lombardi, D., Soldati, T., Riederer, M.A., Goda, Y., Zerial, M. & Pfeffer, S.R. (1993) Rab9 functions in transport between late endosomes and the trans Golgi network. *EMBO. J.* **12**: 677-682.

Lüllmann-Rauch, R. (1989) Experimental mucopolysaccharidosis: preservation and ultrastructural visualization of intralysosomal glycosaminoglycans by use of the cationic dyes Cuprolinic Blue and Toluidine Blue. *Histochemistry* **93**: 149-154.

Luzio, J.P. & Banting, G. (1993) Eukaryotic membrane traffic: retrieval and retention mechanisms to achieve organelle residence. *Trends Biochem. Sci.* **18**: 395-398.

Maniatis, T., Fritsch, E.F. & Sambrook, J. (eds.) (1982) *Molecular cloning: a laboratory manual.* Cold Spring Harbour Laboratory Press, Cold Spring Harbour, New York.

Maroteaux, P. & Lamy, M. (1965) Hurler's disease, Morquio's disease, and related mucopolysaccharidoses. *J. Pediat.* **67**: 312-323.

Maroteaux, P., Lêveque, B., Marie, J. & Lamy, M. (1963) Une nouvelle dysostose avec élimination urinaire de chondroitine-sulfate B. *Press Méd.* **71**: 1849-1852.

Matalon, R., Arbogast, B. & Dorfman, A. (1974) Deficiency of chondroitin sulfate N-acetylgalactosamine 4-sulfate sulfatase in Maroteaux-Lamy syndrome. *Biochem. Biophys. Res. Comm.* **61**: 1450-1457.

Matsudaira, P. (1987) Sequence from picomole quantities of proteins electroblotted onto polyvinylidene difluoride membranes. *J. Biol. Chem.* **262**: 10035-10038.

McDowell, G.A., Cowan, T.M., Blitzer, M.G. & Greene, C.L. (1993) Intrafamilial variability in Hurler Syndrome and Sanfilippo syndrome type A: implications for evaluation of new therapies. *Am. J. Med. Gen.* **47**: 1092-1095.

McGovern, M.M., Mandell, N., Haskins, M. & Desnick, R.J. (1985) Animal model studies of allelism: characterization of arylsulfatase B mutations in homoallelic and heteroallelic (genetic compound) homozygotes with feline mucopolysaccharidosis VI. *Genetics* **110**: 733-749.

McGovern, M.M., Vine, D.T., Haskins, M.E. & Desnick, R.J. (1982) Purification and properties of feline and human arylsulfatase B isozymes: evidence for feline homodimeric and human monomeric structures. *J. Biol. Chem.* **257**: 12605-12610.

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., Kaplan, D., Wise, D., Hanley, W.B., Suddarth, S.B., Sevick, M.E. & Maumanee, A.E. (1965) The genetic mucopolysaccharidoses. *Medicine* **44:** 445-483.

Mellman, I., Fuchs, R. & Helenius, A. (1986) Acidification of the endocytic and exocytic pathways. *Annu. Rev. Biochem.* **55**: 663-700.

Melton, D.A., Krieg, P.A., Rebagliati, M.R., Maniatis, T., Zinn, K. & Green, M.R. (1984) Efficient *in vitro* synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. *Nucleic Acids Res* . **12**: 7035-7056.

Metcalf, P. & Fusek, M. (1993) Two crystal structures for cathepsin D: the lysosomal targeting signal and active site. *EMBO J.* **12**: 1293-1302.

Miura, N., Ohtsuka, E., Yamaberi, N., Ikehara, M., Uchida, T. & Okada, Y. (1985) Use of the deoxyinosine-containing probe to isolate and sequence cDNA encoding the fusion (F) glycoprotein of Sendai virus (HVJ). *Gene* **38**: 271-274.

