PATHOPHYSIOLOGY OF ACUTE INTERMITTENT PORPHYRIA

PATRICIA MEI LIN ONG

Thesis submitted to
The Faculty of Medicine, University of Glasgow
For the degree of
Doctor of Philosophy (Ph.D.)

The Duncan Guthrie Institute of Medical Genetics
University of Glasgow

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

DECLARATION

I certify that this thesis does not contain material previously published or written by any other person except where referred to in the text and that the results in this thesis have not been submitted for any other degree or diploma.

PATRICIA M.L. ONG

In ever loving memory of my beloved grandfather,

Mr. Fong Peck Yew

(3 July 1903 - 24 March 1994)

THE TREASURY OF TRUTH (DHAMMAPADA)

Chapter 20: Magga Vagga (The Path)
Verse 280: The Slothful Miss The Path

Utthānakālamhi anutthahānō
yuvā bali alasiyam upetō
Samsanna sankappamanō kusitō
pannāya maggam alasō na vindati.
(20:8)

Though time to strive, not striving,
while young and strong yet indolent,
weak-minded and irresolute:
one finds not wisdom's way.

Unenterprising, youthful but lazy, irresolute and weak, fail in their way to wisdom.

ACKNOWLEDGEMENTS

I would like to thank the following people who have helped me one way or another during the course:

Professor J.M. Connor for accepting me into the Ph.D. programme to allow me to develop and pursue my research interests, and also for his helpful comments, guidance and encouragement throughout my period of study;

Dr. George Lanyon for his guidance, supervision, patience and understanding, and for giving me the opportunity to develop my teaching skills;

Professor Michael Moore for coordinating this project with our South African counterparts, Drs. Richard Hift and Janet Halkett, who provided patient samples;

Dr. Charles Mgone who introduced the various technologies and patiently guided me with his expertise;

Mrs. Elizabeth O'Hare for oligonucleotide synthesis, Miss Ann Ross who always ensured that the orders arrived promptly, and Mr. Shaban Javed for his friendship and always having a listening ear;

Dr. David Aitken and Mr. Gordon Graham for their expertise, guidance and helpful discussions in the various protein biochemistry technologies;

To Dr. Carol Chu, Lulwa Al-Jeryan and Eunice Stefanou for their everlasting friendship;
Everyone in the CMG II and Dev. Gen. Labs;
All the clinicians for their collection of samples, and the patients for their willingness to participate in this study;
To all other staff members of DGI for accommodating me in one way or another;

To Grandma, Auntie Vivien and Mum for their support, trust and understanding throughout;
To Auntie Gloria, Uncle Chris, Reuben and Joshua for their concern;
To Robert, Sabrina, Bernard, Serene and baby Carolyn for always being there;

Last, but not least, to the University of Glasgow for a Postgraduate Research Scholarship.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

A	adenine
Å	angstrom
ADP	ALA dehydratase deficiency porphyria
AIP	acute intermittent porphyria
ALA	aminolaevulinic acid
APS	ammonium persulphate
AUFS	absorbance units full scale
bp	base pair
BPS	branch-point sequence
BSA	bovine serum albumin
C	cytosine
°C	degrees Celsius
CCM	chemical cleavage of mismatches
cDNA	complementary deoxyribonucleic acid
CEP	congenital erythropoietic porphyria
Ci	Curie (3.7×10^{10} Becquerels)
cm	centimetre
cpm	counts per minute
CRIM	cross-reacting immunological material
CV	consensus value
ddH_2O	double distilled water
ddNTP	dideoxyribonucleoside triphosphate
DEPC	diethyl pyrocarbonate
DGGE	denaturing gradient gel electrophoresis
dH_2O	deionised water
DMD	Duchenne muscular dystrophy
DNA	deoxyribonucleic acid
dNTP	2'-deoxyribonucleoside triphosphate
DRPLA	dentatorubral pallidoluysian atrophy
DTT	dithiothreitol
E. coli	*Escherichia coli*
EDTA	ethylenediamine tetraacetic acid
EPP	erythropoietic protoporphyria
G	guanine
g	gram
HA	heteroduplex analysis
HCP	hereditary coproporphyria
HEP	hepatoerythropoietic porphyria
HMB-S	hydroxymethylbilane synthase
HPLC	high performance liquid chromatography
hr	hour
IEF	isoelectric focusing
Ig	immunoglobulin

IPTG	isopropylthio-β-D-galactopyranoside
kb	kilobase
kD	kilodalton
l	litre
LC	liquid chromatography
M	molar
μg	microgram
μl	microlitre
μM	micromolar
μm	micrometre
5'-mC	5'-methylcytosine
mg	milligram
ml	millilitre
mM	millimolar
mm	millimetre
MMLV	Moloney Murine Leukaemia Virus
MOPS	3-(N-morpholino) propanesulfonic acid
mRNA	messenger ribonucleic acid
ng	nanogram
nm	nanometre
OD	optical density
PBG	porphobilinogen
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PCT	porphyria cutanea tarda
pI	isoelectric point
pmol	picomole
pM	picomolar
pol II	polymerase II
PTT	protein truncation test
RFLP	restriction fragment length polymorphism
RNA	ribonucleic acid
RNase	ribonuclease
rpm	revolutions per minute
rRNA	ribosomal ribonucleic acid
RT-PCR	reverse transcriptase PCR
SDS	sodium dodecyl sulphate
SDS-PAGE	SDS polyacrylamide gel electrophoresis
SDW	sterile distilled water
snRNA	small nuclear ribonucleic acid
snRNP	small nuclear ribonucleoprotein
ss	splice site
SSCP	single strand conformation polymorphism
T	thymidine
T_m	melting temperature

TEMED	N,N,N,N-tetramethylethylenediamine
TGGE	temperature gradient gel electrophoresis
Tris	tris (hydroxymethyl) aminomethane
tRNA	transfer ribonucleic acid
UTR	untranslated region
UV	ultraviolet
V	volt
VP	variegate porphyria
W	watt

LIST OF PUBLICATIONS AND PRESENTATIONS

PUBLICATIONS

1. Patricia M.L. Ong, W. George Lanyon, Richard J. Hift, Janet Halkett, Michael R. Moore, Charles S. Mgone & J. Michael Connor (1996). Detection of four mutations in six unrelated South African patients with acute intermittent porphyria. *Molecular and Cellular Probes* **10**, 57-61.

SPOKEN PRESENTATIONS

1. **British Human Genetics Conference**- York; 16 - 18 September 1996
   Alternative splicing of hydroxymethylbilane synthase mRNA excludes exons 3 and 12.
   Ong, Patricia; Lanyon, WG; Connor, JM
   Abstract: *Journal of Medical Genetics* (1996) vol. 33 supplement 1 pS2 (No. SP7)

2. **British Medical Genetics Conference**- York; 11 - 13 September 1995
   Detection of mutations in the porphobilinogen deaminase gene: a new occurrence involving two mutations on a single chromosome.
   P Ong, G Lanyon, R Hift, J Halkett, M Moore, J Connor

3. **British Medical Genetics Conference**- York; 12 - 14 September 1994
   Identification of Three Mutations in the Porphobilinogen Deaminase Gene in Four South African Patients with Acute Intermittent Porphyria.
   PML Ong, WG Lanyon, MR Moore, JM Connor
   Abstract: *Journal of Medical Genetics* (1995) vol. 32 no. 2 p141.

POSTER PRESENTATIONS

1. **European Society of Human Genetics Conference-** London; 11 - 13 April 1996
   Exon 8 skipping due to a novel acceptor splice site mutation in the hydroxymethylbilane synthase gene.
   Ong, Patricia; Lanyon, WG; Cramp, CE; Connor, JM.
   Abstract: *European Journal of Human Genetics* (1996) vol. 4 supplement 1 p95 (No. 5.200)

2. **Human Genome Meeting-** Heidelberg, Germany; 22 - 24 March 1996
   The spectrum of mutations detected in the hydroxymethylbilane synthase gene in seven South African patients.
   P.M.L. Ong, W.G. Lanyon, R.J. Hift, J. Halkett, M.R. Moore, J.M. Connor

3. **International Symposium on Porphyrias and Heme Related Disorders: Molecular Basis, Diagnostic and Clinical Aspects-** Helsinki, Finland; 28 June - 2 July 1995
   Detection of Three Mutations in Four Patients with Acute Intermittent Porphyria.
   Ong P.M.L., Lanyon W.G., Hift R., Halkett J., Moore M.R., Louie G.V., Paterson J.R., Wilcox D.E., Connor J.M.
   Abstract: *The Scandinavian Journal of Clinical and Laboratory Investigation* (1995) vol. 55 supplement 223

4. **Tetrapyrrole Discussion Group Meeting-** Queen Mary and Westfield College, University of London; 11 - 12 July 1994
   Identification of Three Mutations in the Porphobilinogen Deaminase Gene in Four South African Patients with Acute Intermittent Porphyria.
   PML Ong, WG Lanyon, MR Moore, JM Connor

Note: Some of the data in this study have been presented at the various conferences and published in the journal as listed above. The abstracts and reprint of the published work are bound at the back of this thesis.

SUMMARY

Acute intermittent porphyria (AIP) is an inborn error of metabolism which is caused by deleterious mutations in the hydroxymethylbilane synthase (*HMB-S*) gene. The overall aims of the present study were to search for and characterise mutations in the *HMB-S* gene and to develop an *in vitro* expression system to confirm the pathogenic nature of these mutations.

Thirty individuals from fifteen families with AIP from South Africa, Scotland, England and the Republic of Ireland were studied. Initially, the entire coding sequence of the *HMB-S* gene was screened using heteroduplex analysis (HA) and chemical cleavage mismatch (CCM) analysis. Any differences in the band patterns, when compared to the negative control in both techniques, were characterised by direct sequencing. The same region of the genomic DNA was also sequenced. This step was especially important in characterising splice site mutations responsible for exon skipping. Using the above strategies, six missense mutations, R22C, R26C, R26H, R116W, R173Q and L177R, one splice site mutation (345-1G→A) leading to the skipping of exon 8 and two frameshift mutations (771insT and $delG_{1002/1003/1004}$) were identified. Four of these mutations, R22C, 345-1G→A, 771insT and $delG_{1002/1003/1004}$, are novel. All the missense mutations resulted in the substitution of highly conserved residues, hence are considered to be potentially pathogenic. In addition, alternative splicing of exons 3 and 12, resulting in their exclusion from a proportion of mRNA transcripts, was identified in thirty AIP patients and forty-four healthy individuals.

The R22C mutation (R7 in the *E. coli* model) in exon 3 affects the highly conserved Arg22 residue and is caused by a C to T substitution at nucleotide 64. Arginine, a basic residue, is substituted by cysteine, a neutral polar amino acid and this is predicted to destabilise the protein. This mutation, initially detected in two Scottish siblings, was subsequently identified in three other family members by employing a simple restriction enzyme analysis using *Hinf* I. R26C (R11 in *E. coli*) is the result of a C to T transition in exon 3 at position 76. The highly conserved Arg26 residue in twelve species is substituted by the polar neutral cysteine residue. This mutation was present in two Irish siblings and was identified in three further relatives by performing a simple *Aci* I restriction enzyme

analysis. R26H is caused by a G to A transition of the adjacent base following that causing R26C and was confirmed by an *Aci* I digestion. R116W (R101 in *E. coli*) in exon 8 was identified in three unrelated South African patients and raised the possibility of these patients having a common Dutch ancestry, due to the high prevalence of this mutation in the Dutch population. R116W is a consequence of a C to T transition at nucleotide 346 which was confirmed by an *Aci* I digestion. A destabilisation of the protein is expected with the substitution of the basic amino acid arginine with tryptophan, a neutral non-polar amino acid. The mutation in exon 10, R173Q (R155 in *E. coli*), is caused by a G to A transition at position 518 and a loss of an *Nci* I recognition site indicates a base change. In this case, the basic residue arginine is substituted with glutamine, a polar neutral residue. This replacement inhibits substrate binding and chain elongation. L177R (L159 in *E. coli*), in exon 10, is the result of a T to G transversion at nucleotide 530, causing the non-polar neutral residue leucine to be replaced by the basic amino acid arginine. This introduces a destabilising effect in the mature protein and impairs its normal function. The *Aci* I restriction enzyme was employed to identify this mutation in five members of a Scottish family.

345-1G→A is a 3' acceptor splice site mutation at position -1 of intron 7 which causes the skipping of exon 8 in three related individuals from England. 26 amino acids are predicted to be absent from the protein product but the translational reading frame is not altered. However, the codon at the junctions of exons 7 and 8 of the normal mRNA is now altered and encodes for asparagine instead of lysine. The application of *Eco* RII confirmed this mutation.

771insT was identified in exon 12 of a South African patient which results in a shift in the reading frame and the alteration of the thirty-two amino acids following this insertion before reaching a premature termination of translation at codon 290. This mutation affected a *Sty* I recognition site, which was used for confirmation. A protein product containing 289 amino acids is expected instead of the 361 amino acids normally found in a healthy individual, equivalent to the loss of approximately 20% of the protein. The mutation, $delG_{1002/1003/1004}$, also results in a termination codon, this time at codon 343, and

consequently a protein containing 342 amino acids is expected. This corresponds to a loss of approximately 5% of the mature protein.

Five of the missense mutations, R22C, R26C, R26H, R116W and R173Q, occur at the hypermutable CpG dinucleotide, a hotspot for mutations. The normal occurrence of alternative splicing of exons 3 and 12 had not been previously reported. Both exons are spliced out from the mRNA transcripts to various extents in thirty AIP patients and forty-four healthy individuals. A study of the frequency of occurrence of each base at each position of the 5' and 3' splice junctions of the *HMB-S* gene, and an analysis of the consensus values (CV) at both splice sites, according to the method of Shapiro and Senapathy (1987), did not reveal any irregularities.

The sensitivity and efficiency of the two screening techniques, HA and CCM, were compared. HA, although relatively simple and rapid, does not involve hazardous substances and was not able to detect one of the mutations identified in this study, thus giving an efficiency of about 90%. On the other hand, while CCM involves potentially hazardous materials and is time-consuming, it was able to detect all the mutations including the one missed by HA. Therefore, CCM was regarded as a better screening method for use in this project.

The second objective of this project was to develop an *in vitro* gene expression system. The pTrc 99A (Pharmacia) plasmid expression vector was used for expression studies and the accompanied protocol was followed. The appropriate fragments of interest were ligated to the vector and subcloned in the bacterial strain, JM109, of *E. coli*. The cloned inserts were then screened, by CCM, for extraneous misincorporations by Taq DNA polymerase and sequenced to confirm their integrity. Of the 126 clones screened, 44 (34.9%) were found to have maintained their integrity and were suitable for expression studies. 5 of these, each harbouring a mutant allele, were subjected to *in vitro* expression studies. The clones with the mutations R26H, R116W and R173Q, were introduced into the expression system to determine their pathophysiology in AIP. The average specific activities of the three mutations, as determined by the *in vitro* expression system, were 0.83 X 10^6 pmol/mg protein/hr for R173Q, 0.93 X 10^6 pmol/mg protein/hr for R116W and 1.07 X 10^6 pmol/mg protein/hr for R26H.

CHAPTER 1:

INTRODUCTION

CHAPTER 1: INTRODUCTION

1.1 THE SPECTRUM OF MUTATIONS IN GENETIC DISEASE

The gene is the unit of heredity. It is composed of deoxyribonucleic acid (DNA), which contains genetic information that determines the order of amino acids in a peptide chain, and hence the three-dimensional structure of the protein which it encodes. Enzymes and other biologically important proteins consist of one or more peptide chains folded into a three-dimensional structure, the exact shape of which is crucial for their normal function.

There are various types of mutations in the human genome. The study of naturally occurring gene mutations is important in the understanding of the origins of genetic variation and the mechanisms of evolution. The various types of mutations (single base-pair substitutions, deletions, insertions, inversions and duplications) have been detected and characterised in many human genes, thus enabling the study of the underlying mutational mechanisms.

1.1.1 SINGLE BASE-PAIR SUBSTITUTIONS

Single base-pair substitutions may affect the transcription of the gene into messenger ribonucleic acid (mRNA), the translation of the mRNA into protein, the processing events following transcription or the accuracy of splicing. Point mutations can be divided into three groups, viz. missense mutations where the single nucleotide alteration results in the substitution of one amino acid for another, nonsense mutations where the substitution of a single nucleotide results in a premature termination codon and the formation of a truncated protein product, and sense mutations which change a termination codon to one that codes for amino acids and result in the translation of a protein beyond the normal termination signal until another termination codon is reached (Weatherall, 1991). Two variant forms of single base substitutions exist which have no apparent effect on the protein product: (i) the amino acid is not changed in the mature protein- they are known as silent mutations; and (ii) the altered amino acid is from the same class as the normal amino acid (i.e.

neutral/basic/acidic, hydrophilic/hydrophobic) but the activity of the protein is not affected and the phenotypic effects are insignificant- these are known as neutral substitutions or conservative changes.

Missense mutations can be regarded as disease-causing from one or more of the following sources: (1) occurrence of the mutation in a region of known functional importance; (2) a conserved residue is affected; (3) an unrelated patient has the same mutation; (4) the absence of the mutation in a large number of healthy individuals; (5) the inheritance of the mutation and disease phenotype in a family; (6) demonstration that the mutant protein produced *in vitro* possesses similar biochemical properties and characteristics as its *in vivo* counterpart; and (7) reversal of the pathological phenotype in the patient/cultured cells by replacement of the mutant gene/protein with its wild-type counterpart.

CpG dinucleotides have long been recognised to be hotspots for mutations. In eukaryotic genomes, 5'-methylcytosine (5'-mC) occurs predominantly in CpG dinucleotides, the majority of which seem to be methylated (Cooper, 1983). Methylation of cytosine results in a high level of mutations due to the tendency of 5'-mC to undergo deamination to form thymine (Figure 1.1), which is not recognised by the DNA repair mechanisms and thus results in mutation (Cooper and Krawczak, 1991).

A survey of those single base-pair mutations responsible for human genetic disease, showed that 35% occurred within CpG dinucleotides (Cooper and Youssoufian, 1988), with C→T or G→A transitions occurring with a frequency of over 90%.

Single base-pair substitutions affecting mRNA splicing are non-randomly distributed, which can be related to the phenotypic consequences of mutation (Krawczak et al, 1992). Three types of point mutations exist that affect mRNA splicing: (i) mutations within the 5' or 3' consensus splice sites which usually reduce the amount of correctly spliced mature mRNA and/or lead to the utilisation of alternative splice sites in the vicinity. The resulting mRNAs either lack a portion of the coding sequence (exon skipping) or contain additional sequence of intronic origin (cryptic splice site utilisation); (ii) mutations within an intron or exon that may serve to activate cryptic splice sites and lead to the production of aberrant mRNA species; and (iii) mutations within a branch-point sequence (BPS).

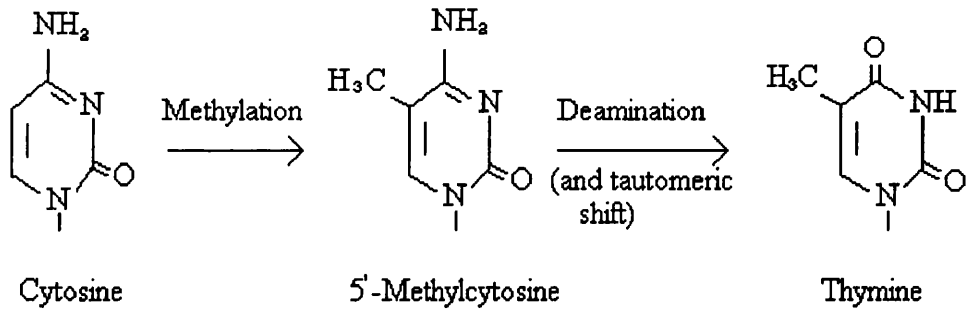


Figure 1.1: Schematic representation of the molecules for cytosine, 5'-methylcytosine and thymine and the chemical events for the transformation of cytosine to thymine.

1.1.2 DELETIONS

Gene deletions have been reported to be responsible for at least 159 different human disorders (Cooper and Krawczak, 1993) and can be classified according to the length of DNA deleted, either one or a few base-pairs, or several hundred kilobases. Gross gene deletions may arise either by (i) homologous unequal recombination mediated by related gene sequences or repetitive sequence elements; or (ii) non-homologous recombination involving DNA with minimal sequence homology. The former involves cleavage and rejoining of non-sister chromatids at homologous but non-allelic DNA sequences, resulting in fusion genes if the recombination breakpoints are intragenic. The latter, however, involves non-homologous or illegitimate recombination between two sites which show little or minimal sequence homology. Short gene deletions (<20bp) have been studied by Krawczak and Cooper (1991) to understand the mechanisms involved. Direct repeats, palindromes and symmetric elements (e.g. CTGAAGTC, GGACAGG) could be potential mechanisms involved in gene deletions.

1.1.3 INSERTIONS

Insertions, like deletions, can involve either one or a few bases, or several hundred kilobases. The largest fragment inserted into a gene is 220kb in the Duchenne muscular dystrophy (*DMD*) gene (Bettecken and Müller, 1989). Studies by Cooper and Krawczak (1991) using short insertion-type mutations (<10bp) revealed that such a mutation is a non-random process which is dependent on the local DNA sequence environment. The majority of insertional mutations can be explained by (i) slipped mispairing; (ii) inverted repeats; or (iii) symmetric elements.

1.1.4 INVERSIONS

Inversions are an extremely rare form of gene mutation. The β-globin gene cluster in Indian $^{A}\gamma\delta\beta$-thalassemia involves complex rearrangement (Jennings et al, 1985). Two fragments, 0.83kb and 7.46kb, were deleted and the intervening segment was simultaneously inverted and reintroduced between the $^{A}\gamma$ and δ gene loci. Jennings et al

proposed that this event could be possible by the folding pattern of the chromatin in the cluster region bringing the $^{A}\gamma$ gene into close proximity with the δ- and β-globin genes. This rearrangement enhanced the expression of the upstream fetal $^{G}\gamma$ gene.

1.1.5 DUPLICATIONS

The duplication of either whole genes or their constituent exons have an important role in the evolution of the mammalian genome. However, they may also result in disease. Duplication of several amino acids may alter the protein conformation, resulting in an unstable protein, or may result in a frameshift in the mRNA producing a defective protein. The largest duplication reported to date is a 400kb internal duplication of the *DMD* gene involving exons 13 to 42 (Angelini et al, 1990), but the patient manifested the relatively mild Becker form of muscular dystrophy. Gene duplications may arise due to homologous or non-homologous recombination. For example, homologous recombination may occur between repetitive sequences such as the *Alu* elements. Non-homologous recombination between non-homologous sequences may result in duplication by random chromatid cleavage and rejoining events.

1.1.6 EXPANSION OF UNSTABLE REPEAT SEQUENCES

The expansion of unstable trinucleotide repeat sequences is one of the recently recognised mechanisms of mutagenesis. Examples of the disorders known to be due to the expansion of such repeats are Fragile X (Fu et al, 1991), spinobulbar muscular atrophy (La Spada et al, 1991), myotonic dystrophy (Fu et al, 1992), Huntington's disease (The Huntington's Disease Collaborative Research Group, 1993), spinocerebellar ataxia I (Orr et al, 1993), dentatorubral-pallidoluysian atrophy (DRPLA) (Koide et al, 1994), Machado-Joseph disease (Maciel et al, 1995) and Haw River syndrome (Burke et al, 1994), the latter of which has been shown to be caused by an expanded repeat in the same allele as that for DRPLA. Patients suffering from any one of these diseases manifest the condition when the size of the repeats expand beyond the normal size range.

1.2 INBORN ERRORS OF METABOLISM- DISORDERS OF PORPHYRIN METABOLISM

Inborn errors of metabolism are metabolic diseases in which the characteristic classical pathological and biochemical abnormalities can be attributed to a deficient or defective gene product (protein) which, in turn, is due to the presence of a particular abnormal gene.

1.2.1 THE PORPHYRIAS

The porphyrias are a group of rare and complex metabolic disorders which affect the heme biosynthetic pathway. They may be inherited or acquired, and each results in a reduction in the activities of the respective enzymes in the pathway. Consequently, there is an excessive production of porphyrins or their precursors proximal to the block (Brodie et al, 1977), which can be detected in the urine, faeces or blood (Kappas et al, 1989) and results in the various forms of porphyria.

1.2.2 PORPHYRINS AND THE HEME BIOSYNTHETIC PATHWAY

Porphyrins are cyclic tetrapyrroles which are required for the formation of the essential heme-containing compounds such as cytochromes, catalase and peroxidase. Therefore, porphyrins are required as building blocks for heme biosynthesis. The porphyrins formed in the heme biosynthetic pathway are in the reduced forms of porphyrins (i.e. porphyrinogens), which are rapidly oxidised by air to form porphyrins. With the exception of the terminal step of the pathway involving the synthesis of heme from protoporphyrin and catalysed by ferrochelatase, all other enzymatic steps require porphyrinogens as substrates.

Heme is produced by a chemical pathway which consists of eight steps, each of which is enzymatically controlled (Figures 1.2a and 1.2b). Eight molecules of glycine and eight molecules of succinyl CoA are the initial substrates required to synthesise one molecule of heme. The first and last three enzymes of the pathway are localised in the mitochondria, while the intermediate enzymes are localised in the cytosol. The rate-limiting step is at the level of 5-aminolaevulinic acid (ALA) synthase. A negative feedback exists whereby heme

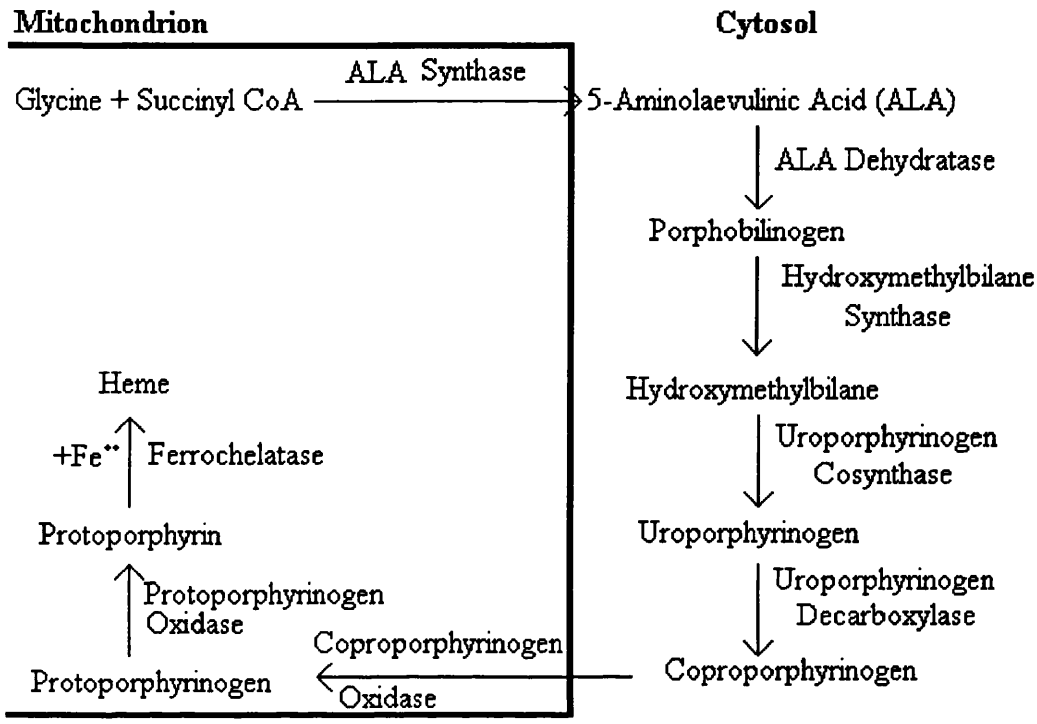


Figure 1.2a: The heme biosynthetic pathway.

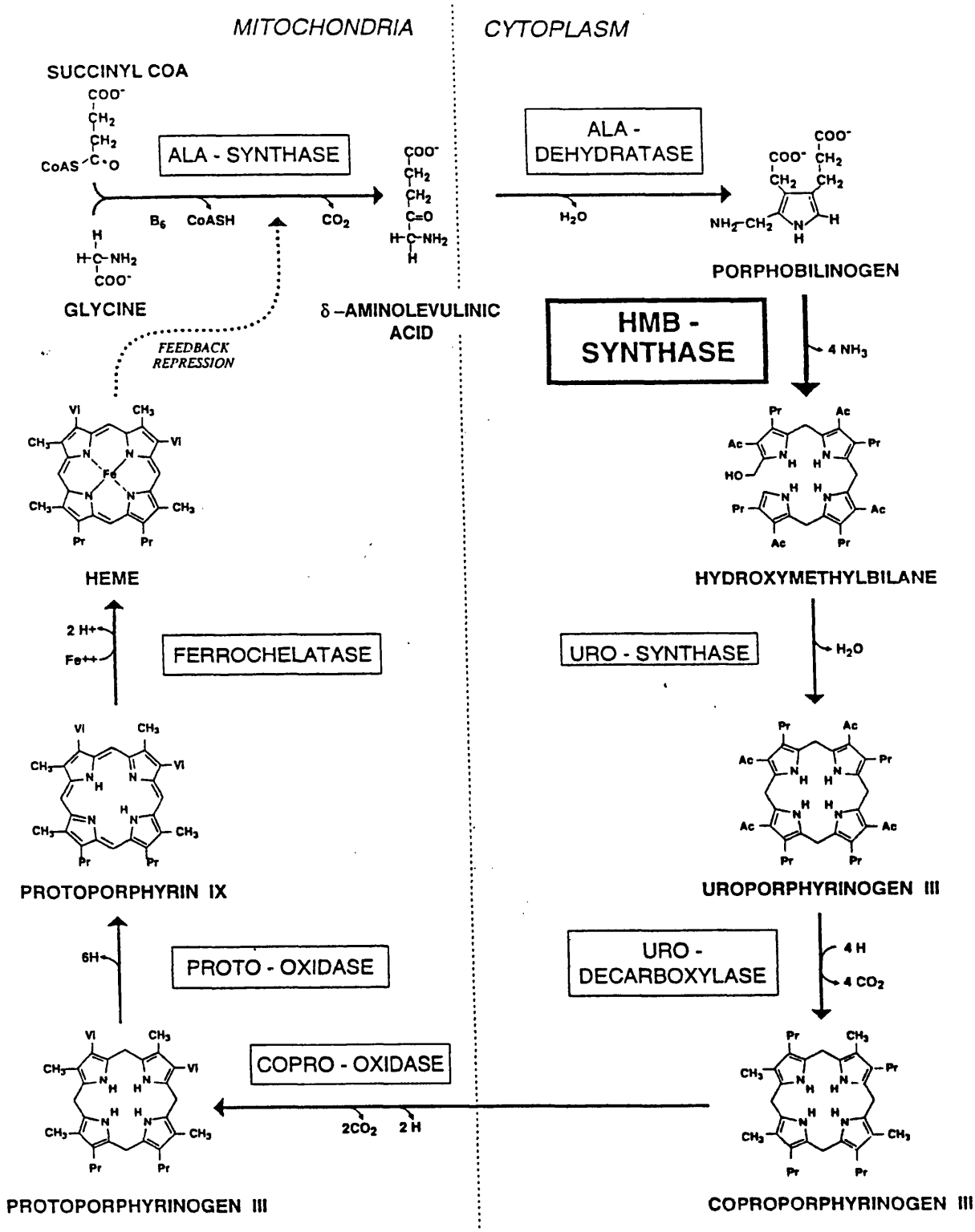


Figure 1.2b: Diagrammatic representation of the chemical events occurring at each step of the heme biosynthetic pathway. Reproduced from Astrin and Desnick (1994).

acts as a negative regulator of the accumulation of ALA synthase. This means that the rate of accumulation of ALA synthase increases greatly in the absence of heme and is diminished in its presence. However, the rate-limiting feature of this enzyme occurs only in the liver and not in the erythroid cells. The erythron increases heme synthesis when there in an increase in cell number.

Two isoforms of ALA synthase and hydroxymethylbilane synthase (HMB-S) exist, one of which is expressed only in the erythrocytes and the other is expressed in all the other cells. Both isoforms are derived from a single gene. The transcripts of all other genes encoding the remaining enzymes in the heme biosynthetic pathway are identical both in the erythroid and non-erythroid cells.

1.2.3 CLASSIFICATION OF THE PORPHYRIAS

The porphyrias are classified as either hepatic or erythroid, depending on the site of expression of the enzymatic defect in each disorder (Table 1.1). The cardinal feature of the erythroid porphyrias is skin photosensitivity. With the exception of ALA dehydratase deficiency porphyria (ADP) and acute intermittent porphyria (AIP), where the main symptoms are neurovisceral attacks because they do not accumulate the photodynamically active porphyrins, and porphyria cutanea tarda (PCT) where patients suffer from skin photosensitivity, all the other hepatic porphyrias may involve neurovisceral attacks coupled with photosensitivity.

1.3 ACUTE INTERMITTENT PORPHYRIA (AIP)

1.3.1 INTRODUCTION

Acute intermittent porphyria (AIP) is the most common autosomal dominant form of the acute porphyrias. It has been estimated to have a prevalence of 1 in 10000-20000 in Europe (Goldberg et al, 1987), with the highest incidence of 1 in 1000 occurring in Swedish Lapland (Waldenström, 1957). Hence AIP is also known as Swedish porphyria. A

Table 1.1: Classification of the Porphyrias.

Classification	Defective enzyme	Main presenting symptom(s)
Erythropoietic		
Congenital erythropoietic porphyria (CEP)	Uroporphyrinogen III cosynthase	Photosensitivity
Erythropoietic protoporphyria (EPP)	Ferrochelatase	Photosensitivity
Hepatic		
ALA dehydratase deficiency porphyria (ADP)	ALA dehydratase	Neurovisceral
Acute intermittent porphyria (AIP)	Hydroxymethylbilane synthase	Neurovisceral
Hereditary coproporphyria (HCP)	Coproporphyrinogen oxidase	Neurovisceral and/or photosensitivity
Variegate porphyria (VP)	Protoporphyrinogen oxidase	Neurovisceral and/or photosensitivity
Porphyria cutanea tarda (PCT)	Uroporphyrinogen decarboxylase	Photosensitivity
Hepatoerythropoietic porphyria (HEP)	Uroporphyrinogen decarboxylase	Photosensitivity and/or neurovisceral

much higher prevalence of 2 to 5 per thousand psychiatric patients has also been reported (Tishler et al, 1985).

AIP is a consequence of a defect in the gene encoding the third enzyme of the heme biosynthetic pathway, hydroxymethylbilane synthase (HMB-S) (EC 4.3.1.8), resulting in a partial deficiency in its activity and an accumulation of the porphyrin precursors 5-aminolaevulinic acid (5-ALA) and porphobilinogen (PBG), leading to disease manifestations. Unlike most other porphyrias, which accumulate porphyrins, patients with AIP do not develop cutaneous photosensitivity, since they accumulate porphyrin precursors but not porphyrins, which are photodynamically inactive.

1.3.2 CLINICAL FEATURES OF AIP

AIP is a disorder of variable penetrance (Gates, 1946). Approximately 90% of carriers of the defective gene remain asymptomatic throughout their lives, while the remaining 10% manifest the disease to varying degrees of severity. The frequency of attacks may vary from one individual to another, and is more common in females than in males (Goldberg, 1959).

The clinical onset of AIP occurs during or after puberty and the main presenting symptoms are abdominal pain, tachycardia, hypertension, vomiting, neurological disturbances and psychiatric symptoms. Abdominal pain, which may be generalised or localised, is almost invariably present and is usually the initial symptom of an acute attack. Tachycardia and hypertension have been reported in more than half of all attacks (Moore et al, 1987). In up to 40% of patients, hypertension may become sustained between attacks. Severe vomiting results in dehydration and hyponatremia, accompanied by constipation (Moore and Brodie, 1990). Any type of neuropathy may occur, which may be symmetric, asymmetric or focal (Poser and Edwards, 1978). Muscle weakness can be severe due to a prolonged attack, and is due to the effects of porphyria on nerves that control muscles. It often begins proximally in the legs but may involve the arms or the distal extremities. The higher rate of occurrence of AIP in the psychiatric institutions could be attributed to the misdiagnosis of such patients as suffering from a genetic disorder.

1.3.3 PRECIPITANTS OF AIP

Approximately 90% of the defective gene carriers remain asymptomatic, although they may manifest the condition either by endogenous or exogenous environmental factors. The precipitants of AIP include certain therapeutic drugs, alcohol, infections, caloric deprivation and endocrine factors.

Barbiturates are the most potent of all drugs known to exacerbate acute porphyria. They are still widely used for the induction of anaesthesia during surgery and therefore it is vital that the patient's condition is known well in advance of a surgical procedure. Alcohol and infections (both bacterial and viral) are known to contribute to the induction of an acute attack. Urinary 5-ALA and PBG excretion have been found to increase in AIP patients with reduced calorie intake. However, this situation is reversed if calories from a carbohydrate-rich diet are introduced into the diet. The clinical onset of AIP occurs during or after puberty, indicating that endocrine factors play a role in the genesis of an acute attack.

All the precipitating factors result in the induction of the rate-limiting enzyme of the pathway, ALA synthase (Kappas et al, 1995), which causes an overproduction of 5-ALA and PBG. A reduced activity of HMB-S is unable to prevent the accumulation of these compounds, thereby leading to the symptoms of AIP. It can be seen, therefore, that it is vital to identify carriers of this defective gene so that they can be appropriately advised to avoid these precipitating factors.

1.3.4 MOLECULAR GENETICS OF HYDROXYMETHYLBILANE SYNTHASE (*HMB-S*)

The human hydroxymethylbilane synthase (*HMB-S*) gene has been localised to 11q24.1→q24.2 (Namba et al, 1991) by *in situ* hybridisation and gene dosage studies. The gene is divided into 15 exons (Figure 1.3) and is spread over 10kb of genomic DNA (Chretien et al, 1988; Yoo et al, 1993). The length of individual exons ranges from 39 to 438 base pairs (bp). The largest intron spans 2913bp and the smallest 87bp.

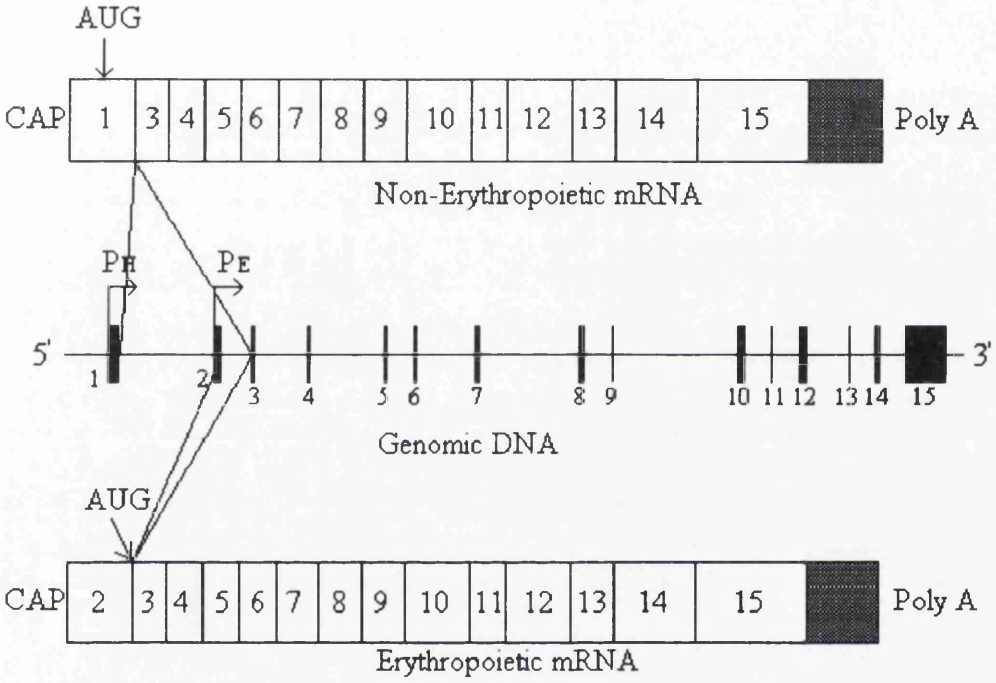


Figure 1.3: The organisation of the human hydroxymethylbilane synthase (*HMB-S*) gene and its two transcripts. The vertical boxes in the genomic DNA represent the exons, drawn to scale. P_H and P_E indicate the promoters of the housekeeping gene and erythroid cell respectively.

HMB-S is encoded by a single gene. Two distinct forms of the enzyme exist: a housekeeping isoenzyme of 44 kilodaltons (kD) which is present in all cells, and an erythroid-specific isoenzyme of 42kD. The two isoforms differ at the N-terminus where the housekeeping enzyme possesses 17 additional amino acids.

These two isoforms are a consequence of alternative splicing of two primary transcripts arising from two distinct promoters (Grandchamp et al, 1987; Chretien et al, 1988). The housekeeping promoter is located in the 5' untranslated region, and its transcript comprises exons 1 and 3 to 15 while the erythroid promoter, located 3kb downstream, is in intron 1 and its transcript contains exons 2 to 15 (Figure 1.3). Two erythroid-specific transcription factors, GATA-1 and NF-E2, recognise the sequences in the *HMB-S* erythroid promoter. The translation initiation codon is not present in exon 2, therefore, the translated portion of the erythroid mRNA begins in exon 3.

The isolation and sequencing of the cDNAs encoding the two isoforms have been performed (Raich et al, 1986; Grandchamp et al, 1987) and revealed that the cDNA encoding the erythroid form has an open reading frame of 1032bp which codes for 344 amino acids, while that for the housekeeping isoenzyme is 1083bp, coding for 361 amino acids. The complete coding sequence of the *HMB-S* gene is shown in Appendix 2.

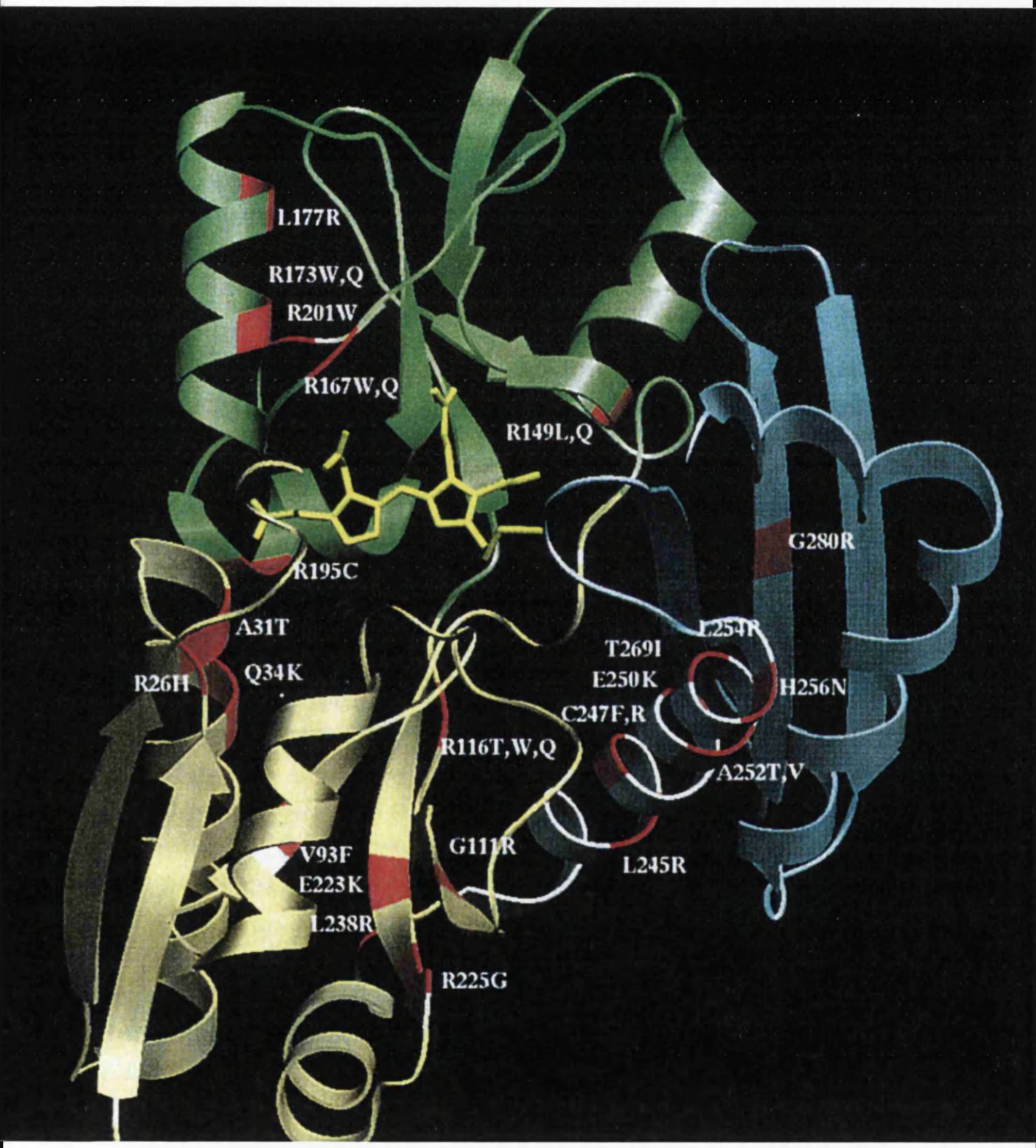
1.3.5 THE THREE-DIMENSIONAL STRUCTURE OF HMB-S

The three-dimensional structure of wild-type HMB-S from *E. coli* has been elucidated by Louie et al (1996) at a resolution of 1.76Å (Figure 1.4). The overall dimensions of the protein molecule are 57Å X 43Å X 32Å. The polypeptide chain of HMB-S is folded into three α/β domains of approximately equal size and the overall topology of domains 1 (residues 3-99) and 2 (residues 105-193) were found to be similar based on the doubly-wound, five-stranded mixed β-sheet. At the interface of these two domains is an active site cleft with dimensions 15Å X 13Å X 12Å. The major links between these two domains are two hinge regions. Domain 3 (residues 222-307), on the other hand, has an approximately equal interaction with the other two domains through hydrogen bonds and salt-bridges. The C-terminal of domain 3 is an open-faced, anti-parallel sheet consisting of three strands

Figure 1.4: The three-dimensional structure of the *E. coli* hydroxymethylbilane synthase (HMB-S) protein.

The three domains of the enzyme are illustrated as different colours, with domain 1 being represented in gold, domain 2 in green and domain 3 in blue. The dipyrromethane cofactor, which is situated in a deep active site cleft between domains 1 and 2, is shown in yellow. Examples of some of the mutations identified so far are shown according to the human numbering system and the sites of the mutations are depicted in red. The β-strands are represented by wide arrows, α-helices as ribbons and loops as thin ropes. (Reproduced from Warren et al, 1996).

Figure 1.4: The three-dimensional structure of the *E. coli* hydroxymethylbilane synthase (*HMB-S*) protein.



which forms a short segment with domain 1 through residues 218-221. The dipyrromethane cofactor is important in the assembly of the tetrapyrrole product. It is assembled by the apoenzyme from two molecules of PBG and is covalently attached, through a thioether bond, to the side chain of cysteine-242 located on a flexible loop of domain 3. The cofactor is important in the stabilisation of the protein structure since, once it is assembled, the protein undergoes a permanent and irreversible conformational change which maintains the cofactor at the active site. It forms hydrogen bonds and salt-bridges between its acetate and propionate side groups and the polypeptide chain. The second pyrrole ring (C2) of the cofactor is located internally in the active site cleft before oxidation. Upon oxidation, it was found to be externally located which may correspond to the site of substrate binding and polypyrrole chain elongation.

The majority of the invariant residues are clustered in the vicinity of the active site cleft. These residues are mainly involved in bond formation (Louie et al, 1996), particularly salt-bridges and hydrogen bonds, with the dipyrromethane cofactor and the PBG substrate, and also in the elongation of the product. Aspartic acid 84 (Asp84) has been identified as the key catalytic residue of HMB-S due to its strategic location in the vicinity of the active site cleft within the molecule. Replacement of this residue has been found to either reduce the specific activity of the protein (Glu84 has less than 1% of the wild-type specific activity) or completely abolish the catalytic activity (Ala84 and Asn84). The invariant residues located some distance away from the active site cleft are more important in stabilising the protein structure than interacting with the cofactor and product. Any disruption to these highly conserved residues will render the protein inactive or unstable and lead to the formation of a non-functional protein.

1.3.6 MOLECULAR PATHOLOGY OF AIP

To date, seventy-three different mutations have been reported to be responsible for AIP, indicating the immense molecular heterogeneity of this condition. Deletions, insertions, missense, nonsense and other point mutations causing splicing defects have been characterised (Table 4.1).

Immunoassays have been developed to measure the antigenic concentration of erythrocyte HMB-S. AIP patients can be divided into two categories based on the measurements of the enzyme activity and immunoreactive concentration in erythrocytes. Patients who show the presence of a stable erythroid HMB-S but with 50% enzyme activity are classified as cross-reacting immunological material (CRIM)-positive, because antigenic material is present in excess of activity (Nordmann et al, 1990). CRIM-positive AIP is further subdivided into two types, Type 1 and Type 2, based on the ratio of the immunologic to enzymatic material (Table 1.2). Approximately 85% of AIP patients are classified as CRIM-negative, where the ratio of immunoreactive concentration to enzymatic activity in erythrocytes is 1.0. CRIM-negative AIP can be subdivided into Type 1, where there is a reduction in the amounts of immunologic and enzyme material, and Type 2, where the activity and immunoreactive material are present at normal levels. The latter group of patients do not possess any mutations in their erythroid cells, therefore, their erythrocyte HMB-S activities are normal and the immunoreactive material is recognised by the antibodies raised against them. These patients can be differentiated from the healthy population as they have an increased urinary 5-ALA and PBG. The CRIM-status of the different mutations are stated in Table 4.1.

