APPLICATION OF DIRECT cDNA SEQUENCING FOR THE CHARACTERISATION OF MOLECULAR PATHOLOGY IN ACUTE INTERMITTENT PORPHYRIA

Charles S Mgone C M.D., M Med (Paed).

THESIS SUBMITTED FOR THE DEGREE OF DOCTOR OF PHILOSOPHY
TO THE UNIVERSITY OF GLASGOW, FACULTY OF MEDICINE

Duncan Guthrie Institute of Medical Genetics
Yorkhill
GLASGOW

December, 1991

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

To Michael, Jane, Charles, Harold and Joyce Mgone

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

Charles Stephen Mgone

ACKNOWLEDGEMENTS.

I would like to extend my gratitude and appreciation to Dr. W.G. Lanyon and Prof. J.M. Connor, for their supervision of the study and for reviewing the manuscript. I would also like to thank Dr. M.R. Moore for his help in collecting the patients, Prof. B. Grandchamp for supplying the primers, the sequence of the porphobilinogen deaminase cDNA and the protocol for the polymerase chain reaction and to Dr. F Milner-White and Prof T. L. Blundell for their help in interpretating the secondary structure alterations likely to result from the detected mutations. I am also grateful to Ms K. Evans, Mrs N. Tariq and Mrs M. Crouch for their help in establishing the lymphoblastoid cell lines, and to the Sir Tayabali Karimjee Education Trust and the Overseas Research Student Award Scheme, for their generous grants. I am also thankful to all those patients and their general practitioners for agreeing to take part in this study and to Drs K.A. Himid and H. Kasimbazi for their help and views. Last but not least I am very thankful to all members of the institute for their help and support in one way or another, and to my family for enduring the three year period of separation, allowing me to pursue my ambition.

C.S.M.

TABLE OF CONTENTS.

CHAPTER THREE.

CHAPTER FOUR.

Figures

Tables

ABBREVIATIONS.

A	adenine (only used as part of a sequence)
AIP	acute intermittent porphyria
ALA	aminolevulinic acid
AMD	amplification and mismatch detection
APS	ammonium persulphate
ARMS	amplification refractory mutation system
ASO	allele specific oligonucleotide
ATP	2'-adenosine 5'-triphosphate
BMD	Becker muscular dystrophy
bp	base pair(s)
Bisacrylamide	*N,N'*-methylenebisacrylamide
C	cytosine
CCM	chemical cleavage of mismatch
Ci	Curie (3.7×10^{10} Becquerel)
cpm	counts per minute
cm	centimetre(s)
cDNA	complementary deoxyribonucleic acid
dATP	2'-deoxyadenosine 5'-triphosphate
dCTP	2'-deoxycytidine 5'-triphosphate
ddNTP	dideoxyribonucleoside triphosphate
DEPC	diethyl pyrocarbonate
DGGE	denaturing gradient gel electrophoresis
dGTP	2'-deoxyguanosine 5'-triphosphate
dH_2O	distilled water
dITP	2'-deoxyinosine 5'-triphosphate
DMD	Duchenne muscular dystrophy

DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DNAase	deoxyribonuclease
dNTP	deoxyribonucleoside triphosphate
DTT	dithiothreitol
dTTP	2'-deoxythymidine 5'-triphosphate
EDTA	ethylenediaminetetraacetic acid
FCS	foetal calf serum
G	guanine
g	gram(me)
g	gravitational (centrifugal) force
h	hour(s)
HPLC	high pressure liquid chromatrography
HPRT	hypoxanthine-guanine phosphoribosyltransferase
kb	kilobase(s)
kDA	kiloDalton
L	litre
lb/sq in	pounds per square inch
LTR	long terminal repeat
M	molarity
min	minute
MOPS	3-(*N*-morpholino)propanesulfonic acid
mg	milligram(me)
ug	microgram(me)
ml	millilitre(s)
ul	microlitre(s)

mM	millimolar
mRNA	messenger ribonucleic acid
NACS	nucleic acids chromatography system
NADH	dihydropteridine reductase
NP-40	Nonidet P-40
nm	nanometre
ng	nanogram(s)
OD	optical density
PAH	phenylalanine hydroxylase
PCR	polymerase chain reaction
PBG	porphobilinogen
PBG-D	porphobilinogen deaminase
PKU	phenylketonuria
PNK	polynucleotide kinase
p mol	pico mole
poly $(A)^+$	polyadenylated
RFLP	restriction fragment length polymorphism
RNA	ribonucleic acid
RNAase	ribonuclease
rRNA	ribosomal ribonucleic acid
rpm	revolutions per minute
R-T PCR	reverse-transcriptase polymerase chain reaction
s	second(s)
SDS	sodium dodecyl sulphate
SSCP	single-strand conformation polymorphism
T	thymidine
TBE	Tris-borate-ethylenediaminetetraacetic acid

tRNA	transfer ribonucleic acid
u	unit(s)
TE	Tris-ethylenediaminetetraacetic acid
TEMED	*N,N,N,N'*-tetramethylethylenediamine
T_m	melting temperature
TMACl	tetramethylammonium chloride
Tris	Tris (hydroxymethyl) aminomethane
U.V.	ultra violet

CHAPTER ONE: SUMMARY.

1. SUMMARY.

Acute intermittent porphyria (AIP) is an autosomal dominant disorder caused by partial deficiency of the enzyme porphobilinogen deaminase (PBG-D). It is heterogeneous and commonly gene carriers of the disorder remain asymptomatic and may not always be diagnosed by conventional biochemical methods. Since detection of gene carriers is of central importance to the management of this condition, an alternative method of diagnosis is essential. Mutations causing acute intermittent porphyria have not however, been fully characterised. In the current study, direct cDNA sequencing of polymerase chain reaction (PCR) amplified templates has been developed and applied for the characterisation of mutations associated with this disorder.
The procedure was developed by amplifying RNA from various sources including human placenta, chorion, lymphocytes and lymphoblastoid cells. The amplification was performed by a technique referred to as reverse-transcriptase polymerase chain reaction (R-T PCR) in which both the first strand cDNA synthesis and the subsequent amplification are performed in the same reaction mixture. Two approaches to the R-T PCR amplification were employed and compared. In the first approach, the first strand cDNA synthesis was carried out with one of the primers complementary to the non-erythroid PBG-D mRNA and in the second, by using oligo$(dT)_{12\text{-}18}$. Both methods were successful and comparable, but the later was preferred because it could be modified in asymmetric PCR to directly produce either the sense or the anti-sense strand.

The PBG-D cDNA synthesised and amplified by the R-T PCR was either directly sequenced as double-stranded (ds) templates or eluted and reamplified by asymmetric PCR to produce single-stranded templates. Alternatively, single-stranded templates were produced directly by 'asymmetric' R-T PCR. Prior to sequencing, the PCR amplified templates were concentrated, desalted and purified to remove excess deoxyribonucleoside triphosphates (dNTPs) and amplification primers. Several purification methods were employed and their efficacy compared. These included, spun-column chromatography, nucleic acid chromatography system (NACS) purification, centrifuge-driven dialysis, geneclean^{TM} purification, gel fractionation and selective precipitation in ammonium acetate and propan-2-ol. Selective precipitation with ammonium

acetate and propan-2-ol was found to be the simplest and most effective method of template purification. In addition it was also inexpensive, reliable and convenient.

Dideoxy sequencing of both double-stranded and single-stranded (ss) templates was performed with either Sequenase T7 DNA polymerase or *Taq* DNA polymerase. Sequencing of the single-stranded templates, especially when produced by asymmetric reamplification of cDNA gave the most consistent and reliable results. For routine sequencing, there was no difference in the performance of the two sequencing enzymes used, although *Taq* DNA polymerase was better than Sequenase T7 DNA polymerase in handling templates with complex secondary structures.

The procedure of direct sequencing was applied on asymetrically amplified templates of thirty patients with acute intermittent porphyria (AIP) and ten normal controls. The diagnosis of acute intermittent porphyria was based on increased excretion of aminolevulinic acid (ALA) and porphobilinogen (PBG) in urine and decreased activity of erythrocyte porphobilinogen deaminase (PBG-D) coupled with a clinical history of one or more acute attacks. The mean erythrocyte porphobilinogen activity in the acute intermittent porphyria patients was 22.3 nmol/h/ml erythrocytes. The normal adult activity range for the enzyme is 25-42 nmol/h/ml erythrocytes in the females and 30-48 in males. After optimisation of the R-T PCR, correct sized products were obtained from the amplification of all samples, indicating absence of any major deletions. Sequencing of these products revealed seven point mutations in twelve patients with acute intermittent porphyria and none in the control subjects. All mutations were due to single base substitutions, four of which were associated with amino acid substitutions and are likely to be the cause of AIP in these individuals. The remaining three were silent mutations without change of amino acid and are therefore regarded as neutral polymorphisms. The detected mutations were Q34K (C_{100}->A) seen in two related individuals, L177R (T_{530}->G) also observed in two unrelated individuals, R167Q (G_{500}->A) and H256N (C_{766}->A) each seen separately in single subjects. The silent mutation L42L (G_{117}->A) was seen in one individual whereas S45S (G_{135}->A) and V202V (G_{606}->T) were seen in two and four individuals respectively. With the exception of the mutation R167Q which has been previously

reported in four other individuals, the rest of the mutations were novel, emphasising the heterogeneity of this condition. Among the four mutations that were associated with amino acid substitutions, three of them (Q34K, R167Q and L177R) occurred in highly conserved domains of the PBG-D cDNA, involving amino acids which are identical in all six species in which the gene has been sequenced. The remaining mutation (H256N), involved a conserved charge. Three of the mutations namely, L177R, H256N and S45S abolish recognition sites for restriction enzymes, *Alu*I, *Ban*I and *Rsa*I respectively, which can be used for tracking of these mutations in affected family members and possibly in the community at large. Prediction of the secondary structure changes likely to be caused by these mutations was performed using the PEPTIDESTRUCTURE and PLOTSTRUCTURE computer programmes. From this it was predicted that among the four delineated mutations, the Q34K was the only mutation that was likely to be associated with a configurational change in the protein structure. This mutation was predicted to lead to the disruption of β-strand. It must, however, be emphasised that three of these mutations affected highly conserved amino acids and one a conserved charge. Moreover, despite the lack of prediction in alteration of the secondary structure, previous expression studies have demonstrated that mutation R167Q is associated with a drastically reduced optimal pH of the porphobilinogen deaminase.

Finally, it was possible to simplify this procedure of direct sequencing of PCR amplified products by doing the reverse-transcription and the subsequent direct sequencing from crude RNA preparations of lymphocytes, thus bypassing the need for elaborate RNA extractions. This modification, greatly saved time and labour and and in addition, was found to be reliable. Furthermore, this procedure of direct sequencing of PCR amplified productcs can be used as a prototype to investigate the molecular pathology of any other single gene disorder as shown in this study by being applied to demonstrate the G551D ($G_{1,784}$->A) mutation in cystic fibrosis.

CHAPTER TWO: INTRODUCTION.

2. INTRODUCTION.

2.1. Molecular pathology of single gene disorders.

Single gene disorders generally arise from two main types of molecular pathology, namely gross aberrations and point mutations. Gross aberrations are caused by insertions, deletions or rearrangements of genes and these may be large enough to be visible under light microscopy if the change involves more than 4,000 kilobases (kb). These aberrations are most commonly observed in tandemly duplicated genes (i.e. two genes with a similar nucleotide sequence on the same chromosomal location) because of non-homologous pairing and crossing over during meiosis (Antonarakis, 1989). In genes like the beta-globin and factor VIII gene that have only one copy per chromosome, deletions or rearrangements are relatively rare and account for 1 to 5 % of mutations (Antonarakis et al., 1985a, Gitschier et 1986; Youssoufian et al., 1987). Deletions were first described in the human globin gene clusters in relation to thalassaemias and hereditary persistence of foetal haemoglobin. In some forms of α-thalassaemia, different-sized deletions remove one α-globin gene and leave one functional gene per haploid genome. These deletions arise from unequal crossing over between homologous pairs of chromosome 16, leaving one α-globin on one of the chromosomes and three on the other (Embury et al., 1980; Antonarakis et al., 1985b; Weatherall, 1985). Deletions have also been described in the dystrophin gene (Monaco et al., 1985; Kunkel, 1986; Read et al., 1988; Chamberlain et al., 1988; Cooke et al ., 1990) where they account for up to 60% of mutations of this gene. These mutations lead to either Duchenne muscular dystrophy (DMD) or its milder allelic form, Becker muscular dystrophy (BMD). Other conditions in which deletions have also been described include; factor IX deficiency (Gianelli et al., 1984), antithrombin III deficiency (Prochovnik et al ., 1983), familial hypercholesterolaemia (Sudhof et el., 1985; Lehman et al., 1986) and Lesch-Nyhan syndrome (Yang et al., 1984; Wilson et al., 1986). Rearrangements like duplication and translocation leading to DMD have also been described in the dystrophin gene (Ray et al., 1985; Cooke et al., 1990). Insertions have been reported in dihydropteridine reductase (NADH) deficiency (Howells

et al., 1990) and hypoxanthine-guanine phosphoribosyltransferase (HPRT) deficiency (Davidson et al., 1991).

Point mutations, on the other hand, are sub-microscopic lesions caused by substitutions, deletions, or insertion of one, or a few nucleotides in critical regions of genes. These mutations are inherited as mendelian traits (McKusick, 1990). Single base substitutions may either be due to transitions or transversions. Transitions refers to those changes where one purine is substituted by another purine or when a pyrimidine is substituted for a pyrimidine. On the other hand, in a transversion, a pyrimidine is substituted by a purine or vice versa. Although point mutations may affect single nucleotides, such changes can still have profound effects on the gene product, either by affecting the transcription of the gene into mRNA or the processing of mRNA during the production of its mature form or even the translation of mRNA into protein. Point mutations may bring about these adverse effects by causing important amino acid substitutions in the protein product (Antonarakis, 1989), by prematurely introducing stop codons leading to the production of truncated proteins (Scobie et al., 1990; Lee and Anvret, 1991), by altering splicing sites with or without subsequent skipping of exons (Grandchamp et al., 1989a; Grandchamp et al., 1989b; Delfau et al., 1990), or by altering regulatory 'boxes' and poly A addition signals (Weatherall, 1965).

The mechanism of the exon-intron junction splicing is not fully known. Primary transcripts are large mRNA precursors which correspond to the entire gene, including exons and introns. Before release these mRNA precursors are processed in the cell where the mRNA corresponding to introns are excised and the exons spliced together. There are at least three different splicing systems involving rRNA, tRNA and mRNA. In the case of the mRNA precursors, the 5' exon-intron junction has a conserved segment of nine nucleotides in which introns tend to begin with GT and end with AG (consensus sequences), obeying the so called GT-AG rule. The nine-nucleotide consensus for the 5' junction extends three residues upstream and six residues downstream from the splice point. The conserved sequence of the 3' splice junction is on the other hand composed of a pyrimidine-rich region of a variable length (at least ten bases long)

followed by a short consensus sequence extending three bases upstream and one base downstream from the splice point (Mount, 1982). The pyrimidine-rich region is devoid of the sequence AG, which is the invariant dinucleotide at the end of the 3' intron-exon junction. Thus the 5' exon-intron junction always has the bases G and T, whereas the 3' junction has A and G. Use of the 5' splice site is abolished by changes in the invariant GT sequence and is decreased by mutations elsewhere in the in the nine conserved bases (Watson et al., 1987). 5' splice mutations, besides abolishing splice sites, may activate new sites (cryptic splice sites). Splice site mutations have been described in phenelalanine hydroxylase (PAH) gene (Dianzani et al., 1991) and the PBG-D gene (Grandchamp et al.,1991a, 1991b, 1991c). In the PBG-D gene, one of these mutations, a G to A transition within the donor splicing site of intron 12 has been observed to lead to the skipping of exon 12 during post-transcriptional maturation of the primary transcripts. As the junction of exon 11 to exon 13 does not disrupt the reading frame, the mutant RNA is translated into a truncated protein missing 40 amino acid residues encoded by exon 12 (Grandchamp et al., 1991c). Splice site mutations leading to activation of cryptic splice sites and hence abnormal proteins have been described in β-thalassaemia and haemoglobin E. These includes G to A transitions at position 110 of the first intron and exon of the β-globin gene respectively (Weatherall, 1985). In the first, the mutation produces a new splice acceptor site which tends to be used more frequently then the original site resulting in the production of mRNA containing intron sequences. This abnormal RNA cannot be used as a template for the β-globin chain synthesis and therefore results in β-thalassaemia. In the second mutation, a G to A substitution in exon 1 of the β-globin gene activates a cryptic splice site which is utilised at a lower rate. This leads to the production of some abnormal mRNA lacking exon sequences and having a premature stop codon resulting in haemoglobin E phenotype. Thus despite of their similarity, these two processing mutations of the β-globin gene lead to different conditions.

Processing defects may also occur when mutations involve other consensus domains like regulatory 'boxes' and poly(A)$^+$ addition signals. Towards the 3' end of DNA sense strands (upstream), genes have consensus sequences called enhancer and promoter regions which are

responsible for the regulation of gene activity. The promoter is involved in the attachment of RNA polymerase to the DNA sense strand during transcription. Many promoters have an element called a TATA box (Goldberg-Hogness box) with the sequence TATAAA, 25-30 bp upstream from the start. Further upstream there are often one or more promoter elements including the sequences CCAAT and GGGCGG. The activity of many promoters is modulated by an enhancer, a separate regulatory element located on the same DNA molecule. This can be situated, either upstream or downstream and may be as far as 1 Kb from the promoter region. At the 5' end of the gene, there is a consensus sequence AATAAA, which signals the addition of the poly(A)$^+$ tail. Mutations of both the TATA box and the poly(A)$^+$ additional signal leading to mRNA processing defects has been described. These for instance, include mutations involving the TATA and CCAAT boxes resulting in β-thalassaemia phenotype (Orkin et al., 1983; Weatherall, 1985). Poly(A)$^+$ addition signal mutations have been described in both α- and β-thalassaemias where an AATAAA to AATAAG mutation of the α-2-globin gene causing α-thalassaemia has been reported. In this mutation, presumably the A to G transition at the last position of the consensus poly(A)$^+$ additional signal interferes with mRNA polyadenalation. This will therefore, prevent the transportation of the mRNA into the cytoplasm (Higgs et al., 1983). A similar mutation, AATAAA to AACAAA in the β-globin gene has been reported as a cause of β-thalassaemia (Weatherall, 1985). Finally, there is yet another type of mRNA processing mutation whereby the abolishing of the termination codon leads to the production of an abnormal protein. An example of this is the haemoglobin Constant Spring, which is a cause of α-thalassaemia phenotype (Weatherall, 1985). In this mutation, a T to C transition in the α-globin gene, converts the termination codon TAA (UAA) to CAA encoding for glutamine. This results in the production of an unstable α-globin chain variant with 31 additional amino acids residues at the C-terminal end.

On the basis of the amino acid or codon changes, point mutations have also been classified into missense and nonsense mutations, whereby missense mutations include those in which one amino acid is substituted for another and nonsense mutations refer to changes that lead to a

premature introduction of a stop codon. Given that there are only three stop codons, it therefore follows that most of the point mutations are likely to be of missense rather than the nonsense type. As any new protein arising from a missense mutation contains only a single amino acid substitution, it often possesses some of the biological activities of the original protein. Quite often, missense proteins fail to function only at a higher than normal temperature and are therefore known as temperature-sensitive mutations. Missense mutations have been described in many conditions including, β-thalassaemia (Ristaldi et al., 1989), sickle cell anaemia (Chang and Kan, 1982), α-1-antitrypsin deficiency (Graham et al., 1990), factor IX deficiency (Montadon et al., 1990), cystic fibrosis (Cutting et al., 1990), factor VIII deficiency (Youssoufian et al., 1988; Higuchi et al., 1990), phenylketonuria (Konecki et al., 1991), AIP (Lee et al., 1990, Delfau et al., 1990, 1991) and HPRT deficiency (Gibbs et al., 1989; Davidson et al., 1991).

Due to degeneracy of the genetic code, up to a quarter of point mutations lead to no change in the amino acid (Connor and Fergusson-Smith, 1991). Such changes are referred to as silent or same-sense mutations. Usually these mutations lead to no pathology. However, in some highly expressed genes, there are certain codons which may be preferred to others in the translation of proteins and that such codons even when changed to another same-sense codon, result in inefficient production of the encoded amino acids. Silent mutations may also lead to alterations of recognition sites for restriction enzymes, making it possible to directly detect mutations. Alternatively, such silent mutations may create restriction fragment length polymorphism (RFLPs) which may also be used to detect mutations through linkage analysis. Of particular interest to this study, such silent mutations have also been described in the PBG-D gene where in some cases they have been useful in the tracking of mutations of this gene (Grandchamp et al., 1987; Gu, 1991; Lee, 1991a; Picat et al., 1991)

On the other hand, as nonsense mutations lead to premature termination of the protein product, the length of the final product and theoretically the severity of the deficiency will therefore depend on the site of the mutation. Nevertheless, most prematurely terminated proteins (irrespective of the size), lead to reduced or no biological activity. Such nonsense mutations have

been described in porphobilinogen deaminase gene (Scobie et al., 1990; Delfau et al., 1990, Delfau et al., 1991; Lee and Anvret 1991a), factor VIII gene (Youssofian et al., 1886 and 1988; Higuchi et al., 1990), α- and β thalassaemias (Weatherall, 1985; Wong et al., 1987), factor IX gene (Montandon et al., 1990), cystic fibrosis gene (Cutting et al., 1990), HPRT gene (McKeran et al., 1975; Davidson et al., 1991), and PAH gene (Wang et al., 1989, 1890, Svensson et al., 1990). Nonsense mutations may also arise from insertions or deletions of one or a few bases leading to alteration of the open reading frame, the so called frame-shift mutations. This may also lead to a premature introduction of a stop codon. These insertions or deletions are thought to arise from displacement of bases from either the growing strand in the case of insertions, or from the template strand in the case of deletions. Such mutations tend to occur particularly at sites with runs of identical bases in a DNA fragment, where the displacement can be stabilised by normal base pairing beyond the unpaired base. When the miscopied DNA strand acts as a template, the insertion or deletion is usually copied accurately, thus fixing the mutation. Since frame-shift mutations alter the open reading frame, they tend to cause complete disruption of protein synthesis (Ghosal and Saedler 1978). Examples of these mutations have been described in β-thalassaemia (Orkin et al., 1881; Wong et al, 1990), factor VIII gene (Higuchi et al., 1990) and cystic fibrosis (White et al., 1990; Ianuzzi et al., 1991) and PBG-D gene (Nordmann et al., 1991). In the case of cystic fibrosis, two mutations have been identified in exon 7, one caused by a two-nucleotide insertion at position 1154 and the other caused by a one-nucleotide deletion at position 1213, both causing frame-shifts. This leads to the premature insertion of termination codons at residues 369 and 368 respectively (Ianuzzi, 1991). A frame-shift mutation of the cystic fibrosis gene caused by a two-nucleotide insertion at exon 13, leading to the introduction of a stop codon at residue 821 has also been described (White). Therefore, in many conditions, like α-thalassaemia, β-thallassaemia, pheylketonuria, haemophilia A, haemophilia B, Lesch-Nyhan syndrome and AIP, the same phenotype may be caused by different mutations of the same gene.

Although error rates for different bases do vary by a hundred fold, on average a nucleotide is likely to be changed by mistake about once for every 10^9 times it is replicated. These mutations

do not occur at the same rate in all sites. There are some areas in certain genes that are more likely to under go mutations and are generally referred to as 'hot spots' (Benzer, 1961). This often occurs at CpG dinucleotides, which in higher animals seem to play a major role in mutagenesis (Mandel and Chambon, 1979; Bird, 1980; Nussinov, 1981; Barker and White, 1982; Barker et al., 1984; Youssoufian et al., 1988; Delfau et al, 1990). Evidence in support of this includes the underrepresentation of this dinucleotide in vertebrate genomes (Nussinov, 1981), a higher frequency of polymorphism observed with restriction enzymes that contain CpG in their recognition sequences (Barker et al., 1984) and the observation of many mutations in different disorders involving this dinucleotide. Mutations that involve CpG dinucleotide have been for instance, described in α_1-antitrypsin gene (Kidd et al., 1983), adenosine deaminase (Bonthron et al., 1989), insulin (Shibasaki et al., 1985), antithrombin III (Duchange et al., 1986), factor IX deficiency (Bentley et al., 1886), protein C (Romeo et al., 1987), factor VIII deficiency (Youssoufian et al., 1986 1988), ornithine transcarbamylase (Lee and Nassbaum, 1989) and porphobilinogen deaminase gene (Grandchamp et al., 1989b; Delfau et al., 1990; Lee 1991b).

2.2. Methods used in studying molecular pathology of monogenic disorders.

In recent years many techniques have been developed for the manipulation and study of genes, leading to tremendous progress in the understanding of structure and function of human genes. These achievements have led to the understanding of the biochemical mechanisms and basic diagnosis at the DNA level of many diseases, especially those due to single gene defects (Antonarakis, 1989). Among these new techniques, the polymerase chain reaction (PCR), has undoubtly contributed most to the attainment of these achievements (Vosberg, 1989).

2.2.1. Polymerase chain reaction.

Polymerase chain reaction is an enzyme-catalyzed *in vitro* method of selective multiplication of DNA or RNA segments of interest, starting from small amounts of material. This is done by flanking the segment of interest (template) with two synthetic oligonucleotides commonly referred to as amplification primers, in such a way that their 3' ends point to each other when hybridized to the corresponding strands of the template. Through an enzyme controlled polymerisation, the amplification primers are then made to extend towards each other with repetition of this process many times, usually between 30 to 40 cycles as illustrated in figure 1. The essential steps in each cycle are, thermal denaturation of double stranded target molecules followed by annealing of the amplification primers to each strand and finally the polymerase controlled DNA synthesis. Since both strands of a given target DNA are used as templates, the number of target sequences multiplies exponentially ending with large amounts of the targeted DNA segment as defined by the 5' ends of the amplification primers (White et al., 1989). This feature of PCR, does not only account for its sensitivity but its specificity as well, since the selective amplification of the targeted region conversely reduces the background of unwanted sequences.

Moreover, PCR can also be used to amplify RNA transcripts. This is made possible after converting mRNA to first strand cDNA with reverse transcriptase (Saiki et al., 1988) as illustrated diagrammatically in figure 2. This can be achieved by using either oligo$(dT)_{12\text{-}18}$, random hexamers or by using a forward (upstream) primer, the primer that is complementary to the original mRNA.

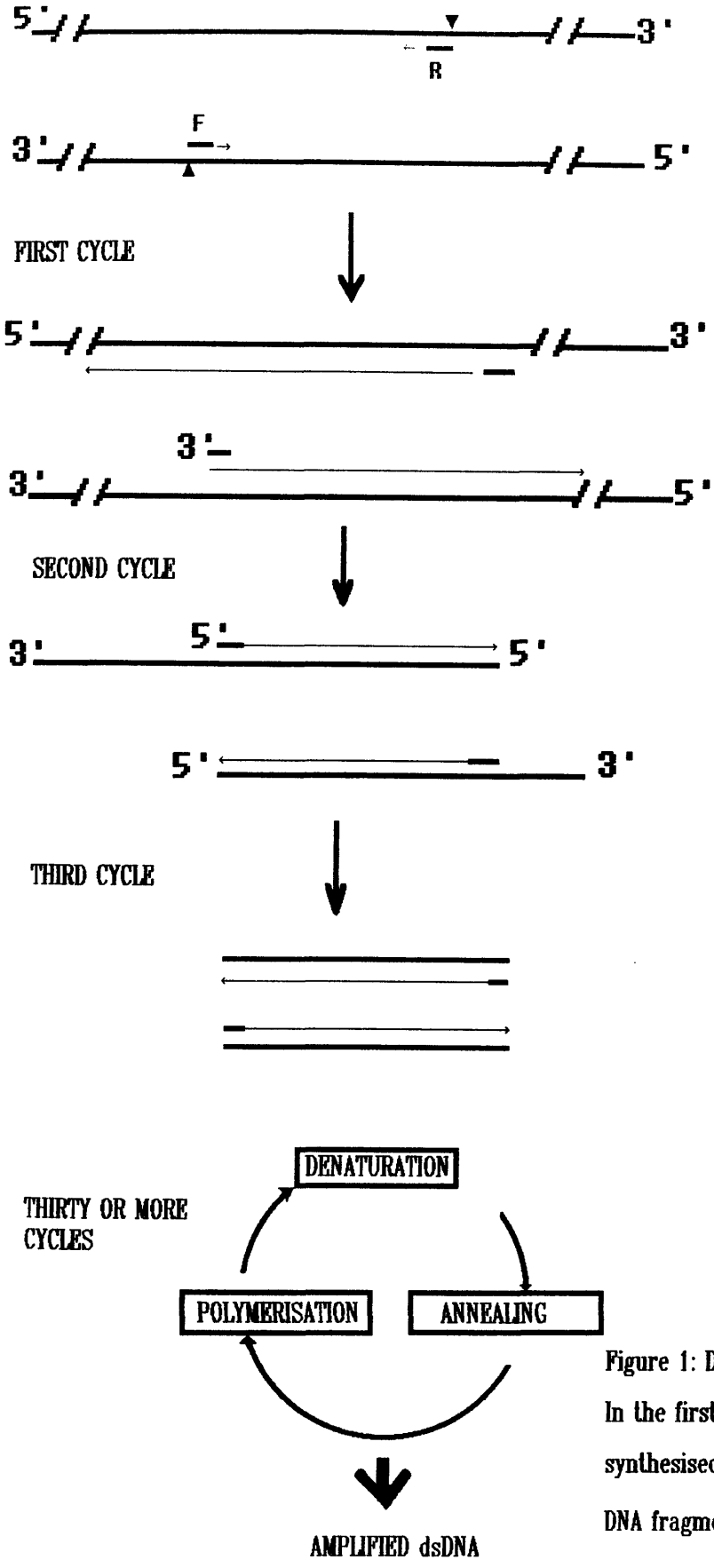


Figure 1: Diagramatic representation of a standard PCR. In the first cycle, a fragment larger than the targeted DNA is synthesised. However, after two amplification cycles targeted DNA fragment is produced and eventually it becomes the predominant product. F is a forward, and R, a reverse primer.

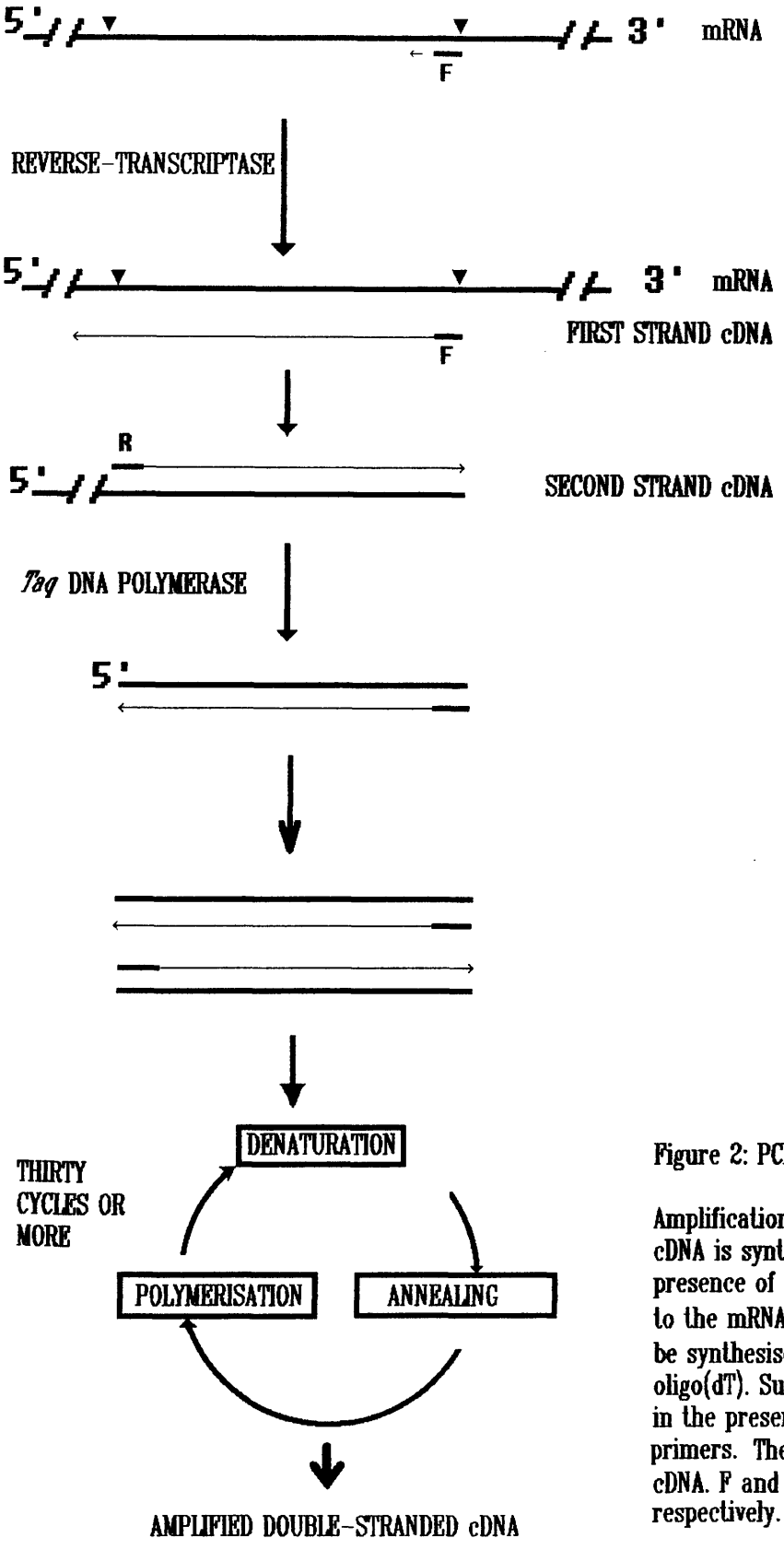


Figure 2: PCR amplification of RNA transcripts.

Amplification of mRNA is performed after first strand cDNA is synthesised via reverse-transcription in the presence of a forward (upstream) primer, complementary to the mRNA. Aternatively, the first strand cDNA can be synthesised by using either random hexamers or oligo(dT). Subsequent PCR amplification is performed in the presence of both, forward and reverse (downstream) primers. The latter is complementary to the first strand cDNA. F and R represents forward and reverse primers respectively.

Polymerase chain reaction is sensitive enough to enable amplification of single DNA molecule and is now routinely used in the extraction of single-copy genes from complex genomic mixtures (Kim and Smithies, 1988, Li et al., 1988). For example, Saiki and co-workers (1988), were able to amplify certain targeted sequences of the β-globin gene with a homozygous deletion contained in a 10^{-6} dilution of genomic DNA. This suggests that PCR is able to amplify target sequences that are present only once in 10^5 to 10^6 cells. This high sensitivity has been shown in the handling of mRNA as well (Harbath and Vosberg, 1988), where indeed, amplification has been possible from the RNA content of a single cell (Rappolee et al., 1988). It is the combination of this sensitivity and specificity that makes PCR a very versatile procedure. The technique can be performed on virtually any DNA or RNA-containing material and has for instance been successfully carried out from individual hair roots (Higuchi et al., 1988a), single sperm (Li et al., 1988), buccal cells, (Lench et al., 1988) and epithelial cells in urine (Gasparini et al., 1989). This technique can also be performed on partially degraded DNA (Paabo et al., 1988) and on paraffin-embedded tissues (Impraim et al., 1987; Shibata et al., 1988; Jackson et al 1989).

The discovery that restriction site sequences could easily be introduced into DNA fragments through PCR has further increased the scope of applications of this procedure. Scharf et al., (1986) found that this could be done by merely attaching the restriction sequences to the 5' end of the amplification primers. Despite these sequences being mismatched to the DNA template, they seem to interfere with neither the efficiency nor the specificity of the PCR. As strands initiated by these primers are copied including their added on 5' tails, the restriction sites sequences become incorporated in the PCR products. In this way fragments of up to 45 bases can be added onto DNA fragments, as for instance the addition of a G+C rich segment to augment mutation detection in denaturing gradient gel electrophoresis (Sheffield et al., 1989). The introduction of restriction sites onto PCR products facilitates their further handling such as in cloning (Scharf et al., 1986). Besides addition of restriction sites, these 5' tails can be used to incorporate a phage promoter necessary for reverse-transcriptase mediated RNA or DNA sequencing (Stoflet et al., 1988; Sarkar and Sommer 1988; Sarkar and Sommer 1989). Also

based on this principle, several other methods of modifying DNA fragments have been described. At its simplest the principle can be applied to specifically label a particular strand or both strands of PCR products with a radioactive, biotin or fluorescent tag (Lo et al., 1988; Gyllensten, 1989; Hutman et al., 1989). It can also be used to introduce defined mutations at specific sites in DNA segments (Higuchi et al., 1988; Vallete et al., 1989) and in DNA recombination procedures (Horton et al., 1989).

This diversity in protocols and applications of PCR entails the need for different amplification conditions. Quite often there is a need for optimisation of the PCR conditions before they can be applied routinely in a particular protocol (Saiki, 1989a; Innis et al., 1990). Variables which need to be optimised include the concentrations of templates, magnesium ions, deoxynucleoside triphosphates, enzyme and amplification primers. Specificity and efficiency of the amplification is also influenced by the cyclic thermal profile and the number of cycles. Failure of optimisation may lead to poor or no amplfication of the desired product as well as non-specific amplification. At times, the amplification of these non-specific products may be so efficient as to become equally or more predominant than the desired product (Saiki et al., 1988; Feldman et al., 1988; Haqqi et et al., 1988; Anwar et al., 1990). Lack of optimisation may also lead to misincorporation of bases (Innis and Gelfand, 1989). Another problem associated with PCR is the presence of false positive results. This is linked to the high sensitivity of the procedure, since the commonest source of these false positive reactions tend to be contaminations carried over from previous positive amplifications (Li et al., 1988; Sarkar and Sommer, 1990).

Currently, there are several PCR based methods for direct detection of nucleotide variations in genetic conditions. These include restriction fragment length polymorphism analysis (Saiki et al., 1985; Kogan et al., 1987; Lee et al., 1990), single-stranded conformation polymorphism analysis (Orita et. al., 1989a; Orita et al., 1989b), denaturing gradient gel electrophoresis (Cariello et al., 1988; Traystman et al., 1990), RNase cleavage analysis (Myers et al., 1988) chemical cleavage of mismatch detection (Cotton et al., 1988; Grompe et al., 1991), allele specific oligonucleotide hybridisation (Embury et al., 1987; Lo et al., 1988), amplification refractory mutation system

analysis (Newton et al., 1989) and direct genomic sequencing (Wrishchnik et al., 1987; Wong et al., 1987; Newton et al., 1988).

2.2.2. Restriction fragment length polymorphism (RFLP) analysis.

Until recently, the usual method for the identification of mutations through linkage analysis required amounts of undegraded DNA in the order of 5-10 ug, to be cleaved, gel electrophoresed, Southern blotted and the RFLP haplotypes probed with radioactive DNA. Using PCR, however, one can now amplify the region of interest containing the polymorphic site starting with genomic material of 1 ng or less. The presence of the diagnostic site is then directly demonstrated by cleavage or absence of cleavage of the amplified DNA fragment. Heterozygosity can be identified by the demonstration of both alleles (Feldman et al., 1988). Furthermore, PCR being an *in vitro* synthetic process, gives rise to a product that does not contain 5-methylcytosine residue and cannot therefore reflect any methylation pattern present in the original template. This has led to the identification of new RFLPs, which had previously gone unnoticed because of inhibition of restriction enzyme cleavage by the methylation of cytosine. This has, for instance been used in the identification of *Hha*I and *Apa*I RFLPs in the factor IX and PBG-D genes respectively (Winship et al., 1989; Picat, et al., 1991). Direct detection of mutations through demonstration of RFLPs has for example, been applied in the diagnoses of haemophilia A (Kogan et al. 1987) and sickle cell anaemia (Saiki et al.,1985). Direct RFLP analysis however requires prior knowledge of the sequence of the DNA segment of interest. Unfortunately only 5 to 10 percent of point mutations can be directly detected by this method (Antonarakis 1989).

Where direct detection is not possible, linkage analysis can still be very valuable in the tracking of mutations. PCR based RFLP linkage studies have been, for example used in prenatal diagnosis of α-1-antitrypsin deficiency (Abbot et al., 1988) and in the diagnosis of haemophilia B (Reiss et al., 1990), cystic fibrosis (Northrup et al., 1989; Rosenbloom et al., 1989; Huth et al., 1989) as well as acute intermittent porphyria (Lee, 1991a). Linkage studies can benefit from the high sensitivity of the PCR technology by the fact that it can be used to process minute amounts or

even partially degraded material of key family members who may otherwise be unavailable for analysis because of death or other reasons. This is due to the fact that PCR can still be performed on their preserved tissues when available, as amplification can be successfully carried out on formalin-fixed or paraffin-embedded tissues (Impraim et al., 1987) and on Guthrie spots (Williams et al., 1988).

Nevertheless, it being an indirect method, RFLP linkage analysis will always have the disadvantage of depending on key family members whose specimens may at times be unavailable to the investigator. Other problems associated with linkage analysis include recombination, non-paternity and non-informative key family members. In RFLP linkage analysis, key family members who may include, parents and some of their relatives, grandparents or previously born children, need to be studied to establish linkage and phase before results can be useful for the subsequent detection of mutations. In some cases both parents may have the same RFLP and make it impossible to establish linkage with the sought mutation. Such parents are said to be non-informative. In seeking to establish autosomal linkage, the ideal mating is a mating where one partner is heterozygous for both the disease and the RFLP and the other is homozygous for the RFLP and the normal allele (a double backcross mating). In addition to this, for an RFLP to be useful, it must be tightly linked to the mutant gene to avoid problems due to recombination. Besides, with this method one cannot detect mutations if present in less than 5% of the studied cells, as may occur in mosaicism (Haliassos et al., 1989).

PCR technology has simplified and increased the sensitivity of RFLP analysis for both direct and indirect detection of genetic variation, but the procedure still retains some of the inherent limitations associated with linkage analysis. Furthermore, there are many single base mutations that do not alter any restriction enyzme sites and therefore, are not amenable to direct detection by this method. Detection of these mutations will require alternative methods as described in subsequent sections.

2.2.3. Single-stranded conformation polymorphism (SSCP) analysis.

The electrophoretic mobility of single-stranded nucleic acids depends not only on size but also on their sequences. Under non-denaturing conditions, ssDNA may have a folded conformation that is stabilized by intrastrand interaction. Any base change can interrupt this conformity and alter mobility which can be detected by comparing the mutant sample along with the wild-type as a control (Orita et al., 1989a; Orita et al., 1989b, Labrune et al., 1991). The sequences to be examined are amplified and labelled simultaneously by PCR using either labelled primers or a labelled nucleotide. In a typical PCR, as up to 10% of the amplifying primers and deoxynucleotides as substrates are incorporated in the amplification products, the efficiency of labelling of the targeted molecules is very high compared to that in RFLP detection by Southern blotting. The labelled PCR products are then denatured and resolved in polyacrylamide gels containing 10% glycerol for the SSCP analysis. This is based on the fact that, the conformation of a single-stranded nucleic acid is presumably determined by the balance between thermal fluctuation and weak local stabilizing forces such as short intrastrand base pairings and base stacking. Therefore, changes in the gel conditions, such as the running temperature or presence of glycerol will be liable to cause conformational changes and will be detected as an electrophoretic mobility shift. This effect of sequence change on electrophoretic mobility is, however, unpredictable and in some cases may not be significant enough so as to be detectable (Orita et al., 1989b; Labrune et al., 1991).

Although SSCP analysis is faster than Southern blotting, it still has some limitations. These include the fact that as a prerequisite, sequences of the targeted DNA need to be known, both for the construction of amplification primers and the preparation of suitable templates. This is necessary, especially for long DNA segments because of a limitation in the size that can be analysed by this method (Orita et al., 1989). SSCP analysis is most suitable for the screening of DNA fragments of less than 200 base pairs (bp) in length. (Orita et al., 1989b). Moreover SSCP analysis can only suggest the presence of a nucleotide variation, but not reveal its exact nature or location. Following the detection of such changes DNA sequencing

needs to be done to determine the exact nature of the pathology. Therefore, at its best, SSCP is suitable only in the screening for nucleotide variations including polymorphisms and to some extent for the tracking of known mutations within family members.