Mizusawa, S., Nishimura, S. & Seela, F. (1986) Improvement of the dideoxy chain termination method of DNA sequencing by use of deoxy-7-deazaguanosine triphosphate in place of dGTP. *Nucleic Acids Res.* **14**: 1319-1324.

Modaressi, S., Rupp, K., von Figura, K. & Peters, C. (1993) Structure of the human arylsulphatase B gene. *Biol. Chem. Hoppe-Seyler* **374**: 327-335.

Morgan, R.A. & Anderson, W.F. (1993) Human gene therapy. *Annu. Rev. Biochem.* **62**: 191-217.

Morreau, H., Galjart, N.J., Willemsen, R., Gillemans, N., Zhou, X.Y. & d'Azzo, A. (1992) Human lysosomal protective protein: glycosylation, intracellular transport, and association with β-galactosidase in the endoplasmic reticulum. *J. Biol. Chem.* **267**: 17949-17956.

Moullier, P., Bohl, D., Heard, J.-M. & Danos, O. (1993) Correction of lysosomal storage in the liver and spleen of MPS VII mice by implantation of genetically modified skin fibroblasts. *Nature Genet.* **4**: 154-159.

Muenzer, J. (1986) Mucopolysaccharidoses. *Adv. Pediatr.* **33**: 269-302.

Muenzer, J., Neufeld, E.F., Constantopoulos, G., Caruso, R.C., Kaiser-Kupfer, M.I., Pikus, A., Danoff, J., Berry, R.R., McDonald, H.D., Thompson, J.N., Rodén, L. & Zasloff, M.A. (1992) Attempted enzyme replacement using human amnion membrane implantations in mucopolysaccharidoses. *J. Inher. Metab. Dis.* **15**: 25-37.

Murooka, Y., Ishibashi, K., Yasumoto, M., Sasaki, M., Sugino, H., Azakami, H. & Yamashita, M. (1990) A sulfur- and tyramine-regulated *Klebsiella aerogenes* operon containing the arylsulfatase (*atsA*) gene and the *atsB* gene. *J. Bacteriol.* **172**: 2131-2140.

Murphy, H.R., Kalman, M. & Cashel, M. (1992) Identification of the *gppB* locus as two convergent arylsulphatase-like genes, *aslA* and *aslB*, capable of suppressing a guanosine pentaphosphate phosphatase missense mutation in *Escherichia coli*. (unpublished GenBank entry ECOASLAB).

Murray, V. (1989) Improved double-stranded DNA sequencing using the linear polymerase chain reaction. *Nucleic Acids Res.* **17**: 8889.

Myers, R.M, Larin, Z. & Maniatis, T. (1985) Detection of single base substitutions by ribonuclease cleavage at mismatches in RNA:DNA duplexes. *Science* **230**: 1242-1246.

Myers, R.M., Sheffield, V.C. & Cox, D.R. (1989) Mutation detection by PCR, GC-clamps, and denaturing gradient gel electrophoresis. In Erlich, H.A. (ed.); *PCR technology: principles and applications for DNA amplification.* Stockton Press, New York, pp. 71-88.

Nakamaye, K.L. & Eckstein, F. (1986) Inhibition of restriction endonuclease *Nci*I cleavage by phosphorothioate groups and its application to oligonucleotide-directed mutagenesis. *Nucleic Acids Res.* **14**: 9679-9698.

Neote, K., Bapat, B., Dumbrille-Ross, A., Troxel, C., Schuster, S.M., Mahuran, D.J. & Gravel, R.A. (1988) Characterization of the human *HEXB* gene encoding lysosomal β-hexosaminidase. *Genomics* **3**: 279-286.

Neufeld, E.F. (1991) Lysosomal storage diseases. *Annu. Rev. Biochem.* **60**: 257-280.

Neufeld, E.F. & Muenzer, J. (1989) The mucopolysaccharidoses. In Scriver, C.R., Beaudet, A.L., Sly, W.S. & Valle, D. (eds.); *The Metabolic Basis of Inherited Disease.* McGraw-Hill, New York, pp. 1565-1587.