A large number of mutations occur at CpG dinucleotides, in line with the fact that they are known hotspots for mutations due to the spontaneous deamination of 5'-methylcytosine to thymine (Barker et al, 1984; Cooper and Youssoufian, 1988). This is true for the mutations responsible for AIP, with more than half of them involving mutations in CpG dinucleotides.

1.3.7 DIAGNOSIS OF AIP

All AIP patients, except those classified as CRIM-negative Type 2, and latent gene carriers, have approximately 50% HMB-S activity in their erythrocytes as compared to normal individuals. 5-ALA and PBG are excreted in large amounts in the urine between attacks in clinically expressed and some latent AIP patients. However, increased levels of 5-ALA and PBG are also observed in hereditary coproporphyria (HCP) and variegate porphyria (VP). These porphyrias can produce the same symptoms as AIP and the pattern

Table 1.2: Classification of AIP based on the ratio of immunoreactive concentration to enzyme activity in erythrocytes (adapted from Desnick et al, 1985).

	Immunoreactive Concentration (%) (A)	Enzymatic Activity (%) (B)	Ratio (A/B)
Control	100	100	1.0
CRIM-negative			
Type 1	50	50	1.0
Type 2	100	100	1.0
CRIM-positive			
Type 1	85	50	1.7
Type 2	280	50	5.7

of inheritance is also similar to AIP. The HMB-S activity is normal in HCP and VP, thus differentiating these two conditions from AIP.

The diagnosis of AIP, thus, depends on a reduced activity of HMB-S in erythrocytes. The diagnosis of asymptomatic heterozygotes, who possess normal levels of 5-ALA and PBG, has been difficult since there is a significant overlap of erythrocyte HMB-S activity between the affected and normal values (Lamon et al, 1979; McColl et al, 1982; Bonnaiti-Pellie et al, 1984; Pierach et al, 1987; Kappas et al, 1989). Therefore, a more definitive approach in the diagnosis of AIP is to use molecular techniques to identify specific mutations in AIP patients.

1.3.8 LINKAGE ANALYSIS

Ten diagnostically useful intragenic polymorphisms have been detected in the *HMB-S* gene (Astrin and Desnick, 1994). Since correct diagnosis is essential for the prevention of acute attacks, restriction fragment length polymorphism (RFLP) analysis provides a valuable diagnostic test for AIP at the DNA level. This test is especially valuable in AIP families with normal erythrocyte HMB-S activity.

Linkage analysis of AIP families using RFLPs has demonstrated close linkage of the disease with the *HMB-S* gene (Llewellyn et al, 1987; Lee et al, 1988; Grandchamp et al, 1989c). Four intragenic RFLPs in intron 1, 1345G/A (*Msp* I), 1500T/C (*Pst* I), 2377C/A (*Apa* L1) and 2479A/G (*Bst* NI), have been shown to be in linkage disequilibrium with the *HMB-S* gene (Scobie et al, 1990a; Lee et al, 1991), which makes these RFLPs less useful in the diagnosis of AIP.

Linkage analysis may prove to be useful in the diagnosis of presymptomatic AIP heterozygotes in families whose molecular pathologies are not known. However, a key individual known to suffer from the condition has to be available to allow the RFLPs to be tracked through the generations. Another important point is that the individuals analysed should be informative for the markers. These limitations can be overcome by performing direct mutation detection using one of the methods described in section 1.4.

1.4 METHODS AVAILABLE FOR MUTATION DETECTION

The detection of sequence changes in human DNA is of great importance in the understanding of the molecular pathologies involved in genetic disease. Many methods have been developed for the detection of base alterations but no single method is suitable for all situations. Each technique has its advantages and disadvantages for a given situation, therefore, the method of choice should be evaluated to suit the length of template available (e.g. length of cDNA to be analysed), number of samples to be screened, or the urgency of the results.

1.4.1 POLYMERASE CHAIN REACTION (PCR)

The polymerase chain reaction (PCR) is an *in vitro* nucleic acid amplification method which has revolutionised modern molecular genetics. Saiki et al (1985) were the first to describe the technique, which consequently, became more widely used as compared to cloning strategies.

PCR comprises repeated thermal cycling of the reaction mixture. The thermal cycling serves to dissociate the products of the previous thermal cycle, then allowing the association of these dissociated products with further reaction starting materials for another phase of synthesis. Specifically, the template DNA is denatured at a high temperature to give single-stranded molecules. This is followed by short oligonucleotide primers (amplimers) annealing to specific, nucleotide sequence-defined regions of the template at a lower temperature. The positions where the amplimers anneal to the template define the 'target'. The thermal cycle is concluded by the amplimers being enzymatically extended by the coupling of appropriately base-paired 2'-deoxynucleoside triphosphates (dNTPs) on the template at an intermediate temperature, thus producing another double-stranded DNA (dsDNA) copy of the original target.

Each PCR cycle comprises a set of time- and temperature-controlled incubations. The function of each incubation is to:

(i) denature the target nucleic acid at a high temperature, usually in the region of 94°C;

(ii) anneal amplimers at a temperature dependent on their calculated annealing temperature (usually in the range 30 - 65°C);

(iii) extend the primers by using the thermostable enzyme DNA polymerase which catalyses the addition of nucleotides to the 3' end of each primer at a temperature of about 72°C.

Each cycle results in the exponential amplification of the number of copies of the target sequence.

All the mutation detection methods currently used are PCR-based. These include single strand conformation polymorphism (SSCP), heteroduplex analysis (HA), denaturing gradient gel electrophoresis (DGGE), temperature gradient gel electrophoresis (TGGE), chemical cleavage mismatch (CCM) analysis, RNase A cleavage analysis, protein truncation test (PTT) and direct sequencing.

1.4.2 SINGLE STRAND CONFORMATION POLYMORPHISM (SSCP) ANALYSIS

Single strand conformation polymorphism (SSCP) analysis allows the detection of single base changes in short DNA fragments due to the mobility differences of single-stranded DNA molecules. It depends on the ability of the fragment to fold back on itself to give a unique conformation by intramolecular base pairing. It was first described by Orita et al (1989) and is based on the principle that single stranded DNA molecules assume a three-dimensional conformation which is highly dependent on the primary sequence. If there is a sequence difference between the wild-type and mutant DNA, each of the single strands will assume a different conformation from the wild-type strands after denaturation and migrate differently on a non-denaturing polyacrylamide gel (Figure 1.5). Under non-denaturing gel conditions, the conformation of the single strands is stabilised by intrastrand interactions.

SSCP has a detection rate of about 70 - 95% for segments about 200bp in length, which is a drawback if a large region is to be analysed. Another disadvantage of SSCP analysis is that it does not allow the location of the mutation within the fragment.

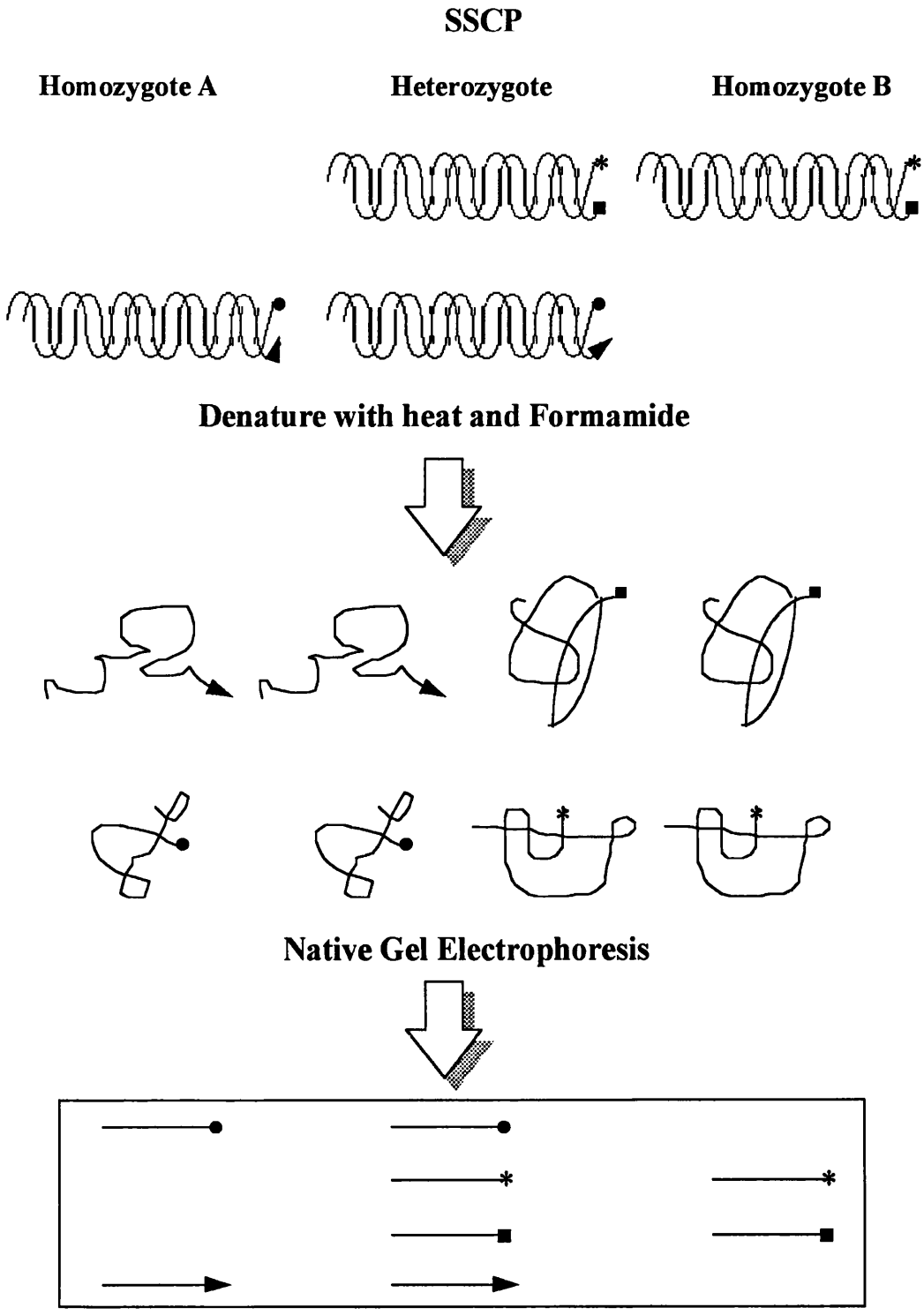


Figure 1.5: Diagrammatic representation of the principle of SSCP.

(Adapted from Flowgen FMC Bioproducts).

1.4.3 HETERODUPLEX ANALYSIS (HA)

Heteroduplex analysis (HA) is a relatively simple and rapid method for detecting mismatches between double-stranded DNA containing one wild-type strand and a complementary strand with an altered nucleotide sequence (Figure 1.6). The heteroduplexes are formed during the later stages of PCR (Nagamine et al, 1989). Fragments between 200-600bp are optimal for resolution of the homo- and heteroduplexes. Its detection rate is 80 - 90% for fragments less than 300bp (Grompe, 1993). Therefore, larger fragments will have to be screened by another method.

1.4.4 DENATURING GRADIENT GEL ELECTROPHORESIS (DGGE)

Denaturing gradient gel electrophoresis (DGGE) is a rapid screening technique which allows the resolution of DNA fragments differing by as little as a single nucleotide change. The method is based on the differential electrophoretic mobilities of wild-type and mutant DNA through linearly increasing concentrations of a denaturing agent (formamide and urea are common denaturants) (Fischer and Lerman, 1983; Myers et al, 1985). Another method of generating a denaturing gradient is to use temperature, and the method is known as temperature gradient gel electrophoresis (TGGE) (Meyer et al, 1991). The melting properties of DNA molecules with a single base difference will cause them to migrate differently into discrete sequence-dependent domains of low melting temperature. When the concentration of the denaturing agent equals the melting temperature (T_m) of the lowest melting domain of the DNA fragment, its mobility will eventually be retarded. The T_m of a melting domain is highly dependent on its nucleotide sequence because the stacking interactions between adjacent bases have a considerable effect on the stability of the double helix. Therefore, when the wild-type and mutant DNA fragments differing by a single base are electrophoresed in a polyacrylamide gel that contains an increasing linear denaturing gradient from the top to the bottom, the molecules will melt accordingly. The wild-type and mutant DNA can be visualised as discrete bands and any differences between the wild-type and mutant DNA can be easily identified.

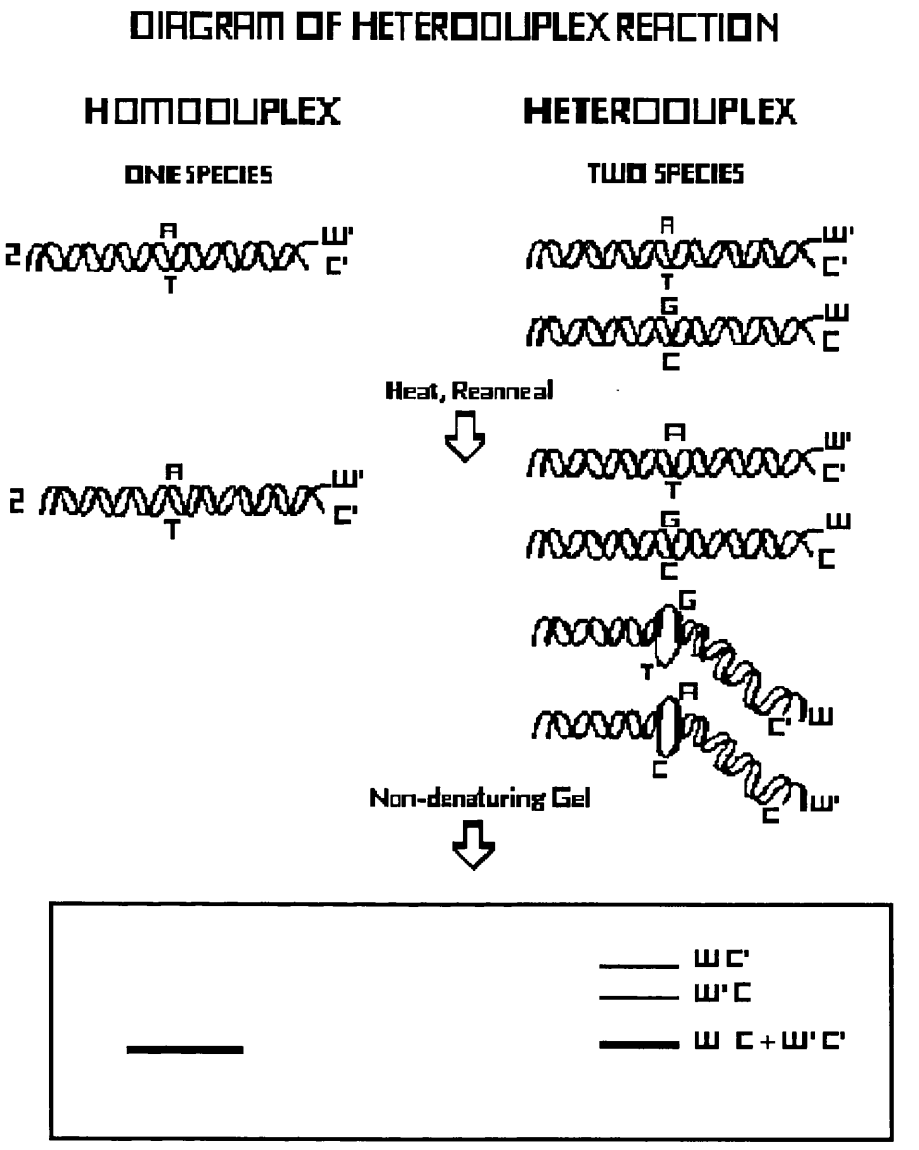


Figure 1.6: Diagrammatic representation of the principle of HA.

(Adapted from Flowgen FMC Bioproducts).

DGGE has been reported to have a 95% accuracy in PCR products up to 600bp (Grompe, 1993). In order for DGGE to function efficiently, a "GC-clamp" of about 30-40 nucleotides long is required at the 5' end of the primer to provide a high T_m domain to prevent complete denaturation (Myers et al, 1985a). Like SSCP, DGGE does not localise the base change within a fragment.

1.4.5 RNase A CLEAVAGE ANALYSIS

The principle of heteroduplex analysis is also applicable to RNase A cleavage analysis. In this technique, an RNA:DNA heteroduplex between a radioactive wild-type riboprobe and mutant DNA, generated by PCR, is subjected to RNase A cleavage (Myers et al, 1985b). This enzyme will recognise and cleave single-stranded RNA at the points of mismatch. The reaction is analysed by electrophoresis and autoradiography. The mutation can be localised by this method but the use of radioactively labelled RNA is required and has a detection rate of only approximately 50%. Because of this, it has been replaced by chemical cleavage mismatch (CCM) analysis.

1.4.6 CHEMICAL CLEAVAGE MISMATCH (CCM) ANALYSIS

The chemical cleavage mismatch (CCM) technique was first described by Cotton et al (1988) and is based on the principle of generating a heteroduplex between a radiolabelled wild-type DNA and mutant DNA by boiling and reannealing. If a base alteration is present in the mutant DNA, a heteroduplex is formed between the control and test DNAs. The site of mismatch is chemically modified with hydroxylamine and osmium tetroxide, and the site of modification is cleaved by piperidine. This is followed by denaturing polyacrylamide gel electrophoresis and autoradiography. Osmium tetroxide modifies mispaired thymines and hydroxylamine modifies mispaired cytosines. Adenosine and guanosine mismatches of the wild-type sense strand are detected by labelling the antisense strand of wild-type DNA in the heteroduplex. CCM is a very sensitive technique with a detection rate of >95% when only the wild-type DNA is labelled and 100% when both the mutant and wild-type DNAs are labelled (Forrest et al, 1991). The advantages of this method are threefold: (i) up to 2kb segments can be screened for sequence alterations; (ii) the type of base change is known

from the cleaving reagent; and (iii) the localisation of the change is indicated by the size of the cleavage products. However, the entire CCM procedure is lengthy and involves the use of hazardous chemicals and radioactivity. This technique has been improved with the use of silver staining instead of [α-^{32}P]dCTP (Saleeba et al, 1992) as an alternative method of detection, thus eliminating the use of radioactivity.

1.4.7 PROTEIN TRUNCATION TEST (PTT)

One of the latest techniques developed for mutation detection is the protein truncation test (PTT), first described by Roest et al (1993). This technique is useful for detecting mutations leading to premature translation termination and is not applicable to other single base alterations. A relatively large segment of coding sequence (about 2.4kb) can be screened and the site of mutation is localised, hence only a small region needs to be sequenced.

1.4.8 DIRECT SEQUENCING

Direct sequencing is a technique used to analyse the sequence of PCR products, either using DNA or cDNA as templates for PCR without an additional prior subcloning step into sequencing vectors. For genes which are relatively small, such as the hydroxymethylbilane synthase (*HMB-S*), direct sequencing of the coding region is used as the main method of mutation detection (Mgone et al, 1992).

Two methods are available for sequencing- the Maxam and Gilbert method (1977), which is based on the chemical modification of a particular base, and the Sanger method (1977), also known as the dideoxy method, which is an enzyme-based method.

In the Maxam and Gilbert method, the template is cut with a restriction enzyme into fragments which are then sorted by size. Batches of identical fragments, which have been radioactively labelled at one end, are then divided into four separate reaction tubes and treated with specific reagents which modify one or two of the bases. Under the conditions used for cleavage, each template molecule is probably cut at a small number of sites but in the total mixture, every position should be cleaved. The fragments are separated by high

resolution denaturing polyacrylamide gel electrophoresis and visualised by autoradiography.

The Sanger method, or dideoxy method, involves the synthesis of the complementary strand of a single-stranded template using DNA polymerase. The procedure consists essentially of three steps: (i) annealing; (ii) labelling; and (iii) termination. The primer is initially annealed to the template, followed by the incorporation of the four nucleotides into the growing strand catalysed by DNA polymerase. A radioactively labelled nucleotide is included in the mixture for visualisation by autoradiography after separation on denaturing polyacrylamide gels. The termination reaction involves the addition of dideoxynucleoside triphophates (ddNTPs) into the four separate reaction mixtures. ddNTPs lack a 3'-OH group which is necessary for chain elongation, resulting in fragments of various lengths. These can be visualised as a sequence of bands upon autoradiography after polyacrylamide gel electrophoresis.

Direct sequencing provides the exact information on the location and base change involved, therefore, has to be performed as a final confirmatory step in all mutation detection methods.

Therefore, the overall strategy adopted in the detection of mutations in the *HMB-S* gene was to screen the entire coding sequence using HA and CCM, followed by direct sequencing. This strategy was chosen because both screening technologies enable the rapid analysis of the cDNA, which is short, as well as allowing the detection of point mutations, which are predominant in AIP. Based on the features of AIP and the *HMB-S* gene, SSCP and PTT are not suitable for use in this study.

1.5 *IN VITRO* GENE EXPRESSION SYSTEMS

Five major expression systems are available to researchers for application in *in vitro* gene expression: (i) *E. coli*; (ii) *Bacillus subtilis*; (iii) yeast; (iv) mammalian cells; and (v) baculovirus. Proteins can be grouped into four classes to simplify the selection of expression systems (Goeddel, 1990): (i) small peptides (<80 amino acids) are easily

expressed as fusion proteins in *E. coli*; (ii) secreted proteins (e.g. enzymes, cytokines and hormones), 80 to 500 amino acids, can be subjected to direct expression in all five systems. Proteins 100 to 200 amino acids in size can be efficiently expressed directly in *E. coli*; (iii) large (>500 amino acids) secreted proteins and cell surface receptor proteins are best expressed in mammalian cell expression systems; and (iv) non-secreted proteins (>80 amino acids) which have not yet been studied thoroughly.

One of the purposes for expressing a new gene or cDNA is to ensure that the isolation of the correct sequence has been successful. If this is the main purpose of the experiment, and the protein being expressed is derived from a higher eukaryote, transient expression in mammalian cells can be easily performed, giving results rapidly and usually a biologically active protein is produced.

A growing interest in *in vitro* gene expression is the production of a mutant protein, either for structure-function experiments, or the correlation of the activity of a mutant enzyme with the mutation present. Therefore, it is important to select an appropriate expression system which will allow the generation of mutants in an assayable form.

The *E. coli* expression system is the most widely used system for *in vitro* expression. A high level of gene expression is a key requirement for basic research and the production of useful biological materials. The amount of proteins produced in culture at the time of harvest is of utmost importance and this requirement is achieved by using *E. coli* which is able to attain very high cell densities. *E. coli* has been considered to be a good expression system for producing proteins to generate antibodies. The antigen can be expressed rapidly in large quantities in *E. coli*, making it an ideal choice for this purpose.

Although a large number of proteins can be produced in *E. coli*, a number of limitations exist for the protein produced. It should not be too small, too large, too hydrophobic or contain too many cysteines. These guidelines should generally be taken into account if a stable protein with the correct conformation is to be produced. Fusion protein expression strategies usually permit good translation initiation and allows one to generate stable protein products with small peptides.

1.6 AIMS OF THE PRESENT STUDY

The main aims of this project were:

(I) To detect and fully characterise mutations in the hydroxymethylbilane synthase (*HMB-S*) gene in patients suffering from acute intermittent porphyria (AIP).

(II) To develop an *in vitro* gene expression system to correlate the activities of the enzyme with the severity of the mutations detected using the *HMB-S* gene as a model.

(I) Detection and characterisation of mutations in the *HMB-S* gene

(a) To amplify the *HMB-S* coding sequence for subsequent mutation screening by heteroduplex and CCM analyses.

(b) To characterise the alterations detected by direct sequencing.

(c) To predict the effects of the mutations identified on the structure and function of the protein product.

(II) Development of an *in vitro* expression system using the *HMB-S* gene as a model

(a) To introduce the *HMB-S* coding sequence into a plasmid expression vector containing a strong *trc* promoter and a strong transcription termination signal (*rrn*B).

(b) To verify the sequence of the clones by CCM analysis and direct sequencing before proceeding to gene expression studies.

(c) To determine the specificity of the custom-made antibody for the human form of HMB-S in order to separate it from the *E. coli* form of the enzyme.

(d) To separate the human HMB-S from the *E. coli* form using a variety of protein separation techniques.

(e) To measure the activity of the human enzyme in the *in vitro* expression system and to correlate with the mutation present.

(f) To compare, where possible, the activity levels of the enzyme in the erythrocyte with that obtained in an *in vitro* expression system.

CHAPTER 2:

MATERIALS AND METHODS

CHAPTER 2: MATERIALS AND METHODS

2.1 PATIENTS

Nine AIP samples from nine different families from the University of Cape Town, South Africa, were kindly provided by Dr. Richard Hift. In addition, three samples from one family from England where the index case was individual III:2 of Pedigree 13519 (Appendix 1), seven samples from two families from the Republic of Ireland where the index cases of Pedigree 13168 were individuals III:1 and III:3 (Appendix 1) (the pedigree of the other family was not available) and eleven samples from three families from Scotland where the index cases of Pedigree 9532 were individuals III:8 and III:9 (Appendix 1) and that of Pedigree 13543 was individual III:6 (Appendix 1) (the individual from the third family was an isolated case), were sent to us for molecular analysis. All these patients were identified as AIP sufferers on the basis of at least one of the following criteria: (i) increased urinary excretion of PBG, decreased activity of erythrocyte HMB-S and at least one acute attack; (ii) a relative suffering from AIP; or (iii) equivocal biochemical values. The biochemical data of the patients from South Africa and members of Pedigree 13168 are shown in Tables 2.1 and 2.2 respectively. The other family members analysed in the four pedigrees available were: IV:1 and IV:2 (Pedigree 13519) (II:2 and III:5 were found to be affected based on their biochemical values which were not provided); II:1 to II:6, II:8, III:2, III:5 to III:8, IV:1 and IV:2 (Pedigree 13168); II:2, III:3, III:5, III:7 and III:11 (Pedigree 9532) (II:7 was found to be affected based on his biochemical values which were not sent to us); and II:3, III:2, III:3 and III:5 (Pedigree 13543) (the biochemical data of II:1 was not available).

To confirm that any sequence alterations detected in any of the above patients was not a polymorphism, twenty healthy individuals from the normal population were analysed for the same base change.

Table 2.1: Biochemical data of the South African patients under study.

Patient	Clinical Phenotype	Urinary PBG (μmol/10mmol creatinine)	Erythrocyte HMB-S activity (nmol PBG/g Hb/hr)	% of HMB-S activity of patients as compared to normal individuals
644	2 previous acute attacks but has since remained well	263.0	5.05	20.5 - 37.9
635	Asymptomatic	233.0	12.1	49.1 - 90.7
638	Quiescent	52.1	5.41	22.0 - 40.6
639	1 very mild attack but is in complete remission	88.6	ND	-
637	Severe attacks requiring 2-3 monthly admissions	282.4	8.11	32.9 - 60.8
632	Very severe attacks proceeding to paralysis and respiratory failure; tuberculosis complicated treatment	894.0	9.41	38.2 - 70.5
Normal		<16.0	13.34 - 24.62	

ND = not determined

Table 2.2: Biochemical data of five members of Pedigree 13168.

Individual	Erythrocyte HMB-S activity (nmol/ml RBC/hr)
II:6	20.6
III:1	18.4
III:2	28.7
III:3	19.8
III:5	29.5
Reference range	22 - 46

2.2 EXTRACTION OF GENOMIC DNA FROM BLOOD

Genomic DNA was extracted from peripheral blood lymphocytes according to a variation of the method described by Kunkel et al (1977). 10ml of peripheral blood collected in tubes containing ethylenediamine tetraacetic acid (EDTA) as an anticoagulant was transferred to a labelled 50ml Falcon tube and 40ml of cold lysis buffer (0.32M sucrose, 10mM Tris-HCl pH 7.5, 5mM $MgCl_2$, 1% Triton X-100) was added. The contents of the tube were mixed gently by inversion and centrifuged at 2800rpm for 10 minutes at 4°C. The supernatant was discarded and the resulting pellet was resuspended in 3ml of nuclei lysis buffer (10mM Tris-HCl pH 8.0, 0.44M NaCl, 2mM EDTA pH 8.2), followed by the addition of 100µl of proteinase K (10mg/ml) to catalyse the digestion of histones and other proteins, and 200µl of 10% SDS. The contents were thoroughly mixed by inversion (the mixture should become dark and viscous) and incubated overnight at 37°C. 1ml of 6M NaCl was added to the overnight mixture to remove proteins and other impurities and shaken vigorously for 15 seconds. The tube was spun at 2500rpm for 15 minutes. The supernatant was transferred to a fresh Falcon tube and 3ml of phenol/chloroform was added. The tube was inverted gently for 2 to 3 minutes before spinning at 1000rpm for 20 minutes. The upper aqueous layer was transferred to a fresh tube and the DNA was precipitated with 2 volumes of absolute ethanol. The DNA was spooled out using a sealed Pasteur pipette, washed in 70% ethanol, air-dried and resuspended in 500µl of TE buffer (10mM Tris-HCl pH 7.5, 1mM EDTA). The DNA was stored at 4°C until required.

2.3 EXTRACTION OF TOTAL CELLULAR RNA FROM LYMPHOCYTES

All necessary precautions were taken to avoid contamination by RNases. Glassware and plasticware used for the storage and preparation of RNA were soaked overnight in 0.1% diethyl pyrocarbonate (DEPC) in sterile distilled water (SDW), rinsed in SDW and autoclaved.

The lymphocytes were separated from whole blood as follows:

10ml of whole blood was collected in tubes containing EDTA and gently layered over an equal volume of Histopaque®-1077 (SIGMA) in a Universal. The tube was centrifuged at 1400rpm for 30 minutes at room temperature. The interphase layer, consisting of mononuclear cells including lymphocytes, was carefully removed with a pastette and transferred to a fresh Universal. The cells were washed with 20ml of cold phosphate buffered saline (PBS) (SIGMA; each tablet contains 0.01M phosphate buffer, 0.0027M KCl, 0.137M NaCl, pH 7.4 at 25°C) and pelleted by centrifugation at 1400rpm for 10 minutes. The supernatant was discarded and the pellet was subjected to total cellular RNA extraction using one of the following methods.

2.3.1 THE ACID-GUANIDINIUM THIOCYANATE-PHENOL-CHLOROFORM METHOD

The cells were given another wash with PBS and RNA extraction was performed according to the method described by Chomczynski and Sacchi (1987). The resulting pellet was resuspended in 500µl of solution D (denaturing solution) (4M guanidinium thiocyanate, 25mM sodium citrate pH 7.0, 0.5% sarcosyl, 0.1mM β-mercaptoethanol). At this stage, DEPC-treated tips and tubes are used. [To minimise the handling of the hazardous compound guanidinium thiocyanate, 250g of guanidinium thiocyanate (Fluka) was dissolved directly, without weighing, in 293ml of water in the manufacturer's bottle, 17.6ml of 0.75M sodium citrate, pH 7.0 and 26.4ml of 10% sarcosyl at 65°C]. The mixture was transferred to an autoclaved and DEPC-treated 1.5ml eppendorf tube and to it was added 50µl of 2M sodium acetate (pH 4.0), 500µl of water-saturated phenol and 100µl of chloroform-isoamyl alcohol (49:1 ratio). The mixture was vortexed and placed on ice for 15 minutes, after which it was centrifuged at 10000rpm for 20 minutes at 4°C. The aqueous phase was transferred to a fresh 1.5ml DEPC-treated eppendorf tube and 1ml of isopropanol was added. The contents were mixed well and kept at -20°C for at least 1 hour or at 4°C overnight. The tube was centrifuged and the supernatant discarded. The pellet was dissolved in 300µl of solution D and 300µl of isopropanol was added. The tube

was left at -20°C for 1 hour and then spun for 10 minutes at 4°C. The pellet was washed in 250µl of 75% ethanol, air-dried and dissolved in 100µl of DEPC-treated water at 65°C for 10 minutes. The RNA sample was stored at -20°C.

2.3.2 THE TRIZOL® METHOD

1ml of TRIZOL® reagent (Gibco BRL) was added to the pellet to lyse the cells by passing the suspension several times through a DEPC-treated pipette tip. The lysate was transferred to a 1.5ml DEPC-treated eppendorf tube and incubated for 5 minutes at room temperature to permit the complete dissociation of nucleoprotein complexes. 200µl of chloroform per 1ml TRIZOL® reagent was added and shaken vigorously by hand for 15 seconds. The mixture was incubated at room temperature for 2 to 3 minutes, after which it was centrifuged at no more than 12000g for 15 minutes at 4°C. The colourless upper aqueous phase, containing the RNA, was transferred to a fresh 1.5ml eppendorf tube. [The lower red, phenol-chloroform phase can be used for DNA extraction if required]. 500µl of isopropanol per 1ml of TRIZOL® reagent used for the initial homogenisation was added to precipitate the RNA from the aqueous phase and incubated at room temperature for 10 minutes. The tube was centrifuged at 12000g for 10 minutes at 4°C, the supernatant was discarded and the RNA pellet was washed once with 1ml of 75% ethanol per 1ml of TRIZOL® reagent used for the initial homogenisation. The tube was vortexed and centrifuged at no more than 7500g for 5 minutes at 4°C to pellet the RNA. The supernatant was discarded, the RNA pellet was air-dried for 5 to 10 minutes and dissolved in an appropriate volume of RNase-free water. The tube was incubated at 55-60°C for 10 minutes to ensure that the pellet had completely dissolved.

2.3.3 RNeasy™ TOTAL RNA KIT

The RNeasy™ total RNA kit was obtained from QIAGEN Ltd. 350µl or 600µl (depending on the number of cells available) of lysis buffer RLT (10µl of β-mercaptoethanol was added to 1ml of lysis buffer RLT before use) was added to the pellet and transferred to a 1.5ml RNase-free eppendorf tube. The tube was centrifuged for 3

minutes to pellet any insoluble material, 1 volume (350µl or 600µl) of 70% ethanol was added to the lysate, mixed by pipetting and applied onto an RNeasy spin column. The column was spun for 15 seconds at 10000rpm and the flow-through was discarded. The column was next washed with 700µl of wash buffer RW1, followed by 500µl of wash buffer RPE (the concentrated RPE solution was diluted with 4 volumes of ethanol). The flow-through was discarded and the column was washed with 500µl of wash buffer RPE and spun for 2 minutes to dry the membrane. The RNA was eluted with 30 to 50µl of DEPC-treated water into a 1.5ml collection tube by spinning for 60 seconds at 8000g. The RNA was stored at -20°C.

2.4 VISUALISATION OF THE EXTRACTED RNA

To ascertain the quality of the extracted RNA, a 30ml 1% agarose checking gel was prepared by weighing out 0.45g of agarose into 22ml of dH_2O. After cooling the gel in a fume hood to 55°C, 5ml of (37%) formaldehyde and 3ml of 10X MOPS buffer [3-(N-morpholino) propanesulfonic acid] were added and the gel was immediately poured before setting. 1µl of RNA was mixed with 5µl of formamide, 1.65µl of (37%) formaldehyde, 1µl of 10X MOPS and 1.3µl of dH_2O. The preparation was heated to 55°C for 10 minutes, quenched on ice and 2µl of gel loading buffer (0.25% bromophenol blue, 0.25% xylene cyanol FF, 40% sucrose in water) was added. 1µg of *Escherichia coli* (*E. coli*) ribosomal RNA (Boehringer Mannheim) was also loaded as size markers and electrophoresed at 75V for 1 hour with 1X MOPS as running buffer. After electrophoresis, the gel was soaked in SDW for 1 hour to remove the formaldehyde and stained in 5µg/ml of ethidium bromide. The gel was destained overnight in water, after which it was viewed under UV light to check for the integrity of the ribosomal bands and to estimate the concentration of the RNA.

2.5 DETERMINATION OF NUCLEIC ACID CONCENTRATION

The concentration of DNA and RNA was determined by spectrophotometry (Pharmacia GeneQuant II RNA/DNA calculator). The optical density (OD) was measured at 260nm and 280nm. Each sample was diluted by adding 5µl of sample to 995µl of dH_2O. An OD reading of 1 corresponds to approximately 40µg/ml of RNA and 50µg/ml of DNA. A value of 2 for the ratio of the absorbances at 260nm to 280nm indicates good quality nucleic acid.

2.6 DESIGN, SYNTHESIS AND PURIFICATION OF OLIGONUCLEOTIDES

The primers required for genomic DNA and RT-PCR amplifications were designed with the aid of the OLIGO™ Version 3.4 (©Wojciech Rychlik; MEDPROBE). The primers were selected on the basis of the following criteria: (i) the average G+C content was around 50% and consisted of a random base distribution; (ii) the calculated T_m of both primers, where an addition of 2°C was used for A or T and 4°C for G or C was in the range of 50-75°C; and (iii) the length of the primers was about 20 to 30 bases, and palindromic sequences as well as a rich distribution of Cs or Gs at the 3' end were avoided. In addition, self-complementarity between the primers was avoided i.e. no internal duplex structures were allowed. The sequences of the primers used in this study are shown in Table 2.3.

All the primers were synthesised on the APPLIED BIOSYSTEMS 391 DNA SYNTHESISER. The primers required a deprotection stage to remove the oligonucleotide from the solid support and, hence, the column. This is done by using the double syringe method to flush the column several times (Sawadago and Van Dyke, 1991). This treatment removes the cyanoethyl phosphate protecting groups. 1ml of ammonia solution was pushed into the column over three 10-minute intervals and collected at the other end by another 1ml syringe and placed in a Nunc tube. An additional 1ml of ammonia solution was added and the tube was incubated overnight at 55°C to remove the protecting groups on

Table 2.3: Sequence of the primers used in this study showing their locations in the *HMB-S* gene.

Location	Primer	Sequence (5' → 3')
5' UTR→E1	019H(F)	tac ttt cca agc gga gcc ATG
E1	377H(F)	CAT GTC TGG TAA CGG CAA TGC GG
E3	750D(F)	GAT TCG CGT GGG TAC CCG CA
E4	Phe4-(R)	GAG GCT TTC AAT GTT GCC AC
E5	F161(F)	TTG CTA TGT CCA CCA CAG GGG
E6	Phe6+(F)	ATG CCC TGG AGA AGA ATG A
E7	R341(R)	CAG ATG GCT CCG ATG GTG AAG
E8	F356(F)	CTC ATG ATG CTG TTG TCT TTC
E8→E9	F413(F)	CAG AGA AGA GTG TGG TGG GAA
E9	R478(R)	GCG GGA ACT TTC TCT GCA GCT
E9→E10	F482(F)	TGG AGT TCA GGA GTA TTC GGG
E10	Phe10+(F)	GGA AGC TGG ACG AGC AGC AGG A
E11	Phe11-(R)	CAC AGC ATA CAT GCA TTC CTC A
E12	F666(F)	GGA AGT GCG AGC CAA GGA CCA
E12	F2CA(F)	TGC TTC GCT GCA TCG CTG AA
E13	Phe13-(R)	CTA CTG GCA CAC TGC AGC CTC
E13→E14	F808(F)	GCT ATG AAG GAT GGG CAA CTG
E14	749D(R)	CAT GGA TGG TAG CCT GCA TGG TCT C
E15	F951(F)	CAT CAC TGC TCG TAA CAT TCC
3' UTR	508D(R)	cat ctg tgc ccc aca aac cag
3' UTR	020H(R)	atg tag gca ctg gac agc agc
I7	I7(F)	aat acc agt gag ttg gca atc g
I8	I8(R)	tcc ctg cat ctt ctg ggc aca t

UTR, E, I: untranslated region, exon, intron; (F), (R): forward, reverse; lower case: intronic sequence; upper case: exonic sequence.

the exocyclic amines of the bases. The ammonia was evaporated off overnight until the solution was odourless. The OD was read at 260nm and an appropriate concentration of primers was used for PCR amplification.

2.7 PCR TECHNOLOGY

2.7.1 GENOMIC DNA-PCR

All PCR reactions were performed in 100µl which contained 1X PCR buffer [Perkin Elmer; 10mM Tris-HCl pH 8.3, 50mM KCl, 1.5mM $MgCl_2$, 0.001% (w/v) gelatin], 200µM of each dNTP and 0.5µM of each primer. The mixture was UV-irradiated for 10 minutes in the GRI-Amplirad to minimise contamination. 0.5-1µg of DNA and 2 units of Taq DNA polymerase were added, overlaid with 50µl of light mineral oil and subjected to PCR, which was performed either on the TECHNE PHC1 automated thermocycler or the HYBAID OmniGENE. 35 cycles of PCR was carried out, each comprising a denaturing step at 95°C for 1 minute, annealed at 60°C for 1 minute and primer extension at 72°C for 1.5 minutes. A final extension step of 10 minutes at 72°C was done after the completion of the 35 cycles. The PCR products were visualised on a 2% agarose gel containing ethidium bromide (10µg/ml) run in 1X TBE buffer (10X TBE: 108g Tris base, 55g boric acid, 9.3g EDTA and made up to 1 litre). The samples were prepared for loading by mixing 5µl of each sample with 1µl of loading buffer (30% glycerol, 0.25% bromophenol blue).

2.7.2 REVERSE TRANSCRIPTASE PCR (RT-PCR)

PCR requires a double-stranded template to function. RNA is not recognised by Taq DNA polymerase to be a suitable template for extension, therefore, it has to be converted to complementary DNA (cDNA) for Taq DNA polymerase to function in PCR. This is achieved by RT-PCR, a sensitive and rapid method for screening mRNA molecules for size differences caused by deletions, insertions or splicing defects which lead to exon skipping.

In this study, three different methods of priming were used to generate cDNA, as outlined below.

2.7.2.1 RT-PCR USING OLIGO(dT_{12-18})

0.5-1μg of RNA diluted in DEPC-treated water was heated at 95°C for 5 minutes for denaturation and immediately quenched on ice. 0.2μg of oligo (dT_{12-18}) (Pharmacia) was added to prime the reaction at 65°C for 10 minutes and immediately quenched on ice. 2μl of 0.1M DTT, 2μl of 250μM of each dNTP, 4μl of 5X reverse transcriptase buffer (Gibco BRL; 250mM Tris-HCl pH 8.3, 375mM KCl, 15mM $MgCl_2$) and 200 units of Moloney Murine Leukaemia Virus (MMLV) reverse transcriptase (Gibco BRL) were added. The volume was adjusted to 20μl and the mixture was incubated at 37°C for 1 hour. After reverse transcription, the mixture was heated at 95°C for 5 minutes to inactivate the enzyme (Sellner et al, 1992) and to disrupt the mRNA:cDNA hybrid.

2.7.2.2 RT-PCR USING DOWNSTREAM PRIMERS

The methodology is exactly the same as in section 2.7.2.1 except that 0.1μM of a gene-specific primer was used in place of oligo (dT_{12-18}).

2.7.2.3 RT-PCR USING RANDOM HEXAMERS

0.2μg of random hexamers (Gibco BRL) was used to prime the RT reaction instead of oligo (dT_{12-18}) or the downstream primer.

The cDNA synthesised by all three methods is ready for use as a template for PCR analysis. The entire reaction mixture was used for PCR which consisted of 1X PCR buffer, 0.5μM final concentration of each primer (except for the reverse primer used in section 2.7.2.2 where 0.4μM should be added) and 2 units of Taq DNA polymerase. The volume was adjusted to 100μl, vortexed briefly and centrifuged quickly. The mixture was overlaid with light mineral oil (SIGMA) and PCR was performed as in section 2.7.1. If a good yield

of PCR product was obtained as seen on an agarose gel (section 2.7.1), the remainder of the sample was electrophoresed, the appropriate band was excised from the gel and the cDNA was eluted by soaking for 24 hours in 100µl of TE buffer at -20°C. The supernatant, containing the eluted cDNA, was transferred to a fresh tube for further analyses.

2.7.3 GENERATION OF RADIOACTIVELY LABELLED PCR PRODUCTS

2.7.3.1 5' END-LABELLED PCR PRODUCTS

This was necessary to generate the radioactive probe for CCM analysis. This was achieved by 5' end-labelling, in a total volume of 10µl, 50pmol of each primer with 5-10µCi [γ-^{32}P]dATP (Amersham; 3000Ci/mmol; 10Ci/ml) in 1X T4 polynucleotide kinase buffer (Pharmacia; 10X buffer: 100mM Tris-acetate, 100mM magnesium acetate, 500mM potassium acetate) and 10 units of T4 polynucleotide kinase (Pharmacia). The mixture was incubated at 37°C for ½ -1 hour, after which it was ready for use in PCR.

2.7.3.2 INTERNALLY-LABELLED PCR PRODUCTS

To internally label PCR products for use in CCM analysis, PCR was carried out as before (section 2.7.1) except that the concentration of cold dCTP was reduced from 200µM to 6µM and 20pmol of [α-^{32}P]dCTP (60µCi) (3000Ci/mmol; 10Ci/ml) was added. The number of PCR cycles was reduced to prevent the generation of non-specific products by the misincorporation of alternate dNTPs as [dCTP] decreased.

2.7.4 ASYMMETRIC PCR

Asymmetric PCR was carried out to generate single-stranded templates for direct sequencing (Gyllensten and Erlich, 1988) as described by Mgone et al (1992). 10µl of the gel-eluted double-stranded PCR product was used as a template for asymmetric PCR, where one primer was 50 to 100 times less concentrated than the other (0.5 to 1.0pM) in a total volume of 100µl. The other components were the same as for symmetric PCR but the number of cycles was increased to 40 to 45 cycles as there is a linear increase in the

amount of PCR products due to the limiting concentration of one primer. The asymmetric PCR product yield was monitored by agarose gel electrophoresis as previously described (section 2.7.1), with a double-stranded control and a suitable marker [Gibco BRL; 1kb ladder (1.01μg/μl) or 100bp ladder (1μg/μl)] run at the same time to determine the position and amount of the single strands. If sufficient single strands are available, they can be used for direct sequencing.

2.7.5 PCR SCREENING FOR RECOMBINANT CLONES

To screen bacterial colonies for the presence of the successful ligation of the insert to the plasmid after transformation, primers were designed where one spanned across the plasmid-insert junction (pTrc 99A, Table 2.4) and the other was a primer corresponding to the sequence of the insert (R341, Table 2.3). This was to ensure that the colonies which gave positive PCR results are due to the successfully ligated fragments.

The PCR reaction mixture contained the same concentrations of PCR buffer, dNTPs and primers as before (section 2.7.1). The template used here was that of the bacterial colony, scraped (one-sixth of the colony) from the agar plate with a toothpick, and dipped into the PCR mixture to release the cells. The mixture was vortexed and heated at 99°C for 10 minutes. The tube was immediately placed on ice and Taq DNA polymerase was added. PCR was carried out as in section 2.7.1, except that the annealing temperature was 57°C. At the same time, the colonies were grown in a 5ml culture containing 100μg/ml of ampicillin. Colonies which yielded a positive PCR result had the plasmid DNA extracted from the corresponding cultures (section 2.9.6). Confirmatory steps of restriction mapping, CCM and direct sequencing were performed before proceeding to large scale preparation of plasmid DNA (section 2.9.7).

Table 2.4: Primers used for *in vitro* expression studies.

Primer	Orientation	Sequence (5'→3')
pTrc 99A	Forward	TTG ACA ATT AAT CAT CCG GC
p214	Forward	CGT ATA ATG TGT GGA ATT GTG
p383	Reverse	GCG TTC TGA TTT AAT CTG TAT C
PheN	Forward	GAG CCA TGG CTG GTA ACG GCA ATG C
PheH	Reverse	CCC AAG CTT CTG TGC CCC ACA AAC CA

pTrc 99A was used in conjunction with R341 (Table 2.3) to amplify across the plasmid-insert junction. p214 and p383 were used to amplify part of the region of the plasmid flanking the insert to enable sequencing across the plasmid-insert junctions at both the 5' and 3' ends. PheN and PheH incorporate the *Nco* I and *Hind* III restriction sites respectively to enable ligation of the insert into the expression vector for further analysis.

2.7.6 RESTRICTION ENDONUCLEASE ANALYSIS OF RESTRICTION SITES

2.7.6.1 DETERMINATION OF RESTRICTION SITES

The flanking region of any sequence alteration was analysed on the GCG Package for any creation or destruction of restriction sites. Table 3.3 lists the various restriction enzymes employed in this study for the confirmation of a nucleotide change or to determine the carrier status of a relative of an affected individual where possible.

2.7.6.2 ETHANOL PRECIPITATION OF PCR PRODUCTS

To prepare the PCR products for restriction enzyme digestion, ethanol precipitation was carried out to purify the products. 40µl of PCR product was mixed with one-tenth its volume of 3M sodium acetate (pH 5.2) and 2 volumes of absolute ethanol. The tube was kept at -20°C for 10 minutes, spun for 10 minutes and washed in 70% ethanol for 5 minutes. The resulting pellet was air-dried and resuspended in 10µl of dH_2O. 2µl of the appropriate enzyme buffer and 1 unit of the restriction enzyme was added in a total volume of 20µl. The tube was incubated at 37°C for at least 1 hour. The digests were checked on an agarose gel to ensure that the digestion was successful and that it had gone to completion.

2.8 MUTATION DETECTION TECHNIQUES

In this project, three different mutation detection methods were employed for the initial screening and characterisation of mutations. These were: (i) heteroduplex analysis (HA); (ii) chemical cleavage mismatch (CCM) analysis; and (iii) direct sequencing. All three techniques are PCR-based.

2.8.1 HETERODUPLEX ANALYSIS (HA)

Heteroduplex analysis (HA) was performed using the MDE Heteroduplex Kit from FLOWGEN. The kit consists of 500ml of 2X MDE gel solution, 60µl of a 350bp heteroduplex control in triple dye loading buffer and 1.1ml of 6X triple dye loading buffer. The PCR conditions for the products generated for subsequent HA must be fully optimised to ensure that no non-specific products are present which may interfere with the heteroduplex bands.

At the end of the final PCR cycle, EDTA was added to a final concentration of 5mM (1µl of 0.5M EDTA per 100µl reaction) to inactivate the DNA polymerase. 5µl of each sample was heated at 95°C for 3 minutes, and cooled slowly to 37°C over a period of 30 minutes for the formation of heteroduplexes. 1µl of the triple dye loading buffer was added to each sample tube and mixed well.