2.2.4. Denaturing gradient gel electrophoresis (DGGE).

DGGE allows the separation of DNA molecules differing by single base changes (Fischer and Lerman 1983; Myers et al., 1985a). This separation is based on the fact that when in solution, DNA molecules melt in discrete segments called melting domains, which may vary from 25 base pairs (bp) up to several hundred bp. The melting temperature (T_m) of each melting domain is highly dependent on its nucleotide sequence (Myers et al., 1989). In the DGGE system, DNA fragments are separated according to their melting properties by electrophoresis through a linearly increasing gradient of denaturants. When a DNA fragment is run through such a system it remains double stranded until it reaches the concentration of denaturants equivalent to a T_m that causes the lower-temperature melting domains of the fragment to melt. When this occurs, the molecule begins to branch, sharply decreasing the mobility of the DNA fragment in the gel and eventually separating the two segments differing in as little as a single nucleotide (Sheffield et al., 1989). Strand separation will, however, not occur in DNA fragments located in the highest melting domain, because of the loss of sequence-dependent migration caused by the completion of strand separation (Myers et al., 1985a). Because of this limitation, it has been estimated that on average the method is able to detect only 50% of the possible single base changes in fragments ranging from 50 bp to several hundred bp in size (Myers et al., 1985b). To circumvent this, a G+C rich segment, commonly referred to as a GC-clamp can be attached to a segment that melts in two domains (Myers et al., 1985b; 1985c). This G+C segment, consisting of around 40 bases, is attached to the 5' end of one of the primers used in PCR for the amplification of the targeted segment. The PCR based DGGE system with GC-clamps, not only improves the detection rate of mutations but also simplifies the method considerably, obviating the necessity for radioactive labelling. Prior to this improvement, the standard DGGE required 5-10 ug of

genomic DNA for each analysis, whereas the PCR based DGGE allows analysis to be done on samples with less than 5 ng of genomic DNA. Moreover, PCR based DGGE enables results to be examined directly by ethidium bromide staining, thus increasing the speed of detection (Sheffield et al., 1989).

Although the method can be used to detect mutations in DNA segments of up to 1000 bases long, it is most suitable for screening fragments of around 100-500 bases. The limitation in the size of DNA fragments that can be screened by DGGE is due to poor resolution and presence of unpredictable secondary structures in long DNA fragments. Further to this, DNA fragments longer than a few hundred base pairs travel very slowly in polyacrylamide gels so that impractical times are required to resolve them (Sheffield et al., 1987). Another and probably more serious problem associated with this technique is the presence of spurious PCR products, since successful detection of mutations in amplified DNA fragments by DGGE analysis requires the presence of single DNA species. In some cases however, different molecules other than the targeted DNA are produced and may even be in excess of the desired product. In such cases, the additional fragments will interfere with the analysis. Moreover, misincorporation may occur during the PCR, producing extra background bands in the DGGE. Although this varies from template to template and depends on the PCR conditions, it has been estimated to occur at the rate of 0.25% in a 30 cycle PCR using *Thermus aquaticus* (Taq) DNA Polymerase (Saiki et al., 1988a). It must be mentioned that the synthesis of these long primers with GC-clamps adds to the cost of this technique.

Like the previously mentioned techniques DGGE analysis can neither reveal the type of the mutation nor its exact location. Therefore unless solely used for linkage analysis it needs to be complemented with DNA sequencing (Cariello et al., 1988; Myers et al.,1989).

2.2.5. RNase cleavage analysis.

Single base changes in DNA can be detected by cleavage of mismatches in RNA:DNA duplexes with the enzyme RNase (Myers et al.,1985d; Winter et al., 1985; Gibbs and Caskey 1987). This is done by mixing a uniformly labelled ssRNA probe with double-stranded test DNA. The ssRNA probes are synthesized as run-off transcripts from cloned DNA templates while the source of test dsDNA can either be from genomic, cloned or PCR amplified products. The RNA:DNA mixture is heated to melt the DNA strands. This is then followed by an annealing reaction whereby the labelled RNA probe is allowed to anneal with its complementary strand in the test DNA fragment forming an RNA:DNA duplex. The presence of any base changes in the test DNA will cause base mismatches in the RNA:DNA duplex. The annealed mixture is then treated with the enzyme RNase A to cleave the RNA strand at the mismatched sites. After the cleavage reaction, the duplex is treated with denaturants to separate the strands and then run on a denaturing gel to separate the RNA fragments according to their sizes and visualized by radioautography. This procedure can efficently detect 30 to 40% of all possible mismatches in the RNA:DNA duplexes. The detection rate can however, be improved to detect up to 60 to 70% of the possible base changes, by testing a DNA fragment with each of its two corresponding labelled RNA probes in separate cleavage reactions (Meyer et al., 1985d). This method is most suitable for screening DNA fragments between 100 and 1000 bp long. Fragments longer than 1000 bp are difficult to screen because of random cleavages which occur even in perfectly matched bases and may be numerous enough to interfere with the results. Moreover, analysis of such long RNA fragments requires the use of denaturing agarose gels (Meyers et al., 1988). Therefore, the most obvious limitation of this method is that it cannot detect all possible mismatches. Furthermore, the method requires extra cloning for probe production and that the screening is limited to DNA fragments of up to 1000 bp only.

2.2.6. Chemical cleavage of mismatch (CCM) detection.

CCM, also known as amplification and mismatch detection (AMD), had been developed to screen for point mutations but can also be used for the detection of deletions and insertions that are otherwise too small to be demonstrated by ordinary gel electrophoretic techniques. The method is based on the formation of heteroduplexes between wild-type and mutant DNA or RNA molecules. Any base changes in the mutant molecules will create mismatches in these heteroduplexes. When chemically modified, the mismatched bases become sensitive to cleavage with piperidine and their positions can then be located by running the samples in denaturing polyacrylamide gels (Cotton et al., 1988; Cotton and Campbell, 1989; Montandon et al., 1989). In the method, reference DNA probe is mixed with excess test DNA or RNA, and the mixture heated to cause strand separation, then cooled to allow reannealing and hence heteroduplex formation in which mismatched or unmatched base pairs occupy the mutation site. The probe is then modified at the mismatched C or T base by reacting with hydroxylamine or osmium tetroxide respectively and subsequently cleaved by piperidine treatment. The samples are finally run on denaturing polyacrylamide gels similar to sequencing gels, to locate the site of cleavage and hence the mutation (Cotton, 1989). As mismatched A or G bases are transposed to mismatched T and C respectively, they should be detectable by the probe of the opposite strand. Moreover, matched bases adjacent or close to mismatched or unmatched bases become reactive by transmission of the distortion, thus indirectly allowing mutation detection (Cotton et al., 1988; Cotton 1989). This also enables detection of insertions (Cotton 1989). Deletions are also detectable because of the reactivity of unmatched C and T bases. Therefore, one of the strong points of this method, is its potential ability to detect all classes of point mutations. (Grompe et al., 1989).

This PCR-based technique can be performed by using either uniformly labelled or end labelled probes. Either of these can be used when the investigated DNA or RNA fragment is expected to have a single mutation in the whole region covered by the probe. However, in the case where multiple mutations are expected, it is advisable to use end labelled probes as they will generate a

single specific band for each reactive C or T base. CCM analysis is therefore useful in screening for mutations by not only demonstrating the site but also the base changes. The method has been used successfuly in scanning DNA segments of up to 1200 bp, localizing mutations to within 30 bp (Grompe et al., 1989). Therefore, unlike all other methods mentioned above, CCM analysis is able to predict the site and nature of mutations. In addition, it is not affected by misincorporation of bases due to the infidelity of *Taq* DNA Polymerase or reverse-transcriptase used in PCR or cDNA synthesis, as both the probe and test DNA would represent a population of PCR products in which alterations would be randomly distributed through out the sequence (Grompe et al., 1989). It should however, be emphasised that CCM analysis is merely a screening method developed to circumvent the need to sequence large fragments of DNA in search of unknown mutations. With this technique it is therefore, possible to limit DNA sequencing to the regions around the mismatched sites with the added high degree of confidence that the detected changes are genuine rather than artefacts. Once a mutation has been detected and characterized by DNA sequencing, subsequent detection in individuals, families or populations should be carried out using other simpler mutation detection methods such as oligonucleotide hybridisation (Grompe etal., 1989; Cotton 1989). The main disadvantage of this procedure is that it requires the use of very toxic chemicals and a complex protocol.

2.2.7. Allele specific oligonucleotide hybridisation.

In this method, oligonucleotides specifically constructed to recognize wild-type or mutant alleles are labelled with a radioisotope and used as probes. Under very stringent conditions, such probes will only hybridise to their perfectly homologous sequences and not to those that vary even by a single nucleotide residue. Thus a normal gene can be detected using the wild-type probe and the mutant gene with the probe containing the same mutation (Studencki and Wallace., 1984). In practice, for the routine analysis "dot-blot" filter hybridisation is used and detection is done by radioactive probes (Saiki et al., 1986), though biotinylated and enzyme-labelled probes can be used as well (Bugawan et al., 1988; Saiki et al., 1988b; Lo et al., 1988). This method also allows simultaneous amplification and analysis of multiple polymorphic sites by using ASO probes of identical length and tetramethylammonium chloride (TMACl). When washed in 3 M TMACl, perfectly matched oligonucleotides of the same length tend to dissociate from the hybridised test DNA sample at the same temperature thus leaving behind the mismatched probes (Wood et al., 1985; Kogan et al.,1987). Alternatively a "reverse dot-blot" procedure can be used in which the oligonucleotide probe is immobilised on a membrane and hybridised to labelled PCR products (Saiki et al., 1989b). In this method a panel of different probes can be used to screen simultanously for several mutations in a given PCR product.

There are, however, some limitations associated with ASO hybridisation analysis. Firstly, as a prerequisite for the oligonucleotide hybridisation analysis, the mutations to be studied need to have been precisely defined previously. Secondly, in many autosomal dominant disorders the mutations tend to be unique to each family thus limiting the usefulness of each probe to a particular family.

2.2.8. Amplification refractory mutation system (ARMS) analysis.

The amplification refractory mutation system analysis has been developed for the detection of known mutations in genomic DNA. The method entails PCR amplification of the target DNA with a normal or mutant primer and a common end primer. The normal and the mutant primer differ at their 3' ends by only a single base which corresponds to the normal and mutant alleles (Newton et al., 1989). The technique is based on the concept that template amplification through PCR in which one of the amplification primers is mismatched at the 3' end is not possible. This is attributed to the lack of a 3' exonucleolytic proof-reading activity of *Taq* DNA polymerase (Tindall and Kunkel, 1988). The technique is performed by amplifying a test DNA with a mutant amplification primer and a common primer. As a control, internal primers are added in the same reaction. Similarly, another set of reaction is performed using normal and common primers. In the same way as before, as a control reaction, internal primers are also added to this reaction. After PCR amplifications, the products are size fractionated in agarose gels and visualised after staining with ethidium bromide. With a normal sample, amplification will only be positive in the reaction with the normal and the common primer as well as with the control internal primers but not in the reaction containing the mutant primer. On the other hand, both the mutant and the normal primers when used with the common end primer will show positive amplifications of samples which are heterozygous for the mutation being investigated, whereas no amplification will be seen in homozygous mutant samples amplified with normal primers. This technique is therefore, rapid, non-isotopic and can distinguish heterozygotes from homozygotes. The method is, however, limited to the investigation of mutations that have already been characterised or those that are linked to a characterised polymorphism. As a means of direct detection of mutations, this technique is therefore useful only in the tracking of known mutations within family members in much the same way as ASO.

2.2.9. Direct genomic sequencing.

Almost without exception, the complete characterization of any mutation depends on sequencing, as this will determine the exact variation and its position in the gene (Cotton, 1989). The two most popular techniques of genomic sequencing are based on the chemical method developed by Maxam and Gilbert (1977) and the enzymatic method of Sanger et al., (1987). In the Maxam and Gilbert method, the DNA is cleaved into fragments which are then radioactively labelled at one end, and subsequently divided up into four batches, each of which is treated differently by chemicals to modify a particular base or bases resulting in very small DNA fragments, some of which will be end-labelled. The fragments are then resolved in denaturing polyacrylamide gels resulting in base specific bands on autoradiography. The Sanger method, on the other hand is an *in vitro* DNA synthesis in which the growing chain of the DNA synthesised is stopped randomly at frequent points leading to numerous fragments. The DNA synthesis starts at one site where the sequencing primer anneals to the template and is facilitated by DNA polymerases in the presence of deoxynucleoside triphosphates (dNTPs). The synthesis is terminated by the incorporation of dideoxynucleoside triphosphates (ddNTPs) present in the mixture. Dideoxynucleoside triphosphates bring about this termination because they lack the 3'-hydroxyl residue necessary for chain elongation. With the right mixtures of dNTPs and each of the ddNTPs, the polymerisation will terminate at each point of ddDNTP incorporation, thus giving full sequence information. To enable interpretation of the results a radioactively labelled nucleoside is included in the synthesis, so that labelled chains of various length can be visualized by autoradiography following separation in denaturing polyacrylamide gels. Of the two methods, Sanger's method popularly known as dideoxy sequencing is the most widely used. This method is diagrammatically represented in figure 3, using sequencing of the PBG-D cDNA to illustrate the principles involved. The illustration represents an actual sequencing of the PBG-D cDNA anti-sense strand of a control subject in whom the junctional region around exons 12 and 13 was sequenced with primer R4.

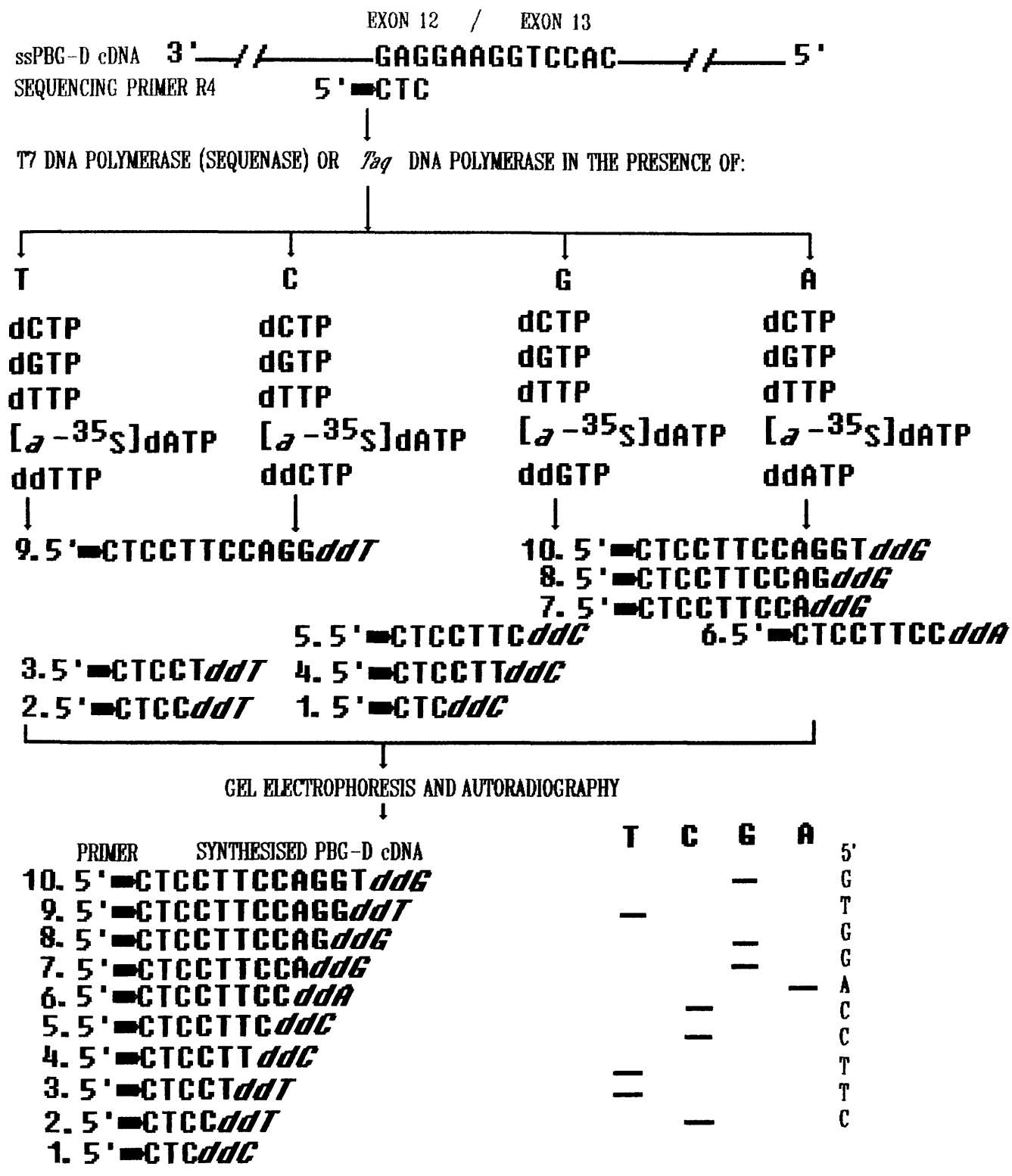


FIGURE 3: DIAGRAMMATIC REPRESENTATION OF THE DIDEOXY-MEDIATED SEQUENCING METHOD.
THE NUMBERS ON THE LEFT REPRESENT THE ORDER IN WHICH EXTENSION-TERMINATION REACTIONS TAKE PLACE IN THE GROWING CHAIN OF THE SYNTHESISED DNA OR cDNA. THE LAST G BAND AT THE 5' (TOP) END OF THE GEL CORRESPONDS TO NUCLEOTIDE 776 IN WHICH A C TO A TRANSVERSION HAS BEEN DETECTED IN AN AIP PATIENT IN THIS STUDY.

The main drawback of genomic sequencing, no matter how comprehensive and definitive it may be, is that it is a labour intensive procedure. This is especially so when used for characterisation of unknown mutations in long DNA segments. The advent of PCR and subsequent direct sequencing of the PCR products has to a great extend reduced the time and labour required in RNA or DNA sequencing. Direct sequencing as opposed to conventional sequencing based on cloning, is faster, simpler and more reliable since for each sample, only a single sequence needs to be determined. This is based on the fact that it virtually abolishes potential errors due to *Taq* DNA polymerase infidelity in the amplification process since such errors would not be detectable against the much greater signals of the consensus sequence. Moreover it is amenable to automation because being an *in vitro* method it is independent of any cultured organisms (Gyllensten, 1989; Gibbs et al., 1989).

There are two main approaches to direct sequencing of PCR products. The first is based on a method of sequencing dsDNA as produced by a standard PCR and involves a denaturation step of boiling or alkaline treatment prior to annealing of the template with the sequencing primer (Wrishchnick et al., 1987; Higuchi et al., 1988a; Newton et al., 1988; Gyllensten 1989). In the heat denaturation method, the template is melted by boiling in the presence of the sequencing primer and the mixture snap frozen at -70^0C. To prevent renaturation, the sequencing reaction is then done immediatly following the shortest possible annealing stage. Although the annealing step seems to be very crucial in direct sequencing of PCR products, there have been very few attempts at its optimimisation prior to the work of Casanova and co-workers (1990). Recommended durations, incubation temperatures and the template:primer ratios for the annealing step seem to vary from author to author (Wrischnik et al., 1987; Newton et al., 1988; Higuchi et al., 1988; Kretz et al., 1989; Winship 1989). For optimum results it is, however, recommended to perform the primer-template annealing by first denaturing the template in the presence of the primer at 100^0C for 2 min and then immediately freeze at -70^0C, for 15 to 45 seconds using a primer:template ratio of approximately 1:20 (Casanova et al., 1990). Alkaline denaturation of PCR products prior to primer-template annealing is based on methods that have

annealing is based on methods that have been developed for the sequencing of supercoiled plasmid DNA (Chen and Seeburg, 1985; Hattori and Sakaki, 1986; Lim and Pene, 1988; Saiki et al., 1988c; Hsiao, 1991). The DNA template is denatured by incubation in NaOH at 37^0C., for 30 minutes, neutralized with sodium acetate or HCl, then precipitated and washed in alcohol. The DNA pellet is finally dissolved in an appropiate volume of either water or TE buffer ready for annealing to the sequencing primer. Alkaline denaturation of ds templates has been claimed to be superior to the heat denaturation method (Chen and Seeburg, 1985). Sequencing of ds PCR products is, however, not invariably satisfactory because of the fact that short linear PCR products tend to reassociate very rapidly thus preventing primer-template annealing or extension (Kreitman and Landweber, 1989).

One way of improving the quality of dsPCR sequencing is by using 5' end-labelled sequencing primer. Unfortunately this entails the use of ^{32}P-label, sacrificing the safer ^{35}S with its superior resolution. Moreover, the kinasing reaction necessary for end labelling introduces an additional enzymatic step (Kreitman and Landweber, 1989). Finally, a new method of sequencing dsPCR products with *Taq* DNA polymerase in cyclic reactions has been described (Lee, 1991b). This method which combines the techniques of PCR and dideoxy sequencing, involves performing cyclic sequencing reactions on PCR products in a thermal cycler in the presence of excess amounts of sequencing primer and a radioactive isotope such as (α-^{35})S d ATP. The repeated cyclic reactions performed in this method enables good sequencing ladders to be generated from minute template quantities. The method also obviates the need for denaturation of the double-stranded templates (Lee, 1991b) and additionally, the use of *Taq* DNA Polymerase in sequencing, allows high annealing and extension temperatures which is beneficial for some templates, especially those with strong secondary structures (Innis et al., 1988). Despite these improvements sequencing of dsPCR products remain problematic and to alleviate this a second approach based on sequencing of ssPCR products has been developed.

Several methods of generating ssPCR templates for sequencing have been described. These include methods that preferentially produce ss template in the PCR as well as those that at first

produce dsPCR products which are subsequently manipulated to generate ss templates. Single-stranded PCR products are produced by a method that has come to be known of as asymmetric PCR (Glyllesten and Erlich, 1988; Saiki et al., 1988c; Kreitman and Landweber, 1989). Asymmetric PCR is performed by using unequal amounts of amplifying primers at a ratio of between 50-100:1, in otherwise unchanged amplification parameters. During the first 20 to 25 cycles this will generate ds products in the usual manner, but following the inevitable exhaustion of the limiting primer, the subsequent products will be single stranded complementary to the limiting primer. The production of the ss template is not as efficent as dsPCR but nevertheless is sufficient for direct sequencing (Saiki et al., 1988c). Alternatively ss templates can be generated by at first performing the standard dsPCR, followed by ethanol precipitation and then reamplification of the products using only one of the amplifying primers, thus by default making the absent primer limiting. In both cases, the generated ss template is then sequenced using either the limiting primer or a nested (internal) primer complementary to the template.

Single-stranded templates suitable for sequencing can also be prepared by removal or separation of one of the strands produced by the normal dsPCR. Several methods have been developed to achieve this, ranging from physical separation to enzymatic or chemical reactions. Strand separation and the subsequent DNA sequencing has for instance been accomplished by using biotinylated primers and magnetic beads (Schofield et al., 1989; Hultman et al., 1989). In this method one of the amplifying primers is biotinylated at the 5' end. After amplification, the amplified product is mixed with magnetic beads to bind the biotinylated strand. The non-biotinylated strand is then eluted with alkali to yield the ssDNA template immobilised at the 5' end, to be used in solid phase sequencing.

Another approach to preparing ss templates is through enzymatic digestion of one of the strands whilst protecting the other. One of such methods is based on the ability of 5'-O-(1-thiotriphosphates) to block the nucleolytic action of exonuclease III when incorporated into substrate DNA (Putney et al., 1981). In this procedure the PCR amplification products are thiolated using Klenow Polymerase and then cleaved with a restriction enzyme to produce

asymmetrically protected fragments. Finally, the cleaved products are digested to completion with exonuclease III which will only digest the unprotected strand (Ward et al., 1989). Alternatively, single-stranded templates can be derived from dsPCR products by chemical means. Nakamaye and co-workers (1989) have described a method based on the incorporation of deoxynucleotide analogues carrying a phosphorothioate substitution in the α-position (dNTPαS) during PCR. The phosphorothioate-containing DNA is subsequently chemically cleaved by alkylating agents such as 2,3-epoxy-1-propanol or 2-iodoethanol. In the method, four separate PCRs are performed each with a different dNTPαS and the radioactive label being introduced in a strand specific manner by either 5' end-labelling of one of the primers or by labelling the PCR products. In the latter method, amplification is carried out with only one of the primers being phosphorylated. At the end of the PCR, the non-phosphorylated primer will then be able to accept radioisotope through its free 5'-end hydroxyl group. To sequence the opposite strand a second round of PCR amplification is done in which the phosphorylation or end labelling of the primer is reversed (Nakamaye et al., 1988).

Alternatively, following dsPCR, one of the strands may be made suceptible to enzyme digestion with lamda-exonuclease by phosphorylating its 5' end (Little et al., 1967). To facilitate this one of the amplifying primers is either synthesized with a 5' terminal phosphate or is kinased prior to the PCR. At the end of the PCR amplification, the products are digested with lamda-exonuclease, a 5' to 3' nuclease which attacks dsDNA only when there is a 5' terminal phosphate. This would therefore, digest the strand derived from the phosphorylated primer leaving behind the opposite strand suitable for sequencing (Higuchi and Ochman, 1988).

Direct sequencing of PCR products can also be done through the incorporation of phage promoters onto the PCR products. This has been done for the sequencing of both DNA and RNA templates (Stoflet et al., 1988; Sarkar and Sommer, 1988; Sarkar and Sommer, 1989). It is achieved by attaching phage promoter sequences onto at least one of the amplification primers. In the direct genomic sequencing, the segments amplified by PCR are transcribed to provide single-stranded templates for reverse trascriptase-mediated sequencing. The transcription

reaction besides providing single-stranded templates, also increases the sensitivity of the method by augmenting the PCR amplification and in deed at times, may compensate for a sub-optimal PCR amplification. With slight modification this method has been used to directly sequence amplify RNA segments. At first cDNA is synthesized with either oligo(dT) or an mRNA specific oligonucleotide primer and then amplified by PCR, in which at least one of the amplification primers contains a phage promoter attached to a sequence complementary to the region to be amplified. The PCR products are subsequently transcribed with the phage promoter and the transcripts used in reverse transcriptase-mediated dideoxy sequencing using nested primers (Sarkar and Sommer, 1988, 1989). Use of phage promoters in the sequencing of either DNA or RNA has the advantage of producing a second round of amplification through the transcription reaction obviating the need for purification of the PCR products. It also provides up to a billionfold amplification, enabling detection of mRNAs present at less than one copy per cell (Sarkar and Sommer, 1989). Phage incorporation however increases the cost of amplifying primers especially when both strands need to be sequenced. Moreover, the need to use reverse transcriptase for sequencing, precludes the use of other sequencing enzymes.

Single-stranded templates can also be prepared from dsPCR products by removal of a complementary strand with ssDNA of an M 13 clone (Gal and Hohn, 1989). The procedure requires the cloning of one variant of the sequence into an M 13 vector to be used in all subsequent sequencing of the related DNA segments. Excess of this M 13 ssDNA is added to the dsPCR products to remove its complementary strand, thus allowing the sequencing primer to anneal to the single-stranded template thus generated. The usefulness of the cloned ssDNA in this procedure is limited to the particular DNA segment for which it was tailored. In addition both the amplification and the sequencing primers must be located outside the cloned region so as to have no homology with the M 13 ssDNA.

Double-stranded PCR productcs can also be directly cloned into M 13 vectors. To facilitate this, primers are constructed in such a way as to include suitable restriction sites (linkers) near their 5' ends. These 5' tails as mentioned before, do not seem to interfere with PCR amplification. After

amplification the PCR products are cleaved with the appropriate restriction enzymes, dialysed to remove inhibitors of ligation and ligated into the M 13 vectors. Unlike standard cloning protocols, cloning of PCR products circumvents the construction and screening of full genomic libraries. In addition it allows sequencing from nanogram amounts of DNA (Scharf et al., 1986). The main disadvantage of the procedure as mentioned earliear, is the need to sequence several clones, firstly to ensure the sequencing of both alleles and secondly to distinguish changes that may arise from the infidelity of *Taq* DNA Polymerase used in the PCR amplification.

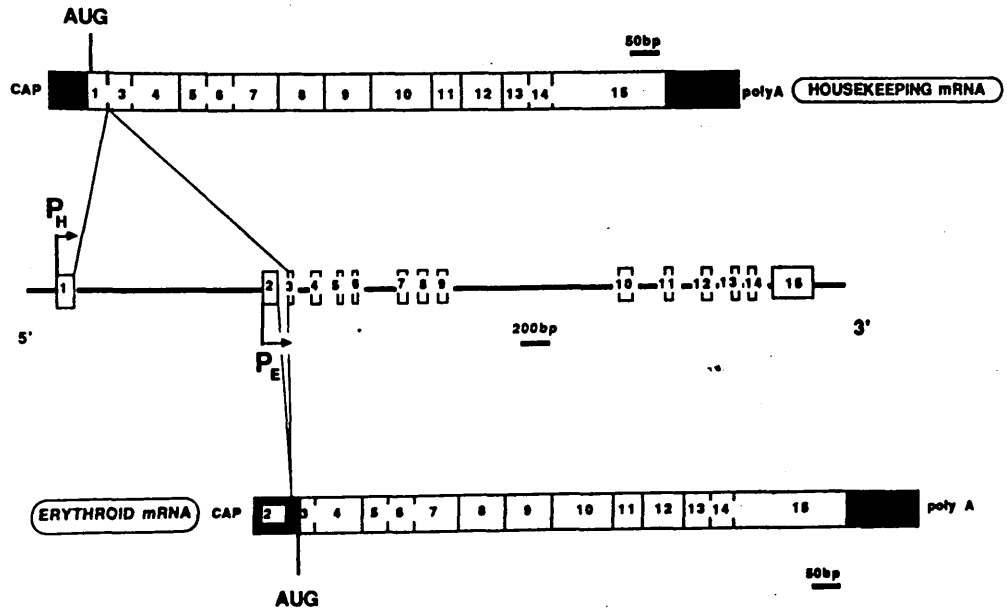
The presence of many different procedures for the direct detection of mutations indicates that currently there is not a single method suitable for all methods. The choice of the method to be used will depend on the length of the DNA to be analysed, type of mutations and the purpose of the test. The approach to detection of mutations in a long DNA fragment with unknown number of mutations will obviously be different from that of a short fragment with a known single base change. Similarly, a choice of method needed for the characterisation of unknown mutations, by necessity, needs to be different from a mere screening test or a test needed for confimation of an already known mutation. Evidence is now emerging that characterisation of unknown mutations in a large gene would require a combination of at least two of these methods. A sound strategy would for example, be to scan such a gene by means of CCM analysis and then characterise the mutations by direct sequencing of the affected regions. For subsequent detection of these mutations in families of the affected individuals, a simple method like ASO analysis by dot blot hybridisation or ARMS analysis can then be used (Cotton, 1989). In a small gene, however, it may be possible and even advisable to characterize all mutations cost effectively by direct sequencing alone.

2.3. Application of methods of mutation detection in acute intermittent porphyria.

Acute intermittent porphyria is an autosomal dominant disorder caused by reduced activity of porphobilinogen deaminase (PBG-D, EC4.1.3.8) the third enzyme in the heme biosynthetic pathway. PBG-D catalyses the condensation of four porphobilinogen monomers to form hydroxymethylbilane (preuroporphobilinogen), a linear tetrapyrole, prior to cyclisation to uroporphyrinogens I or III (Battersby et al., 1980). The deficiency of PBG-D activity causes reduction in heme synthesis and increased activity of amninolevulinic (ALA) synthase, since the later is under negative feedback control by heme. The combination of increased ALA synthase activity and the partial enzymatic block, leads to the accumulation of the porphyrin precursors, delta-aminolevulinic acid and porphobilinogen which being highly water soluble are excreted in urine, particularly during acute attacks and tend to decrease with the clinical improvement (Tschudy et al., 1975). Most carriers of the AIP gene (approximately 90%) however, do not demonstrate excess urinary excretion of delta-aminolevulinic and porphobilinogen and remain clinically latent (Kappas et al., 1983).

Using mouse-human clones, Meisler et al. (1980, 1981) showed that PBG-D is determined by a gene in chromosome 11 and Wang et al., 1981) assigned the locus to the long arm in the region 11q23-qter. This was supported by de Verneuil et al. (1982) who studied the dosage effect in three children with trisomy 11qter and confirmed the assignment to region 11q23.2-11qter. The PBG-D gene measures 10 kb in length and is split into 15 exons (Raich et al., 1986). It is transcribed into two distinct mRNAs through alternative splicing of two primary transcripts arising from two promoters located approximately three kilobases apart. The upstream promoter with a housekeeping function is active in all cells, whereas the downstream promoter is active only in erythroid cells (Chretien et al., 1988). The 5' terminal segment of the ubiquitous PBG-D mRNA is transcribed from exon 1, which lies 3.1 kb upstream from the common region of the two mRNAs. Thus in the ubiquitous (housekeeping) mRNA, the primary transcript is spliced so that exon 1 is linked to exon 3 as shown in figure 4. In contrast, the 5' terminal segment of the

erythroid PBG-D mRNA is transcribed from exon 2, which lies 2.9 kb 3' to exon 1 and 175 bp upstream from exon 3. The primary transcript is then spliced and joined in such a way that exon 2 is joined to exon 3 without exon 1 (Chretien et al., 1988. It has been deduced that the protein sequence of the two isoforms of the PBG-D, differ by the presence of an additional stretch of 17 amino acid residues at the amino terminus of the non-erythroid form (Grandchamp et al., 1987). This may also explain the difference in their molecular weights, whereby the erythroid PBG-D is 40 kDA whilst the non-eryrthroid PBG-D is 42 kDA (Lannfelt et al., 1989). The reading frame of the non-erythropoid PBG-D gene is 1083 bp long and encodes 361 amino acid residues whereas the erythroid form encodes 344 residues (Grandchamp et al.,1987).



Structure of the human PBGD gene. The exons and introns are drawn to scale. Dashed boxes represent untranslated regions.

Figure 4: Structure of human PBG-D gene.

(From Chretien et al., 1988). Alternative splicing of the PBG-D gene for erythroid and non-erythroid mRNA is indicated. P_H stands for the housekeeping promoter and P_E the erythroid promoter. The non-coding regions of the erythropoietic cDNA comprise 81 bases at the 5' end preceding the initiation methionine codon, and 267 bases at the 3' end followed by 14 bases of the consensus poly(A)$^+$ additional signal and poly(A)$^+$ tail respectively (Lee, 1991a).

Acute intermittent porphyria occurs in all races, though the actual prevalence in each is not known and seem to variable in different communities. The highest prevalence is in Lapland, Northern Sweden where it has been estimated as 1/10,000. The general prevalence in subjects of European stock is, however, estimated at 1/20,000 (Goldberg et al., 1987).

Acute intermittent porphyria is a heterogeneous disorder as has been demonstrated by immunological methods, whereby four different forms of PBG-D deficiency have been delineated (Mustajoki and Desnick, 1985). These different classes of AIP have been identified on the basis of presence and amounts of the immunological cross-reactive enzyme protein commonly referred to as cross-reactive immunological material or CRIM. The majority of AIP gene carriers (approximately 85%) are CRIM-negative and the remaining 15% are CRIM positive (Anderson et al., 1981; Desnick et al., 1985). The CRIM-negative group is further subdivided into two types depending on the erythrocyte PBG-D activity, whereby in CRIM-negative type 1 there is half the normal and in type 2, normal activity. Similarly, the CRIM positive group is divisible into two types. CRIM-positive type 1 consists of individuals with a CRIM/PBG-D activity ratio of 1.6 and type 2 with a ratio of 5.6. This suggests that in CRIM-positive type 2 there is a greater formation of the inactive enzyme protein. Within affected families, each member has the same CRIM reaction and type. Although it has been postulated that CRIM-positive type 2 is associated with a milder form of AIP (Mustajoki and Desnick 1985), no systematic evaluation of severity of this condition in relation to the CRIM phenotype has been made. Furthermore, the heterogeneity of the molecular pathology seem to be more complex than this, since the same CRIM phenotype and subtype may be caused by different mutations of the PBG-D gene. Theoretically, it would seem logical that, nonsense, deletions and frame-shift mutations which generate premature stop codons and therefore truncated and unstable proteins would be associated more with CRIM-negative phenotype. In contrast, CRIM-positive phenotype would be expected to be associated more with missense mutations, since these are likely to produce a kinetically defective protein differing from PBG-D by only a single amino acid. In practice, this has not strictly proved to be the case. Among the three CRIM

positive mutations that have been reported to date, two are of the missense type with the substitution of arginine to glutamine (Delfau et al., 1990) and the other is a splice defect which leads to the skipping of exon 12 which encodes for 40 amino acid residues (Grandchamp et al., 1980c). CRIM-negative mutations have proved to be more heterogenous as none of the several mutations described in this group, has so far been reported in more than one or two index families. Molecular pathology which cause CRIM-negative phenotypes include, splice site mutations (Grandchamp et al., 1989a, 1989b), missense mutations (Lee et al., 1990, Nordmann et al., 1990; Delfau et al., 1991), nonsense mutations (Scobie et al., 1990; Lee and Anvret, 1991a; Delfau et al., 1991), a frame-shift mutation and a nine base deletion (Nordmann et al, 1991).

The clinical course and natural history of AIP is very variable. Patients with manifest disease may suffer periodic attacks with normal intervals between. The frequency of attacks varies markedly with some having very few attacks, while others may have regular episodes sometimes with hardly any remissions. The severity of attacks also vary widely; from mild episodes of abdominal discomfort to fulminant attacks with fatal outcome. In addition, clinical manifestations of acute attacks also vary between individuals and even in the same individual from one attack to another. Commonly gene carriers in AIP, remain asymptomatic unless exposed to precipitants which include certain drugs (Tschudy et al., Kappas et al 1983), alcohol, calorie restriction (Felsher and Redeker 1967), infections (Dudzinski and Weinstein 1984) and endocrine factors (Levit et al., 1957; Welland et al., 1964). These can result in neurovisceral crises with abnominal pain, vomiting, mental dysfunction and peripheral neuoropathy. The latter occurs in two-thirds of acute porphyric attacks and may result in permanent disability or be fatal if the respiratory muscles become involved (Goldberg, 1959; Stein and Tschudy, 1970; Becker and Kramer 1977). Identification of gene carriers within these families and counselling about avoidance of precipitants is thus of central importance to the management of this condition. Symptomatic patients excrete large amounts of delta-aminolevulinic acid and porphobilinogen in their urine and often have reduced levels (to approximately 50% of normal) of porphobilinogen deaminase in erythrocytes. Accurate identification or exclusion of carrier status by conventional

biochemical methods is however not always possible (Bonaiti-Pellie et al., 1984; Pierach et al., 1987; Kauppinen et al., 1990). Asymptomatic carriers for instance do not consistently excrete abnormal amounts of porphyrins or the porphyrin precursors, delta-aminolevulinic acid and porphobilinogen (Kappas et al., 1983). Moreover, determination of the erythrocyte PBG-D activity may not always be reliable or conclusive as there is an overlap between the normal and the affected, whereby up to 20% of then affected individuals may have normal activity (Lammon et al., 1979; Pierach et al., 1987). PBG-D activity may also be affected by the presence of other diseases. The activity is for instance reduced in uremia (Andriola et al., 1980), chronic polyarthritis (Blum et al., 1978) and increased in haemolytic disorders (Anderson et al 1977), hepatic diseases (Blum et al., 1978) and haematological malignancies (Epstein et al., 1983). PBG-D activity is also raised in the neonatal period (Hughes and Rifkind 1981). In addition there is a group of AIP patients with normal erythrocyte PBG-activity (Desnick et al., 1985; Grandchamp et al, 1989a; Grandchamp et al., 1989b). Therefore, molecular analysis offers an alternative approach to carrier detection.

Three PBG-D intragenic DNA polymorphisms are available for gene tracking within affected families (Llewellyn et al., 1987; Lee et al., 1988; Kauppinen et al., 1990). This can be used in conjuction with biochemical carrier detection and appears to be more reliable. For example, in the study of Kuappinen et al. (1990), of the 62 persons at 1 in 2 risk, 30 were shown to have inherited the disease-linked haplotype yet five of these had normal results on biochemical testing and in a further patient the result was equivocal. This approach is, however, limited by the need to study multiple key family members some of whom may be deceased or otherwise unavailable as mentioned earlier.

RNase protection assays have been used to study the molecular pathology of the PBG-D gene. In their study Llewellyn et al. (1988) reported a deletion in which exon 13 was suspected to have been skipped. Direct sequencing of PCR-amplified material has been attempted for the characterisation of mutations in AIP but without much success (Grandchamp et al., 1989b). Most of the known mutations in AIP to date have been determined through sequencing of cloned PCR

amplified cDNA in association with ASO hybridisation by 'dot-blot' analysis. The described mutations include splicing defects in exon 1 (Grandchamp et al., 1989a), intron 1 (Grandchamp et al., 1989b), exon 12 (Grandchamp et al., 1989c), exon 10 (Delfau et al., 1991) and a stop codon mutation in exon 9 (Scobbie et al., 1990a) all being caused by single base substitutions. Other single base substitutions with subsequent amino acid changes have been described in exon 8 (Lee et al., 1990), and in exons 9, and 12 (Nordmann et al., 1990, Delfau et al., 1991). The majority of these mutations, totalling five, have in fact been reported in exon 10 (Delfau et al., 1990; Nordmann et al., 1990; Lee et al., 1991; Delfau et al., 1991). Finally a T deletion in exon 14 causing a frame shift of a stop codon has been described (Nordmann et al., 1990; Delfau et al., 1991). The diversity of these mutations confirms the heterogeneity of this condition. This is further underscored by the fact that virtually all mutations described todate have been unique to the original family of detection. Exception to this, are two G to A mutations at nucleotide positions 500 and 518 in exon 10 (respectively).

2.4. Aims of the present study.

The main aim of this study was to develop a method of direct genomic sequencing of PCR amplified templates for the detection and characterisation of molecular pathology in monogenic disoders and apply the method to investigate the porphobilinogen deaminase gene in patients with acute intermittent porphyria.

Specific aims of the study were:

2.4.1. To extract total cellular RNA and synthesise non-erythroid PBG-D cDNA.

2.4.2. To optimise and evaluate different methods of R-T PCR amplifications of the non-erythroid PBG-D cDNA.

2.4.3. To compare different methods of purifying PCR amplified templates and determine the most cost-effective method.

2.4.4. To develop protocols for the direct sequencing of PCR amplified products.

2.4.5. To evaluate the efficacy of different methods of direct sequencing of PCR amplified templates.

2.4.6. To characterise mutations that cause AIP by direct sequencing of the PCR amplified PBG-D cDNA.

2.4.7. To determine the significance of these mutations and compare them to those that have been reported in the literature.

CHAPTER THREE: METHODOLOGY.

3. PATIENTS, MATERIALS AND METHODS.

3.1. Patients and controls.

Families with biochemically confirmed AIP were identified through the records of the Porphyrias Research Unit at the Western Infirmary, Glasgow. After obtaining appropiate consents from both the patients and their general practitioners, blood samples were collected and immortalised lymphoblastoid cell line established. Controls were randomly selected from individuals who presented with problems other than AIP. The first thirty patients with AIP, two of whom were related and ten control subjects were included in the study. Diagnosis of AIP was based on increased excretion of delta-aminolevulinic and porphobilinogen in urine and decreased activity of erythrocyte PBG-D coupled with a clinical history of one or more acute attacks (Moore et al., 1987). Clinical assessement was made according to the methods given by McColl et al (1986). Porphyrins in urine and faeces were measured by the method of Moore (1983) and the erythrocyte PBG-D activity according to Piepkotn et al (1983). The mean erythrocyte PBG-D activity in the AIP patients was 22.3 nmol/h/ml erythrocytes, ranging from 11.7 to 24.7 nmol/h/ml erythrocytes. The normal adult activity range for the enzyme is 25-42 nmol/h/ml erythrocytes in females and 30-48 in males.