Nolan, C.M. & Sly, W.S. (1989) I-cell disease and pseudo-Hurler polydystrophy: disorders of lysosomal enzyme phosphorylation and localization. In Scriver, C.R., Beaudet, A.L., Sly, W.S. & Valle, D. (eds.); *The Metabolic Basis of Inherited Disease.* McGraw-Hill, New York, pp. 1589-1601.

O'Brien, J.F., Cantz, M. & Spranger, J. (1974) Maroteaux-Lamy disease (mucopolysaccharidosis VI), subtype A: deficiency of a N-acetylgalactosamine-4-sulfatase. *Biochem. Biophys. Res. Comm.* **60**: 1170-1177.

Ohshima, Y. & Gotoh, Y. (1987) Signals for the selection of a splice site in pre-mRNA: computer analysis of splice junction sequences and like sequences. *J. Mol. Biol.* **195**: 247-259.

Orita, M., Iwahana, H., Kanazawa, H., Hayashi, K., & Sekiya, T. (1989) Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms. *Proc. Natl. Acad. Sci. U.S.A.* **86**: 2766-2770.

Oshima, A., Kyle, J.W., Miller, R.D., Hoffmann, J.W., Powell, P.P., Grubb, J.H., Sly, W.S., Tropak, M., Guise, K.S. & Gravel, R.A. (1987) Cloning, sequencing, and expression of cDNA for human β-glucuronidase. *Proc. Natl. Acad. Sci. USA* **84**: 685-689.

Peters, C., Rommerskirch, W., Modaressi, S. & von Figura, K. (1991) Restoration of arylsulphatase B activity in human mucopolysaccharidosis-type-VI fibroblasts by retroviral-vector-mediated gene transfer. *Biochem. J.* **276:** 499-504.

Peters, C., Schmidt, B., Rommerskirch, W., Rupp, K., Zühlsdorf, M., Vingron, M., Meyer, H.E., Pohlmann, R. & von Figura, K. (1990) Phylogenetic conservation of arylsulfatases: cDNA cloning and expression of human arylsulfatase B. *J. Biol. Chem.* **265**: 3374-3381.

Pfeffer, S.R. (1988) Mannose 6-phosphate receptors and their role in targeting proteins to lysosomes. *J. Membrane Biol.* **103**: 7-16.

Proia, R.L. & Soravia, E. (1987) Organization of the gene encoding the human β-hexosaminidase α-chain. *J. Biol. Chem.* **262**: 5677-5681.

Quon, D.V.K., Proia, R.L., Fowler, A.V., Bleibaum, J. & Neufeld, E.F. (1989) Proteolytic processing of the β-subunit of the lysosome enzyme, β-hexosaminidase, in normal human fibroblasts. *J. Biol. Chem.* **264**: 3380-3384.

Raleigh, E.A., Lech, K. & Brent, R. (1993) Selected topics from classical bacterial genetics. In Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A. & Struhl, K. (eds.); *Current Protocols in Molecular Biology*. Wiley, New York, pp. 1.4.8-1.4.9.

Ratazzi, M.C., Marks, J.S. & Davidson, R.G. (1973) Electrophoresis of arylsulfatase from normal individuals and patients with metachromatic leukodystrophy. *Am. J. Hum. Genet.* **25**: 310-316.

Reed, K.C. & Mann, D.A. (1985) Rapid transfer of DNA from agarose gels to nylon membranes. *Nucleic Acids Res.* **13**: 7207-7221.

Robertson, D.A., Callen, D.F., Baker, E.G., Morris, C.P. & Hopwood, J.J. (1988a) Chromosomal localization of the gene for human glucosamine-6-sulphatase to 12q14. *Hum. Genet.* **79**: 175-178.