The gel was made using the Bio-Rad Sequencing Cell apparatus with 1mm spacers and comb. The gel was composed of 50ml of 2X MDE Gel Solution, 6ml of 10X TBE buffer, 15g of urea and made up to 100ml with dH_2O. 400µl of freshly prepared 10% ammonium persulphate (APS) and 40µl of N,N,N,N-tetramethylethylenediamine (TEMED) were added to facilitate the polymerisation of the gel. 0.6X TBE buffer was used as the running buffer. The samples, together with the appropriate DNA size markers and the 350bp control heteroduplex, were loaded onto the gel and electrophoresed overnight for 16-20 hours at a constant voltage of 20V/cm of the gel. For a 50cm gel, a constant voltage of 1000V was used.

2.8.1.1 SILVER STAINING OF POLYACRYLAMIDE GELS

After the gel run, silver staining was carried out. For all the steps, the gel should be kept totally submerged in solution with gentle shaking. The gel was fixed in a solution of 10% ethanol/0.5% acetic acid for 10 minutes. The solution was discarded and the gel was stained in 0.1% silver nitrate for 15 minutes, after which it was briefly rinsed twice in ddH_2O. The bands were developed in a solution of 1.5% sodium hydroxide/0.1% formaldehyde for 20 minutes, or until the bands were of the desired intensity. The gel was

fixed in 0.75% sodium carbonate for 10 minutes, sealed in a plastic bag and photographed immediately.

2.8.2 CHEMICAL CLEAVAGE MISMATCH (CCM) ANALYSIS

This technique is able to detect point mutations, small deletions and insertions. It is based on the cleavage of mismatched bases which are modified chemically (Cotton et al, 1988).

There are five major steps in this method: (i) preparation of the labelled DNA probe and unlabelled test DNA; (ii) formation of heteroduplexes; (iii) chemical modification; (iv) chemical cleavage of mismatched base pairs; and (v) electrophoresis and autoradiography.

(i) PREPARATION OF LABELLED DNA PROBE AND UNLABELLED TEST DNA

The probe and test samples required in this step were generated by PCR amplification as described in sections 2.7.3 and 2.7.1 respectively. The PCR products were purified to remove ethidium bromide, primers and unincorporated dNTPs, either directly or after electrophoresis in a low melting point agarose gel (Nusieve from FMC Bioproducts), depending on whether non-specific bands were present. For this, the GENECLEAN II® KIT (Bio 101) was used. To the PCR product or gel slice (excised from the low melting point agarose gel) was added 3 volumes of 6M NaI solution, the latter was incubated at 45-55°C until the gel slice was dissolved. 10µl of the glassmilk suspension was added to allow the DNA to adsorb onto the glassmilk matrix. The tubes were vortexed and kept on ice for 15 minutes, after which they were centrifuged for 5 seconds. The glassmilk-DNA complex was washed three times with the NEW WASH buffer (diluted as instructed by the manufacturer). The pellet was resuspended in 15-20µl of TE buffer (section 2.2) and incubated at 65°C for 10 minutes to elute the DNA. The mixture was centrifuged for 10 minutes and the supernatant was transferred to a fresh tube. Approximately 100ng/µl of DNA was typically recovered. The procedure was repeated for the radiolabelled probe.

(ii) FORMATION OF HETERODUPLEXES

Two sets of labelled 1.5ml siliconised eppendorf tubes were prepared for each sample, one for the hydroxylamine reaction and the other for the osmium tetroxide reaction. 10-15ng of labelled probe (10^5 cpm) was mixed with a 10-fold excess of unlabelled test DNA (100-150ng), 18µl of 10X hybridisation buffer (1M Tris-HCl pH 8.0, 3M NaCl) and the total reaction volume adjusted to 180µl with TE buffer. 100µl of light mineral oil was overlaid onto the mixture which was boiled for 5 minutes and immediately transferred to a water-bath set at 65°C for a 5- to 16-hour incubation for the formation of heteroduplexes. 750µl of precipitation mix (63mM sodium acetate, 20µM EDTA, 80% ethanol) and 3µl of mussel glycogen (Boehringer Mannheim; 20mg/ml) were added and the mixture was kept on dry ice for 15 minutes, after which it was spun at 12000rpm for 10 minutes. The pellet was washed with 70% ethanol, air-dried and resuspended in 7µl of TE buffer.

(iii) CHEMICAL MODIFICATION

HYDROXYLAMINE REACTION

20µl of hydroxylamine solution [1.39g of hydroxylamine hydrochloride (BDH), 1.6ml of pre-warmed (37°C) ddH_2O, adjusted to pH 6.0 with approximately 1.6ml of diethylamine] was added to the designated tubes containing 7µl of the heteroduplex and incubated at 37°C for 2 hours.

OSMIUM TETROXIDE REACTION

18µl of osmium tetroxide solution [1.5µl of 4% osmium tetroxide solution (Aldrich) which is stable at 4°C for up to 3 months, 6.75µl of pyridine (BDH), 154µl of TE buffer] was added to the other set of tubes designated for the osmium tetroxide reaction. The tubes were incubated at 37°C for 2 hours.

Both the hydroxylamine and osmium tetroxide reactions were stopped by adding 750µl of precipitation mix and kept on dry ice for 10 minutes. The tubes were centrifuged for 10 minutes at 12000rpm and the pellet was washed with 70% ethanol and air-dried.

(iv) PIPERIDINE CLEAVAGE OF MISMATCHED BASE PAIRS

50µl of a 10% piperidine solution (Fluka) was added to each pellet, vortexed for 30 seconds and incubated at 90°C for 30 minutes. The reaction was stopped with the addition of 750µl of precipitation mix and incubated on dry ice for 10 minutes. The tube was spun as before and the pellet was resuspended in 10µl of formamide loading buffer (95% formamide, 10mM EDTA, 10mg/ml bromophenol blue, 10mg/ml xylene cyanol FF).

(v) GEL ELECTROPHORESIS AND AUTORADIOGRAPHY

The samples were heated at 95°C for 5 minutes, snap-chilled on ice and subjected to electrophoresis on a 0.4mm-thick 8% denaturing polyacrylamide gel containing 7M urea as denaturant, prepared by using the sequencing gel apparatus (Bio-Rad; Sequigen) at 35W until the xylene cyanol is at the bottom of the gel. The gel solution was made up of 12ml of a 40% acrylamide:bisacrylamide (19:1) (NBL Ltd.), 25.3g urea, 6ml of 10X TBE buffer and dH_2O to adjust the volume to 60ml. 100µl of a freshly prepared 25% APS and TEMED were added. 1X TBE buffer was used as the running buffer. A labelled DNA marker (Gibco BRL; 1kb or 100bp) was run together with the samples to allow the determination of the size of the cleavage products. The gel was transferred onto a piece of 3mm Whatman paper, cut to the size of the gel, overlaid with SARAN WRAP™ and autoradiographed using KODAK X-Omat AR film for 12 to 16 hours.

The labelled DNA marker was prepared as follows:

1µl of the appropriate DNA marker was mixed with 1µl of 10X T4 polynucleotide kinase buffer (Pharmacia), 7µl of dH_2O, 5 units of T4 polynucleotide kinase (Pharmacia) and 5µCi [γ-^{32}P]dATP. The mixture was incubated at 37°C for ½ - 1 hour and the reaction was

stopped with 50µl of formamide loading buffer [section 2.8.2(iv)]. It was denatured together with the samples for CCM analysis prior to loading.

2.8.3 DIRECT SEQUENCING

Direct sequencing of the HMB-S DNA or cDNA was carried out using the dideoxy chain terminating method (Sanger et al, 1977) using the Sequenase™ Version 2.0 sequencing kit (Amersham). Single-stranded templates required for sequencing were generated by asymmetric PCR (section 2.7.4). The asymmetric PCR products were purified by adding an equal volume of 4M ammonium acetate and 2 volumes of isopropanol. The mixture was incubated at room temperature for 10 minutes, centrifuged for 10 minutes, washed in 70% ethanol and resuspended in 14µl of dH_2O. 7µl was used for each sequencing reaction.

Sequencing was performed according to the manufacturer's instructions. 7µl of the single-stranded template was mixed with 2µl of 5X Sequenase™ reaction buffer (to give a final concentration of 40mM Tris-HCl pH 7.5, 20mM $MgCl_2$, 50mM NaCl) and 1pmol of the appropriate sequencing primer. The mixture was annealed by incubating at 65°C for 2 minutes and then allowed to cool slowly to less than 35°C over a period of 30 minutes. [While cooling, four tubes (for each of the termination mixtures) for each sample were labelled, 2.5µl of each termination mixture (containing 80µM of each dNTP and 8µM of each ddNTP) was transferred to the appropriate tube and pre-warmed at 42°C]. The mixture was immediately placed on ice and to it was added 1µl of 0.1M DTT, 2µl of a 1:5 dilution of labelling mix (7.5µM of each dNTP except dATP), 0.5µl of [α-^{35}S]dATP (1000Ci/mmol; 10µCi/µl), 1µl of Mn buffer (0.15M sodium isocitrate, 0.1M $MnCl_2$) if sequences close to the primer are to be read, and 2µl (1.5 units) of a 1:8 dilution of T7 DNA polymerase [25µl of Sequenase™ Version 2.0, 13 units/µl; diluted with 25µl of inorganic pyrophosphatase (5 units/ml in 10mM Tris-HCl pH 7.5, 0.1mM EDTA, 50% glycerol) and 150µl of glycerol enzyme dilution buffer (20mM Tris-HCl pH 7.5, 2mM DTT, 0.1mM EDTA, 50% glycerol)]. The mixture was incubated at room temperature for 2 to 5 minutes. 3.5µl of the labelling reaction was then added to each termination mix at

42°C and incubated at 42°C for a further 5 minutes, after which 4µl of formamide-dye stop solution (95% formamide, 20mM EDTA, 0.05% bromophenol blue, 0.05% xylene cyanol FF) was added to terminate the reaction. The samples can be stored at -20°C until required.

Because the Sequenase™ enzyme was diluted in glycerol-containing buffer, the buffer used in the polyacrylamide gel should be a glycerol tolerant gel buffer (20X TTE: 216g Tris base, 72g taurine, 4g $Na_2EDTA.2H_2O$ and made up to 1 litre). The gel used to run the sequencing reactions was an 8% polyacrylamide gel containing 7M urea with 1X TTE buffer as the gel and running buffers. The gel was pre-warmed to 50°C before sample loading. Once the gel had reached the desired temperature (as determined by a temperature indicator), the samples were heated at 75°C for 2 minutes, snap-chilled on ice and loaded onto the gel. The gel was electrophoresed at 50°C for 2 to 4 hours, after which it was fixed in 10% methanol/10% acetic acid for 10 minutes, dried in a vacuum gel drier (Bio-Rad) at 80°C for at least 30 minutes and exposed for 16 to 18 hours to KODAK X-Omat AR film in a cassette using intensifying screens.

2.9 TECHNIQUES IN RECOMBINANT DNA

2.9.1 RESTRICTION DIGESTS OF DOUBLE-STRANDED DNA AND VECTOR

The primers used for the amplification of the fragments to be cloned have each been modified to contain a restriction site, *Nco* I for the 5' primer, and *Hind* III for the 3' primer (primers PheN and PheH respectively, Table 2.4) for ligation with the expression vector, pTrc 99A (Pharmacia). The amplified products were visualised on a 1% agarose gel and ethanol precipitated as in section 2.7.6.2. Both restriction enzymes were obtained from Pharmacia and the digestion was carried out on both the DNA fragments and vector for 2 hours at 37°C using 5 units of each enzyme.

2.9.2 LIGATION

The digested fragments and vector were added in a molar ratio of 3:1 of insert to vector. In a 0.5ml sterile eppendorf tube, 100-200ng of the linearised vector was added to approximately one-tenth of the purified digested insert, 1µl of 10X ligase buffer (Promega; 300mM Tris-HCl pH 7.8, 100mM $MgCl_2$, 100mM DTT, 5mM ATP) and 0.1 to 1 Weiss unit of T4 DNA ligase. The volume was brought to 10µl with dH_2O. The reaction mixture was incubated at 15°C for 16 hours as recommended by the manufacturer.

2.9.3 TRANSFORMATION OF COMPETENT CELLS

The strain of *E. coli* competent cells used was JM109 (Promega). 5µl of the ligation mixture was added to 100µl of ice-cold JM109 competent cells and incubated for a further 30 minutes on ice. The mixture was given a heat-shock treatment at 42°C for 1 minute and then chilled on ice for 2 minutes. 500µl of L-broth medium (10g bactotryptone, 5g yeast extract, 10g NaCl, pH 7.5) was added followed by an incubation at 37°C for 1 hour in an orbital shaker, after which 100µl of the mixture was spread onto agar plates (L-broth medium + 1.5% agar) containing 100µg/ml of ampicillin. The plates were incubated at room temperature for 5 minutes before incubation overnight in a 37°C incubator.

2.9.4 SCREENING FOR PUTATIVE RECOMBINANT CLONES

The colonies formed on the agar plates were screened for the presence of the insert as described in section 2.7.5.

2.9.5 PREPARATION OF GLYCEROL STOCKS

Colonies which did not contain any extraneous misincorporations were immortalised as glycerol stocks. This was done by growing the colony in 10ml of L-broth medium containing 100µg/ml of ampicillin overnight at 37°C with shaking, after which a 15% glycerol stock was made and stored at -70°C.

2.9.6 MINIPREPARATION OF PLASMID DNA

Three different methods were used in the minipreparation of plasmid DNA: (i) alkali-lysis method; (ii) QIAprep™ Plasmid Kit; and (iii) INSTA-MINI-PREP™ kit.

(i) ALKALI-LYSIS METHOD

Minipreparation of plasmid DNA was carried out by the alkali-lysis method according to Sambrook et al (1989). A single bacterial colony was inoculated in 5ml of L-broth medium containing 100μg/ml of ampicillin overnight at 37°C with shaking. 1.5ml of this culture was spun for 15 seconds to pellet the cells which were resuspended in 100μl of solution A (25mM Tris-HCl pH 8.0, 10mM EDTA, 50mM glucose). 2mg/ml of lysozyme was freshly added. 200μl of solution B (0.2M NaOH, 1% SDS) was next added and the tube was mixed by rocking. The tube was left on ice for 5 minutes. 150μl of cold solution C (5M potassium acetate, 11.5% acetic acid, pH 4.8) was added and the contents were mixed by inversion, vortexed quickly and held on ice for 5 minutes to precipitate the DNA and proteins. The tube was centrifuged at maximum speed for 1 minute and the supernatant was transferred to a fresh 1.5ml eppendorf tube where one-tenth its volume of 3M sodium acetate (pH 5.0) and an equal volume of isopropanol were added and the tube was incubated at -40°C for 1½ to 2 hours. The tube was centrifuged for 10 to 15 minutes to precipitate the RNA and plasmid DNA and the pellet was washed with 70% ethanol. The pellet was resuspended in 50μl of TE buffer (10mM Tris, 0.25mM EDTA, pH 7.4) and treated with RNase at a final concentration of 10μg/ml for 30 minutes at 37°C.

(ii) QIAprep™ PLASMID KIT

The kit used was obtained from QIAGEN Ltd. and the protocol was carried out according to the manufacturer's instructions. A single bacterial colony was inoculated in 1-5ml of L-broth medium containing 100μg/ml of ampicillin overnight at 37°C with shaking. The culture was centrifuged for 15 minutes to pellet the cells and resuspended in 250μl of buffer P1 (containing RNase A and stored at 4°C). Following this, 250μl of buffer P2 was

added and the tube was gently inverted 4 to 6 times to mix. 350µl of buffer N3 was next added and the tube was immediately inverted gently 4 to 6 times and centrifuged for 10 minutes. During centrifugation, a QIAprep spin column placed in a 2ml collection tube was prepared. The supernatant was applied to the column, which was then centrifuged for 30 to 60 seconds and the flow-through was discarded. The QIAprep spin column was washed with 0.5ml of buffer PB, centrifuged for 30 to 60 seconds, and the flow-through was discarded. The column was next washed with 0.75ml of buffer PE and spun for 30 to 60 seconds. The flow-through was discarded and the column was spun for a further 1 minute to remove residual wash buffer. The column was placed in a clean 1.5ml eppendorf tube and the DNA was eluted with 50µl of 10mM Tris-HCl (pH 8.5) or H_2O. The column was kept at room temperature for 1 minute, centrifuged for 1 minute and 1µl was used for checking on a 1% agarose gel.

(iii) INSTA-MINI-PREP™ KIT

This kit was purchased from 5 Prime→3 Prime, Inc. and minipreparation of plasmid DNA was performed as outlined by the manufacturer. A single bacterial colony was inoculated in 1-3ml of L-broth medium containing 100µg/ml of ampicillin overnight at 37°C with shaking. The unopened INSTA-MINI-PREP™ tube was centrifuged at 12000g for 10 to 30 seconds to pellet the INSTA-MINI-PREP™ gel. 1.5ml of the overnight bacterial culture was transferred to a 1.5ml eppendorf tube and centrifuged at 12000 to 16000g for 30 seconds. This step can be repeated if necessary. 50-100µl of either molecular biology grade water, TE (10mM Tris-HCl, 1mM EDTA, pH 8.0) or STE (100mM NaCl, 10mM Tris-HCl, 1mM EDTA, pH 8.0) was added to the pellet and vortexed to fully resuspend the bacterial cells. The entire suspension was transferred to the pre-spun INSTA-MINI-PREP™ tube by pipetting, followed by the addition of 300µl of shaken PCI solution (50 Phenol: 49 Chloroform: 1 Isoamyl alcohol) and mixed thoroughly by repeated vigorous inversion. The INSTA-MINI-PREP™ tube was centrifuged at 12000g for 30 seconds, after which 300µl of CI solution (49 Chloroform: 1 Isoamyl alcohol) was added and the two liquid upper phases were mixed by repeated vigorous inversion (do not

vortex or allow the lower organic phase to mix with the upper aqueous phase). The contents were centrifuged at 12000g for 30 seconds and the topmost aqueous phase was transferred to a fresh eppendorf tube.

2.9.7 LARGE SCALE PREPARATION OF PLASMID DNA

Large scale plasmid DNA preparation was achieved with the QIAGEN PLASMID MAXI KIT. A single bacterial colony was inoculated in 10ml of L-broth medium containing 100μg/ml of ampicillin. The culture volume was increased to 500ml and inoculated with 1ml of the overnight culture. Another overnight incubation at 37°C with shaking was carried out and the cells were harvested by centrifugation at 6000rpm for 15 minutes at 4°C in a Sorvall rotor. All traces of supernatant were removed by inverting the tube before proceeding to the next step.

The bacterial pellet was resuspended in 10ml of buffer P1 (100μg/ml RNase A, 50mM Tris-HCl, 10mM EDTA, pH 8.0). 10ml of buffer P2 (200mM NaOH, 1% SDS) was next added, mixed gently by inverting 4 to 6 times and incubated at room temperature for 5 minutes. 10ml of chilled buffer P3 (3.0M potassium acetate, pH 5.5) was next added and the contents were mixed immediately but gently before incubating on ice for 20 minutes. The mixture was centrifuged at 15000rpm for 30 minutes at 4°C and the supernatant was discarded. During the centrifugation, a QIAGEN-tip 500 was equilibrated with 10ml of buffer QBT (750mM NaCl, 50mM MOPS, 15% ethanol, pH 7.0, 0.015% Triton X-100) and the column was allowed to empty by gravity flow. The supernatant was applied to the column which was allowed to empty as before. The column was washed twice, each with 30ml of buffer QC (1.0M NaCl, 50mM MOPS, 15% ethanol, pH 7.0). The plasmid DNA was eluted with 15ml of buffer QF (1.25M NaCl, 50mM Tris-HCl, 15% ethanol, pH 8.5) and precipitated with 0.7 volumes of isopropanol. The mixture was immediately centrifuged at 6000rpm for 30 minutes at 4°C, and the pellet was washed in 15ml of cold 70% ethanol. The pellet was dissolved in 500μl of TE buffer.

2.9.8 RELEASING THE DNA FRAGMENT FROM THE VECTOR

The isolated plasmid DNA from sections 2.9.6 and 2.9.7 was digested with *Nco* I and *Hind* III to release the insert from the plasmid. Further manipulations can be performed on the released insert.

2.9.9 CHECKING THE INTEGRITY OF THE INSERT

In order to rule out a base change as being due to misincorporation by Taq DNA polymerase, several independent clones were screened by CCM (section 2.8.2) and direct sequencing (section 2.8.3).

2.10 *IN VITRO* GENE EXPRESSION STUDIES: PROTEIN ANALYSIS

2.10.1 *IN VITRO* GENE EXPRESSION

Colonies which have maintained their integrity, as determined by CCM and direct sequencing, were suitable for *in vitro* gene expression. The protocol followed was exactly as described for the pTrc 99A expression kit (Pharmacia). A loopful of bacterial cells (stored as a glycerol stock) was inoculated in 5ml of L-broth medium containing 100μg/ml of ampicillin overnight at 37°C with shaking. 500μl of the overnight culture was inoculated in 50ml of L-broth medium containing 100μg/ml of ampicillin and the cells were grown at 37°C to mid-log phase i.e. A_{600} = 0.6→1.0. Expression was induced with the addition of isopropylthio-β-D-galactopyranoside (IPTG; Gibco BRL) to a final concentration of 1.0mM. The culture was spun at 1500g for 10 minutes at 4°C and the pelleted cells were washed with cold 10mmol/l Tris-HCl (pH 8.0). This was repeated twice. The cells were resuspended in 10mmol/l Tris-HCl (pH 8.0) and sonicated using the MSE Soniprep, after which the tube was centrifuged at 12000g for 30 minutes and the supernatant was filtered through a 0.22μm filter unit.

2.10.2 QUANTIFYING PROTEINS

This was performed according to the method of Lowry et al (1951). A fresh solution of 98% of solution A (2% Na_2CO_3 in 0.1M NaOH), 1% of solution B1 (1% $CuSO_4$ in ddH_2O) and 1% of solution B2 (2% Na K tartrate in ddH_2O) was prepared, and a 1:1 dilution of Folins Reagent (BDH) in ddH_2O was made. At the same time, a 500µg/ml standard solution of bovine serum albumin (BSA; SIGMA) was diluted down in distilled water to give a 500µg to 50µg standard curve. To 800µl of the A+B1+B2 solution was added 80µl of sample, which was then left at room temperature for 10 minutes. 80µl of the diluted Folins Reagent was next added, while mixing, and the tube was left in the dark for 45 minutes, after which the absorbance at 750nm was read and compared with the standard curve.

2.10.3 PRODUCTION OF POLYCLONAL ANTIBODIES

The antibody used in this study was custom-made by GENOSYS BIOTECH, Inc. using rabbits as hosts and following a 90-day standard immunisation schedule. The choice of the peptide sequence to raise the antisera was based on the following criteria set out by GENOSYS BIOTECH, Inc.: (i) N- or C-terminal sequences were preferred, although internal sequences which represent regions that are surface-exposed in the intact protein are acceptable; (ii) 10 to 15 amino acids with less than 6 hydrophobic residues and no more than 4 adjacent hydrophobic residues should be selected. If conditions (i) and (ii) are adhered to, the peptides are likely to be soluble in aqueous solvents, easier to couple to carriers and more likely to be immunogenic.

The choice of peptide sequence used in this project was made by comparing the sequences of the human and *E. coli*, and the sequence of the peptide used to raise the antibody was **HTAMKDGQLYLTGG**, corresponding to the amino acids His-Thr-Ala-Met-Lys-Asp-Gly-Gln-Leu-Tyr-Leu-Thr-Gly-Gly. This sequence (i) fulfilled the criteria set out by GENOSYS BIOTECH, Inc.; and (ii) was found to contain the least sequence homology between the two species.

2.10.4 PURIFICATION OF ANTISERA

The antisera supplied by GENOSYS BIOTECH, Inc. were unpurified and contained the different isotypes of immunoglobulins (Ig) i.e. IgA, IgG, IgM. The fraction of interest is the IgG isotype, therefore, IgG has to be separated from the total Ig content and this has been achieved using the Protein A Sepharose® 4 Fast Flow (Pharmacia).

The 20% ethanol solution in which the gel was packed was removed by transferring the gel to a column and allowing the ethanol to flow through until just above the level of the beads. Ten column volumes (i.e. 10 X 1ml) of 0.1mol/l sodium phosphate (pH 7.0) was passed through the column, after which the unpurified antiserum was applied gently to the column. 1ml fractions were immediately collected and their absorbances at 280nm were read and graphically represented. The absorbance should gradually increase and reach a maximum level for approximately 5 X 1ml fractions before decreasing to baseline level. This indicates that all proteins, apart from IgG which has bound to the Protein A beads, have been eluted from the column. Once this has been achieved, the elution buffer (0.1M glycine-HCl, pH 3.0) was applied to the column. Again, 1ml fractions were immediately collected into tubes containing 2mol/l Tris-HCl (pH 8.0) to neutralise the pH and their absorbances were read at 280nm. A similar profile should be observed as before since the IgG is now eluted from the column. Fractions showing a high protein content, as determined by the absorbances at 280nm, were combined and dialysed overnight at 4°C in phosphate buffered saline (0.05mol/l potassium phosphate, 0.1mol/l NaCl, pH 7.4). After dialysis, the amount of IgG present was quantified as in section 2.10.2.

2.10.5 ANALYSIS OF *IN VITRO* GENE EXPRESSION

In order to determine if *in vitro* gene expression was successful, the total proteins present in the lysate were separated by sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) according to the method described by Laemmli U.K. (1970) followed by Western blotting with a specific antibody.

2.10.5.1 SDS-PAGE

The Protean® II xi Cell (Bio-Rad) was assembled according to the manufacturer's instructions. A 10-12% separating gel [10-12ml of 30% acrylamide/bisacrylamide (29:1), 7.8ml of solution A (1.5M Tris, 0.4% SDS, pH 8.8) and made up to 30ml with dH_2O] was first prepared, and 300µl of freshly prepared 10% APS and 30µl of TEMED were added just before pouring the gel. The gel was poured to a level such that the space above the separating gel corresponded to the length of the teeth of the comb and an additional 1cm. The spacers and comb used were 0.75mm thick. The gel was overlaid with dH_2O to exclude air which slows down the polymerisation process. After polymerisation (about 1 hour), the overlay was poured off and a 5% stacking gel [1.6ml of 30% acrylamide/bisacrylamide (29:1), 2.6ml of solution B (0.5M Tris, 0.4% SDS, pH 6.8) and made up to 10ml] was poured over the separating gel and the comb was immediately inserted. After polymerisation, the comb was removed and the wells were cleaned with the running buffer (10X TGS: 250mM Tris, 2.5M glycine, 1% SDS). 100µg of each sample and the Rainbow™ coloured protein molecular weight markers (Amersham) were diluted in the ratio 5 parts 2X gel loading buffer: 4 parts sample/marker: 1 part 1M DTT (2X gel loading buffer: 1.2ml of 1M Tris-HCl pH 6.8, 2ml of glycerol, 4ml of 10% SDS, 0.01% bromophenol). The samples and marker were heated at 90°C for 5 minutes prior to loading. The gel was electrophoresed for 4 to 6 hours at a constant voltage of 250V. The apparatus was disassembled and Western blotting was performed.

2.10.5.2 WESTERN BLOTTING

The gel was pre-equilibrated in cold transfer buffer (0.025M Tris, 0.192M glycine, 20% methanol) for 15 minutes. A sheet of HYBOND ECL™ membrane (Amersham) was initially soaked in dH_2O and then equilibrated in transfer buffer for 15 minutes. The electroblotting cassette (Bio-Rad) was assembled according to the manufacturer's instructions. Two pieces of filter paper and the fibre pads were soaked in transfer buffer and a gel-membrane sandwich was set up in the following order:

cathode side →fibre →filter →gel →HYBOND →filter →fibre→anode side of
of holder pad paper ECL™ paper pad holder

The Trans Blot™ apparatus (Bio-Rad) was assembled and transfer was carried out overnight at 4°C at a constant voltage of 30V. After transfer, the filter was subjected to immunodetection while the gel was stained with Coomassie Brilliant Blue [50% methanol, 10% acetic acid, 0.05% Coomassie Brilliant Blue (R250)] to assess the transfer efficiency.

2.10.5.3 IMMUNODETECTION

The filter was rinsed briefly in Tris buffered saline (TBS: 20mM Tris, 137mM NaCl, pH 7.6), after which it was blocked in blocking buffer [TBS, 0.2% (v/v) Tween 20, 5% (w/v) BSA, pH 7.6] for 2 hours. The blocking buffer was discarded and the filter was incubated overnight with shaking in the wash buffer [TBS-T: TBS, 0.1% (v/v) Tween 20] containing the primary antibody (obtained from section 2.10.4) diluted 1:10000 to 1:5000 and 0.2% (w/v) BSA. The solution was then discarded and the filter was washed 8 X 15 minutes with TBS-T. The secondary antibody [horseradish peroxidase conjugated goat anti-rabbit IgG (whole molecule); SIGMA] was added at a dilution of 1:40000 and the filter was incubated at room temperature for 2 hours with shaking, after which it was washed 8 X 15 minutes as before.

The detection system employed was based on chemiluminescence using the ECL™ Western Blotting Detection Reagents (Amersham) and performed according to the manufacturer's instructions. Equal volumes of solutions 1 and 2 were placed in two separate Universals (the final volume required is 0.125ml/cm^2 filter). Excess wash buffer was drained off from the filter and placed on a piece of SARAN WRAP™ with the protein side up. The two detection solutions were mixed and applied to the protein side of the filter, ensuring that the filter does not become uncovered. The filter was incubated for 1 minute at room temperature without agitation. The excess detection solutions were drained off from the filter which was wrapped in SARAN WRAP™ with the protein side down to form an envelope, avoiding pressure on the filter. The membrane was placed in a film

cassette with the protein side up and exposed to a sheet of HYPERFILM ECL (Amersham), initially for 15 seconds, then exposed to another sheet of film accordingly.

2.10.6 HIGH PERFORMANCE LIQUID CHROMATOGRAPHY (HPLC)

High performance liquid chromatography (HPLC) is an analytical technique used in the separation of macromolecules and small molecules by passing them through a column containing a separation matrix under high pressure. The desired fractions collected can then be subjected to further investigations.

HPLC has several advantages over other forms of liquid chromatography (LC): (i) the HPLC column can be used repeatedly without regeneration (Hamilton and Sewell, 1982); (ii) less hands-on time required, hence less dependence on the operator's skill; (iii) a higher resolution is achieved with greater reproducibility; (iv) automation and quantitation take control over operating conditions such as column temperature, inlet pressure, flow rate and separation time; and (v) shorter analysis times are possible.

The HPLC system used in this project consisted of: (i) LKB 2152 HPLC controller; (ii) 2158 Uvicord SD; (iii) LKB 2150 HPLC pump; (iv) LKB 2210 Recorder (2-channel); and (v) Mono Q® anion exchange column (HR 5/5, 0.5 X 5cm, Pharmacia). The column was equilibrated with buffer A (10mmol/l Tris-HCl, pH 8.0). 100µl of sample was injected into the system and the run was started. The elution buffer B was 1M NaCl in buffer A. The gradient elution programme (Table 2.5) was initiated from the controller with the profile being monitored at a wavelength of 280nm and visualised on the chart recorder at a speed of 5mm/minute. The flow rate of the sample was set at 1ml/minute and the sensitivity was set at 0.2 absorbance units full scale (AUFS).

2.10.7 ISOELECTRIC FOCUSING (IEF)

Isoelectric focusing (IEF) is an electrophoretic technique used in the separation of a mixture of proteins based on the differences in their net charge. Proteins, being amphoteric (zwitterionic) compounds, are either positively or negatively charged as they contain both acidic and basic residues. The charge of a protein is derived from the pH-dependent

Table 2.5: Gradient elution programme for the separation of the human and *E. coli* HMB-S proteins.

Time (min)	Flow rate (ml/min)	% Buffer A	% Buffer B
0.0	1.0	100	0
6.0	1.0	99	1
26.0	1.0	85	15
28.0	1.0	0	100
30.0	1.0	100	0

ionisation of the amino acid side-chain carboxyl and amino groups ($-COOH \leftrightarrow COO^- + H^+$; $-NH_2 + H^+ \leftrightarrow NH_3^+$).

Each protein has a pH at which its net charge is zero. This pH is known as its isoelectric point (pI) and is the point where the number of positive charges is equal to the number of negative charges. Factors which affect the pI of a protein are ionic strength, temperature and native or denaturing conditions. In IEF, the proteins migrate through a pH gradient until they reach their pI where migration ceases. If the local environment is below the pI of the protein, the protein will migrate towards the negative electrode (cathode). If the protein is in a solution which has a pH higher than its pI, it will migrate towards the positive electrode (anode).

(i) POLYACRYLAMIDE GEL PREPARATION

The gel was prepared by adding 10ml of 29.1% acrylamide, 10ml of 0.9% N,N-methylenebisacrylamide, 7ml of 87% (v/v) glycerol and 1.5ml of ampholine (Pharmacia; pH 3.5-9.5) to 31.5ml of ddH_2O. The solution was degassed for 10 minutes and 1.5ml of freshly prepared 1% APS and 50µl of TEMED were added.

(ii) CONSTRUCTION OF GEL MOULD

A 3mm thick glass plate (125mm X 260mm) was bordered by a 2mm thick rubber gasket and a second glass plate was placed on top to form the gel mould, which was held together by twelve clamps. A portion of the gel mould gasket was eased out at one corner and the freshly prepared gel solution was carefully introduced into the mould with a 60ml syringe attached to a short length of rubber tubing. The gasket was reclosed and clamped. After polymerisation (about 1½ hours), the clamps and gasket were removed and the plates gently prised apart with a spatula. The plate carrying the gel was placed on top of the LKB Multiphor apparatus which has cold water (4°C) circulating. The electrode paper wicks were soaked in 1M NaOH (cathode) and 1M H_3PO_4 (anode) and placed along the appropriate edges of the gel. Either sample pads (1.5cm X 0.5cm of filter paper) or the applicator strip (Pharmacia), applied close to the cathodic side of the gel, could be used for

sample loading. 15µl of sample was loaded onto the sample pad or applicator strip, together with the pI markers (Pharmacia, pI 3.5-9.5), and the electrode lid was placed over the gel ensuring that the platinum electrode wires were in contact with the electrode strips. The gel was focussed at 30W for 1½ to 2 hours, after which it was subjected to one of the following staining methods.

(iii) IEF POLYACRYLAMIDE GEL STAINING

(iiia) COOMASSIE BRILLIANT BLUE STAINING

Before staining, the gel was fixed in fixing solution (115g trichloroacetic acid, 34.6g sulphosalicylic acid and made up to 1 litre) for 15 to 30 minutes, in a stainless steel tray and covered, to precipitate the proteins and wash out the ampholines. The gel was then immersed in staining solution (500ml ethanol, 160ml acetic acid, 0.115% Coomassie Brilliant Blue R250 and made up to 2 litres) overnight, and destained with several changes of destaining solution (staining solution without Coomassie Brilliant Blue R250).

(iiib) SILVER STAINING

This was achieved using the Silver Stain Plus (Bio-Rad) for detecting proteins. The gel was initially fixed in fixing solution as previously described [section 2.10.7(iiia)], and then fixed in the Fixative Enhancer Solution [50% (v/v) methanol, 10% (v/v) acetic acid, 10% (v/v) fixative enhancer concentrate, 30% deionised distilled water] for 30 minutes with gentle agitation. The gel was stained in the staining solution (prepared 5 minutes before use) [175ml deionised distilled water, 25ml silver complex solution, 25ml reduction moderator solution, 25ml image development reagent, and 50ml of room temperature Development Accelerator Solution added immediately before use (50g Development Accelerator Reagent, 950ml deionised distilled water and made up to 1 litre, stored at 4°C)] for approximately 20 minutes or until the desired staining intensity was achieved. The reaction was stopped with 5% acetic acid for 15 minutes and rinsed in deionised

distilled water for 5 minutes. The gel was sealed in a plastic bag and photographed immediately.

2.10.8 HMB-S ASSAY

The assay was adopted according to the method described by de Rooij et al (1987). Tris-HCl buffer (50mmol/l, pH 8.0) containing BSA (1g/l) and PBG (100μmol/l, SIGMA) was added to samples of 50 to 200μl to a final volume of 600μl. The mixture was incubated for 60 minutes at 37°C and 600μl of trichloroacetic acid (1.5mol/l) was added to stop the reaction. The mixture was exposed to UV light at 350nm for 30 minutes to convert the uroporphyrinogen to uroporphyrin, after which it was centrifuged at room temperature for 10 minutes at 1500g. The fluorescence was measured on the Perkin Elmer fluorescence spectrophotometer MPF-44B at an excitation wavelength of 399nm and an emission wavelength of 640nm. Coproporphyrin I (SIGMA) was used as a standard and the activity was expressed as pmol/ml/hr at 37°C.

CHAPTER 3:

RESULTS

CHAPTER 3: RESULTS

3.1 STRATEGY FOR DETECTING MUTATIONS WITHIN THE *HMB-S* GENE

Due to the small size of the coding sequence of the *HMB-S* gene (1.1kb), total cellular RNA was extracted from peripheral blood lymphocytes of the thirty patients with AIP and analysis was performed directly on the coding sequence as shown in Figure 3.1. If the patient was found to carry a mutation leading to exon skipping, then the strategy depicted in Figure 3.2 was followed to characterise the mutation present. All the mutations detected at the RNA level were also analysed at the DNA level by amplifying the corresponding exon and sequencing was performed.

3.1.1 RT-PCR OF THE *HMB-S* mRNA

The entire coding sequence of the *HMB-S* gene was amplified by RT-PCR as a single segment. Primers located in the 5' and 3' untranslated regions (019H and 020H respectively, Table 2.3) were used in the initial amplification process to yield a fragment size of 1.3kb and nested primers were employed on the amplified product for screening by HA and CCM, characterisation by direct sequencing of both strands was done and restriction endonuclease digestion was carried out as a final confirmatory step. CCM with osmium tetroxide consistently failed to reveal any cleavage products, therefore, further screening by CCM was carried out using hydroxylamine for modification.

The mRNA was also subjected to exon-by-exon amplification to locate the region of exon skipping or deletion in products showing a shorter transcript.

3.1.2 AMPLIFICATION OF SELECTED EXONS OF GENOMIC DNA

DNA was used as a template for PCR when size differences were detected at the mRNA level after RT-PCR. Primers were designed using OLIGO™ Version 3.4 according to the criteria stated in section 2.6 to amplify across the 5' and 3' exon-intron junctions for

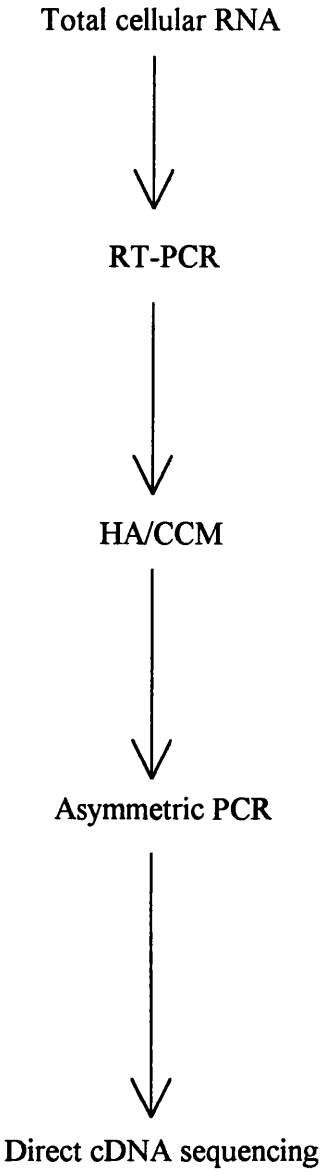


Figure 3.1: Strategy adopted when using RNA as a template.

This is the principal strategy followed in this study in the identification of the molecular pathology of AIP. Any mutation detected was then analysed at the DNA level as shown in Fig. 3.2.

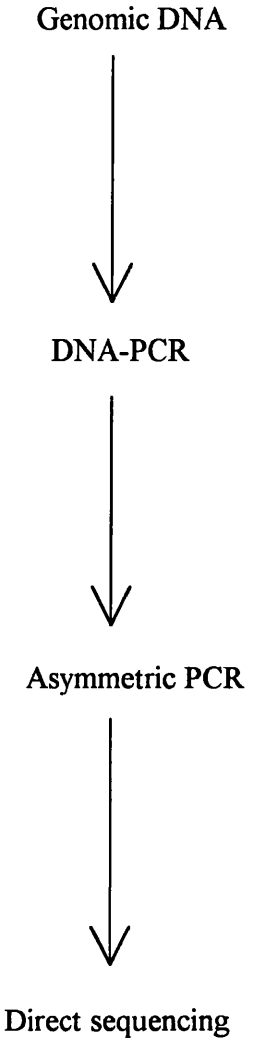


Figure 3.2: Strategy adopted when using genomic DNA as a template.

This strategy was followed to analyse the sequences at the exon-intron boundaries if exon skipping was detected at the RNA level, as well as to detect the presence of the same mutation found after amplification using RNA as the initial template.

subsequent characterisation of any base change by direct sequencing of both strands.

3.2 SUMMARY OF MUTATIONS DETECTED IN THE *HMB-S* GENE

By employing the strategies outlined in section 3.1, nine different mutations, of which four were novel, and two polymorphisms, were identified. In addition, the phenomenon of exons 3 and 12 undergoing alternative splicing was also encountered. The list of mutations detected in this study are tabulated in Table 3.1 and sub-divided according to the origins of the families as shown in Table 3.2.

The base changes detected were confirmed by using the appropriate restriction enzyme to see if the recognition site of the enzyme was created or abolished. The GCG Package "MAP" programme was used to search for the relevant restriction enzymes. The use of restriction enzyme digestion is simple and rapid in determining the carrier status of family members of an affected individual, as well as to screen a large number of healthy individuals to find out if the sequence alteration was polymorphic. The list of restriction enzymes employed in this study as a further confirmatory step is shown in Table 3.3.

A total of nine different and potentially pathogenic mutations identified in this study have been fully characterised. These include six missense mutations, one insertion, one deletion and one splice site mutation leading to exon skipping. The missense mutations include R22C, R26C, R26H, R116W, R173Q and L177R, the insertional mutation 771insT, the deletion mutation $delG_{1002/1003/1004}$ and the splice acceptor site mutation was in intron 7 (345-1G→A). Five of the missense mutations occurred at the hypermutable CpG dinucleotide. In two of the South African patients, no sequence alterations were identified in their coding sequence. An analysis of the promoter region of the DNA would have been done, however, the DNA samples of these patients were not sent to us for further analysis.

Table 3.1: Summary of *HMB-S* mutations and polymorphisms detected in this study.

The table summarises the mutations and polymorphisms identified in this study, showing the method of screening used in the identification of each base change, the location of each sequence alteration, the base change involved and the effects of these alterations.

*denotes novel mutations

#denotes mutations involving the hypermutable CpG dinucleotide

lower case nucleotide: intronic sequence

upper case nucleotide: exonic sequence

The underlined bases are the bases which have been altered.

Table 3.1: Summary of *HMB-S* mutations and polymorphisms detected in this study.

Mutation	Method of Screening	Exon/ Intron	Nucleotide position	Base change	Effect of mutation
R22C*#	CCM, sequencing	E3	64	CGC→TGC	Arg→Cys
R26C#	CCM, HA, sequencing	E3	76	CGC→TGC	Arg→Cys
R26H#	CCM, HA, sequencing	E3	77	CGC→CAC	Arg→His
S45S#	sequencing	E4	135	TCG→TCA	Ser→Ser
345-1G→A*	RT-PCR, HA, sequencing	I7	345-1	cag→caa	E8 skipping
R116W#	CCM, HA, sequencing	E8	346	CGG→TGG	Arg→Trp
R173Q#	CCM, HA, sequencing	E10	518	CGG→CAG	Arg→Gln
L177R	CCM, HA, sequencing	E10	530	CTG→CGG	Leu→Arg
V202V	sequencing	E10	606	GTG→GTT	Val→Val
771insT*	HA, sequencing	E12	771	T insertion	Stop codon 33aa away
$delG_{1002/1003/1004}$*	CCM, HA, sequencing	E15	1002, 1003 or 1004	G deletion	Stop codon 8aa away

Table 3.2: Mutations identified according to the origins of the families.

Origin	Mutation	Number of families
South Africa	R26H	1
	R116W*	3
	R173Q	1
	771insT	1
	$delG_{1002/1003/1004}$	1
Scotland	L177R	2
	R22C	1
England	E8 skipping	1
Ireland	R26C	1
	R116W*	1

*denotes a recurrence of mutation in two different populations.

Table 3.3: Summary of restriction enzymes used in this study to confirm the mutations detected.

Mutation	Exon/ Intron	Nucleotide change	Primers used for PCR	Product size (bp)	Restriction fragment sizes (bp)	Enzyme
R22C	E3	$C_{64}\rightarrow T$	019H + R341	359	**204*, 132, 126,** 78, 23	*Hinf* I
R26C	E3	$C_{76}\rightarrow T$	377H + R341	338	**310*, 266,** 44, 28	*Aci* I
R26H	E3	$C_{77}\rightarrow T$	377H + R341	338	**310*, 266,** 44, 28	*Aci* I
E8 skipping	I7	345-1G→A (last base of intron 7)	Phe6+ + I8R	868	**346*, 254, 220, 208, 92,** 39, 22, 21, 9, 3	*Eco* RII
R116W	E8	$C_{346}\rightarrow T$	Phe6+ + Phe11-	398	**228*, 170, 130, 98**	*Aci* I
R173Q	E10	$G_{518}\rightarrow A$	F413 + Phe11-	233	**189*, 104, 85, 44**	*Nci* I
L177R	E10	$T_{530}\rightarrow G$	F413 + Phe11-	233	**170, 117*, 63,** 53*	*Aci* I
771insT	E12		F666 + 508D	443	**431, 340*,** 91*, 12	*Sty* I

Numbers in bold are the major fragments observed on the gel.

* denotes the fragments generated due to the mutation.

Enzyme in red shows its wide application for the different mutations in this study.

3.3 ANALYSIS OF MUTATIONS IN THE *HMB-S* CODING SEQUENCE

3.3.1 SMALL ALTERATIONS

3.3.1.1 MISSENSE MUTATIONS

In 27 of the 30 patients studied, RT-PCR products of the expected size (1.3kb) were observed upon electrophoresis in a 2% agarose gel, thus ruling out the possibility of large deletions or exon skipping. Abnormal band patterns were observed on heteroduplex (Figure 3.3) and CCM analyses (using hydroxylamine for chemical modification), which were subsequently characterised by direct sequencing and reconfirmed by restriction enzyme digestion, where applicable.

(i) R22C (EXON 3)

This novel mutation was initially detected in two Scottish siblings (Appendix 1, Pedigree 9532, individuals III:8 and III:9) by CCM analysis with hydroxylamine (Figure 3.4a). Upon sequencing in the region of the cleavage site, a C to T transition at nucleotide position 64 (Figure 3.4b) was observed in both siblings, which was confirmed after sequencing the opposite strand and by *Hinf* I digestion (Figures 3.4c and 3.4d). Subsequent analysis of five family members (II:2, III:3, III:5, III:7 and III:11) by *Hinf* I digestion revealed that three of them carry the mutation. This mutation causes arginine to be substituted by cysteine. Arg22 (equivalent to Arg7 in *E. coli*), which is highly conserved in ten species (Louie et al, 1996), is the residue affected. Its high degree of conservation is an indication that it has an important role in the proper functioning of the enzyme. Arginine, a basic residue, is substituted by cysteine, a neutral, polar amino acid which is predicted to destabilise the protein. However, the precise role of this residue is not known at the present time.

There is a family history of this condition. II:9 had succumbed to the disease and II:7 is a known carrier of the defective gene. The R22C mutation was transmitted from individual II:7 to three of his daughters. III:8 and III:9 have reduced HMB-S activity (21.0 and 22.0 units respectively) (normal range 30 - 54 units). His fourth daughter, III:10, had

Figure 3.3: Heteroduplex analysis of 10 patients with different mutations.

Heteroduplex analysis was performed on 10 patients with different mutations. Each of the mutations revealed a different heteroduplex pattern on the MDE Gel solution (Flowgen).

The primers used for amplifying the products in lanes 1 to 9 and 12 to 13 were 019H and R478 (Table 2.3) to yield a fragment of 496bp and the primers Phe6+ and Phe11- were used to amplify the products in lanes 10 and 11 to give a product of 398bp.

L: 1kb ladder (Gibco BRL)

HC: heteroduplex control (350bp) supplied by Flowgen

Lanes 1 to 3: negative controls

Lanes 4 and 5: R26C

Lanes 6 to 8: R116W

Lane 9: R26H

Lane 10: 771insT

Lane 11: $delG_{1002\ 1003\ 1004}$

Lane 12: R173Q

Lane 13: L177R

A non-specific band was observed in lanes 2 to 5 and 9 as indicated by the arrow.