3.2. Total cellular RNA extraction.

In the handling of RNA, proper precautions were observed at all the times to avoid contamination with RNAases. Glassware and plasticware used for the preparation and storage of RNA were soaked in 0.1% diethyl pyrocarbonate (DEPC) in water for 14-16 h and rinsed several times with distilled water before being autoclaved for 15 min at 15 lb/sq in, in liquid cycle. When autoclaving was impossible the utensils were rinsed with chloroform. All glassware and plasticware for RNA work were kept aside from the general use. Electrophoresis tank used for running RNA gels was strictly reserved for that purpose. Prior to RNA electrophoresis, the tank was filled with 3% solution of hydrogen peroxide (H_2O_2) and left to stand at room temperature for 10 min, then thoroughly rinsed with DEPC treated water. Additionally all solutions used for handling of RNA were prepared using autoclaved water and RNAase free utensils. Whenever possible such solutions were treated for at least 12 h with 0.1% DEPC in water, prior to autoclaving. Tris containing solutions were however not treated with DEPC but had to be used freshly prepared because DEPC readily reacts with amines. Furthermore, gloves were worn at all the times while preparing reagents and utensils for use in RNA procedures. This was adhered to even more rigorously when handling the RNA itself.

Total cellular RNA was extracted from buccal epithelial cells, lymphocytes and lymphoblastoid cells. Extraction of RNA from buccal cells and lymphocytes was modified from methods described by Lench et al. (1988) and Sherman et al. (1989), whereas extraction of total cellular RNA from lymphoblastoid cells was based on methods of Glisin et al (1974), Ullrich et al. (1977) and MacDonald et al. (1987).

3.2.1. Extraction of RNA from buccal epithelial cells.

Subjects rinsed their mouths first with water then with 15 ml of 0.9% saline for about 10 s. Buccal epithelial cells were pelleted by centrifugation at 500 *g* for 10 min. Supernatants were discarded and cell pellets washed in cold phosphate buffered glucose (6 mM Na_2HPO_4, 1mM KH_2PO_4, 300 mM glucose, pH 7.4) and resuspended in 50 ul 0.1% DEPC in water in microfuge

tubes. The cells were lysed by immersing the tubes in a water bath at 100^0C for 5 min and cooling quickly on ice. Debris was removed by centrifugation for 5 min and supernatants transferred to 0.5 ml microfuge tubes to be used in reverse-transcriptase PCR (R-T PCR).

3.2.2. Extraction of RNA from lymphocytes.

20 ml of venous blood was collected in heparinised bottles and diluted 1:1 with 10% foetal calf serum (FCS). The diluted blood was carefully layered over 8 ml of lymphoprep (9.6% sodium metrizoate/5.6% ficoll) without breaching the interface and centrifuged at 400 *g* for 40 min at room temperature. After centrifugation mononuclear cells including lymphocytes form a distinct band between the sample layer and lymphoprep. The layer of mononuclear cells was carefully removed with a pasteur pipette and washed in 20 ml of 10% FCS, then pelleted by centrifugation at 400 *g* for 20 min. The supernatant was dispensed in chloros and the cell pellet resuspended in phosphate buffered glucose (pH 7.4) to determine cell count. Four million cells were then removed and washed in phosphate buffered glucose and pelleted in a microcentrifuge by centrifugation at 1000 *g* for 10 min. The cell pellet was resuspended in 50 ul of 0.1% DEPC in water and the cells lysed by boiling in a water bath at 100^0C for 5 min and then quickly chilled on ice. The debris was removed by centrifugation and 50 ul of the cell lysate transferred to a 0.5 ml microfuge tube to be used in PCR amplification.

3.2.3. Extraction of total cellular RNA from lymphoblastoid cells.

Two different methods using chaotropic agents were used to extract total cellular RNA from lymphoblastoid cells. In the first method RNA was extracted with guanidinium thiocyanate followed by equilibrium centrifugation with caesium chloride according to Glisin et al. (1977) and Ullrich et al. (1977), whereas in the second method, extraction was done with guanidine hydrochloride and organic solvents according to MacDonald et al. (1987).

3.2.3.1. Extraction of RNA with guanidinium thiocyanate and centrifugation in caesium chloride.

Harvested cells from lymphoblastoid cell lines were stored at -80^0C till required. To each cell pellet in a 30 ml Sorvall tube, five volumes of guanidinium thiocyanate homogenisation buffer (4 M guanidinium thiocyanate, 0.1 M Tris.Cl pH 7.5, 1% β-mercaptoethanol) was added. As β-mercaptoethanol is unstable in solutions, to prepare the homogenisation buffer, 50 g of guanidinium thiocyanate was dissolved in 10 ml of 1 M Tris.HCl (pH 7.5) and brought up to 100 ml with water. The solution was filtered and stored at room temperature and β-mercaptoethanol was added to a final concentration of 1% (0.14 M) just before use. Homogenisation was done by carefully vortex-mixing the cell pellet in the homogenisation buffer and then passing the homogenate several times through a 23-gauge needle until the cell lysate was no longer viscous.

After homogenisation, sodium lauryl sarcosinate was added to the preparation to a final concentration of 0.5% and mixed thoroughly. The mixture was centrifuged at 5000 *g* for 10 min at room temperature and the supernatant transferred to a fresh tube leaving behind any cellular debris. Using a hypodermic syringe fitted with a 23-gauge needle, 6 ml of the cell lysate was carefully layered onto a 7 ml cushion of 5.7 M CsCl, 0.01 M ethylenediamine tetraacetic acid (pH 7.5) in a 13 ml pollyalomer ultracentrifuge tube. The tubes were weighed and balanced by the addition of more guanidinium thiocyanate whenever necessary.

RNA was pelleted by ultracentrifugation at 125,000 *g* for 16 h, at room temperature in a swinging-bucket rotor with the brakes turned off. To achieve this a TH641 rotor was used at the speed of 32,000 rpm. At the end of centrifugation the fluid above the caesium chloride cushion was carefully removed with a pasteur pipette stopping as soon as the DNA band had been removed. This usually presents as a white visible band. The remaining fluid was then carefully decanted, and the tubes inverted to drain on paper towels for 5 min, taking care not to disturb or drop the RNA pellets. Using a red-hot sterile scalpel, the tubes were cut off approximately 1 cm from the bottom forming small cups which were placed on ice. The RNA pellets were then dissolved in 100 ul of TE buffer (10 mM Tris.HCl, 1 mM ethylenediamine tetraacetic acid pH

7.6) with repeated pumping of the pipette and the RNA solution transferred to a fresh microfuge tube. This was repeated with another 100 ul aliquot of TE buffer. The tubes were finally rinsed with 50 ul of TE buffer and the aliquots combined in a sterile microcentrifuge tube giving 250 ul of RNA solution. To ensure that the RNA was completely dissolved, the mixture was thoroughly vortexed then incubated in a water bath at 65^0C for 10 min and finally vortexed once again. To the RNA solution 0.1 volume of 2 M potassium acetate and 2.5 volumes of ice-cold ethanol was added, mixed and chilled at -20^0C for a minimum of 2 h. After precipitation the RNA was collected by centrifugation at 5,000 *g* for 20 min and carefully decanting the supernatant. The pellets were briefly dried and redissolved in 200 ul of TE buffer (pH 7.6) and divided into 100 ul aliquots. To one aliquot, 0.1 volume of 2 M potassium acetate and 2.5 volume of ethanol was added and the RNA solution kept at -80^0C for long term storage. 10 ul was taken from the remaining aliquot to determine the RNA concentration and the rest stored at -20^0C for daily use.

3.2.3.1. RNA extraction with guanidine HCl and organic solvents.

To a cell pellet, 10 volumes of guanidine HCl homogenisation buffer I (8 M guanidium HCl, 0.1 sodium acetate pH 5.2, 5 mM dithiothreitol, 0.5% sodium lauryl sarcosinate) were added and the pellet homogenised by drawing and passing through a 23-gauge needle several times until the lysate was no longer viscous. The homogenate was centrifuged at 5000 *g* for 10 min at room temperature and the supernatant transferred to a fresh tube, then 0.1 volume 3 M sodium acetate (pH 5.2) added and mixed well. To this 0.5 volume of ice-cold ethanol was added and thoroughly mixed. The mixture was then stored for at least 2 h at 0^0C.

The RNA was pelleted by centrifugation at 5000 *g* for 10 min at room temperature and recovered by discarding the supernatant. After drying, the RNA pellet was dissolved in guanidine HCl homogenisation buffer II (8 M guanidine HCl, 0.1 M sodium acetate pH 7.0, 1 mM dithiothreitol, 20 mM ethelenediaminetetraacetic acid pH 8.0). To this, 0.5 volume of ice-cold ethanol was added, mixed and the solution kept at -20^0C for at least 2 h. The RNA was recovered by centrifugation at 5000 *g* for 10 min, discarding the supernatant and allowing the

pellet to dry in air. The ethanol precipitation was repeated twice more. After the third precipitation, the pellet was dissolved by adding 2.5 ml of 0.02 M ethelenediaminetetraacetic acid (pH 8.0) and thoroughly vortexed. The mixture was then centrifuged at 3000 *g* and the supernatant saved. This was repeated with another 2.5 ml of ethelenediaminetetraacetic acid (EDTA) and the supernatants pooled. To this an equal volume of chloroform:1-butanol (4:1) was added, vortex-mixed and the mixture centrifuged at 5000 *g* for 10 min at room temperature. The aqueous (upper) phase was then transferred to a fresh tube and the chloroform:1-butanol extraction repeated. After the second extraction, the aqueous phase was transferred to a fresh tube and 3 volumes of 4 M sodium acetate added. The mixture was then stored for at least 1 h at -20^0C and the RNA recovered by centrifugation at 5000 g for 20 min at 0^0C after removing the supernatant. The RNA pellet was washed once with 3 M sodium acetate (pH 7.0) at 4^0C and centrifuged at 5000 *g* for 20 min at 0^0C. After centrifugation, as much as possible of the supernatant was removed and the pellet dissolved in 500 ul of 0.2% sodium dodecyl sulphate (SDS), 0.05 M EDTA (pH 8.0). To this, 2 volumes of ice-cold ethanol was added and the mixture kept at 0^0C for at least 2 h. The RNA was finally recovered by centrifugation at 5000 *g* for 10 min at 4^0C and the pellet washed with 70% ethanol, then allowed to dry in air. After drying, the pellet was dissolved in 200 ul of TE buffer and divided into 100 ul aliquots. To one of the aliquots, 3 volumes of ethanol was added and the RNA stored at -70^0C for later use. 10 ul were removed from the second aliquot for estimation of the RNA concentration and the rest stored at -20^0C as described previously.

3.2.4. Determination of RNA concentration.

The RNA quantitation was done by spectrophotometry at wave lengths of 260 and 280 nm. To do this, 10 ul of RNA solution was diluted in 990 ul of TE buffer in 1 ml cuvettes and the optical density (OD) determined. An OD of 1 at 260 nm corresponds to approximately 40 ug/ml of RNA. The ratio between readings at 260 and 280 nm was used to determine the purity of the RNA solution, whereby a pure RNA solution was expected to have an OD_{260}/OD_{280} value of 2.

3.2.5.. Qualitative assessment of RNA.

The quality of RNA was assessed by agarose gel electrophoresis. A 30 ml mini-gel was prepared by weighing out 0.45 g of agarose and adding 22 ml of water in a flask. The flask and contents were weighed and the agarose boiled until fully dissolved. The flask was then re-weighed, evaporated water replaced and the gel allowed to cool to 55° C. After cooling, 5 ml of formaldehyde and 3 ml of 10X 3-(*N*-morpholino) ethanesulfonic acid (MOPS) were mixed and quickly poured before setting. Meanwhile, the RNA samples were prepared by mixing 1 ul of RNA solution with 5 ul of formamide, 1.65 ul of formaldehyde, 1 ul of 10X MOPS and 1.3 ul of dH_2O. The mixture was heated to 55^0C for 10 min, quenched on ice and to it 2 ul of gel loading buffer (0.25% bromophenol blue, 0.25% xylene cyanol FF, 40% sucrose in water) added. The samples were then immediately loaded and electrophoresed at 75 V until bromophenol blue reached the bottom of the gel. After electrophoresis, the gel was soaked in water for 1 h, changing the water twice or thrice to wash out formaldehyde. The gel was stained for 5 min in a 0.5 ug/ml solution of ethidium bromide in running buffer, then destained for 2 hr or when necessary overnight and viewed under U.V. light to check for the integrity of ribosomal bands. As a marker the samples were run alongside 1 ug of *Escherichia coli* ribosomal RNA (Boehringer) treated in the same way as the RNA samples.

3.3. Polymerase chain reactions.

To develop a method of direct sequencing, PCR amplifications were done on various templates. These included DNA and RNA fragments from the plasmid pBR322, bacteriophage lambda DNA, Duchenne muscular dystrophy gene, cystic fibrosis gene, chorionic RNA, placental RNA, HeLa cell RNA and porphobilinogen deaminase gene. All PCRs were done on a Techne PHC 1 thermal cycler.

3.3.1. PCR amplification of pBR322.

Primers flanking a 500 bp region of pBR322 DNA were constructed and used in amplification of this region by PCR. Both coiled and linear pBR322 DNA were used in these experiments. To linearise pBR322 DNA, the plasmid was digested to completion with *Eco*RI. The digestion was carried out by incubating at 37^0C for 2 h, 5 ug of plasmid DNA with 50 u of *Eco*RI in a reaction mixture containing 50 mM Tris-HCl (pH 8.0), 10 mM $MgCl_2$, 100 mM NaCl and 5 mM spermidine. At the end of the digestion the digests were ethanol precipitated in the presence of 0.3 M sodium acetate (pH 5.2). This was done by adding to the digests, 0.1 volume of 3 M sodium acetate (pH 5.2), mixing well and then adding 2 volumes of ice-cold ethanol. The mixture was stored on ice at 0^0C for 30 min, then centrifuged at 12,000*g* for 10 min to recover the DNA pellet. The pellet was washed once in 70% ethanol by adding 500 ul of ethanol and recentrifuging at 12,000*g* for 2 min. After removal of the supernatant the pellet was dried and dissolved in 10 ul of TE buffer.

Amplification was performed by addding 1 ug of the plasmid DNA to a 100 ul reaction mixture containing 50 mM KCl, 10 mM Tris.HCl (pH 8.4), 1.5 mM $MgCl_2$, 100 ug/ml gelatine, 200 uM of each dNTP (dATP, dCTP, dGTP, dTTP) and 0.5 uM of each primer. To this, 2.5 u of *Taq* DNA Polymerase was added, the mixture overlaid with 100 ul of light mineral oil and amplified for 30 cycles. Usually a master mix of reagents was prepared whereby to a 1.0 ul microfuge tube were added all components for PCR except DNA and the enzyme. The mixture was then put in a U.V. box (Amplirad II) for 10 min. After U.V. treatment, DNA was added and thoroughly mixed by vortexing, following which *Taq* DNA polymerase was added and briefly vortexed. The reaction mixture was overlaid with light mineral oil and briefly centrifuged to bring down any droplets so formed to the bottom of the tube. The first denaturation was done at 93^0C for 3 min and thereafter for 1 min in the remaining cycles. Annealing was performed at 55^0C for 2 min and extension at 72^0C for 1.5 min except for the last extension which was carried out for 7 min. At the end of PCR, samples were briefly centrifuged and 5 ul of each reaction mixture taken from below the oil and resolved in a 1.4% SeaKam agarose gel in 0.5X TBE buffer (44.5 mM Tris-

borate, 1 mM). Electrophoresis was done at 10-15 V/cm until bromophenol blue migrated two thirds the distance of the gel. The gel was stained by soaking for 30 min at room temperature in the running buffer containing 0.5 ug/ml of ethidium bromide. When necessary the gel was destained for 20 min before visualisation under U.V. light.

3.3.2. PCR amplification of bacteriophage lambda DNA.

A 500 bp segment of bacteriophage lambda DNA supplied in the GeneAmp^{TM} DNA amplification kit (Perkin Elmer Cetus) was amplified using the supplied control amplification primers. The sequences of primers complementary to the sense and antisense strands were: 5'-GGTTATCGAAATCAGCCACAGCGCC-3' and 5'-GATGAGTTCGTGTCCGTACAACIGG-3' respectively. 1 ng of the target DNA fragment and 1 uM of each primer was used in PCR using otherwise the same conditions as those described for the amplification of pBR322 DNA. Similarly, analysis of the amplified products was done in the same way.

In attempts to produce single-stranded templates for direct sequencing, asymmetric PCR was performed on bacteriophage lambda DNA. In asymmetric PCR, similar parameters as those for standard PCR were followed, except that the number of PCR cycles performed and the amounts of dNTPs and amplification primers used were changed. The asymmetric PCRs were performed with a reduced concentration of each dNTP to 50 uM instead of 200 uM. To produce the sense strand, 50 pmol of its complementary primer were used along with 0.5 pmol of the primer complementary to the opposite strand, whereas the antisense strand was produced by reversing the ratio of the primers used. Another departure from the standard PCR was that 40 cycles were done instead of 30.

3.3.3. Asymmetric PCR amplification of cystic fibrosis gene.

A DNA fragment consisting of 493 bp, spanning across the $G_{1,784}$->A (G551D) mutation on the cystic fibrosis gene was amplified by asymmetric PCR to produce templates for sequencing. Sequences of the forward and reverse primers used were, 5'-TTCAGCAATGTTGTTTTGACCAAC-3' and 5'-CACAGATTCTGAGTAACCATAATC-3' respectively. Amplifications were performed on genomic DNA in 100 ul volumes containing 50 mM KCl, 10 mM Tris.HCl (pH 8.4), 2.5 mM $MgCl_2$, 100 ug/ml gelatine, 50 uM of each dNTP, 0.5 uM of the forward primer, 0.05 uM of the reverse primer, 1 ug of DNA and 3.5 u of *Taq* DNA polymerase. To produce the opposite strand, a second set of reaction mixtures was prepared in which the amounts of the primers were reversed to contain 0.05 uM of forward and 0.5 uM of reverse primer.The actual procedure followed in the preparation of the reactions, was as described previously above. Amplification was performed by initially heating the mixture to 95^0C for 10 min prior to the addition of *Taq* DNA polymerase. The samples were then denatured at 91^0C for 0.5 min, annealed at 52^0C for 1 min and extended at 72^0C for 0.5 min. 40 cycles were performed in which the last extension step was done for 10 min. At the end of the PCR, the amplification products were briefly centrifuged and 5 ul removed under the oil cap to be analysed as described previously.

3.3.4. Reverse-transcriptase PCR.

Reverse-transcriptase PCR amplifications of the PBG-D cDNA were performed on various RNA samples e.g. from placenta, chorion, HeLa cells, buccal epithelial cells and from total cellular RNA extracted from human lymphocytes. In doing R-T PCR, two main approaches were used in the synthesis of first strand cDNA. In the first approach, a forward (upstream) primer was used whereas in the second, oligo(dT) was used. In both cases, reverse transcriptase and the PCR were performed in the same tube. Oligonucleotide primers used in the amplification and subsequent sequencing of the PBG-D gene and their relative positions along the PBG-D cDNA are as as shown in figure 5. The oligonucleotide primers were synthesised on an Applied Biosystem Incorporation (ABI) 380B DNA synthesiser at small scale cycle, using ABI cyanolethyl phosphoramidites. After synthesis, all primers were deprotected with ammonia, purified by reverse-high pressure liquid chromatography (HPLC) and dissolved in 1 ml of dH_2O. The oligonucleotide primers were quantified by measuring their OD at 260 nm.

F1: 5'- ACT TTC CAA GCG GAG CCA TG -3'
F2: 5'- CAT GTC TGG TAA CGG CAA TGC GG -3'
F3: 5'- GAT TCG CGT GGG TAC CCG CA -3'
R6: 5'- GAG GCT TTC AAT GTT GCC AC -3'
F4: 5'- ATG CCC TGG AGA AGA ATG A -3'
F5: 5'- GGA AGC TTG ACG AGC AGC AGG A -3'
R5: 5'- AGC ATA CAT GCA TTT CTC A -3'
R4: 5'- CTA CTG GCA CAC CTG CAG CCT C -3'
R3: 5'- CAT GGT AGC CTG CAT GGT CT -3'
R2: 5'- CTG TGC CCC ACA AAC CAG TTA AT -3'
R1: 5'- TAG GCA CTG GAC AGC AGC AA -3'

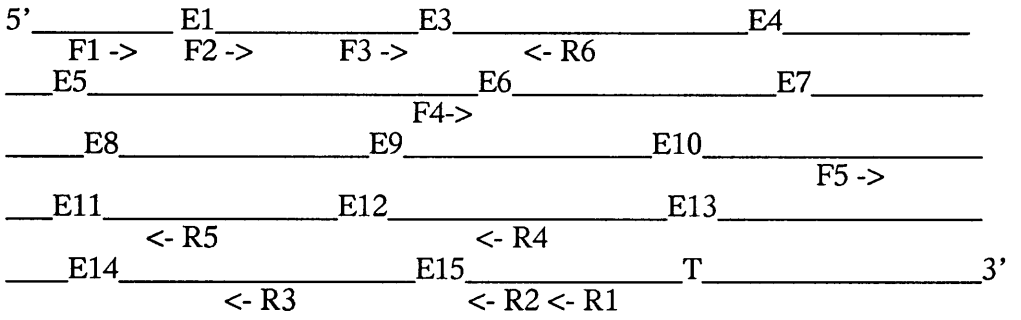


Figure 5: Oligonucleotide primers used in PCR amplification and sequencing and their relative positions on the PBG-D cDNA.

Sequences of oligonucleotide primers used in PCR amplifications and sequencing reactions. All primers were fully deprotected by ammonium treatment and purified by reverse-HPLC. F and R represents forward and reverse primers respectively. E stands for exon and T the termination codon. The PBG-D map is not drawn to scale.

3.3.4.1. Reverse-transcriptase PCR using forward (upstream) primers.

RNA isolated from one of the above described methods was reverse transcribed by using with a forward (upstream) primer complementary to mRNA and the cDNA thus synthesised amplified directly by PCR. To accomplish this, 1 ug of RNA was heated to 65^0C for 10 min, quenched on ice and then reverse transcribed for 1 hr at 37^0C in a 20 ul reaction mixture containing 200 uM each dNTP (dATP, dCTP, dGTP, dTTP), 10 mM DTT, 75 mM KCl, 50 mM Tris.HCl (pH 8.3), 3 mM $MgCl_2$, 100 ug/ml bovine serum albumin, 0.5 uM forward primer and 200 u of murine malony leukaemia virus-reverse transcriptase. At the end of the reverse transcription the mixture was heated to 95^0C for 5 min, quenched on ice and the reverse (downstream) primer added to the final concentration of 0.5 uM. The volume of the reaction mixture was adjusted to 100 ul and 2.5 u of *Taq* DNA polymerase added. The samples were then briefly vortexed and capped with 100 ul of light mineral oil. Finally after a brief centrifugation, the cDNA was amplified during 30 cycles of PCR. The initial denaturation was done at 95^0C for 3 min and thereafter at the same temperature for 1 min. Annealing was performed at 60^0C for 1 min and extension at 72^0C for 1.5 min except for the last extension step which was increased to 7 min.

At the end of amplification the samples were drawn below the oil cap and resolved in 1.4% SeaKem agarose gels and visualise under U.V. light as described previously. The appropriate band was excised and the cDNA eluted by soaking for 3 h in 100 ul of TE buffer. The cDNA in solution was removed from the agarose by pipetting and stored at -20^0C or used directly in asymmetric PCR without further purification.

3.3.4.2. Reverse-transcriptase PCR using 0ligo (dT).

The R-T PCR was performed in a similar way as above, except 0.2 ug oligo $(dT)_{12\text{-}18}$ was used instead of a forward primer in the synthesis of the first strand cDNA. At the end of the reverse transcription, forward and reverse primers were added to a final concentration of 0.5 uM of each and amplification carried out as described above.

A modification of this was also performed using unequal amounts of primers as in asymmetric PCR designed to directly synthesise single-stranded cDNA. In 'asymmetric' R-T PCR the amplification was carried out with 50 pmol of one primer and 0.5 pmol of the second primer.

In another modification, total cellular RNA was first treated with RNAase-free DNAase I, before being amplified in R-T PCR. 89 ul of RNA solution at 500 ug/ml were mixed with 1 ng of RNAase-free DNAase in a reaction mixture containing 50 mM Tris.HCl (pH 7.6), 10 mM $MnCl_2$, 0.1 mg/ml bovine serum albumin and incubating at 37^0C for 20 min. After incubation, the reaction mixture was extracted once with 200 ul phenol:chloroform and the aqueous phase removed to a fresh microfuge tube. The RNA was precipitated by adding 0.1 volume of 3 M sodium acetate and 2.5 volumes of ethanol, chilled on ice for 15 min, then recovered by centrifuging at 12,000 *g* for 15 min. The RNA pellet was washed once in 70% ethanol, resuspended in 90 ul dH_2O and 3 ul used in each R-T PCR as described before.

3.3.5. Asymmetric PCR of PBG-D cDNA.

Porphobilinogen deaminase cDNA synthesised by R-T PCR was reamplified by asymmetric PCR to produce templates suitable for direct sequencing. Conditions for the asymmetric PCR were modified in such a way that a minimum amount of dNTPs were used. 5 ul of PBG-D cDNA estimated to be 0.5-1.0 ug by ethidium bromide fluorescence intensity, was amplified in a 100 ul reaction mixture containing 50 mM KCl, 10 mM Tris.HCl (pH 8.4), 1.5 mM $MgCl_2$, 100 ug/ml gelatine, 50 uM of each dNTP (dATP, dCTP, dGTP, dTTP), 0.5 uM forward primer and 0.05 uM reverse primer. To synthesise the opposite strand, the primer concentrations were reversed. As before master mixes were prepared by adding all components except for the cDNA and the enzyme. The mixture was then irradiated in a U.V. box (Amplirad II) for 10 min. After irradiation, cDNA was added and the mixture thoroughly vortexed. Finally, 2.5 u of *Taq* DNA polymerase were added and after vortexing briefly, the reaction mix was capped with 100 ul of mineral oil. PCR was done using the same parameters as in the standard PCR except that 40 instead of 30 cycles were performeed.

3.4. Purification of PCR products.

At the end of PCR, the amplified DNA or cDNA was concentrated and purified by removing excess primers, dNTPs and salts. Various methods were used and compared. These included Sephadex™ spun-column chomatography (Sambrook et al., 1989b), agarose gel purification followed by nucleic acid chromatography system (NACS) purification (BRL), Geneclean purification (Vogelstein and Gillespie, 1979), centrifuge-driven dialysis (Saiki et al.,1988a; Sambrook et al., 1989a), selective alcohol precipitation (Gyllensten, 1989; Brow M-A D, 1990) and acrylamide gel purification.

3.4.1. Spun-column chromatography.

At the end of PCR, 5 ul of the PCR products were removed below the oil cap and analysed as mentioned before. On finding that the appropriate products had been amplified, the rest of the sample was removed and extracted once with chloroform to remove the mineral oil. This was effected by mixing the PCR products with 150 ul of chloroform and centrifuging the mixture at 12,000 *g* for 15 seconds. The aqueous phase was removed and re-extracted in butanol and this time the lower (aqueous) phase retained. The remaining butanol was driven off by immersing the tubes in a 37^0C water bath with the caps off for 5 to 10 min. The DNA or cDNA was then precipitated by addition of 0.1 volume of 3 M sodium acetate and 2.5 volumes of ethanol as described before. After recovery by centrifugation the DNA pellet was resuspended in 100 ul of O.5X TE buffer pH (8.0).

In the meantime, spun columns were prepared by plugging the bottom of 1 ml disposable syringes with a small amount of sterile glass wool and completely filling them with Sephadex G-50 (DNA grade) equilibrated in 0.5X TE buffer. The syringes were then inserted onto 20 ml universal containers and centrifuged at 1600 *g* for 4 min at room temperature in a swinging bucket rotor. More Sephadex G-50 was added and the centrifugation repeated until the volume of the packed columns was approximately 0.9 ml and remained unchanged on further centrifugation. 100 ul of 0.5X TE buffer was then carefully applied to the top of each column

which was subsequently centrifuged at 1600 *g* for 4 min and repeated twice. At the end of the third wash, the PCR amplified DNA samples in 100 ul 0.5X TE buffer were applied to the top of the columns. The columns were then inserted into decapped 1 ml microfuge tubes, placed in fresh universal containers as shown in figure 6, and centrifuged at 1600 *g* for 4 min at room temperature to collect 100 ul of the effluent. The eluted cDNA was concentrated by precipitating in 0.1 volume of 3 M sodium acetate and 2 volumes of ethanol, then washed once in 70% ethanol as described earlier. After drying, the cDNA pellet was dissolved in 14 ul of TE buffer (pH 7.4) for use in sequencing reactions.

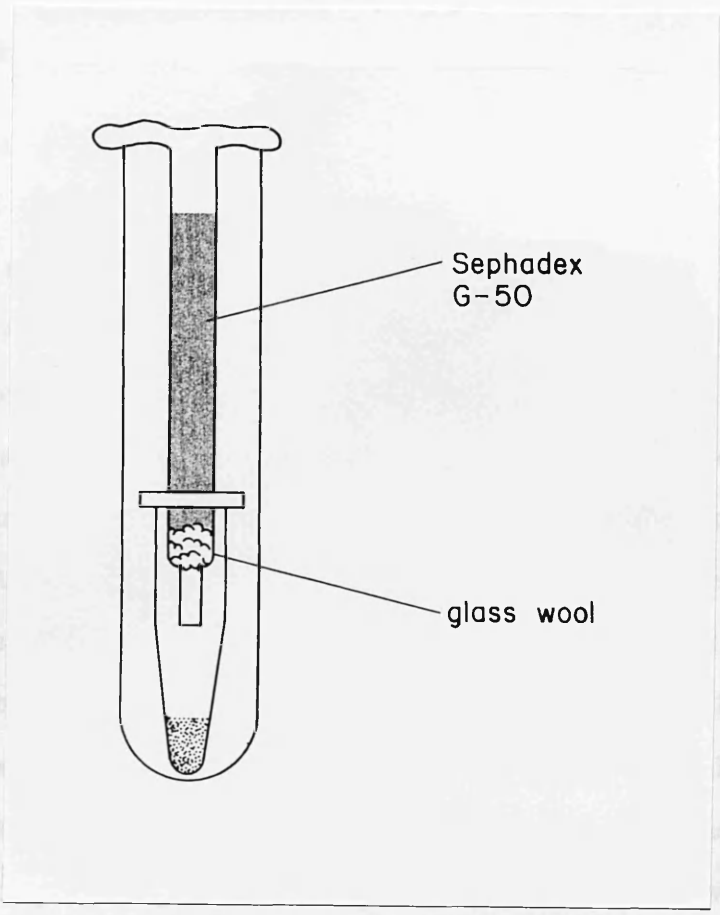


Figure 6: Spun-column chromatography.

(From Sambrook et al., 1989b).

Spun columns are prepared by packing 1 ml disposable syringes with Sephadex G-50™ beads (DNA grade) and equilibrating them with 0.5X TE. Such prepared spun-columns, can be stored in upright position at 4^0C for up to one month or more before use. They should, however, be rehydrated with 0.5X TE just before use. Samples are purify by carefully applying them at the top of the Sephadex columns and centrifuging in a swinging bucket rotor, collecting the eluates in decapped 1 ml microfuge tubes.

3.4.2. Agarose gel fractionation followed by NACS column chromatography.

At the end of PCR amplification, the products were electrophoresed in a 1.5% low melting NuSieve agarose gel and the appropriate band excised. The gel slice was placed in a microfuge tube and completely melted by placing in a water bath at 70^0C. The volume of the molten gel was determined and 4 volumes of 0.25 M NaCl in TE buffer (pH 7.2) added. The diluted molten gel was then incubated at 70^0C for 10 min. In the meantime the resin in NACS columns was hydrated by being washed thrice with 2.0 M NaCl in TE buffer (pH 7.2) and equilibrated with 0.2 M NaCl in TE buffer (pH 7.2). The diluted molten gel was then loaded onto NACS colums and the bound cDNA washed with warm 0.2 M NaCl in TE buffer (pH 7.2) heated to 42^0C, to remove gel impurities. The cDNA was eluted by adding 100 ul of 1.0 M NaCl in TE buffer (pH 7.2) and collecting the effluent by gravity. This was repeated twice more and the three effluents pooled. The cDNA was precipitated by mixing with 10 ug of tRNA and 600 ul of 95% cold ethanol (-20^0C) and freezing the mixture at –70^0C for 10 min. At the end of the precipitation, the cDNA was recovered by centrifugation at 12,000 *g* for 15 min and the pellet washed once in 80% ethanol, dried and resuspended in 14 ul of TE buffer (pH 7.4).

Alternatively the PCR products were directly purified in the NACS columns without any prior gel purification. Using this alternative method, the PCR products were extracted once in chloroform as described previously and loaded directly onto the equilibrated NACS columns to be processed as above.

3.4.3. Geneclean^{TM} purification.

Agarose gel electrophoresis followed by Geneclean^{TM}II (BIO 101 Inc) purification was also used to clean up PCR products prior to sequencing. At the end of PCR amplification, the products were resolved in 1.5% low melting NuSieve agarose gel as mentioned before. The appropriate band was excised and DNA eluted using geneclean according to the manufacturer's recommendations. Briefly, the excised gel slice was weighed and assuming 1 g is equivalent to 1 ml, 0.5 volume of 'TBE modifier' and 4.5 volumes of the supplied stock solution of NaI were

added giving a final concentration of at least 4 M. The mixture was incubated at 50^0C for 5 min to dissolve the agarose gel. At the end of incubation, 5 ul of the provided 'glassmilk' suspension was added, mixed and the mixture kept on ice for 5 minutes to bind the DNA molecules. The mixture was then centrifuged in a microcentrifuge for 5 seconds and the supernatant removed. To ensure that most of the liquid had been removed, the tubes were recentrifuged and as much as possible of the supernatant removed. The pellet was then resuspended in 10-50 volumes of the 'NEW WASH' provided, centrifuged for 5 seconds and the supernatant discarded. This washing was repeated twice more and after the third wash the pellet was resuspended in 7 ul of TE buffer and incubated at 50^0C for 3 min. The mixture was then centrifuged for 30 seconds and the supernatant containing eluted DNA was transferred to a fresh microfuge tube. This was repeated once and the two eluates pooled to be used in sequencing reactions.

3.4.4. Centrifuge-driven dialysis method.

After removing 5 ul for checking, the remaining 95 ul of the PCR products were extracted once with 150 ul of chloroform to remove mineral oil. The aqueous layer was transferred to a fresh tube and 2 ml of dH_2O added. This was then loaded to the reservoir chamber of a Centricon 30 microconcentrator as shown in figure 7 and centrifuged at 3000 *g* (6500 rpm in a Sorvall SA 600 rotor) for 30 min at room temperature. At the end of centifugation the filtrate was discarded and the unit inverted with the retentate cup placed at the bottom of the rotor. The sample was then centrifuged at 1000 *g* (3000 rpm in a Sorvall SA 600 rotor) to collect the concentrated products. The amplified cDNA was recovered by ethanol precipitation and after washing once with 70% ethanol was dried and dissolved in 14 ul of TE buffer (pH 7.4).

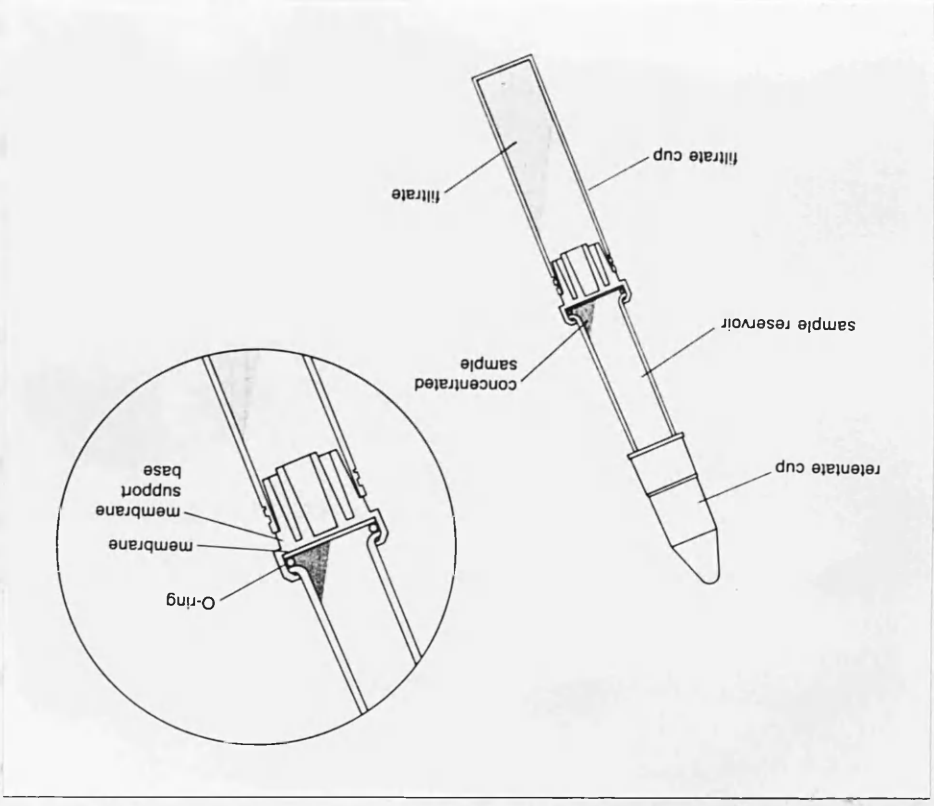


Figure 7: Centrifuge-driven dialysis using microconcentrators.

(From Amicon publication number 1-259G).

The microconcentrator device consists of two chambers, a filtrate cup and a sample reservoir separated by a filtration membrane. Samples are desalted and concentrated by centrifuging, after their dilution and application into the sample reservoir. The filtrate collects in the filtrate cup which can be used for storage. This is useful for checking sample losses if necessary, as for instance when dealing with precious material. The retentate cup is then attached to the sample reservoir and the device inverted to collect the sample by centrifugation.

3.4.5. Selective alcohol precipitation.

Two methods were used, based on the fact that 2.0-2.5 M ammonium acetate reduces the coprecipitation of dNTPs with DNA (Okayama and Berg, 1982). At the end of PCR amplification, the samples were analysed and mineral oil extracted as described before. An equal volume of 4 M ammonium acetate (pH 5.2) was then added, mixed and to this, 3 volumes of ethanol added and again mixed. The mixture was left at room temperature for 5 min, then centrifuged at 12,000 *g* for 10 min to recover the DNA. The DNA pellet was finally washed in 70% alcohol, dried in air and dissolved in 14 ul of TE buffer (pH 7.4).

Alternatively, the PCR products were mixed with an equal volume of 4 M ammonium acetate (pH 5.2), and precipitated in 2 volumes of propan-2-ol for 10 min at room temperature and the cDNA recovered by centrifuging at 12,000 *g* for 10 min. The cDNA pellet was then washed in 70% alcohol, dried and dissolved in 14 ul of TE buffer (pH 7.4) as described before.

3.4.6. Purification and strand separation in acrylamide gel.

Simultaneous purification and strand separation of PCR products was performed on 5% acrylamide gels. The gels were prepared by dissolving 2.5 g of acrylamide and 0.05 g of *N,N'*-methylene bisacrylamide in 1X TBE buffer (89 mM Tris-borate, 2 mM EDTA, pH 8.0). The gel was polymerised by adding 90 ul 25% ammonium persulphate (APS) and mixed by swirling followed by addition of 25 ul *N,N,N',N'*-tetramethyethylenediamine (TEMED) which was mixed in the same manner. The gel solution was allowed to stay for a period of 2 h, to ensure complete polymerisation of acrylamide. The polymerised gel was then pre-run for 2 h at 10 V/cm.

After standard PCR amplification of PBG-D cDNA as described earlier, dsPCR products were removed under the mineral oil, checked and extracted once with chloroform. The aqueous phase was then precipitated with ethanol and washed twice in 70% ethanol. After the second washing the cDNA pellet was dissolved in 40 ul of strand separation buffer containg, 30% dimethyl sulphoxide (DMSO), 1 mM EDTA, 0.05% bromophenol blue, 0.05% xylene cyanol FF and heated to 90^0C for 2 min. After heating, the samples were quickly chilled on ice and immediately

applied to the gel. Electrophoresis was carried out until bromophenol blue reached the bottom of the gel. The gel was stained with ethidium bromide and appropriate bands excised. The separated strands were recovered using NACS columns as described above.

3.5. Genomic sequencing.

The PCR amplified DNA and cDNA were directly sequenced using both double- and single-stranded templates involving various sequencing protocols.

3.5.1. Sequencing of dsPCR products with end-labelled primers.

Double-stranded PCR products were purified using one of the described methods above and used directly in sequencing reactions without cloning. Sequencing reactions were done with either ^{32}P end-labelled primers or uniformly ^{35}S-labelled templates by the dideoxy method of Sanger (1977) using Sequenase T7 DNA polymerase or *Taq* DNA polymerase enzymes. Sequencing primers were (^{32}P) 5' end-labelled using 5.3 u of T4 polynucleotide kinase by mixing 20 pmol of the primer with 25 pmol of (gamma-^{32}P)ATP (3000 Ci/mmol) in a 20 ul reaction mixture containing 50 mM Tris.HCl, 10 mM $MgCl_2$, 5 mM dithiothreitol, 0.1 mM spermidine HCl and 0.1 mM EDTA (pH 8.0). The reaction mixture was incubated at 37^{0}C for 30 min. In practice, the reaction mixtures were mixed, briefly centrifuged and incubated at 37^{0}C for 15 min. After 15 min the samples were again mixed, recentrifuged and incubated for another 15 min. At the end of incubation the T4 polynucleotide kinase was inactivated by heating the samples to 90^{0}C for 10 min. The amount of radioactivity incorporated into the primers was checked by diluting 0.5 ul of the reaction mixture in 4.5 ul of water and spotting 2 ul onto two DE-81 filters. The filters were allowed to dry and one of them washed four times by swirling in 250 ml of 0.5 M Na_2HPO_4 (pH 7.0), each wash lasting for 5 min. This was followed by two, 1 min washes in water and a final 1 min wash in 95% ethanol. The filter was allowed to dry and then both the washed and unwashed filters were placed in scintillation vials and 5 ml of toluene-based scintillation fluid added. The amount of the radioactivity was determined and the

proportion of the incorporated radioactivity calculated by the following formula: cpm in washed filter/cpm in unwashed filter = proportion of the radioactivity incorporated. Incorporation rate of 20% or more was usable for the subsequent sequencing reactions.

The double stranded templates had to be denatured before being used in sequencing reactions. This was done either by alkaline treatment or by boiling at 100^0C. In the alkaline denaturation method, 17 ul of the purified PCR products were denatured in 0.2 M NaOH, 0.2 mM EDTA at 37^0C for 30 min and the mixture neutralised by adding 0.1 volume of 3 M sodium acetate (pH 5.2). The DNA was then precipitated by adding 2 volumes of ethanol and stored at -70^0C for 15 min. After precipitation, the DNA was recovered by centrifugation and washed in 70% ethanol as described before. In Sequenase T7 DNA polymerase mediated sequencing, the denatured DNA was redissolved in 14 ul of water and 7 ul used in each sequencing reaction, whereas in *Taq* DNA polymerase sequencing, the pellet was resuspended in 20 ul and 10 ul used the sequencing reaction.

In the Sequenase T7 DNA polymerase sequencing reactions, template-primer annealing was done by adding 1 pmol (1 ul) of the ^{32}P end-labelled sequencing primer to 5 ul of the alkaline denatured template in a 10 ul reaction mixture containing 40 mM Tris.HCl, 20 mM $MgCl_2$, 50 mM NaCl (pH 7.5) and heating the mixture to 65^0C for 2 min then cooling slowly to below 30^0C over a period of 20 min. The annealed template-primer mixture was briefly centrifuged and immediately kept on ice. As sequencing primer, one of the amplification primers or a nested primer complementary to the strand to be sequenced was used.

Alternatively dsPCR templates and sequencing primers in the reaction mixture described above, were denatured and annealed by melting at 100^0C for 5 min, then snap quenched in dry ice-ethanol bath at -70^0C for 30 seconds and immediately thawed by a 15 s centrifugation. The annealed template-primer mixture was then placed on ice and immediately used in extension-termination reactions for sequencing.

To the ice-cold annealed template-primer mixture were added 1 ul 0.1 M dithiothreitol (DTT), 2.5 ul dH_2O, and 2 ul (1.5 u) of a 1:8 dilution of Sequenase T7 DNA polymerase enzyme in TE

buffer. 3.5 ul of this mixture were immediately added to 2.5 ul of dideoxy A, C, G, and T termination mixes which had been pre-warmed to 37^0C for at least 1 min. Dideoxy A termination mix contains 80 uM dATP, 80 uM dCTP, 80 uM dGTP, 80 uM dTTP, 8 uM ddATP and 50 mM NaCl. Dideoxy C, G and T termination mixes are similarly constructed with the appropriate ddNTPs; ddCTP, ddGTP, ddTTP respectively. The mixtures were incubated at 37^0C for 5 min and the reactions terminated by adding 4 ul of a stop mix containing 95% formamide, 20 mM EDTA, 0.05% bromophenol blue, 0.05% xylene cyanol FF.