Robertson, D.A., Freeman, C., Morris, C.P. & Hopwood, J.J. (1992) A cDNA clone for human glucosamine-6-sulphatase reveals differences between arylsulphatases and non-arylsulphatases. *Biochem. J.* **288**: 539-544.

Robertson, D.A., Freeman, C., Nelson, P.V., Morris, C.P. & Hopwood, J.J. (1988b) Human glucosamine-6-sulfatase cDNA reveals homology with steroid sulfatase. *Biochem. Biophys. Res. Comm.* **157**: 218-224.

Rodén, 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.

Rommerskirch, W., Fluharty, A.L., Peters, C., von Figura, K. & Gieselmann, V. (1991) Restoration of arylsulphatase A activity in human-metachromatic-leucodystrophy fibroblasts via retroviral-vector-mediated gene transfer. *Biochem. J.* **280**: 459-461.

Rommerskirch, W. & von Figura, K. (1992) Multiple sulfatase deficiency: catalytically inactive sulfatases are expressed from retrovirally introduced sulfatase cDNAs. *Proc. Natl. Acad. Sci. USA* **89**: 2561-2565.

Roughley, P.J. & White, R.J. (1992) The dermatan sulfate proteoglycans of the adult human meniscus. *J. Orthop. Res.* **10**: 631-637.

Roy, A.B. (1953) The sulphatase of ox liver. I. the complex nature of the enzyme. *Biochem. J.* **53**: 12-15.

Roy, A.B. (1976) Sulphatases, lysosomes and disease. *Aust. J. Exp. Biol. Med. Sci.* **54**: 111-135.

Ruben, Z., Rorig, K.J. & Kacew, S. (1993) Perspectives on intracellular storage and transport of cationic-lipophilic Drugs. *Proc. Soc. Exp. Biol. Med.* **203**: 140-149.

Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B. & Erlich, H.A. (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239**: 487-491.

Sambrook, J., Fritsch, E.F. & Maniatis, T. (eds.) (1989) *Molecular cloning: a laboratory manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbour, New York.

Sanger, F., Nicklen, S. & Coulson, A.R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**: 5463-5467.

Sasaki, H., Yamada, K., Akasaka, K., Kawasaki, H., Suzuki, K., Saito, A., Sato, M. & Shimada, H. (1988) cDNA cloning, nucleotide sequence and expression of the gene for arylsulfatase in the sea urchin (*Hemicentrotus pulcherrimus*) embryo. *Eur. J. Biochem.* **177**: 9-13.

Sawadogo, M. & Van Dyke, M.W. (1991) A rapid method for the purification of deprotected oligodeoxynucleotides. *Nucleic Acids Res.* **19**: 674.

Schimmel, P. (1993) Functional analysis suggests unexpected role for conserved active-site residue in enzyme of known structure. *Proc. Natl. Acad. Sci. USA* **90**: 9235-9236.

Schuchman, E.H., Jackson, C.E. & Desnick, R.J. (1990) Human arylsulfatase B: MOPAC cloning, nucleotide sequence of a full-length cDNA, and regions of amino acid identity with arylsulfatases A and C. *Genomics* **6**: 149-158.

Schuler, G.D., Altschul, S.F. & Lipman, D.J. (1991) A workbench for multiple alignment construction and analysis. *Proteins: Structure, Function, and Genetics* **9**: 180-190.

Scott, H.S., Anson, D.S., Orsborn, A.M., Nelson, P.V., Clements, P.R., Morris, C.P. & Hopwood, J.J. (1991) Human α-L-iduronidase: cDNA isolation and expression. *Proc. Natl. Acad. Sci. USA* **88**: 9695-9699.

Scott, H.S., Litjens, T., Hopwood, J.J., & Morris, C.P. (1992) A common mutation for mucopolysaccharidosis type I associated with a severe Hurler syndrome phenotype. *Human Mutation* **1**: 103-108.