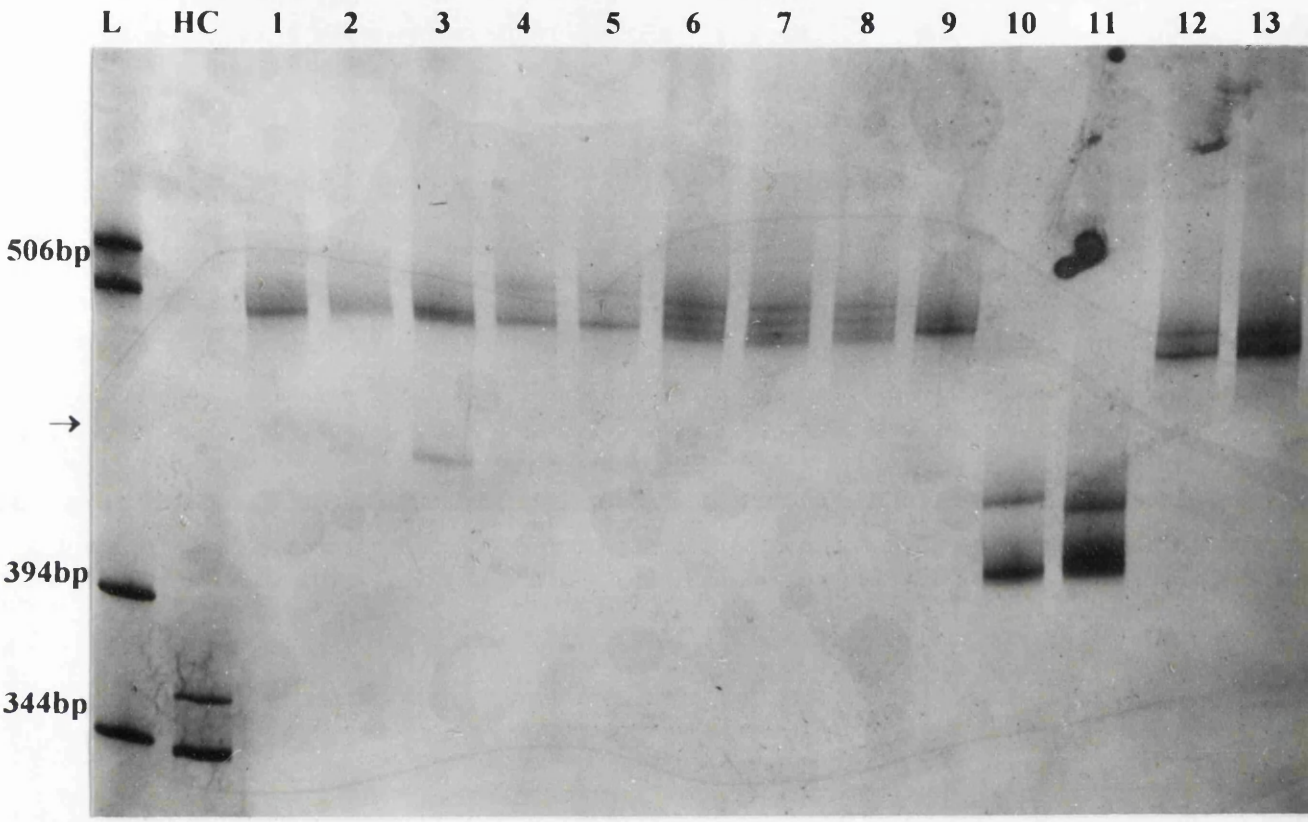


Figure 3.3

Figure 3.4a: Chemical cleavage mismatch analysis showing the cleavage product due to the R22C mutation.

The fragment amplified for CCM analysis with the primers 019H and 020H (Table 2.3) in the 2 siblings (III:8 and III:9 of Pedigree 9532 in Appendix 1) and 20 healthy individuals gave a product of 1.3kb. A cleavage product of 82bp was observed in both siblings as shown by the arrow, indicating a sequence alteration of C to T at nucleotide position 64 of the sense strand from the translation initiation codon since the primer 019H is located 18bp upstream from the translation initiation site. P = patient, C = normal control.

Figure 3.4b: Sequence analysis of exon 3 (R22C).

Partial sequence of the *HMB-S* cDNA in the region of the cleavage product in Figure 3.4a confirmed a $C_{64} \rightarrow T$ transition as indicated in both individuals III:8 and III:9 of Pedigree 9532 (Appendix 1) but not in 20 healthy individuals. P = patient, C = normal control.

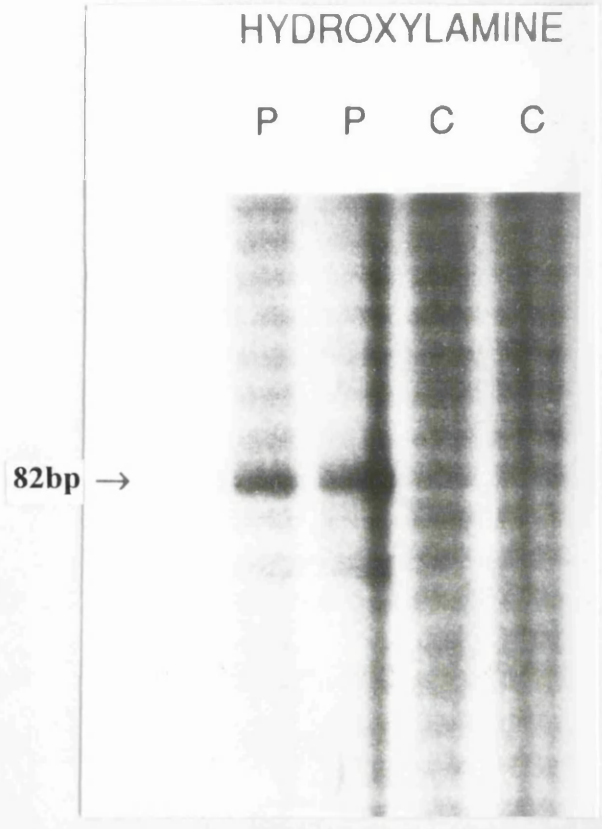


Figure 3.4a

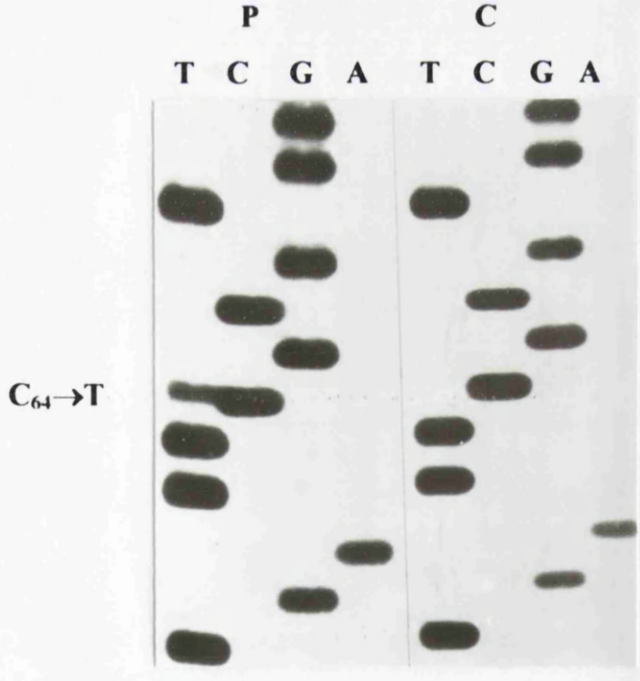


Figure 3.4b

Figure 3.4c: Restriction map of *Hinf* I to detect the R22C mutation.

Restriction map of the amplified cDNA fragment using primers 019H and R341 (Table 2.3) showing the restriction fragment sizes expected for the normal (132bp, 126bp, 78bp and 23bp) and mutant (132bp, 204bp and 23bp) alleles upon digestion with *Hinf* I. The $C_{64}\rightarrow T$ transition abolishes the *Hinf* I* recognition site to yield a fragment of 204bp from the mutant allele.

Figure 3.4d: Restriction analysis of R22C with *Hinf* I.

Restriction enzyme digestion with *Hinf* I on a 3% SeaKem (GTG) agarose gel. In Pedigree 9532, a 204bp fragment, which indicates the loss of the *Hinf* I* restriction site due to the mutation, was seen in individuals II:2, III:5, III:8 and III:9 and III:11 but not in individuals III:3 and III:7 and 20 normal controls. L = 100bp ladder (Gibco BRL), P = digested PCR products from the patient, C = normal control.

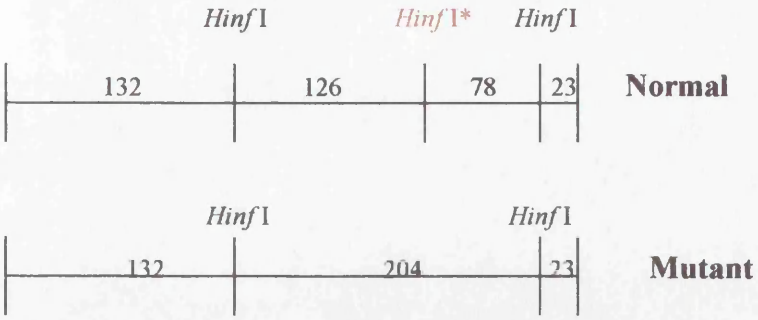


Figure 3.4c

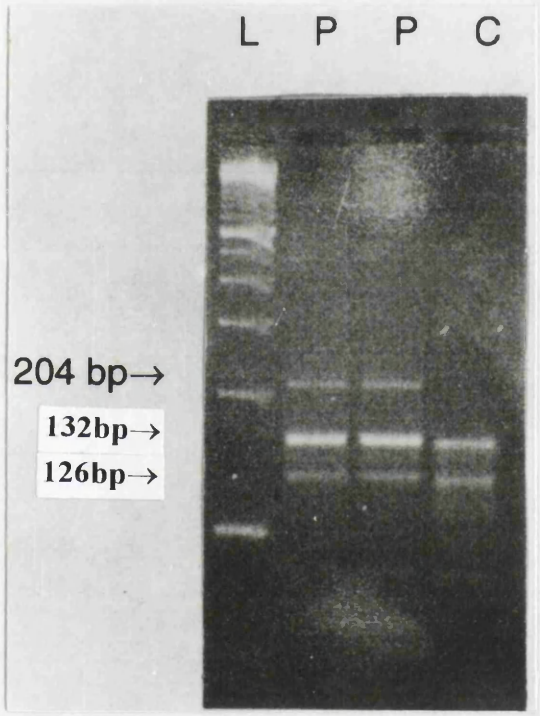


Figure 3.4d

not been analysed by molecular techniques but her HMB-S activity was below the normal range (22.9 units).

(ii) R26C (EXON 3)

This mutation was initially identified in two of three Irish siblings (Appendix 1, Pedigree 13168, individuals III:1 and III:3 but not III:2). The cleavage products and heteroduplexes observed on hydroxylamine modification in CCM analysis (Figure 3.5a) and HA (Figure 3.3) respectively were due to a missense mutation as revealed by direct sequencing. A C to T transition at position 76 was present (Figure 3.5b) which caused arginine to be substituted by cysteine. This mutation abolishes an *Aci* I restriction site (Figures 3.5c and 3.5d) which was employed to rapidly determine the carrier status of a further thirteen members of this family (II:1 to II:6, II:8, III:5 to III:8, IV:1 and IV:2), three of whom (individuals II:8, IV:1 and IV:2) were found to be carriers and one (II:7), who is deceased, was inferred to be a carrier of the defective gene from the results obtained for the other family members.

Arg26 (Arg11 in *E. coli)*, which is a conserved residue in twelve species, serves to form a salt-bridge with the acidic side chain of the second pyrrole ring of the cofactor. This mutation affects the β-sheet conformation and results in a defective enzyme being produced. Since the basic arginine residue is replaced by the polar uncharged cysteine, this is predicted to destabilise the protein.

Family studies indicated that II:7 had most probably succumbed to the disease, therefore, all her offspring were analysed for the presence of a mutation in the *HMB-S* gene. III:1 and III:3 have lower than normal HMB-S activity levels (Table 2.2), which indicate that they could be carriers of the defective gene. This was confirmed by molecular analysis and restriction enzyme digestion (Figure 3.5). Both III:2 and III:5 have activity levels within the normal range, thus ruling out the possibility that they could be harbouring the defective gene. II:6 has a low HMB-S activity but no sequence alterations were detected in his coding sequence. One possibility could be that he has a much higher proportion of alternatively spliced exons 3 and 12 in his coding sequence, which is a normal occurrence in the general population (as seen in section 3.4), thus resulting in a

Figure 3.5a: Chemical cleavage mismatch analysis showing the cleavage product due to the R26C mutation.

The fragment amplified for CCM analysis with the primers 019H and 020H (Table 2.3) in three Irish siblings (III:1 to III:3 of Pedigree 13168 in Appendix 1) and 20 healthy individuals gave a product of 1.3kb. A cleavage product of 94bp was observed in III:1 and III:3 but not in III:2 as indicated by the arrow, thus indicating a C to T transition at nucleotide position 76 of the sense strand from the translation initiation codon since the primer 019H is located 18bp upstream from the translation initiation site. P = patient, C = normal control.

Figure 3.5b: Sequence analysis of exon 3 (R26C).

Partial sequence of the *HMB-S* cDNA in the region of the cleavage product in Figure 3.5a confirmed a $C_{76} \rightarrow T$ transition as indicated in both individuals III:1 and III:3 of Pedigree 13168 but not in III:1 and 20 healthy individuals. P = patient, C = normal control.

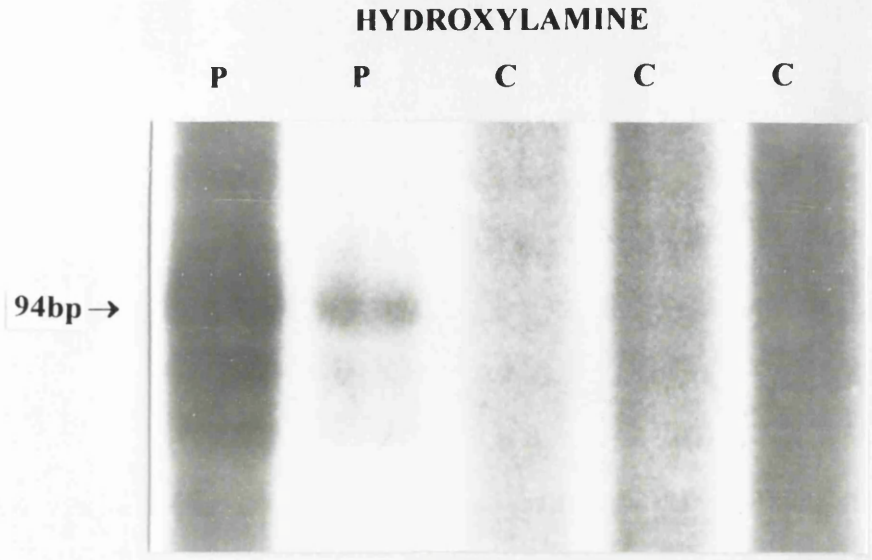


Figure 3.5a

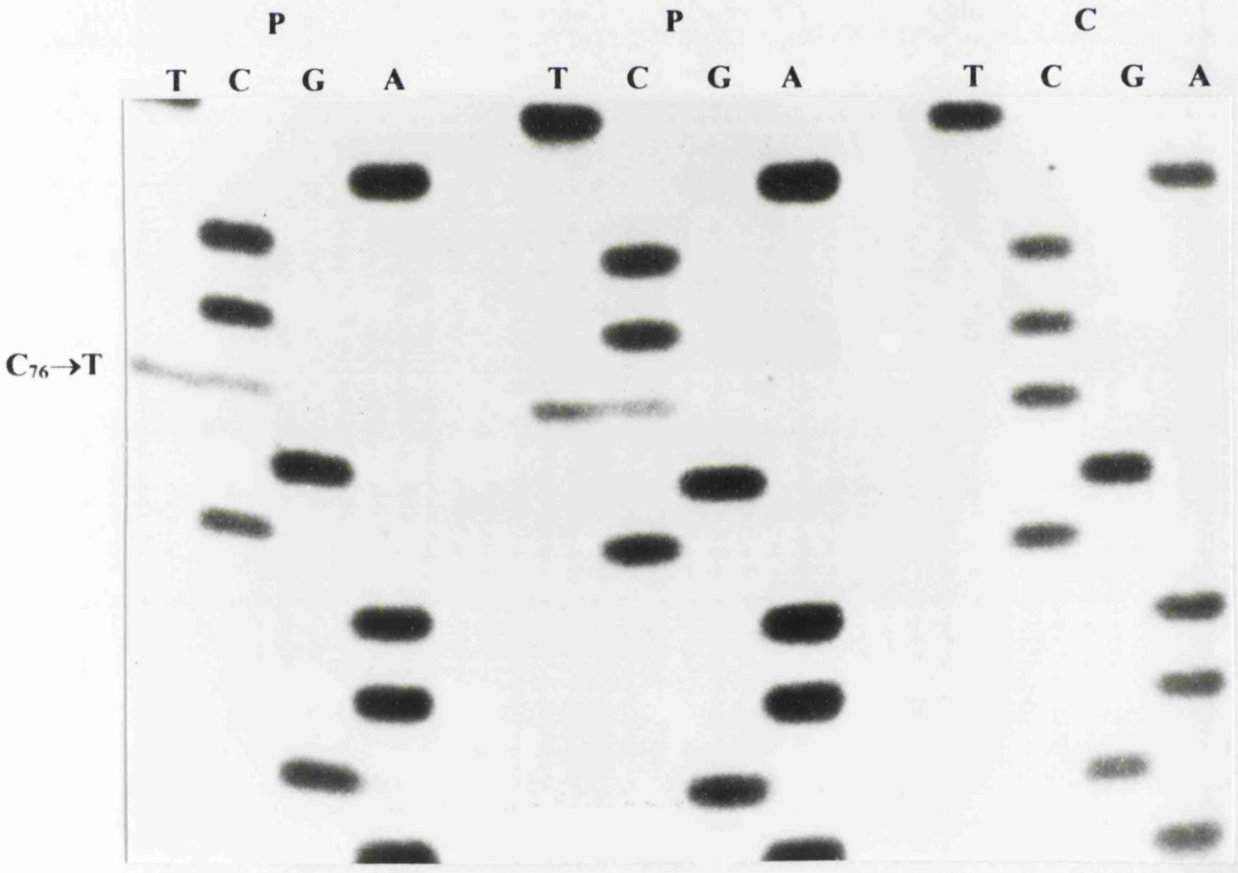


Figure 3.5b

Figure 3.5c: Restriction map of *Aci* I to detect the R26C mutation.

Restriction map of the amplified cDNA fragment using primers 377H and R341 (Table 2.3) showing the restriction fragment sizes expected for the normal (266bp, 44bp and 28bp) and mutant (310bp and 28bp) alleles upon digestion with *Aci* I. The *Aci* I* restriction site is removed as a result of the mutation, yielding a 310bp fragment.

Figure 3.5d: Restriction analysis of R26C with *Aci* I.

Restriction enzyme digestion with *Aci* I on a 3% SeaKem (GTG) agarose gel. In Pedigree 13168, a 310bp fragment, which indicates the loss of the *Aci* I* restriction site due to the mutation, was seen in individuals II:8, IV:1 and IV:2 but not in II:1 to II:6 and III:5 to III:8 and 20 healthy individuals. L = 100bp ladder (Gibco BRL), C = normal control, P = digested PCR products from the patient.

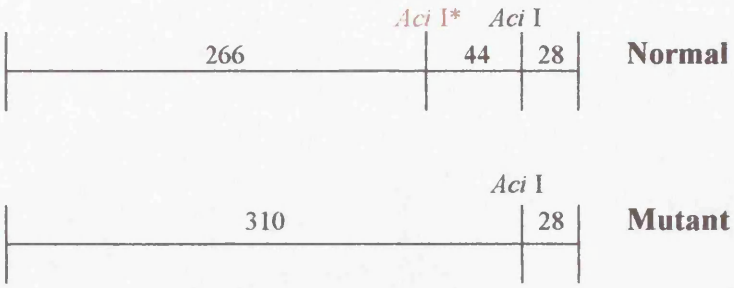


Figure 3.5c

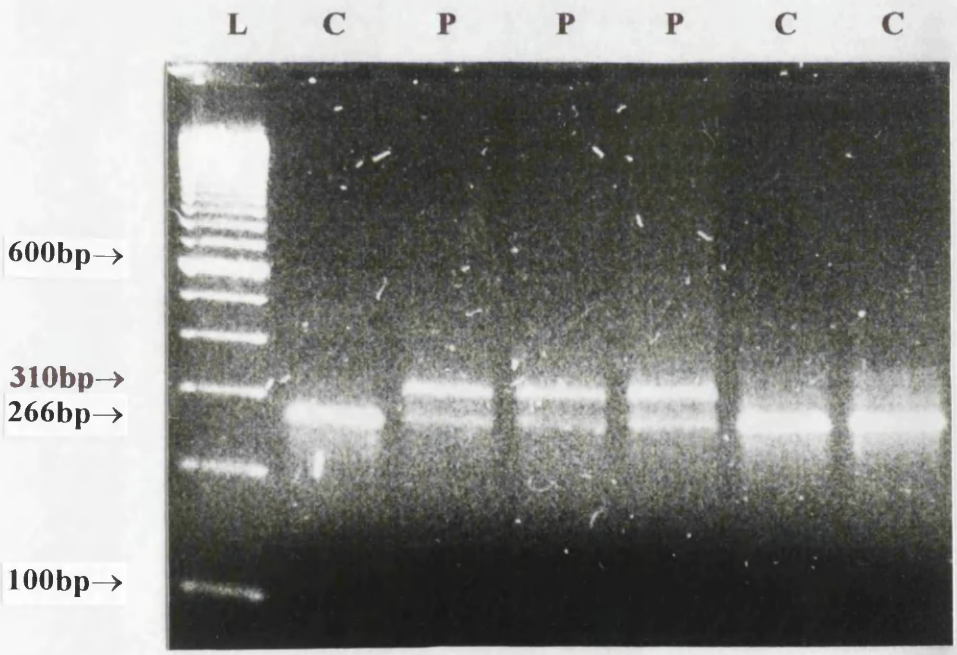


Figure 3.5d

lower HMB-S activity. It would be extremely useful and important to compare his measurement with those of his siblings to aid in the clarification of his carrier status.

From Pedigree 13168, individual II:7 is deduced to be a carrier for the mutant allele because (i) her husband does not have the mutation; (ii) two of her children are carriers of the mutant allele; and (iii) she has an affected sister (individual II:8). Another interesting observation was that II:7 also had the V202V polymorphism because two of her unaffected children had the V202V allele, while the other two affected children had the mutant allele but not the polymorphic allele. It is fortunate that II:6 is homozygous for the G allele, thus allowing the deduction of the genotype of the deceased individual II:7.

Presymptomatic testing was carried out for IV:1 and IV:2, who are 6 and 3 years old respectively. This is extremely vital since both individuals were found to be carrying the defective gene. Consequently, they can be appropriately advised to avoid the precipitants of AIP and subsequently lead normal lives without manifesting the condition.

(iii) R26H (EXON 3)

This mutation is in the same codon as R26C but the mutated nucleotide is the second nucleotide of codon 26 whereas the mutated nucleotide for R26C is the first nucleotide. Positive CCM and HA results (Figures 3.6a and 3.3 respectively) were obtained for this missense mutation which, upon sequencing, was found to be due to a G to A transition at nucleotide 77 (Figure 3.6b), resulting in the substitution of arginine with histidine. *Aci* I was used again to confirm the base change (Figures 3.6c and 3.6d). Arg26 is situated with its side chain pointing into the interior of the active site cleft and is close to the negatively charged carboxylate groups on ring C2 of the dipyrromethane cofactor (Jordan and Woodcock, 1991). The replacement of arginine with histidine severely inhibits substrate binding and chain elongation. Both arginine and histidine are basic amino acids but, because they affect key residues which are highly conserved in the protein, they are predicted to be the primary cause of this condition. The South African patient with this mutation had suffered two previous attacks but has remained well since then.

Figure 3.6a: Chemical cleavage mismatch analysis showing the cleavage product due to the R26H mutation.

The fragment amplified for CCM analysis with hydroxylamine using the primers 019H and 020H (Table 2.3) in the South African patient 644 and 20 normal controls gave a product of 1.3kb. A cleavage product of 95bp was observed in the patient as shown by the arrow, indicating a C to T transition had occurred at nucleotide position 77 from the translation initiation codon since the primer 019H is located 18bp upstream from the translation initiation site. C = normal control, P = patient.

Figure 3.6b: Sequence analysis of exon 3 (R26H).

Partial sequence of the *HMB-S* cDNA in the region of the cleavage product in Figure 3.6a confirmed a $G_{77} \rightarrow A$ transition on the antisense strand as indicated by the arrow but was not detected in 20 healthy individuals. P = patient, C = normal control.

The quality of the sequence around this region was the same despite repeated attempts.

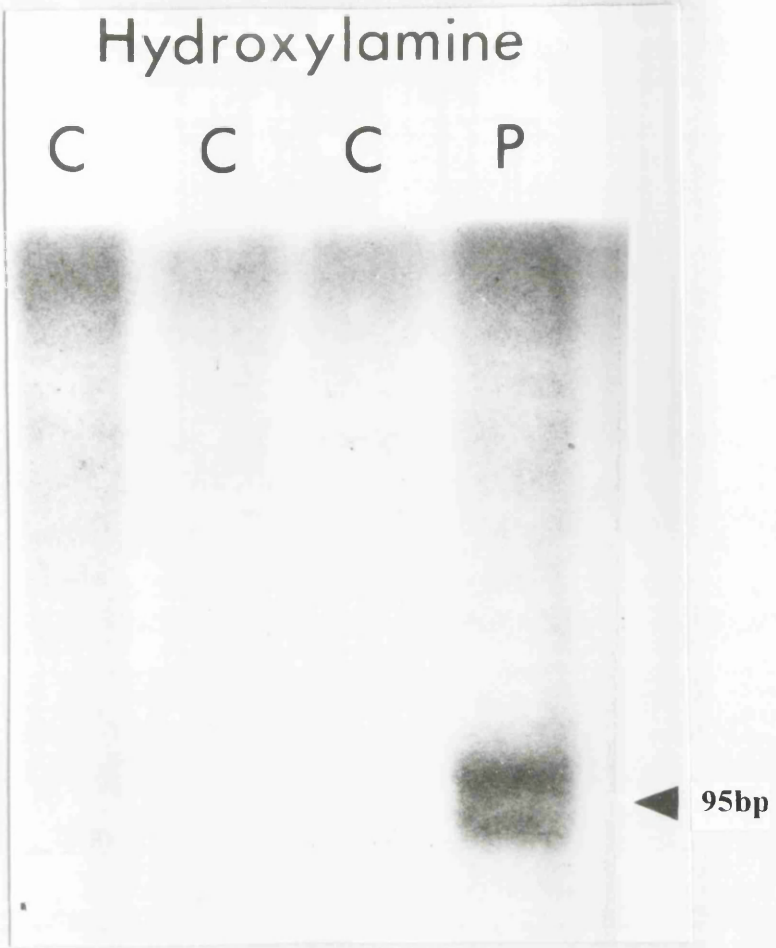


Figure 3.6a

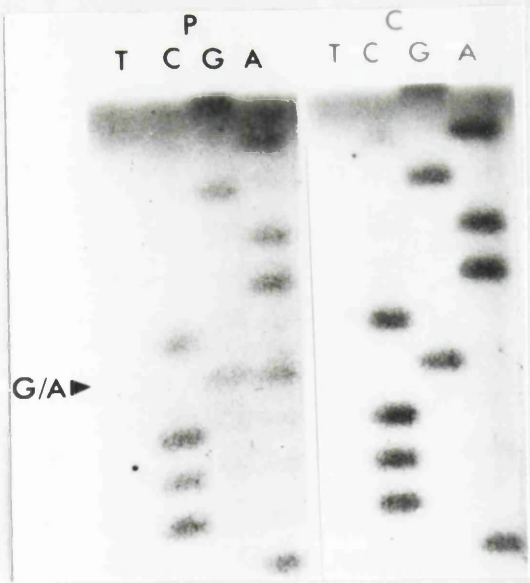


Figure 3.6b

Figure 3.6c: Restriction map of *Aci* I to detect the R26H mutation.

Restriction map of the amplified cDNA fragment using primers 377H and R341 (Table 2.3) showing the restriction fragment sizes expected for the normal (266bp, 44bp and 28bp) and mutant (310bp and 28bp) alleles upon digestion with *Aci* I. The *Aci* I* restriction site is removed as a result of R26H, yielding a 310bp fragment.

Figure 3.6d: Restriction analysis of R26H with *Aci* I.

Restriction enzyme digestion with *Aci* I on a 3% SeaKem (GTG) agarose gel. The South African patient 644 revealed a 310bp fragment which indicates the loss of the *Aci* I* restriction site due to the mutation. The 310bp fragment was not present in any of the 20 healthy individuals analysed. L = 100bp ladder (Gibco BRL), P = digested PCR products from the patient, C = normal control.

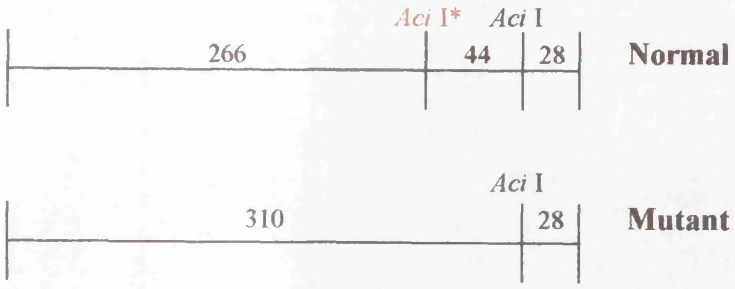


Figure 3.6c

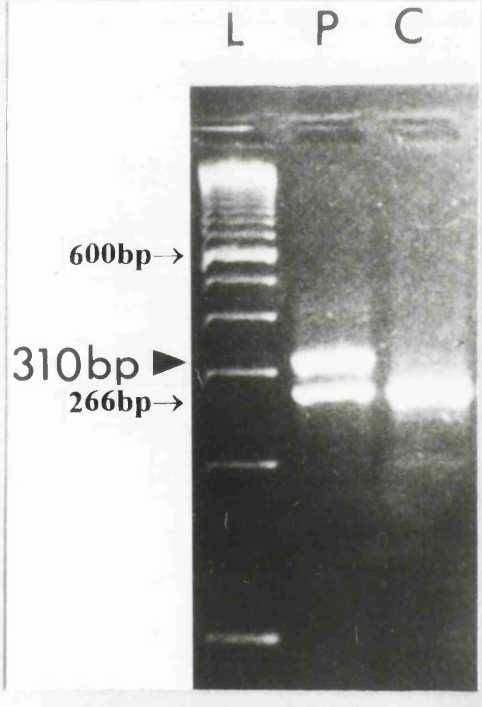


Figure 3.6d

(iv) R116W (EXON 8)

This missense mutation was detected in three unrelated South African patients and two members of a family from the Republic of Ireland (pedigree not available). Its presence was initially seen on CCM and heteroduplex analyses (Figures 3.7a and 3.3 respectively). The substitution of the highly conserved arginine with tryptophan was the result of a C to T transition at nucleotide 346 (Figure 3.7b). It abolishes the restriction site for *Aci* I (Figures 3.7c and 3.7d) to yield the restriction fragments shown in Table 3.3. A further three members of the Irish family were found to be negative for this mutation after employing the *Aci* I restriction enzyme.

This mutation was first reported by Lee et al (1990) in a Swedish family with AIP. The residue involved, Arg116 (Arg101 in *E. coli*) is substituted by tryptophan. The consequence of the substitution of the basic amino acid arginine with tryptophan, a neutral, non-polar amino acid, is the destabilisation of the protein. Arg116, which is conserved in twelve species, interacts with the interdomain hinge region by forming a salt-bridge with Glu231 and a hydrogen bond with the backbone carbonyl group of residue 198. Its replacement with tryptophan results in the destabilisation of the protein, possibly due to the loss of these stabilising interactions.

In the three unrelated South African patients (635, 638 and 639) who carry this mutation, they did not suffer any serious attacks. Their erythrocyte HMB-S activity varied very widely (patients 635 and 638, Table 2.1) but they have remained well. This mutation was also identified in two members of an Irish family (pedigree not available). R116W was previously reported in patients of French and Dutch origins (Gu et al, 1993). Approximately 30% of the Dutch patients analysed had the R116W mutation as compared with 3% of the French patients, indicating the high frequency of this mutation in the Dutch population.

Figure 3.7a: Chemical cleavage mismatch analysis showing the cleavage product due to the R116W mutation.

The fragment amplified for CCM analysis with the primers 019H and 020H (Table 2.3) in the South African patients 635, 638 and 639 and 20 healthy individuals gave a product of 1.3kb. A cleavage product of 364bp was observed in all three patients as indicated by the arrow, indicating a C to T transition at nucleotide position 346bp from the translation initiation codon since the primer 019H is located 18bp upstream from the translation initiation site. L = radiolabelled 1kb ladder (Gibco BRL), P = patient, C = normal control.

Figure 3.7b: Sequence analysis of exon 8 (R116W).

Partial sequence of the *HMB-S* cDNA showed the $C_{346}\rightarrow T$ transition in the region of the cleavage product in Figure 3.7a as indicated by the arrow. This sequence alteration was not observed in 20 healthy individuals. M = mutant sample, N = normal control.

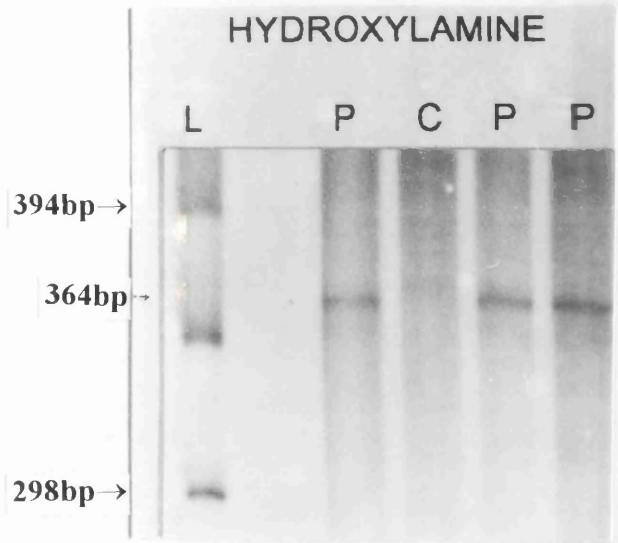


Figure 3.7a

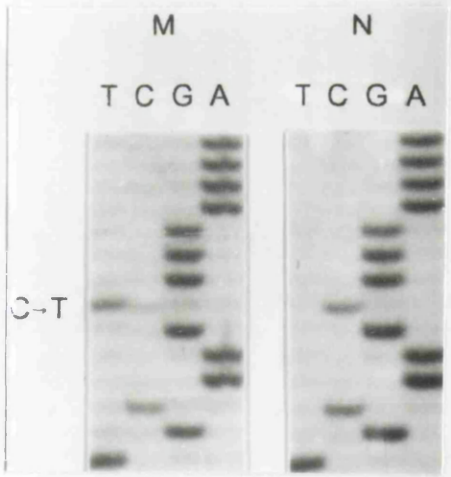


Figure 3.7b

Figure 3.7c: Restriction map of *Aci* I to detect the R116W mutation.

Restriction map of the amplified cDNA fragment using primers Phe6+ and Phe11- (Table 2.3) showing the restriction fragment sizes expected for the normal (98bp, 130bp and 170bp) and mutant (228bp and 170bp) alleles upon digestion with *Aci* I. The *Aci* I* restriction site is removed as a result of R116W, yielding a 228bp fragment.

Figure 3.7d: Restriction analysis of R116W with *Aci* I.

Restriction enzyme digestion with *Aci* I on a 3% SeaKem (GTG) agarose gel. The three South African patients, 635, 638 and 639, and a further two members of an Irish family revealed an extra 228bp fragment upon *Aci* I digestion which was not seen in 20 healthy individuals. A 228bp fragment indicates the loss of the *Aci* I* restriction site due to the mutation. L = 100bp ladder (Gibco BRL), P = digested PCR products from the patient, C = normal control.

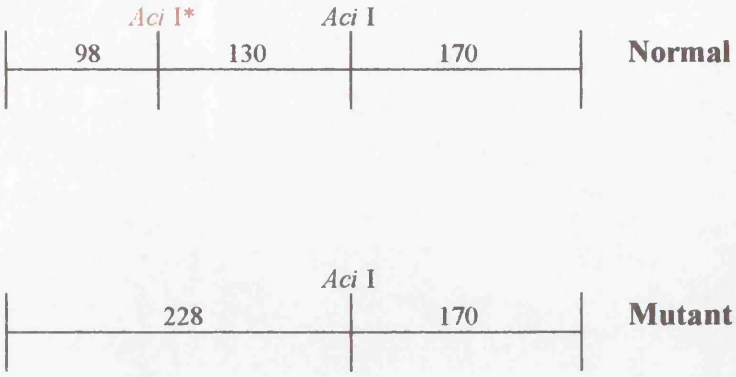


Figure 3.7c

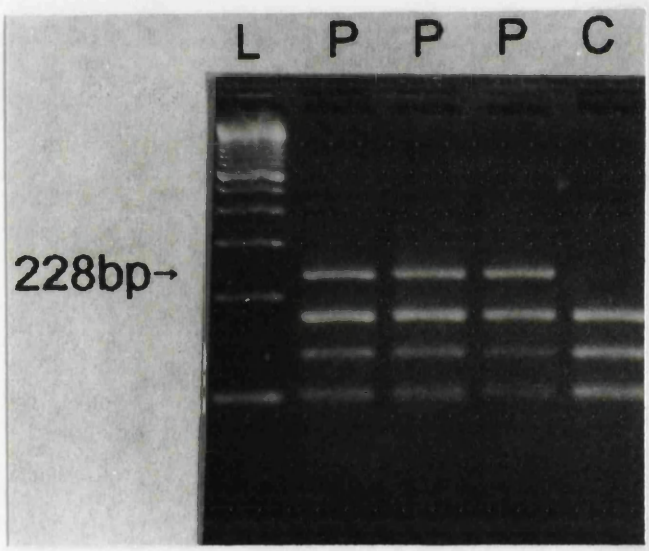


Figure 3.7d

(v) R173Q (EXON 10)

This mutation was detected by CCM and heteroduplex analyses (Figures 3.8a and 3.3 respectively) and was found to be due to a G to A transition at position 518 on sequence analysis (Figure 3.8b). It destroys the recognition site for *Nci* I (Figures 3.8c and 3.8d), resulting in the presence of a larger restriction fragment in the mutant sample and not in the negative control. Arg173 (Arg155 in *E. coli*) is a very highly conserved residue, being maintained in thirteen different species. It forms salt-bridges with the propionate group of the first pyrrole ring of the cofactor and the carboxylate side group of the substrate molecule. The substitution of the basic residue, arginine, with the polar, neutral residue, glutamine prevents binding interactions with the substrate, leading to the inhibition of substrate elongation. The affected patient, of South African origin, suffered severe attacks of the condition which required two to three monthly hospital admissions.

(vi) L177R (EXON 10)

Another highly conserved residue to be substituted is Leu177. It was detected in five members of a Scottish family (Appendix 1, Pedigree 13543, individuals II:3, III:2, III:3, III:5 and III:6) and another unrelated Scottish patient. Its replacement with arginine was due to a T to G transversion at nucleotide 530, initially detected by CCM and heteroduplex analyses (Figures 3.9a and 3.3 respectively) and characterised by sequencing (Figure 3.9b). This mutation results in the creation of an *Aci* I restriction site (Figures 3.9c and 3.9d), which was employed to screen IV:1, IV:3 and IV:4, all of whom have affected parents, but were fortunately found to be negative for this mutation. Leu177 (Leu159 in *E. coli*) is a highly conserved residue in eleven species. It forms part of the hydrophobic core of the molecule and its replacement with arginine would be structurally deleterious. This is because leucine, which is a non-polar, neutral amino acid, is being replaced with a basic amino acid, arginine. This is predicted to have an adverse effect on the hydrophobic part of the molecule by destabilising the protein structure due to the loss of stabilising interactions with the introduction of a charged group within the apolar core.

Figure 3.8a: Chemical cleavage mismatch analysis showing the cleavage product due to the R173Q mutation.

The fragment amplified for CCM analysis using hydroxylamine with the primers 019H and 020H (Table 2.3) in the South African patient 637 and 20 normal controls gave a product of 1.3kb. A cleavage product of 536bp was observed in the South African patient as shown by the arrow, thus indicating a C to T transition at nucleotide position 518 from the translation initiation codon since the primer 019H is located 18bp upstream from the translation initiation site. P = patient, C = normal control.

Figure 3.8b: Sequence analysis of exon 10 (R173Q).

Partial sequence of the *HMB-S* cDNA in the region of the cleavage product in Figure 3.8a confirmed a $G_{518} \rightarrow A$ transition as indicated in the South African patient but not in 20 healthy controls. N = normal control, M = mutant sample.

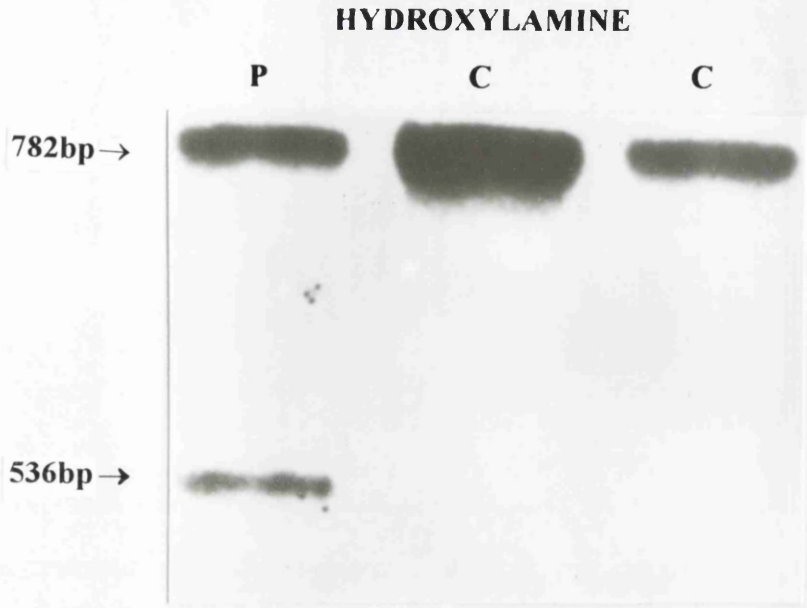


Figure 3.8a

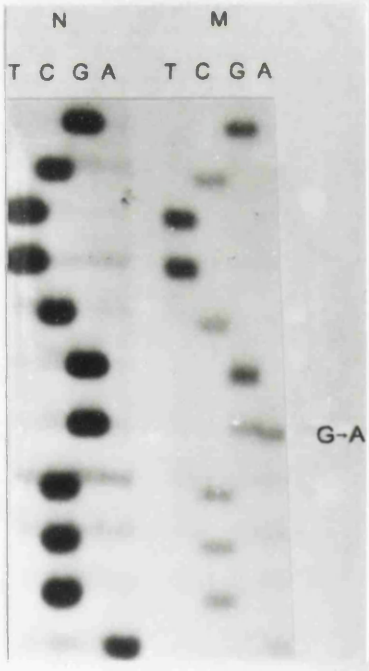


Figure 3.8b

Figure 3.8c: Restriction map of *Nci* I to detect the R173Q mutation.

Restriction map of the amplified cDNA fragment using primers F413 and Phe11- (Table 2.3) showing the restriction fragment sizes expected for the normal (104bp, 85bp and 44bp) and mutant (189bp and 44bp) alleles upon digestion with *Nci* I. The *Nci* I* restriction site is removed as a result of R173Q to produce a 189bp fragment.

Figure 3.8d: Restriction analysis of R173Q with *Nci* I.

Restriction enzyme digestion with *Nci* I on a 3% SeaKem (GTG) agarose gel. The South African patient 637 revealed a 189bp fragment upon digestion with *Nci* I, which was not present in a further 20 normal controls. A 189bp fragment indicates the loss of the *Nci* I* restriction site due to R173Q. L = 100bp ladder (Gibco BRL), P = digested PCR products from the patient, C = normal control.

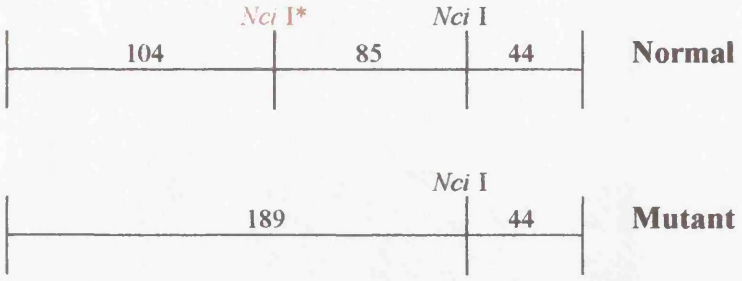


Figure 3.8c

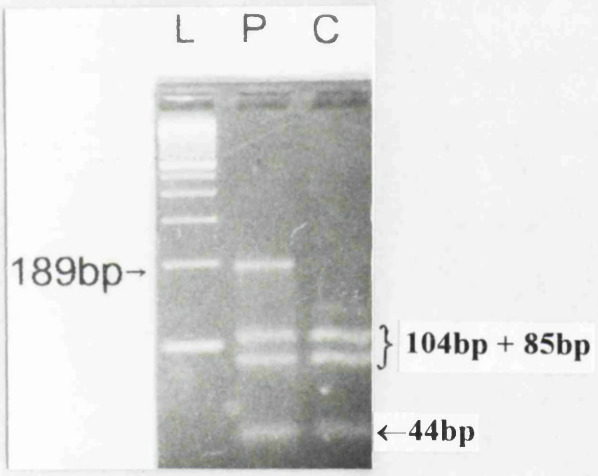


Figure 3.8d

Figure 3.9a: Chemical cleavage mismatch analysis showing the cleavage product due to the L177R mutation.

The fragment amplified for CCM analysis using hydroxylamine with the primers 019H and 020H (Table 2.3) is 1.3kb. The L177R mutation was identified in an isolated case from a Scottish family and five members of Pedigree 13543 (individuals II:3, III:2, III:3, III:5 and III:6 in Appendix 1) but not in 20 normal controls. A cleavage product of 548bp was observed in these patients, indicating a cytosine mismatch at position 530bp on the sense strand from the translation initiation codon since the primer 019H is located 18bp upstream from the translation initiation site. C = normal control, P = patient.

Figure 3.9b: Sequence analysis of exon 10 (L177R).

Partial sequence of the *HMB-S* cDNA in the region of the cleavage product in Figure 3.9a revealed a $T_{530}\rightarrow G$ transversion as indicated in all the patients mentioned in Figure 3.9a but not in 20 normal controls. M = mutant sample, N = normal control.

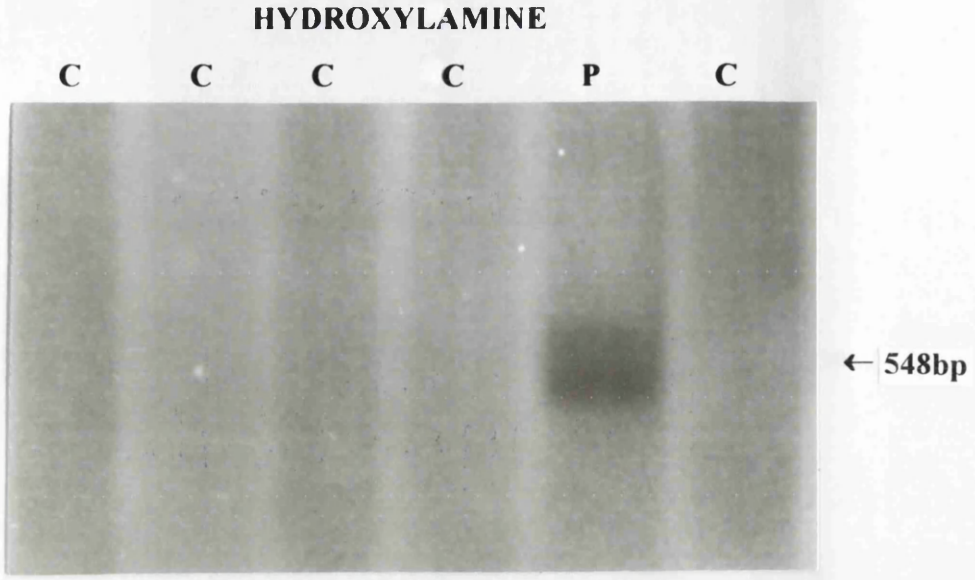


Figure 3.9a

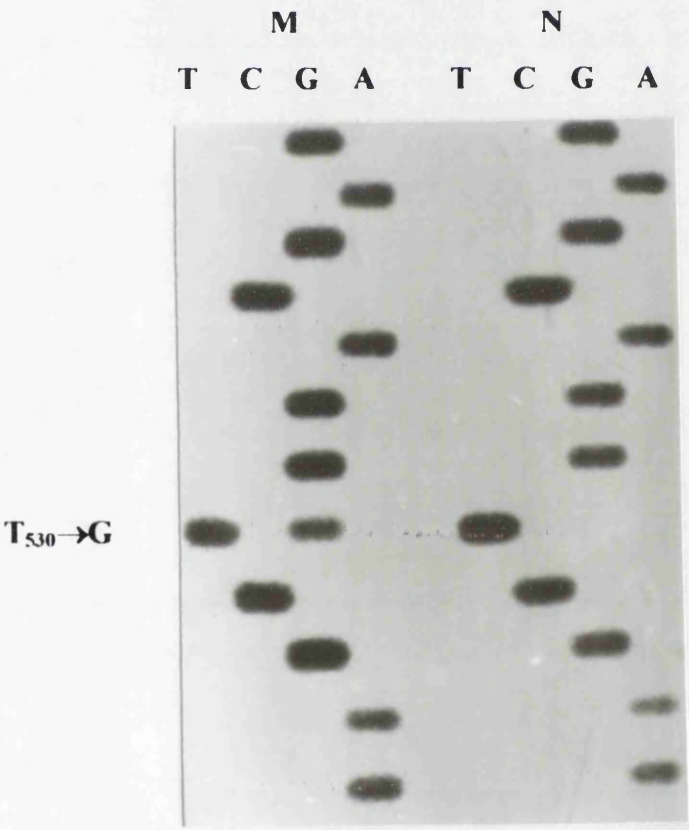


Figure 3.9b

Figure 3.9c: Restriction map of *Aci* I to detect the L177R mutation.