3.5.2. Sequencing of dsPCR products by incorporation labelling.

Double-stranded PCR amplified templates were prepared and annealed to the sequencing primers as described above. To 10 ul of the annealed template-primer mixture on ice, were added 1 ul 0.1 M DTT, 2 ul 1.5 uM each dCTP, dGTP, dTTP, 0.5 ul (α-^{35}S)dATP (1000 Ci/mmol) and 2 ul (1.5 u) of a 1:8 dilution of Sequenase T7 DNA polymerase version 2.0, diluted in 10 mM Tris.HCl (pH 7.5), 5 mM DTT and 0.5 mg/ml bovine serum albumin. The solution was mixed thoroughly avoiding formation of bubbles, then briefly centrifuged and kept at room temperature for 3 min. Finally 3.5 ul of the labelled mixture were transferred to the termination mixtures and the reactions performed as described above.

3.5.3. Sequencing of ssPCR products by incorporation labelling.

Single-stranded templates were synthesised by asymmetric PCR amplification and purified by one of the methods previously described. Sequencing reactions were performed using either Sequenase T7 DNA polymerase or *Taq* DNA polymerase enzymes. In the Sequenase T7 DNA polymerase based reactions, 7 ul of the purified PCR products and 1 pmol of sequencing primer in a 10 ul reaction mixture containing 40 mM Tris.HCl (pH 7.5), 20 mM $MgCl_2$ and 50 mM NaCl were annealed by heating to 65^0C for 2 min and then allowed to cool down to below 30^0C over a period of 20 min. Thereafter the rest of the reactions were carried out as described above in the sequencing of the double-stranded templates.

When using *Taq* DNA polymerase, the purified PCR products were resuspended in 20 ul dH_2O and 10 ul of which were mixed with 1 pmol of a sequencing primer in a 13 ul reaction mixture containing 50 mM Tris.HCl, 7 mM $MgCl_2$ (pH 8.8). Annealing was accomplished by heating the mixture at 70^0C for 2 min, then cooled to below 30^0C over a period of 20 min. The annealed template-primer was briefly centrifuged, put on ice, and to it were added 2 ul of 1.5 uM each dCTP, dGTP, dTTP, 0.5 ul (α-^{35}S)d ATP (1000 Ci/mmol) and 2 ul (2 u) of *Taq* DNA polymerase diluted from 5 u/ul to 1u/ul in 25 mM Tris.HCl (pH 8.8), 0.1 mM EDTA, 0.15% Tween 20 and 0.15% Nodidet P-40. The mixture was mixed thoroughly and then incubated at 45^0C for 5 min to label the template. At the end of the labelling incubation, 4 ul of the labelled reactions were transferred to 4 ul of dideoxy termination mixtures and incubated at 70^0C for 5 min. The dideoxy termination mixes all contain 50 mM Tris.HCl (pH 8.8), 7 mM $MgCl_2$, and 20 mM of each dNTP (dATP, dCTP, dGTP and dTTP). In addition, individual mixes contain the following: A-mix 800 mM ddATP, C-mix 400 mM ddCTP, G-mix 60 mM ddGTP and T-mix 800 dTTP. After 5 min at 70^0C, the samples were cooled to room temperature and the reactions stopped by adding 4 ul of 95% formamide, 20 mM EDTA, 0.05% bromophenol blue and 0.05% xylene cyanol FF and mixing thoroughly. The samples were then briefly centrifuged and denatured at 70^0C for 5 min. After denaturation the samples were quickly placed on ice and immediately loaded and electrophoresed on polyacrylamide sequencing gels. Samples labelled with ^{35}S can be stored at -20^0C for up to one week with little degradation.

Sequencing gels contained 7 M urea, 8% acrylamide and were prepared by dissolving 25.3 g urea and 16 ml of 30% acrylamide solution (acrylamide: *N,N'*-methylenebisacrylamide at 19:1) in 1X TBE buffer (89 mM Tris, 89 mM borate, 2 mM EDTA pH 8.3) to a final volume of 60 ml. The urea was dissolved by stirring at room temperature or by briefly warming in a water bath at 50^0C in a fume hood. The gel solution was then filtered through a 0.45 micron mesh (millipore) filter and degassed under strong vacuum for 10 min. The gel was moulded in a Sequi-Gen$^{(R)}$ nucleic acid sequencing cell (Bio-Rad) consisting of an upper buffer chamber and a glass plate bound together with a permanent adhesive at the sides and the bottom edges, an outer plate and a

pair of clamps, thus obviating the need of using tape for sealing. Prior to casting the gel, both the bound and the outer plates were cleaned with warm soapy water and rinsed with deionised water. The plates were then dried and polished with ethanol. The bound glass plate was siliconised in a fume hood and allowed to dry. On drying, spacers were inserted between the two plates and held together by the side clamps. The bottom of the sequencing cell was sealed by pouring 10 ml of the gel solution into a beaker and adding 50 ul each of 25% APS and TEMED in that order. This was mixed by swirling and quickly poured onto and completely saturating the sealing paper strip in a provided casting tray. The sequencing cell was then immediately placed on the paper strip allowing the gel solution to move up between the glass plates by capillary action. The sequencing apparatus was held upright in this position for 2 min until the thin film of gel at the bottom completely polymerised thus sealing the squencing cell. To the remaining gel solution were added 90 ul of 25% APS and 90 ul TEMED, then mixed and poured between the sealed plates of the sequencing cell. After inserting a comb the sequencing cell was rested inclined at an angle of about 5^0 and the gel allowed to polymerise over a period of 30 min, when the comb was removed and the wells immediately washed with 1X TBE buffer using a pasteur pipette. The casting tray and sealing paper strip were removed and the gel left to age for at least 3 h. In most cases the gels were conveniently prepared in the evening and left to age in running buffer overnight, covered with Saran wrap to prevent drying. Such gels can be kept for up to 20 hr before running.

Prior to loading of the samples, the gels were pre-run at approximately 2000 V for 30 min to stabilise the gel temperature at 55^0C, or alternatively the pre-run was done for 1-2 hr at a gel temperature of 45^0C, monitored by gel temperature indicator strips placed on the outer glass plate of the sequencing cell. The sequencing samples were then denatured at 70^0C for 5 min. Immediately before the end of the denaturation the wells were once again cleaned with 1X TBE buffer to get rid of leached urea. At the end of denaturation the samples were placed on ice and 1.5 ul of each were immediately applied on the hot gel in the order of T, C, G and A. The electrophoresis was conducted at a gel temperature of 50^0C for 2-4 h depending on the size of the

DNA fragment to be sequenced. Routinely for convenience, staggered loading was performed, whereby the samples were loaded at different intervals such that one batch was run for 2 hr and the other for 4 hr on the same gel.

At the end of electrophoresis the glass plates were gently prised apart with a blade and the gel, usually stuck to the outer glass plate, was fixed for 10 min in 1 L 10% methanol/10% acetic acid. After 15 min the gel on the glass plate was gently removed from the fixative and the excess fixative drained or carefully wiped with paper towels. The gel was then lifted off the plate using Whatman 3MM paper and covered with Saran wrap. Air bubbles and wrinkles were smoothed out by gently rolling with a 10 ml plastic pipette. The edges were then trimmed with a scalpel to fit the slab gel drier and the gel dried at 80^0C for 30 min with the Saran wrap side up. When dry, the Saran wrap was removed and the gel directly exposed to Kodak X-Omat XAR5 film for 16-18 hr at -70^0C with an intensifying screen.

3.5.4. Variations in sequencing protocols.

Depending on the need, the sequencing protocols and gel running conditions mentioned above were modified from time to time. This depended on factors like the need to sequence closer or further from the sequencing primers or on encountering regions with strong secondary structures leading to unresolved compressions on sequencing gels.

To emphasise and read sequences close to sequencing primers with Sequenase T7 DNA polymerase, two approaches were used. In the first method, the labelling reaction was done in the presence of reduced concentrations of dNTPs (0.75 mM instead of 1.5 mM each dCTP, dGTP and dTTP) and its duration strictly limited to 3 min. In addition, the quantity of the templates was doubled, thus in this case the purified PCR templates were resuspended in 10 ul and the whole amount used in the sequencing reactions. Alternatively, the sequencing reactions were done in the usual manner except that in addition, 1 ul of Mn^{++} buffer (0.15 M sodium isocitrate, 0.1 M $MnCl_2$) was added just before the addition of Sequenase T7 DNA polymerase (Tabor and Richardson, 1989). The gels were run for 2.5 h under the same conditions described above. These methods highlight 200 nucleotides or less from the sequencing primers. To read sequences of about 400 nucleotides beyond sequencing primers, the concentration of dNTPs was increased to 7.5 uM for each dNTP and 2 ul of (α-^{35}S)dATP (1000 Ci/mmol) were used in the labelling reaction. In addition, the labelling reactions were increased to 5 min. The samples were resolved in wedge gels using spacers that were thin at the top (0.25 mm) and thicker at the bottom (0.4 mm). Such gels were run at a gel temperature of 40^0C for 6.5 h, fixed in 10% acetic acid/10% methanol for 20 min and dried for 45 min.

Compressions were resolved by at first doing termination reactions at 45-50^0C with Sequenase T7 DNA polymerase and when still unresolved, switching to *Taq* DNA polymerase. Rarely it was necessary to resort to the dNTP analogues dITP and 7-deaza-dGTP. The former was used with Sequenase T7 DNA polymerase and the later with *Taq* DNA polymerase. These analogues were used in place of the regular dGTP.

Another variation was the use of pyrophosphatase in sequencing reactions. Pyrophosphatase improves the intensity of the sequencing ladder by preventing pyrophosphorolysis which occurs naturally with Sequenase T7 DNA polymerase (Tabor and Richardson, 1990). To prevent this 0.0003 u of pyrophosphatase was added for every unit of Sequenase T7 DNA polymerase. This was either added immediately before the addition of Sequenase T7 DNA polymerase enzyme or by directly mixing 0.33 ul of pyrophosphatase in 10 mM Tris.HCl, 0.1 mM EDTA and 50% glycerol with every 1 ul of the Sequenase T7 DNA polymerase stock solution.

3.6. Secondary structure analysis.

PEPTIDESTRUCTURE and PLOTSTRUCTURE computer programmes were used to predict for alterations in secondary structure and hydropathy likely to be caused by the detected mutations (Wolf et al., 1988). Changes in the secondary structure were determined from analyses of two-dimensional plots of the mutant protein against the normal PBG-D. Hydrophilicity changes were determined for each mutation by comparing the mean hydrophilic value at and around the substituted amino acid with normal values (Chou and Fasman, 1978). Crystallographic studies were kindly performed by Professor Blundell after aligning the sequences of the PBG-D from various species including the human and *Escherichia coli.* Highly conserved amino acids were defined as those amino acids that were present at the same residue position in all species for which their sequence has been described. These include, *Escherichia coli* (Jordan et al., 1988), rat, (Beaumont et al., 1988; Blundell, 1991; personal communication) mouse (Beaumont et al., 1989) bacillus (Blundell, 1991; personal communication) *Euglina glacilis* (Sharif et al., 1989) and human (Raich et al., 1886; Grandchamp et al., 1987; Chretien et al., 1988).

CHAPTER FOUR: RESULTS.

4. RESULTS.

4.1. Direct sequencing of dsPCR amplified DNA templates.

4.1.1. Sequencing with ^{32}P end-labelled primers.

Before its application on clinical cases, the direct sequencing of PCR amplified products was first performed on various templates including DNA fragments of the plasmid pBR322, the dystrophin gene, bacteriophage lambda and the gene responsible for cystic fibrosis. This was necessary for the establishment, development and optimisation of various procedures used in the detection of mutations. Initially the study concentrated on the amplification and sequencing of double stranded DNA templates. Early experiments were performed on plasmid, where a 500 bp pBR322 DNA fragment was amplified through a standard polymerase chain reaction and the dsPCR products directly sequenced using one of the amplification primers, 5' end-labelled with ^{32}P. Purification of the templates prior to sequencing was carried out by spun-column chromatography. In the sequencing reactions, template-primer annealing was performed by denaturing the template-primer mixture at 100^0C for 2 min and then snap-freezing at -70^0C for 30 s. Using either linearised or supercoiled pBR322 the expected 500 bp PCR product was obtained as shown in figure 8a.

Purification of the PCR products with spun column chromatography gave inconsistent results in terms of DNA recovery and primer removal. For instance when using spun column chromatography for desalting and separation of primers from DNA templates, it is recommended to collect the first eluate (Sambrook et al., 1989b). In the current study, however, the appropriate products were sometimes recovered in the second or even the third eluate. The template-primer separation was also not always effective as in some cases the products were eluted along with the amplification primers even in the first eluate. Moreover, on some occasions the sephadex columns collapsed with the subsequent loss of the entire samples. Figure 8b shows 500 bp fragments of spun column purified PCR amplified DNA of pBR322 and dystrophin gene after two, three and four rounds of elution.

Sequencing of dsPCR amplified pBR322 DNA purified by spun columns persistently failed. In most cases the results showed only a dark background without any sequencing ladder. To circumvent this, attempts were made at producing single-stranded templates by gel separation as shown in figure 9a and 9b.

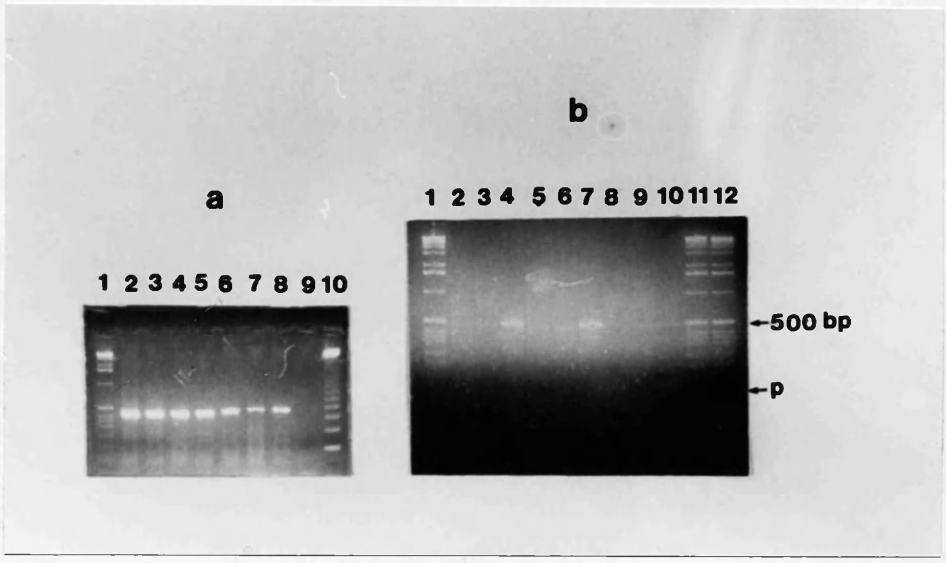


Figure 8: Spun-column chromatography in the purification of PCR products.

8a: Lane 1, 1 kb DNA ladder, lanes 2-5, pBR322 and 6-8, 500 bp fragment of the dystrophin gene. Lanes 9 and 10, negative control and 123 bp DNA ladder respectively.

8b: Lanes 1, 11 and 12, 1 kb DNA ladder, lanes 2 and 3, purified PCR amplified products of dystrophin DNA collected as second spun column eluates. Lane 4, purified PCR amplified pBR322 DNA also collected as a second spun column eluate. Lanes 5-6 and 8-9, third and fourth eluates collected from purified PCR amplified dystrophin products whereas lanes 7 and 10 are third and fourth pBR322 eluates. In lane 10 amplification primers (p) can be seen as well.

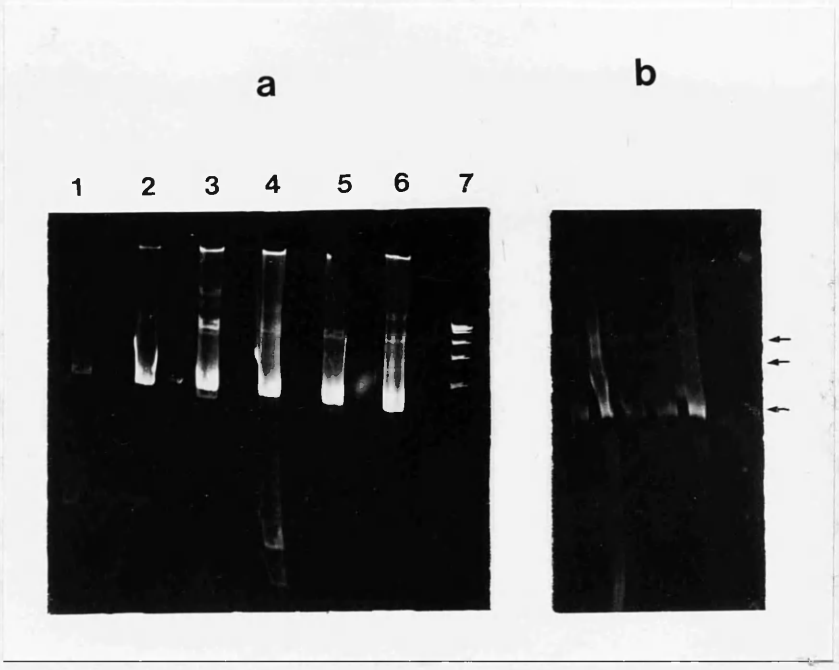


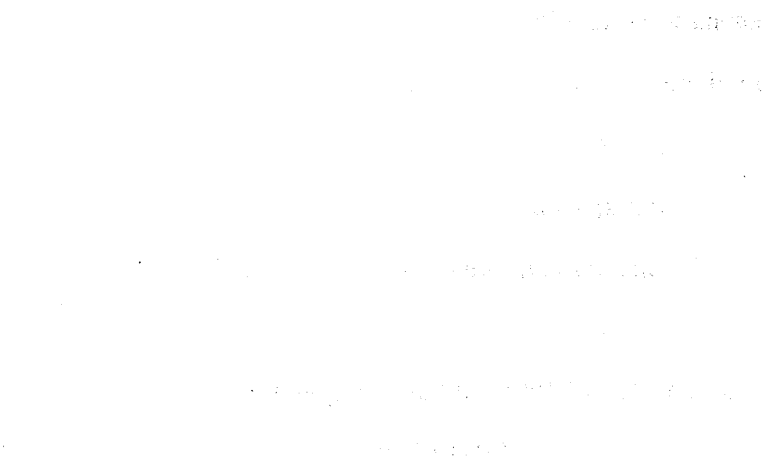
Figure 9: Strand separation gel electrophoresis.

9a: PCR amplified DNA fragments. Lanes 1 and 2, from dystrophin gene and 3-6, pBR322. Lane 7, 1 kb DNA ladder.

9b shows that following strand separation gel electrophoresis, three main bands were visualised. The bottom band was presumably due to the undenatured DNA and the top two bands represented the separated DNA strands.

4.1.2. Incorporation sequencing using ^{35}S.

Following the failure of sequencing dsPCR products using ^{32}P end-labelled primers, internal labelling was attempted. To achieve this, a 500 bp long segment of bacteriophage lambda DNA was amplified and sequenced using two different approaches. In the first approach, the dsDNA templates and sequencing primers were annealed after the templates were denatured by boiling at 100^0C for 2 min in the presence of the primers and immediately frozen at -70^0C for 30 s to prevent renaturation. In the second approach, template denaturation was performed by alkaline treatment and the template-primer annealing carried out by heating the mixture to 65^0C and slowly cooling to below 30^0C. Prior to sequencing reactions, the PCR products were purified by NACS columns either directly or after an initial fractionation in agarose gels. In order to determine the optimal template:primer ratio, different quantities of DNA were used as sequencing templates. This ranged from 50% of a PCR mixture to two reactions. Figure 10a shows dsPCR amplified products of bacteriophage lambda DNA and 10b and 10c sequencing ladders generated from these products.



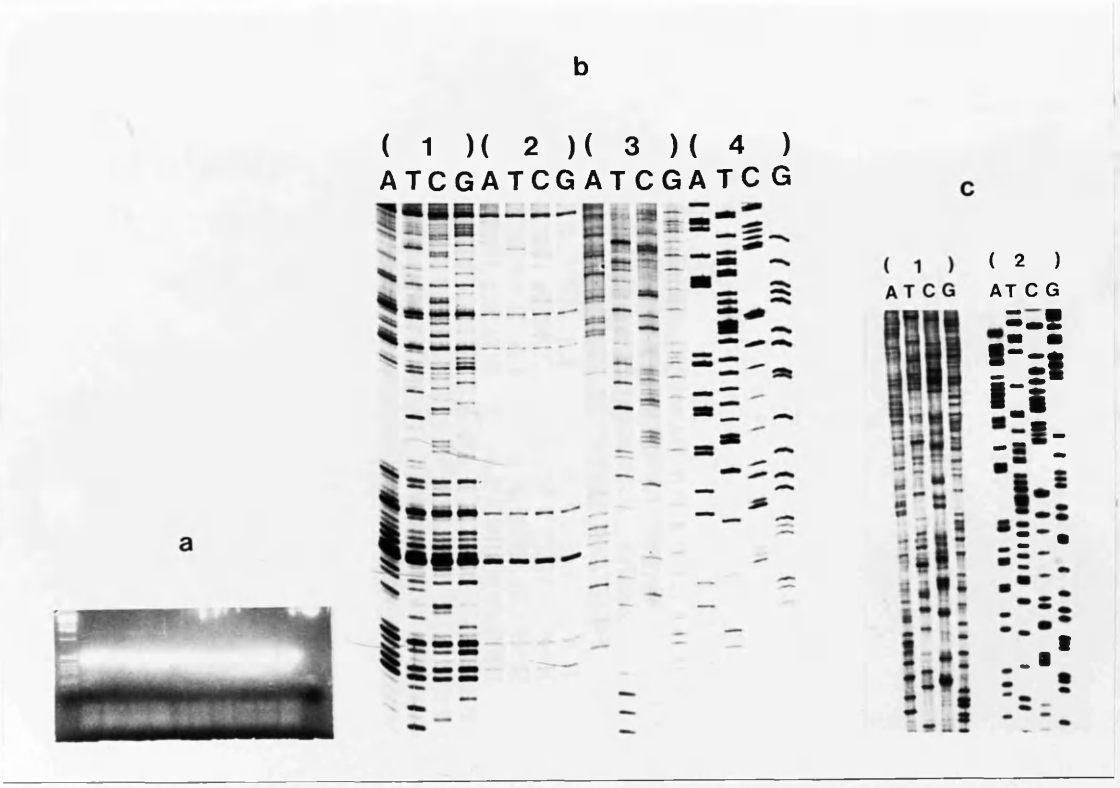


Figure 10: Sequences from bacteriophage lambda dsDNA templates.

10a: Double-stranded PCR products derived from amplification of bacteriophage lambda DNA.

10b: Sequence set 1, sequencing ladder derived from products combined from two polymerase chain reactions (200 ul), purified directly by NACS columns. Set 2, sequencing ladder derived from one PCR (100 ul) after both gel and NACS column separations and 3, from one PCR (100 ul) directly purified by NACS column chromatography. Sequence set 4, sequencing ladder of the control M13mp18 cloned DNA.

10c: Sequence set 1, sequence ladder generated from 50% of PCR products purified directly using NACS column and set 2 from M13mp18 DNA.

4.2. Direct sequencing of ssPCR amplified DNA templates.

Single-stranded PCR products were generated through asymmetric polymerase chain reaction by amplifying a 493 bp fragment of cystic fibrosis gene spanning across the $G_{1,784}$->A (G551D or Gly 551 to Asp) mutation. This was performed on an individual with a heterozygous G551D mutation and on a known normal control. The PCR products were purified by selective precipitation in 2 M ammonium acetate and propan-2-ol. Sequencing was performed with *Taq* DNA polymerase using 50% of each purified PCR product and 1 pmol of the limiting primer as a sequencing primer. The sequencing reactions were stagger loaded and electrophoresed for 2 and 3 h. Both the sense and anti-sense strands were generated and sequenced. It was possible to unequivocally demonstrate the putative mutation in the cystic fibrosis carrier as shown in figure 11.

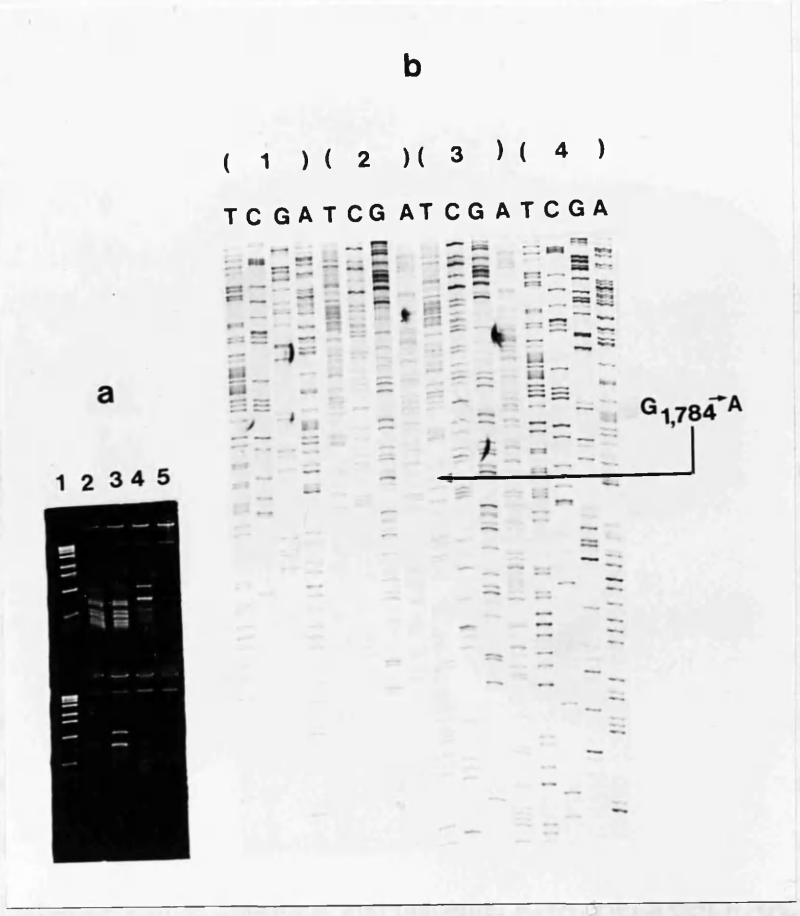


Figure 11: Partial sequences of cystic fibrosis gene demostrating the G1,784->A mutation.

11a: Lane 1, 1 kb DNA ladder, lanes 2-5, top row PCR products with forward primer limiting and bottom row with reverse primer limiting. Asymmetric PCR tends to generate a lot of spurious products as seen here. These are presumably drop-off products, homologous to the target products and do not interfere with sequencing reactions.

11b: Sets 1 and 2, sequencing ladders from a cystic fibrosis carrier of a $G_{1,784}$->A mutation and 3 and 4 from a normal control. Sets 1 and 4 were run for 2 h whereas 2 and 3 for 3 h.

The generated sequences of the individual heterozygous for the G551D mutation (with the lower case letters representing intronic and the capitals exonic sequences) is as follows:

5'- aatgcagatg caatgttcaa aatttcaact gtggttaaag caatagtgtg

atatgattac attgcaagga agatgtcctt tcaaattcag attgagcata

ataaaagtga ctctctaatt ttctattttt gctaatagGA CATCTCCAAG

1730 TTTGCAGAGA AAGACAATAT ATTCTTGGAG AAGGTGGAAT CACACTGAGT

1780 GGAGG/ATCAAC GAGCAAGAAT TTCTTTAGCA AGgt

4.3. Application of direct sequencing of PCR products in the detection of PBG-D gene mutations.

4.3.1. PCR amplification and 'one-step' direct sequencing of PBG-D cDNA.

Prior to the handling of the clinical samples from AIP patients, experiments were performed to determine the optimal conditions required for the synthesis, PCR amplification and direct sequencing of PBG-D cDNA. These optimisation experiments were carried out on various sources of RNA including human chorionic and placental RNA, HeLa cell RNA, and total cellular RNA from lymphocytes of normal subjects. Both double- and single-stranded PCR amplified cDNA were produced and sequenced. Using primers F2 and R2 which flank a 1,108 bp region including the entire PBG-D cDNA, reverse transcription and PCR amplification was performed according to Grandchamp et al (1989c) whom had kindly supplied us with the primers. On following the recommended conditions a 500 bp PCR product was obtained instead of the expected product measuring 1,108 bp. The amplification of this 500 bp product was persistent and was seen in all samples irrespective of the RNA source. This 500 bp product was amplified from human chorionic, placental and blood lymphocyte RNA as well as from HeLa cell RNA. In addition it was obtained from both crude and caesium chloride grade RNA preparations. Furthermore, this product was also persistently amplified when PBG-D cDNA

synthesis and amplification was performed through RT-PCR primed with either oligo$(dT)_{12-18}$ or F2 (forward primer) as described under methods.

Single-stranded cDNA was produced by 'asymmetric' R-T PCR whereby synthesis of the first strand cDNA was performed by priming with 50 pmol of F2, a forward primer complementary to the PBG-D m RNA. After reverse transcription, different PCR amplifications were performed by the addition of 0.5 pmol, 0.66 pmol and 1 pmol of R2, a reverse primer complementary to the PBG-D cDNA, giving F2:R2 ratios of 100:1, 75:1 and 50:1 respectively and 40 amplification cycles performed. Results of both asymmetric and standard PCR with the amplification of the 500 bp product are shown in figure 12a. Both the single- and double-stranded products were subsequently directly sequenced.

A 'One-step' direct sequencing was performed at first on purified double-stranded PCR products. Following R-T PCR amplification, the PCR products were desalted and concentrated by centrifuge-driven dialysis method using Centricon 30 (Amicon). The purified products were then annealed to the sequencing primers F3 and R3 by heating to 100^0C for 2 min and immediately placing on dry-ice ethanol bath at -70^0C for 30 s. Sequencing was performed with Sequenase T7 DNA polymerase, whereby termination reactions were carried out at 50^0C, instead of being done at the usual temperature of 37^0C. Sequencing ladder obtained from this experiment was faint and at some places ambiguous as shown in figure 12b. Partial sequence of the open reading frame read as follows:

3'-ATCACCCTTC TAAATATCCA TCTCCGCTGT TTGGAATGGC TCGGACCACT

ATCGCCAACA GGTTCTATCT TAGAATCAAG TTGAAATTTA AACGTGTTC –5'.

This as well, did not correspond to the already published sequences of either the erythroid or the non-erythroid PBG-D cDNA (Raich et al., 1986; Grandchamp et al., 1987).

To improve the results obtained from sequencing of dsPCR amplified templates, single-stranded products were used instead. Samples were purified by centrifuge-driven dialysis method using Centricon 30 (Amicon) microconcentrators. Sequencing was done with 1 pmol of the limiting

primer R2, using Sequenase T7 DNA polymerase and carrying out termintion reactions at 37^0C. To determine the optimal template:primer ratio for the annealing reaction, 30, 40, 50 and 75% of the PCR products were sequenced with 1 pmol of the sequencing primer. The results of these sequencing reactions are shown in figure 13.

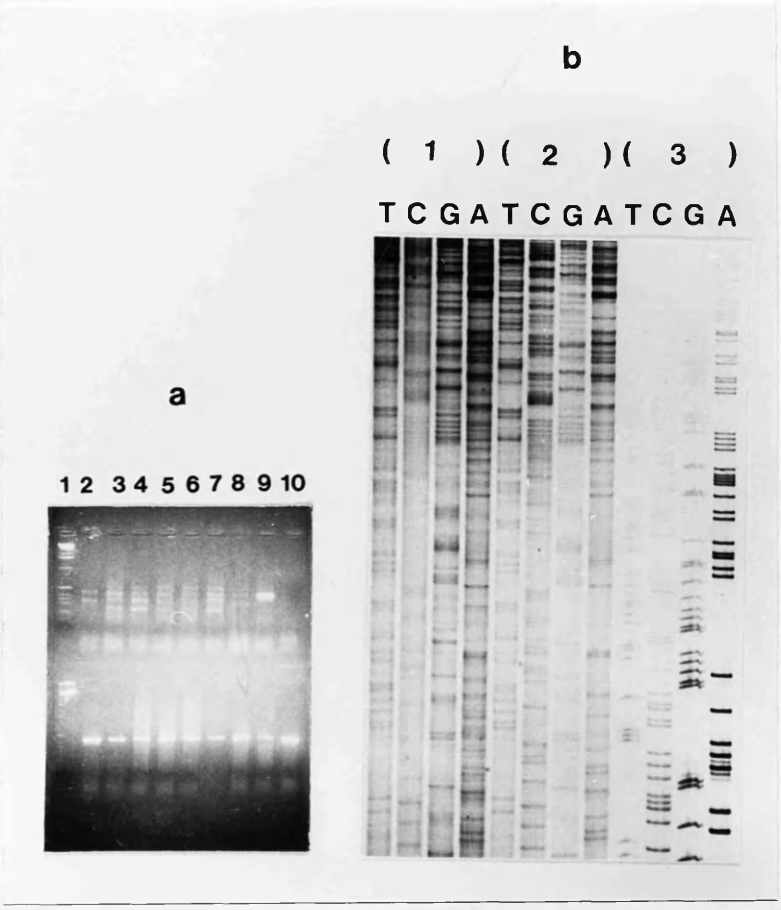


Figure 12: 'One step' direct sequencing of dsPBG-D cDNA

12a top row: lane 1, 1 kb ladder, lanes 2 and 8, R-T PCR performed with an F2:R2 ratio of 100:1, lanes 4 and 7 with an F2:R2 ratio of 75:1 and lanes 5 and 6 with an F2:R2 ratio of 50:1. In all these cases the first strand cDNA synthesis was carried out by priming with F2. Lane 3, R-T PCR amplification using oligo$(dT)_{12\text{-}18}$ with an R2:F2 ratio of 100:1 whereas lane 9 with equal amounts of F2 and R2 and lane 10, negative control.

12a bottom row: Lane 1, 1 kb ladder and lanes 2-10, standard R-T PCR of PBG-D cDNA.

12b: Sequence set 1 and 2, PBG-D cDNA sequenced with nested primers F3 (5 pmol) and R3 (10 p mol) respectively. Sequence set 3, M13mp18 control DNA.

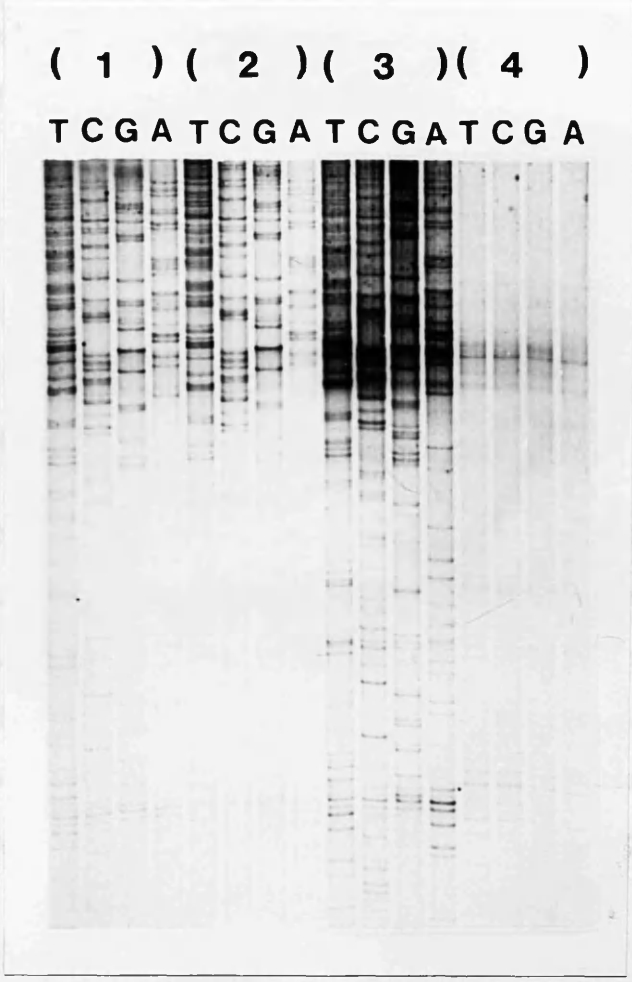


Figure 13: One step direct sequencing of ssPBG-D cDNA.

13: Sequence sets 1-4, Single-stranded PBG-D cDNA sequenced with 1 pmol of the limiting primer R2, using 30, 40, 50 and 75% of PCR products respectively. Sequencing reactions were performed in the presence of Sequenase T7 DNA polymerase. The sequence ladders generated from this and the previous experiment were rather faint, implying inadequate DNA templates in the one step direct sequencing of PCR amplified products.

Sequencing ladders generated from direct sequencing of ssPCR products were much improved and easier to read than those from dsPCR. Despite these improvements, there were however, still some regions in the sequencing ladders with bands appearing in all four lanes, albeit to a lesser extent when compared with sequencing of dsPCR products. The sequence obtained from this one step direct sequencing method using single-stranded templates was similar to that produced from double-stranded PCR products shown above and read as follows:

3'————AT ACATCGTTTT ATCACCCTTC TAAATATCCA TCTCCGCTGT
TTGGAATGGC TCGGACCACT ATCGACCAAC AGGTTCTACT TTAGAATCAA
GTTGAAATTT AAACGTGTCT CTGGAGA –5'

R-T PCR was repeated using total cellular RNA which had been previously treated with RNAase free DNAase. This still yielded a 500 bp product and on sequencing with an internal sequencing primer R3, the same results as above were obtained. To circumvent this the annealing step in PCR was raised to 60^0C and the amplification and subsequent sequencing of the products were performed in two steps. At first RT-PCR was performed to synthesise cDNA which was then reamplified to produce single-stranded templates suitable for sequencing.

4.3.2. Asymmetric reamplification and sequencing of PBG-D cDNA.

Optimisation experiments were done using HeLa cell and human placental RNA by altering various conditions starting from the reverse transcriptase reaction to the eventual PCR. PCR amplified products of expected size, were eventually obtained by using new primers (F1 and R1) and raising the annealing step to 60^0C as described under methodology. Primers R1 and F1 flank a region 1,150 bp in length which includes the non-coding region of the PBG-D gene thus enabling amplification of the entire cDNA. Even under these new conditions there was still, in some cases amplification of the 500 bp band. The products were therefore, fractionated in agarose gels and the appropriate band excised. The cDNA was then eluted and reamplified by asymmetric PCR. Alternatively, amplification of the right sized products was possible by amplifying the cDNA with primers F3 and R2 or with primers F2 and R3. These primers amplify overlapping cDNA products and therefore span the entire PBG-D cDNA. Figures 14a to 14d show results of these PCR optimisation experiments.

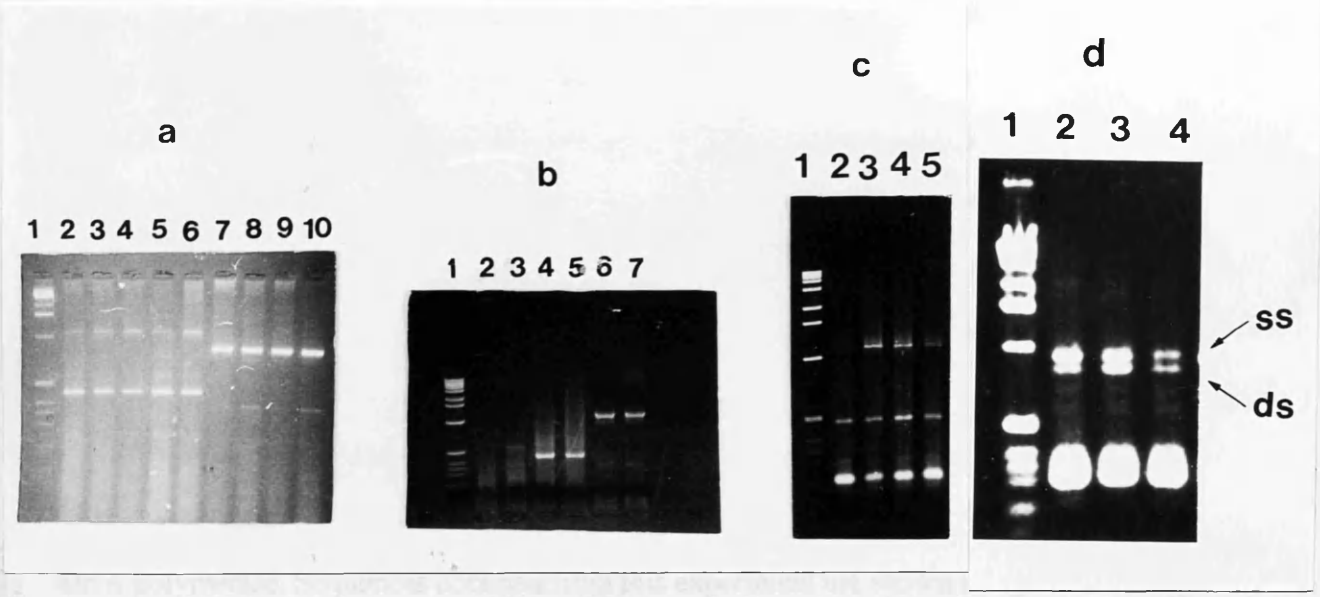


Figure 14: Optimisation for PCR conditions.

14a: Lanes 1, 1 kb ladder, Lanes 2-6, R-T PCR amplification with primers F2 and R2 and lanes 7-10 with primers F2 and R3. The amplification conditions were as described in the methodology with the annealing temperature being at 60^0C.

14b: Lane 1, 1 kb ladder, lanes 2-5, R-T PCR amplification with primers R2 and F2. In lanes 2 and 3 the amplication was done by an asymmetric PCR, whereas in 4 and 5 by a standard PCR. Lanes 6 and 7, standard R-T PCR with primers F1 and R1. In lanes 1-5 annealing was performed at 55^0C and 6 and 7 at 60^0C.

14c: R-T PCR amplification with primers F1 and R1 in which annealing was performed at 55^0C.

14d: Asymmetric reamplification of products in figure 14b lanes 6 and 7, using internal primers F2 and R3 demonstrating presence of single-stranded cDNA.

Having optimised for the PCR conditions, RT-PCR was performed on various templates including HeLa cells, human placenta cells, lymphocytes and lymphoblastoid cells of normal individuals. The R-T PCR was done with primers F1 and R1 and the cDNA eluted as described previously. The eluted cDNA was then reamplified with internal primers F2 and R3 in an asymmetric PCR and the products purified by different methods to compare their efficacy. Figure 15a demonstrates PCR products obtained from R-T PCR using primers F1 and R1 and figure 15b, the results of asymmetrical reamplification of these products with internal primers F2 and R3. The single-stranded PCR products obtained from the reamplification of the PBG-D cDNA were purified either by selective precipitation in 2 M ammonium acetate and propan-2-ol or by centrifuge-driven dialysis using Centricon 30 (Amicon). After purification, 50% of each PCR was sequenced with either the limiting primer F2 or the internal primer F3, using Sequenase T7 DNA polymerase. Sequences obtained from this experiment are shown in figure 15c. These sequences correspond with the published sequences of the PBG-D cDNA.