Scott, H.S., Nelson, P.V., Litjens, T., Hopwood, J.J & Morris, C.P. (1993) Multiple polymorphisms within the α-L-iduronidase gene (*IDUA*): implications for a role in modification of MPS-I disease phenotype. *Hum. Mol. Genet.* **2**: 1471-1473.

Scriver, C.R., Beaudet, A.L., Sly, W.S. & Valle, D. (1989) (eds.); *The Metabolic Basis of Inherited Disease.* 6th edition, McGraw-Hill, New York.

Selden, R.F. & Chory, J. (1987) Electrophoresis onto DEAE-cellulose paper. In Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A. & Struhl, K. (eds.); *Current Protocols in Molecular Biology*. Wiley, New York, pp. 2.6.5-2.6.6.

Shankaran, R., Ameen, M., Daniel, W.L., Davidson, R.G. & Chang, P.L. (1991) Characterization of arylsulphatase C isozymes from human liver and placenta. *Biochim. Biophys. Acta* **1078**: 251-257.

Shapiro, L.J. (1989) Steroid sulfatase deficiency and X-linked icthyosis. In Scriver, C.R., Beaudet, A.L., Sly, W.S. & Valle, D. (eds.); *The Metabolic Basis of Inherited Disease*. McGraw-Hill, New York, pp. 1945-1964.

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

Shi, Z., Byeon, I.-J. L., Jiang, R.-T. & Tsai, M.-D. (1993) Mechanism of adenylate kinase. What can be learned from a mutant enzyme with minor perturbation in kinetic parameters ? *Biochemistry* **32**: 6450-6458.

Shigematsu, Y., Chikahide, H., Nakai, A., Kuriyama, M., Kikawa, Y., Konishi, Y., Sudo, M. & Konishi, K. (1991) Mucopolysaccharidosis VI (Maroteaux-Lamy syndrome) with hearing impairment and pupillary membrane remnants. *Acta Paediatr. Jpn.* **33**: 476-481.

Shipley, J.M., Grubb, J.H. & Sly, W.S. (1993) The role of glycosylation and phosphorylation in the expression of active human β-glucuronidase. *J. Biol. Chem.* **268**: 12193-12198.

Sibille, A., Eng, C.M., Kim, S.-J., Pastores, G. & Grabowski, G.A. (1993) Phenotype/genotype correlations in Gaucher disease type I: clinical and therapeutic implications. *Am. J. Hum. Genet.* **52**: 1094-1101.

Southern, E.M. (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* **98**: 503-517.

Spranger, J.W., Koch, F., McKusick, V.A., Natzschka, J., Wiedemann, H.-R. & Zellweger, H. (1970) Mucopolysaccharidosis VI (Maroteaux-Lamy's disease). *Helv. Paediat. Acta* **25**: 337-362.

Staden, R. (1980) A new computer method for the storage and manipulation of DNA gel reading data. *Nucleic 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.

Steckel, F., Hasilik, A. & von Figura, K. (1983) Biosynthesis and maturation of arylsulfatase B in normal and mutant cultured human fibroblasts. *J. Biol. Chem.* **258**: 14322-14326.

Stein, C., Gieselmann, V., Kreysing, J., Schmidt, B., Pohlmann, R., Waheed, A., Meyer, H.E., O'Brien, J.S. & von Figura, K. (1989a) Cloning and expression of human arylsulfatase A. *J. Biol. Chem.* **264**: 1252-1259.

Stein, C., Hille, A., Seidel, J., Rijnbout, S., Waheed, A., Schmidt, B., Geuze, H. & von Figura, K. (1989b) Cloning and expression of human steroid sulfatase: membrane topology, glycosylation and subcellular distribution in BHK-21 cells. *J. Biol. Chem.* **264**: 13865-13872.

Stein, M., Zijderhand-Bleekemolen, J.E., Geuze, H., Hasilik, A. & von Figura, K. (1987) M_r 46 000 mannose 6-phosphate specific receptor: its role in targeting of lysosomal enzymes. *EMBO J.* **6**: 2677-2681.