Restriction map of the amplified cDNA fragment using primers F413 and Phe11- (Table 2.3) showing the restriction fragment sizes expected for the normal (170bp and 63bp) and mutant (117bp, 53bp and 63bp) alleles upon digestion with *Aci* I. The *Aci* I* restriction site is created as a result of L177R, producing two additional fragments, 117bp and 53bp.

Figure 3.9d: Restriction analysis of L177R with *Aci* I.

Restriction enzyme digestion with *Aci* I on a 10% non-denaturing polyacrylamide gel. All the patients with the L177R mutation, i.e. individuals II:3, III:2, III:3, III:5 and III:6 from Pedigree 13543 (Appendix 1), and the isolated case from a Scottish family, revealed the additional 117bp fragment, indicating a sequence alteration. This fragment was not present in IV:1, IV:3 and IV:4 (Pedigree 13543) and 20 healthy individuals. L_1 = 1kb ladder (Gibco BRL), L_2 = 100bp ladder, C = normal control, P = digested PCR products from the patient.

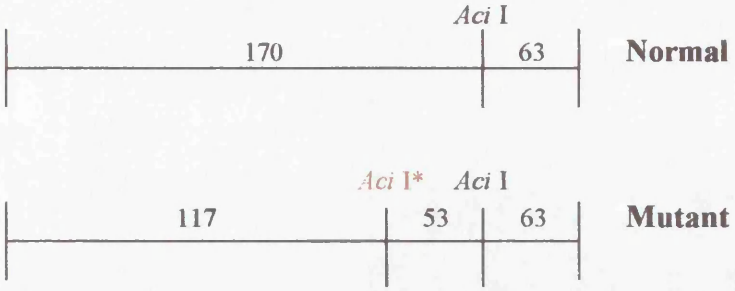


Figure 3.9c

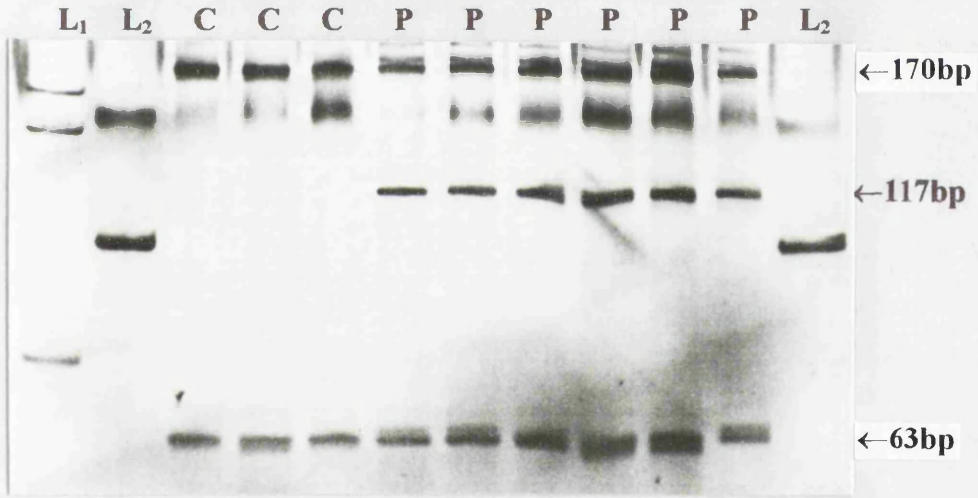


Figure 3.9d

This mutation was identified in a Scottish patient who manifested the main presenting symptoms of this condition which included abdominal pain and vomiting. It is suspected that these symptoms had arisen due to calorie restriction. Her erythrocyte HMB-S activity was found to be 18 nmol/hr/ml (normal range 30 - 67 nmol/hr/ml) upon biochemical analysis, which is very much reduced from the normal range.

Several members of a Scottish family were also found to carry the L177R mutation. As seen from Pedigree 13543 (Appendix 1), all the offspring of II:3 have the defective gene. Initially, III:6 was thought to be a sporadic case of AIP. Subsequent screening of family members revealed that this was not the case. Her siblings, mother and mother's half-sister (who was not analysed in this study but was diagnosed as suffering from AIP based on her biochemical values) were identified as carriers of the defective gene. IV:3 and IV:4, 12 and 9 years old respectively, had a presymptomatic testing performed and were found not to have inherited the defective gene.

3.3.1.2 FRAMESHIFT MUTATIONS

(i) 771insT (EXON 12)

This novel frameshift mutation was detected in one South African patient with a very severe phenotype. This mutation showed a positive result on heteroduplex analysis (Figure 3.3), and direct sequencing revealed a one base-pair insertion at position 771 (Figure 3.10a) which resulted in the formation of a premature stop codon 33 amino acids downstream from the insertion. *Sty* I digestion was done as a confirmatory step (Figures 3.10b and 3.10c). This frameshift mutation (Ong et al, 1996) resulted in the alteration of the thirty-two amino acids following the insertion, prior to the formation of a premature stop codon. The resulting truncated protein, if stable, is anticipated to contain 289 amino acids instead of the normal 361 i.e. a loss of approximately 20% of the translated protein is predicted and this causes impaired protein function.

Gu et al (1994) have reported the largest insertion in the *HMB-S* gene: 8bp were inserted in exon 12, resulting in the formation of a stop codon 10 amino acids away. Large insertions are not a common feature in the molecular pathology of AIP.

Figure 3.10a: Sequence analysis of exon 12 (771insT).

Sequencing the entire *HMB-S* cDNA revealed a 771insT mutation in the South African patient 632 as indicated but not in 20 normal controls. N = normal control, M = mutant sample.

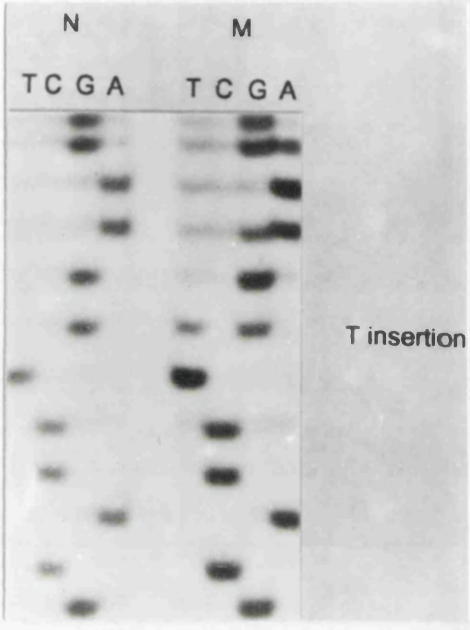


Figure 3.10a

Figure 3.10b: Restriction map of *Sty* I to detect the 771insT mutation.

Restriction map of the amplified cDNA fragment using primers F666 and 508D (Table 2.3) showing the restriction fragment sizes expected for the normal (431bp and 12bp) and mutant (340bp, 91bp and 12bp) alleles upon digestion with *Sty* I. The *Sty* I* restriction site is created as a result of 771insT, producing a 340bp fragment, in addition to the 431bp fragment derived from the normal allele.

Figure 3.10c: Restriction analysis of 771insT with *Sty* I.

Restriction enzyme digestion with *Sty* I on a 3% SeaKem (GTG) agarose gel. L = 100bp ladder (Gibco BRL), C = normal control, P = digested PCR products from the patient. The presence of a 340bp restriction fragment from South African patient 632, in addition to the 431bp fragment derived from the normal allele, is indicative of the mutation. This mutation was not detected in 20 healthy individuals.

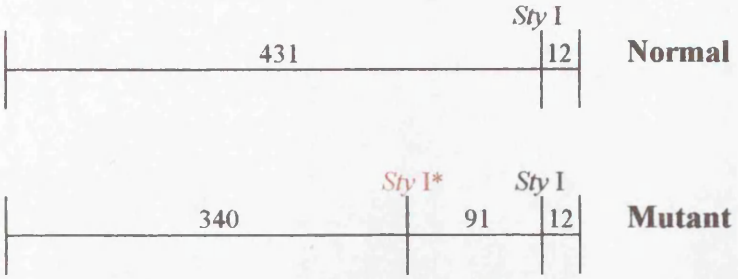


Figure 3.10b

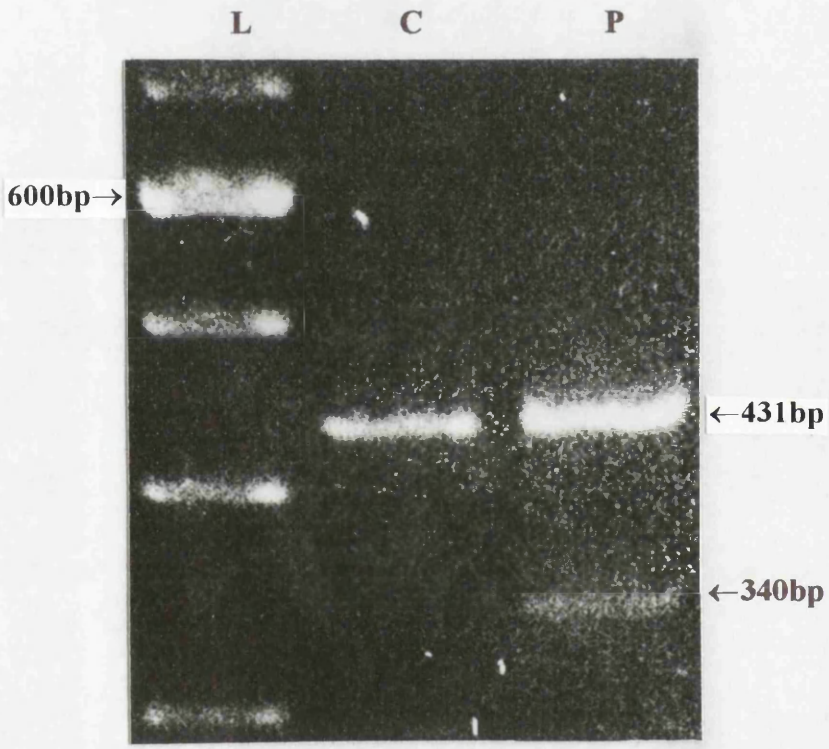


Figure 3.10c

(ii) delG$_{1002/1003/1004}$ (EXON 15)

The patient with this mutation suffered a severe course of the disease and finally succumbed to the disease. CCM and heteroduplex analyses both revealed abnormal bands (Figures 3.11a and 3.3 respectively) and was identified as being due to a deletion of either one of three G nucleotides at positions 1002 to 1004 of the coding sequence (Figure 3.11b). This mutation in exon 15 results in seven altered amino acids following the deletion, resulting in a premature stop codon 8 codons downstream from the point of deletion and is anticipated to result in the loss of approximately 5% of the mature protein. A truncated protein containing 342 amino acids instead of the normal 361 is produced.

This mutation provided some difficulty in assigning exactly which nucleotide was deleted. This is because three consecutive Gs are present in this region. As seen from Figure 3.11b, the sequence read in the sense direction shows that G_{1004} is deleted, whereas G_{1002} is deleted when the sequence is read in the antisense direction. At this stage, the conclusion which can be drawn about this mutation is that one of the three Gs at positions 1002, 1003 or 1004, has been deleted which, consequently, results in the formation of a truncated protein.

From Table 4.1, it is observed that the deletions reported so far involve only one or two bases. Large deletions, like insertions, are not a common feature in the molecular pathology of AIP.

This patient had suffered four acute attacks, two of which were mild and another two were severe. One of the severe attacks was associated with some neuropathy and the other was marked by severe hyponatraemia and convulsions, which required the administration of hypertonic saline. This patient's father and sister are known to carry the defective gene but have remained well. This patient finally succumbed to the disease. This case is proof that AIP is a disease of variable expression even among individuals with the same genotype, thus making the correlation of genotype with phenotype almost impossible.

Figure 3.11a: Chemical cleavage mismatch analysis showing the cleavage product due to the $delG_{1002/1003/1004}$ mutation.

The fragment amplified for CCM analysis using hydroxylamine with the primers 019H and 020H (Table 2.3) in one South African patient and 20 healthy individuals gave a product of 1.3kb. A cleavage product in the region of 1020-1022bp was present in the South African patient but not in 20 healthy individuals analysed. This indicates that a mismatched cytosine or guanine is present in this region of the cDNA. C = control, P = patient.

Figure 3.11b: Sequence analysis of exon 15 ($delG_{1002/1003/1004}$).

Partial sequence of the *HMB-S* cDNA of the South African patient in the region of the cleavage product in Figure 3.11a, both of the antisense (AS) and sense (S) strands, showing the deletion of G_{1002} in the former and G_{1004} in the latter. This sequencing pattern was not observed in 20 normal controls.

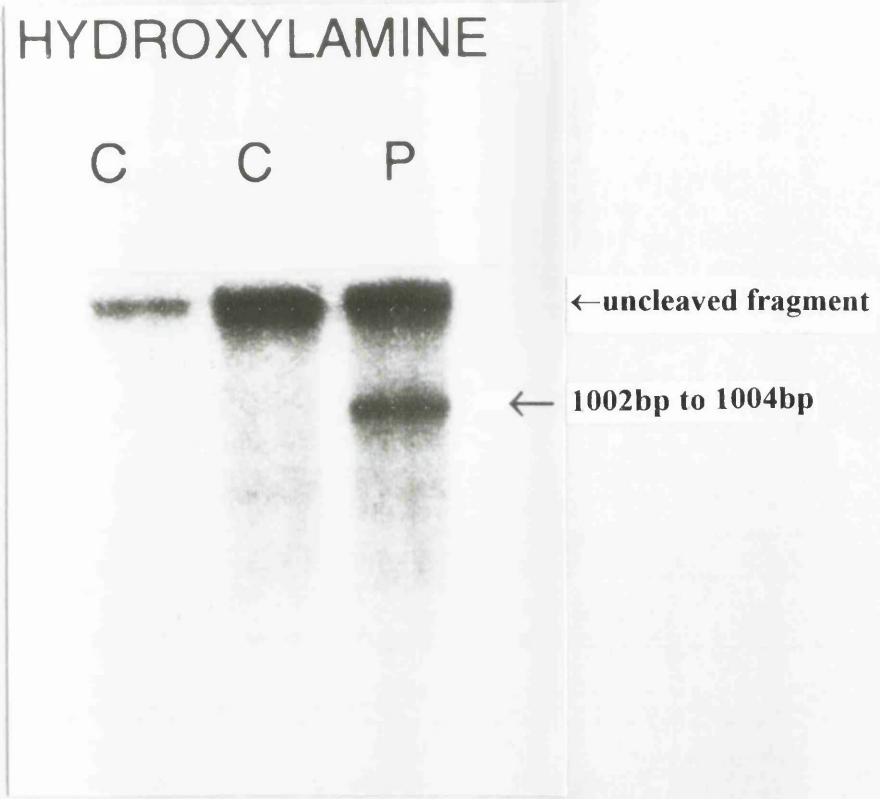


Figure 3.11a

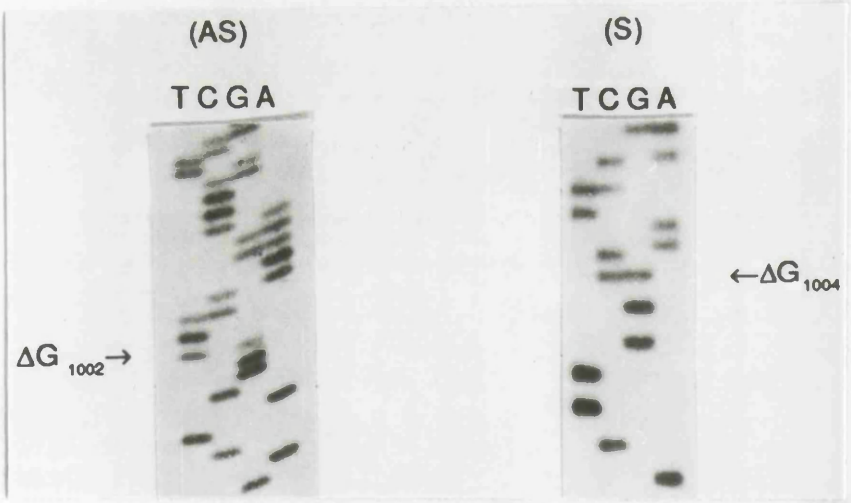


Figure 3.11b

3.3.1.3 SILENT MUTATIONS AND POLYMORPHISMS

(i) S45S (EXON 4)

Direct sequencing revealed a G to A transition at position 135 (Figure 3.12) which maintains the serine residue at position 45. This sequence alteration was not detected in the other 29 patients and therefore, is considered to be a silent mutation. This sequence variation was detected in one patient with the L177R mutation and this phenomenon was also reported by Mgone et al (1992). However, this base change was not detected in the other five patients with the L177R mutation, thus excluding the possibility of this silent mutation being in linkage disequilibrium with L177R.

(ii) V202V (EXON 10)

This is a previously reported intragenic polymorphism (Figure 3.13) reported by Mgone et al (1992) which maintains the valine residue at position 202.

3.3.2 LARGE ALTERATIONS

(i) Exon 8 skipping

The only mutation which showed the presence of a shorter transcript after RT-PCR was that of exon 8 skipping in a father and two of his daughters (Appendix 1, Pedigree 13519, individuals III:2, IV:1 and IV:2). The region surrounding exon 8 was amplified and a shorter product was seen (Figure 3.14a). This shorter product could be due to the absence of either exon 7 or exon 8, both of which are equal in size. Therefore, direct sequencing of the entire segment was carried out and exon 8 was found to be missing. Intronic primers I7F and I8R (Table 2.3), flanking exon 8, were designed and used initially for screening the region flanking exon 8 in the genomic DNA by HA (Figure 3.14b), and then for sequencing across the exon-intron boundaries of genomic DNA. A G to A transition at position -1 of the invariant AG dinucleotide at the 3' acceptor splice site of intron 7 (Figure 3.14c) was responsible for the skipping of exon 8. This mutation was

Figure 3.12: Sequence analysis of the exon 4 silent mutation (S45S).

Partial sequence analysis of the *HMB-S* cDNA showing the $G_{135}\rightarrow A$ transition, resulting in the silent mutation S45S. This silent mutation was identified in a patient with the L177R mutation, and was previously identified independently in another Scottish patient by Mgone et al (1992). Therefore, all the other patients with the L177R mutation, as well as the other AIP patients and 20 normal controls, were also analysed for this silent mutation but was not found in any of the individuals analysed. From the results obtained, this sequence alteration was found to be a silent mutation which was not in linkage disequilibrium with the L177R mutation.

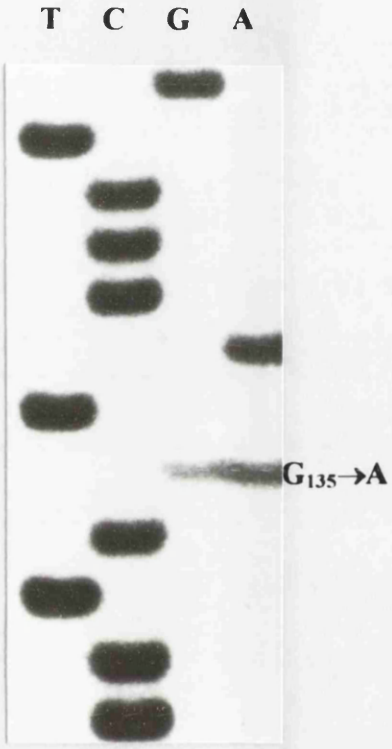


Figure 3.12

Figure 3.13: Sequence analysis of the exon 10 polymorphism (V202V).

Partial sequence analysis of the *HMB-S* coding sequence showing the G/T polymorphism at position 606 as indicated.

Panels 1 and 3 show homozygosity for the G and T alleles respectively and panel 2 shows G/T heterozygosity.

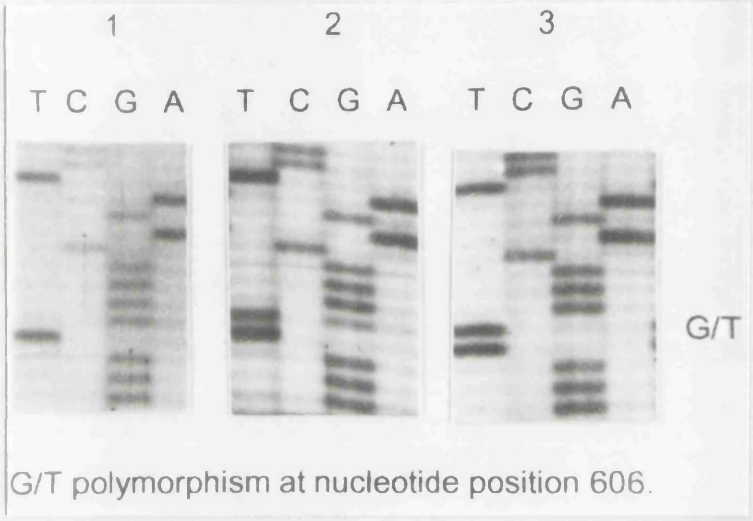

1
2
3
T C G A
T C G A
T C G A
G/T
G/T polymorphism at nucleotide position 606.

Figure 3.13

Figure 3.14a: RT-PCR products showing the absence of exon 8.

RT-PCR products using the primers Phe6+ and R478 (Table 2.3) to yield a product of 231bp. A shorter transcript (154bp) in 3 patients from England (III:2, IV:1 and IV:2 of Pedigree 13519 in Appendix 1) was present but not in 20 negative controls (231bp). L = 100bp ladder (Gibco BRL), C = negative control, P = patient.

Figure 3.14b: Heteroduplex analysis showing positive results in three patients.

Heteroduplex analysis of genomic DNA in the region of the exon skipping using primers I7F and I8R revealed a slower moving heteroduplex in the three patients but not in 20 normal controls. The commercial heteroduplex control (350bp) was run alongside the samples. L = 1kb ladder (Gibco BRL), HC = commercial heteroduplex control (350bp) from Flowgen, C = normal control, P = patient.

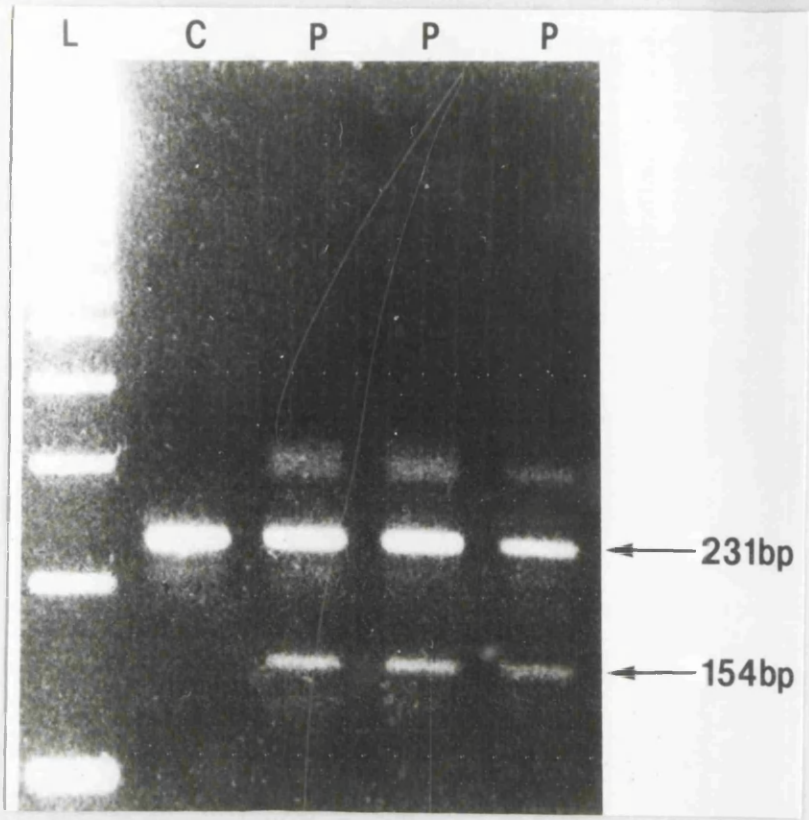


Figure 3.14a

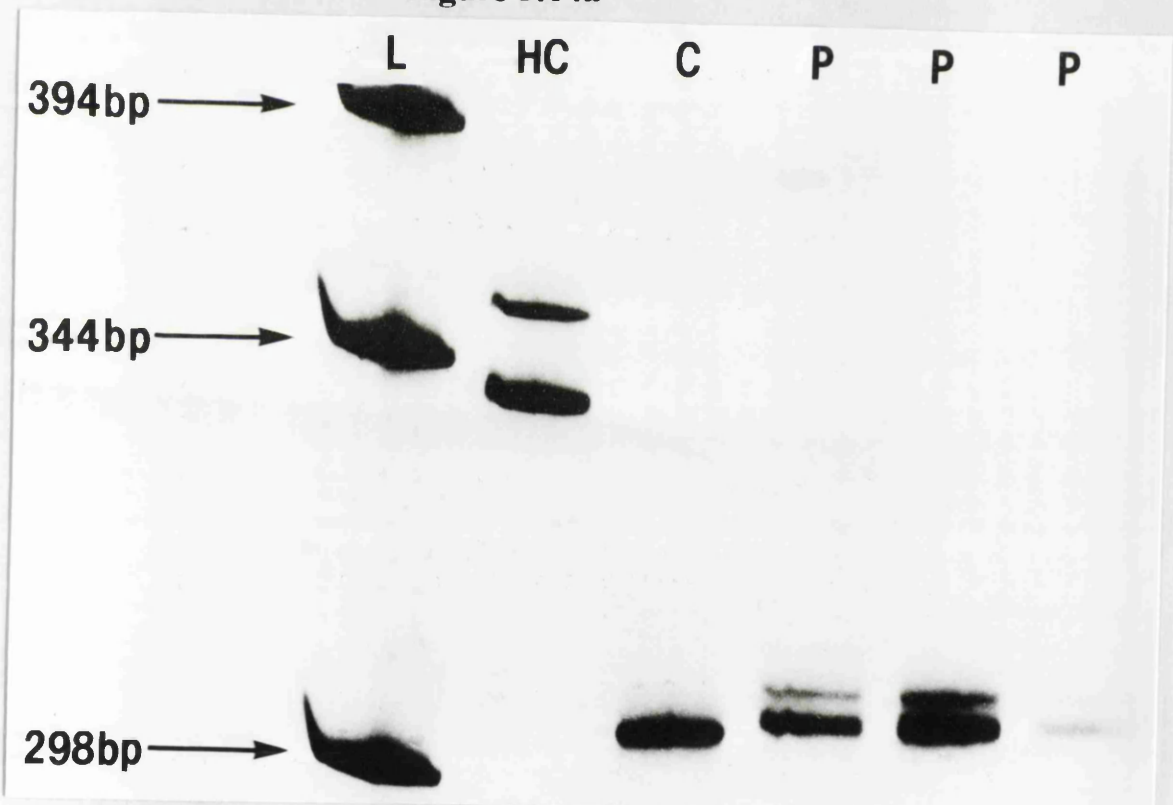


Figure 3.14b

Figure 3.14c: Sequence analysis of intron 7 (345-1G→A).

Partial sequence analysis of intron 7 acceptor splice site of the genomic DNA from the three patients, III:2, IV:1 and IV:2 of Pedigree 13519 (Appendix 1), revealed a G to A transition at the last position of this splice site as indicated. This sequence alteration was not present in 20 healthy individuals. P = patient, C = normal control.

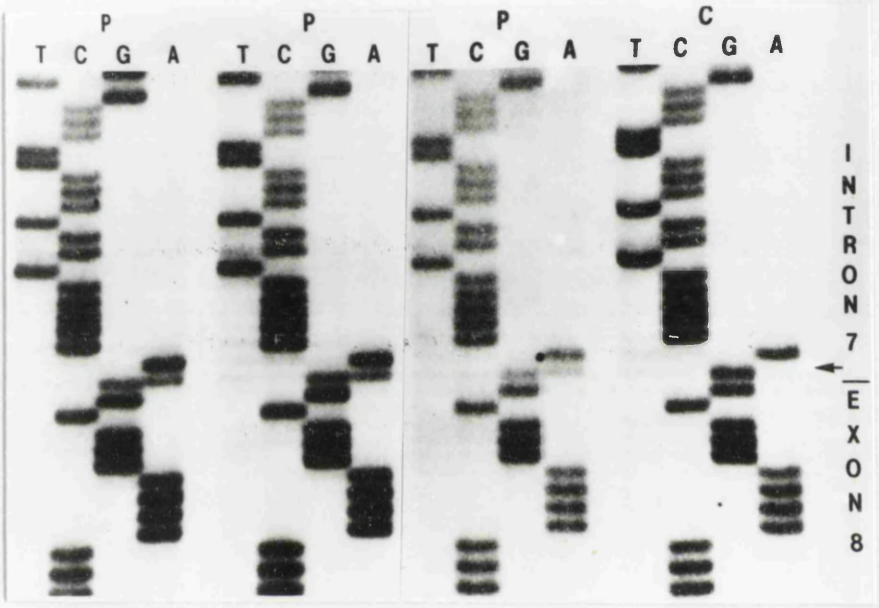


Figure 3.14c

Figure 3.14d: Restriction map of *Eco* RII to detect the 345-1G→A mutation.

Restriction map of the amplified genomic DNA fragment flanking exon 8 using primers Phe6+ and I8R (Table 2.3) showing the restriction fragment sizes expected for the normal (220bp, 208bp, 254bp, 92bp, 39bp, 22bp, 21bp, 9bp and 3bp) and mutant (220bp, 208bp, 346bp, 39bp, 22bp, 21bp, 9bp and 3bp) alleles upon digestion with *Eco* RII. The *Eco* RII* restriction site is removed as a result of the mutation, yielding an additional 346bp fragment.

Figure 3.14e: Restriction analysis of 345-1G→A with *Eco* RII.

Restriction enzyme digestion with *Eco* RII on a 3% SeaKem (GTG) agarose gel. In Pedigree 13519, individuals III:2, IV:1 and IV:2 revealed an additional 346bp fragment upon digestion with *Eco* RII. This fragment was not present in 20 healthy individuals. A 346bp fragment (*) indicates the loss of the *Eco* RII* restriction site due to the mutation. L = 100bp ladder (Gibco BRL), C = normal control, P = digested PCR products from the patient.

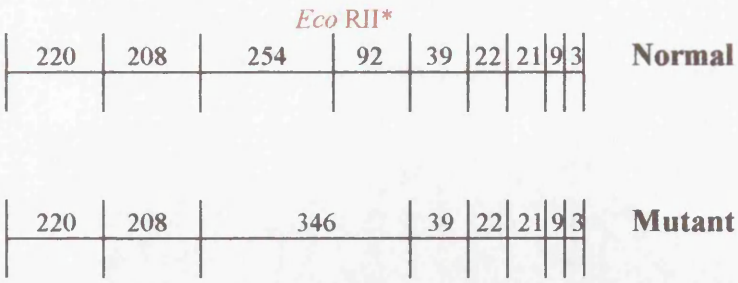


Figure 3.14d

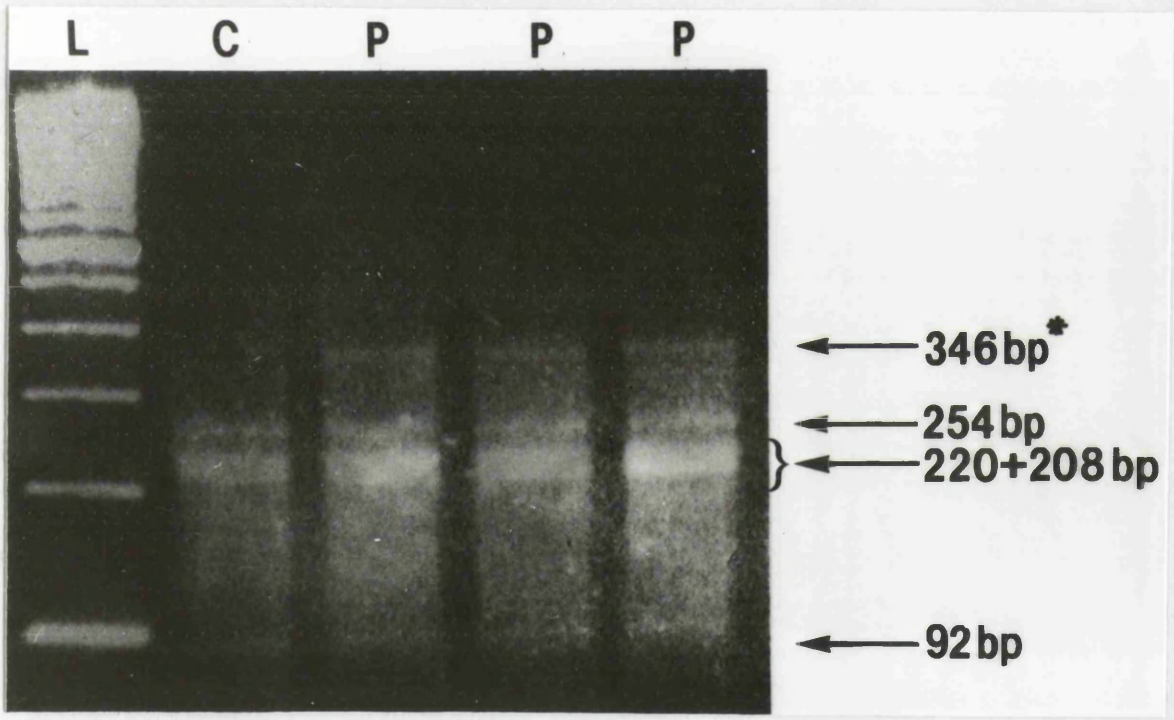


Figure 3.14e

confirmed by an *Eco* RII digestion (Figures 3.14d and 3.14e). This mutation does not disrupt the translational reading frame but is predicted to result in the exclusion of 26 amino acids from the mature protein product. The codon at the junction of exons 7 and 8 of the normal mRNA is now altered i.e. AAG (lysine) to AAT (asparagine). The basic amino acid, lysine, is substituted with asparagine, a neutral amino acid and this is predicted to cause the mature protein to be unstable.

This family was identified as there was a family history of the condition. III:2 had an increased urinary excretion of PBG (97mg/l), compared with the reference samples (0 - 0.3mg/l). He has also suffered recurrent crises of abdominal pain, the cardinal feature of AIP. Presymptomatic testing for IV:1 and IV:2, who are 11 and 8 years old respectively, revealed that they have inherited the mutant allele from III:2. They can, therefore, be appropriately advised to avoid the precipitants of AIP.

From Table 2.1, there is a wide range of erythrocyte HMB-S activities among the South African patients, ranging from 5.05 nmol PBG/g Hb/hr to 12.1 nmol PBG/g Hb/hr. However, a lower erythrocyte HMB-S activity is not associated with a more severe phenotype. Patient 632 has an erythrocyte activity of 9.41 nmol PBG/g Hb/hr, which is higher than that for patients 638 (5.41 nmol PBG/g Hb/hr) and 637 (8.11 nmol PBG/g Hb/hr). The former has the most severe phenotype among all the South African patients. 637 also has a higher erythrocyte activity than two other patients (644 and 638) but presents with a much more severe phenotype. The urinary PBG excretion is not a good indication of the severity of the condition. 632 has the highest level of urinary PBG excretion and this can be a preliminary indication of the degree of severity. However, if the values for the other patients are compared, this correlation is invalid. Patient 639 has a relatively low, although elevated, level of urinary PBG excretion (88.6μmol/10mmol creatinine) but, had nevertheless, previously suffered a very mild attack. Patient 635, on the other hand, has a markedly elevated urinary PBG excretion (233.0μmol/10mmol creatinine) but remains asymptomatic.

The type of mutation the patients have are all very destructive to the mature protein. Despite this, the patients present with a range of phenotypes. A possible reason for this

difficulty in correlating the genotype with the phenotype is that AIP is a disease of variable expression, with environmental factors influencing the degree of manifestation of the condition. This, to some extent, makes this correlation impossible. Another reason could be that insufficient clinical and biochemical data for some of the patients may hinder such a correlation to be made.

A G/T polymorphism in exon 10 at nucleotide 606, previously reported by Mgone et al (1992), was detected in three unrelated South African patients with the R116W mutation and two Irish patients with the R26C mutation. Initially, it was thought that this intragenic polymorphism was in linkage disequilibrium with the R116W mutation. However, this was discounted since it also exists with two other mutations, $G_{500}\rightarrow A$ and $C_{100}\rightarrow A$ (Mgone et al, 1992), and in this study, $C_{76}\rightarrow T$, all of which are likely to be responsible for AIP.

An interesting application of this polymorphism was observed for Pedigree 13168. III:1 and III:3 carry the R26C mutation but do not have the V202V polymorphism. On the other hand, III:2 and III:5 are free from the defective gene but possess the V202V polymorphism. On screening the coding sequence of the *HMB-S* gene of II:6, he was found not to have the polymorphism. Therefore, the genotype of the deceased individual, II:7, can be deduced to be positive for the R26C mutation and the V202V polymorphism, both of which are present on different chromosomes.

The S45S silent mutation was found in a Scottish patient with the L177R mutation. This association was previously reported by Mgone et al (1992) but does not provide sufficient evidence that S45S is in linkage disequilibrium with L177R because it is not detected in the other five patients with the same mutation, thus excluding the possibility of linkage disequilibrium.

3.4 ALTERNATIVE SPLICING OF *HMB-S* mRNA REMOVES EXONS 3 AND 12

In addition to the occurrence of alternative splicing of *HMB-S* mRNA which excludes exon 1 from the erythroid cells and exon 2 from the housekeeping cells, exons 3 and 12 were found, in this study, to have undergone varying degrees of alternative splicing, both in thirty AIP patients and forty-four healthy individuals (Figures 3.15a and 3.16a). Sequencing the cDNA in the region of exon skipping revealed that exon 1 was directly spliced to exon 4 and that exon 11 was precisely spliced to exon 13 (Figures 3.15b and 3.16b respectively). The transcript missing exon 3 is 54bp shorter than the normal transcript and that missing exon 12 is 120bp shorter. The ratio of PCR products generated during the log-phase of amplification for the regions flanking exon 12 (19 cycles of amplification) shown in Figure 3.16c was the same as that generated after 35 cycles (Figure 3.16d), indicating that these ratios reflect the true concentrations of the differentially spliced mRNA molecules *in vivo*. No PCR products were visible on the agarose gel after 12 cycles of amplification. In addition, exon 3 alone was found to be missing in a proportion of the transcripts, as well as in a proportion of those transcripts with exon 12 also missing (Figure 3.17). Approximately 10 to 50% of the transcripts were observed to be alternatively spliced. Sequencing the DNA in the vicinity of the exon-intron boundaries did not reveal any abnormal sequences. Analysis of the other exons of the *HMB-S* mRNA by RT-PCR revealed that none of them had undergone alternative splicing (Figures 3.17a, 3.17b, 3.17c and 3.17d). All the individuals who showed alternative splicing of exons 3 and 12 were analysed on three further occasions and the pattern of alternative splicing was reproducible.

A schematic representation of alternative splicing of exons 3 and 12 is depicted in Figure 3.18. The primers used to detect exon 3 skipping were 377H and Phe4-, while Phe10+ and 749D were used to detect exon 12 skipping.

Figure 3.15: RT-PCR and sequence analysis of the region showing exon 3 skipping.

(a): RT-PCR results of a representative group of individuals showed two bands, 131bp and 77bp, whose size difference corresponds to the absence of exon 3. The upper panel shows the RT-PCR products of 19 AIP patients and the lower panel represents RT-PCR products of 19 normal individuals. In both groups of patients, it was observed that exon 3 was spliced out to varying degrees.

(b): Sequence analysis of the excised upper and lower bands from **(a)** showed that exon 3 is missing from the lower band, resulting in exon 1 being directly spliced to exon 4.

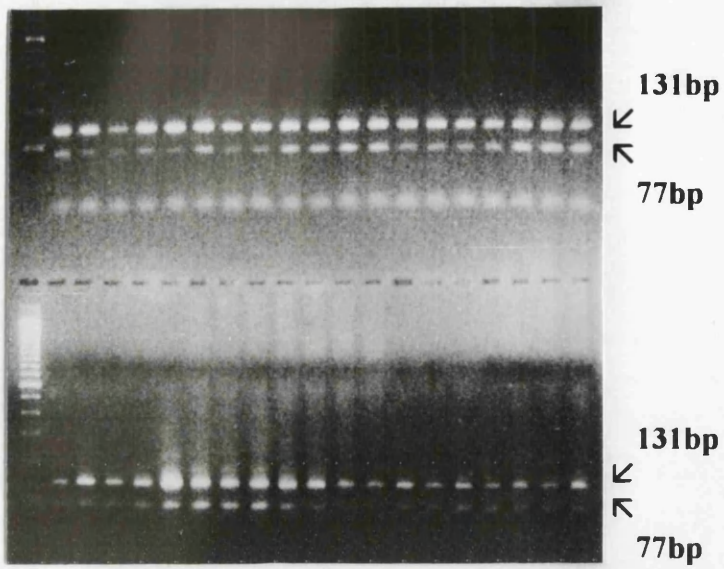


Figure 3.15a

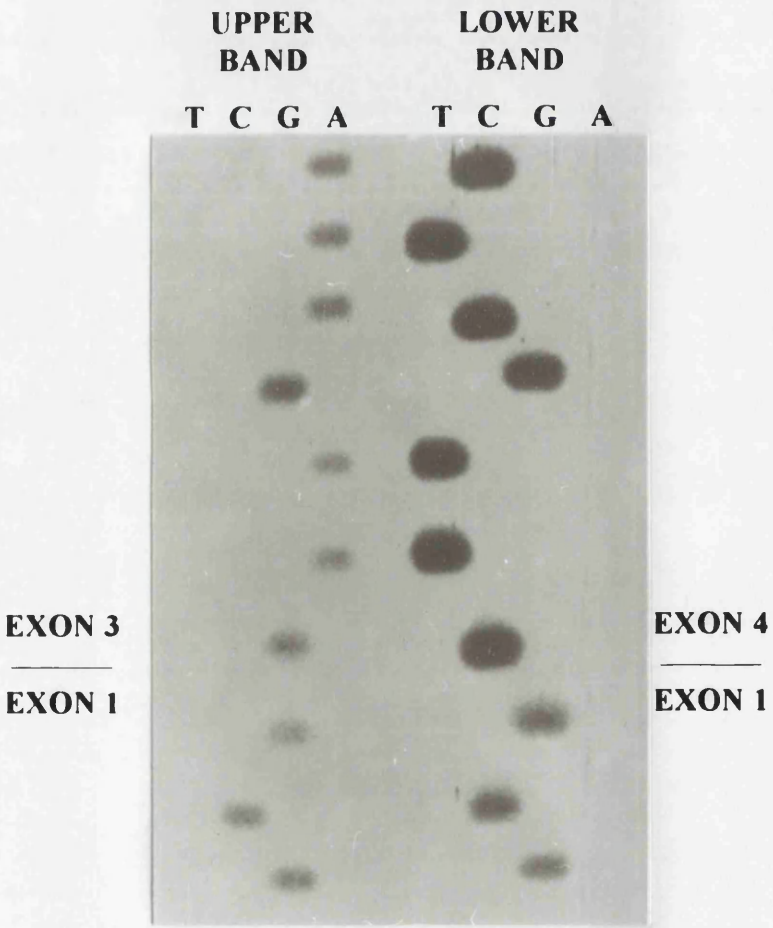


Figure 3.15b

Figure 3.16: RT-PCR and sequence analysis of the region showing exon 12 skipping.

(a): RT-PCR results of a representative group of individuals showed two bands, 378bp and 258bp, whose size difference corresponds to the absence of exon 12. The upper panel shows the RT-PCR products of 19 AIP patients and the lower panel represents RT-PCR products of 19 normal individuals. In both groups of patients, it was observed that exon 12 was spliced out to varying degrees.

(b): Sequence analysis of the excised upper and lower bands from **(a)** showed that exon 12 is missing from the lower band, resulting in exon 11 being directly spliced to exon 13.

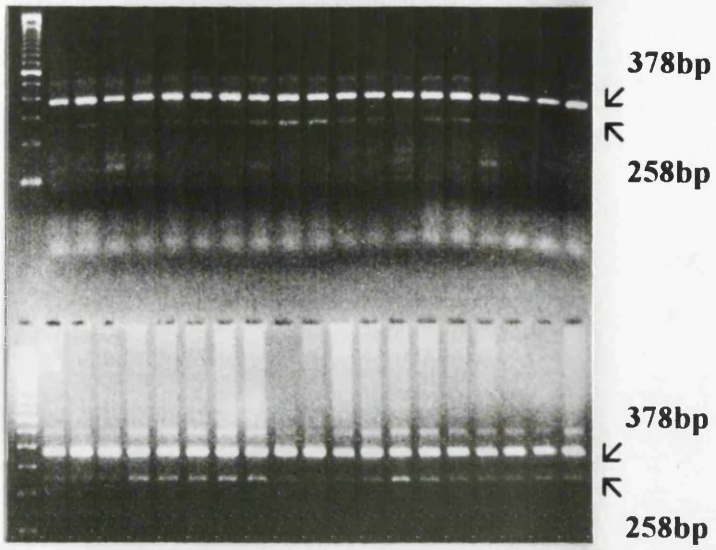


Figure 3.16a

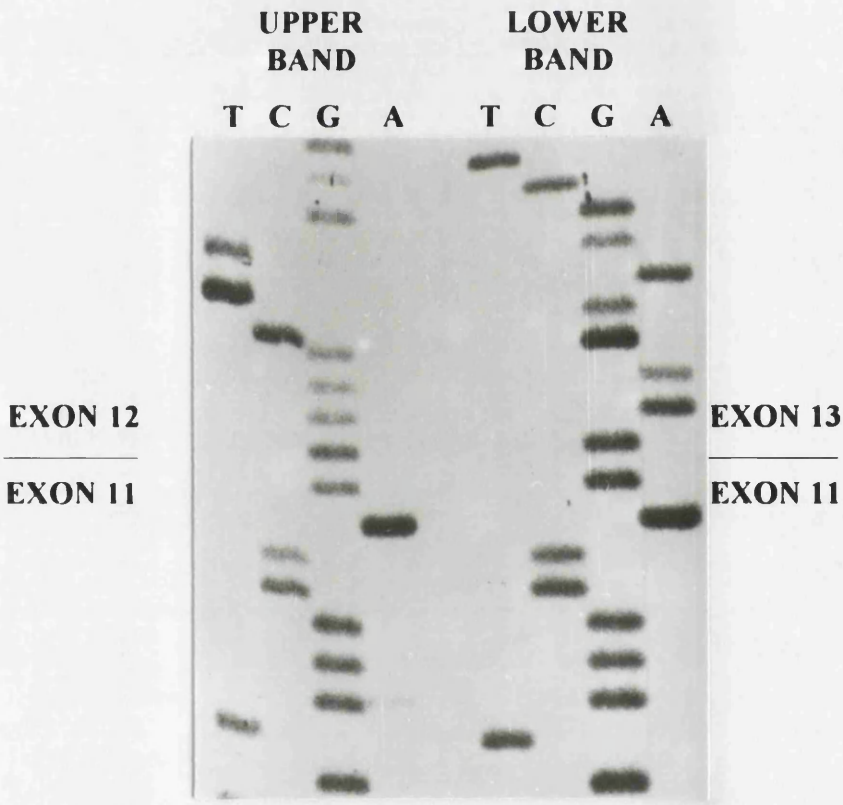


Figure 3.16b

Figure 3.16c: RT-PCR products of four individuals after 19 cycles of amplification.

The four individuals show varying degrees of splicing out of exon 12 after 19 cycles of amplification (log-phase PCR).

Figure 3.16d: RT-PCR products of the four individuals from Figure 3.16c after 35 cycles of amplification.

The RT-PCR products of the individuals in Figure 3.16c reveal a similar ratio of amplification products as in Figure 3.16c after 35 cycles of PCR.

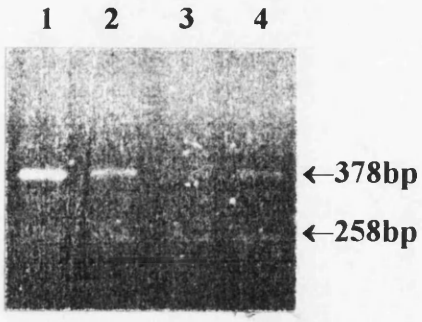


Figure 3.16c

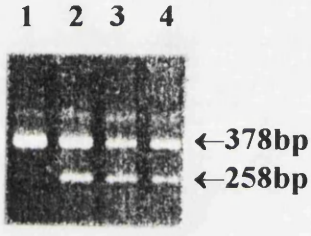


Figure 3.16d

Figure 3.17: Evidence for 4 different mRNA transcripts in the lymphocyte population.

The full length *HMB-S* cDNA was amplified using the primers 019H and 020H (Table 2.3), giving a fragment of 1.3kb. The amplified products were analysed on a 2% Metaphor gel (Flowgen), where 4 different mRNA transcripts were present as described below:

L:	1kb ladder (Gibco BRL)
Lane 1:	intact transcript (i.e. exons 3 and 12 present)
Lanes 2 to 7, 9, 10:	exons 3 and 12 missing from the smallest allele
Lanes 8, 11, 12:	exon 12 missing from the smaller allele
Lane 12:	exon 3 missing from the larger allele

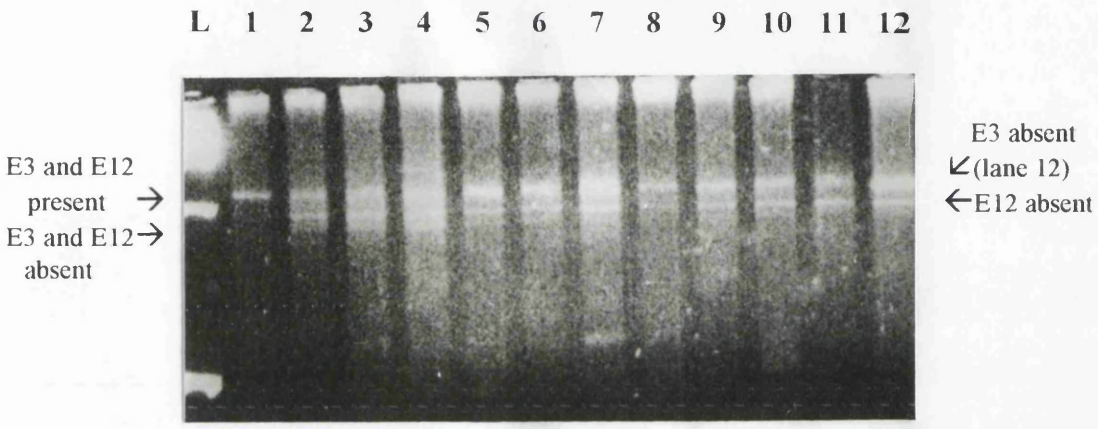

L 1 2 3 4 5 6 7 8 9 10 11 12
E3 and E12 present →
E3 and E12→ absent
E3 absent
↙(lane 12)
←E12 absent

Figure 3.17

Figure 3.17a: RT-PCR products of the region spanning exons 1 to 8 of a number of individuals.