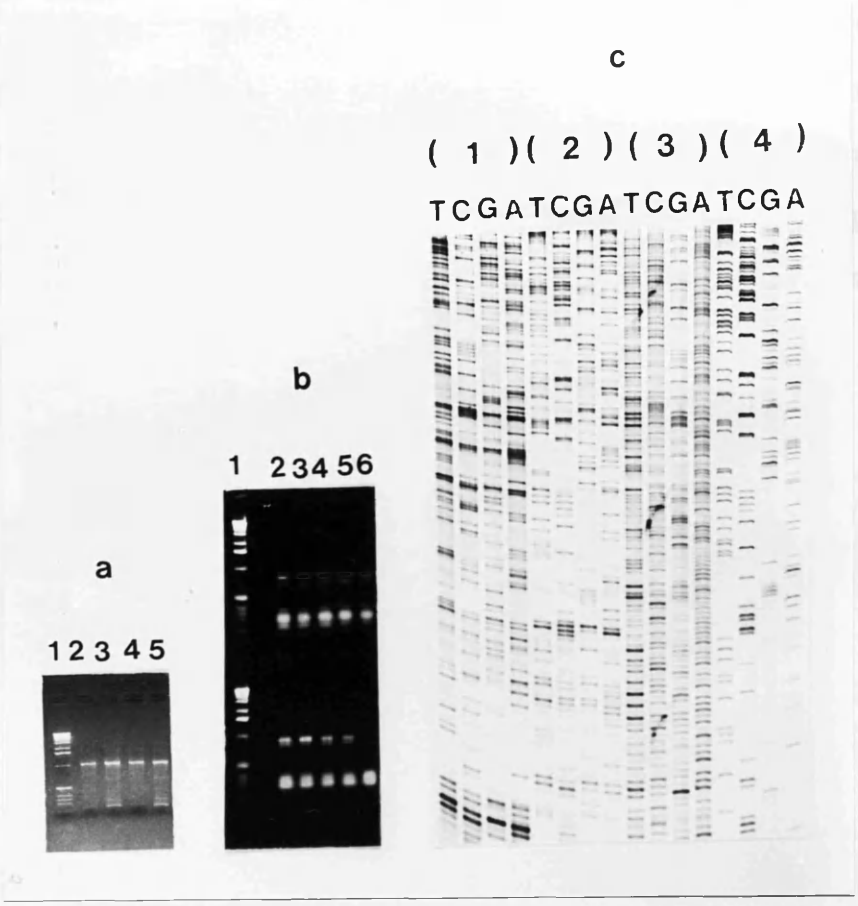


Figure 15: A two step amplification and sequencing of PBG-D cDNA.

15a: Double-stranded PCR products obtained from amplification with primers F1 and R1.

15b: Products of asymmetric PCR amplification of PBG-D cDNA. The original products in figure 15a above, were reamplified with nested primers F2 and R3. Top row, are products produced with F3 as the limiting primers whereas, in the bottom row, the limiting primer was R3.

15c: Sequence sets 1 and 3, sequences of PBG-D cDNA from HeLa cell RNA and lymphocytes of a normal control respectively, following Centricon 30 microconcentrator purification. Sequence sets 2 and 4, sequences obtained from HeLa cells and normal lymphocytes after the purification of PCR products by selective precipitation with ammonium acetate and propan-2-ol.

Set 1 and 3 were sequenced with the limiting primer F2 whereas sets 2 and 4 with the internal primer F3.

Sequencing of PCR products from 15b above, was repeated using *Taq* DNA polymerase as a sequencing enzyme. Sequencing was done with primer F3 following the purification of PCR products by selective precipitation in 2 M ammonium acetate and propan-2-ol. Sequences generated from this experiment are shown in figure 16. These experiments in principle established and standardised the protocol that was used for the detection of mutations of PBG-D gene in patients with AIP.

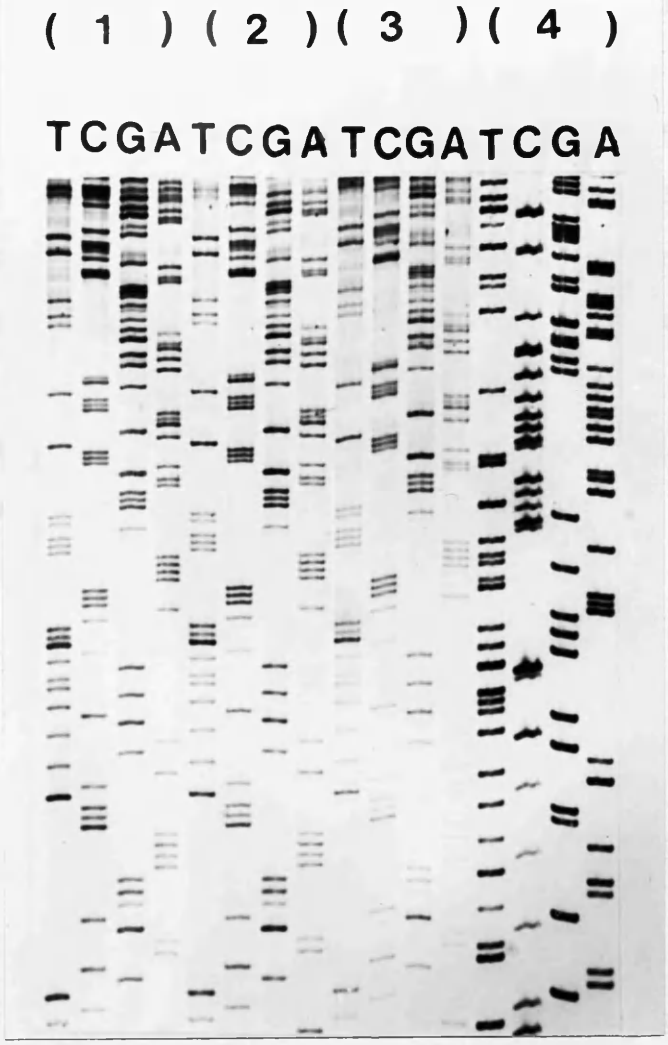


Figure 16: Sequencing of PBG-D cDNA with Taq DNA polymerase.

Set 1-3, Sequences of PCR amplified PBG-D cDNA from human placental RNA, normal control lymphocyte RNA and HeLa cell RNA respectively. Set 4, M13mp18 control DNA. The products were all purified by precipitation in ammonium acetate and propano-2-ol and sequenced with primer F3 in a *Taq* DNA polymerase mediated reaction.

4.3.3 Direct sequencing of PBG-D cDNA from crude cell lysate preparations.

Following the successful sequencing of PBG-D cDNA, modifications were introduced in order to simplify the protocol used for the detection of mutations. These included the use of crude preparations of buccal cells and fresh blood lymphocytes as sources of RNA. Crude RNA preparations were made by lysing buccal cells from mouth washings and direct amplification of PBG-D cDNA attempted from the cell lysates. This approach was however, unsuccessful despite several attempts. 20 ml of venous blood were then collected from volunteers and lymphocytes separated. The lymphocytes were lysed and PBG-D cDNA amplified from crude RNA preparations using primers F1 and R1 as described before. Asymmetric reamplification of the cDNA with internal primers, provided templates suitable for sequencing. In addition asymmetric R-T PCR was performed directly on the cell lysate using primers F1 and R1 and sequenced successfully with internal primers as shown in figure 20.

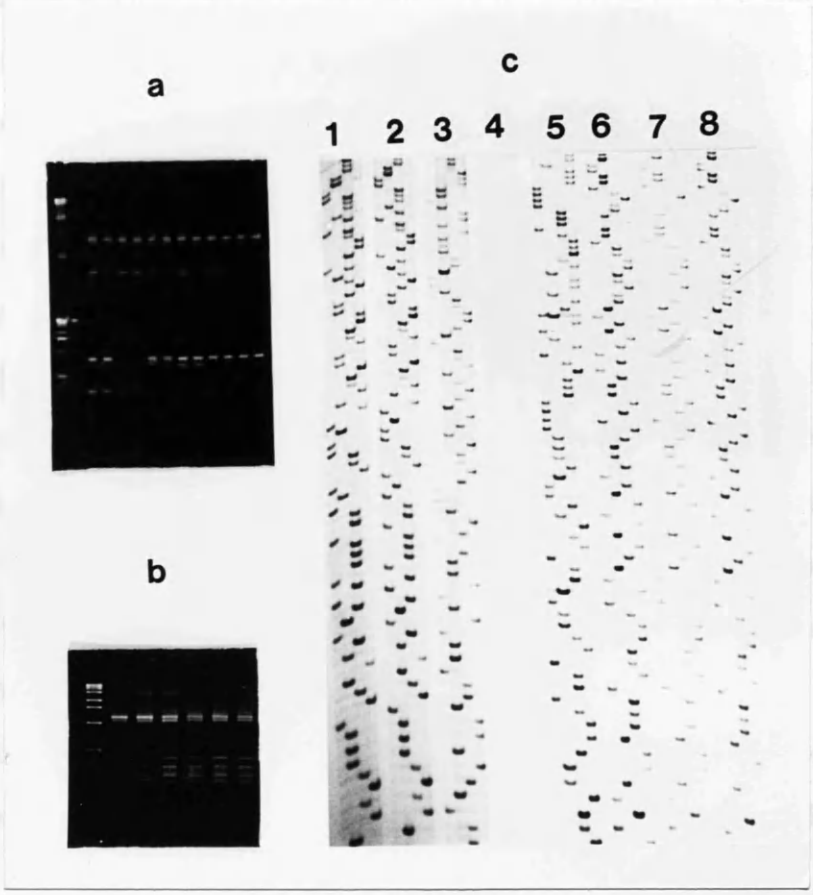


Figure 17: Direct sequencing of PBG-D cDNA derived from uncultured cells.

17a: Asymmetric PCR of PBG-D cDNA derived from uncultured lymphocytes amplified with primers F3 and R3. The top row is with F3 limiting and the bottom row with R3 limiting.

17b: Asymmetric PCR products amplified directly in a R-T PCR using primers F1 and R1.

17c: Sequence sets 1-3, sequences generated from uncultured lymphocytes as shown in figure 17a bottom row, in which sets 1 and 2 were sequenced with primer R4 and set 3 with R5. Sequence set 4, one step sequencing of PCR products amplified by asymmetric R-T PCR and sequenced with a nested primer F3 without reamplification of the cDNA. Sets 5-8, sequences generated from lymphoblastoid cell lines, in which set 5 was sequenced with F3 and the rest with R3.

4.3.4. Problems associated with direct sequencing of PBG-D cDNA.

In the course of the development and optimisation of the protocol for PCR amplification and direct sequencing of PBG-D cDNA, several problems worth mentioning were encountered. Some of these problems would occur from time to time even after the establishment and standardisation of the protocol, mainly during the sequencing reactions. The most commonly encountered problems were the inability to amplify and sequence a particular strand of cDNA and the inability to accurately interpret some results due to the presence of ambiguous sequencing ladders caused by the appearance of the same sequencing bands in all four lanes.

In some cases asymmetric PCR amplification seemed to work for one strand and not the other, making it difficult to sequence a particular cDNA strand even after optimal conditions had been defined. On the other hand, sequencing of a particular cDNA strand was occasionally found to be poor even after what appeared to have been a very successful PCR amplification. These problems are illustrated in figures 18a to 18c.

Sequence ambiguity due to the presence of the same bands appearing in all lanes was encountered in several templates. There were two main types of such bands as shown in figures 19 and 20.

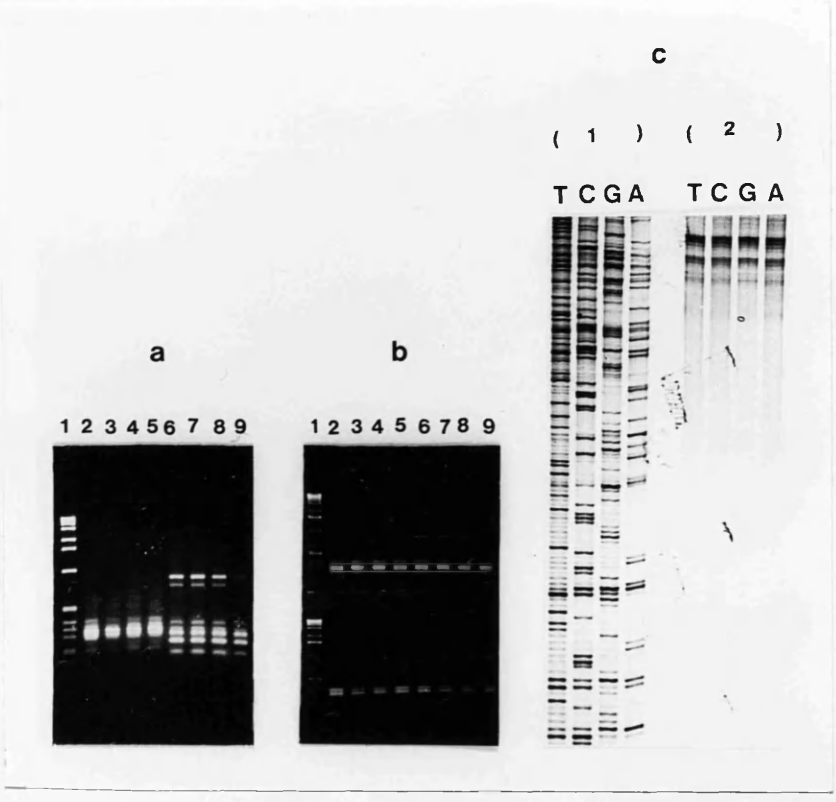


Figure 18: Strand specific problems associated with direct sequencing of PCR amplified products.

18a: Lane 1, 1 kb DNA ladder. Lanes 2-5, asymmertic PCR products amplified with primers F2 and R3 in which R3 was the limiting primer and in lanes 6-9 the limiting primer was F2. PCR amplification in lanes 2-5 is very much inferior when compared with that in lanes 6-9 though they were from the same individual and performed at the same time using the same conditions. 18b: Lanes 1 top and bottom rows, 1 kb DNA ladder and lanes 2-9 top row, asymmetric PCR products amplified with primers F2 and R2 in which F2 was limiting, whereas in the bottom row R2 was limiting. Lanes 6-9 bottom row show the presence of more single stranded products then 6-9 in the top row. This is confirmed in the sequencing ladders shown in 18c.

18c: Sequencing set 1, sequence of PBG-D cDNA in figure 18b bottom row lane 6, sequenced with primer R2. Set 2, sequencing of products from lane 6 of the top row, sequenced with F2. In both cases the sequencing primers were the limiting primers used in asymmetric PCR amplifications. Set 1 also depicts ambiguous sequences due to the presence of multiple bands in several lanes.

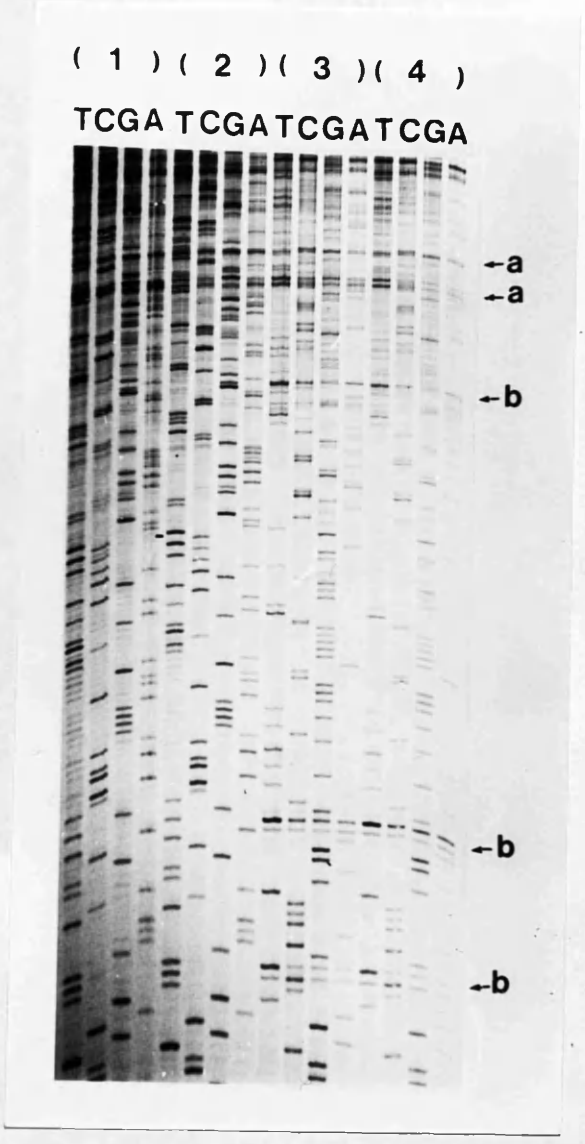


Figure 19: Sequencing artefacts.

Sequence sets 1 and 2, AIP patient and a normal control respectively, sequenced with primer F3. Set 3 and 4, the same individuals sequenced with primer F5. Arrows (a) point out bands which appeared in all tracks, whereas (b) bands were unique to those templates sequenced with primer F5. The gel was run for 3 h at the temperature of 50^0C.

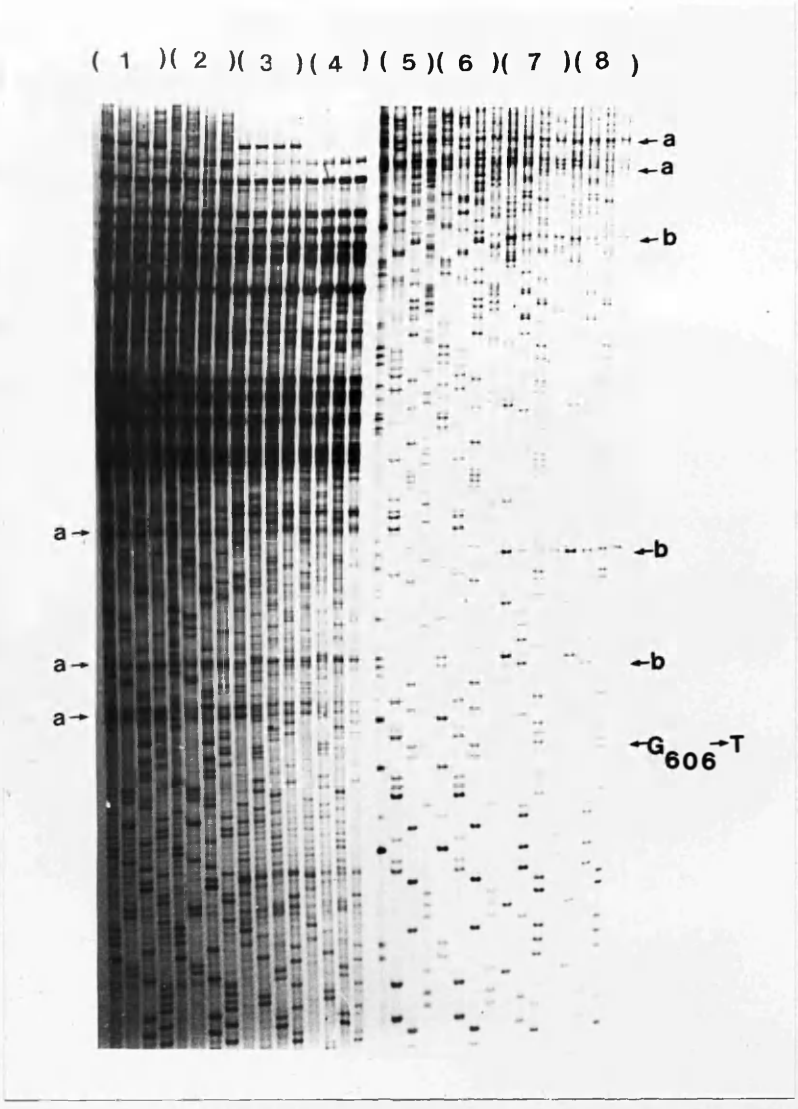


Figure 20: Effect of gel temperature on sequencing artefacts.

Sequencing products in figure 19, were run in a different gel at a running temperature of 60^0C. Sets 1-4 were run for 4 and sets 5-8 for 2 h. Sets 1 and 2, AIP patient and normal control respectively, sequenced with primer F3, whereas 3 and 4, were sequenced with primer F5. Sets 5-8, sequencing products from sets 1-4 respectively, electrophoresed for 2 h. Both types of artefacts (a and b) were persistent in this case, even when the gel was run at a higher temperature of 60^0C. A substitution of G to T at nucleotide position 606 is seen in set 7.

4.3.5. Detection of mutations in PBG-D gene in patients with AIP.

4.3.5.1. RNA extraction and quantitation.

Total cellular RNA was extracted from lymphoblastoid cell lines of 30 patients with AIP and 10 controls. In the clinical cases and controls, total cellular RNA was prepared by extraction with guanidium thiocyanate and equilibrium centrifugation through caesium chloride. The preparations were run in RNA checker gels to determine the integrity of the RNA whereas concentration and purity were determined by spectrophotometry. The mean total cellular RNA concentration was 650 ug/ml and the range 128-1400 ug/ml. The purity as determined by the mean value of OD_{260}/OD_{280} was 1.98. Results of fractionation of the RNA preparations are shown in figures 21a and 21b.

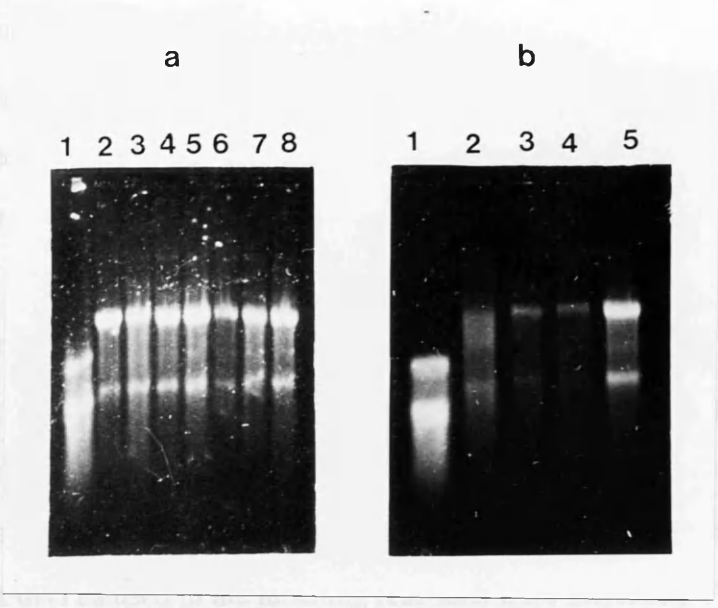


Figure 21: RNA electrophoresis of AIP patients and controls.

21a and 21b: Lane 1, 1 ug of *Escherichia coli* ribosomal RNA used as a marker whereas the remaining lanes in 21a, are RNA samples from AIP patients and 21b from normal controls. The top band in the samples from patients and controls represent 28S rRNA and the bottom one 18S rRNA while the smear denotes the presence of mRNA.

4.3.5.2. Delineation of molecular pathology in AIP.

Total cellular RNA from affected individuals and control subjects were amplified with primers F1 and R1 in a R-T PCR to produce cDNA. In all cases the correct sized fragments were amplified thus excluding any major deletions or truncated mRNA products. The cDNA was then reamplified in asymmetric PCR using primers F2 and R2. Sequencing was done with limiting or nested primers as described in the methods. Sequence changes were determined by comparing with normal control sequencing ladders and confirmed by showing the same change in the opposite strand.

To sequence the entire PBG-D cDNA, at least four different sequencing primers were used in addition to the application of various techniques which highlight different regions of the sequencing ladder. Highlighting of sequences close to the sequencing primers was achieved by using manganese buffer whereas to facilitate sequencing of regions distant from the primers, the concentrations of dNTPs used in the labelling reactions were increased. Figure 22, demonstrates sequencing of the entire PBG-D cDNA in a single experiment, where sequence sets 3, 4, 5 and 8 where from the same individual.

Figure 22: Sequencing of the entire PBG-D cDNA in one experiment.

Sequence set 3 was sequenced with primer R4 using a labelling mix containing 7.5 uM each of dCTP, dGTP and dTTP and 20 uCi of (α-^{32}P)dATP. The reactions were electrophoresed for 4 h. Sequence set 8 was also sequenced with primer R4, except that standard concentrations of dNTPs in the labelling mix were used as described in the methods. In addition, the sequencing

reactions were performed in the presence of manganese buffer and electrophoresed for 2 h only. Sets 4 and 5, were sequenced with primers R5 and F4 respectively and electrophoresis carried out for 4 h. In this screening experiment, Set 3 shows the G_{606}->T mutation in the anti-sense strand.

Using the above method, the PBG-D cDNA was rapidly scanned in both AIP patients and normal controls. Seven mutations summarised in table 1 and shown in figures 23-29 were demonstrated in nine patients with AIP and none in the control subjects. The Q34K and L177R mutations were each seen in two individuals whereas, H256N, R167Q, L42L and S45S were noted in one individual each and V202V in four. All mutations were due to single base substitutions. Four mutations namely Q34K, R167Q, L177R and H256N were associated with amino acid substitutions whereas in the remaining three (L42L, S45S and V202V) there were no changes. Three of the mutation namely L177R, H256N and S45S were associated with alteration in the recognition sites for restriction enzymes. Mutations L117R and H256N lead to the abolishing of recognition sites for the enzymes *Alu*I and *Ban*I respectively whereas the silent mutation S45S abolishes the recognition site for *Rsa*I.

Table 1: Mutations of the PBG-D gene, detected in AIP patients in the current study.

Mutation	Exon	Amino acid change
C_{100}->A (Q34K)	4	Glutamine to lysine.
G_{500}->A (R167Q)	10	Arginine to glutamine.
T_{530}->G (L177R)	10	Leucine to arginine.
C_{766}->A (H256N	12	Histidine to asparagine.
G_{117}->A (L42L)	4	None.
G_{135}->A (S45S)	4	None.
G_{606}->T (V202V)	10	None.

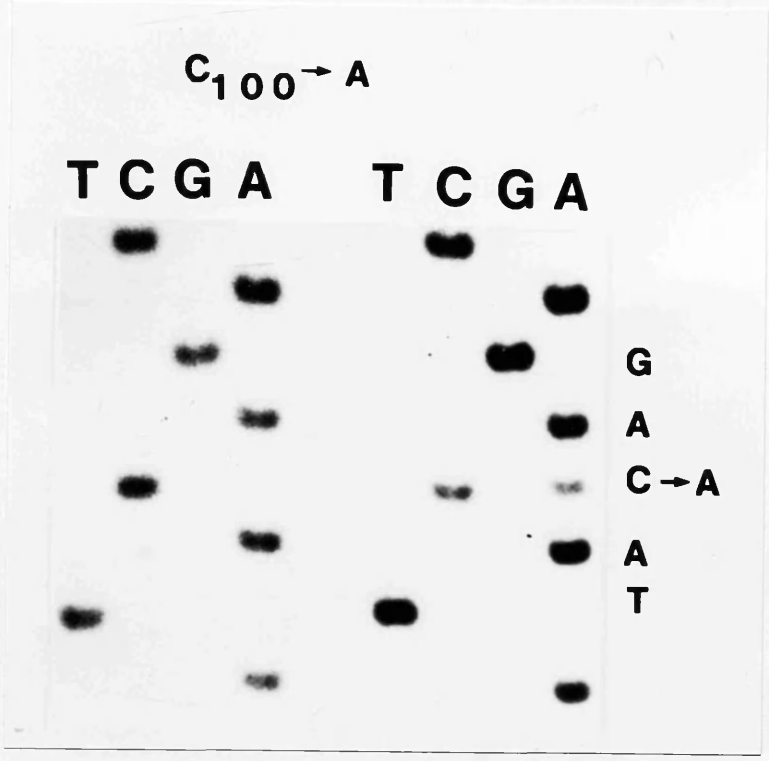


Figure 23: Q34K mutation causing the substitution of glutamine to lysine.

Partial sequencing ladders of the sense strands, comparing sequences from a normal control on the left and a patient with acute intermittent porphyria on the right. The wild-type sequence reads, ATACAGAC, whereas the mutant sequence is, ATAC/AAGAC, denoting a C to A transversion at nucleotide position 100 in exon 4. This is demonstrated by the appearance of both alleles i.e. bands C and A at the same position. The mutation leads to the substitution of a highly conserved amino acid glutamine to lysine at the amino acid residue position 34 (Q34K). In addition, by using PEPTIDESTRUCTURE and PLOTSTRUCTURE computer programmes, it has been predicted that, this change is likely to cause a disruption of the PBG-D β-strand. The mutation was observed in a mother and her daughter.

Figure 24: R167Q mutation causing the substitution of arginine to glutamine.

Partial sequences of exon 10 of the PBG-D gene. Mutant sequences of the sense strands are shown on the left and the wild type on the right. In the section shown, the normal sequence reads, AACTCCAAAGGGGCTTACGA whilst the mutant reads, AACTCCAAAGGGG/ACTTCGA because of a G to A transition at nucleotide position 500. This change causes the substitution of the highly conserved amino acid arginine to glutamine at position 167 (R167Q) of the protein structure. In the figure, it can be observed that in the mutant sequence the mutated A band has greater intensity then the normal G band. To date this is the only mutation of PBG-D that has been reported separately by more than one investigator. This mutation was seen in one patient.

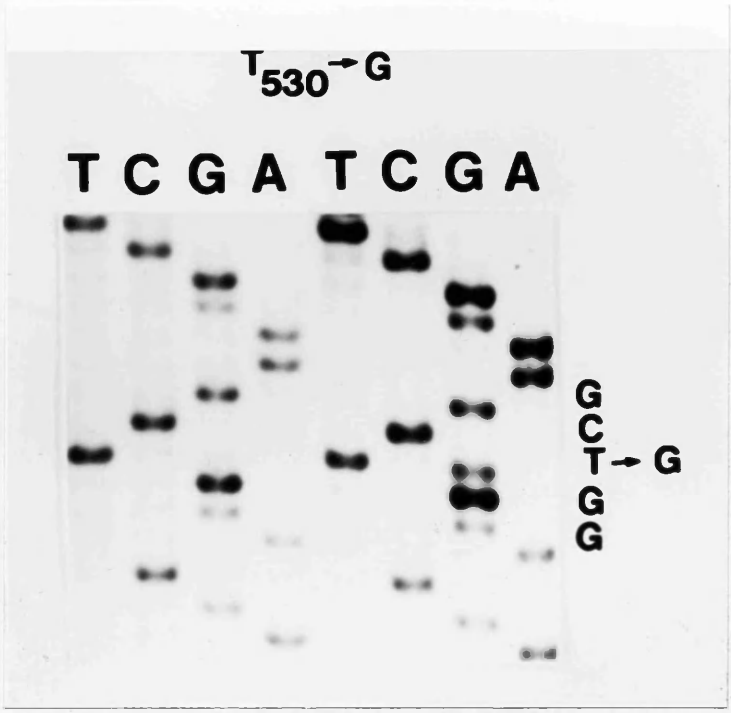


Figure 25: L177R mutation causing the substitution of leucine to arginine.

The normal sequence, AGCAGGTCGAAGGCT shown on the left is part of exon 10 of the PBG-D gene. The mutant sequence, AGCAGGT/GCGAAGGCT on the right arises from a T to G transversion at position 530 as demonstrated in the figure, by the presence of both G and T bands at this position. This results in the substitution of a highly conserved amino acid, leucine to arginine at amino acid residue position 117. The change, which was observed in two unrelated AIP patients, also abolishes a recognition site for the restriction enzyme *Alu*I which can be used to track this mutation within this family.

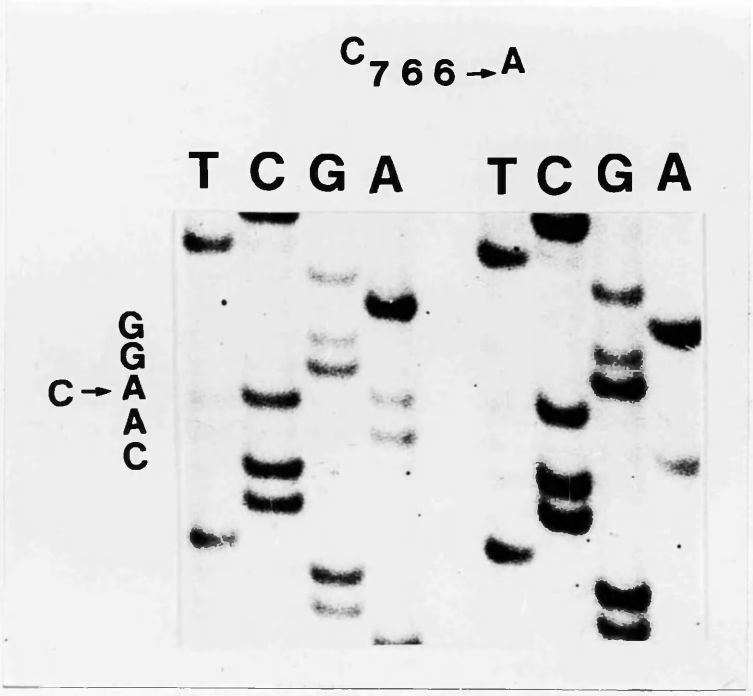


Figure 26: H256N mutation causing the substitution of histidine to asparagine.

Partial sequences of exon 12 of the PBG-D gene. The sequence on the right is from a patient with AIP showing a C to A transversion at nucleotide position 766. The normal sequence on the left reads, GGTCCACGGAGTC whilst the mutant reads, GGTCCAC/AGGAGTC. This mutation substitutes the amino acid histidine for asparagine at the amino acid residue position 256, altering a conserved charge. In addition the change abolishes a recognition site for the restriction enzyme *Ban*I which may be used to track this mutation in this family. The change was observed in one patient.

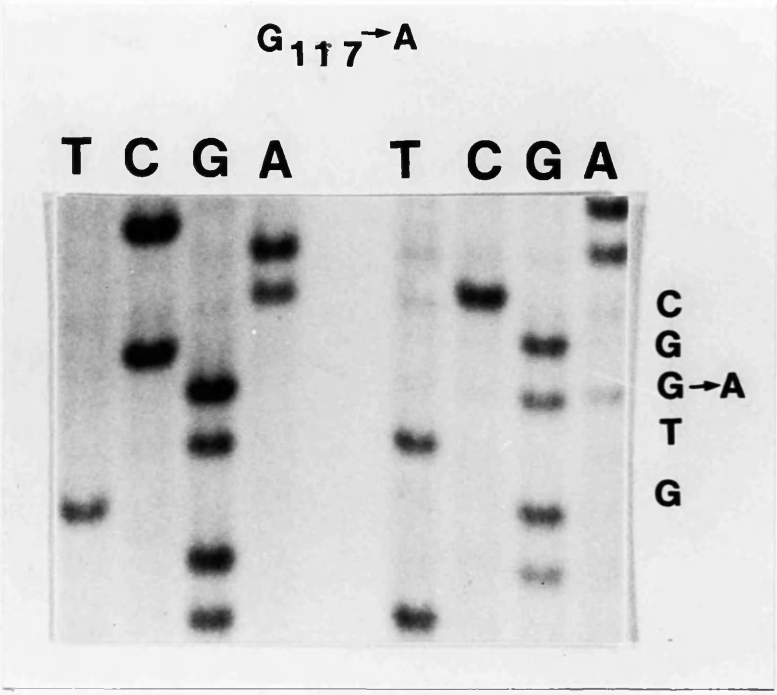


Figure 27: Silent mutation L42L.

Partial sequences of exon 4 of the PBG-D gene, showing a silent mutation at nucleotide position 117. The sequence on the left, GGTGGCAAC is from a normal control and the one on the right from an AIP patient. The mutant sequence reads, TGGTG/AGCAA. This mutation, which was observed in one individual, does not result in any change, either in the amino acid pattern or recognition sites for restriction enzymes.

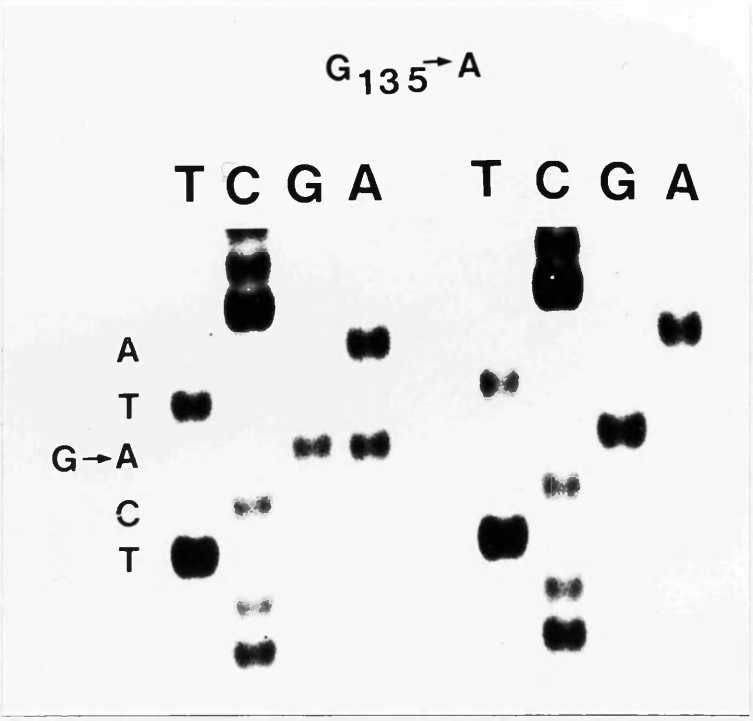


Figure 28: Silent mutation S45S.

The partial sequence on the left is from a patient with AIP with a G to A transition at position 135 in exon 4. The wild-type sequence on the right reads, CCTCCCGTACCC, whereas the mutant is, CCTCG/ATACCC. This mutation which was observed in one subject abolishes a recognition site for the restriction enzyme *Rsa*I, which may be useful in linkage studies for tracking gene carriers within this family. Since the mutation changes codon TCG (UCG) to TCA (UCA), both coding for the same amino acid serine, it does not cause any amino acid substitution.

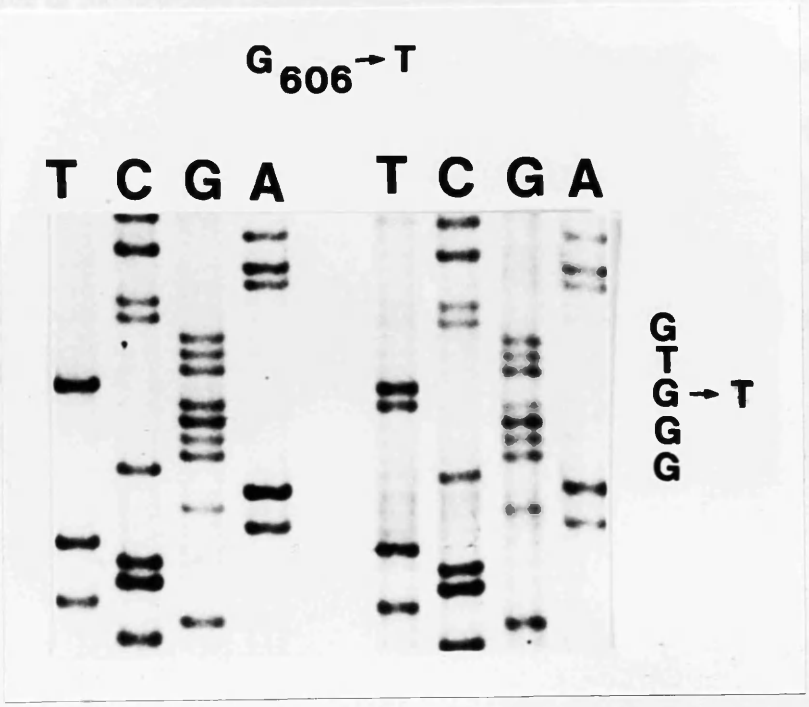


Figure 29: Silent mutation V202V.

Partial sequences of exon 10. The sequencing ladder on the left is the normal and on the right a mutant sequence from a patient with AIP. The normal sequence is, CGTCCTAGACGGGGTGGGCCAACAC whereas the mutant sequence reads, CGTCCTAGACGGGG/TTGGGCCAACAC. This alters codon GTG (GUG) to GTT (GUU) both coding for the same amino acid valine. This mutation, therefore causes no change in the amino acid pattern, and moreover, does not alter any recognition sites for restriction enzymes. The mutation was seen in four patients with AIP.

4.3.6. Analysis of secondary structure alterations resulting from the detected PBG-D mutations.

Alignment of the mutated amino acid sequences with normal sequences of other species revealed that three of the detected mutations affected highly conserved amino acids and one a conserved charge. In addition, crystallographic studies kindly performed by Prof. Blundell (Birbeck College, University of London) revealed that these mutations were likely to result in either, structural or functional aberrations. These results are summarised in table 2. In the table, due alignment of the human PBG-D with sequences of different deaminases from other species, the numbering of the amino acid residues differs from that reported by Grandchamp et al (1987). The number in brackets refer to the original amino acid residue position in the human non-erythroid PBG-D.

Table 2: Analysis of the PBG-D structural and functional changes likely to arise from the detected mutations.

Residue	Structural environment	Possible structural or functional change
Gln19(34)*	Located at base of active site cleft and involved in hydrogen bonding to conserved residues Ser81(96) and Arg176(191)	Change to Lys disrupts hydrogen bonding network, especially in two nearby positively charged residues
Arg149(167)*	Forms direct salt bridge to acetyl side group of ring 2 of cofactor. The position occupied by this ring may become the binding site for the incoming PBG	Change to Gln results in loss of stabilising interaction to carboxylate group of the cofactor, and possibly of the substrate
Leu159(177)*	Forms part of hydrophobic core of the molecule	Change to Arg introduces a destabilising effect of a charged group within the apolar core.
His237(256)	In *E. Coli* protein, Arg237 forms a salt bridge to Glu292. His237(256) and Asn292(322) may form a corresponding hydrogen bond in the human PBG-D	Change to Asn, results in a pair of Asn which may be less able to form hydorogem bond

*Highly conserved amino acid.

CHAPTER FIVE: DISCUSSION.

5. DISCUSSION.

During the course of developm ent and optimisation of direct sequencing of PCR amplified products in the characterisation of PBG-D mutations, it became clear that the two most important factors were the template quality and quantity. These were in turn, influenced mainly by the efficiency of PCR amplification and the subsequent purification of the templates. In the case of RNA amplification, the quality of the RNA as a starting material also mattered.

5.1. Direct sequencing of dsPCR amplified DNA templates with end-labelled primers.

Direct sequencing of PCR amplified products was first performed on ds templates. In these experiments DNA was amplified from 500 bp segments of the DMD gene, the plasmid pBR322 and bacteriophage lambda. In all cases after optimisation for the PCR conditions, the correct sized products were amplified. The amount of DNA amplified as estimated by ethidium bromide fluorescence was judged to be adequate as shown in figure 8a. Prior to sequencing, the products were extracted in chloroform to remove traces of mineral oil used in the polymerase chain reactions, then passed through Sephadex G-50 (Pharmacia) spun columns for concentration and removal of salts, excess dNTPs and amplification primers. Sequencing was performed with the same amplification primers end-labelled with ^{32}P. The primers were annealed to the templates by heating the primer-template mixture to 100^0C for 2-5 min and snap freezing in a dry ice-ethanol bath for 30 s. In some cases following the heat denaturation, the templates were annealed on wet ice at 0^0C for 30 s. The annealed primer-template mixture was then immediately used in sequencing reactions to discourage template renaturation. These early attempts at sequencing invariably failed but were useful in identifying the main pitfalls, which were template purification and primer-template annealing.

Purification of the PCR templates aims at removing the excess amplification primers, dNTPs and salts used in the amplification reactions because of their interference in sequencing reactions. The presence of the excess amplification primers interferes with the annealing and the extension steps in the sequencing reactions. This occurs because the unlabelled amplification primers may

either compete directly with the radiolabelled sequencing primers for binding sites on the template DNA or generate extension products that efficiently exclude the radiolabelled sequencing primers. It is also necessary to remove all dNTPs used in the PCR, so that they do not serve as substrates in the subsequent chain-termination reactions and hence interfere with sequencing reactions. The presence of salts also tend to interfere with sequencing reactions by inhibiting sequencing enzymes. Although the removal of excess dNTPs, amplification primers and salts have been attempted by separation through Sephadex G 50 (Pharmacia) spun columns (Newton et al., 1988), in the current study this method was found to be unsatisfactory. Both the separation of amplification primers from templates and the template recovery were found to be inconsistent. For example, when using spun-column chromatography, it is recommended that the eluted DNA be collected during the first round of centrifugation (Sambrook et al.,1989b), but it was found in the current study that this was not always applicable. The DNA recovery was variable from one experiment to another, and maximum recovery could be anywhere between the first and the the fourth round of elution as shown in figure 8b. Furthermore these experiments revealed that the method was associated with poor recovery of DNA as evidenced by its presence even during the fourth round of elution. Thus after collecting the first or even the second round eluate, there was still a significant amount of DNA trapped in the columns. In addition, primer removal was not always efficient, thus defeating the whole purpose of the purification. Another practical problem associated with this method of purification was the tendency of the columns to collapse, leading to loss of samples. Although this was infrequent, such losses can be catastrophic when precious samples are involved.