Stumpf, D.A., Austin, J.H., Crocker, A.C. & LaFrance, M. (1973) Mucopolysaccharidosis type VI (Maroteaux-Lamy syndrome) I. Sulfatase B deficiency in tissues. *Am. J. Dis. Child.* **126**: 747-755.

Suzuki, Y. (1992) Intracellular turnover of mutant enzymes in lysosomal storage diseases. *Acta Histochem. Cytochem.* **25**: 559-562.

Tan, C.T.T., Schaff, H.V., Miller Jr, F.A., Edwards, W.D. & Karnes, P.S. (1992) Valvular heart disease in four patients with Maroteaux-Lamy syndrome. *Circulation* **85**: 188-195.

Tanaka, A., Ohno, K., Sandhoff, K., Maire, I., Kolodny, E.H., Brown, A. & Suzuki, K. (1990) GM2-gangliosidosis B1 variant: analysis of β-hexosaminidase α gene abnormalities in seven patients. *Am. J. Hum. Genet.* **46**: 329-339.

Taylor, H.R., Hollows, F.C., Hopwood, J.J. & Robertson, E.F. (1978) Report of a mucopolysaccharidosis occuring in Australian aborigines. *J. Med. Genet.* **15**: 455-461.

Taylor, J.A., Gibson, G.J., Brooks, D.A. & Hopwood, J.J. (1990) Human *N*-acetylgalactosamine-4-sulphatase biosynthesis and maturation in normal, Maroteaux-Lamy and multiple-sulphatase-deficient fibroblasts. *Biochem. J.* **268**: 379-386.

Taylor, J.W., Ott, J. & Eckstein, F. (1985) The rapid generation of oligonucleotide-directed mutations at high frequency using phosphorothioate-modified DNA. *Nucleic Acids Res.* **13**: 8764-8785.

Tollersrud, O.K. & Aronson, N.N. (1989) Purification and characterization of rat liver glycosylasparaginase. *Biochem. J.* **260**: 101-108.

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. & Orii, T. (1991) Morquio disease: isolation, characterization and expression of full-length cDNA for human N-acetylgalactosamine-6-sulfate sulfatase. *Biochem. Biophys. Res. Comm.* **181**: 677-683.

Tonnesen, T., Gregersen, H.N. & Güttler, F. (1991) Normal MPS excretion, but dermatan sulphaturia, combined with a mild Maroteaux-Lamy phenotype. *J. Med. Genet.* **28**: 499-501.

Trefz, F.K., Burgard, P., König, T., Goebel-Schreiner, B., Lichter-Konecki, U., Konecki, D., Schmidt, E., Schmidt, H. & Bickel, H. (1993) Genotype-phenotype correlations in phenylketonuria. *Clin. Chim. Acta* **217**: 15-21.

Tsuji, S., Choudary, P.V., Martin, B.M., Stubblefield, B.K., Mayor, J.A., Barranger, J.A. & Ginns, E.I. (1987) A mutation in the human glucocerebrosidase gene in neuronopathic Gaucher's disease. *N. Engl. J. Med.* **316:** 570-575.

Urlaub, G., Mitchell, P.J., Ciudad, C.J. & Chasin, L.A. (1989) Nonsense mutations in the dihydrofolate reductase gene affect RNA processing. *Mol. Cell. Biol.* **9**: 2868-2880.

Van Hoof, F. & Hers, H.G. (1964) L'ultrastructure des cellules hépatiques dans la maladie de Hürler (Gargoylisme). *Compt. Rend. Acad. Sci. Paris* **259:** 1281-1283.

von Figura, K. & Hasilik, A. (1986) Lysosomal enzymes and their receptors. *Annu. Rev. Biochem.* **55:** 167-193.

von Heijne, G. (1986) A new method for predicting signal sequence cleavage sites. *Nucleic Acids Res.* **14**: 4683-4690.