The region spanning exons 1 to 8 of the *HMB-S* mRNA did not show any extraneous bands due to exon skipping, except that due to exon 3 skipping as seen by the presence of a shorter transcript. L = 100bp ladder.

Figure 3.17b: RT-PCR products of the region spanning exons 7 to 11 of a number of individuals.

The region spanning exons 7 to 11 of the *HMB-S* mRNA showed the presence of only one population of mRNA transcript (398bp), thus eliminating the possibility of exon skipping in this region. L = 100bp ladder.

Figure 3.17c: RT-PCR products of the region spanning exons 9 to 11 of a number of individuals.

The RT-PCR products encompassing exons 9 to 11 of the *HMB-S* mRNA showed only the expected amplification products (290bp). L = 100bp ladder.

Figure 3.17d: RT-PCR products of the region spanning exons 13 to 15 of a number of individuals.

The RT-PCR products covering exons 13 to 15 of the *HMB-S* mRNA revealed only amplification products of the expected size (442bp). L = 100bp ladder.

Exon 12 was not analysed here as it had already been shown to undergo varying degrees of alternative splicing (Figure 3.16a).

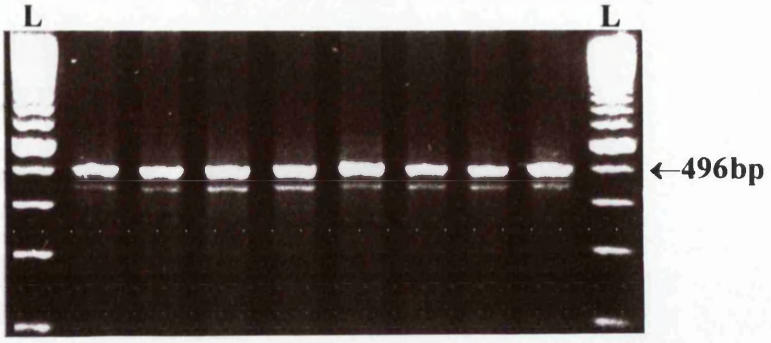


Figure 3.17a

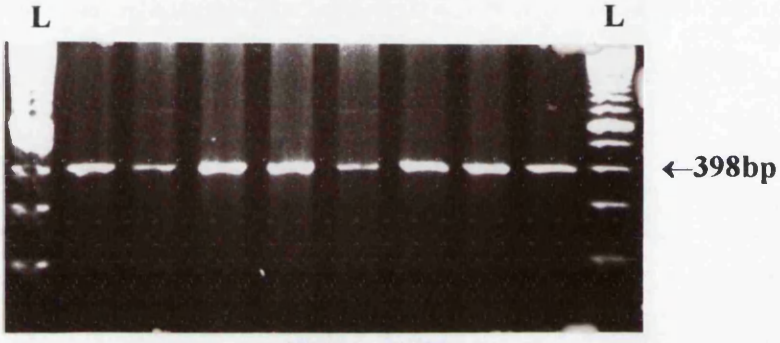


Figure 3.17b

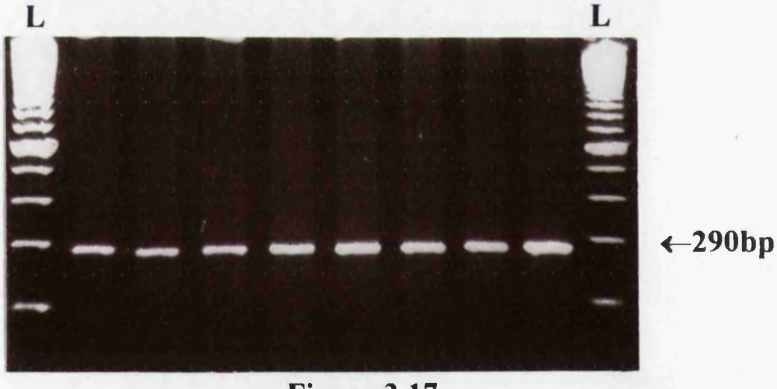


Figure 3.17c

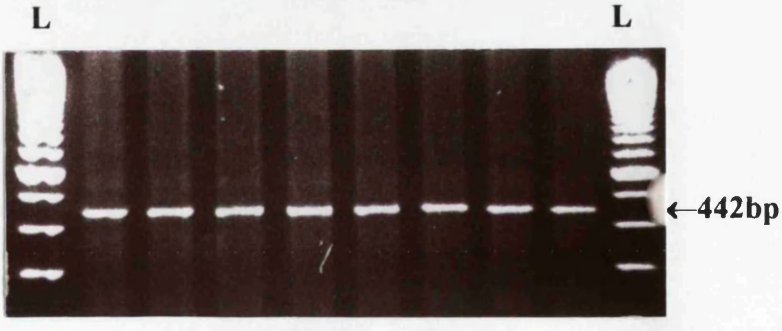


Figure 3.17d

Figure 3.18: Schematic representation of the location of primers used to detect exons 3 and 12 skipping.

A schematic representation showing the location of the primers used in this study (arrowed) to detect the regions of alternative splicing. The upper diagram shows the mRNA transcript of one allele before splicing while the lower diagram shows the transcript after splicing. The exons in bold type are removed as a result of alternative splicing.

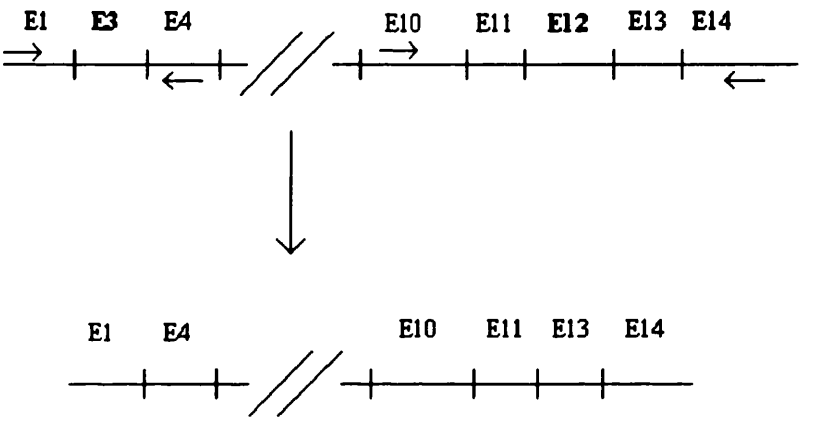


Figure 3.18

3.5 *IN VITRO* GENE EXPRESSION

3.5.1 *HMB-S* RECOMBINANT CLONES

The clones constructed for subsequent *in vitro* expression studies consisted of the plasmid expression vector, pTrc 99A (Pharmacia) and the segment of interest which was amplified by PCR. The expression vector consists of a strong *trc* promoter which directs a controlled, high-levelled expression of foreign inserts. It also possesses a strong transcription termination signal (*rrn*B) downstream.

In this study, three missense mutations, R26H, R116W and R173Q, were successfully expressed in the *in vitro* expression system. An attempt was made to express the other mutations, R22C, R26C, 345-1G→A, L177R, 771insT and $delG_{1002/1003/1004}$, in the system but was unsuccessful.

3.5.2 VERIFICATION OF THE INTEGRITY OF THE CLONES

Since the segment of interest was generated by PCR, its sequence has to be verified before proceeding to expression studies. CCM analysis was the screening method of choice since it enabled the verification of the entire insert in a single analysis. The sequence of the clone was then compared with that of the wild-type. The sequences of the primers used in cloning are shown in Table 2.4.

Of the 126 cloned inserts screened by CCM analysis, 44 (34.9%) did not show any misincorporations by Taq DNA polymerase. 39 of these intact fragments were those of the normal allele while the remaining 5 had incorporated the mutated allele, which were suitable for expression studies. A typical result obtained after CCM analysis consists of a mixture of clones with misincorporated nucleotides (seen as cleavage products of various sizes), clones with the normal allele (no cleavage products) and clones with the mutated allele (various cleavage products corresponding to the location of the mutations present in each clone) (Figure 3.19).

To prove that there were no misincorporations present, the clones not showing any cleavage products, as well as those clones which showed cleavage products of the correct

Figure 3.19: CCM analysis of clones to check for misincorporations by Taq DNA polymerase.

A variety of cleavage products are observed after subcloning the human *HMB-S* coding sequence into the expression vector pTrc 99A, with hydroxylamine being the chemical used for modification.

Lanes 1 to 3, 6:	clones showing random misincorporations
Lanes 4, 5, 7 to 10:	clones not revealing any cleavage products
Lane 11:	positive control showing the cleavage product due to the mutation R116W

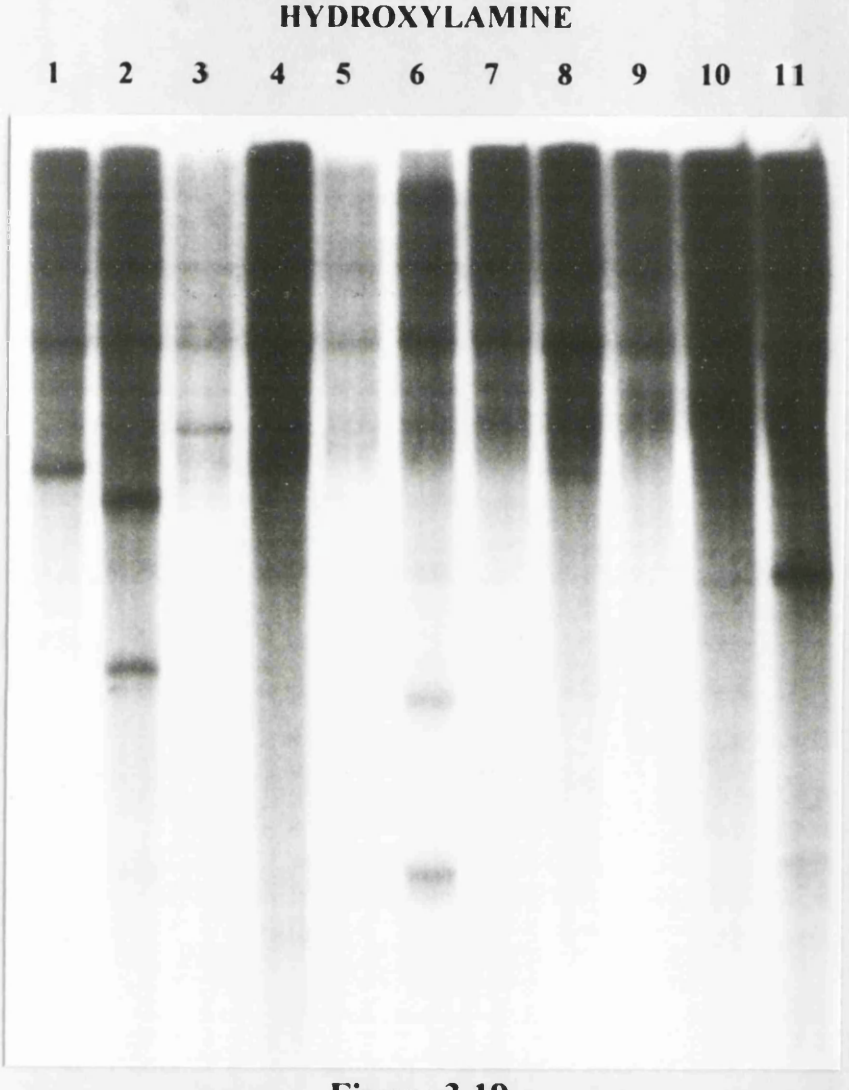


Figure 3.19

sizes, were sequenced to ensure that, in the former case, there were no base changes and in the latter case, the cleavage products were due to the mutation present.

3.5.3 ANALYSIS OF PROTEIN EXPRESSION

After subjecting the *HMB-S* clones to *in vitro* expression, Western blotting was performed to accomplish a threefold purpose: (i) to verify the specificity of the custom-made antibody; (ii) to determine if expression was successful; and (iii) to quantify the amount of proteins generated.

Varying concentrations of primary and secondary antibodies were used and the washing conditions were modified from standard protocols to obtain the highest signal to background ratio (Figures 3.20a and 3.20b). A 1:10000 to 1:5000 dilution of the primary antibody (6.125mg/ml) was found to give the best signal with low background (Figure 3.20b) while a 1:40000 dilution of the secondary antibody was found to be optimal for detection by chemiluminescence. An overnight incubation of the filter in the primary antibody-containing wash buffer and an extended washing time (8 X 15 minutes) also ensured that the primary antibody was specific for the protein against which it was raised.

The success of protein expression in an *in vitro* system was determined by the presence of an additional band at approximately 44kD, corresponding to the size of the housekeeping protein (Figures 3.20a and 3.20b).

The specificity of the antibody for the human HMB-S protein was verified after all the conditions for Western blotting had been optimised (Figure 3.20b).

3.5.4 SEPARATION OF HUMAN AND *E. coli* HMB-S PROTEINS

3.5.4.1 HPLC

An attempt was made to separate the human and *E. coli* HMB-S proteins by HPLC. The gradient elution programme (Table 2.5) was devised based on the results reported by de Rooij et al (1987) and found to give a relatively good separation. From the chromatograms shown in Figures 3.21a and 3.21b, an extra peak was observed for the clone containing the expressed human HMB-S protein but this was not present in the

Figure 3.20: Optimisation of Western blotting conditions.

The effects of both antibody concentrations and washing stringencies are compared. The samples were run on a 12% SDS-PAGE gel.

(a) The conditions employed are those recommended by the manufacturer of the Western Blotting Detection Reagents (Amersham). The protein of interest is shown (44kD), as well as non-specific bands.

(b) The conditions have been modified such that the specificity of the primary antibody for the protein it has been raised against is maximum (section 3.5.3).

- = clones without the human insert

+ = clones with the human insert

Lane 1: R173Q

Lane 2: negative control i.e. no mutation present

Lane 3: R26H

The additional bands below the protein band of interest (44kD) could be due to the presence of internal methionines within the mRNA transcript which may provide additional initiation codons for translation.

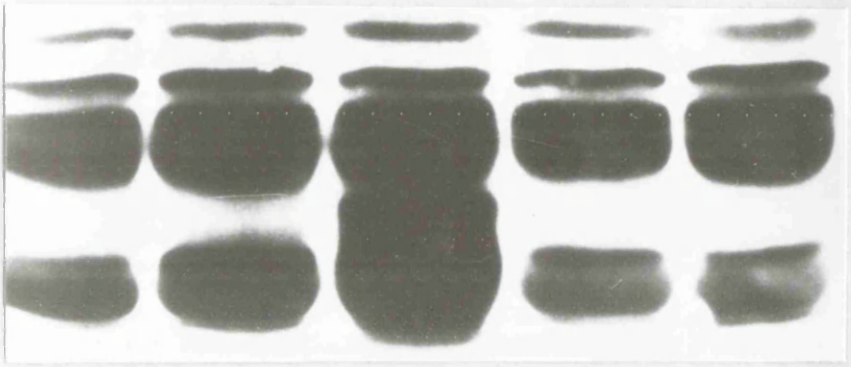


Figure 3.20a

Figure 3.20b

Figure 3.21a: HPLC chromatogram of the clone without the human protein.

The conditions used for this HPLC run are shown in Table 2.5. Three distinct peaks were consistently produced with each run, hence the eluent for each of these peaks was collected for further analysis. The peak, as indicated by the arrow, shows a different profile when compared with **Figure 3.21b**.

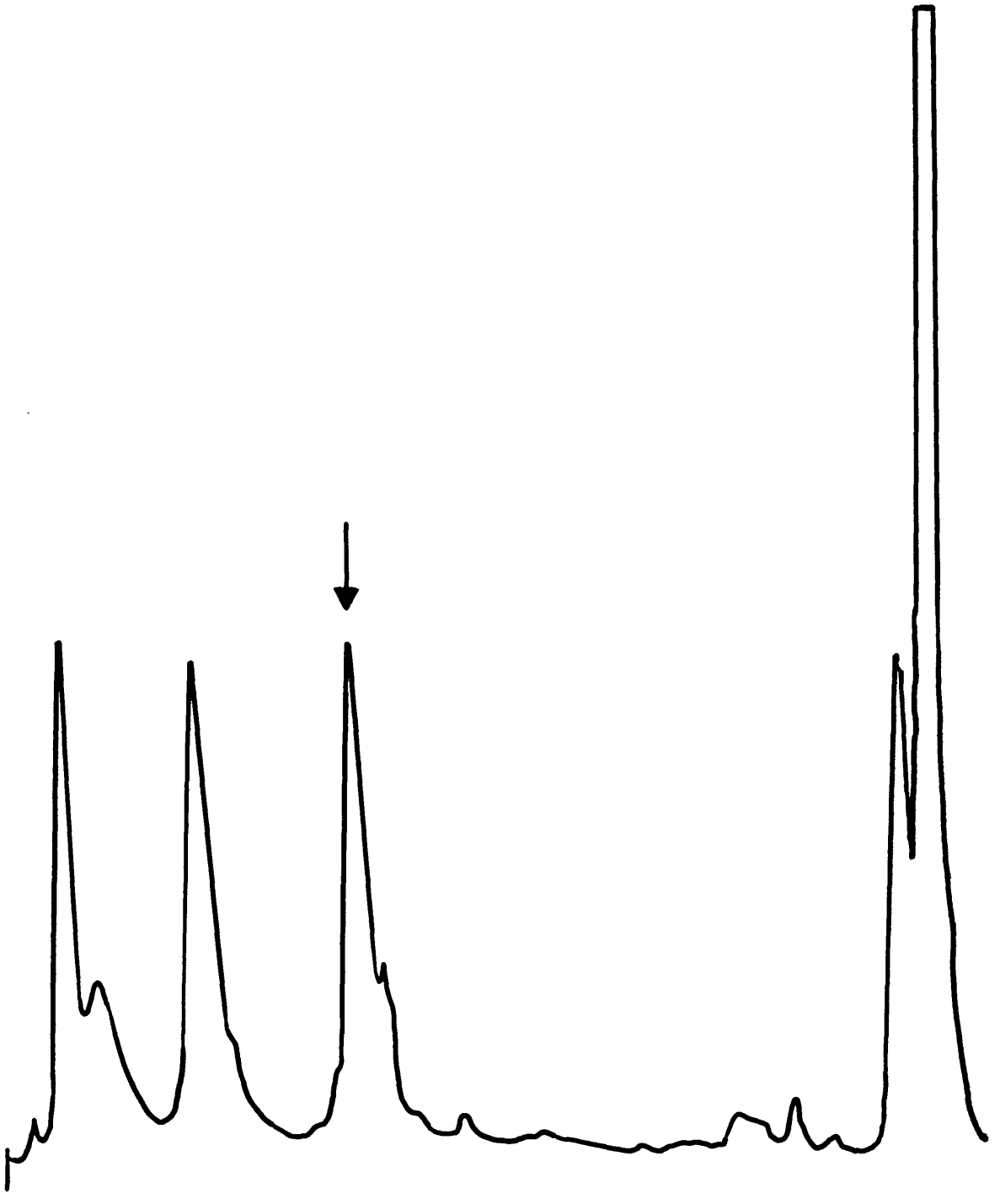


Figure 3.21a

Figure 3.21b: HPLC chromatogram of the clone containing the human protein.

The third peak, as indicated, is significantly different from that in **Figure 3.21a**. Both peaks, in this case, were collected for further analysis.

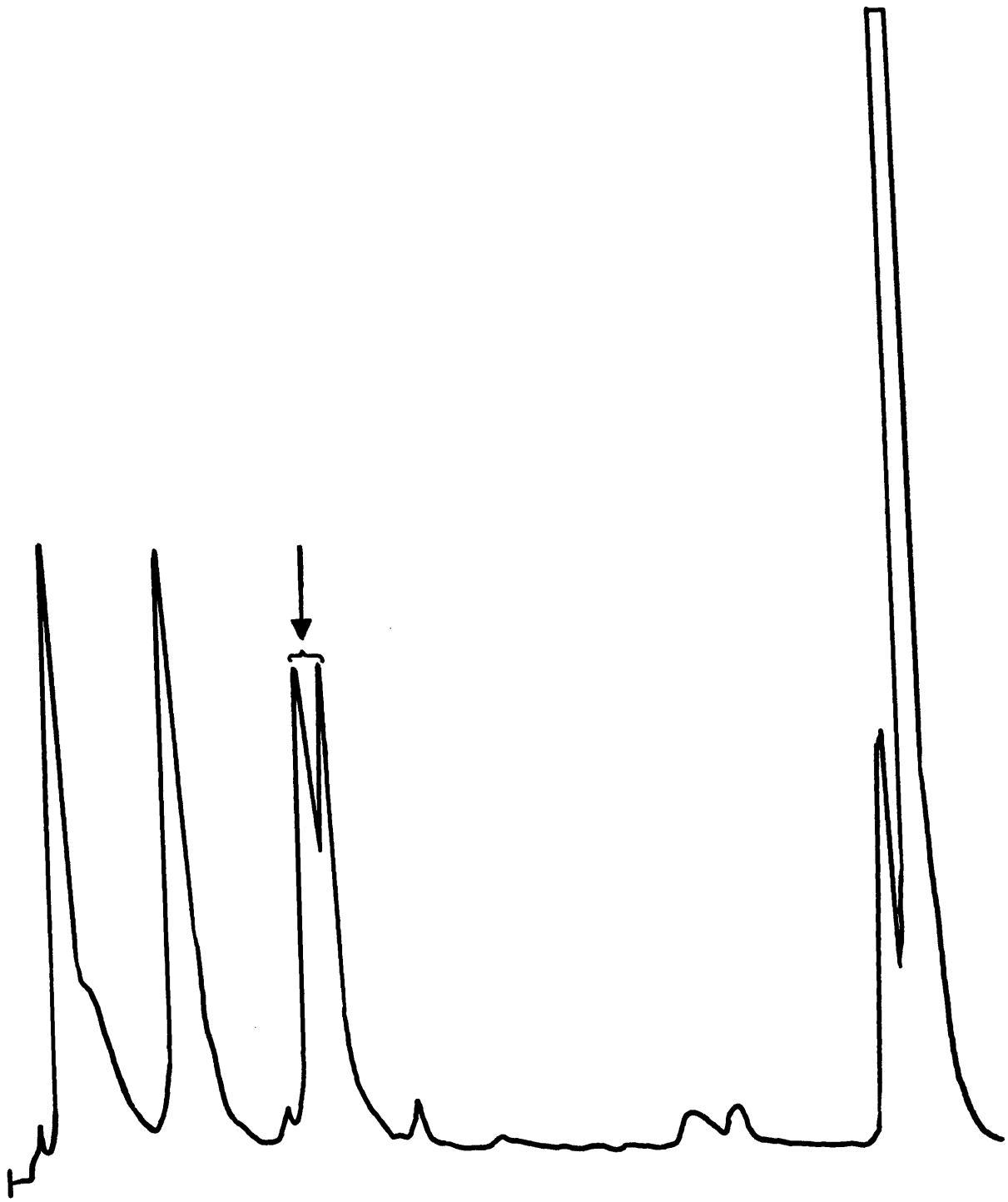


Figure 3.21b

clone without the human protein present. Therefore, the extra peak observed is most likely due to the human form of the protein.

3.5.4.2 IEF

The IEF method of separation has the added advantage in that the proteins can be concentrated in the gel for further manipulations to be done specifically on the protein of interest after separation. The pIs of the two forms of HMB-S protein were first determined and detected by (i) Coomassie Brilliant Blue staining, and (ii) silver staining which is a more sensitive detection method. Coomassie Brilliant Blue staining apparently showed no differences between the samples with and without the expressed human protein (Figure 3.22a). The silver staining method of detection was next employed to determine if the resolution of the bands obtained by Coomassie Brilliant Blue staining could be further improved. Again, no differences were observed among the samples with the human protein present and those without the human protein (Figure 3.22b). The pIs of the two forms of HMB-S protein could be too close for separation by IEF.

3.5.5 HMB-S ACTIVITIES

The total protein lysates of three different mutant clones, viz. R26H, R116W and R173Q, as well as a clone with the normal human sequence and a clone without any human insert, were used for the measurement of HMB-S activity. The latter was performed to establish a baseline level for the activity of the bacterial HMB-S enzyme for comparison with the expression levels of the human enzyme (Table 3.4). For all the clones assayed, the specific activities calculated were the average of two independent assays.

The clones with the three mutations gave relatively similar specific activities. If the specific activity of the clone without the human insert was subtracted from that of the other clones, then it is clear that each of the mutations causes a marked reduction in the activity of the enzyme, yielding an approximate 3 - 4% of residual activity. Therefore, bacterial activity was irrelevant due to the relative efficiency of recombinant human protein expression compared to endogenous bacterial protein expression.

Figure 3.22: Isoelectric focusing of the expressed proteins.

(a) The clone without the human insert was focussed side by side with one containing the human form to look for the presence of an extra band due to the human protein. Coomassie Brilliant Blue staining did not show any differences between the two samples, so silver staining was undertaken to improve the resolution.

Lane 1: clone without the human insert

Lane 2: clone with the human insert

(b) Silver staining of the same samples from **(a)** did not show any improvement.

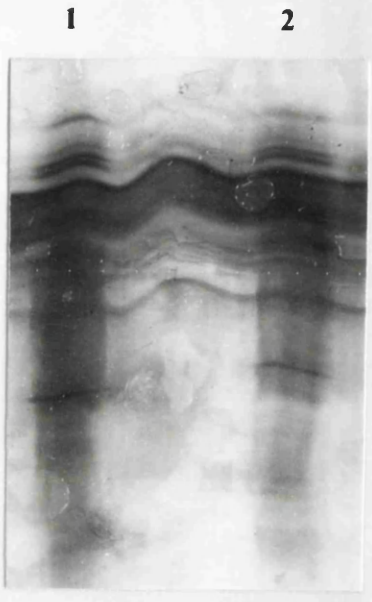


Figure 3.22a

Figure 3.22b

Table 3.4: Correlation of the mutations with their activities as determined by an *in vitro* expression system.

Mutation	Enzyme activity (pmol/ml/hr)	Average specific activity (pmol/mg protein/hr)	Specific activity due to human HMB-S (pmol/mg protein/hr)	% of residual activity
R26H	4.83×10^6	5.8×10^6	1.07×10^6	3.9
R116W	4.72×10^6	5.66×10^6	0.93×10^6	3.4
R173Q	4.63×10^6	5.56×10^6	0.83×10^6	3.0
Normal human insert	26.72×10^6	32.07×10^6	27.34×10^6	100.0
pTrc 99A (no human insert)	3.94×10^6	4.73×10^6	-	-

The enzyme activities of the clones with and without mutations are listed. The HMB-S activity of the clone which does not contain the human insert was also measured to provide a baseline measurement, to which the expression levels of the human enzyme can be normalised.

The specific activity due to the human HMB-S was obtained by subtracting the specific activity of the bacterial enzyme (i.e. pTrc 99A without the human insert) from the clones with and without mutations analysed here.

CHAPTER 4:

DISCUSSION

CHAPTER 4: DISCUSSION

4.1 PATHOPHYSIOLOGY OF AIP

A total of seventy-three mutations have so far been reported in the *HMB-S* gene (Table 4.1). This consolidates the fact that AIP is a heterogeneous disorder. 78.1% of the mutations are single base substitutions (either missense, nonsense or those leading to splicing defects) and the remaining 21.9% involve frameshift mutations (Daimon et al, 1994; Gu et al, 1994; Kauppinen et al, 1995; Mgone et al, 1993; Ong et al, 1996; Schreiber et al, 1994; Schreiber et al, 1995) either due to insertions or deletions (Table 4.1). This is represented diagrammatically as Figure 4.1.

The mutations reported so far in the *HMB-S* housekeeping gene are spread over the entire length of the coding sequence. All the exons have at least one mutation reported, with exon 12 having the highest number of mutations, followed by exon 10 (Figure 4.2). Both exons are, thus, extremely important regions in the HMB-S protein molecule due to their frequent association with the disease (about 38.4%). Therefore, a good strategy for mutation detection of new samples with no previous information would be to focus initially on these two exons before proceeding to the other regions of the coding sequence.

4.1.1 MUTATION ANALYSIS

In this study, nine different mutations were identified and fully characterised in thirteen families. Two families failed to reveal any mutations in their coding sequence. This could be due to the presence of mutations in the promoter regions of the *HMB-S* gene. Six of the detected mutations were missense mutations (R22C, R26C, R26H, R116W, R173Q and L177R), one was a single base substitution affecting a 3' acceptor splice site (345-1G→A), one was a single base deletion ($delG_{1002/1003/1004}$) and one was a single base insertion (771insT). The novel mutations identified in this study were R22C, 345-1G→A, $delG_{1002/1003/1004}$ and 771insT. The mutations which have been previously described were

Table 4.1: Summary of mutations detected in the *HMB-S* gene.

No.	Intron/ Exon	Base change	Nucleotide position	Amino acid change	CRIM-status	Effect of mutation	Reference
1	E1	G→T	33	A11A	–	5' donor splice site mutation	Grandchamp et al (1989b)
2	I1	G→A	33+1	E12K		5' donor splice site mutation	Grandchamp et al (1989c)
3	E3	C→G	66			Skipping of E3	Llewellyn et al (1996)
4	E3	C→T	76	R26C	+	Arg→Cys	Kauppinen et al (1995)
5	E3	G→A	77	R26H	+	Arg→His	Llewellyn et al (1993)
6	I3	G→T	87+1			5' donor splice site mutation	Lundin et al (1995)
7	E4	G→A	91	A31T	+	Ala→Thr	Gu et al (1994)
8	E4	del A	97	97delA	–	Frameshift⇒stop codon 10aa away	Kauppinen et al (1995)
9	E4	C→A	100	Q34K	–	Gln→Lys	Mgone et al (1992)
10	E4	C→T	100	Q34X	–	Gln→Stop	Kauppinen et al (1995)
11	E4	T→A	125	L42X		Leu→Stop	Puy et al (1996)
12	E5	G→T	163	A55S	+	Ala→Ser	Gu et al (1994)
13	E5	del GT	168	168delGT		Frameshift⇒stop codon 64aa away	Schreiber et al (1995)
14	E5	del C	174	174delC	–	Frameshift⇒stop codon 40aa away	Gu et al (1994)
15	E5	ins G	182	182insG	–	Frameshift⇒stop codon 5aa away	Gu et al (1994)
16	I5	G→A	210+1		–	5' donor splice site mutation	Gu et al (1994)
17	E6	del AG	218	218delAG	–	Frameshift⇒stop codon 9aa away	Gu et al (1994)
18	E6	G→C	251	A84C		Ala→Cys	Mgone et al (1993)
19	E7	G→T	277	V93F		Val→Phe	Chen et al (1992)
20	E7	A→G	293	K98R	–	Lys→Arg	Kauppinen et al (1995)
21	E7	ins A	between C_{314} and T_{315}			Frameshift⇒stop codon 121aa away	Schreiber et al (1995)
22	E7	G→A	331	G111R	–	Gly→Arg	Gu et al (1993a)
23	E8	C→T	346	R116W	–	Arg→Trp	Gu et al (1993)
24	E8	G→A	347	R116Q		Arg→Gln	Mgone et al (1994)
25	E8	C→T	356	P119L		Pro→Leu	Lundin et al (1995)

No.	Intron/ Exon	Base change	Nucleotide position	Amino acid change	CRIM-status	Effect of mutation	Reference
26	I8	G→T	422+1			5' donor splice site mutation	Lundin et al (1995)
27	E9	C→T	445	R149X	–	Arg→Stop	Kauppinen et al (1995)
28	E9	G→A	446	R149Q	–	Arg→Gln	Delfau et al (1991)
29	E9	G→T	446	R149L	–	Arg→Leu	Gu et al (1994)
30	E9	C→T	465	Q155X	–	Gln→Stop	Scobie et al (1990)
31	E9	ins A	between A_{470} and G_{471}			Frameshift⇒stop codon 52aa away	Schreiber et al (1994)
32	I9	G→A	499-1			3' acceptor splice site mutation⇒ deletion of E10	Lundin et al (1994)
33	E10	C→T	499	R167W	–	Arg→Trp	Gu et al (1993) Llewellyn et al (1992)
34	E10	G→A	500	R167Q	–	Arg→Gln	Delfau et al (1990) Llewellyn et al (1992) Mgone et al (1992)
35	E10	C→T	517	R173W	–	Arg→Trp	Mgone et al (1994)
36	E10	G→A	518	R173Q	+	Arg→Gln	Delfau et al (1990)
37	E10	T→G	530	L177R		Leu→Arg	Mgone et al (1992)
38	E10	C→T	583	R195C	–	Arg→Cys	Kauppinen et al (1995)
39	E10	G→A	593	W198X	–	Trp→Stop	Lee and Anvret (1991)
40	E10	C→T	601	R201W		Arg→Trp	Lundin et al (1994)
41	E10	del G	604	604delG		Frameshift⇒stop codon 53aa away	Schreiber et al (1994)
42	E10	C→T	610	Q204X		Gln→Stop	Mgone et al (1994)
43	E10	G→T	612	612G→T	–	5' donor splice site mutation, last 9 bases of E10 deleted	Delfau et al (1991)
44	E11	del A	629	629delA		Frameshift⇒stop codon 44aa away	Lee et al (1995)
45	I11	C→G	652-3			3' acceptor splice site mutation⇒ skipping of E12	Puy et al (1996)
46	E12	G→A	667	E223K	–	Glu→Lys	Gu et al (1994)
47	E12	C→G	673	R225G	–	Arg→Gly	Kauppinen et al (1995)
48	E12	C→T	673	R225X	–	Arg→Stop	Kauppinen et al (1995)

No.	Intron/ Exon	Base change	Nucleotide position	Amino acid change	CRIM-status	Effect of mutation	Reference
49	E12	T→G	713	L238R	–	Leu→Arg	Kauppinen et al (1995)
50	E12	del CT	730	730delCT	–	Frameshift⇒stop codon 6aa away	Mgone et al (1993) Gu et al (1994)
51	E12	T→G	734	L245R	–	Leu→Arg	Delfau et al (1991)
52	E12	T→C	739	C247R		Cys→Arg	Mgone et al (1993)
53	E12	G→T	740	C247F	–	Cys→Phe	Chen et al (1992)
54	E12	ins TTCGC TGC	742	742ins8	–	Frameshift⇒stop codon 10aa away	Gu et al (1994)
55	E12	G→A	748	E250K	–	Glu→Lys	Gu et al (1994)
56	E12	G→C	748	E250A		Glu→Ala	Lundin et al (1995)
57	E12	G→A	754	A252T		Ala→Thr	Mgone et al (1993)
58	E12	C→T	755	A252V		Ala→Val	Mgone et al (1993)
59	E12	C→A	766	H256N		His→Asn	Mgone et al (1992)
60	E12	G→A	771	771G→A	+	5' donor splice site mutation⇒ deletion of E12	Grandchamp et al (1989a)
61	E12	G→C	771	771G→C		5' donor splice site mutation⇒ deletion of E12	Daimon et al (1993)
62	E12	ins T	771	771insT		Frameshift⇒stop codon 33aa away	Ong et al (1996)
63	E13	C→T	806	T269I		Thr→Ile	Mgone et al (1994)
64	E13	G→A	820	G274R		Gly→Arg	Mgone et al (1994)
65	E14	G→A	838	G280R	–	Gly→Arg	Kauppinen et al (1995)
66	E14	G→A	848	W283X		Trp→Stop	Mgone et al (1994)
67	E14	G→A	849	W283X		Trp→Stop	Schreiber at al (1995)
68	E14	C→T	886	Q296X	–	Gln→Stop	Kauppinen et al (1995)
69	E14	del T	900	900delT	–	Frameshift⇒stop codon 15aa away	Delfau et al (1991)
70	I14	G→A	912+1	912+1G→A	–	5' donor splice site mutation⇒ deletion of E14	Gu et al (1993a)
71	E15	ins C	913	913insC		Frameshift⇒stop codon 1aa away	Puy et al (1996)

No.	Intron/ Exon	Base change	Nucleotide position	Amino acid change	CRIM-status	Effect of mutation	Reference
72	E15	del A	1073	1073delA	–	Frameshift⇒ protein contains 449aa instead of 361aa	Kauppinen et al (1995)
73	E15	ins C	between -22 and -21 from ter-mination codon			Frameshift⇒stop codon 4aa away	Daimon et al (1994)

Figure 4.1: Percentage of occurrence of different types of mutations in the *HMB-S* gene.

The pie chart shows the frequency of the different types of mutations encountered in the *HMB-S* gene. Missense mutations form the majority of mutations reported so far in the *HMB-S* gene.

Figure 4.1: Percentage of occurrence of different types of mutations in the *HMB-S* gene

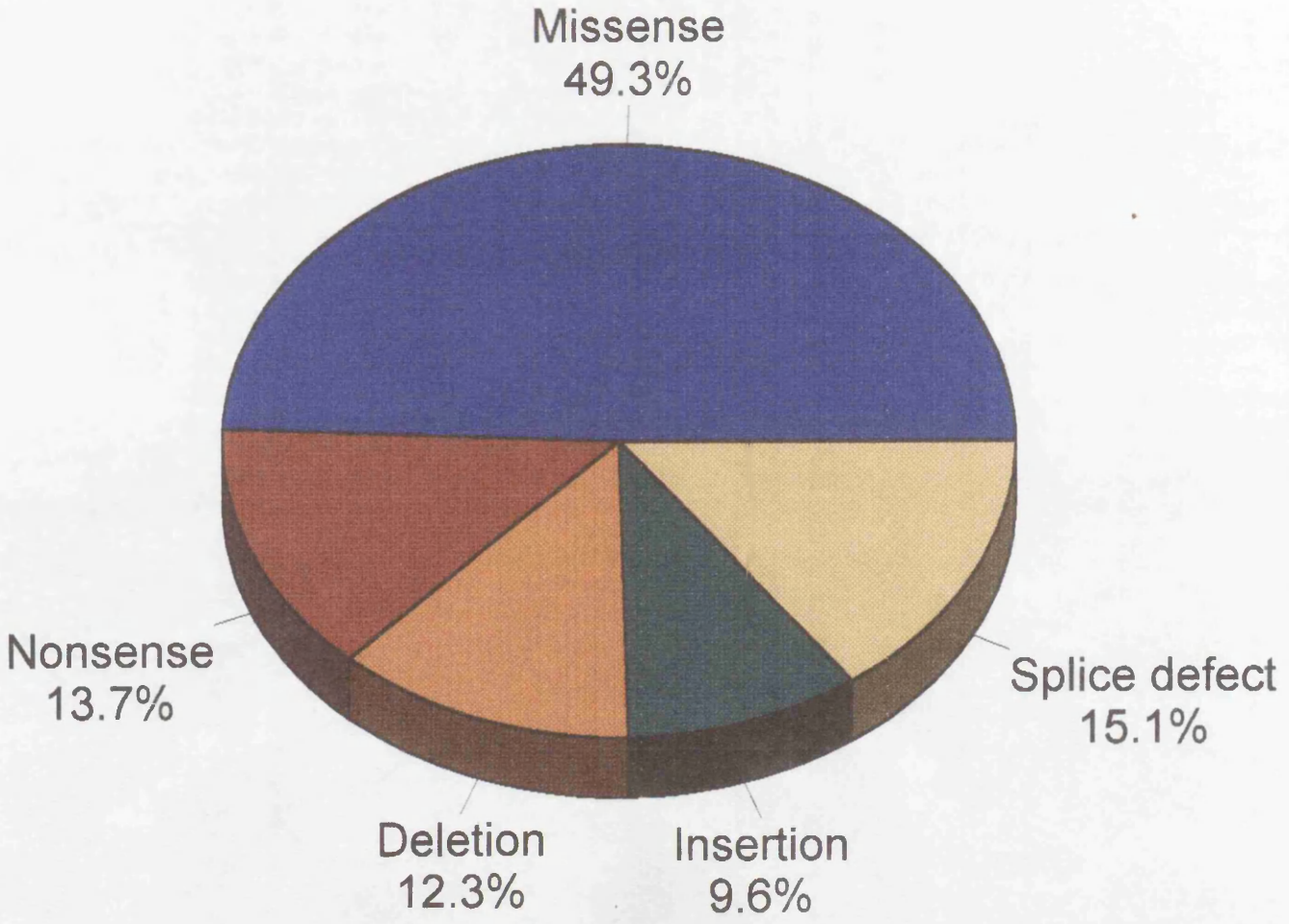
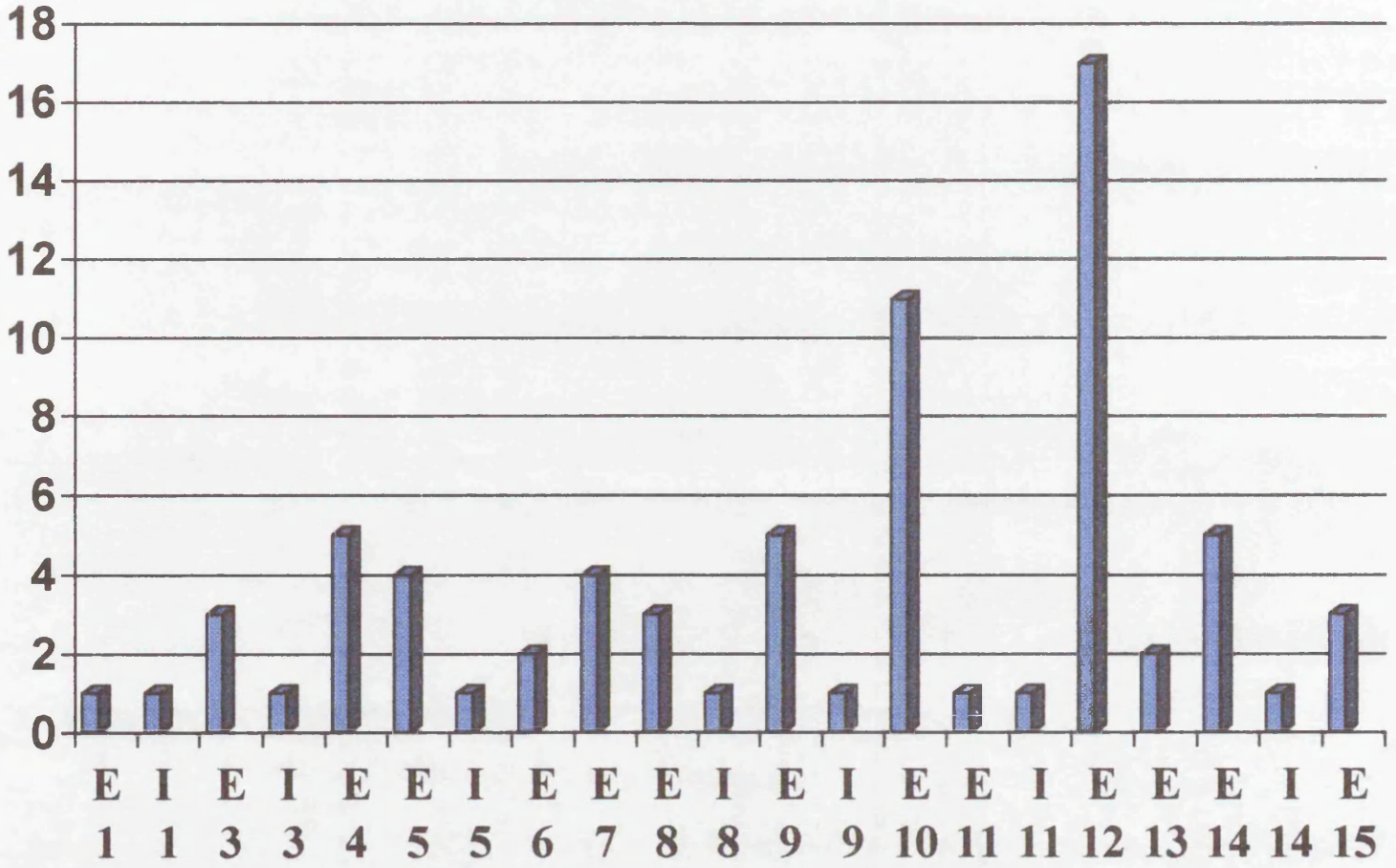


Figure 4.2: Distribution of mutations in the *HMB-S* gene.

The chart shows the distribution of reported mutations in the *HMB-S* gene. All the exons of the housekeeping gene have at least one mutation. Exon 12 has the highest number of mutations reported in the literature, followed by exon 10.

The X-axis represents the location of the mutation within the gene (E = exon; I = intron) and the Y-axis shows the number of mutations identified in each region of the gene.

Figure 4.2: Distribution of mutations in the *HMB-S* gene



R26C (Kauppinen et al, 1995), R26H (Llewellyn et al, 1993), R116W (Gu et al, 1993), R173Q (Delfau et al, 1990) and L177R (Mgone et al, 1992). The heterogeneity of AIP is further emphasised in the South African group of patients where five different mutations were detected in seven apparently unrelated patients. All the mutations detected in this study are potentially pathogenic since (i) a highly conserved residue is affected; (ii) truncated protein products would be formed; or (iii) an exon has been skipped.

AIP has a very high incidence in Sweden with 1/1000 of the population affected (Waldenström, 1957). A number of reported mutations could be shown to originate from Sweden. The G→A transition in exon 10 (W198X) (Lee and Anvret, 1991) identified in a family from Swedish Lapland, was also found in 15 of 33 Swedish families with AIP. 12 of these families with the W198X mutation were related to the index family. Another mutation, C→T in exon 8 (R116W), was found to occur at a high frequency in Dutch patients with the CRIM-negative phenotype (15 of 49 patients or 30%) but only in 1 of 33 French and 1 of 28 Swedish families (Lee et al, 1990; Gu et al, 1993). This mutation was first described in a Swedish patient (Lee, 1991) and was detected in three unrelated South African patients in the present study, whose ancestry revealed that they may have a mixture of Malay, European and Khoi blood, the latter being the original inhabitants of the Cape at the time of European settlement. If the European component of the mixture is Dutch, this might explain why this mutation is present in all three apparently unrelated patients. The L177R mutation has been independently reported by Mgone et al (1992) in a Scottish patient. This mutation was detected again in six Scottish patients (five members of a family and an isolated case). Their ancestry is not known but it could prove to be a common Scottish mutation. R173Q was first identified in French and Dutch patients (Delfau et al, 1990), and it was detected in one of the South African patients, who may have Dutch ancestors. Therefore, a patient's geographical origin may lend assistance in the identification of the mutation responsible for AIP in that family.

The novel R22C mutation was initially identified in individuals III:8 and III:9 of Pedigree 9532 (Appendix 1). Subsequent analysis of individuals II:2, III:3, III:5, III:7 and III:11 by *Hinf* I digestion revealed that II:2, III:5 and III:11 are carriers of the defective

gene. This mutation in exon 3 of the *HMB-S* gene affects the highly conserved Arg22 residue (Arg7 in *E. coli*). Because this residue is highly conserved in ten species (Louie et al, 1996), it is a good indication that it has an important role in the proper functioning of the enzyme. Arginine, a basic residue, is substituted by cysteine, a neutral, non-polar amino acid which, consequently, is predicted to destabilise the protein. However, the precise role of this residue is not known at the present time.

R26C, previously reported by Kauppinen et al (1995), was initially identified in individuals III:1 and III:3 but not in III:2 of Pedigree 13168 (Appendix 1). By employing an *Aci* I digestion for the other members of this family, namely, II:1 to II:6, II:8, III:5 to III:8, IV:1 and IV:2, it was shown that II:8, IV:1 and IV:2 were also carrying the defective gene. Arg26 (Arg11 in *E. coli*) is a conserved residue in twelve species. It functions to form a salt-bridge with the acidic side chain of the second pyrrole ring of the cofactor. Any substitution of this residue will result in a defective enzyme being produced. This mutation results in the basic arginine residue to be replaced by the polar uncharged cysteine, which is predicted to destabilise the protein. In this family, individual II:6 was found to have an erythrocyte HMB-S activity of 20.6nmol/ml RBC/hr (normal range 22-46nmol/ml RBC/hr). An analysis of his entire coding sequence by direct sequencing did not reveal any sequence alterations. He may have a much higher proportion of alternatively spliced exons 3 and 12 in his coding sequence, which has been found, in this study, to be a normal occurrence in the general population. It would be useful to express his mis-spliced transcripts to determine if the degree of splicing out would have an effect on his erythrocyte HMB-S activity.

The R26H mutation affects the same residue as the R26C mutation, and is expected to result in a defective protein being produced.

R116W in exon 8 was identified in three unrelated South African patients and two members of a family from the Republic of Ireland. This mutation was easily confirmed by an *Aci* I digestion. Arg116 (Arg101 in *E. coli*) is conserved in twelve species and serves to interact with the interdomain hinge region by forming a salt-bridge with Glu231 and a hydrogen bond with the backbone carbonyl group of residue 198. Its replacement with tryptophan would probably result in the loss of these stabilising interactions, leading to the

formation of a non-functional protein. Both the South African and Irish families may have Dutch ancestors but no information is currently available for confirmation.

The R173Q mutation in exon 10, identified in a South African patient, affects the highly conserved Arg173 (Arg155 in *E. coli*) residue. This mutation is expected to result in a defective protein since the substitution of the basic residue, arginine, with the polar neutral residue, glutamine, is anticipated to prevent the formation of salt-bridges with the propionate group of the first pyrrole ring of the cofactor and the carboxylate side group of the substrate molecule, thus leading to the inhibition of substrate elongation.