Despite of the fact that successful sequencing of dsPCR products has been reported by several authors (Wrischnick et al., 1977; Wong et al., 1987; Engelke et al.,, 1988; Higuchi et al., 1988a; Newton et al., 1988), this approach can still be problematical. This is due to the fact that the protocols currently available for preparation of double-stranded templates for sequencing were developed for covalently closed circular plasmids (Chen and Seeburg, 1985; Zagursky et al., 1965; Lim and Pene, 1988). Although satisfactory results are obtained for supercoiled plasmid

DNA, rapid reannealing of short, linear, dsDNA templates as produced by PCR is a significant obstacle to sequencing of DNA templates by this approach (Kretman and Landweber, 1989). This problem is compounded by the fact that these protocols, adapted for sequencing of dsPCR products do vary considerably (Wong et al., 1987; Wrischnik et al., 1987; Newton et al., 1988; Higuchi et al., 1988; Saiki et al., 1988; Kretz et al., 1989; Winship, 1989; Bachmann et al., 1990). For example in the direct sequencing of β-globin gene, Wong et al., (1987) denatured the templates by heating at 95^0C for 10 min and performed the primer-template annealing at 37^0C for 2 min at a primer:template ratio of approximately 7:1, whereas Saiki et al. (1988) while investigating the same gene, performed the denaturing at 95^0C for 10 min and annealing at 0^0C for an unspecified duration. On the other hand Newton et al., (1988) while sequencing the α-antitrypsin gene denatured the templates by boiling at 100^0C for 5 min and and immediately placed the primer-template mixture at -70^0C for 5 min. In 1990, Casanova et al., determined the optimal set of conditions for the primer-template annealing. They found out that for best results, the primer-template mixture should be heated to 100^0C and immediately placed at -70^0C for 15-45 s, using a primer:template ratio of 20:1. Among these factors, the duration of the annealing reaction is probably the most crucial parameter. Casanova et al. (1990) were able to show that sequences were readable only when the duration of the annealing reaction was carried out for 0-45 s and that within that range of time, sequencing band intensity increased with time. Sequences became unreadable when primer-template annealing was done for 60 s or more. Similarly sequences were difficult to read when the primer:template ratio was 2:1 or less and non-specific sequencing bands appeared when the ratio exceeded 200:1.

In order to compare efficacy of different methods of denaturation of templates prior to the primer-template annealing, templates were also denatured by alkali treatment. The templates were denatured in 0.2 M NaOH for 5 min, either at room temperature or at 37^0C, put on ice and the reaction neutralised with ammonium acetate as described in the methods section. The DNA was then immediately precipitated by addition of 2.5 vol of ethanol and resuspended in 10 ul of dH_2O or TE buffer before being used in the sequencing reactions. The denatured DNA pellets

prepared by this method can be stored in this form for several weeks and when required dissolved in dH_20 or TE buffer just before use (Hattori and Sakaki, 1986). Despite the fact that alkali denaturation has proved to be superior to heat denaturation in the preparation of ds supercoiled plasmid templates for sequencing (Chen and Seeburg, 1985; Hsiao, 1991), it does not seem to be the preferred method in the sequencing of dsPCR amplified templates. Virtually all reports on successful sequencing of dsPCR products have been carried out using the heat denaturation method (Wong et al., 1978; Wrischnik et al., 1987; Newton et al., 1988; Higuchi et al., 1988; Kretz et al., 1989; Winship, 1989; Bachmann et al., 1990; Casanova et al, 1990). Nevertheless, the main problem associated with both approaches is the rapid reannealing of the templates following the denaturation.

To overcome this problem, Winship (1989) used DMSO. The addition of 10% DMSO in the sequencing reaction mixtures, not only improved the intensity of sequencing ladders but also reduced the background at specific positions. This reduction in the background is thought to be due to the prevention of secondary structure formation. Similar results have also been obtained by Bachmann et al. (1990) through the addition of the detergents nonidet P-40 (NP-40) and Tween 20. By denaturing templates in the presence of 0.5% NP-40 or Tween 20 or a combination of both and including the detergents in the sequencing reactions, the intensity of the sequencing ladders similarly increased and the background at specific positions decreased as well. These modifications, however, were not attempted in the current study.

Another way of improving the quality of sequences generated from dsPCR products is to end-label the sequencing primer with ^{32}P, rather than have it incorporated as a nucleotide during the polymerase-mediated chain extension in the sequencing reactions (Engelke et al., 1987; Wong et al., 1987; Saiki et al.,, 1988). This approach unfortunately requires an additional enzymatic step in the kinasing reaction and more important, obviates the use of ^{35}S with its superior base-ladder resolution (Kraitman and Landweber, 1989). In the current study, the application of spun-column purification and ^{32}P end-labelled primers for sequencing of dsPCR amplified templates was unsuccsseful. In most cases the autoradiographs showed a high background with almost no

sequencing ladder at all. To ensure that the problem did not lie in the radiolabelling of the sequencing primers, the primers were tested for the degree of isotope incorporation before use. This was done using Whatman DE-81 filters as described in the methodology. These filters being positively charged, strongly adsorb and retain nucleic acids, including oligonucleotides. Unincorporated nucleotides bind less strongly to the filters and are removed by washing with sodium phosphate. In most cases the proportion of radioactivity incorporated into the primers was well above 70%. To be effective, the sequencing primers should incorporate at least 20% of the radiolabel.

To circumvent the problems associated with sequencing of dsPCR derived templates, attempts were made to produce single-stranded templates using strand separation gel electrophoresis. Following a standard PCR, the double-stranded products were precipitated in ethanol, washed twice in 70% ethanol and the pellet dissolved in a strand separation buffer containing 30% DMSO, 1 mM EDTA, 0.05% bromophenol blue and 0.05% xylene cyanol FF. The mixture was then heat denatured and fractionated in a 5% non-denaturing polyacrylamide gel as previously described. The gel was stained with ethidium bromide and the DNA visualised under U.V. light. Results of these experiments as shown in figure 9a and 9b, indicate the presence of several bands though three of them appeared to be more prominent then the others. The electrophoretic mobility of ssDNA fragments in non-denaturing polyacrylamide gel is unpredictable because to some extent this depends on secondary structure. In figures 9a and 9b, it was therefore impossible to tell which band belonged to which strand. Moreover, the presence of multiple bands presumably due to spurious products which are otherwise undetected in agarose gels further complicated the interpretation. It was, however, assumed that the fast main band represented the non-denatured dsDNA and the two slower bands the two separated bands. The method was also found to be cumbersome and labour-intensive.

These early attempts at direct sequencing of PCR amplified products were possibly unsuccessful because of a combination of several factors. These factors where in two main areas namely, inefficient purification of the PCR products and poor primer-template annealing and

extension during the sequencing reactions. The inefficient purification was associated with poor DNA recovery and hence inadequate templates for sequencing. In addition, there was inefficient primer removal and this must have contributed to the poor results. Although not proven, it is also very likely that this inefficient purification from using spun columns might have been associated with inadequate removal of salts and dNTPs which may have also contributed to the poor results. Sequencing primer-template annealing was performed at different temperatures and for different durations some of which were sub-optimal. In some experiments the annealing was carried at 0^0C and the duration not timed. The primer:template ratio was also sub-optimal and ranged between 1:1 to 10:1. These variation reflect the diversity of protocols in use for sequencing of dsPCR products.

From these experiments, it was concluded that the most significant problem was in the purification method. To improve this, other PCR product purification methods were used and compared. In addition, in order to improve the resolution of the sequencing ladders, ^{35}S was used instead of ^{32}P.

5.2. Direct sequencing of dsPCR products using incorporation labelling.

Following the failure to sequence dsPCR products by the above methods the approach was revised. Double-stranded products produced by standard PCR were concentrated and purified with NACS columns using two approaches. In one set of experiments, the products were first fractionated in 1.5% low melting NuSieve agarose gel and the appropriate band excised. The gel slice was then melted and the cDNA eluted and purified with the NACS columns. Alternatively, following PCR, the products were directly purified in the NACS columns. NACS is an ion exchange resin that binds nucleic acids in low salt (0.1-0.5 M NaCl) and releases them in high salt solutions (0.7-2.0 M NaCl). The principle of NACS column purification is based on the fact that generally the resin binds larger polynucleotides more tightly than smaller ones. It also binds single-stranded nucleic acids more tightly than double-stranded molecules.

In sequencing reactions, prior to the primer-template annealing, templates were denatured by either alkali or heat treatment as described before. Of the two methods, heat denaturation was found to be more effective than alkali treatment. This is in contrast to the experience of Chen and Seeburg (1985), who on sequencing double-stranded plasmid templates found the opposite to be true. It is of interest, however, to point out that virtually all successful reports on sequencing of dsPCR amplified templates have been based on the heat denaturation method (Wong et al., 1978; Wrischnik et al., 1987; Newton et al., 1988; Higuchi et al., 1988; Kretz et al., 1989; Winship, 1989; Bachmann et al., 1990; Casanova et al, 1990). This may be due to the fact that, dsPCR being short and linear tend to reanneal much faster then the supercoiled plasmid templates. The optimal primer:template ratio was determined by altering the amount of the template used. This was thought to be necessary because of some of the previous reports that have indicated that one of the main problems of direct sequencing of PCR products was due to insufficient template DNA (Saha, 1989). To overcome this problem Saha (1989) recommended to either perform PCRs in 500 ul instead of the usual 100 ul reactions or to pool several PCRs prior to sequencing. In the current study different amounts of dsDNA templates ranging from 50% of a PCR to two pooled PCRs were sequenced with 1 p mol of the sequencing primer as shown in figure 10. In

the experiment depicted in figure 10, bacteriophage lambda DNA was amplified and sequenced with one of the primers used in the amplification. Figure 10b shows that the sequencing ladder generated from templates that were gel fractionated prior to NACS purification had a cleaner background than those that were directly purified by the NACS columns. It also shows higher intensity of the sequencing bands when two PCRs were pooled than when only one PCR was used in the sequencing reactions. Paradoxically, it was the templates that were directly purified by the NACS columns without any prior gel separation that gave the best results in terms of the number of readable bases. Using samples which their purification included gel separation, the sequencing bands were faint and in many areas the same bands appeared in all four lanes. The latter phenomenon often referred to as 'pile-ups', 'full stops' or 'walls' usually occurs when a sequencing enzyme especially Sequenase T7 DNA polymerase, encounters a complex secondary structure in the DNA template being sequenced. Such artefacts may also occur with bad template preparations especially if there is an excess of salt or nicking of template DNA during the purification. The complex process of gel separation, melting and passing through the NACS columns may explain this. In addition, it has been discovered that agarose contains substances that inhibit *Taq* DNA polymerase and it has therefore been suggested that agarose separation should not be used for those samples which are going to be sequenced with this enzyme (Gyllensten, 1989). This has however, been shown not to be the case with Sequenase T7 DNA polymerase by Kert et al., (1989), who were able to directly sequence PCR products from low-melting agarose, thus circumventing the need for DNA elution and further purification. The faint sequencing bands noted in the samples that were first fractionated in agarose gels compared to those that were directly purified through the NACS column may be due to inefficient elution of the DNA templates from agarose using this method. In a similar experiment depicted in figure 10c, a readable sequencing ladder was obtained from the sequencing of as little as 50% of one PCR, indicating that in a single amplification reaction, there is sufficient DNA for sequencing as long as it is processed efficiently in the course of template purification.

The quality of the sequencing ladders obtained from dsDNA was however not always satisfactory. The quality of such templates in terms of the number of bases which can be accurately read depends on the purity of the DNA templates (Murphy and Ward, 1989). Artefacts are seen more often in the sequencing of double- rather than single-stranded templates. The majority of the artefacts are likely to be caused by the greater tendency of the denatured dsDNA templates to form inter-chain cross links through reannealing. Pile-ups, often observed at several points in the sequencing ladder appear to be produced more often in the sequencing of double-stranded templates. These can often be eliminated by performing the chain extension and termination reactions at a higher temperature (42^0C up to 50^0C) or by sequencing the opposite strand (Hatori and Sakaki, 1986; Murphy and Ward, 1989). When performing the chain extension and termination reactions at such high temperatures it may be necessary to add more sequencing enzyme in a chase-mix. Alternatively these pile-ups may be eliminated by using dITP or 7-deaza-dGTP instead of dGTP in the sequencing reactions (Murphy and Ward, 1989) or by sequencing with *Taq* DNA polymerase (Innis et al., 1988; Gyllensten, 1989; Brow, 1990). If this problem persists, it may necessitate the use of ssDNA-binding protein in the sequencing reactions. Single-stranded DNA-binding protein acts by preventing the formation of secondary structures. The disadvantage of using this, is that it requires removal by digestion with proteinase K prior to electrophoresis of the samples as it causes retardation of nucleotide migration (Murphy and Ward, 1989). Template reannealing is also responsible for the high background seen in the sequencing of double-stranded templates.

In the experiments referred to in figure 10, in addition to the problems associated with the sequencing of double-stranded templates, inconveniences were noted with the use of NACS columns. As purification with NACS columns is based on gravitational force, the drip columns tend to be slow and time consuming. Practical problems such as entrapped air bubbles within the columns interfering with elution of the cDNA, was a very common occurance. This hampered the simultaneous handling of many samples. To improve on this, attempts were made to produce

single-stranded PCR templates through PCR and to find simpler and more convenient ways of purifying these templates.

5.3. Direct sequencing of ssPCR amplified templates from genomic DNA.

Single-stranded templates were produced by asymmetric PCR from various sources including a fragment of the cystic fibrosis gene, spanning across the $G_{1,784}$->A (G551D) mutation. In the case of the cystic fibrosis gene, this segment was amplified both from a known cystic fibrosis carrier and a normal control. In these experiments, the amplification primers were used at a ratio of 1:100 to amplify approximately 0.5 ug of genomic DNA. In asymmetric PCR, by using unequal amounts of amplification primers, most of the product generated during the first 20-25 cycles is double-stranded and accumulates in the usual exponential way. As the limiting primer becomes exhausted, the later cycles generate an excess of ssDNA complementary to the limiting primer. However, in contrast to the exponential growth of the dsDNA, ssDNA accumulates linearly (Gyllensten and Erlich, 1988; Kreitman and Landwber, 1988; Saiki et al., 1988c). Because of the fact that ssDNA products appear only in the later cycles of the PCR and that they increase by arithmetic progression, efficiency of asymmetric PCR is low when compared to the standard PCR (Gyllensten, 1989; McCabe, 1990). To compensate for this sub-optimal amplification it has been suggested that the number of PCR cycles be increased (Gyllensten and Erlich, 1988). In these experiments the number of PCR cycles was raised to 40 instead of 30. Alternatively, the low efficiency of the asymmetric PCR can be overcome by adding more *Taq* DNA polymerase in the late cycles of the PCR or by using more of the enzyme right from the outset.

To enable sequencing of both strands two sets of PCRs were done with reciprocal ratios of the limiting primers. The results of these amplifications are shown in figure 11a. One of the features associated with asymmetric PCR and shown in figure 11a, is the production of spurious products. Amplification of these spurious products can be minimised by using the minimum necessary amount of target DNA and dNTPs. A balance should nevertheless be considered between the

DNA yield and the presence of the spurious products. In most cases, however, the presence of these spurious products do not seem to interfere with the sequencing of the templates. This can be explained by the fact that these products have discrete 3' ends and are heterogeneous only at their 5' ends; the ends that are oriented away from sequencing primers (McCabe, 1990). Following the asymmetric PCR, the templates were purified by selective precipitation in 2 M ammonium acetate (pH 5.2) and 2 vol of propan-2-ol at room temperature as described before. Sequencing was performed by Sangers' dideoxy-method, using *Taq* DNA polymerase and limiting primers. From each PCR, only 50% of the purified products were used in the sequencing reactions contrary to other suggestions in which up to five PCRs were pooled (Saha, 1989). The primer-template annealing was performed by heating the primer-template annealing mixture to 65^0C and allowing it to slowly cool to below 30^0C. The samples were stagger-loaded, running one batch for 2 h the other for 3 h. Figure 11b demonstrates the $G_{1,784}$->A mutation in the sample electrophoresed for 3 h from the cystic fibrosis carrier. At this position the cystic fibrosis carrier has both bases G and A denoting heterozygosity while the normal control has only the expected base G. This single base substitution changes the codon GGT for glycine to GAT for aspartic acid. This change which replaces a neutral amino acid with a charged one, is the second commonest mutation in cystic fibrosis. It accounts for about 5% of of all mutations in the cystic fibrosis chromosomes in Caucasians (Cutting, et al., 1990).

This experiment illustrates one of the major advantages of direct sequencing of PCR amplified products in that it enables the demonstration of a heterozygote position in those cases where two alleles differ by a point mutation. Both alleles are are seen in the the same position in the sequencing ladder as in this case the bases G and A. This is because both alleles are amplified in the PCR and hence represented in the amplification reactions. In contrast, in the case of cloned products, several templates would have needed to be sequenced before the heterozygosity was determined. In such cases indeed both alleles need to be seen before heterozygosity can be proved. In addition several templates need to be sequenced or else the mutant allele would be missed if by chance all the clones picked for sequencing consist of the wild-type sequences.

The sequencing ladders generated from ssPCR products from the genomic DNA of cystic fibrosis hetreozygotes were much clearer and more easily read in comparison to the double-stranded templates shown in the preceding experiments. In these particular sequencing ladders, no pile-ups or high backgrounds were observed. Part of the explanation for this may have been due to the use of *Taq* DNA polymerase for sequencing of the templates. The high processivity and heat stability of this enzyme enables it to easily sequence through areas were formation of secondary structures are likely to cause problems (Innis et al., 1988).

The templates were purified simply by selective precipitation in propan-2-ol in the presence of ammonium acetate. Ammonium acetate works by inhibiting the co-precipitation of dNTPs. Two sequential precipitations of DNA in the presence of 2 M ammonium acetate for instance, results in the removal of over 99% of dNTPs from DNA templates (Okayama and Berg, 1982). The precipitation was performed at room temperature and the DNA pellet collected after a 10 min centrifugation in a bench-top centrifuge, making this purification method simple, convenient and inexpensive.

In this experiment, it was therefore possible to successfully produce and sequence ssPCR amplified templates from genomic DNA. The templates generated very satisfactory sequencing ladders free of ambiguities and it was possible to demonstrate a point mutation in a subject heterozygous for the mutation G551D in the cystic fibrosis gene. From these results it was therefore decided to adapt and apply this technique to the detection and characterisation of the mutations responsible for acute intermittent porphyria.

5.4. PCR amplification and direct sequencing of PBG-D cDNA.

PCR amplification and sequencing of the $G_{1,784}$->A mutation in the cystic fibrosis gene described above was performed on genomic DNA. This method was modified to amplify and sequence PBG-D cDNA from its mRNA. Two approaches were used. In the first approach, referred to here as 'one-step' direct sequencing, single-stranded products were amplified directly from R-T PCR. This was modified in a similar way to asymmetric PCR in that unequal molar

quantities of primers were used and hence referred to in the current study as 'asymmetric R-T PCR.' In the second approach which is referred to here as a 'two-step' direct sequencing, regular R-T PCR was performed on RNA samples and the cDNA eluted before it was reamplified by asymmetric PCR to produce single-stranded templates.

5.4.1. 'One-step' direct sequencing of PBG-D cDNA.

PBGD cDNA was produced by a process referred to as RT-PCR. In this procedure, cDNA is initially synthesised either from total cellular RNA or from mRNA by a reverse-transcriptase and then PCR amplified with *Taq* DNA polymerase. Both reactions are done in the same reaction mixture, greatly simplifying the procedure (Sherman et al., 1989; Sambrook et al., 1989a). In performing R-T PCR, three approaches are available based on the method used for the first-strand cDNA synthesis. First-strand cDNA synthesis may be accomplished by extension with random hexamers, oligo$(dT)_{12\text{-}18}$ or an amplification primer complementary to the respective mRNA (upstream primer). In the current study, the last two approaches were employed and compared.

Asymmetric R-T PCR was performed to directly produce single-stranded cDNA from RNA. To achieve this two approaches were employed. In the first approach the first-strand cDNA was synthesised by using an upstream primer, whereas in the second oligo$(dT)_{12\text{-}18}$ was used instead. Both methods worked equally well but the latter was preferred because of its simplicity and adaptability to produce either of the cDNA strands. By using an upstream primer for first-strand cDNA synthesis the only option of performing an asymmetric RT-PCR is with the downstream primer limiting. This therefore, restricts one to the amplification of the sense strand only. On the other hand, by priming the first-strand cDNA synthesis with oligo$(dT)_{12\text{-}18}$ in asymmetric R-T PCR, one can interchange limiting primers to enable amplification of either strand.

Optimisation of both the standard and asymmetric R-T PCR was performed on RNA samples from HeLa, human placenta and chorion cells. This was done, partly to evaluate the robustness of the technique and partly because these RNA samples were already available. In addition RNA

was also extracted from normal volunteers. R-T PCR amplification was performed with primers F2 and R2 kindly supplied by professor Grandchamp. The R-T PCR was performed by first heating the RNA samples at 65^0C for 10 min and placing on ice before the addition of other components. This enhances the cDNA priming step by breaking up aggregates and some secondary structures in the RNA molecules (Kawasaki, 1990). After completion of the reverse-transcriptase step, the samples were heated at 95^0C for 5 min before proceeding with the PCR amplification. The 95^0C heat treatment inactivates the reverse-transcriptase and denatures the RNA-DNA hybrids; a process that improves both the efficiency and specificity of the PCR (Kawasaki and Wang, 1989, Sambrook et, al 1989a). The PCR was performed according to Grandchamp et al., (1989c), whereby denaturation was performed at 95^0C for 1 min followed by annealing at 55^0C for 1 min and extension at 72^0C for 2.5 min. The final extension was performed at 72^0C for 7 min to ensure full strand polymerisation in the PCR products. This led to the amplification of a 500 bp fragment instead of the expected 1,108 bp product as shown in figure 12a bottom row. The amplification of this 500 bp PCR product occurred in all samples no matter the source or method of RNA extraction used. Similar results were also obtained from asymmetric R-T PCR amplifications performed with different primer ratios for the optimisation of single strand production as shown in figure 12a top row. In these optimisation experiments the optimal primer ratio was found to be 100:1 as seen in lanes 2 and 8 of figure 12a top row. This was determined by fractionating the PCR products at 15 V/cm in 1.4% agarose in TBE. Under these condition the single-stranded product runs slower than the double-stranded product. Figure 12a, lanes 2 and 8 in the top, shows two bands with hardly any background from spurious products.

Both the single- and double-stranded PCR amplified templates were sequenced. The purification was done either by centrifuge-driven dialysis using Amicon 30™ microconcentrators or by selective precipitation in propan-2-ol. In centrifuge-driven dialysis, purification and concentration is achieved by ultrafiltration of the samples through an anisotropic membrane with a specific pore size cut-off. This is effected by adding diluted PCR products into

the reservoir chamber of the microconcetrators as described previously and shown in figure 7. Centrifugal force is then used to drive the solvent and low molecular weight solutes through the membrane into the filtrate cup leaving behind macromolecules larger than the membrane pore size, which in the case of Centricon 30 represents a molecular weight of 30,000. Smaller molecules like dNTPs, amplification primers and salts used in PCR can therefore be readily removed from the template DNA to be sequenced.

Sequencing reactions were performed with Sequenase T7 DNA polymerase with chain-termination and extension reactions being carried out either at 37^0C, 42^0C or 50^0C. In the case of double-stranded products, primer-template annealing was performed by heating the samples in the presence of the primers and immediately snap freezing as described in the methods. Sequencing was done with either one of the sequencing primers or with nested primers using different amounts ranging from 1 pmol to 10 pmol. The amount of the template DNA was also varied. This was done to determine the optimal set of conditions required in the sequencing reactions. Results of two of these experiments are illustrated in figure 12b.
Sequencing ladders generated from these experiments were faint and in some cases there were high backgrounds. Pile-ups were also observed at several areas on the sequencing ladders. Most of these problems were similar to those that were encountered and discussed earlier while sequencing other double-stranded templates like the bacteriophage lambda DNA and plasmid pBR322. Figure 12b demonstrates that pile-ups were no2. Figure 12b demonstrates that pile-ups were nc chain-extension and termination reactions were carried out at 50^0C while sequencing with Sequenase T7 DNA polymerase.

To circumvent these problems, single-stranded products in figure 12a, lanes 2 and 8 top row, were purified by centrifuge-driven dialysis and sequenced. The sequencing was performed with Sequenase T7 DNA polymerase by carrying out the chain-extension and termination reactions at 37^0C. To optimise for the primer-template annealing and extension reactions during sequencing, different amounts of DNA templates were used while keeping the amount of the sequencing primers constant at 1 p mol. In the experiment depicted in figure 13, sequencing was performed

using 30, 40, 50 and 75% of one PCR reaction. This is in contrast to the suggestion of Saha (1989) of pooling together up to five PCRs. This suggestion, as previously mentioned was made to ensure that there is sufficient DNA template for sequencing. In figure 13, set 3, it is in fact demonstrated that the best results were obtained when 50% of a PCR was sequenced with 1 pmol of the sequencing primer. It also shows that the sequencing ladder generated from ssPCR products is much superior to that generated from dsPCR products (figure 12). Sequences obtained from both the dsPCR and ssPCR products were, however, the same and did not match the already published sequences of either the erythropoietic or the non-erythropoietic PBG-D cDNAs (Raich et al., 1986; Grandchamp et al., 1987).

From these experiments it emerged that the optimal amount of PCR products to be used in sequencing reactions with 1 p mol of the sequencing primer was 50% of each PCR reaction. The above experiment was therefore repeated using these amounts of template and sequencing primer except that sequencing was performed with a nested primer R3 instead of the limiting primer. This was done because reamplification or sequencing with nested primers greatly improves specificity (Wrischik et al.,1987; Engelke et al., 1988). Sequencing with the nested primer R3 generated the same sequences as those described above, confirmirming that the 500 bp PCR product being amplified was not part of the PBG-D cDNA. This therefore, implied that either the R-T PCR was amplifying a cDNA product completely different from PBG-D cDNA or was amplifying genomic sequences and that the sequences obtained were part of the intronic sequences of the PBG-D gene. The later was unlikely because the amplification primers were known to correspond to different exons spanning across several introns. Nevertheless, to test these hypotheses both strands of cDNA were produced to enable sequencing of as much of the product as possible. This was achieved by using oligo $(dT)_{12\text{-}18}$ in the reverse transcription and performing the PCR amplification with unequal molar amounts of primers as described in the methodology. In these experiments the RNA was first treated with RNAase free-DNAase prior to being used in RT-PCR amplifications. This was done to ensure that the amplification was not arising from contaminating genomic DNA. Despite this modification the same 500 bp product

was obtained. Furthermore, direct sequencing of the RNA samples without prior reverse-transcription was not successful, again proving that the amplification of the 500 bp product was not arising from contamination of the RNA samples with genomic DNA. Using EMBL and GenBank data bank computer programmes, sequences obtained from these experiments were compared to previously reported sequences to determine the degree of homology. The closest homology, albeit very low, was with the starfish *Poacta ochraceus* cytoplasmic actin gene, confirming that the generated sequences were most probably not associated with the PBG-D gene. The conclusion from these experiments was that the primers F2 and R2 were somehow amplifying a mRNA other than the PBG-D mRNA. This may have risen through mispriming or misextension of the primers. Similar problems supposedly to be due to non-specific annealing of amplification primers to templates have been reported before (Haqqi et al., 1988; Fedlman et al., 1988; Anwar et al., 1990). Despite multiple attempts under different conditions and using two different set of primers, Haqqi et al. (1988), were unable to obtain specific amplification for a region of the coding sequence of the mouse V-β-6 receptor gene. This problem was overcome by performing a two-step PCR amplification in which following the first amplification the products were diluted 100,000-fold and reamplified with a second set of primers. A large dilution was found to be necessary as a 1,000-fold dilution still resulted in amplification of non-specific products. In other experiments, while amplifying the KM19 locus in cystic fibrosis gene, Feldman et al (1988) and Anwar et al., (1990) separately detected co-amplification of an extraneous product. On sequencing, no homology was found between this extraneous product and the KM19 product. Moreover, a search of the EMBL, GenBank and NBRF databanks did not reveal any significant homologies (Anwar et al., 1990). The amplification of the extraneous product was thought to arise from non-specific hybridisation of the amplifying primers to L1 repeats elements in the nonfunctional long terminal repeat (LTR) sequence. This problem was solved by either introducing a set of new amplification primers or by raising the annealing temperature from 55^0C to 60^0C, thus performing all steps in the PCR without allowing the

temperature to fall below 60^0C. In the current study, to solve the problem of amplification of the 500 bp extraneous product, optimisation for different PCR parameters were performed.

Specificity in PCR depends on several factors including amplification primers, cycling thermal profile and concentrations of magnesium ion, dNTPs, template and enzyme used. Failure of optimisation can lead to the amplification of non-specific products which at times may even be more predominant than the desired product. One obvious discrepancy between the current study and the work done by Grandchamp et al. (1989c), was the type of heating block used. In their work Grandchamp et al. (1989d) used DNA Thermal Cycler (Perkin-Elmer-Cetus Instruments^{TM}) while in this study a Techne PHC-1 Thermal Cycler^{TM} was used. This difference in the type of the heating block used may have affected the actual ramp temperatures. The ramp time, which is the time taken to change from one temperature to another, depends on the type of equipment used (Saiki, 1989). This is significant because one of the important factors influencing PCR specificity is the annealing step. Although, a temperature of 55^0C is a good starting point for annealing of most 20-base primers with about 50% GC content, in some cases higher temperature may be necessary to improve primer specificity. In addition, PCR specificity may also be improved by minimising the incubation time during the annealing and the extension step because this reduces the chances of mispriming and misextension (Saiki, 1989). After trying different thermal profiles, ultimately the right sized product was obtained when the initial denaturation was performed at 95^0C for 3 min and subsequent denaturations carried out at the same temperature for 1 min, followed by annealing at 60^0C for 1 min and extension at 72^0C for 1.5 min. The final extension was carried out as before at 72^0C for 7 min. Thus, modifications on the original protocol included the introduction of a 3 min initial denaturation step, raising of the annealing temperature from 55^0C to 60^0C and the reduction of the extension step from 2.5 min to 1.5 min. In addition to this, new primers F1 and R1 situated externally to the original primers were constructed. Alternatively, the correct sized products could be amplified by using one of the original primers, F2 or R2 with an internal primer on the opposite end. Thus for example by

performing one set of PCR with F2 and R3 and another with F3 and R2, the right sized products which overlap each other and span across the entire PBG-D cDNA could be amplified.

In PCR, primers are the single most important determinants of the outcome of amplification. In most cases amplification primers work, but for a variety of reasons some of which are not yet fully understood, some primers fail to amplify their intended target. This failure in amplification of the intended product can often be solved simply by moving the primers a few bases either upstream or downstream from the original position (Saiki, 1989). Optimisation experiments illustrated in figure 14a to 14d illustrate these points. In figure 14a the annealing step was performed at 60^0C instead of 55^0C as recommended in the original protocol. With these parameters using primers F2 and R2 the predominant product was approximately 500 bp long and overshadowed the desired 1,108 bp long product as shown in lanes 2-6. On the other hand, by changing one of the primers the expected product became more predominant than that produced by mispriming. This is shown in lanes 7-10, where the same PCR parameters as those used in lanes 2-6 were applied except that amplifications were performed with primers F2 and R3 instead of F2 and R2. Moreover, by performing the annealing reaction at 55^0C with primers F2 and R2 the 500 bp fragment was predominantly amplified both in the asymmetric and the standard PCRs as shown in figure 14b, lanes 2-5 were lanes 1 and 2 are products from asymmetric PCR and lanes 4 and 5 from a standard PCR. Thus specificity was to a large extent improved by merely changing the amplification primers. This modification on the protocol, however, did not completely prevent amplification of the 500 bp product. This problem was solved by performing PCR with the new set of primers F1 and R1, a few bases outside the original positions and modifying the thermal profile as mentioned earlier. This is depicted in figure 14b in lanes 6 and 7, where by performing the annealing reaction at 60^0C, the desired products were predominantly amplified. Thus by both using these new primers and altering the PCR conditions it was possible to amplify the desired product. On the other hand, when these new primers were used and the annealing step performed at 55^0C both the 500 bp and the desired products were amplified as shown in figure 14c.

In some cases, spurious products resulting from non-specific priming become persistent even after the most stringent optimisation. In such cases the use of specificity enhancers like TMAC (Hung et al., 1990) or formamide (Sarkar et al., 1990) may be helpful. These agents were not used in the current study and indeed their use may have not been desirable as they could have encouraged the amplification of the 500 bp product. This is due to the fact that these specific enhancers work by discouraging non-specific priming, but under these PCR conditions there is a preference for the production of the 500 bp product. Alternatively, this intractable problem of spurious products can be overcome by separating the products in a gel and excising the desired band. This is followed by a second round of amplification preferably with internal primers. The second round of amplification can be carried out directly or after elution of the DNA from the first amplification (Gyllensten, 1989). This approach was found to be convenient and was used in this study even after the desirable products were obtained following optimisation as shown in figure 14b lanes 6 and 7 and in figure 15a. This approach was preferred for two reasons. Firstly, after the first PCR amplification, the eluted cDNA could be stored in this form for a long time without resorting to repeated thawing and freezing of the less stable RNA samples. Secondly, the second round of amplification with internal primers increased the specificity and quantity of the cDNA templates.

5.4.2. 'Two-step' sequencing of PBG-D cDNA.

Although direct sequencing of asymmetric R-T PCR amplified PBG-D cDNA templates was successful as shown in figure 13, the sequencing ladders generated were faint in comparison to those that were obtained from direct sequencing of genomic DNA from the cystic fibrosis gene. This can be explained by having too little template cDNA from the porphobilinogen gene. This in turn, may be due to either sub-optimal reverse-transcription or insufficient PBG-D mRNA. To circumvent this, a first round of R-T PCR was performed with primers F1 and R1 in a standard manner to produce double-stranded products. The desired band was then excised and the cDNA eluted and reamplified with internal primers F2 and R3 in an asymmetric PCR. This is shown in

figures 15a and 15b where PCR amplification products of primers F1 and R1 in figure 15a were reamplified with primers F2 and R3 at a ratio of 1:100 (Figure 15b). The adequacy of the single-stranded products was monitored by running in 1.4% agarose gels as described earlier. Thus in this example it was possible to produce both cDNA strands with equal efficiency. The templates were purified by either selective precipitation in propan-2-ol or by centrifuge-driven dialysis methods and sequenced with the limiting primer F2 or an internal primer F3 using Sequenase T7 DNA polymerase. The results of these experiments shown in figure 15c, again demonstrate that purification of PCR templates by selective precipitation in propan-2-ol was as equally effective as the centrifuge-driven dialysis method using Amicon^{TM} concentrators. It also shows that sequencing ladders generated with internal primers were superior to those sequenced with limiting primers. The best results in these experiments are seen in sequencing set 4 in which the PCR amplified templates were purified by the selective precipitation in propan-2-ol and sequenced with the internal primer F3. In addition all sequences generated from these experiments corresponded to the normal sequences of the PBG-D cDNA.

These experiments were repeated by amplifying PBG-D cDNA from human placenta, HeLa cells and from lymphocytes of a normal subject. Following the two-step PCR amplification, templates were purified by the selective propan-2-ol method and sequencing performed with the limiting primer F3 using *Taq* DNA polymerase. The results of this experiment, illustrated in figure 16 (sequencing sets 1-3), demonstrate very clear sequencing ladders comparable to that produced with the cloned M13mp18 control DNA in set 4. These sequencing ladders, devoid of both high backgrounds and band artefacts, were easily readable for up to 200 bases. Although these sequencing ladders were somewhat superior to those generated with Sequenase T7 DNA polymerase, the difference did not appear to be very significant. From these experiments it became clear that the strategy to follow in the sequencing of the PBG-D gene was first to amplify the PBG-D cDNA by a standard R-T PCR and elute the amplified cDNA from an agarose gel. Such eluted cDNA could be stored at -20^0C for several months. The cDNA should then be reamplified asymmetrically and the products purified by either centrifuge-driven dialysis or

selective precipitation in propan-2-ol in the presence of 2 molar ammonium acetate. Sequencing can then be performed with either internal primers or limiting primers using Sequenase T7 DNA polymerase or *Taq* DNA polymerase. In this study, purification by selective precipitation in propan-2-ol was preferred to centrifuge-driven dialysis because it is simple, inexpensive and does not require any special equipment. Besides Amicon^{TM} microconcentrators used in this study, there are a myriad of other commercially available systems for purification of PCR templates prior to sequencing. Success with selective precipitation using propan-2-ol, however, circumvented the need for these costly purification methods. Moreover, as there was no remarkable difference between Sequenase T7 DNA polymerase and *Taq* DNA polymerase as far as clarity and accuracy of the sequencing ladders were concerned, the former was used in all routine sequencing. *Taq* DNA polymerase was reserved only for those cases where pile-ups could not be resolved by Sequenase T7 DNA polymerase even when the chain-extension and termination reactions were carried out at 50^0C. Both internal and limiting primers were used for sequencing, depending on which area was to be sequenced.

In all cases successful R-T PCR was performed without resorting to the use of RNAase inhibitors like RNasin or placental RNAase inhibitor. Although, these are generally included in all protocols that involve PCR amplification of RNA, no systematic evaluation of their importance has been made. In the current study, it has therefore been shown that these RNAase inhibitors can be safely left out of the R-T PCR protocols so long as other measures for reducing exposure to RNAase are enforced. In addition, the R-T PCR does not require the selection of Poly(A)$^+$ mRNA from total cellular RNA.

Following the successful development and application of direct sequencing in the characterisation of mutations of the PBG-D gene, modifications were introduced to simplify the protocol. This, for instance, included attempts at introducing a non-invasive method such as the use of buccal washings as the source of RNA instead of lymphoblastoid cell lines. As cell cultures and RNAs extraction introduce additional steps and delay in the analysis of samples, a method of processing crude cell lysates of uncultured lymphocytes was also employed.

Patients with AIP receive numerous venopunctures, either for investigations or medication. This not only is a cause of discomfort but can pose difficulties for the clinician. A non-invasive diagnostic test will therefore greatly alleviate this problem. In addition such a procedure would encourage corporation from asymptomatic gene carriers or unaffected family members who otherwise may be reluctant to volunteer blood specimen. It also eliminates the need of medical supervision in sample collection and the chance, albeit remote, of transmitting such blood borne infections like hepatitis and HIV (Lench et al., 1988). In the current study buccal epithelial cells were collected from volunteers by mouth washings using normal saline, and R-T PCR done directly on cell lysate without any elaborate RNA extraction. This approach was unsuccessful despite many attempts. Although, there have been reports of successful PCR amplification of genomic DNA from buccal cells, similar RNA amplification may be logistically difficult due to the presence in the saliva of non-specific inhibitors and ribonucleases. In addition, in mouth washings there may be insufficient PBG-D mRNA molecules for reverse transcription and therefore this may be another limiting factor in this approach.

Another attempt at simplification of the protocol for the characterisation of PBG-D mutations, was done by by circumventing the need for cell cultures and elaborate RNA extraction. This was accomplished by performing the R-T directly on blood lymphocytes. Following a venopuncture, the lymphocytes were separated and lysed to release total cellular RNA in the presence of DEPC and R-T PCR performed directly on the cell lysate without any further purification of the nucleic acids (Sherman et al., 1989). Thus, cell lysis, reverse transcription and PCR amplification of the PBG-D cDNA were all performed in the same tube greatly simplifying the procedure. Consequently in a matter of less than eight hours, the cDNA was amplified, eluted and a second round of amplification done with unequal molar amounts of internal primers to produce single-stranded cDNA templates suitable for sequencing. Therefore, starting from the receiving of blood samples to the reading of the results it may take less than two normal working days. Additionally, a further attempt at simplifying this procedure by doing a direct asymmetric R-T PCR was made. In this method single-stranded cDNA was prepared directly without a second

round of amplification thus carrying out the whole procedure from start to reading of the results in less than 24 h. Results of the direct asymmetric R-T PCR depicted in figure 17b show successful amplification, though direct sequencing of these products resulted in a very faint sequencing ladder as shown in the sequence set 4 of figure 17c. The faint sequencing ladder may be due to insufficient amount of single-stranded cDNA in the templates (Gyllensten and Erlich, 1988). As it has been mentioned before, the estimation of ssDNA by ethidium bromide staining may not be consistent. It may also be recalled that direct asymmetric R-T PCR performed previously on different RNA samples similarly gave very faint sequencing ladders and prior to the present optimisation, gave rise to sequences which were not from the PBG-D cDNA. Similar results associated with faint sequencing ladders has been observed by other workers (Gyllensten and Erlich, 1991). It therefore, seems that, in optimisation, there is an improvement in the specificity of the PCR amplification at the expense of its sensitivity. Sub-optimal cDNA synthesis may also be a factor since traces of DEPC are known to inhibit reverse-transcriptase and to modify purine residues in RNA by carboxymethylation. DEPC must therefore be removed prior to the R-T PCR, and in this study this was achieved by boiling the samples during the cell lysis step. Theoretically, it is most likely that this approach may be successful with another template, other then PBG-D cDNA, which produces a more homogeneous PCR amplified product.

Sequencing ladders generated from the second round of amplification were, however, more easily read and comparable to those produced from cultured cells after an elaborate RNA extraction procedure using guanidinium thiocynate and equilibrium centrifugation with caesium chloride. By doing the PCR in two steps and eluting the cDNA in the first step, the technique also provides material that can be stored on a long term basis, allowing further experiments or confirmation of the results. In addition, as discussed before and unlike direct asymmetric PCR, the second round of amplification also allows the generation of both cDNA strands, depending on which primer is chosen to be limiting.

5.4.3. Problems associated with direct sequencing of PCR amplified templates.

Several problems worth mentioning were, however, encountered with this method of direct sequencing of PCR amplified templates. Some of these problems were universal and may be observed in any sequencing project while others are perculiar only to the direct sequencing of PCR amplified templates. Thus, problems like compressions, band artefacts, variation in band intensity, smiling and frowning of bands are not for instance peculiar to this method and will not be discussed in detail except when relevant to the study. Certain problems were, however, particularly associated with direct sequencing of PCR amplified products and these included, inability to sequence certain strands, band artefacts, faint sequencing ladders and high background. Most of these problems were easily solved by optimising for PCR or sequencing reaction conditions and moreover none of them seriously interfered with the interpretation of the results. Other problems were resolved by improving template purification methods and gel electrophoretic conditions.

In a few occasions it was very difficult to sequence a specific strand in some templates. In its extreme form, this problem presented itself as failure in the PCR amplification of that particular strand. For reasons not as yet fully understood, on some occasions asymmetric primer ratios were able to amplify one strand and yet fail with the opposite strand. More commonly and subtly, the reciprocal asymmetric primer ratios may give rise to different amounts of ssDNA (Gyllensten, 1989). This problem was easily solved by simply changing the amplification primers. This is shown in figure 18a and 18b. In 18a, PBG-D cDNA from one patient was amplified in an asymmetric PCR with primers F2 and R3, in which R3 and F2 were limiting in lanes 2-5 and 6-9 respectively. It can be observed that there is a suboptimal amplification in the PCR where primer R3 was limiting. Thus in this experiment there was a more efficient amplification of ssDNA when the forward primer was limiting leading to the production of the antisense strand. Samples from this patient were amplified in another asymmetric PCR, this time with primers F2 and R2 and shown in figure 18b lanes 6-9, in which the limiting primers were F2 in the top and R2 in the bottom row. This time it was possible to demonstrate adequate amplification of ssDNA in the

PCR when the reverse primer (R2) was limiting, therefore amplifying the sense strand. Thus by changing the amplification primers it was possible to overcome difficulties associated with poor amplification of a particular strand. Theoretically, this problem can also be solved by using a different ratio of the primers in the asymmetric PCR. Therefore, in practice it was found to be a good policy to always monitor the amount of ssDNA, by size-fractionation of the PCR products in 1.4% agarose gels as shown in figure 18a and 18b. As mentioned earlier, under these conditions the single-stranded cDNA moved more slowly and just behind the double-stranded fragment. It must, however, be mentioned that ssDNA cannot be consistently quantified from staining with ethidium bromide, because its tendency to form secondary structures and intercalate the dye may vary between templates (Gyllensten, 1989). Nevertheless, this should not be a major problem when sequencing the same DNA or cDNA fragments. The importance of monitoring for ssPCR templates prior to sequencing is further shown in figure 18c, where in the sequence set 1, PCR amplified templates in lane 6 bottom row were successfully sequenced with the limiting primer R2, whereas sequencing of the products in lane 6 top row completely failed as shown in sequence set 2. This is due to the fact that although seemingly positive, the PCR templates in lane 6 top row consists mainly of dsDNA. Like templates in sequencing set 1, products from lane 6 bottom row were also sequenced with the limiting primer F2. Finally, the importance of checking PCR templates is also shown in figure 18b, where in lanes 2-5 top row there is a complete failure in the PCR amplification when primer F2 is limiting and very suboptimal amplification in lanes 2-5 when R2 is the limiting primer.