Voskoboeva, E., Isbrandt, D., von Figura, K., Krasnopolskaya, X. & Peters, C. (1994) Four novel mutant alleles of the arylsulfatase B gene in two patients with intermediate form of mucopolysaccharidosis VI (Maroteaux-Lamy syndrome). *Hum. Genet.* **93**: 259-264.

Wicker, G., Prill, V., Brooks, D., Gibson, G., Hopwood, J., von Figura, K. & 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.

Wilson, P.J. (1992) Molecular genetics of Hunter syndrome. Ph.D. thesis, University of Adelaide, South Australia.

Wilson, P.J., Meaney, C.A., Hopwood, J.J. & Morris, C.P. (1993) Sequence of the human iduronate 2-sulfatase (IDS) gene - short communication. *Genomics* **17**: 773-775.

Wilson, P.J., Morris, C.P., Anson, D.S., Occhiodoro, T., Bielicki, J., Clements, P.R. & Hopwood, J.J. (1990) Hunter Syndrome: isolation of an iduronate-2-sulfatase cDNA clone and analysis of patient DNA. *Proc. Natl. Acad. Sci. USA* **87**: 8531-8535.

Wisselaar, H.A., Kroos, M.A., Hermans, M.M.P., van Beeumen, J. & Reuser, A.J.J. (1993) Structural and functional changes of lysosomal acid α-glucosidase during intracellular transport and maturation. *J. Biol. Chem.* **268**: 2223-2231.

Wolfe, J.H., Sands, M.S., Barker, J.E., Gwynn, B., Rowe, L.B., Vogler, C.A. & Birkenmeier, E.H. (1992) Reversal of pathology in murine mucopolysaccharidosis type VII by somatic cell gene transfer. *Nature* **360**: 749-753.

Wood, W.I., Gitschier, J., Lasky, L.A. & Lawn, R.M. (1985) Base composition-independent hybridization in tetramethylammonium chloride: a method for oligonucleotide screening of highly complex gene libraries. *Proc. Natl. Acad. Sci. USA* **82**: 1585-1588.

Yamada, K., Akasaka, K. & Shimada, H. (1989) Structure of sea-urchin arylsulfatase gene. *Eur. J. Biochem.* **186**: 405-410.

Yanagishita, M. (1993) Function of proteoglycans in the extracellular matrix. *Acta Pathologica Japonica* **43**: 283-293.

Yang, Q., Angerer, L.M. & Angerer, R.C. (1989) Structure and tissue-specific developmental expression of a sea urchin arylsulfatase gene. *Dev. Biol.* **135**: 53-65.

Yanisch-Perron, C., Viera, J. & Messing, J. (1985) Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* **33**: 103-119.

Yen, P.H., Allen, E., Marsh, B., Mohandas, T., Wang, N., Taggart, R.T. & Shapiro, L.J. (1987) Cloning and expression of steroid sulfatase cDNA and the frequent occurence of deletions in STS deficiency: implications for X-Y interchange. *Cell* **49**: 443-454.

Yen, P.H., Marsh, B., Allen, E., Tsai, S.P., Ellison, J., Connoly, L., Neiswanger, K. & Shapiro, L.J. (1988) The human X-linked steroid sulfatase gene and a Y-encoded pseudogene: evidence for an inversion of the Y chromosome during primate evolution. *Cell* **55**: 1123-1135.

Yokota, I., Coates, P.M., Hale, D.E., Rinaldo, P. & Tanaka, K. (1991) Molecular survey of a prevalent mutation, ^{985}A-to-G transition, and identification of five infrequent mutations in the medium-chain acyl-CoA dehydrogenase (MCAD) gene in 55 patients with MCAD deficiency. *Am. J. Hum. Genet.* **49**: 1280-1291.