L177R in exon 10, previously reported by Mgone et al (1992) in the Scottish population, was identified again in two Scottish families in this study, one of which consisted of an isolated case. The other family, Pedigree 13543 (Appendix 1), had five affected members, II:3, III:2, III:3, III:5 and III:6, the latter of which was initially thought to be an isolated case. Subsequent analysis of IV:1, IV:3 and IV:4 revealed that the mutant allele was absent in these individuals. II:1 was not analysed in this study but was diagnosed with AIP based on her biochemical data. Leu177 (Leu159 in *E. coli*) is conserved in eleven species and forms part of the hydrophobic core of the molecule. The introduction of a charged group in this region of the protein by arginine would be structurally deleterious. This mutation could prove to be a common Scottish mutation.

There are fifty-eight invariant residues in the HMB-S protein (Brownlie et al, 1994); twenty-eight of which have been reported to be mutated (48.3%). This comprises 38.4% of all reported mutations. The *E. coli* model has been used to explain the effects of the mutations on the structure and function of the enzyme because it has a high degree of homology with the human form (43%). The location of all seventy-three mutations listed in Table 4.1 is shown in Figure 4.3.

How missense mutations affect the structure, stability and function of a protein has been studied by Pakula and Sauer (1989) and Alber (1989). The resulting phenotypic effects of a mutation will depend on the nature of the amino acid substitution and its location within the protein. Mutations which affect the core residues of the protein structure and those residues involved in maintaining the stability of the protein through the

Figure 4.3: The location of mutations within the hydroxymethylbilane synthase (HMB-S) protein.

The location of the various mutations detected in the HMB-S protein is shown. The numbers refer to the number of the mutation in Table 4.1. The dipyrromethane cofactor is represented in red and is situated in a deep cleft between domains 1 and 2 (shown in yellow and green respectively). Domain 3 is shown in blue.

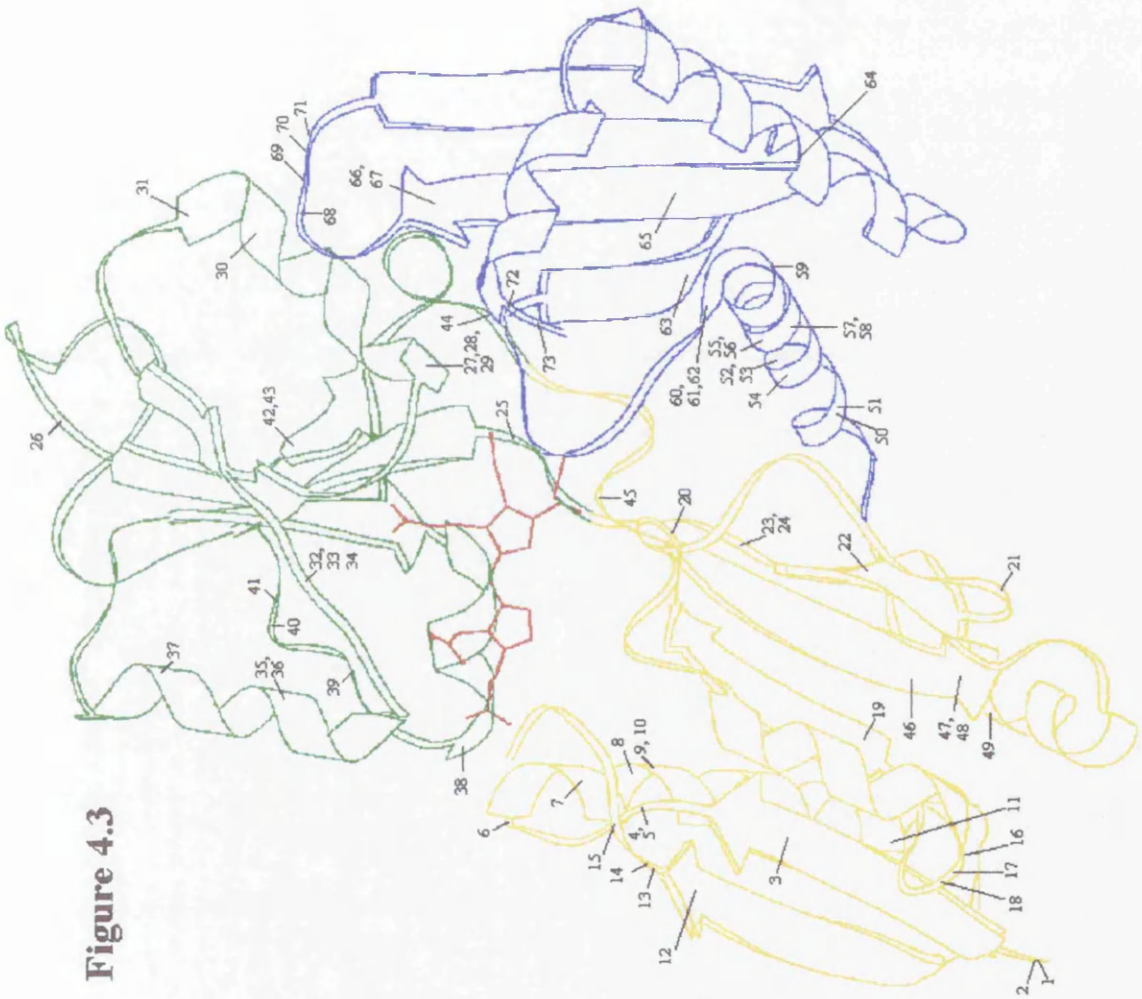

1
2
3
4, 5
6
7
8
9, 10
11
12
13
14
15
16
17
18
19
20
21
22
23, 24
25
26
27,28, 29
30
31
32, 33 34
35, 36
37
38
39
40
41
42,43
44
45
46
47, 48
49
50
51
52, 53
54
55, 56
57, 58
59
60, 61,62
63
64
65
66, 67
68
69
70
71
72
73

Figure 4.3

formation of hydrogen, disulphide and electrostatic bonds will most probably be responsible in the causation of the disease.

Conserved amino acids are usually clustered in important regions of the molecule, namely, the hydrophobic core, sites which are conformationally important and at the active site. Many of the substitutions involve residues situated close to the active site, and this may prove detrimental to the enzyme.

Residues Arg149, Arg167 and Arg173 (Arg131, Arg149 and Arg155 respectively in the *E. coli* model), all of which are located in the vicinity of the active site, are involved in providing stabilising interactions with the acetate and propionate side groups of the cofactor. Arg149 interacts with the side chains of ring C1 of the cofactor and its substitution, by site-directed mutagenesis (Jordan and Woodcock, 1991; Lander et al, 1991) with leucine or histidine, prevents the assembly of the cofactor and the correct folding of the enzyme. This results in the production of an inactive and unstable apoenzyme. Arg173 serves to interact with the side chains of rings C1 and C2 of the cofactor, and its replacement with tryptophan or glutamine (numbers 35 and 36 respectively, Figure 4.3) results in the prevention of substrate binding, the final stages of pyrrole chain extension and the final release of the tetrapyrrole product. The invariant Arg116 is situated on one of the two short strands linking domains 1 and 2 (numbers 23 and 24, Figure 4.3), with the charged side chain forming an ion pair with Glu250. Site-directed mutagenesis at this position does not prove to be deleterious. Arg26, like Arg173, forms interactions with the cofactor at the substrate-binding site, and its replacement with cysteine or histidine (numbers 4 and 5 respectively, Figure 4.3) inhibits the conversion of the enzyme to the enzyme-substrate complex, and eventually the formation of the product.

The mutations which occur at the conserved residues of the hydrophobic core are expected to be detrimental to the protein structure, because they either introduce a charged group to the apolar core of the molecule, or result in steric hindrance with the substitution of another amino acid e.g. L177R and A31T (numbers 37 and 7 respectively, Figure 4.3).

Eleven mutations (15%) have been reported in exon 10 (Table 4.1), indicating its important role in maintaining the proper function of the enzyme. The three highly conserved residues in this exon which have undergone single base substitutions are

Arg167, Arg173 and Leu177. The substitution of these residues with other amino acids is predicted to lead to the formation of a non-functional, unstable protein.

In this study, four of the five residues altered are conserved (Arg26, Arg116, Arg173 and Leu177). The fifth residue affected, Arg22, has not been studied in detail but, from the table of alignment of amino acid sequences (Louie et al, 1996) of thirteen species, it is a highly conserved residue, being maintained in ten of the species listed. The highly conserved nature of these mutations indicates that they have an important role in the protein molecule to ensure its normal function. Any mutation affecting these conserved residues will consequently lead to impaired protein function.

Therefore, the missense mutations identified in the *HMB-S* gene have been shown to impair the normal function of the enzyme because they (i) destabilise the protein through loss of the stabilising interactions within the molecule; (ii) prevent substrate binding and chain elongation; (iii) introduce a charged group into the apolar core of the enzyme; or (iv) cause steric hindrance with the substitution of the new amino acid.

Nonsense and frameshift mutations have also been documented in the *HMB-S* gene (Table 4.1). Ten nonsense mutations (14%) were found throughout the gene and led to the formation of truncated proteins. Large insertions and deletions are not a common feature in the molecular pathology of AIP. The largest number of bases reported to be inserted was eight in exon 12 (Gu et al, 1994) which resulted in the formation of a premature termination of translation ten amino acids away. The novel insertion mutation identified in this study, 771insT in exon 12, was detected in a South African patient. The consequence of this mutation was the alteration of the thirty-two amino acids following this insertion prior to the formation of a termination codon. The largest number of bases reported to be deleted in the *HMB-S* gene was two, 168delGT (Schreiber et al, 1995) in exon 5, 218delAG (Gu et al, 1994) in exon 6 and 730delCT (Mgone et al, 1993; Gu et al, 1994) in exon 12. Both insertions and deletions result in the formation of truncated proteins, which will impair the normal function of the protein.

The majority of the splice site mutations identified in the *HMB-S* gene occur at the 5' donor splice site, all of which led to the skipping of the preceding exon. The only exception to this rule was the identification of a 9bp deletion at the 5' splice site of exon 10 (Delfau et

al, 1991). The relative abundance of the abnormally spliced mRNA was found to be the same as the normal transcript, indicating that the abnormally spliced mRNA was of equal stability as the normal mRNA. Only two 3' acceptor splice site mutations have been reported in the *HMB-S* gene, i.e. in intron 9 (Lundin et al, 1994) and intron 11 (Puy et al, 1996), leading to the skipping of exons 10 and 12 respectively. A similar pattern was also observed for the splice site mutation detected in this study, where the 3' splice site mutation detected at the last position of intron 7 resulted in the skipping of exon 8. Both exons 10 and 12 have recorded the highest number of mutations. Many highly conserved residues are present in both exons e.g. Arg167, Arg173 and Leu177 in exon 10, and Arg225 and Glu250 in exon 12. These residues either provide stabilising interactions within the enzyme molecule or enable substrate binding and chain elongation. Therefore, the absence of either of these exons will result in a defective protein product and, consequently, lead to an increase in susceptibility to AIP if exposed to certain triggering factors. Likewise, the exclusion of exon 8 from the protein is predicted to result in its destabilisation since the three conserved residues, involved in maintaining its stability, Arg116, Asp121 and Leu134, are absent. A calculation of the consensus value (CV) of the 3' acceptor splice site of the mutant intron 7 was 77.59%, which was a reduction of 16.6% from that of the normal sequence (94.19%). These values provide evidence that the sequence alteration is most probably responsible for the splicing error.

In the two South African patients where no sequence alterations were identified in the coding sequence of the *HMB-S* gene, it would be extremely useful to obtain their DNA samples to enable the analysis of the promoter regions.

4.1.2 ALTERNATIVE SPLICING

The phenomenon of alternative splicing of exons 1 and 2 of the *HMB-S* gene in erythroid and non-erythroid tissues is well-documented (Chretien et al, 1988).

In this study, several patients were initially found to exhibit exon 12 skipping in their *HMB-S* gene. Several attempts were made to identify the cause of exon 12 skipping by sequencing the preceding and following introns of this gene but to no avail. On screening forty-four healthy individuals, exon 12 was found to undergo varying degrees of alternative

splicing (Figure 3.16). The other exons were similarly screened for evidence of alternative splicing, and exon 3 was found to be alternatively spliced to varying degrees as well (Figure 3.15). Four different mRNA transcripts are likely to exist in the lymphocyte population (Figure 3.17): those which (i) are intact; (ii) have exon 3 missing; (iii) have exon 12 missing; and (iv) have exons 3 and 12 missing.

Two approaches were undertaken in an attempt to identify a reason for this event. The first approach was to study the frequency of occurrence of each base at each position of the 5' and 3' splice junctions of the *HMB-S* gene and to compare with that of four hundred vertebrate genes analysed by Padgett et al (1986). The 5' and 3' splice site consensus sequences shown in Figure 4.4 were determined after a survey of four hundred vertebrate genes by Padgett et al (1986).

At some positions, the frequency of occurrence of a particular base in the *HMB-S* gene, is very close to that reported by Padgett et al (1986). However, there are significant differences in the frequency at other positions. For example, at position +4 of the 5' donor splice site, the frequency of adenine occurring was reported to be 74% but the frequency was only 50% in the *HMB-S* gene. This creates some difficulty in interpreting how particular sequences at the splice sites might affect the proper splicing of the *HMB-S* mRNA.

Therefore, another approach taken was to calculate the consensus values (CV) at the 5' and 3' splice sites (Tables 4.4 and 4.5 respectively) according to the method of Shapiro and Senapathy (1987). These values were estimated on the basis of the conservation of bases at the eight positions most likely to contribute to 5' splice recognition determinants (positions -2 to +6) and the fourteen positions (positions -14 to +1 but excluding position -4) at the 3' end and were used as relative indicators of the efficacy of splice sites.

The findings that exons 3 and 12 underwent some degree of splicing out led to the investigation of the CVs of their flanking introns. It can be seen, however, that the values do not justify their splicing out from the *HMB-S* pre-mRNA. There are lower CVs for other introns which should, theoretically, make these splice sites relatively inefficient, but this was not observed experimentally. The nucleotides at two and three positions deviate

5' (DONOR) SPLICE SITE CONSENSUS SEQUENCE

-4	-3	-2	-1		+1	+2	+3	+4	+5	+6
C_{29}/A_{34}	C_{38}/A_{35}	A_{62}	G_{77}	:	G_{100}	T_{100}	A_{60}	A_{74}	G_{84}	T_{50}
		EXON			INTRON					

In the *HMB-S* gene, the following was observed (Table 4.2):

-4	-3	-2	-1		+1	+2	+3	+4	+5	+6
C_{29}/A_{21}	C_{50}/A_{21}	A_{64}	G_{64}	:	G_{100}	T_{100}	A_{79}	A_{50}	G_{64}	T_{43}
		EXON			INTRON					

3' (ACCEPTOR) SPLICE SITE CONSENSUS SEQUENCE

-14	-13	-12	-11	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1		+1
Y_{78}	Y_{81}	Y_{83}	Y_{89}	Y_{85}	Y_{82}	Y_{81}	Y_{86}	Y_{91}	Y_{87}	N	C_{78}	A_{100}	G_{100}	:	G_{55}
												INTRON			EXON

Correspondingly, in the *HMB-S* gene (Table 4.3):

-14	-13	-12	-11	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1		+1
Y_{50}	Y_{79}	Y_{86}	Y_{86}	Y_{79}	Y_{93}	Y_{93}	Y_{93}	Y_{86}	Y_{71}	N	C_{86}	A_{93}	G_{93}	:	G_{29}
												INTRON			EXON

Figure 4.4: The percentage of occurrence of the consensus sequences at the 5' and 3' splice junctions. Y = pyrimidine, N = any base, subscript numerals refer to percentage frequency of occurrence. Adapted from Krawczak et al (1992).

Table 4.2: 5' splice site sequences of the *HMB-S* gene.

Consensus sequence of 400 genes	Reported	-4 C_{29}/A_{34}	-3 C_{38}/A_{35}	-2 A_{62}	-1 G_{77}	+1 G_{100}	+2 T_{100}	+3 A_{60}	+4 A_{74}	+5 G_{84}	+6 T_{50}
HMB-S	I1	G	G	C	G	G	T	G	A	G	T
HMB-S	I2	G	A	A	G	G	T	A	C	T	G
HMB-S	I3	C	C	A	G	G	T	**G**	**G**	G	T
HMB-S	I4	A	T	C	A	G	T	G	A	G	T
HMB-S	I5	T	A	A	G	G	T	A	A	C	A
HMB-S	I6	A	T	G	A	G	T	A	A	G	T
HMB-S	I7	G	C	A	A	G	T	A	A	G	A
HMB-S	I8	A	G	A	G	G	T	A	A	G	T
HMB-S	I9	**T**	A	**T**	**T**	G	T	A	**T**	**C**	**C**
HMB-S	I10	G	C	A	G	G	T	A	G	G	G
HMB-S	I11	C	C	A	G	G	T	A	C	A	C
HMB-S	I12	C	C	**T**	G	G	T	A	**G**	G	**G**
HMB-S	I13	G	C	A	A	G	T	A	A	G	T
HMB-S	I14	C	C	A	G	G	T	A	C	C	A
HMB-S	%	C=29 A=21	C=50 A=21	A=64	G=64	G=100	T=100	A=79	A=50	G=64	T=43

The 5' donor splice site sequences of the *HMB-S* gene are compared with the consensus sequence reported by Padgett et al (1986). The consensus sequences of introns 3 and 12 (heavily shaded) are the introns of interest in this study. The nucleotide variants are shown in bold type in the unshaded boxes. The nucleotide variants from the consensus sequence of intron 9 are also shown in bold type (see Discussion).

Table 4.3: 3' splice site sequences of the *HMB-S* gene.

		-14	-13	-12	-11	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	+1
Consensus sequence of 400 genes	Reported	Y_{78}	Y_{81}	Y_{83}	Y_{89}	Y_{85}	Y_{82}	Y_{81}	Y_{86}	Y_{91}	Y_{87}	N	C_{78}	A_{100}	G_{100}	G_{55}
HMB-S	I1	G	T	G	G	G	C	T	T	C	A	G	G	G	C	T
HMB-S	I2	**G**	**A**	C	T	T	T	T	T	C	T	G	C	A	G	G
HMB-S	I3	A	T	C	T	C	T	A	C	C	C	G	C	A	G	C
HMB-S	I4	G	A	C	T	C	T	C	T	C	C	T	C	A	G	T
HMB-S	I5	T	T	T	C	T	G	T	C	C	G	G	C	A	G	A
HMB-S	I6	T	C	C	C	A	T	C	T	C	T	A	T	A	G	A
HMB-S	I7	C	C	C	T	C	C	T	C	C	C	C	C	A	G	G
HMB-S	I8	T	T	C	T	C	T	C	T	G	G	G	C	A	G	T
HMB-S	I9	**G**	C	**A**	**A**	C	T	C	T	C	C	A	C	A	G	**C**
HMB-S	I10	T	T	C	C	G	C	C	T	C	C	A	C	A	G	A
HMB-S	I11	T	**G**	T	T	T	T	T	C	C	**A**	T	C	A	G	G
HMB-S	I12	A	C	T	T	C	T	T	G	T	T	A	C	A	G	G
HMB-S	I13	G	T	T	T	T	C	T	T	C	C	C	C	A	G	C
HMB-S	I14	C	T	T	C	C	C	T	C	A	T	A	C	A	G	C
HMB-S	%	Y=50	Y=79	Y=86	Y=86	Y=79	Y=93	Y=93	Y=93	Y=86	Y=71	N	C=86	A=93	G=93	G=29

The 3' acceptor splice site sequences of the *HMB-S* gene are compared with the consensus sequence reported by Padgett et al (1986). The consensus sequences of introns 2 and 11 (heavily shaded) are the introns of interest in this study. The nucleotide variants are shown in bold type in the unshaded boxes. The nucleotide variants from the consensus sequence of intron 9 are also shown in bold type (see Discussion).

Table 4.4: Calculation of the consensus value (CV) of the 5' donor splice site of the *HMB-S* gene.

5' DONOR SITE	CONSENSUS VALUE
Intron 1	79.58
Intron 2	79.58
Intron 3	83.55
Intron 4	72.97
Intron 5	87.71
Intron 6	82.61
Intron 7	89.22
Intron 8	100
Intron 9	58.03
Intron 10	90.36
Intron 11	78.26
Intron 12	80.91
Intron 13	93.38
Intron 14	82.23

The CVs are calculated according to the method of Shapiro and Senapathy (1987). The shaded boxes are the introns of interest in this study.

Table 4.5: Calculation of the consensus value (CV) of the 3' acceptor splice site of the *HMB-S* gene.

3' ACCEPTOR SITE	CONSENSUS VALUE
Intron 1	40.24
Intron 2	95.85
Intron 3	92.30
Intron 4	95.18
Intron 5	86.13
Intron 6	74.51
Intron 7	94.19
Intron 8	90.99
Intron 9	91.71
Intron 10	90.28
Intron 11	91.71
Intron 12	84.83
Intron 13	95.02
Intron 14	84.12

The shaded boxes are the introns of interest in this study.

from the 5' consensus sequences in introns 3 and 12 respectively (Table 4.2), and at two positions in the 3' consensus sequences of introns 2 and 11 (Table 4.3). However, it is difficult to prove that these changes contribute to the splicing out of exons 3 and 12. For example, the nucleotides at six positions of the 5' splice site of intron 9 deviate from the consensus sequence (Table 4.2), thus contributing to its low CV (58.03) (Table 4.4) but do not cause exon 9 to be spliced out under normal circumstances. Similarly, four variant nucleotides were observed at the 3' splice site of intron 9 and, although its CV is the same as that for intron 11 (91.71) (Table 4.5), it does not result in the removal of exon 10 in the mature transcript. Three possible explanations are proposed: (i) other sequence signals may exist in addition to the splice junction sequence signal; (ii) splice sites with infrequently occurring bases may be recognised by different components of the splicing machinery; (iii) alternative splicing is a consequence of errors made by the splicing system.

Alternative splicing may have important implications; it may have preceded or was concomitant with the appearance of constitutive splicing as a means of facilitating rapid gene evolution. This may still hold true today: genes that remain in flux would retain alternative splicing but those whose products have been optimised for a given cell environment or function would have made all their exons constitutive or eliminated their introns altogether. Alternative splicing would be especially advantageous for genes that are evolving or require flexibility.

The alternatively spliced *HMB-S* transcripts identified will result in new isoforms of the HMB-S protein. The new isoforms will be shorter by 18 and 40 amino acids due to the absence of exons 3 and 12 respectively but the translational reading frame of both isoforms is not disrupted. In Figure 3.15a, the shorter allele, missing exon 3, shows variable intensity among the samples analysed. A similar observation was made for that missing exon 12 (Figure 3.16a). However, in Figure 3.15a, it is obvious that, in some individuals, the proportion of truncated mRNAs is equal to that of full length mRNA. In these cases, if the truncated enzyme has reduced or no activity, there will be a significant reduction in the amount of active enzyme available to the cells. Two conserved residues are located in exon 3, Arg26 and Ser28 (Arg11 and Ser13 respectively in the *E. coli* model). Arg26 has been shown to form a salt-bridge with the side chain of the second pyrrole ring of the cofactor

(Llewellyn et al, 1993). Any disturbance of this residue will disrupt substrate binding and chain elongation. Seventeen mutations (23%) have been recorded in exon 12 (Table 4.1), especially in the region of the $\alpha 1_3$-helix, indicating its importance in the proper functioning of this enzyme. This exon also crosses between domains 1 and 3. The polar groups in this helix are in contact with the two domains, therefore, any mutation which changes the charge of these groups will severely disrupt the conformation of the enzyme and have detrimental effects on the enzyme activity. The residues Ala252, Leu254 and His256 are located within the $\alpha 1_3$-helix. These residues, when altered, cause steric disruption and destabilise the protein. It is postulated that an inverse relationship exists between total HMB-S activity and the number of mRNA molecules with exon skipping. If alternative splicing also occurs in erythrocytes, it might explain the range of HMB-S activities found in the general population.

No conclusions can be made at this point regarding the effects these mis-spliced transcripts have on the activity of the enzyme. It would be useful to express these transcripts in the *in vitro* expression system to study if a correlation exists between the degree of splicing out of exons 3 and 12 and the enzyme activity.

The splicing out of exons 3 and 12 from the *HMB-S* mRNA identified in this study seems to be a true occurrence in the general population. 19 cycles of amplification of the mRNA transcripts (Figure 3.16c), which represents the log-phase of PCR, resulted in the same ratio of amplified products as that generated after 35 cycles (Figure 3.16d). This indicates that the ratio of amplified transcripts observed is a true reflection of the concentration of mRNA species in the lymphocyte population.

An interesting mutation recently reported by Llewellyn et al (1996) which causes exon skipping was that of a C to G transversion at position -22 from the 5' splice site of intron 3, causing the skipping of exon 3. Such rare intraexonic mutations which affect 5' splice site selection have been documented in mitochondrial acetoacetyl-coenzyme A thiolase deficiency (Fukao et al, 1994), in an *in vitro* mutant of the human hypoxanthine-guanine phosphoribosyl transferase gene (Steingrimsdottir et al, 1992) and in bovine thyroglobulin deficiency (Ricketts et al, 1987). These mutations are postulated to affect splice site

selection by altering the secondary structure of pre-mRNA and, consequently, interferes with the assembly of small nucleoproteins into the spliceosome.

Splicing involves the removal of the intron segments from pre-mRNA and the accurate end-to-end joining of exons in the proper arrangement to code for a specific polypeptide chain.

RNA polymerase II (Pol II) transcribes protein-encoding genes. The processing of Pol II transcripts depends on the sequences of three critical regions within the introns- the 5' and 3' ends and the branch-point sequence close to the 3' end. Another determining factor is the presence of all the components of a complex particle, the spliceosome, which comprises a number of ribonucleoprotein subunits, each containing a small RNA molecule.

Studies of a large number of splice junctions revealed that the 5' and 3' ends of almost all introns begin with the sequence GT and terminate with AG respectively. The branch-point nearly always involves a specific adenine-containing residue.

Mammalian cells, unlike those of bacteria, contain numerous small RNA species which are neither transfer RNA (tRNA) nor ribosomal RNA (rRNA). Six different small nuclear RNAs (snRNAs) called U1, U2, U3, U4, U5 and U6, U6 being the smallest (100bp) and U3 the largest (215bp), are present in the cells of all vertebrate species. The snRNAs form stable complexes with certain abundant nuclear proteins called small nuclear ribonucleoproteins (snRNPs). In splicing, the role of snRNAs is to recognise the 5' and 3' splice sites, and the branch-point of each intron in the intron-containing transcript (Green, 1986; Sharp, 1987). U1 RNA is the most abundant of all the snRNAs in vertebrate cells. Its 5' end contains a sequence that is precisely complementary to the nine-nucleotide consensus for 5' splice junctions, while the U2 snRNA binds to the BPS (Watson et al, 1987), both events being necessary for the proper splicing of pre-mRNA.

Signals important for splicing of mammalian pre-mRNAs reside at the splice junctions themselves. The nucleotide sequences at the 5' junction of introns almost invariably begin with GT and end with AG at the 3' junction. The nine-nucleotide consensus sequence for 5' junctions extends three residues upstream and six residues downstream from the splice point (Watson et al, 1987). The average 5' junction matches this consensus in at least seven positions; very rarely will there be only a five-residue match. The 3' splice junction

consensus sequence differs from the 5' consensus in that it comprises a pyrimidine-rich region of variable length, usually greater than ten nucleotides, followed by a short consensus sequence extending only four residues upstream and one residue downstream from the splice point. A schematic diagram is shown below:

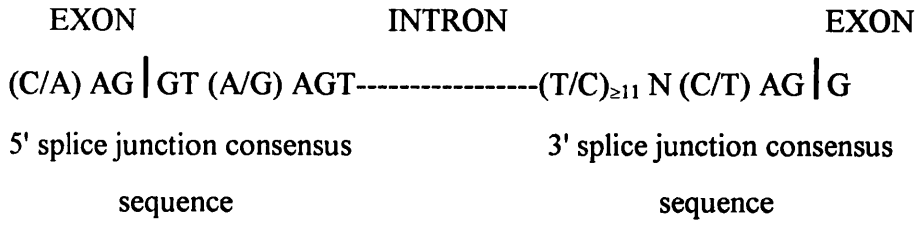


Figure 4.5: Consensus sequences at the 5' and 3' splice junctions.

Splice junction sequences from genes representing the entire range of eukaryotes (yeast to human) involve the same 5' and 3' consensus sequences, thus reflecting the highly conserved nature of the splicing elements which recognise these signals. In addition, the conserved recognition element known as the branch-point sequence (BPS) is located 18 - 40 base pairs upstream of the 3' splice site. It is important in the formation of a branch with the 5' terminus of the intron in the splicing process but seems to have a weak consensus sequence.

The conserved sequences at the splice junctions are utilised in such a way that introns are precisely recognised by the splicing elements and accurately spliced out by two transesterification reactions. The first step, which involves cleavage at the 5' splice site, requires prior U1 snRNA binding and ATP hydrolysis. This is accompanied by the subsequent formation of an unusual 2' - 5' phosphodiester bond between the 5' terminal G residue of the intron and the 2'-hydroxyl of the adenylate residue at the branch-point, about thirty nucleotides upstream from the 3' end of the intron. This results in a loop structure known as a lariat intermediate. The second step involves cleavage at the 3' splice site and subsequent joining of the two exons by another phosphodiester bond, resulting in the removal of the intron in a lariat conformation. The RNA fragments that are removed during splicing remain in the nucleus to be degraded into ribonucleotides by enzymes. These

ribonucleotides can be used again in the synthesis of more RNA after being converted back into the triphosphate form (Figure 4.6).

Thus, it can be seen that the efficiency of mammalian pre-mRNA splicing is governed by four sequence elements within the introns, viz. the invariant GT dinucleotide at the 5' splice junction, the highly conserved AG dinucleotide at the 3' end, the weakly conserved BPS and a pyrimidine-rich region known as the polypyrimidine [poly(Y)] tract (Padgett et al, 1986; Sharp, 1987). However, no sequence variations at these four sequence elements which could be responsible for the splicing out of exons 3 and 12 in the *HMB-S* gene were observed.

In the processing of RNA in higher eukaryotes, the amount of discarded RNA ranges from 50 to nearly 90% of the primary transcripts (Freifelder, 1987). Introns are excised one by one and the two exons are spliced together before the next intron is excised. Thus, it is expected that a large number of different nuclear RNA molecules would be present at any one instant.

If the excision of introns were to take place in a specific order, then only particular RNA intermediates should result. However, if intron excision occurs randomly, then all intermediates would be expected. For example, in the chick ovomucoid gene, which has seven introns (A to G), intron excision was found to take place in a preferred order. Random excision would result in more than 300 intermediates but only a few are observed. Intron E or F is usually lost in intermediates lacking one intron, and both E and F are lost in intermediates lacking two introns. Introns G and D are preferentially removed next, followed by B, and then A and C. Such preferential excision could be due to intramolecular base pairing which determines the order of removal. Once excision occurs, the conformation of the RNA would be altered, making it suitable for another intron to be excised as new pairs of splicing sites become available. It is expected that conformational changes would take place after each splicing event until all the introns are removed.

Eukaryotic differentiation is characterised by the regulated expression of structurally distinct, developmentally regulated and cell type-specific protein isoforms (Breitbart et al, 1987). This protein diversity could arise either due to (a) preferential selection of one of the members of a multigene family for expression in a particular cell, developmental stage

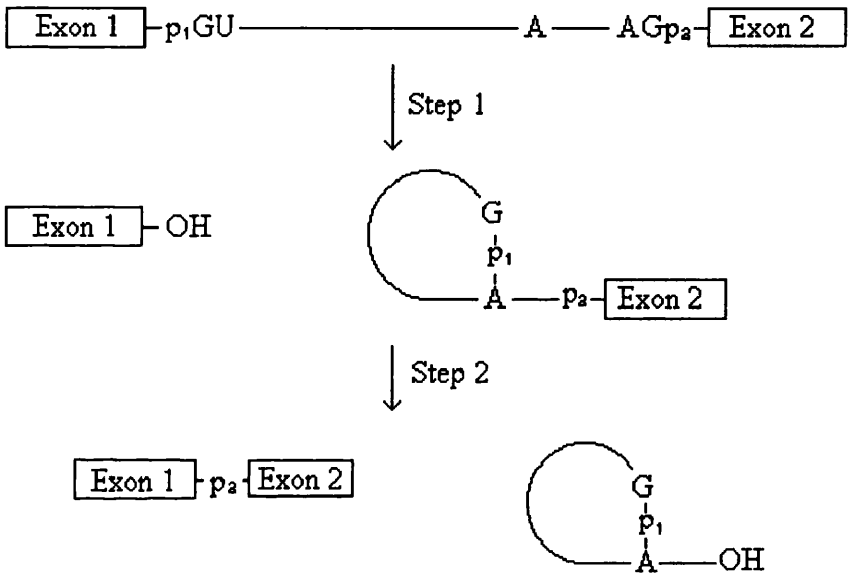


Figure 4.6: The pre-mRNA splicing pathway.

Step 1 involves the 5' splice site cleavage and lariat formation and step 2 involves the 3' splice site cleavage and exon joining. p indicates a phosphodiester bond.

or physiological condition; or (b) the generation of several different proteins from a single gene, where the molecular mechanism involved is alternative splicing. This leads to the differential expression of genomic sequences, thereby producing numerous protein isoforms from a single gene, varying only in their specific domains. This could allow the fine modulation of protein function (Smith et al, 1989). Through such changes, most aspects of protein function, ranging from the determination of cellular and subcellular localisation to the modulation of enzyme activity, can be altered. The number of different mRNAs, and hence protein isoforms encoded by a given gene, increases exponentially as a function of the number of exons that participate in alternative splicing. This, in turn, leads to phenotypic variability derived from a single gene. Alternative splicing can also quantitatively regulate gene expression by giving rise to prematurely truncated open reading frames, or by regulating mRNA stability or translational efficiency via variability in the untranslated regions.

A single base substitution within an intron can alter the splicing from a normal donor or acceptor site to a cryptic one that competes successfully for the splicing factors. Most cases of alternative splicing, involving internal donor or acceptor sites, as well as retained introns, are variant forms of this concept. If the resulting proteins produced by these alternative splice sites are advantageous, they will be retained. If a further mutation results in a new splice site behaving more strongly than the original one, it could be used constitutively where the coding capacity of the exons can be either increased by incorporating intronic sequences, or lost when part of the exon is excised.

The regulation of alternative splicing must reside in information encoded in the gene transcript (*cis*) but may require additional control by diffusible factors (*trans*) that may be cell-specific (Andreadis et al, 1987). Certain genes generate heterogeneous primary transcripts with either different promoters or polyadenylation sites. The information in *cis* may be sufficient to control the different splicing routes these different transcripts follow, hence specific *trans* factors may not be necessary. Other genes may, however, produce variant mRNAs from a single unique primary transcript that is initiated at the single promoter site and terminated at a single polyadenylation site. The execution of tissue-specific or developmentally regulated splicing in these identical transcripts cannot be

encoded in the *cis* elements only and requires additional control by cell-specific *trans* factors.

At this stage, neither the biological significance of alternative splicing nor the contribution of the missing exons 3 and 12 to the condition of patients with other *HMB-S* mutations is known. However, the presence of this phenomenon must be taken into account when providing clinical molecular genetic analysis of AIP i.e. the absence of exons 3 and 12 from RT-PCR products does not necessarily lead to the conclusion that the molecular pathology of the disorder has been defined.

4.1.3 *IN VITRO* GENE EXPRESSION

Of the six missense mutations detected in this study, three (R22C, R26C and L177R) consistently failed to express the desired protein in the *in vitro* system. A possible reason for the failure of these mutant clones to express the mutant proteins could be that the mutation results in an unstable protein product. Therefore, only three missense mutations, R26H, R116W and R173Q, were suitable for studying their effects on the HMB-S enzyme activity by using the expression system develop in this study.

The functional importance of three missense mutations identified in this study has been determined by using an *in vitro* expression system. From Table 3.4, the specific activities of the three missense mutations analysed, due to the human HMB-S protein, are 1.07 X 10^6 pmol/mg protein/hr (R26H), 0.93 X 10^6 pmol/mg protein/hr (R116W) and 0.83 X 10^6 pmol/mg protein/hr (R173Q), corresponding to 3.9%, 3.4% and 3% residual activity respectively. All these mutations have a very destructive effect on the mature protein, therefore, it was expected that a very much reduced activity would be observed. The HMB-S activities of all three mutated proteins are very similar to one another, probably indicating that the severity of these mutations down-regulates the enzyme activity to a similar extent. *In vitro* expression studies have been done by other groups and have found that the R116W mutation did not produce any enzyme activity in an *in vitro* system (Lee et al, 1990) and R173Q yielded a specific activity of 0.6% of the normal enzyme (Delfau et al, 1990). Therefore, the damaging nature of these mutations has been confirmed by *in vitro* expression studies.

Both the *in vivo* and *in vitro* systems have a similar pattern of expression, thus indicating that the *in vitro* system could be used as a representative of the *in vivo* system. However, these findings should be supported with further mutation analyses to provide convincing data that what is observed in the *in vitro* system can be correlated with the *in vivo* system.

The *in vitro* expression system can be further applied to the analysis of the other six mutations identified in this study to assess their effects on protein stability and activity. In addition to that, the transcripts missing exons 3 and 12 could also be expressed in the expression system to ascertain their effects on protein stability and enzyme activity.

4.2 FUTURE PERSPECTIVES

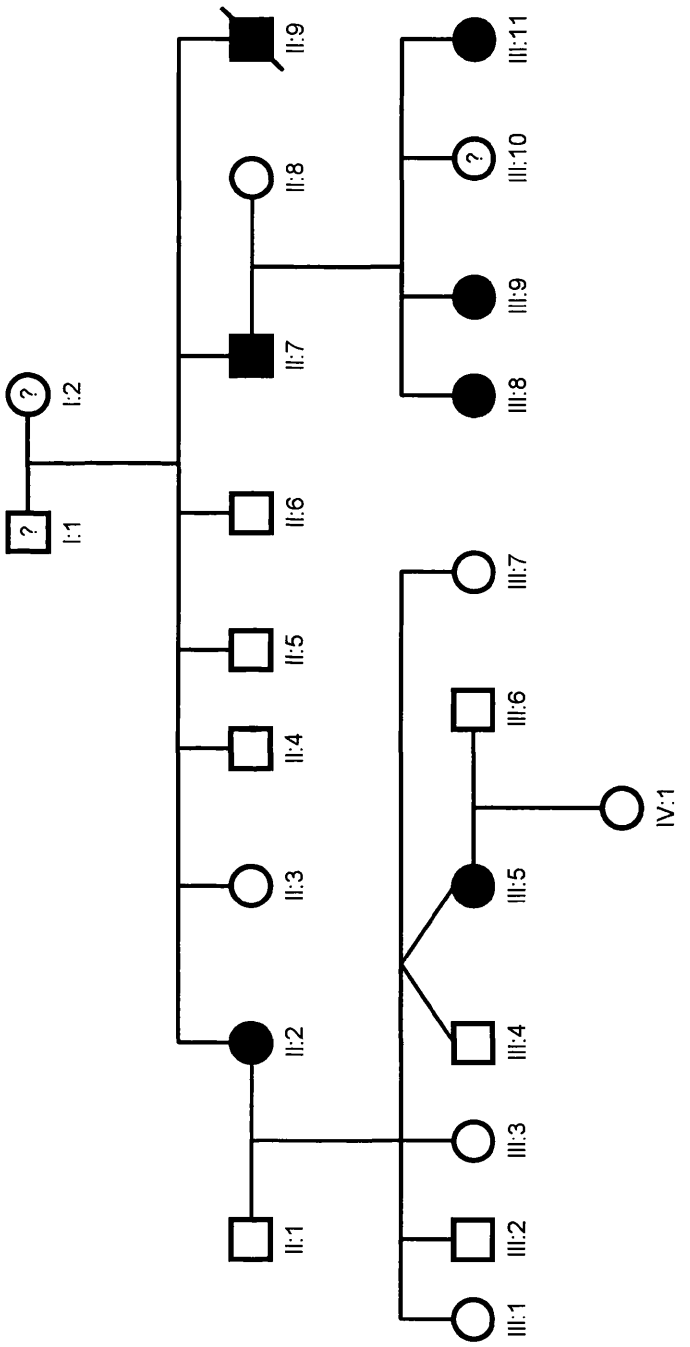
The mutations which have been identified during this project and by others will continue to throw light on the necessary structure of HMB-S, required for full enzyme activity. *In vitro* expression of these naturally occurring mutations will be particularly useful in this respect. For example, expression of the missense mutations (R22C, R26C and L177R) described above will help in this process. Also, our ability to target precise regions of the enzyme by site-directed mutagenesis should increase more rapidly, our understanding of the mechanisms involved in both the assembly of the dipyrromethane cofactor and the tetrapyrrolic product, hydroxymethylbilane. There is also the question of substrate binding and release of the tetrapyrrole from HMB-S.

The discovery, in this project, that exons 3 and 12 of HMB-S can be differentially spliced out in non-erythroid cells generates some interesting questions. For example, erythrocyte HMB-S activities should be measured in the general population and the proportion of alternatively spliced *HMB-S* mRNAs in these cells, determined. This would show whether or not there is a relationship between the range of enzyme activities known to occur and the level of alternative mRNA splicing present. It would also be intriguing to know if individuals with a higher proportion of exons 3 and 12 skipping and who also have a pathological mutation in HMB-S, have a higher risk of AIP crisis.

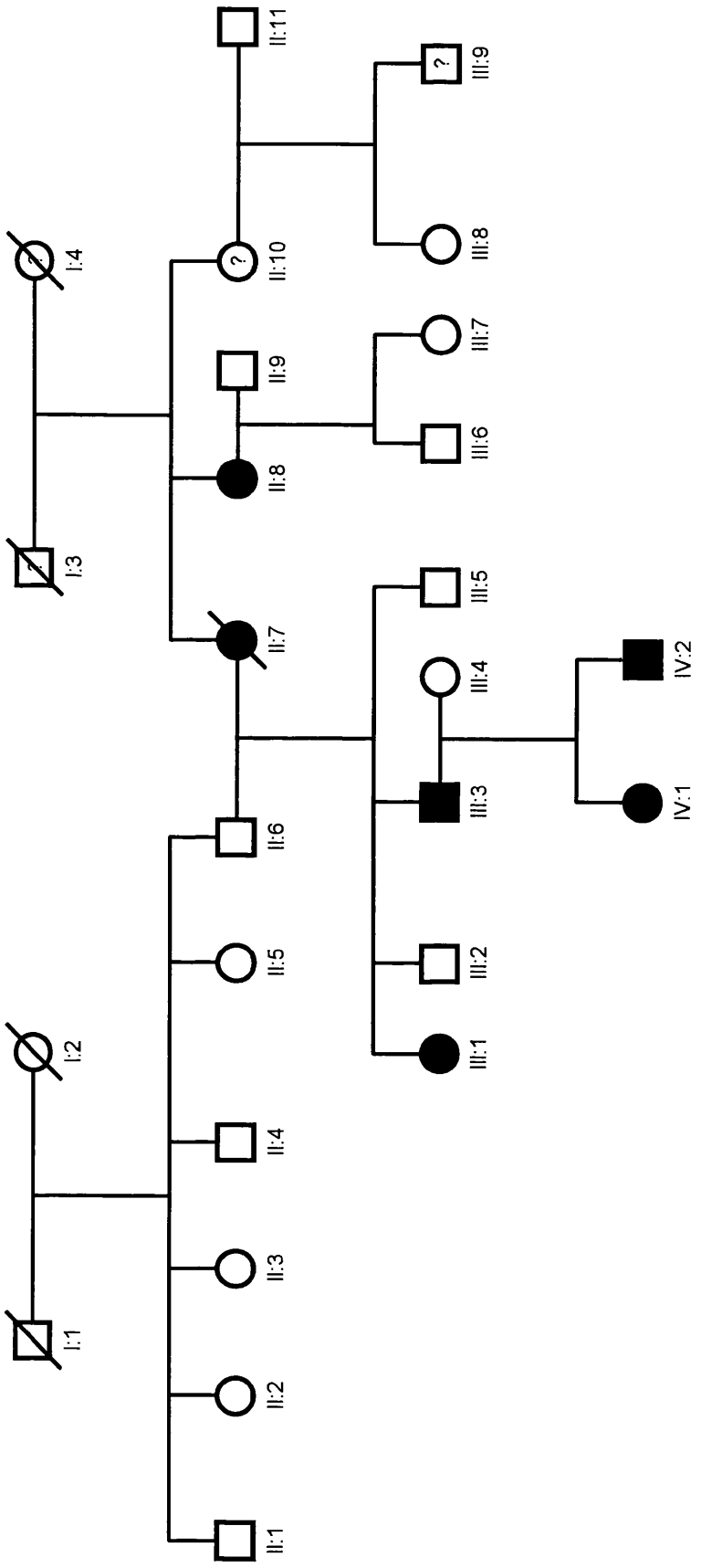
The detection of carriers of mutations in the *HMB-S* gene will continue to be a problem for diagnostic laboratories, given that there are no common mutations or marker haplotypes associated with the disorder. A reverse transcriptase/PCR analysis of the mRNA should be the first approach and will quickly reveal exon skipping in the cDNA. However, as discussed earlier, it is important to identify the splice signal mutations responsible for this. The small size of the cDNA (1083bp) means that its PCR product can be sequenced directly, paying particular attention to exons 10 and 12 which are known hotspots for mutation. Sequencing is, of course, the labour-intensive part of any diagnosis and therefore, until the technology has been simplified, it will not be used routinely in many molecular genetics diagnostic laboratories.