The sequencing ladder in set 1 of figure 18c, demonstrates two other problems. First there is a fairly high background and secondly there are some sections in the sequence where two or three bands are appearing in the same position. In this case it is possible that both problems are arising from the same cause. High background and band artefacts of this sort may arise from the reannealing of dsDNA present in the templates or from bad template preparation. Bad template preparation may lead to nicking or contamination with salts. Both template nicking and salt contamination may cause bands to appear at the same position in two or three lanes and in

addition, the former may give rise to a high background as well. Salt contamination can be easily detected from the appearance of the cDNA pellet because of the presence of salt crystals. Sometimes the only clue to this, is the appearance of a larger cDNA pellet than usual. When salt contamination is suspected, it is advisable to wash the cDNA pellet twice in 70% ethanol, otherwise in most cases a single wash after purification by precipitation in ammonium acetate and propan-2-ol is sufficient. These band artefacts may also arise from inadequate mixing of the reagents in the sequencing reactions or may be due to genuine compressions. The last two causes of such band artefacts are not peculiar to the sequencing of PCR amplified templates and may be observed in any sequencing experiment.

The other common problem was the presence of bands appearing at the same position in all four lanes. This problem, slightly different from the above problem, presented with artefacts of two types as shown in figures 19 and 20. The first type (a) is a gel artefact and is commonly seen in the sequencing of PCR amplified products (Brow, 1990). Bands of this sort vary from gel to gel and their relative positions of migration, change with the gel running temperature. These bands are not dependent on the sequence of the templates as demonstrated by their appearance across the entire gel width as shown in figures 19 and 20. This was not found to be a serious problem in the present study since such band artefacts tended to occur at the extreme 5' end of the sequences, that is at the top-most end of the gels where sequences were difficult to read in any case. Nevertheless, when necessary this problem can be easily solved by running the same samples in another gel at a different running temperature and duration as shown in figure 20. By altering these gel running conditions the band either becomes more diffuse, or migrates at a different rate thus enabling the reading of the sequence segment that had been previously obscured by the band.

The second type (b) of artefact bands are sequence dependent and seem to be due to pauses and mild compressions. Sequenase T7 DNA polymerase quite often pauses when it encounters exceptional secondary structure. In its severe form, this can lead to pile-ups. Genuine compressions are observed in G-C rich segments because this leads to the formation of G-C

hairpins which in turn form localised secondary structures in the DNA template, which persist during electrophoresis. These secondary structures cause oligonucleotides to behave as though they are shorter than is actually the case, thus causing them to migrate faster. This causes sequencing bands to run close together and sometimes to be even superimposed, usually with an increased inter-band space in the region immediately above this position. In many cases this was solved by performing the chain extension-termination reactions at 50^0C as mentioned earlier. Alternatively the problem was resolved by using *Taq* DNA polymerase for sequencing, thus enabling extension-termination reactions to take place at 70^0C. For this reason, the most consistent sequencing results with PCR-generated templates have been obtained using *Taq* DNA polymerase (Innis et al., 1988). In refractory cases dITP was substituted for dGTP in the chain extension-termination reactions. This analogue forms I-C base pairs which, having two hydrogen bonds instead of the three found in G-C pairing, are weaker than the latter and therefore their secondary structures are less stable. It must be mentioned that although this problem is common in the direct sequencing of PCR amplified products, it not peculiar to this method. Any complex secondary structure can give rise to this. In addition, practical problems like bad template preparation, failure to properly mix reagents and performing the labelling reaction at temperatures higher than 20^0C or prolonging the reaction beyond 5 min may all lead to this. In the case of the direct sequencing of PCR amplified products, this problem may be intractable due to reannealing of the short linear templates especially when sequencing dsPCR templates but can also be seen in the asymetrically amplified products as they are bound to contain certain amounts of dsDNA. Unless absolutely necessary to sequence a particular strand as in the confirmation of a mutation, this problem can also be solved by simply sequencing the opposite strand.

Another problem observed occasionally was faint sequencing ladders. The commonest cause for this is simply an inadequate amount of template cDNA available for sequencing. This can arise from suboptimal PCR amplification or from loss of DNA template in the course of purification. The presence of double-stranded cDNA and its competitive participation in the sequencing reaction may also contribute to this. Similarly the presence of salt contamination in

the course of DNA preparation may cause a dramatic inhibition of the sequencing enzyme leading to faint bands (Ward and Howe, 1989). To ensure the presence of adequate cDNA in the templates the PCR was performed as a two-step reaction as described earlier. This, besides guaranteeing the adequacy of the template cDNA, also provides specificity in the sequencing ladders. Faint sequencing ladders may also be caused by low sequencing primer concentrations which may arise through degradation due to repeated freezing and thawing. This can easily be prevented by storing the sequencing primers in small aliquots and keeping stock solutions at -20^0C or below.

Sequence dependent variations in band intensity occur during dideoxy-mediated sequencing of templates from any source and are not necessarily confined to direct sequencing of PCR amplified products. In this type of sequencing, single C bands are usually weaker then single bands from other nucleotides and in a homopolymeric run the first (bottom band) C tends to appear much weaker then the rest. On the other hand in a homopolymeric run of A bands, the first A (bottom band) generally appears the strongest. Finally, G bands appear weak when they are preceded by a T. Besides these universal variations in band intensity, in the direct sequencing of PCR amplified templates from heterozygote individuals, the two bands at the mutation site may not appear to have the same intensity. This is because in a heterozygote at a particular locus one transcript may be more frequent in the product then the other (Cotton, 1989). This type of band intensity variation may at times cause difficulties in intrepretation of the results especially when the sequencing ladder has other artefacts or a high background. The problem can be easily solved by repeating the sequencing and using the opposite strand.

5.5. Characterisation of mutations in patients with AIP.

Included in this study were 30 individuals from 29 families with acute intermittent porphyria and 10 normal controls. Diagnosis of AIP was based on increased excretion of delta-aminolevulinic acid and porphobilinogen in urine and decreased activity of erythrocyte PBG-D coupled with a clinical history of one or more acute attack (Moore et al., 1987).

RNA was extracted mainly from lymphoblastoid cell lines in all patients and controls. The extraction was performed by the guanidinium thiocyanate and caesium chloride equilibrium centrifugation method (Glisin, et al., 1974, Ullrich et al., 1977). Guanidinium thiocynate, a chaotropic salt is included in the primary extraction buffer to protect the RNA from degradation by RNAases (Chirgwin et al., 1979). This is necessary because the isolation of intact RNA is made difficult by the release of ribonucleases during cell lysis. Guanidinium thiocynate, a very potent denaturing agent, readily dissolves proteins, thus disintegrating cellular structures causing nucleoproteins to dissociate from nucleic acids as the protein secondary structure is lost (Sambrook et al., 1889c). RNAase are inhibited by the presence of 4 M guanidinium as well as reducing agents like β-mercaptoethanol (Sela et al., 1957). This combination of reagents was used in the extraction of RNA in this study. Following extraction in guanidinium thiocynate buffer the RNA was purified and collected by equilibrium centrifugation through a caesium chloride cushion. Under the conditions described and used in this study, RNA collects at the bottom of the tube while DNA remains floating as a ring near the top of the caesium cushion with proteins above it. In all cases no visible RNA pellet was seen but the quantity was more than sufficient for several experiments. The centrifugation was done in a swinging-bucket rotor with the centrifuge brake off. A swinging-bucket is preferred to a fixed-angle rotor because it allows the RNA to be deposited at the bottom of the tubes rather then along the walls where it may come into contact with the cell lysate and RNAases. The brake was turned off to prevent disturbance of the minute RNA pellets. In addition, throughout the extraction processes and during subsequent handling of RNA, precautions were constantly taken to prevent contamination of the samples with nucleases. Thus disposable sterile plasticware and DEPC treated plastic and glassware were always used as mentioned earlier. Whenever possible solutions were also treated with DEPC or made up in DEPC-treated water. Furthermore, gloves were worn at all times to protect the RNA samples from skin nucleases.

This extraction method was found to be satisfactory and convenient, the rotor dealing with six samples simultaneously. Usually the extraction was performed at the end of the day allowing the

centrifugation to be carried out overnight and the RNA harvested first thing in the morning. The RNA recovery was totally adequate and concentrations ranged from 128-1400 ug/ml with a mean of 650 ug/ml. In fact, in all cases that had an RNA concentration below 650 ug/ml, sample spillage in the course of their handling was the cause. This occurred during the shearing of sample nuclear DNA which was carried out by passing the samples several times up and down a 23-gauge needle. The RNA extracted by this method was also of very good quality and free of protein contamination as determined by the mean OD_{260}/OD_{280} value of 1.98. Pure RNA preparations, have an OD_{260}/OD_{280} value of 2.0. The RNA integrity was checked by electrophoresis in agarose gels containing formaldehyde. Prior to running of the samples, the RNA was denatured by heating at 55^0C for 10 min in the presence of formamide and formaldehyde. This is necessary in order to disrupt aggregates and hydrogen-bonded structures. At the end of the electrophoresis, the gels were washed to remove formaldehyde for better resolution of the rRNA bands. Under these denaturing conditions as is shown in figure 21, there is twice as much of the 28S rRNA (top band) as 18S rRNA (bottom band) in all samples. Typically undegraded RNA exhibits a 2:1 ratio of 28S:18S. The smear seen in all lanes indicate presence of mRNA and possibly rRNA degradation products.

The RNA was divided into two aliquots. To one aliquot, 0.1 volume of 2 M potassium acetate and 2.5 volumes of ethanol were added and the mixture stored at -80^0C until required. To recover the RNA the mixture was simply centrifuged at 5000*g* for 10 min. The use of chloride containing salts such as lithium chloride were avoided in these preparations because of their inhibitory effect on the reverse-transcriptase enzyme. The other aliquot was stored at -20^0C and used for R-T PCR amplification. The amplified PBG-D cDNA was also stored at -20^0C until required in the second round of amplification in asymmetric PCR for the production of suitable single-stranded templates. Conveniently, several single-stranded templates were amplified and purified by selective precipitation in propan-2-ol and ammonium acetate. The cDNA pellets were resuspended in either 10 ul or 14 ul of water and stored at -20^0C until required in sequencing reactions. In each sequencing reaction, a half of these aliquots was used, taking 10 ul aliquots for

Sequenase T7 DNA polymerase and 14 ul aliquots for the *Taq* DNA polymerase mediated sequencing reactions.

To characterise all types of mutations in the coding sequence of the PBG-D gene, the entire cDNA had to be sequenced. This required sequencing with at least four different sequencing primers complementary to the cDNA and mRNA of the PBG-D gene. The use of these primers made it possible to sequence the entire PBG-D cDNA in one experiment. To achieve this, several techniques which highlight different parts of the template sequence were utilised. For instance, the sequencing of regions close to sequencing primers were facilitated by the use of manganese buffer. On the other hand to enable sequencing of regions more distant to the sequencing primers, the concentrations of dNTPs used in the labelling mixture was increased. Furthermore, staggered loading was used, making it possible to run the sequencing reactions for different lengths of time. This enabled the reading of sequences from different segments of the gene all at once. This is demonstrated in figure 22, where sets 3, 4, 5 and 8 were from the same patient highlighting different regions of the PBG-D cDNA.

One of the major problems in sequencing regions close to the primers is the presence of very faint bands. This problem is particularly seen when there is only a small amount of template DNA. To circumvent this problem, two approaches were employed. In the first approach manganese buffer (0.15 M Sodium isocitrate, 0.1 Manganese chloride) was addeed in the sequencing reaction in addition to the usual magnesium based reaction buffer. In the second approach, the amount of nucleotides used in the labelling reactions was reduced to favour chain termination closer to the sequencing primers. The addition of manganese buffer works by potentiating chain termination properties of the dideoxynucleotides. When using T7 DNA polymerase and substituting Mn^{++} for Mg^{++}, discrimination against dideoxynucleotides incorporation by this enzyme is reduced four fold, thus allowing dideoxynucleotides to be incorporated virtually at the same rate as deoxynucleotides. Isocitrate is included in the buffer to expand the effective catalytic range of the Mn^{++}(Tabor and Richardson, 1989). Thus the addition of manganese buffer to sequencing reactions reduces the average length of DNA synthesised in

the termination step, intensifying bands close to the sequencing primer. With manganese buffer, sequences from less than 20 bases from the sequencing primers to approximately 200 bases can be clearly read even with reduced amount of template DNA. Alternatively, the region close to the sequencing primer was highlighted by diluting the labelling mix and restricting the labelling reaction duration to under 5 min as described in the methods. Use of reduced concentrations of nucleotides favours chain termination closer to the starting point. For this method to work effectively, it requires an adequate amount of template DNA, usually not less than 1 ug or else, the sequencing ladders generated will be very faint. This problem was overcome by concentrating the 10 ul aliquots of asymmetrically amplified templates mentioned before to 5 ul, thus effectively using an entire PCR mixture for each sequencing reaction. Of the two methods, the addition of manganese buffer was found to be more convenient and effective and therefore routinely used in this study.

Reading of long sequences or distant regions from the primers was facilitated by using more nucleotides and label in the labelling mixture as described earlier. Under normal circumstances, even when using high-resolution electrophoresis and cloned single-stranded templates, sequence-specific bands generated by the standard sequencing protocol begin to fade at about 600-800 bases from the primer. The exact point where sequence information fades out depends on many factors including the concentrations of template DNA, sequencing primers and label. In the direct sequencing of PCR generated templates this point where sequences begin to fade is observed earlier than with cloned templates. In most cases this problem was solved by changing sequencing primers, using a primer that was nearer to the region of interest. In some cases, however, a region of particular interest especially at the extreme ends of the PBG-D cDNA fragments, could not be reached by the available primers. This was seen in a few instances where a base change needed to be substantiated by resequencing of a specific region or confirmed by sequencing of the opposite strand. It is in these instances that this method of sequencing regions distant from the primers became very useful. This method, however, generated faint sequences close to the primers. To counter this, the manganese buffer method for sequencing regions close

to the primers was used alongside it, and the sequences generated from both were run in the same gel in a staggered fashion. The samples that were prepared to read distant regions were loaded first and electrophoresed for up to 4 h before the other samples were applied on the gel. Following the second loading, both samples were run together for a further 2 h. In most cases the electrophoresis was carried out using normal sequencing gels as mentioned in the methods. Under normal experimental conditions, using 50 cm sequencing gels as was done in this study, a 2.5 h run is expected to generate a sequencing ladder of 250 bases. In order to read more sequences, longer durations of electrophoresis can be used but at the expense of losing the early part of the sequencing ladder through running off the gel. To circumvent this, wedge gels or buffer gradient gels can be used. In wedge gels, the thickness of the gel is increased linearly down its length from top to bottom, causing the voltage drop to decrease in accordance with Ohm's law. This is due to the fact that electrical resistance of polyacrylamide gels in TBE, decreases with increasing cross-sectional area. In such gels, the small DNA molecules slow down towards the bottom of the gel thereby reducing the space between bands, whereas the larger chains are fractionated through a greater length of the thin area of the gel. Thus this technique reduces the tendency of the small DNA chains to be run off the gel enabling long sequencing ladders to be generated. The main drawbacks of this technique is that the bottom bands tend to be ill defined and that the gels take slightly longer to be well fixed and dry. Alternatively, the field strength gradient can be created by using buffer gradient gels (Sambrook et al., 1989a). In buffer gradient gels, the gels are poured in such a manner that there is an increasing concentration of buffer towards the the base of the gel (Biggin et al., 1983; Hong, 1997). Buffer gradient gels are easy to run and generate sequencing bands of good even resolution, but are difficult to pour. Taking all these factors into consideration, it was found that wedge gels were more convenient to use than buffer gradient gels. Therefore, in this study, particularly when specific distant regions needed to be sequenced, the electrophoresis was performed in wedge gels at a low gel running temperature of 40^0C.

Using these methods, cDNA from patients with AIP were sequenced to delineate the molecular pathology in this disorder. For comparison the AIP samples were always run with normal controls. This served two purposes; firstly, for comparison of the sequences to ensure that any changes observed were genuine and not due to artefacts and secondly, to compare the incidence of these changes in affected and normal individuals.

Seven point mutations were detected in nine patients with AIP and none in the controls. Four of these mutations are associated with amino acid substitutions and it is proposed that these are responsible for the causation of AIP, whereas in the remaining three, there are no amino acid changes and are therefore regarded as neutral polymorphisms. With the exception of the R167Q which has been previously described by Delfau et al (1990) all the remaining mutations are novel.

Q34K (C_{100}->A) caused by a C to A transversion in exon 4, changes codon CAG for glutamine to AAG for lysine at nucleotide 100 or amino acid residue position 34. At this position, glutamine is a highly conserved amino acid, present in the PBG-D from all species for which sequence data are available including *Escherichia coli*, rat, mouse, *bacillus*, *euglina gracilis* and human as shown in appendix C (Raich et al., 1986; Thomas and Jordan, 1886; Grandchamp et al., 1987; Chretien et al., 1988; Stubnicer et al., 1988; Beaumont et al., 1989; Sharif et al., 1989). In addition, according to prediction by the PEPTIDESTRUCTURE and PLOTSTRUCTURE programmes from the Genetics Computer Group (Chou and Fasman, 1978) this change is most likely to lead to the disruption of β-strand. Moreover, the mutation affects an amino acid which is located at the base of an active site cleft and is involved in hydrogen bonding to conserved amino acids, serine and arginine located at residue positions 96 and 191 respectively. The change is, therefore, likely to result in disruption of hydrogen bonding network, and in particular of the residues, arginine at position 32 and isoleucine at position 33, in its close proximity. This mutation, however, does not cause any alteration in the recognition sites of restriction enzymes. The mutation was seen in two related patients, a mother and her daughter one of whom had also the silent mutation G_{606}->T (V202V). The fact that this mutation is

predicted to cause disruption of the β-strand and that it affects a highly conserved amino acid at an active site, strongly support the view that this change is most likely to be the putative cause of AIP in this family. This claim is made stronger by the fact that after sequencing of the entire PBG-D cDNA no other changes were seen in one member of this family, while in the other there was a silent mutation in which the codon GUG (GTG) for valine was altered to GUU (GTT) coding for the same amino acid.

R167Q (G_{500}->A) mutation, altering codon CGG for arginine to CAG for glutamine at nucleotide 500 and amino acid residue position 167 was seen in one individual. This change affects a conserved amino acid which forms a direct salt bridge to the acetyl side group of ring 2 of the cofactor. The position occupied by this ring may become the binding site for the incoming substrate porphobilinogen. Mutation at this site result in loss of stabilising interaction to carboxylate group of the cofactor and possibly of the substrate. In addition, this mutation has been described before and is known to be associated with abnormal PBG-D activity (Delfau, et al., 1990). In a similar study, Delfau and co-workers (1990) starting with mRNA from AIP patients, amplified and cloned in a procaryotic expression vector the PBG-D cDNA containing the entire coding region of the gene and were able to demonstrate a G to A transition at position 500 from the translation initiation codon in several clones that had failed to express human PBG-D activity. This was originally noted in two unrelated patients but on further investigation using the ASO hybridisation technique, two additional unrelated patients with this mutation were discovered. Delfau et al. (1990), also showed that the mutated protein produced by this change was immunoreactive, indicating that the mutation was CRIM positive. In addition, through Western blot analysis it was demonstrated that the migration pattern of the mutated protein in SDS polyacrylamide gels was similar to that of the normal enzyme. This mutated protein was, however, shown to have reduced PBG-D activity with a specific activity of approximately 0.7% of the normal enzyme when measured at pH 8. In addition, the optimal pH for the mutated PBG-D was found to be dramatically reduced in comparison to the normal. This mutation has also been described previously in a retrospective diagnosis of a probable compound heterozygote

carrier of this defect (Beukeveld et al., 1990; Picat et al., 1990). This was reported retrospectively in a girl who died at the age of seven, from a severe and unusual presentation of at AIP (Beukeveld et al., 1990). The girl had presented with porencephaly associated with mental retardation and persistent excessive amounts of delta-aminolevulinic acid, porphobilinogen and uroporphyrin in her urine. It was inferred that she must have inherited both abnormal alleles with the G_{500}->A mutation coming from her father and the G_{518}->A from her mother (Picat et al., 1990). Although, in the CRIM negative group, the same mutation has never been described in more than one family, this does not apply to the CRIM positive group (Nordmann et al.,1990). It is possibe that the G_{500}->A may turn out to be one of the commonest mutation responsible for acute intermittent porphyria, though further studies are required before this conclusion can be made with certainty (Delfau et al., 1990). Nevertheless, this is the first time in AIP that the same mutation has been described independently from two different centres. In this mutation, it is interesting to note that the G to A substitution is occurring within a CpG dinucleotide. These kinds of mutations are thought to arise from the oxidative deamination of methylated cytosines and have previously been detected with a high frequency in several genetic diseases (Youssoufian et al., 1988; Grandchamp et al., 1989b; Lee and Nassbaum, 1989; Delfau et al., 1990). This mutation like the G_{100}->A mutation is involving a highly conserved amino acid as shown in appendix C. Also like the G_{100}->A mutation, it is not associated with alteration of a recognition site of any restriction enzyme.

The L177R (T_{530}->G), mutation caused by a T to G transversion at nucleotide 530, was seen in two unrelated patients with AIP. This changes codon CUG (CTG) to CGG thus substitution arginine for leucine at amino acid residue 177. Like the other two previously described mutations in this study, this transversion also affects a highly conserved amino acid. The involved amino acid leucine, forms part of the hydrophilic core of the PBG-D. Therefore, replacement of this residue by arginine introduces a charged group resulting in destabilisation of the core. In addition, this mutation abolishes a recognition site for the *Alu*I/*CviJ*I restriction enzymes allowing these to be used for screening purposes. *Alu*I is a frequent cutter of PBG-D cDNA,

dividing the cDNA into seven fragments in the normal and six in this mutation. For easier interpretation it may, however, be convenient to amplify a 549 bp fragment using primers F4 and R4. In the normal allele, *Alu*I digests this fragment into four segments of 190, 91, 43 and 225 bp. In this mutation, the abolishing of the recognition site between the 91 and 43 bp segments leads to the production of three segments of 190, 134 and 225 bp. The screening can be performed simply by amplifying the cDNA using the R-T PCR technique and fractionating the products in agarose gels as previously described. Following ethidium bromide staining, the PCR products can then be directly visualised under U.V. light. Using such a method, it is possible to quickly scan individual members in affected families. Because the change is intragenic and caused by the mutation, the diagnosis is made directly. This information is useful in the counselling of the affected family members. Depending on the frequency of the mutation this method can also be useful in screening for the mutation in unrelated individuals (Delfau et al., 1990; Lee, 1991a).

Mutation H256N (C_{766}->A), a transversion of C to A at nucleotide 766 leading to the substitution of histidine to asparagine at amino acid residue 256 was detected in one patient. The mutation changes codon CAC to AAC which in addition to the above mentioned amino acid substitution, also abolishes a recognition site for *Ban*I, situated one base upstream from the mutation site. The mutation is situated towards the end of a helix and does not affect a conserved amino acid. The substitution of the positively charged histidine by asparagine does, however, affect a conserved charge. In *E. Coli* protein, the corresponding residue argenine at this position, forms a salt bridge to glutamine. It is possible that in the human, the corresponding hydrogen bond is formed by histidine at this position with asparagine at residue 322. Therefore, substitution of histidine by asparagine results in pairing of the two asparagines, which may be less efficient at forming hydrogen bonds. Since a search through the entire PBG-D did not reveal any other mutation, it is most likely that this change is responsible for the disorder in this patient. The abolishing of the recognition site for the enzyme *Ban*I offers an alternative method of tracking this mutation. In the normal, this enzyme cuts the entire PBG-D cDNA and the immediate intervening sequences twice giving three segments of 87, 694 and 369 bp. This

mutation abolishes the restriction site between the 694 and the 369 bp segments leading to the production of two segments consisting of 87 and 1,063 bp. Similar mutation (L245R or T_{734}-G) also involving a non-conserved amino acid in exon 12 has been reported (Delfau et al., 1991). This mutation caused by a T to G transversion at position 734, is close to the H265N mutation and like it, does not involve a conserved amino acid.

The remaining three mutations, L42L, S45S and V202V were not associated with any change in amino acid residues and are therefore regarded as silent mutations. In some cases, these so called silent mutations may still be responsible for pathology. This especially occurs in highly expressed genes where despite degeneracy of the code, a particular codon or a set of codons may be preferred for certain amino acid than others. In such cases, changing of a codon or a set of codons to another, even if still codes for the same amino acid, may lead to inefficient incorporation of that particular amino acid residue. These mutations in the PBG-D gene did not appear to be of this type, since in all cases the frequency of the original and the changed codons in the PBG-D cDNA appeared to be the same.

V202V was the most frequent mutation and was seen in four unrelated individuals, three of whom had other mutations including L117R, R167Q and Q34K. This mutation alters codon GUG (GTG) to GUU (GTT), but as both code for the same amino acid valine, there is no change in the protein structure. The mutation is not associated with any alteration in the recognition sites of restriction enzymes. Similarly, the L42L mutation due to the change of codon UUG (TTG) to CUG (CTG) seen in one patient in this study, does not lead to any change in the amino acid residues or recognition sites for restriction enzymes. Therefore, this mutation which is most probably a neutral polymorphism, is very unlikely to be the cause of AIP. In addition, since the mutation does not alter a recognition site for any restriction enzyme, it is not useful in the diagnosis or linkage studies for tracking of this condition in the affected family.

The remaining mutation S45S, which was also seen in one patient, is not associated with any amino acid change. The mutation alters codon UCG (TCG) to UCA (TCA) both coding for serine. The change, however, abolishes the recognition site for the enzyme *Rsa*I. This enzyme

digests the normal PBG-D cDNA and its immediate intervening sequences into two fragments of 136 and 1114 bp. As in the mutant allele this recognition site is abolished, the amplified cDNA will fail to cut. This may prove to be useful in linkage studies within affected families. To date, only five RFLPs have been recognised in the human PBG-D locus. These are due to *Msp*I, *Pst*I, *Apa*LI and a polymorphism recognised by *Bst* NI and *Scr*FI (Llewellyn et al., 1987; Lee and Anvret, 1987; Lee et al., 1988; Lee 1991b; Picat et al., 1991). Interestingly, four of these RFLPs have been mapped within a region of 1.5 kb in the first intron of the PBG-D gene which spans approximately 3 kb (Lee, 1991a). The fifth RFLP has been located in exon 1 at position –64 relative to the initiation translational codon (Picat et al., 1991).

DNA polymorphism may arise either from point mutations, tandem repeats of short DNA sequences, or chromosome rearrangements caused by deletions and insertions (Kidd et al., 1989). In the human genome, sequence differences between any two alleles, on average, occur at the rate of 1 in 500 nucleotides (Jeffreys, 1979; Antonorakis et al., 1982). However, only about 5% of such sequence differences in the human DNA are detectable by conventional Southern blot analysis after restriction enzyme digestion (Antonoarakis, 1989). These polymorphisms can, however, be detected by direct visualisation of the amplified products after digestion with the relevant restriction enzymes. Demonstration of polymorphism by using PCR in linkage studies requires considerably less time and effort. Within a day, target sequences can be amplified using primers flanking the polymorphic site, digested with the relevant restriction enzymes, fractionated and visualised on gels (Kogan et al 1987). Haplotyping analysis using RFLPs in the diagnosis of acute intermittent porphyria has been done before and found to be especially useful in the diagnosis of those subjects whose enzyme activity lies within the overlap zone (Lee et al.,1988; Lee et al., 1991a). The use of these intragenic RFLPs in the tracking of PBG-D gene within affected families thus allows asymptomatic carriers and normal individuals to be identified with greater certainty than can be achieved by the conventional biochemical methods (Llewellyn et al., 1987; Lee et al., 1988; Grandchamp and Nordman 1988; Lee et al., 1991). However, this approach is limited to the families of patients who have both potentially

informative genotypes and sufficient unequivocally affected, living relatives to enable linkage studies to be used (Scobie et al., 1990b). In a study involving 47 unrelated AIP patients of mainly European stock, Scobie et al. (1990b) found a marked linkage disequilibrium between the polymorphic sites for *Msp*I, *Pst*I and *Bst*NI/*Scr*FI. They were able to identify only four out of the eight theoretically possible haplotypes from these polymorphic sites. Furthermore, the frequency of one of the haplotypes was less than 2% making the three RFLPs less informative then they would have been when inherited independently. The usefulness of the polymorphic site for *Rsa*I caused by the transition of G to A at position 135 detected in this study is yet to be determined. Theoretically these RFLPs or direct detection of the above described mutations which alter restriction sites can both be used for prenatal diagnosis of AIP. In practice, however, this will probably be considered only in those rare occurrences where both parents are affected. AIP being an autosomal disorder, it is possible that homozygosity is non-viable or may lead to a severe and crippling form of the disease (Beukeveld et al., 1990; Picat et al., 1990).

Among the main problems in the management of AIP, is the variability in the clinical presentation and heterogeneity both at protein and DNA level. As the majority of gene carriers of AIP are not symptomatic, follow up of affected families may not be easy. This problem is compounded by the heterogeneity of the molecular pathology responsible for this condition. To date, 13 confirmed mutations responsible for both CRIM-positive and and –negative AIP have been reported in the literature. Most of these mutations have so far been unique to the index family in which they were first reported. This particularly applies to the common CRIM negative phenotype, where so far the same mutation has never been found in more than one family (Nordmann et al., 1990). With the exception of one mutation caused by a deletion of T, the rest of the reported mutations are due to single base substitution. These mutation summarised in table 2, have different effects on PBG-D cDNA and/or protein. These include splicing defects leading to skipping of exons, premature introduction of stop codons with the subsequent production of truncated proteins, missense and nonsense mutations resulting in the synthesis of abnormal

proteins and frameshift with production of truncated products. Also included in the table, are mutations detected in this study.

Table 3: Reported mutations of the PBG-D gene in AIP.

Putative pathological mutations:

Mutation	Exon	Amino acid change	Reference
1. G_{33}->T	1	None (splicing defect)	(Grandchamp et al., 1989a).
2. G_{34}->A	-1	Glu to Lys (splicing defect).	(Grandchamp et al., 1989b).
3. C_{100}->A	4	Gln to Lys	(Mgone et al., 1991).
4. C_{346}->T	8	Arg to Trp	(Lee et al., 1990).
5. G_{446}->A	9	Arg to Gln	(Delfau et al., 1991).
6. C_{463}->T*	9	Gln to stop	(Scobbie et al., 1990).
7. G_{500}->A	10	Arg to Gln	(Delfau et al., 1990).
8. C_{517}->T	10	Arg to Trp	(Lee et al., unpublished).
9. G_{518}->A	10	Arg to Gln	(Delfau et al., 1990).
10. T_{530}->G	10	Leu to Arg	(Mgone et al., 1991).
11. G_{593}->A	10	Trp to stop	(Lee and Anvret, 1991).
12. G_{612}->T	10	del of 9 bases	(Delfau et al., 1991).
13. T_{734}->G	12	Leu to Arg	(Delfau et al., 1991).
14. T_{766}->A	12	His to Asn	(Mgone et al., 1991).
15. G_{771}->A	12	skipping of exon 12	(Grandchamp et al., 1989c).
16. del T_{900}	14	frame shift of stop codon	(Delfau et al., 1991).

Neutral polymorphisms:

17. C_{-64}->T	1	-	(Picat et al., 1991).
18. G_{117}->A	4	Leu to Leu	(Mgone et al., 1991).
19. G_{135}->A	4	Ser to Ser	(Mgone et al., 1991).
20. G_{606}->T	10	Val to Val	(Mgone et al., 1991).
21. T_{633}->G	-	-	(Grandchamp et al., 1987).
22. C_{731}->T	-	-	(Grandchamp et al., 1987).

–1 refers to the first intron.

*The numbering referred to here, is from the translational initiation site, whereas in the the cited reference, this mutation is actually reported as C_{412}->T.

The first two mentioned mutations in table 2 above, involve the first exon-intron junction changing the consensus splice junction sequences. In these two splicing defects, since the aberration affects only the the coding sequences of the non-erythroid isoform of PBG-D, the enzyme defect is found to be restricted to non-erythropoietic cells. The first mutation caused by a G to T transversion in the last position of the first exon, was observed in a Finnish kindred. Although this change does not modify the amino acid alanine specified in the last codon, since the change of codon GCG to GCU (GCT) still encodes for the same amino acid, it does, however, interfere with the normal splicing by changing the normal splice consensus sequence 5'CGGTGAGAGT 3' to 5'CTGTGAGT 3' (Grandchamp et al 1989a). The second mutation occurring only one nucleotide 3' to the first, was described in a Dutch family. The mutation is caused by a G to A transition in the first position of the first intron, causing substitution of the amino acid glutamic acid to lysine and at the same time changing the consensus splice sequences, 5'CGGTGAGT 3' to CGATGAGT 3'. This like the first mutation leads to the defective splicing of the primary transcripts initiated at the upstream promoter of the gene without affecting the expression of the PBG-D gene in the erythroid cells which utilise the downstream promoter found 2.8 kb 3' to these. Thus affected individuals with any of these two mutations will have normal erythrocyte PBG-D activity. Both mutations were characterised by cloning of the mutant alleles and the subsequent detection in other individuals performed by ASO hybridisation after PCR amplification. These mutations result in CRIM-negative phenotype.

The fourth shown mutation in table 2, was originally observed in one Swedish AIP patient with hepatoma (Lee et al., 1990). This mutation is confined the index family and a further search for the mutation in 28 other affected Swedish families proved fruitless. In this mutation the transversion of base C to T at nucleotide position 346 changes the codon CGG for arginine to UGG (TGG) for tryptophan at the residue position 116 (R116W). Although this mutation is not likely to result in any predictable major secondary structural change, it does, however, affect a highly conserved amino acid as shown in appendix C. This mutation is very similar to the ones

reported in the current study and like the first and second described mutations, was characterised by sequencing of cloned PCR amplified templates and results in CRIM-negative phenotype. Further screening for the mutation in the affected family members and other individuals was performed by ASO hybridisation after PCR amplification.

The fifth mutation, a G to A transition at nucleotide position 446 in exon 9 changes codon CGA for argenine to CAA for glutamine at amino acid residue position 149 (R149Q). This affects a highly conserved amino acid and moreover is likely to disrupt the helix in the secondary structure of the protein. The mutation which results in CRIM-negative phenotype, has been so far described in one individual after sequencing of cloned PCR amplified products (Delfau et al., 1991). The sixth mutation in table 2, referred to in the original publication as C_{421}->T is actually C_{463}->T, counting the nucleotide position from the translation initiation site. The transition of C to T at this position converts codon CAG for glutamine to the stop codon UAG (TAG). This mutation does not affect a conserved amino acid, but the premature introduction of a stop codon prevents the translation of about 60% of the coding sequence leading to the production of a catalytically inactive truncated protein lacking the cysteine residue necessary for of the attachment of the pyrromethane cofactor. The mutation which results in CRIM-negative phenotype was observed in only one patient in spite of investigating 43 unrelated individuals. It was demonstrated by sequencing of cloned PCR products rather than direct sequencing as performed in the current study.

Mutation G_{500}->A (R167Q), observed in one patient in this study, has been previously described in four different families (Beukeveld et al., 1990; Picat et al., 1990; Delfau et al., 1990). It is probably the most common mutation in AIP. Mutation G_{518}->A (R173Q) occurring 18 nucleotides downstream to this, has been observed in two unrelated individuals. The transition of G to A in this mutation leads to the alteration of codon CGG for arginine to CAG for glutamine and at the same time, the supression of an *Msp*I restriction site. Both this and the previous mutation result in CRIM-positive phenotype. Moreover, Lee and co-workers (unpublished data) have observed another mutation affecting this codon due to transition of C to

T changing the codon CGG for arginine this time to UGG (TGG) coding for trytophan. This change also creates a recognition site for the restriction enzyme *Bst*NI which can therefore be used for screening of the mutation in affected families. Although both C_{517}->T (R173T) and G_{518}->A (R173Q) are occurring at the same codon, it is interesting to note that while the former leads to CRIM negative the later causes CRIM positive phenotype. Therefore, the same CRIM phenotype may arise from different mutations and conversely, similar mutations arising from the same codon may result in different phenotype subtypes. Although all these three mutations involving exon 10 affect highly conserved amino acids, they do not seem to significantly alter secondary structure of the PBG-D. In their report Delfau et al. (1990), characterised two of these mutations (G_{500}->A and G_{518}->A) by sequencing of cloned PCR products. Subsequent detection in other affected individuals was then performed by restriction analysis, in the case of the G_{518}->A mutation which suppresses an *Msp*I site and by ASO hybridisation in the G_{500}->A mutation. The C_{517}->T mutation was on the other hand characterised by direct sequencing of PCR products.

The eleventh mutation depicted in table 2, is caused by a G to A transition at nucleotide position 593 in exon 10. Although this does not affect a conserved amino acid, it changes codon UGG (TGG) for tryptophan at amino acid residue 198 to a stop codon UAG (TAG) leading to the production of inactive truncated protein. This nonsense mutation also creates a new restriction site for the enzyme *Nhe*I and leads to CRIM negative phenotype. The mutation was observed in 16 families in Sweden, accounting for 40% of AIP cases in the investigated community. Genealogical studies, however, have revealed that almost all these families were related to the index family implying a 'founder effect' of this mutation in these families and confirming the high degree of heterogeneity in the CRIM negative phenotype. This and the G_{517}->A are the only PBG-D mutations in the literature that have been fully characterised by exclusively using PCR amplification and direct sequencing. Unlike in this study, the amplification was performed by a standard PCR producing double-stranded products which where then directly sequenced with *Taq* DNA polymerase using a novel technique referred to as cycling sequencing (Lee and

Anvret, 1991; Lee, 1991b). In this new technique, the sequencing reactions are performed on a programmable heating block using one of the primers that had been originally used in the PCR. The primer is mixed with the appropriate extension-termination mixes in four separate reactions and overlaid with light mineral oil to prevent evaporation just like in a PCR. After a 3 min denaturation step, cycling is performed by heating the samples in a cyclic manner at temperatures of 94^0C for 1 min, 60^0C for 1 min and 72^0C for 40 sec. This is repeated for for 25 cycles and at the end of the reaction the samples are concentrated to 4 ul, a stop mix added and electrophoresed in the usual way. Following characterisation of the mutation in the index case, family members were screened using another novel technique of cyclic reactions. In this technique, employing the principles of dideoxy DNA sequencing, PCR amplified templates and appropriate termination mixes in a final volume of 10 ul, are mixed with end-labelled primers designed to generate the first termination at the mutation site. Thus for example if there is a G to A mutation, termination mix A containing ddATP, dCTP, dGTP and dTTP is used. The reaction mixture is then amplified for 15 cycles using *Taq* DNA polymerase. At the end of the amplification, 4 ul of a stop solution containing 95% formamide, 0.01% bromophenol blue, 0.01% xylene is added and 2.5 ul of the mixture, denatured and electrophoresed in a 10% denaturing polyacrylamide gel. The results are read on autoradiographs in which homozygote normal individuals will show a single band different from affected homozygotes and heterozygotes will show two bands. This method unlike ASO hybridisation, requires the use of only one primer and no additional equipment other then a programmable heating block.

The twelfth mutation, in table 2, also involves exon 10. It is due to a G to T transversion of the last base of exon 10 (nucleotide position 612) which results in a splicing defect and a new restriction site for *Bam*HI. This mutation which has been partially characterised by direct sequencing of PCR amplified products, leads to a deletion of the last nine bases of this exon, activating a cryptic site three codons upstream of the normal site. This results in the production of a stable but abnormal mRNA which is CRIM-negative. It is interesting to note that the deletion of these nine bases involve three amino acids residues, valine, glycine and glutamine at

positions 202 to 204. All these amino acid residues are not conserved including valine at residue 202 which has been described in one of silent mutations in the current study. It is therefore difficult to explain how this mutation results in CRIM-negative phenotype. The absence of immunoreactive protein is compatible with the possibility that the enzyme is present but its three-dimensional structure is sufficiently altered to prevent recognition by antibodies. Among the possible explanations given include, the fact that the PBG-D activity depends on the covalent binding of a dipyrrolmethane as a cofactor which protects the enzyme from degradation (Umanoff et al., 1988). It has been postulated that, this mutation may prevent binding of the cofactor leading to an unstable protein (Delfau et al., 1991). As a counter argument, it is important to note that the mutation, however, does not involve the cysteine residue 261 of the human PBG-D which is highly conserved through the evolution and is most likely to be the cofactor binding site. The cysteine residue 242, located at a homologous position in the *Escherichia coli* PBG-D has indeed been shown to bind the cofactor (Miller et al., 1988).

The thirteenth mutation, T_{734}-G (L245R), alters codon CTT (CUU) for leucine to CGT (CGU) for arginine at position 245. This mutation does not involve a conserved amino acid. Moreover, being located towards the end of a helix, is not likely to result in any predictable secondary structure change. Therefore, this mutation resembles the nearby G_{766}-T (H256N) mutation, observed in the same exon in the current study. It may be possible that, despite this region not being a conserved domain, nevertheless, is important in the function of the human PBG-D. It is interesting to note that this region is close to the highly conserved cysteine residue 261, which may be the binding site for the cofactor and that the H256N mutation described in this study is actually closer to this cysteine residue than mutation L245R. This mutation which results in CRIM-negative phenotype has been detected in one individual via sequencing of cloned PCR amplified products.

The fifteenth mutation in table 2, is a G to A transition at nucleotide position 117, causing a splice defect and skipping of exon 12 (Grandchamp et al., 1989c). This was observed in one patient out of six investigated using PCR. In their study, Grandchamp et al. (1989c), on

amplifying the entire PBG-D cDNA from these patients, observed that in one individual there was an additional band of about 120 bp, suggesting the presence of a truncated protein besides the normal PBG-D. Direct sequencing of the eluted abnormal product confirmed this to be due to the skipping of exon 12, but the presence of several ambiguities within the sequencing ladder precluded definite characterisation of the mutation. To circumvent this, the abnormal PCR product was then cloned into a plasmid vector and subsequently sequenced. This revealed the above mentioned G to A transition at the last position of exon 12 within the donor splicing site of the intron 12, causing this exon to be skipped in such a way that the exon 11 was directly joined to exon 13. The resultant abnormal protein although stable, is catalytically inactive and gives rise to CRIM positive phenotype. This mutation also suppress a recognition site for the restriction enzyme *Bst*NI.

The sixteenth mutation is a single base deletion, affecting the nucleotide T at position 900. This results in a premature introduction of a stop codon located 15 codons downstream from the deletion. The mutation has been characterised by sequencing of cloned PCR amplified products. It leads to CRIM-negative phenotype.

Silent mutations involving the PBG-D gene have also been reported before. For instance, Grandchamp et al. (1987) have reported T_{633}->G and C_{731}->T mutations. Unfortunately due to lack of detail regarding to the location of these mutations in relation to the translational initiation codon, it is not possible to compare them to any great detail to the ones observed in the current study. More recently, two more silent mutations of the PBG-D gene due to G to T transversion in exon 10 (Gu et al., 1991) and C to T substitution in exon 1 (Picat et al., 1991) have been reported. In the present study, a G to T transversion in exon 10 has also been observed, but as the mutation detected by Gu and his co-workers is yet to be described fully, no comparison is possible at the moment. On the other hand, the C to T substitution in exon 1 described by Picat et al (1991), occurs at position –64 relative to the initiation translational codon and can be detected by digestion with the enzyme *Apa*I.

Therefore, data from both previous and the present work indicates that the majority of mutations are clustered around exon 10. Of the sixteen well documented putative disease causing mutations described to date (including those from this study), six of them involve this exon indicating this to be a hot spot for mutations of the PBG-D gene. All mutations identified so far have been due point mutations resulting in either splicing defects (Grandchamp et al., 1989a; 1989b; 1989c), nonsense mutations leading to synthesis of inactive truncated protein (Scobie et al.,1990; Lee and Anvret, 1991) and missense mutations with single amino acid substitutions (Delfau et al., 1990; Lee et al 1990). Of the missense mutations, all except two, including one seen in the current study, affect highly conserved amino acids. In all these cases the amino acid concerned was conserved in all six species for which information on the PBG-D sequences are available. In the exceptional case seen in the current study, the mutation, nevertheless, affected a conserved charge. The commonest involved amino acid is arginine, which is affected in eight out of the ten reported cases of missense mutations of the PBG-D gene. In six cases this involve the replacement of this amino acid to another and in two cases it is replaced by another amino acid.