Yoshida, M., Ikadai, H., Maekawa, A., Takahashi, M. & Nagase, S. (1993b) Pathological characteristics of mucopolysaccharidosis VI in the rat. *J. Comp. Path.* **109**: 141-153.

Yoshida, M., Noguchi, J., Ikadai, H., Takahashi, M. & Nagase, S. (1993a) Arylsulfatase B-deficient mucopolysaccharidosis in rats. *J. Clin. Invest.* **91**: 1099-1104.

Young, R.A. & Davis, R.W. (1983) Efficient isolation of genes by using antibody probes. *Proc. Natl. Acad. Sci. USA* **80**: 1194-1198.

Zlotogora, J. (1987) Intrafamilial variability in lysosomal storage disorders. *Am. J. Med. Gen.* **27**: 633-638.

PUBLICATIONS RESULTING FROM THIS THESIS

Litjens, T., Baker, E.G., Beckmann, K.R., Morris, C.P., Hopwood, J.J. & Callen, D.F. (1989) Chromosomal localization of *ARSB*, the gene for human N-acetylgalactosamine-4-sulphatase. *Hum. Genet.* **82**: 67-68.

Litjens, T., Morris, C.P., Gibson, G.J., Beckmann, K.R. & Hopwood, J.J. (1991) Human N-acetylgalactosamine-4-sulphatase: protein maturation and isolation of genomic clones. *Biochem. Int.* **24**: 209-215.

Litjens, T., Morris, C.P., Robertson, E.R., Peters, C., von Figura, K. & Hopwood, J.J. (1992) An N-acetylgalactosamine-4-sulfatase mutation (ΔG_{238}) results in a severe Maroteaux-Lamy phenotype. *Human mutation* **1**: 397-402.

PUBLICATIONS RESULTING FROM COLLABORATION

The mutation identification techniques described in Chapter 4 were also utilized to identify MPS-I patient mutations, while employed as a research assistant;

Brooks, D.B, Robertson, D.A., Bindloss, C., Litjens, T., Anson, D.A., Peters, C., Morris, C.P. & Hopwood, J.J. (1994) Mucopolysaccharidosis VI (Maroteaux-Lamy syndrome): two site directed mutations C91S and C91T, abrogate enzyme activity but have different effects on N-acetylgalactosamine 4-sulfatase protein levels. *J. Biol. Chem.*, submitted.

Hopwood, J.J., Vellodi, A., Scott, H.S., Morris, C.P., Litjens, T., Clements, P.R., Brooks, D.A., Cooper, A. & Wraith, J.E. (1993) Long-term clinical progress in bone marrow transplanted mucopolysaccharidosis type I patients with a defined genotype. *J. Inher. Metab. Dis.* **16**: 1024-1033.

Scott, H.S., Litjens, T., Hopwood, J.J., & Morris, C.P. (1992) 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. & Morris, C.P. (1992) PCR detection of two RFLP's in exon I of the α-L-iduronidase (IDUA) gene. *Human Genet.* **90**: 327.

Scott, H.S., Litjens, T., Nelson, P.V., Brooks, D.A., Hopwood, J.J., & Morris, C.P. (1992) α-L-Iduronidase mutations ($Q_{70}X$ and $P_{533}R$) associate with a severe Hurler syndrome phenotype. *Human Mutation* **1**: 333-339.

Scott, H.S., Litjens, T., Nelson, P.V., Thompson, P.R., Brooks, D.A., Hopwood, J.J., & Morris, C.P. (1993) Identification of mutations in the α-L-Iduronidase gene (*IDUA*) that cause Hurler and Scheie syndromes. *Am. J. Hum. Genet.* **53**: 975-986.

Scott, H.S., Nelson, P.V., Litjens, T., Hopwood, J.J & Morris, C.P. (1993) Multiple polymorphisms within the α-L-iduronidase gene (*IDUA*): implications for a role in modification of MPS-I disease phenotype. *Hum. Mol. Genet.* **2**: 1471-1473.