APPENDICES

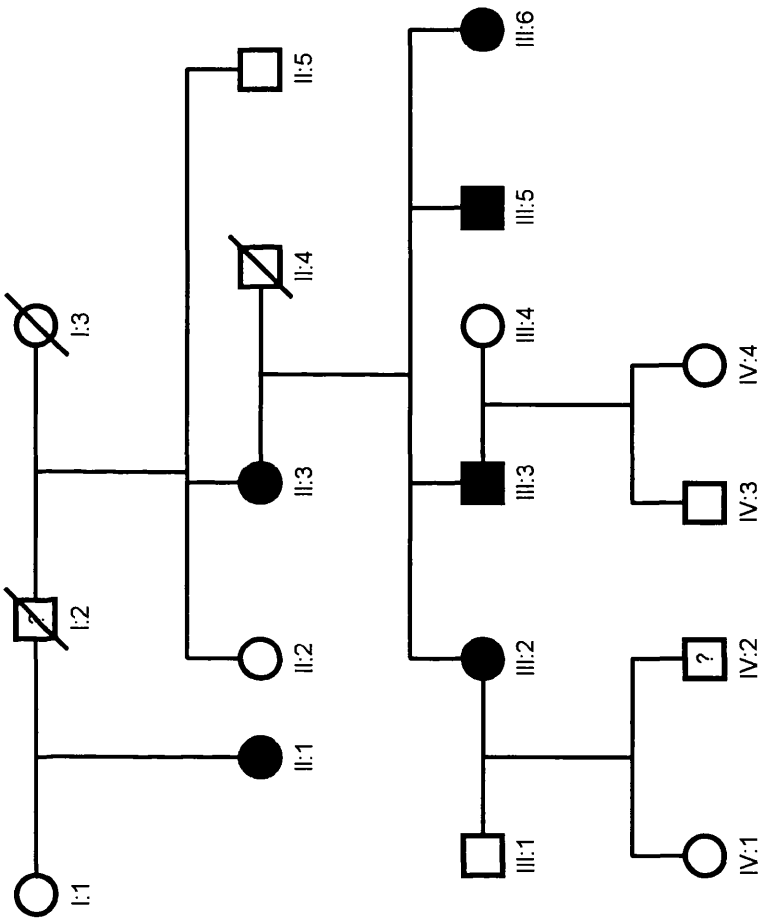
APPENDIX 1



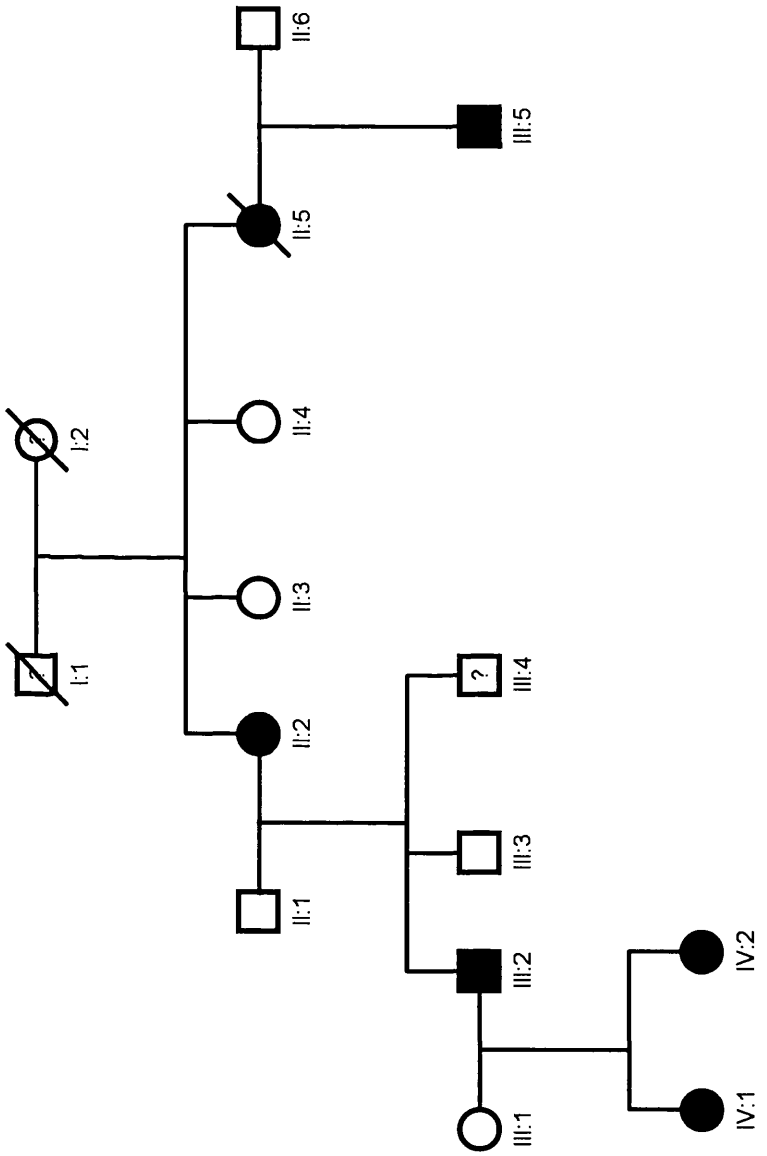
PEDIGREE 9532



PEDIGREE 13168



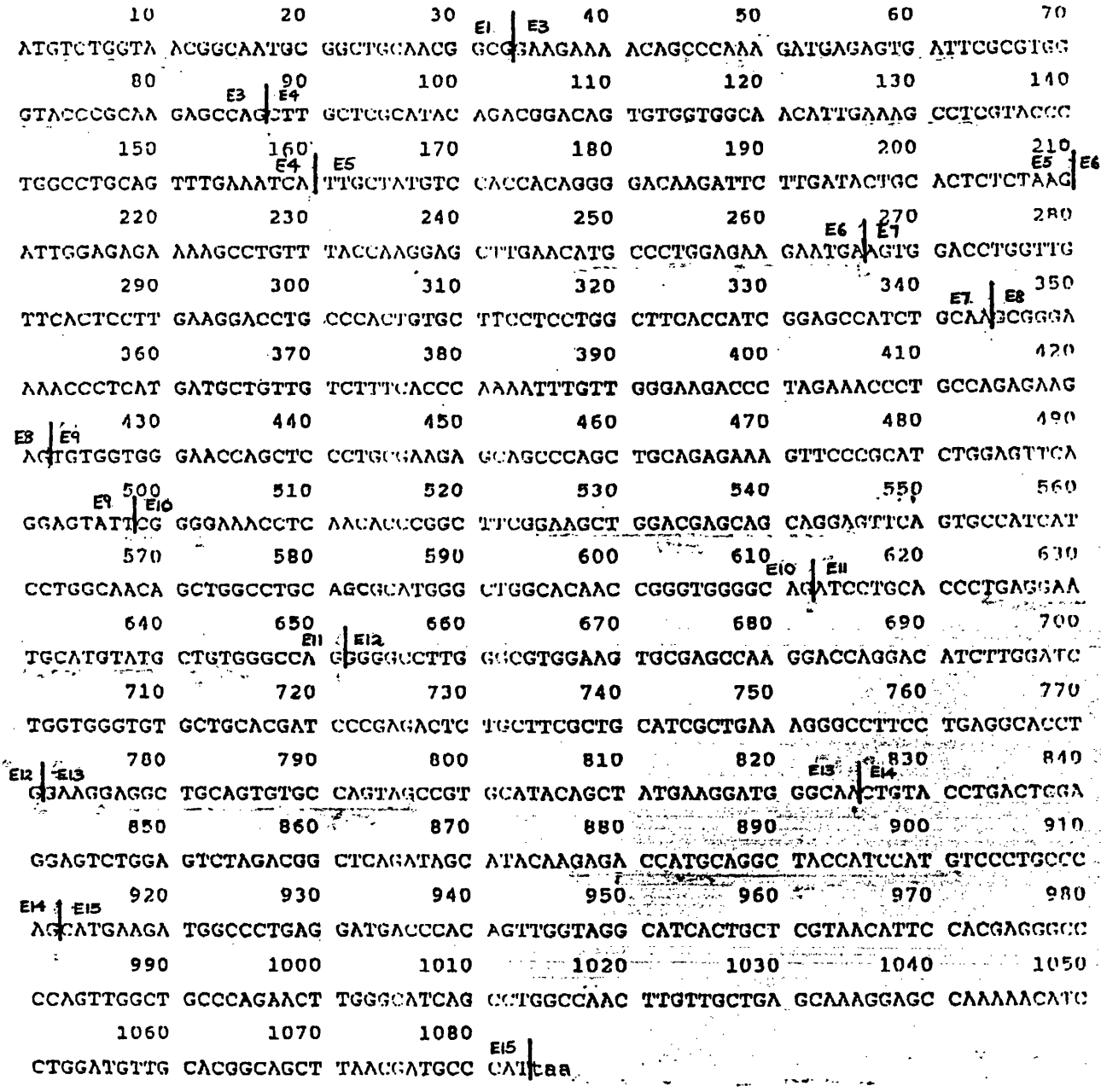
PEDIGREE 13543



PEDIGREE 13519

APPENDIX 2

Complete coding sequence of the hydroxymethylbilane synthase (*HMB-S*) gene.



NUCLEOTIDE POSITIONS	CODON	AMINO ACID	AMINO ACID SYMBOL	AMINO ACID NO.	EXON
1-3	ATG	MET	M	1	1
4-6	TCT	SER	S	2	1
7-9	GGT	GLY	G	3	1
10-12	AAC	ASN	N	4	1
13-15	GGC	GLY	G	5	1
16-18	AAT	ASN	N	6	1
19-21	GCG	ALA	A	7	1
22-24	GCT	ALA	A	8	1
25-27	GCA	ALA	A	9	1
28-30	ACG	THR	T	10	1
31-33	GCG	ALA	A	11	1
34-36	GAA	GLU	E	12	3
37-39	GAA	GLU	E	13	3
40-42	AAC	ASN	N	14	3
43-45	AGC	SER	S	15	3
46-48	CCA	PRO	P	16	3
49-51	AAG	LYS	K	17	3
52-54	ATG	MET	M	18	3
55-57	AGA	ARG	R	19	3
58-60	GTG	VAL	V	20	3
61-63	ATT	ILE	I	21	3
64-66	CGC	ARG	R	22	3
67-69	GTG	VAL	V	23	3
70-72	GGT	GLY	G	24	3
73-75	ACC	THR	T	25	3
76-78	CGC	ARG	R	26	3
79-81	AAG	LYS	K	27	3
82-84	AGC	SER	S	28	3
85-87	CAG	GLN	Q	29	3
88-90	CTT	LEU	L	30	4
91-93	GCT	ALA	A	31	4
94-96	CGC	ARG	R	32	4
97-99	ATA	ILE	I	33	4
100-102	CAG	GLN	Q	34	4
103-105	ACG	THR	T	35	4
106-108	GAC	ASP	D	36	4
109-111	AGT	SER	S	37	4
112-114	GTG	VAL	V	38	4
115-117	GTG	VAL	V	39	4
118-120	GCA	ALA	A	40	4
121-123	ACA	THR	T	41	4
124-126	TTG	LEU	L	42	4
127-129	AAA	LYS	K	43	4
130-132	GCC	ALA	A	44	4
133-135	TCG	SER	S	45	4
136-138	TAC	TYR	Y	46	4
139-141	CCT	PRO	P	47	4
142-144	GGC	GLY	G	48	4
145-147	CTG	LEU	L	49	4

148-150	CAG	GLN	Q	50	4
151-153	TTT	PHE	F	51	4
154-156	GAA	GLU	E	52	4
157-159	ATC	ILE	I	53	4
160-162	ATT	ILE	I	54	4-5
163-165	GCT	ALA	A	55	5
166-168	ATG	MET	M	56	5
169-171	TCC	SER	S	57	5
172-174	ACC	THR	T	58	5
175-177	ACA	THR	T	59	5
178-180	GGG	GLY	G	60	5
181-183	GAC	ASP	D	61	5
184-186	AAG	LYS	K	62	5
187-189	ATT	ILE	I	63	5
190-192	CTT	LEU	L	64	5
193-195	GAT	ASP	D	65	5
196-198	ACT	THR	T	66	5
199-201	GCA	ALA	A	67	5
202-204	CTC	LEU	L	68	5
205-207	TCT	SER	S	69	5
208-210	AAG	LYS	K	70	5
211-213	ATT	ILE	I	71	6
214-216	GGA	GLY	G	72	6
217-219	GAG	GLU	E	73	6
220-222	AAA	LYS	K	74	6
223-225	AGC	SER	S	75	6
226-228	CTG	LEU	L	76	6
229-231	TTT	PHE	F	77	6
232-234	ACC	THR	T	78	6
235-237	AAG	LYS	K	79	6
238-240	GAG	GLU	E	80	6
241-243	CTT	LEU	L	81	6
244-246	GAA	GLU	E	82	6
247-249	CAT	HIS	H	83	6
250-252	GCC	ALA	A	84	6
253-255	CTG	LEU	L	85	6
256-258	GAG	GLU	E	86	6
259-261	AAG	LYS	K	87	6
262-264	AAT	ASN	N	88	6
265-267	GAA	GLU	E	89	6-7
268-270	GTG	VAL	V	90	7
271-273	GAC	ASP	D	91	7
274-276	CTG	LEU	L	92	7
277-279	GTT	VAL	V	93	7
280-282	GTT	VAL	V	94	7
283-285	CAC	HIS	H	95	7
286-288	TCC	SER	S	96	7
289-291	TTG	LEU	L	97	7
292-294	AAG	LYS	K	98	7
295-297	GAC	ASP	D	99	7
298-300	CTG	LEU	L	100	7

301-303	CCC	PRO	P	101	7
304-306	ACT	THR	T	102	7
307-309	GTG	VAL	V	103	7
310-312	CTT	LEU	L	104	7
313-315	CCT	PRO	P	105	7
316-318	CCT	PRO	P	106	7
319-321	GGC	GLY	G	107	7
322-324	TTC	PHE	F	108	7
325-327	ACC	THR	T	109	7
328-330	ATC	ILE	I	110	7
331-333	GGA	GLY	G	111	7
334-336	GCC	ALA	A	112	7
337-339	ATC	ILE	I	113	7
340-342	TGC	CYS	C	114	7
343-345	AAG	LYS	K	115	7-8
346-348	CGG	ARG	R	116	8
349-351	GAA	GLU	E	117	8
352-354	AAC	ASN	N	118	8
355-357	CCT	PRO	P	119	8
358-360	CAT	HIS	H	120	8
361-363	GAT	ASP	D	121	8
364-366	GCT	ALA	A	122	8
367-369	GTT	VAL	V	123	8
370-372	GTC	VAL	V	124	8
373-375	TTT	PHE	F	125	8
376-378	CAC	HIS	H	126	8
379-381	CCA	PRO	P	127	8
382-384	AAA	LYS	K	128	8
385-387	TTT	PHE	F	129	8
388-390	GTT	VAL	V	130	8
391-393	GGG	GLY	G	131	8
394-396	AAG	LYS	K	132	8
397-399	ACC	THR	T	133	8
400-402	CTA	LEU	L	134	8
403-405	GAA	GLU	E	135	8
406-408	ACC	THR	T	136	8
409-411	CTG	LEU	L	137	8
412-414	CCA	PRO	P	138	8
415-417	GAG	GLU	E	139	8
418-420	AAG	LYS	K	140	8
421-423	AGT	SER	S	141	8-9
424-426	GTG	VAL	V	142	9
427-429	GTG	VAL	V	143	9
430-432	GGA	GLY	G	144	9
433-435	ACC	THR	T	145	9
436-438	AGC	SER	S	146	9
439-441	TCC	SER	S	147	9
442-444	CTG	LEU	L	148	9
445-447	CGA	ARG	R	149	9
448-450	AGA	ARG	R	150	9
451-453	GCA	ALA	A	151	9

454-456	GCC	ALA	A	152	9
457-459	CAG	GLN	Q	153	9
460-462	CTG	LEU	L	154	9
463-465	CAG	GLN	Q	155	9
466-468	AGA	ARG	R	156	9
469-471	AAG	LYS	K	157	9
472-474	TTC	PHE	F	158	9
475-477	CCG	PRO	P	159	9
478-480	CAT	HIS	H	160	9
481-483	CTG	LEU	L	161	9
484-486	GAG	GLU	E	162	9
487-489	TTC	PHE	F	163	9
490-492	AGG	ARG	R	164	9
493-495	AGT	SER	S	165	9
496-498	ATT	ILE	I	166	9
499-501	CGG	ARG	R	167	10
502-504	GGA	GLY	G	168	10
505-507	AAC	ASN	N	169	10
508-510	CTC	LEU	L	170	10
511-513	AAC	ASN	N	171	10
514-516	ACC	THR	T	172	10
517-519	CGG	ARG	R	173	10
520-522	CTT	LEU	L	174	10
523-525	CGG	ARG	R	175	10
526-528	AAG	LYS	K	176	10
529-531	CTG	LEU	L	177	10
532-534	GAC	ASP	D	178	10
535-537	GAG	GLU	E	179	10
538-540	CAG	GLN	Q	180	10
541-543	CAG	GLN	Q	181	10
544-546	GAG	GLU	E	182	10
547-549	TTC	PHE	F	183	10
550-552	AGT	SER	S	184	10
553-555	GCC	ALA	A	185	10
556-558	ATC	ILE	I	186	10
559-561	ATC	ILE	I	187	10
562-564	CTG	LEU	L	188	10
565-567	GCA	ALA	A	189	10
568-570	ACA	THR	T	190	10
571-573	GCT	ALA	A	191	10
574-576	GGC	GLY	G	192	10
577-579	CTG	LEU	L	193	10
580-582	CAG	GLN	Q	194	10
583-585	CGC	ARG	R	195	10
586-588	ATG	MET	M	196	10
589-591	GGC	GLY	G	197	10
592-594	TGG	TRP	W	198	10
595-597	CAC	HIS	H	199	10
598-600	AAC	ASN	N	200	10
601-603	CGG	ARG	R	201	10
604-606	GTG	VAL	V	202	10

607-609	GGG	GLY	G	203	10
610-612	CAG	GLN	Q	204	10
613-615	ATC	ILE	I	205	11
616-618	CTG	LEU	L	206	11
619-621	CAC	HIS	H	207	11
622-624	CCT	PRO	P	208	11
625-627	GAG	GLU	E	209	11
628-630	GAA	GLU	E	210	11
631-633	TGC	CYS	C	211	11
634-636	ATG	MET	M	212	11
637-639	TAT	TYR	Y	213	11
640-642	GCT	ALA	A	214	11
643-645	GTG	VAL	V	215	11
646-648	GGC	GLY	G	216	11
649-651	CAG	GLN	Q	217	11
652-654	GGG	GLY	G	218	12
655-657	GCC	ALA	A	219	12
658-660	TTG	LEU	L	220	12
661-663	GGC	GLY	G	221	12
664-666	GTG	VAL	V	222	12
667-669	GAA	GLU	E	223	12
670-672	GTG	VAL	V	224	12
673-675	CGA	ARG	R	225	12
676-678	GCC	ALA	A	226	12
679-681	AAG	LYS	K	227	12
682-684	GAC	ASP	D	228	12
685-687	CAG	GLN	Q	229	12
688-690	GAC	ASP	D	230	12
691-693	ATC	ILE	I	231	12
694-696	TTG	LEU	L	232	12
697-699	GAT	ASP	D	233	12
700-702	CTG	LEU	L	234	12
703-705	GTG	VAL	V	235	12
706-708	GGT	GLY	G	236	12
709-711	GTG	VAL	V	237	12
712-714	CTG	LEU	L	238	12
715-717	CAC	HIS	H	239	12
718-720	GAT	ASP	D	240	12
721-723	CCC	PRO	P	241	12
724-726	GAG	GLU	E	242	12
727-729	ACT	THR	T	243	12
730-732	CTG	LEU	L	244	12
733-735	CTT	LEU	L	245	12
736-738	CGC	ARG	R	246	12
739-741	TGC	CYS	C	247	12
742-744	ATC	ILE	I	248	12
745-747	GCT	ALA	A	249	12
748-750	GAA	GLU	E	250	12
751-753	AGG	ARG	R	251	12
754-756	GCC	ALA	A	252	12
757-759	TTC	PHE	F	253	12

760-762	CTG	LEU	L	254	12
763-765	AGG	ARG	R	255	12
766-768	CAC	HIS	H	256	12
769-771	CTG	LEU	L	257	12
772-774	GAA	GLU	E	258	13
775-777	GGA	GLY	G	259	13
778-780	GGC	GLY	G	260	13
781-783	TGC	CYS	C	261	13
784-786	AGT	SER	S	262	13
787-789	GTG	VAL	V	263	13
790-792	CCA	PRO	P	264	13
793-795	GTA	VAL	V	265	13
796-798	GCC	ALA	A	266	13
799-801	GTG	VAL	V	267	13
802-804	CAT	HIS	H	268	13
805-807	ACA	THR	T	269	13
808-810	GCT	ALA	A	270	13
811-813	ATG	MET	M	271	13
814-816	AAG	LYS	K	272	13
817-819	GAT	ASP	D	273	13
820-822	GGG	GLY	G	274	13
823-825	CAA	GLN	Q	275	13
826-828	CTG	LEU	L	276	14
829-831	TAC	TYR	Y	277	14
832-834	CTG	LEU	L	278	14
835-837	ACT	THR	T	279	14
838-840	GGA	GLY	G	280	14
841-843	GGA	GLY	G	281	14
844-846	GTC	VAL	V	282	14
847-849	TGG	TRP	W	283	14
850-852	AGT	SER	S	284	14
853-855	CTA	LEU	L	285	14
856-858	GAC	ASP	D	286	14
859-861	GGC	GLY	G	287	14
862-864	TCA	SER	S	288	14
865-867	GAT	ASP	D	289	14
868-870	AGC	SER	S	290	14
871-873	ATA	ILE	I	291	14
874-876	CAA	GLN	Q	292	14
877-879	GAG	GLU	E	293	14
880-882	ACC	THR	T	294	14
883-885	ATG	MET	M	295	14
886-888	CAG	GLN	Q	296	14
889-891	GCT	ALA	A	297	14
892-894	ACC	THR	T	298	14
895-897	ATC	ILE	I	299	14
898-900	CAT	HIS	H	300	14
901-903	GTC	VAL	V	301	14
904-906	CCT	PRO	P	302	14
907-909	GCC	ALA	A	303	14
910-912	CAG	GLN	Q	304	14

913-915	CAT	HIS	H	305	15
916-918	GAA	GLU	E	306	15
919-921	GAT	ASP	D	307	15
922-924	GGC	GLY	G	308	15
925-927	CCT	PRO	P	309	15
928-930	GAG	GLU	E	310	15
931-933	GAT	ASP	D	311	15
934-936	GAC	ASP	D	312	15
937-939	CCA	PRO	P	313	15
940-942	CAG	GLN	Q	314	15
943-945	TTG	LEU	L	315	15
946-948	GTA	VAL	V	316	15
949-951	GGC	GLY	G	317	15
952-954	ATC	ILE	I	318	15
955-957	ACT	THR	T	319	15
958-960	GCT	ALA	A	320	15
961-963	CGT	ARG	R	321	15
964-966	AAC	ASN	N	322	15
967-969	ATT	ILE	I	323	15
970-972	CCA	PRO	P	324	15
973-975	CGA	ARG	R	325	15
976-978	GGG	GLY	G	326	15
979-981	CCC	PRO	P	327	15
982-984	CAG	GLN	Q	328	15
985-987	TTG	LEU	L	329	15
988-990	GCT	ALA	A	330	15
991-993	GCC	ALA	A	331	15
994-996	CAG	GLN	Q	332	15
997-999	AAC	ASN	N	333	15
1000-1002	TTG	LEU	L	334	15
1003-1005	GGC	GLY	G	335	15
1006-1008	ATC	ILE	I	336	15
1009-1011	AGC	SER	S	337	15
1012-1014	CTG	LEU	L	338	15
1015-1017	GCC	ALA	A	339	15
1018-1020	AAC	ASN	N	340	15
1021-1023	TTG	LEU	L	341	15
1024-1026	TTG	LEU	L	342	15
1027-1029	CTG	LEU	L	343	15
1030-1032	AGC	SER	S	344	15
1033-1035	AAA	LYS	K	345	15
1036-1038	GGA	GLY	G	346	15
1039-1041	GCC	ALA	A	347	15
1042-1044	AAA	LYS	K	348	15
1045-1047	AAC	ASN	N	349	15
1048-1050	ATC	ILE	I	350	15
1051-1053	CTG	LEU	L	351	15
1054-1056	GAT	ASP	D	352	15
1057-1059	GTT	VAL	V	353	15
1060-1062	GCA	ALA	A	354	15
1063-1065	CGG	ARG	R	355	15

1066-1068	CAG	GLN	Q	356	15
1069-1071	CTT	LEU	L	357	15
1072-1074	AAC	ASN	N	358	15
1075-1077	GAT	ASP	D	359	15
1078-1080	GCC	ALA	A	360	15
1081-1083	CAT	HIS	H	361	15

REFERENCES

- Alber, T. (1989) Mutational effects of protein stability. *Ann. Rev. Biochem.* **58**, 765-798.

- Andreadis, A., Gallego, M.E. & Nadal-Ginard, B. (1987) Generation of protein isoform diversity by alternative splicing: mechanistic and biological implications. *Ann. Rev. Cell Biol.* **3**, 207-242.

- Angelini, C., Beggs, A.H., Hoffman, E.P., Fanin, M. & Kunkel, L.M. (1990) Enormous dystrophin in a patient with Becker muscular dystrophy. *Neurology* **40**, 808-812.

- Astrin, K.H. & Desnick, R.J. (1994) Molecular basis of acute intermittent porphyria: Mutations and polymorphisms in the human hydroxymethylbilane synthase gene. *Hum. Mut.* **4**, 243-252.

- Barker, D.F., Schafer, M. & White, R. (1984) Restriction sites containing CpG show a higher frequency of polymorphism in human DNA. *Cell* **36**, 131-138.

- Bettecken, T. & Müller, C.R. (1989) Identification of a 220kb insertion into the Duchenne gene in a family with an atypical course of muscular dystrophy. *Genomics* **4**, 592-596

- Bonnaiti-Pellie, C., Phung, L. & Nordmann, Y. (1984) Recurrence risk estimation of acute intermittent porphyria based on analysis of porphobilinogen deaminase activity: A Bayesian approach. *Am. J. Med. Genet.* **19**, 755-762.

- Breitbart, R.E., Andreadis, A. & Nadal-Ginard, B. (1987) Alternative splicing: A ubiquitous mechanism for the generation of multiple protein isoforms from single genes. *Ann. Rev. Biochem.* **56**, 467-495.

- Brodie, M.J., Moore, M.R. & Goldberg, A. (1977) Enzyme abnormalities in the porphyrias. *Lancet* **2**, 699-701.

- Brownlie, P.D., Lambert, R., Louie, G.V., Jordan, P.M., Blundell, T.L., Warren, M.J., Cooper, J.B. & Wood, S.P. (1994) The three-dimensional structures of mutants of porphobilinogen deaminase: Toward an understanding of the structural basis of acute intermittent porphyria. *Protein Science* **3**, 1644-1650.

- Burke, J.R., Wingfield, M.S., Lewis, K.E., Roses, A.D., Lee, J.E., Hulette, C., Pericak-Vance, M.A. & Vance, J.M. (1994) The Haw River Syndrome: Dentatorubropallidoluysian atrophy (DRPLA) in an African-American family. *Nature Genet.* **7**, 521-524.

- Chen, C.H., Warner, C.A. & Desnick, R.J. (1992) Acute intermittent porphyria: Identification of novel missense mutations in human hydroxymethylbilane synthase gene. *Am. J. Hum. Genet.* **51**, A45.

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

- Chretien, S., Dubart, A., Beaupain, D., Raich, N., Grandchamp, B., Rosa, J., Goossens, M. & Romeo, P-H. (1988) Alternative transcription and splicing of the human porphobilinogen deaminase gene result either in tissue-specific or in housekeeping expression. *Proc. Natl. Acad. Sci. USA* **85**, 6-10.

- Cooper, D.N. (1983) Eukaryotic DNA methylation. *Hum. Genet.* **64**, 315-333.

- Cooper, D.N. & Krawczak, M. (1991) Mechanisms of insertional mutagenesis in human genes causing genetic disease. *Hum. Genet.* **87**, 519-560.

- Cooper, D.N. & Krawczak, M. (1993) Human Gene Mutation, Bios Scientific Publishers.

- Cooper, D.N. & Youssoufian, H. (1988) The CpG dinucleotide and human genetic disease. *Hum. Genet.* **78**, 151-155.

- 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.

- Daimon, M., Yamatani, K., Igarashi, M., Fukase, N., Morita, Y., Ogawa, A., Tominaga, M. & Sasaki, H. (1994) Acute intermittent porphyria caused by a single base insertion of C in exon 15 of the porphobilinogen deaminase gene that results in a frame shift and premature stopping of translation. *Hum. Genet.* **93**, 533-537.

- Daimon, M., Yamatani, K., Igarashi, M., Fukase, N., Ogawa, A., Tominaga, M. & Sasaki, H. (1993) Acute intermittent porphyria caused by a G to C mutation in exon 12 of the porphobilinogen deaminase gene that results in exon skipping. *Hum. Genet.* **92**, 549-553.

- de Rooij, F.W.M., Hamer, C.M. & Wilson, J.H.P. (1987) Purification of porphobilinogen deaminase from human erythrocytes by fast protein liquid chromatography. *Clin. Chim. Acta* **162**, 61-68.

- Delfau, M.H., Picat, C., de Rooij, F.W.M., Hamer, K., Bogard, M., Wilson, J.H.P., Deybach, J.C., Nordmann, Y. & Grandchamp B. (1990) Two different point G to A mutations in exon 10 of the porphobilinogen deaminase gene are responsible for acute intermittent porphyria. *J. Clin. Invest.* **86**, 1511-1516.

- Delfau, M.H., Picat, C., de Rooij, F., Voortman, G., Deybach, J.C., Nordmann, Y. & Grandchamp, B. (1991) Molecular heterogeneity of acute intermittent porphyria: Identification of four additional mutations resulting in the CRIM-negative subtype of the disease. *Am. J. Hum. Genet.* **49**, 421-428.

- Desnick, R.J., Ostasiewicz, L.T., Tishler, P.A. & Mustajoki, P. (1985) Acute intermittent porphyria: Characterization of a novel mutation in the structural gene for porphobilinogen deaminase. *J. Clin. Invest.* **76**, 865-874.

- Fischer, S.G. & Lerman, L.S. (1983) DNA fragments differing by a single base-pair substitution are separated in denaturing gradient gels: Correspondence with melting theory. *Proc. Natl. Acad. Sci. USA* **80**, 1579-1583.

- Forrest, S.M., Dahl, H.H., Howells, D.W., Dianzani, I. & Cotton, R.G. (1991) Mutation detection in phenylketonuria by using chemical cleavage of mismatch: importance of using probes from both normal and patient samples. *Am. J. Hum. Genet.* **49**, 175-183.

- Freifelder, D. (1987) in Molecular Biology. Boston: Jones and Bartlett Publishers, Inc.

- Fu, Y.H., Kuhl, D.P.A., Pizzuti, A., Pierreti, M., Sutcliffe, J.S., Richards, S., Verkerk, A.J.M.H., Holden, J.J.A., Fenwick, R.G., Warren, S.T., Oostra, B., Nelson, D.L. & Caskey, C.T. (1991) Variation of CGG repeat at the fragile X site results in genetic instability: resolution of the Sherman paradox. *Cell* **67**, 1047-1058.

- Fu, Y.H., Pizzuti, A., Fenwick, R.G., King, J., Rajanarayan, S., Dunne, P.W., Dubel, J., Nasser, G., Ashiwaza, T., De Jong, P., Wieringa, B., Korneluk, R., Perryman, M.B., Epstein, H. and Caskey, T. (1992) An unstable triplet repeat in a gene related to myotonic muscular dystrophy. *Science* **255**, 1256-1258.

- Fukao, T., Yamaguchi, S., Wakazono, A., Orii, T., Hoganson, G. & Hashimoto, T. (1994) Identification of a novel exonic mutation at -13 from 5' splice site causing exon skipping in a girl with mitochondrial acetoacetyl-coenzyme A thiolase deficiency. *J. Clin. Invest.* **93**, 1035-1041.

- Gates, R.R. (1946) Human Genetics. p530. New York: Macmillan.

- Goeddel, D.V. (1990) Gene Expression Technology. Academic Press.

- Goldberg, A. (1959) Acute intermittent porphyria: A study of 50 cases. *Q. J. Med. N.S.* **28**, 183-209.

- Goldberg, A., Moore, M.R., McColl, K.E.L. & Brodie, M. J. (1987) in Oxford Textbook of Medicine. pp 9136-9145. Oxford University Press.

- Grandchamp, B., De Verneuil, H., Beaumont, C., Chretien, S., Walter, O. & Nordmann, Y. (1987) Tissue-specific expression of porphobilinogen deaminase: Two isoenzymes from a single gene. *Eur. J. Biochem.* **162**, 105-110.

- Grandchamp, B., Picat, C., de Rooij, F., Beaumont, C., Wilson, P., Deybach, J.C. & Nordmann, Y. (1989a) A point mutation G→A in exon 12 of the porphobilinogen deaminase gene results in exon skipping and is responsible for acute intermittent porphyria. *Nucleic Acids Res.* **17**, 6637-6649.

- Grandchamp, B., Picat, C., Kauppinen, R., Mignotte, V., Peltonen, L., Mustajoki, P., Romeo, P.H., Goossens, M. & Nordmann, Y. (1989b) Molecular analysis of acute intermittent porphyria in a Finnish family with normal erythrocyte porphobilinogen deaminase. *Eur. J. Clin. Invest.* **19**, 415-418.

- Grandchamp, B., Picat, C., Mignotte, V., Wilson, J.H.P., Te Velde, K., Sandkuyl, L., Romeo, P.H., Goossens, M. & Nordmann, Y. (1989c) Tissue-specific splicing mutation in acute intermittent porphyria. *Proc. Natl. Acad. Sci. USA* **86**, 661-664.

- Green, M.R. (1986) Pre-mRNA splicing. *Ann. Rev. Genet.* **20**, 671-708.

- Grompe, M. (1993) The rapid detection of unknown mutations in nucleic acids. *Nature Genet.* **5**, 111-117.

- Gu, X-F., de Rooij, F., Lee, J-S., Te Velde, K., Deybach, J.C., Nordmann, Y. & Grandchamp, B. (1993) High prevalence of a point mutation in the porphobilinogen deaminase gene in Dutch patients with acute intermittent porphyria. *Hum. Genet.* **91**, 128-130.

- Gu, X-F., de Rooij, F., de Baar, E., Bruyland, M., Lissens, W., Nordmann, Y. & Grandchamp, B. (1993a) Two novel mutations of the porphobilinogen deaminase gene in acute intermittent porphyria. *Hum. Mol. Genet.* **2**, 1735-1736.

- Gu, X-F., de Rooij, F., Voortman, G., Te Velde, K., Deybach, J.C., Nordmann, Y. & Grandchamp, B. (1994) Detection of eleven mutations causing acute intermittent porphyria using denaturing gradient gel electrophoresis. *Hum. Genet.* **93**, 47-52.

- Gyllensten, U.B. & Erlich, H.A. (1988) Generation of single-stranded DNA by the polymerase chain reaction and its application to direct sequencing of the HLA-DQA locus. *Proc. Natl. Acad. Sci. USA* **85**, 7652-7656.

- Hamilton, R.J. & Sewell, P.A. (1982) Introduction to high performance liquid chromatography. London: Chapman and Hall.

- Huntington's Disease Collaborative Research Group (1993) A novel gene containing a trinucleotide repeat that is expanded and stable on Huntington's disease chromosomes. *Cell* **72**, 971-983.

- Jennings, M.W., Jones, R.W., Wood, W.G. & Weatherall, D.J. (1985) Analysis of an inversion within the human β-globin gene cluster. *Nucleic Acids Res.* **13**, 2897-2906.

- Jordan, P.M.. & Woodcock, S.C. (1991) Mutagenesis of arginine residues in the catalytic cleft of Escherichia coli porphobilinogen deaminase that affects dipyrromethane cofactor assembly and tetrapyrrole chain initiation and elongation. *Biochem. J.* **280**, 445-449.

- Kappas, A., Sassa, S., Galbraith, R.A. & Nordmann, Y. (1989) in The Metabolic Basis of Inherited Disease, ed. Scriver, C.R., Beaudet, A.L., Sly, W.S. & Valle, D. Ch.52, pp 1305-1365. New York: McGraw-Hill.

- Kappas, A., Sassa, S., Galbraith, R.A. & Nordmann, Y. (1995) in The Metabolic and Molecular Bases of Inherited Disease, ed. Scriver, C.R., Beaudet, A.L., Sly, W.S. & Valle, D. Ch. 66, pp 2103-2159. New York: McGraw-Hill.

- Kauppinen, R., Mustajoki, S., Pihlaja, H., Peltonen, L. & Mustajoki, P. (1995) Acute intermittent porphyria in Finland: 19 mutations in the porphobilinogen deaminase gene. *Hum. Mol. Genet.* **4**, 215-222.

- Koide, R., Ikeuchi, T., Onodera, O., Tanaka, H., Igarashi, S., Endo, K., Takahashi, H., Kondo, R., Ishikawa, A., Hayashi, T., Saito, M., Tomoda, A., Miike, T., Naito, H., Ikuta, F. & Tsuji, S. (1994) Unstable expansion of CAG repeat in hereditary dentatorubral-pallidoluysian atrophy (DRPLA). *Nature Genet.* **6**, 9-13.

- Krainer, A.R. & Maniatis, T. (1985) Multiple factors including small nuclear riboproteins U1 and U2 are necessary for the pre-mRNA splicing *in vitro*. *Cell* **42**, 725-736.

- Krawczak, M. & Cooper, D.N. (1991) Gene deletions causing human genetic disease: Mechanisms of mutagenesis and the role of the local DNA sequence environment. *Hum. Genet.* **86**, 425-441.

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

- Kunkel, L.M., Smith, K.D., Boyer, S.H., Borgaonkar, D.S., Watchel, S.S., Miller, O.J., Breg, W.R., Jones, H.W. Jr. & Ray, J.M. (1977) Analysis of human Y chromosome-specific reiterated DNA in chromosome variants. *Proc. Natl. Acad. Sci. USA* **74**, 1245-1249.

- La Spada, A.R., Wilson, E.M., Lubahn, D.B., Harding, A.E. & Fischbeck, K.H. (1991) Androgen receptor gene mutations in X-linked spinal muscular atrophy. *Nature* **352**, 77-79.

- Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680-685.

- Lamon, J.M., Frykholm, B.C. & Tschudy, D.P. (1979) Family evaluations in acute intermittent porphyria using red cell uroporphyrinogen I synthetase. *J. Med. Genet.* **16**, 134-139.

- Lander, M., Pitt, A.R., Alefounder, P.R., Bardy, D., Abell, C. & Battersby, A.R. (1991) Studies on the mechanism of hydroxymethylbilane synthase concerning the role of arginine residues in substrate binding. *Biochem. J.* **275**, 447-452.

- Lee, G.Y., Astrin, K.H. & Desnick, R.J. (1995) Acute intermittent porphyria: A single-base deletion and a nonsense mutation in the human hydroxymethylbilane synthase gene, predicting truncations of the enzyme polypeptide. *Am. J. Med. Genet.* **58**, 155-158.

- Lee, J-S. (1991) Molecular genetic investigation of the human porphobilinogen deaminase gene in acute intermittent porphyria. Ph.D. thesis, Stockholm.

- Lee, J-S. & Anvret, M. (1987) A *Pst* I polymorphism for the human porphobilingen deaminase gene. *Nucleic Acids Res.* **15**, 6307.

- Lee, J-S. & Anvret, M. (1991) Identification of the most common mutation within the porphobilinogen deaminase gene in Swedish patients with acute intermittent porphyria. *Proc. Natl. Acad. Sci. USA* **88**, 10912-10915.

- Lee, J-S., Anvret, M., Lindsten, J., Lannfelt, L., Gellerfors, P., Wetterberg, L., Floderus, Y. & Thunell, S. (1988) DNA polymorphisms within the porphobilinogen deaminase gene in two Swedish families with acute intermittent porphyria. *Hum. Genet.* **79**, 379-381.

- Lee, J-S., Grandchamp, B. & Anvret, M. (1990) A point mutation of the human porphobilinogen deaminase gene in a Swedish family with acute intermittent porphyria. *Am. J. Hum. Genet.* **47**, A162 (Supplement).

- Lee, J-S., Lundin, G., Lannfelt, L., Forsell, L., Picat, C., Grandchamp, B & Anvret, M. (1991) Genetic heterogeneity of the porphobilinogen deaminase gene in Swedish families with acute intermittent porphyria. *Hum. Genet.* **87**, 484-488.

- Llewellyn, D.H., Elder, G.H., Kalsheker, N.A., Marsh, O.W.M., Harrison, P.R., Grandchamp, B., Picat, C., Nordmann, Y., Romeo, P.H. & Goossens, M. (1987) DNA polymorphism of human porphobilinogen deaminase gene in acute intermittent porphyria. *Lancet* **2**, 706-708.

- Llewellyn, D.H., Scobie, G.A., Urquhart, A.J., Whatley, S.D., Roberts, A.G., Harrison, P.R. & Elder, G.H. (1996) Acute intermittent porphyria caused by defective splicing of porphobilinogen deaminase RNA: a synonymous codon mutation at -22bp from the 5' splice site causes skipping of exon 3. *J. Med. Genet.* **33**, 437-438.

- Llewellyn, D.H., Smyth, S.J., Elder, G.H., Hutchesson, A.C., Rattenbury, J.M. & Smith, M.F. (1992) Homozygous acute intermittent porphyria: compound heterozygosity for adjacent base transitions in the same codon of the porphobilinogen deaminase gene. *Hum. Genet.* **89**, 97-98.

- Llewellyn, D.H., Whatley, S. & Elder, G.H. (1993) Acute intermittent porphyria caused by an arginine to histidine substitution (R26H) in the cofactor-binding cleft of porphobilinogen deaminase. *Hum. Mol. Genet.* **2**. 1315-1316.

- Louie, G.V., Brownlie, P.D., Lambert, R., Cooper, J.B., Blundell, T.L., Wood, S.P., Warren, M.J., Woodcock, S.C. & Jordan, P.M. (1992) Structure of porphobilinogen deaminase reveals a flexible multidomain polymerase with a single catalytic site. *Nature* **359**, 33-39.

- Louie, G.V., Brownlie, P.D., Lambert, R., Cooper, J.B., Blundell, T.L., Wood, S.P., Malashkevich, V.N., Hädener, A., Warren, M.J. & Shoolingin-Jordan, P.M. (1996) The three-dimensional structure of *Escherichia coli* porphobilinogen deaminase at 1.76-Å resolution. *Proteins: Structure, Function and Genetics* **25**, 48-78.

- Lowry, O.H., Rosebrough, N.J., Farr, A.L. & Randall, R.J. (1951) Protein measurement with the Folin phenol reagent. *J. Biol. Chem.* **193**, 265-275.

- Lundin, G., Hashemi, J., Floderus, Y., Thunell, S., Sagen, E., Lægreid, A., Wassif, W., Peters, T. & Anvret, M. (1995) Four mutations in the porphobilinogen deaminase gene in patients with acute intermittent porphyria. *J. Med. Genet.* **32**, 979-981.

- Lundin, G., Wedell, A., Thunell, S. & Anvret, M. (1994) Two new mutations in the porphobilinogen deaminase gene and a screening method using PCR amplification of specific alleles. *Hum. Genet.* **93**, 59-62.

- Maciel, P., Gaspar, C., DeStefano, A.L., Silveira, I., Coutinho, P., Radvany, J., Dawson, D.M., Sudarsky, L., Guimarães, J., Loureiro, J.E.L., Nezarati, M.M., Corwin, L.I., Lopes-Cendes, I., Rooke, K., Rosenberg, R., MacLeod, P., Farrer, L.A., Sequeiros, J. & Rouleau, G.A. (1995) Correlation between CAG repeat length and clinical features in Machado-Joseph disease. *Am. J. Hum. Genet.* **57**, 54-61.

- Maxam, A.M. & Gilbert, W. (1977) A new method for sequencing DNA. *Proc. Natl. Acad. Sci. USA* **74**, 560-564.

- McColl, K.E.L., Moore, M.R., Thompson, G.G. & Goldberg, A. (1982) Screening for latent acute intermittent porphyria: the value of measuring both leucocyte d-aminolaevulinic acid synthase and erythrocyte uroporphyrinogen-I-synthase activities. *J. Med. Genet.* **19**, 271-276.

- Meyer, C.G., Tannich, E., Herders, J., Henco, K. & Horstmann, R.D. (1991) Direct sequencing of variable HLA gene segments after in vitro amplification and allele separation by temperature-gradient gel electrophoresis. *J. Immunol. Methods* **142**, 251-256.

- Mgone, C.S., Lanyon, W.G., Moore, M.R. & Connor, J.M. (1992) Detection of seven point mutations in the porphobilinogen deaminase gene in patients with acute intermittent porphyria, by direct sequencing of in vitro amplified cDNA. *Hum. Genet.* **90**, 12-16.

- Mgone, C.S., Lanyon, W.G., Moore, M.R., Louie, G.V. & Connor, J.M. (1993) Detection of a high frequency in exon 12 of the porphobilinogen deaminase gene in patients with acute intermittent porphyria. *Hum. Genet.* **92**, 619-622.

- Mgone, C.S., Lanyon, W.G., Moore, M.R., Louie, G.V. & Connor, J.M. (1994) Identification of five novel mutations in the porphobilinogen deaminase gene. *Hum. Mol. Genet.* **3**, 809-811.

- Moore, M.R., McColl, K.E.L., Rimington, C. & Goldberg, A. (1987) in Topics in Haematology. New York: Plenum Publishing Corporation.

- Moore, M.R. & Brodie, M.J. (1990) in Medicine International **81**, 3349-3355.

- Myers, R.M., Lumelski, N., Lerman, L.S. & Maniatis, T. (1985) Detection of single base substitutions in total genomic DNA. *Nature* **313**, 495-498.

- Myers, R.M., Fischer, S.G., Lerman, L.S. & Maniatis, T. (1985a) Nearly all single base substitutions in DNA fragments joined to a GC-clamp can be detected by denaturing gradient gel electrophoresis. *Nucleic Acids Res.* **13**, 3131-3145.

- Myers, R.M., Larin, Z. & Maniatis, T. (1985b) Detection of single base substitutions by ribonuclease cleavage at mismatches in RNA:DNA duplexes. *Science* **230**, 1242-1246.

- Nagamine, C.M., Chan, K. & Lau, Y.F. (1989) A PCR artifact: Generation of heteroduplexes. *Am. J. Hum. Genet.* **45**, 337-339.

- Namba, H., Narahara, K.T., Yokoyama, Y. & Seino, Y. (1991) Assignment of human porphobilinogen deaminase to 11q24.1→q24.2 by in situ hybridization and gene dosage studies. *Cytogenet. Cell Genet.* **57**, 105-108.

- Nordmann, Y., de Verneuil, H., Deybach, J-C., Delfau, M-H. & Grandchamp, B. (1990) Molecular genetics of porphyrias. *Ann. Med.* **22**, 387-391.

- Ong, P.M.L., Lanyon, W.G., Hift, R.J., Halkett, J., Moore, M.R., Mgone, C.S. & Connor, J.M. (1996) Detection of four mutations in six unrelated South African patients with acute intermittent porphyria. *Mol. Cell. Probes* **10**, 57-61.

- Orita, M., Iwahana, H., Kanazawa, H. & Sekiya, T. (1989) Detection of polymorphisms of human DNA by gel electrophoresis as single strand conformation polymorphisms. *Proc. Natl. Acad. Sci. USA* **86**, 2766-2770.

- Orr, H.T., Chung, M.Y., Banfi, S., Kwiatkowski, T.J., Servadio, A., Beaudet, A., McCall, A.E., Duvick, L.A., Ranum, L. & Zoghbi, H. (1993) Expansion of an unstable trinucleotide CAG repeat in spinocerebellar ataxia type-1. *Nature Genet.* **4**, 221-226.

- Padgett, R.A., Grabowski, P.J., Konarska, M.M., Seiler, S. & Sharp, P.A. (1986) Splicing of messenger RNA precursors. *Ann. Rev. Biochem.* **5**, 1119-1150.

- Pakula, A.A. & Sauer, R.T. (1989) Genetic analysis of protein stability and function. *Ann. Rev. Genet.* **23**, 289-310.

- Pierach, C.A., Weimer, M.K., Cardinal, R.A., Bossenmaier, C. & Bloomer, R. (1987) Red blood cell porphobilinogen deaminase in the evaluation of acute intermittent porphyria. *J.A.M.A.* **257**, 60-61.

- Poser, C.M. & Edwards, K. (1978) Transient monoparesis in acute intermittent porphyria. *Arch. Neurol.* **35**, 550.

- Puy, H., Deybach, J-C., Lamoril, J., Robreau, A.M. & Nordmann, Y. (1996) Detection of four novel mutations in the porphobilinogen deaminase gene in French Caucasian patients with acute intermittent porphyria. *Hum. Hered.* **46**, 177-180.

- Raich, N., Romeo, P.H., Dubart, A., Beaupain, D., Cohen-Solal, M. & Goossens, M. (1986) Molecular cloning and complete primary sequence of human erythrocyte porphobilinogen deaminase. *Nucleic Acids Res.* **14**, 5955-5968.

- Ricketts, M.H., Simons, M.J., Parma, J., Mercken, L., Dong, Q. & Vassart, G. (1987) A nonsense mutation causes hereditary goitre in the Afrikaner cattle and unmasks alternative splicing of thyroglobulin transcripts. *Proc. Natl. Acad. Sci. USA* **84**, 3181-3184.

- Roest, P.A.M., Roberts, R.G., Sugino, S., Van Ommen, G.T.B. & den Dunnen, J.T. (1993) Protein truncation test (PTT) for rapid detection of translation terminating mutations. *Hum. Mol. Genet.* **2**, 1719-1721.

- Saiki, R.K., Scharf, S., Faloona, F., Mullis, K.B., Horn, G.T., Erlich, H.A. & Arnheim, N. (1985) Enzymatic amplification of beta-globin genomic sequences and restriction site analysis of sickle cell anaemia. *Science* **230**, 1350-1354.

- Saleeba, J.A., Ramus, S.J. & Cotton, R.G.H. (1992) Complete mutation detection using unlabelled chemical cleavage. *Hum. Mut.* **1**, 63-69.

- Sambrook, J., Fritsch, E.F. & Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual. New York: Cold Spring Harbour Laboratory Press.

- Sanger, F., Nicklen, S. & Coulson, A.R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**, 5463-5467.

- Sawadago, M. & Van Dyke, M.W. (1991) A rapid method for the purification of deprotected oligodeoxynucleotides. *Nucleic Acids Res.* **19**, 674.

- Schreiber, W.E., Fong, F. & Jamani, A. (1994) Frameshift mutations in exons 9 and 10 of the porphobilinogen deaminase gene produce a crossreacting immunological material (CRIM)-negative form of acute intermittent porphyria. *Hum. Genet.* **93**, 552-556.

- Schreiber, W.E., Fong, F., Nassar, B.A. & Jamani, A. (1995) Heteroduplex analysis detects frameshift and point mutations in patients with acute intermittent porphyria. *Hum. Genet.* **96**, 161-166.

- Scobie, G.A., Llewellyn, D.H., Urquhart, A.J., Smyth, S.J., Kalsheker, N.A., Harrison, P.R. & Elder, G.H. (1990) Acute intermittent porphyria caused by a C→T mutation that produces a stop codon in the porphobilinogen deaminase gene. *Hum. Genet.* **85**, 631-634.

- Scobie, G.A., Urquhart, A.J., Elder, G.H., Kalsheker, N.A., Llewellyn, D.H., Smyth, S.J. & Harrison, P.R. (1990a) Linkage disequilibrium between DNA polymorphisms within the porphobilinogen deaminase gene. *Hum. Genet.* **85**, 157-159.

- Sellner, L.N., Coelen, M. & McKenzie, I. (1992) Reverse transcriptase inhibits Taq polymerase activity. *Nucleic Acids Res.* **20**, 1487-1490.

- 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.

- Sharp, P.A. (1987) Splicing of messenger RNA precursors. *Science* **235**, 766-771.

- Smith, C.W.J., Patton, J.G. & Nadal-Ginard, B. (1989) Alternative splicing in the control of gene expression. *Ann. Rev. Genet.* **23**, 527-577.

- Steingrimsdottir, H., Rowley, G., Dorado, G., Cole, J. & Lehmann, A.R. (1992) Mutations which alter splicing in the human hypoxanthine-guanine phosphoribosyltransferase gene. *Nucleic Acids Res.* **20**, 1201-1208.

- Tishler, P.V., Woodward, B., O'Connor, J., Holdbrook, D.A., Seidman, L.J., Hallett, M. & Knighton, D.J. (1985) High prevalence of acute intermittent porphyria in a psychiatric patient population. *Am. J. Psychiatry* **142**, 1430-1436.

- Waldenström, J. (1957) The porphyrias as inborn errors of metabolism. *Am. J. Med.* **22**, 758-773.

- Warren, M.J., Jay, M., Hunt, D.M., Elder, G.H. & Röhl, J.C.G. (1996) The maddening business of King George III and porphyria. *TIBS* **21**, 229-234.

- Watson, J.D., Hopkins, N.H., Roberts, J.W., Steitz, J.A. & Weiner, A.M. (1987) in Molecular Biology of the Gene. pp 632-675. California: The Benjamin/Cummings Publishing Company, Inc.

- Weatherall, D.J. (1991) The New Genetics and Clinical Practice. Oxford University Press.

- Yoo, H-W., Warner, C.A., Chen, C-H. & Desnick, R.J. (1993) Hydroxymethylbilane synthase: Complete genomic sequence and amplifiable polymorphisms in the human gene. *Genomics* **15**, 21-29.