In the current study as in several others (Grandchamp et al., 1989a; 1989b; 1989c; Delfau et al., 1990 and 1991) the starting material was mRNA. By starting with mRNA and using R-T PCR, it allows the *in vitro* amplification of the entire coding sequence of the PBG-D gene to be performed in a single step. When using this method some of the possible mutations of the CRIM-negative phenotype may, however, be missed since for this approach to be successful it requires the mutant gene to be transcribed and the resultant abnormal mRNA to be present in the cells. In the case of CRIM positive mutations this problem does not arise, since in demonstrating the evidence of the presence of abnormal protein, it can be assumed that a pathological mRNA should also be present in the nucleated cells of the affected individuals, and that the mutations should lie in the coding sequence of the mRMA (Grandchamp et al., 1989c). It is likely, however, that many CRIM negative mutations can also be detected using mRNA, including most splicing mutations and amino acid substitution that result in protein instability (Delfau et al., 1990). Working with mRNA as the starting material has also been necessary because it is only

recently that the genomic sequence of the PBG-D gene including introns has been published (Lee, 1991a).

Methods used in the detection and characterisation of the PBG-D gene have evolved from the sequencing of partial libraries of cloned cDNA (Grandchamp et al., 1989a; Grandchamp et al., 1989b) through the sequencing of cloned PCR amplified cDNA and DNA (Lee et al., 1990; Scobie et al., 1990; Delfau et al., 1990) to the direct sequencing of PCR amplified cDNA as reported in this study. Construction of genomic or cDNA libraries is time consuming and labour intensive. This problem has partially been alleviated by sequencing of cloned PCR amplified templates. In this technique, the cDNA or DNA fragments of interest are amplified with primers that add restriction sites to the template to enable cloning into vectors. As mentioned earlier, this method has three main disadvantages. Firstly, any misincorporation caused by *Taq* DNA polymerase during PCR amplification will be passed onto the cloned products and require several clones to be examined before such mistakes are recognised. Secondly, in an autosomal condition like AIP, several clones need to be examined and both alleles demonstrated before the heterozygosity can be confirmed. Lastly, the cloning is an additional step increasing labour and time. To circumvent this, attempts albeit unsuccessful, have been made before at characterisation of mutations in the PBG-D gene by direct sequencing (1989c). Recently, Lee (1991a) by using an alternative DNA method which simultaneously applied PCR and dideoxy sequencing, was able to demonstrate a new RFLP due to the suppression of an *Apa*LI restriction site created by a C to A transversion in the PBG-D gene. By using the same method Lee and Anvret (1991) have also been able to detect and characterise the G_{593}->A mutation mentioned earlier. The method, however, requires elaborate optimisation of several parameters which include the cycling thermal profile and different sequencing primer:template and dideoxynucleotide:deoxynucleotide ratios. This precludes the use of standard or commercially available chain termination-extension mixes. In the current study these problems were circumvented by direct sequencing of asymmetric PCR amplified templates. Through this method it was possible to characterise four possible pathological mutations and a new RFLP caused by the suppression an *Rsa*I restriction site in

exon 4 as shown in figures 23-29. The method was found to be reliable and in comparison to other methods that have been used before in studying this condition, was time saving and less labour-intensive. This method can be applied to investigate other monogenic disorders and should be particularly useful in the investigation of conditions which have molecular heterogeneity similar to AIP.

In practice, after characterisation of the mutation in the index case, subsequent tracking of the mutation in affected families can be done either by ASO hybridisation or by haplotype analysis if new restriction sites are created by the mutations as discussed before. The main problem, however, is to prove that the observed change is the putative cause of the condition being investigated rather than a neutral polymorphism. This is particularly true for missense mutations, where the only change may be a substitution of a single amino acid residue. The problem can be partially solved by comparing the prevalence of the mutation in the affected and normal controls, as has been done in this study. If the mutation can be demonstrated to occur in the affected individuals and to be absent in the controls, it is then most likely to be the cause of the disorder. This approach is, however, of limited use when investigating very rare conditions. Mutations may also be indirectly proved to be the putative cause of diseases if shown to alter the structure or to affect critical or conserved regions of the concerned protein molecule. In this study for instance, PEPTIDESTRUCTURE and PLOTSTRUCTURE computer programmes were used to predict the likely changes in the protein molecule secondary structure resulting from the mutations. In addition it was demonstrated that three of the mutations affect highly conserved amino acids. Alternatively, expression studies can be performed to demonstrate abnormal properties of the mutated protein. Thus, for example, the functional consequence of the mutated protein can be be studied by cloning PCR amplified cDNA containing the mutation from affected individuals, into procaryotic expression vectors and determine its activity. Using this method, Delfau et al. (1990), were able to demonstrate a causal association between the G_{500}->A mutation and an abnormal PBG-D. Moreover, in some cases, the causal effects of mutations may be

obvious, for example, when they cause premature insertion of stop codons or frame-shifts, resulting in severely truncated or grossly abnormal protein products.

The delineation of molecular pathology may also be used to study genes of unknown function. This can be achieved by studying these genes in cloned cDNA or genomic DNA. To achieve this, several approaches can be used. First, by using what are referred to as transient expression systems, one can introduce the normal or mutant genes and determine the quantity and structure of their transcripts. Alternatively, the same may be achieved by introducing the genes into appropiate cell lines, as for example human haemoglobin genes into mouse erythroleukaemia cell lines or by transfecting them into embryos so that their patterns of integration and expression can be studied over several generations (Weatherall, 1985). Such studies may, threfore, lead to understanding of the pathogenesis of some monogenic disorders.

In this study, a system for the characterisation of molecular pathology has been developed and applied to delineate mutations in AIP. This has be accomplished through direct sequencing of PCR amplified cDNA derived from total cellular RNA. The method has been found to be reliable and can be used as a prototype to investigate molecular pathology of other monogenic disorders.

6. REFERENCES.

1. Abbott CM, McMahon CJ, Whitehouse DB, Povey S. Prenatal diagnosis of α-1-antitrypsin deficiency using polymerase chain reaction. *Lancet* 1988;**I**:763-4.
2. Anderson KE, Sassa S, Peterson CM, Kappas A. Increased erythrocyte uroporphyrinogen-I-synthetase, delta-aminolevulinic acid dehydratase and protoporphyrin in hemolytic anemias. *Am J Med* 1977;**63**:359-74.
3. Anderson PM, Reddy RM, Anderson KE, Desnick RJ. Characterization of the porphobilinogen deaminase deficiency in acute intermittent porphyria. *J Clin Invest* 1981;68:**1**-12.
4. Andriola A, Mocelin AJ, Stella SR, Ajzen H, Ramos OL. Determination of erythrocyte uroporphyrinogen I synthase activity in chronic renal failure. *Clin Chim Acta* 1980;**104**:241-4.
5. Antonarakis SE, Boehm CD, Giardina PJV, Kazazian HH Jr. Non-random association of polymorphic restriction sites in the β-globin gene cluster. *Proc Natl Acad Sci USA* 1982;**79**:137-41.
6. Antonarakis SE, Waber PG, Kitur SD, Patel AS, Kazazian HH Jr, Mellis MA, Counts RB, Stamatoyannopoulos G, Bowie EJW, Fass DN, Pitmann DD, Wozney JM, Toole JJ. Hemophilia A: molecular defects and carrier detection by DNA analysis. *N Engl J Med* 1985a:**213**:842-8.
7. Antonarakis SE, Kazazian HH Jr, Orkin SH. DNA

polymorphism and molecular pathology of the human globin gene clusters. *Hum Genet* 1985b;**69**:1-14.

8. Antonarakis SE. Diagnosis of genetic disorders at DNA level. *N Engl J Med* 1989;**320**:153-63.
9. Anwar R, Murray K, Hedge PJ, Smith JC, Markham AF. DNA sequence analysis of the KM19 locus linked to cystic fibrosis. Design of a new oligonucleotide to remove non-specific PCR products. *Hum Genet* 1990;**85**:319-23.
10. Barker D, White R. More base pair change polymorphism at sites containing CpG. *Cytogenet Cell Genet* 1982;**32**:253.
11. Backer D, Schafer M, White R. Restriction sites containing CpG show a higher frequency of polymorphism in the human DNA. *Cell* 1984;**36**:131-8.
12. Bachmann B, Luke W, Hunsmann G. Improvement of PCR amplified DNA sequencing with the aid of detergents. *Nucleic Acids Res* 1990;**18**:1309.
13. Battersby AR, Fookes CJR, Matcham GWJ, McDonald E. Biosynthesis of the pigments of life: Formation of the macrocycle. *Nature* 1980;**285**:17-21.
14. Beaumont C, Porcher C, Picat C, Nordmann Y, Grandchamp B. The mouse porphobilinogen deaminase gene: structural organization, sequence and transcriptional analysis. *J Biol Chem* 1989;**264**:14829-34.
15. Becker DM, Kramer S. The neurological manifestations of porphyria: a review. *Medicine* 1977;**56**:411-23.
16. Bentley AK, Rees DJG, Rizza C, Brownlee GG. Defective propeptide processing of blood clotting factor IX caused

by mutation of arginine to glutamine at position –4. *Cell* 1986,**45**:243-8.

17. Benzer S. On the topography of the genetic fine structure. *Proc Natl Acad Sci USA* 1961;**47**:403-15.
18. Beukeveld GJJ, Wolthers BG, Nordmann Y, Deybach JC, Grandchamp B, Wadman SK. A retrospective study of a patient with homozygous form of acute intermittent porphyria. *J Inher Metad Dis* 1990;**13**:673-83.
19. Biggin MD, Gibson TJ, Hong GF. Buffer gradient gels and ^{35}S label as an aid to rapid DNA sequence determination. *Proc Natl Acad Sci* 1983;**80**:3963-5.
20. Bird AP. DNA methylation and frequency of CpG in animal DNA. *Nucleic Acids Res* 1980;**8**:1499-1504.
21. Blum M, Koel C, Abecassis J. Variations in erythrocyte uroporphyrinogen I synthase activity in nonporphyrias. *Clin Chim Acta* 1978;**87**:119-125.
22. Blundell TL. Porphobilinogen deaminase (EC 4.3.1.8) sequence alignments. Personal communication, 1991.
23. Bonaiti-Pellie C, Phung L, Nordmann Y. Recurrence risk estimation of acute intermittent porphyria based on analysis of porphobilinogen activity: a Bayesian approach. *Am J Med Genet* 1984;**19**:755-62.
24. Bonthron DT, Markham AF, Ginsburg D, Orkin SH. Identification of a point mutation in the adinosine deaminase gene responsible for immunodeficiency. *J Clin Invest* 1985;**75**:894-7.
25. Brow M-A D. Sequencing with *Taq* DNA polymerase. In:

Innis MA, Gelfand DH, Sninsky JJ, White TJ, eds. *PCR Protocols:A Guide to Methods and Applications.* San Diego: Academic Press, Inc., 1990;189-96.

26. Bugawan TL, Saiki RK, Levenson CH, Watson RM, Erlich HA. The use of non-radioactive oligonucleotide probes to analyze enzymatically amplified DNA for prenatal diagnosis and forensic HLA typing. *Biotechnology* 1978;**6**:943-7.

27. Cariello NF, Scott JK, Kat AG, Thilly WG, Keohavong P. Resolution of a missense mutant in human genomic DNA by denaturation gradient gel electophoresis and direct sequencing using in vitro DNA amplification: $HPRT_{Munich}$ *Am J Human Genet* 1988;**42**:726-34.

28. Casanova J-L, Pannetier C, Jaulin C, Kourilsky P. Optimal conditions for directly sequencing double-stranded PCR products with Sequenase. *Nucleic Acids Res* 1990;**18**:4028.

29. Chamberlain JS, Gibbs RA, Ranier JE, Nguyen PN, Caskey CT. Deletion screening of Duchenne muscular dystrophy locus via multiplex DNA amplification. *Nucleic Acids Res* 1988;**16**:11141-55.

30. Chen EJ, Seeburg PH. Supercoil sequencing: A fast and simple method for sequencing plasmid DNA. *DNA* 1985;**4**:165-70.

31. Chou PY, Fasman GD. Empirical predictions of protein conformation. *Annu Rev Biochem* 1978;**47**:251-76.

32. Chirgwin JM, Przybyla AE, MacDonald RJ, Rutter WJ. Isolation of biologically active ribonucleic acid from

sources enriched with ribonuclease. *Biochemistry* 1979;**18**:5294-9.

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

34. Connor JM, Ferguson-Smith MA. Chromosome aberrations. In: *Essential Medical Genetics.* 3rd ed, Oxford: Blackwell Scientific Publications, 1991:9-21.

35. Cooke A, Lanyon WG, Wilcox DE, Dornan ES, Kataki A, Gillard EF, McWhinnie AJM, Morris A, Fergusson-Smith MA, Connor JM. Analysis of Scottish Duchenne and Becker muscular dystrophy families with dystrophin cDNA probes. *J Med Genet* 1990;**27**:292-7.

36. Cotton RGH, Rodrigues NR, Campbell RD. 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* 1988;**85**:4397-4401.

37. Cotton RGH, Campbell RD. Chemical reactivity of matched cytosine and thymine bases near mismatched and unmatched bases in a heteroduplex between DNA strands with multiple differences. *Nucleic Acids Res* 1989;**17**:4223-33.

38. Cotton RGH. Detection of mutations in DNA and RNA by chemical cleavage. In: Mathew C, ed. *Molecular biology* in medicine- Methods in molecular biology Vol 7.

1989a:1-7.

39. Cotton RGH. Detection of single base changes in nucleic acids. *Biochem J* 1989b;**263**:1-10.

40. Cutting GR, Kasch LM, Rosenstein BJ, Zielenski J, Tsui L-C, Antonarakis SE, Kazazian HH. *Nature* 1990;**346**:366-9.

41. Delfau MH, Picat C, de Rooij FWM, Hamer K, Bogard M, Wilson JH, Deybach JC, Nordmann M, Grandchamp B. Two different point G to A mutations in exon 10 of porphobilinogen deaminase gene are responsible for acute intermittent porphyria. *J Clin Ivest* 1990;**86**:1511-6.

42. Dianzani I, Forrest SM, Camaschella C, Saglio G, Ponzone A, Cotton RGH. Screening for mutations in the phenylalanine hydroxylase gene from Italian patients with phenylketonuria by using the chemical cleavage method: a new splice mutation. *Am J Hum Genet* 1991;**48**:631-5.

43. Desnick RJ, Ostasiewicz L, Tishler PA, Mustajoki P. Acute intermittent porphyria: characterization of a novel mutation in the structural gene for porphobilinogen deaminase. *J Clin Invest* 1985;**76**:505-9.

44. de Verneuil H, Phung N, Nordmann Y, Allard D, Leprince F, Jerome H, Aurias A, Rethore MO. Assignment of human uroporphyrinogen I synthase locus to region 11qter by gene dosage effect. *Hum Genet* 1982;**60**:212-3.

45. Duchange N, Chasse J-F, Cohen GN, Zankin NM. Antithrombin gene III tours gene: identification of a

point mutation leading to an arginine to cysteine replacement in a silent deficiency (abstract) *Nucleic* Acids Res 1986;**14**:2408.

46. Dudzinski B, Weinstein AJ. Infectious mononucleosis presenting as acute intermittent porphyria. *South Med J* 1984;**77**:523-5.

47. Embury SH, Miller JA, Dozy AM, Kan YW, Chan V, Todd D. Two different molecular organisations account for the single α-globin gene of the α-thalassemia-2 phenotype. *J Clin Invest* 1980;**66**:1319-25.

48. Embury SH, Scharf SJ, Saiki RK, Gholson MA, Golbus M, Arnheim N, Erlich HA. Rapid prenatal diagnosis of sickle cell anaemia by a new method of DNA analysis. *New Engl J Med* 1987;**316**:656-61.

49. Epstein O, Lahav M, Schoenfeld N, Nemesh L, Shaklai M, Atsmon A. Erythrocyte uroporphyrinogen synthase as a possible diagnostic aid in the diagnosis of lymphoproliferative diseases. *Cancer* 1983;**52**:823-32.

50. Feldman GL, Williamson R, Beaudet AL, O'Brien WE. Prenatal diagnosis of cystic fibrosis by DNA amplification for detection of KM-19 polymorphism. *Lancet* 1988;**II**:102.

51. Felsher BF, Redeker AG. Acute intermittent porphyria: Effect of diet and griseofulvin. *Medicine* 1967;46:217-23.

52. Fischer SG, Lerman LS. DNA fragments differing by single-base pair substitutions are separated in

denaturing gradient gels: correspondence with melting theory. *Proc Natl Acad Sci USA* 1983;**80**:1579-83.

53. Gal S, Hohn B. Direct sequencing of double-stranded DNA PCR products via removing the complementary strand with single-stranded DNA of an M13 clone. *Nucleic Acid Res* 1990;**18**:1076.

54. Gasparini P, Savoia A, Pignatti PF, Dallapicola B, Novelli G. Amplification of RNA from epithelial cells in urine. *N Engl J Med* 1989;**320**:809.

55. Ghosal D, Saedler H. DNA sequence in the mini insertion IS2-6 and its relation to the sequence of IS2. *Nature* 1978;**275**:611-7.

56. Gianelli F, Anson DS, Choo KH, Rees DJG, Winship PR, Ferrari N, Rizza CR, Brownlee GG. Characterisation and use of an intragenic polymorphic marker for detection of carriers of haemophilia B (factor IX) deficiency). *Lancet* 1984;**i**:239-41.

57. Gibbs RA, Caskey CT. Identification and localization of mutation at the Lesch-Nyahan locus by ribonuclease A cleavage. *Science* 1987;**236**:303-5.

58. Gibbs RA, Nguyen P-N, McBride LJ, Koepf SM, Caskey CT. Identification of mutations leading to Lesch-Nyhan syndrome by automated direct DNA sequencing of *in vitro* amplified cDNA. *Proc Natl Acad Sci USA* 1989;86:1919-23.

59. Gitschier J, Wood WI, Tuddenham EGD, Shuman MA, Goraka TM, Chen EY. Lawn RM. Detection and sequence of mutation in the factor VIII gene of haemophiliacs.

Nature 1985;**315**;427-30.

60. Glisin V, Crkvenjakov R, Byus C. Ribonucleic acid isolated by caesium chloride centrifugation. *Biochemistry* 1974;**13**:2633-7.

61. Goldberg A. Acute intermittent porphyria. *Quart J Med* 1959;**28**:183-209.

62. Goldberg A, Moore MR, McColl KEL, Brodie MJ. Porphyrin metabolism and the porphyrias, In: Ledingham JGG, Warrell DA, Weatherall, eds. *Oxford text book of medicine.* 2nd ed, Oxford: Oxford University Press, 1987:136-45.

63. Graham A, Kalsherker NA, Bamforth FJ, Newton CR, Markham AF. Molecular characterisation of two α-1-antitrypsin deficiency variants: proteinase inhibitor (Pi) $Null_{Newport}$ (Gly^{115}->Ser) and (Pi) Z Wrexham (Ser^{-19}->Leu). *Hum Genet* 1990;**85**:537-40.

64. Grandchamp B, de Verneuil H, Beaumont C, Chretien S, Walter O, Nordman Y. Tissue-specific expression of porphobilinogen deaminase: two isoenzymes from a single gene. *Eur J Boichem* 1987;**162**:105-10.

65. Grandchamp B, Nordmann Y. Enzymes of heme biosynthesis pathway: recent advances in molecular genetics. *Semin Hematol* 1988;**25**:303-11.

66. Grandchamp B, Picat C, Mignotte V, Wilson JHP, Te Velde K, Sandkuyl L, Romeo PH, Gossens M, Nordmann Y. Tissue-specific splicing mutation in acute intermittent porphyria. *Proc Natl Acad Sci USA* 1989a;**86**:661-4.

67. Grandchamp B, Picat C, Kauppien R, Mignotte V,

Peltonen L, Mustajoki PH, Romeo PH, Gossens M, Nordmann Y. Molecular analysis of acute intermittent porphyria in a Finnish family with normal erythrocyte porphobilinogen deaminase. *Eur J Clin Invest* 1989b;**19**:415-8.

68. Grandchamp B, Picat C, de Rooij F, Beaumont C, Wilson P, Deybach JC, Nordmann Y. 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* 1989c;**17**;6637-49.
69. Grandchamp B, Delfau MH, Picat C, de Rooij FWM, Nordmann Y. Heterogeneity of molecular defects in acute intermittent porphyria. *Am J Human Genet* (Supplement) 1990;**47**:Abstr. A.156 (0612) 6.2.
70. Grompe M, Muzny DN, Caskey CT. Scanning detection of mutations in human ornithine transcarbamoylase by chemical mismatch cleavage. *Proc Natl Acad Sci USA* 1989;**86**:5888-92.
71. Grompe M, Caskey CT, Fenwick RG. Improved diagnosis for ornithine transcarbamylase deficiency. *Am J Hum Genet* 1991;**48**:212-22.
72. Gu XF, Lee J-S, Delfau MH, Grandchamp B. PCR detection of a G/T polymorphism at exon 10 of the porphobilinogen deaminase gene. *Nucleic Acids Res* 1991 (In press).
73. Gyllensten UB, Ehrich HA. 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* 1988;**85**:7652-6.

74. Gyllensten U. Direct sequencing of *in vitro* amplified DNA. In: Erlich HE, ed. *PCR Technology: Principles and Application for DNA Amplification.* New York: Stockton Press. 1989;45-60.

75. Haliassos A, Chomel JC, Grandjouan S, Kruth J, Kaplan JC, Kitzis A. Detection of minority point mutations by a modified PCR technique: a new approach for a sensitive diagnosis of tumor-progression markers. *Nucleic Acids Res* 1989;**17**:8093-9.

76. Haqqi TM, Sarkar G, David CS, Sommer SS. Specific amplification with PCR of a refractory segment of genomic DNA. *Nucleic Acids Res* 1988;**16**:11844.

77. Harbath P, Vosberg H-P. Enzymatic amplification of myosin heavy-chain mRNA sequences *in vitro*. *DNA* 1988;**7**:297-306.

78. Hattori M, Sakaki Y. Dideoxy sequencing method using denatured plasmid templates. *Anal Bioichem* 1986;**152**:232-8.

79. Higuchi R, von Beroldigen CH, Sensabaugh GF, Erlich HA. DNA typing from single hairs. *Nature* 1988a;**332**:543-6.

80. Higuchi R, Krummel B, Saiki RK. A general method of *in vitro* preparation and specific mutagenesis of DNA fragments: study of protein and DNA interactions. *Nucleic Acids Res* 1988b;**15**:7351-67.

81. Higuchi RG, Ochman H. Production of single-stranded DNA templates by exonuclease digestion following the

polymerase reaction. *Nucleic Acids Res* 1989;**17**:5865.

82. Higuchi M, Wong C, Kochhan L, Olek K, Aronis S, Kasper CF, Kazazian HH Jr. Characterization of mutations in the factor VIII gene by direct sequencing of amplified genomic DNA. *Genomics* 1990;**6**:65-71.
83. Hong GF. The use of DNase I buffer gradient gel, and ^{35}S label for DNA sequencing. *Methods Enzymol* 1987;**155**:93.
84. Horton RM, Hunt HD, Ho SN, Pullen JK, Pease LR. Engineering hybrid genes without the use of restriction enzymes: gene splicing by overlap extension. *Gene* 1989;**77**:61-8.
85. Howells DW, Forrest SM, Dahl H-H M, Cotton RGH. Insertion of an extra codon for threonine is a cause of dihydropteridine reductase deficiency. *Am J Hum genet* 1990;**47**:279-85.
86. Hsiao K-C. A fast and simple procedure for sequencing double-stranded DNA with Sequenase. *Nucleic Acids Res* 1991;**19**:2787.
87. Hughes MJ, Rifkind AB. Danazol, a new steroidal inducer of delta-aminolevulinic acid synthetase. *J Clin Endocrinol Metab* 1981;**52**:549-52.
88. Hultman T, Stahl S, Hornes E, Uhlen M. Direct solid phase sequencing of genomic and plasmid DNA using magnetic beads as solid support. *Nucleic Acids Res* 1989;**17**:4937-46.
89. Hung T, Mak K, Fong K. A specificity enhancer for polymerase chain reaction. *Nucleic Acids Res*

1990;18:4953.

90. Huth A, Estivill X, Grade K, Billwitz H, Speer A, Rosenthal A, Williamson R, Ramsay M, Coutelle C. Polymerase chain reaction for detection of the pMP6d-9/*Msp*I RFLP, a marker closely linked to the cystic fibrosis mutation. *Nucleic Acids Res* 1989;**17**:7118.

91. Iannuzi MC, Stern RC, Collins FS, Hon CT, Hidaka N, Strong T, Becker L, Drumm ML, White, MB, Gerrard B, Dean M. Two frameshift mutations in the cystic fibrosis gene. *Am J Hum Genet* 1991;**48**:227-31.

92. Impraim CC, Saiki RK, Erlich HA, Teplitz RL. Analysis of DNA extracted from formalin-fixed, formalin-embedded tissues by enzymatic amplification and hybridization with sequence-specific oligonucleotides. *Biochem Biophys Res Commun* 1987;**142**:710-6.

93. Innis MA, Myambo KE, Gelfand DH, Brow MAD. DNA sequencing with *Thermus aquaticus* DNA polymerase and direct sequencing of polymerase chain reaction-amplified DNA. *Proc Natl Acad Sci USA* 1988;**85**:9436-40.

94. Innis MA, Gelfand DH. Optimization of PCRs. In: Innis MA, Gelfand DH, Sninsky JJ, White TJ, eds. *PCR Protocols: a Guide to Methods and Applications.* San Diego: Academic Press Inc., 1990:3-12.

95. Jackson DP, Quirke P, Lewis F, Boylston AW, Sloan JM, Robertson D, Taylor GR. Detection of measles virus RNA in paraffin-embedded tissue. *Lancet* 1989;**I**:1391.

96. Jeffreys AJ. DNA sequence variants in the G_{gamma}- and

A_{gamma}- and delta- and β- globin genes of man. *Cell* 1979;**18**:1-18.

97. Kappas A, Sassa S, Anderson KE. The porphyrias. In: Stanbury JB, Wyngaarden JB, Frederickson DS, Goldstein JL, Brown MS, eds. *The Metabolic Basis of Inherited Disease.* 5th ed. New York: McGraw-Hill, 1983:1301-84.

98. Kauppinen R, Peltonen L, Palotie A, Mustajoki P. RFLP analysis of three different types of acute intermittent porphyria. *Hum Genet* 1990;**85**:160-4.

99. Kawasaki ES. Amplification of RNA. In: Innis MA, Gelfand DH, Sninsky JJ, White TJ, eds. *PCR Protocols: a Guide to Methods and Applications.* San Diego: Academic Press Inc., 1990:21-7.

100. Kawasaki ES, Wang AM. Detection of gene expression. In: Erlich H, ed. *PCR Technology: Principles and Applications for DNA Amplification.* New York: Stockton Press, 1989:89-97.

101. Kidd KK, Bowcock AM, Person PL, Schmidtke J, Willard HF, Track RK, Ricciuti F. Report of the committee on the human gene mapping by recombinant DNA techniques. *Cytogenet Cell Genet* 1989;**49**:132-218.

102. Kidd VJ, Wallace RB, Itakura K, Woo SLC. α_1-antitrypsin deficiency detection by direct analysis of the mutation in the gene. *Nature* 1983;**304**:230-4.

103. Kim H-S, Smithies O. Recombinant fragment assay for gene targeting based on polymerase chain reaction. *Nucleic Acids Res* 1988;**16**:8887-903.

104. Kogan SC, Doherty M, Gitschier J. An improved method for prenatal diagnosis of genetic diseases by analysis of amplified DNA sequences. *N Engl Med J* 1987;**317**:985-90.

105. Konecki DS, Lichter-Konecki U. The phenylketonuria locus: current knowledge about alleles and mutations of the phenylalanine hydroxylase gene in various populations. *Hum Genet* 1991;**87**:377-88.

106. Kreitman M, Landweber LR. A stratergy for producing single-stranded DNA in the polymerase chain reaction: a direct method for genomic sequencing. *Gene Anal Techn* 1989;**6**:84-8.

107. Kretz KA, Carson GS, O'Brien JS. Direct sequencing from low-melt agarose with Sequenase. *Nucleic Acids Res* 1989;**17**:5864.

108. Kunkel LM. Analysis of deletions in DNA from patients with Becker and Duchenne muscular dystrophy. *Nature* 1986;**322**:73-7.

109. Lamon JM, Frykholm BC, Tschudy DP. Family evaluations in acute intermittent porphyria using red cell uroporphyrinogen-1-synthase. *J Med Genet* 1979;**16**:134-9.

110. Lannfelt L, Wetterberg L, Lilius L, Thunnel S, Jornvall H, Pavlu B, Wielburski A, Gellerfors P. Porphobilinogen deaminase in human erythrocytes: purification of two forms with apparent molecular weights

of 40 kDA and 42 kDA. *Scand J Clin Lab Invest* 1989;**49**:677-84.

111. Lee J-S. In: *Molecular genetic investigation of the* human porphobilinogen deaminase gene in acute intermittent porphyria. Stockholm: Repro Print, 1991a:8-50.

112. Lee J-S. Alternative dideoxy sequencing of double-stranded DNA by cyclic reactions using *Taq* Polymerase. *DNA* 1991b;**16**:67-73.

113. Lee J-S and Anvret M. A *Pst*I polymorphism for human porphobilinogen deaminase gene. *Nucleic Acids Res* 1987;**15**:6307.

114. Lee J-S, Anvret M, Lindsten J, Lannfelt L, Gellerfors P, Wetterberg L, Floderus L, Thunell S. DNA polymorphisms within the porphobilinogen deaminase gene in two Swedish families with acute intermittent porphyria. *Hum Genet* 1988;**79**:379-81.

115. Lee J-S, Lindsten J, Anvret M. Haplotyping of the human porphobilinogen deaminase gene in acute intermittent porphyria by polymerase chain reaction. *Hum Genet* 1990;**84**:241-3.

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

117. Lee J-S, Anvret M. Identification of a major mutation in human porphobilinogen deaminase gene in Swedish patients with acute intermittent porphyria. In: *Molecular genetic investigation of the human porphobilinogen deaminase gene in acute intermittent porphyria.* Stockholm: Repro Print, 1991:1-8.

118. Lee JT, Nussbaum L. An arginine to glutamine mutation in residue 109 of human ornithine transcarbamylase completely abolishes enzymatic activity in Cos cells. *J Clin Invest* 1989;**84**:1762-6.

119. Lehrman MA, Russell DW, Goldstein JL, Brown MS. Exon-Alu recombination deletes 5 kilobases from the low density lipoprotein receptor gene, producing a null phenotype in familial hypercholesterolemia. *Proc Natl Acad Sci USA* 1986;**83**:3679-83.

120. Lench N, Stanier P, Williamson R. Simple non-invasive method to obtain DNA for gene analysis. *Lancet* 1988;**I**:1356-8.

121. Levit EJ, Noddine JH, Perloff WH. Progesterone-induced porphyria. *Am J Med* 1957;**22**:831-3.

122. Li H, Gyllensten UB, Cui X, Saiki RK, Erlich HA, Arnheim N. Amplification and analysis of DNA sequences in single human sperm and diploid cells. *Nature* 1988;**335**:414-7.

123. Lim HM, Pene JJ. Optimal conditions for supercoil DNA sequencing with *Escheria coli* DNA Polymerase I large fragment. *Gene Anal Techn* 1988;**5**:32-9.

124. Little JW, Lehman AR, Kaiser AD. An exonuclease induced by bacteriophage lambda. *J Bio Chem* 1967;**242**:672-8.

125. Llewellyn DH, Elder GH, Kalsheker NA, Marsh OMW, Harrison PR, Grandchamp B, Picat C, Nordmann Y, Romeo PH, Goossens M. DNA polymorphism of human porphobilinogen deaminase gene in acute intermittent porphyria. *Lancet* 1987;**II**:706-8.

126. Llewellyn DH, Urquhart A, Scobie G, Elder GH, Kalsheker NA, Harrison PR. Molecular analysis of acute intermittent porphyria. *Biochem Soc Trans* 1989;**16**:799-800.

127. Lo Y-MD, Mehal WZ, Fleming KA. Rapid production of vector-free biotinylated probes using the polymerase chain reaction. *Nucleic Acids Res* 1988;**16**:8719.

128. MacDonald RJ, Swift JH, Przybyla AE, Chirgwin JM. Isolation of RNA using guanidium salts. *Methods Enzymol* 1987;**152**:219-27.

129. Mandel JL, Chambon P. DNA methylation: organ specific variations in the methylation pattern within and around ovalbumin and other chicken genes. *Nucleic Acids Res* 1979;**7**:2081-3.

130. Maniatis T, Fritsch EF, Sambrook J. Strand separation gels. In: *Molecular cloning: A laboratory manual.* Cold Spring Harbor Laboratory, Cold Spring Harbor, New York: 1982:179-81.

131. Maxam AM, Gilbert W. A new method for sequencing DNA. *Proc Natl Acad Sci USA* 1977;**74**:560-4.

132. McCabe PC. Production of single-stranded DNA by asymetric PCR. In: Innis MA, Gelfand DH, Skinsky JJ, White TJ, eds. *PCR protocols: A guide to methods and applications.* San Diego: Academic Press, 1990:76-83.

133. McColl KEL, Moore MR, Goldberg A. The Porphyrias. In: Rakel RE, ed. *Conn's Current Therapy.* Philadelphia: Saunders, 1986:349-57.

134. McKeran RO, Andrews TM, Howells L, Gibbs DA, Chinn S, Watts RWE. The diagnosis of carrier state for the Lesch-Nyhan syndrome. *Q J Med* 1975;**44**:189-205.

135. McKusick VA. *Mendelian inheritance in man.* 9th ed. Baltmore: Johns Hopkins University Press, 1990.

136. Meisler M, Wanner L, Eddy RE, Shows TB. The UPS locus encoding uroporphyrinogen I synthase is located on chromosome 11. *Biochem Biophys Res Commun* 1980;**95**:170-6.

137. Meisler MH, Wanner LA, Kao FT, Jones C. Localization of the uroporphyrinogen I synthase locus to human chromosome region 11q13-qter and interconversion of enzyme isomers. *Cytogenet Cell Genet* 1981;**31**:124-8.

138. Mgone CS, Lanyon WG, Moore MR, Connor JM. Detection of seven point mutations in porphobilinogen deaminase gene in patients with acute intermittent porphyria, by direct sequencing of *in vitro* amplified cDNA. (In press, 1991).

139. Miller AD, Hart GJ, Packman LC, Battersby AR. Evidence that the pyrromethane cofactor of hydroxylmethylbilane synthase (porphobilinogen deaminase) is bound the protein through the sulphur atom of cysteine-242. *Biochem J*

1988;**254**:915-18.

140. Monaco AP, Bertelson CJ, Middleswoth W, et al. Detection of deletions spanning the Duchenne muscular dystrophy locus using a tightly linked DNA segment. *Nature* 1985;**316**:842-5.

141. Montandon AJ, Green PM, Gianelli F, Bentley DR. Direct mutations of point mutations by mismatch analysis: application to haemophilia B. *Nucleic Acids Res* 1989;**17**:3347-58.

142. Montandon AJ, Green PM, Bentley DR, Ljung R, Nilsson IM, Gianelli F. Two factor IX mutations in the family of an isolated haemophilia B patient: direct carrier diagnosis by amplification mismatch detection (AMD). *Hum Genet* 1990;**85**:200-4.

143. Moore MR. Laboratory investigations of disturbances of porphyrin metabolism. *Association of Clinical Pathologists Broadsheet 109.* London: British Medical Association. 1983:1-16.

144. Moore MR, McColl, Rimington C, Goldberg A. The acute porphyrias. In: *Disorders of porphyrin metabolism.* London: Plenum Press, 1987:73-117.

145. Mount SM. A catalogue of splice junction sequences. *Nucleic Acids Res* 1982;**10**:459-72.

146. Murphy G, Ward ES. Sequencing of double-stranded DNA. In: Howe CJ, Ward ES, eds. *Nucleic acid sequencing: a* practical approach. Oxford: IRL Press, 1989:99-115.

147. Mustajoki P, Desnick RJ. Genetic heterogeneity in acute intermittent porphyria: characterisation and frequency of porphobilinogen deaminase mutations in Finland. *Br Med J* 1985;**291**:505-9.

148. Myers RM, Lumelsky N, Lermam LS, Maniatis T. Detection of single base substitutions in total genomic DNA. *Nature* 1985a;**313**:395-8.

149. Myers RM, Fischer SG, Maniatis T, Lerman LS. Modification of the melting properties of duplex DNA by attachment of a GC-rich DNA sequence as determined by denaturing gradient gel electrophoresis. *Nucleic Acids Res* 1985b;**13**:1311-29.

150. Myers, Fischer SG, Lerman LS, Maniatis T. Nearly all single base substitutions in DNA fragments joined to a GC-clamp can be detected by denaturing gradient gel electrophoresis. *Nucleic Acids Res* 1985c;**13**:3131-45.

151. Myers RM, Larin Z, Maniatis, T. Detection of single base substitutions by ribonuclease cleavage at mismatches in RNA:DNA duplexes. *Science* 1985d;**230**:1242-6.

152. Myers RM, Maniatis T, Lerman LS. Detection and localization of single base changes by denaturing gradieent gel electrophoresis. *Methods Enzymol* 1987;**155**:501-27.

153. Myers RM, Sheffield VC, Cox DR. Detection of single base changes in DNA: ribonuclease cleavage and denaturation gradient gel electrophoresis. In: Davies KE, ed. *Genome analysis: a practical approach.* Oxford: IRL Press, 1988:95-139.

154. Myers RM, Sheffield VC, Cox DR. Mutation detection by PCR, GC-clamps and denaturing gradient gel electrophoresis. In: Erlich HA, ed. *PCR Technology: Principles and Applications for DNA Amplification.* New York: Stockton Press. 1989: 71-87.

155. Nakamaye KL, Gish G, Eckstein F, Vosberg H-P. Direct sequencing of polymerase chain reaction amplified DNA fragments through the incorporation of deoxynucleoside α-thiotriophosphates. *Nucleic Acids Res* 1988;**16**:9947-59.

156. Newton CR, Kalsheker N, Graham A, Powell S, Gammack A, Riley J, Markham AF. Diagnosis of α_1-antitrypsin defficiency by enzymatic amplification of human genomic DNA and direct sequencing of polymerase chain reaction products. *Nucleic Acid Res* 1988;**16**:8233-43.

157. Newton CR, Graham A, Heptinstall LE, Powell SJ, Summers C, Kalsherker N, Smith JC, Markam JF. Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS). *Nucleic Acids Res* 1989;**17**:2503-16.

158. Nordmann Y, de Verneuil H, Deybach JC, Delfau MH,

Grandchamp B. Molecular genetics of porphyrias. *Ann Med* 1990;**22**:387-91.

159. Northrup H, Rosenbloom C, O'Brien WD, Beaudet AI. Additional polymorphism for D7S8 linkage to cystic fibrosis including detection by DNA amplification. *Nucleic Acids Res* 1989;**17**:1784.

160. Nussinov R. Eukaryotic dinucleotide preference rules and their implications for degenerate codon usage. *J Mol Biol* 1981;**149**:125-31.

161. Okayama H, Berg P. High efficiency cloning of full-length cDNA. *Mol Cell Biol* 1982;**2**:161-70.

162. Orita M, Iwahana H, Kanazawa H, Hayashi K, Sekiya T. Detection of polymorphism of human DNA by gel electrophoresis as single-strand conformation polymorphysims. *Proc Natl Acad Sci USA* 1989a;**86**:2766-70.

163. Orita M, Suzuki Y, Sekiya T, Hayashi K. Rapid and sensitive detection of point mutations and DNA polymorphysim using the polymerase reaction. *Genomics* 1989b;**5**:874-9.

164. Orkin SH, Goff SC. Nonsense and frameshift mutations in β-thalassemia detected in in cloned β-globin genes. *J Biol Chem* 1981;**256**:9782-4.

165. Orkin SH, Sexton JP, Cheng, TC. et al. ATA box transcription mutation in β-thalassemia. *Nucleic Acids Res* 1983;**11**:4727-34.

166. Paabo S, Gifford JA, Wilson AC. Mitochondrial DNA sequences from a 7000-year old brain. *Nucleic Acids*

Res 1988;**16**:9775-87.

167. Picat C, Delfau MH, de Rooij FWM, Beukeveld GJJ, Wolthers BG, Wadman SK, Nordmann Y, Grandchamp B. Identification of mutations on the parents of a patient with a putative compound heterozygosity for acute intermittent porphyria. *J Inher Metab Dis* 1990;**13**:684-6.

168. Picat C, Bourgeois F, Grandchamp B. PCR detection of a C/T polymorphism in exon 1 of the porphobilinogen deaminase gene (PBGD). *Nucleic Acids Res* 1991;**18**:5099.

169. Piepkorn MW, Hamernyik P, Labbe RF. Modified erythrocyte uroporphyrinogen-1-synthase assay and its clinical interpretation. *Clin Chem* 1978;**24**:1751-4.

170. Pierach CA, Weimer MK, Cardinal RA, Bossenmaier C, Blommer R. Red blood cell porphobilinogen deaminase in the evaluation of acute intermittent porphyria. *JAMA* 1987;**257**:60-1.

171. Prochovnick EV, Antonarakis S, Bauer KA, Rosenberg RD, Fearon ER, Orkin SH. Molecular heterogeneity of inherited antithrombin III deficiency. *N Engl J Med* 1983;**308**:1549-52.

172. Putney SD, Benkovic SJ, Schimmel PR. A DNA fragment with an α-phosphorothioate nucleotide at one end is asymmetrically blocked from digestion by exonuclease III and can be replicated *in vivo*. *Proc Natl Sci USA* 1981;**78**:7350-4.

173. Raich N, Romeo PH, Dubart A, Beaupain D, Cohen-Solal M, Goossens M. Molecular cloning and complete primary

sequence of human erythrocyte porphobilinogen deaminase. *Nucleic Acid Res* 1986;**14**:5955-68.

174. Rappolee DA, Mark D, Banda MJ, Werb Z. Wound macrophages express TGF-*a* and other growth factors *in vivo*: analysis by mRNA phenotyping. *Science* 1988;**241**:708-12.

175. Ray PN, Belfall B, Duff C, Logan C, Kean V, Thompson MW, Sylvester JE, Gorski JL, Schmickle RD, Worton RG. Translocation associated with Duchenne muscular dystrophy. *Nature* 1985;**318**:672-5.

176. Read AP, Mountford RC, Forrest SM, Kenwrick SJ, Davies KE, Harris R. Patterns of exon deletion in Duchenne and Becker muscular dystrophy. *Hum Genet* 1988;**80**:152-6.

177. Reiss J, Neufeldt U, Wieland K, Zoll B. Diagnosis of haemophilia B using the polymerase chain reaction. *Blut* 1990;**60**:31-6.

178. Ristaldi MS, Pirastu M, Rosatelli C, Monni G, Erlich H, Saiki R, Cao A. Prenatal diagnosis of β-thalassaemia in mediterranean populations by dot blot analysis with DNA amplification and allele specific oligonucleotide probes. *Prenat Diag* 1989;**9**:629-38.

179. Romeo G, Hassan HJ, Staempfli S, Roncuzzi L, Cianetti L, Leonard A, Vincente V, Mannucci PM, Bertina R, Peschle C, Cortese R. Hereditary thrombophilia: identification of nonsense and missense mutations in protein C gene. *Proc Natl Acad Sci USA* 1987;**84**:2829-32.

180. Rosenbloom CL, Kerem BS, Rommens JM, Tsui LC,

Wainwright B, Williamson R, O'Brein WE, Beaudet AL. DNA amplification for detection of the XV-2c polymorphism linked to cystic fibrosis. *Nucleic Acids Res* 1989;**17**:7117.

181. Saha, BK. Amplification of DNA by PCR for direct sequencing. *Comments* 1989;**16**:17-8.

182. Saiki RK, Scharf S, Faloona F, Mullis KB, Horn GT, Erlich H, Arnheim N. Enzymatic amplification of β-globin genomic sequences and restriction site analysis of sickle cell anemia. *Science* 1985;**230**:1350-4.

183. Saiki RK. Bugawan TL, Horn GT, Mullis KB, Erlich HA. Analysis of enzymatically amplified β-globin and HLA-DQ*a*DNA with allele-specific oligonucleotide probes. Nature 1986;324:163-8.

184. Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB, Erlich HA. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* 1988a;**239**:487-91.

185. Saiki RK, Chang C-A, Levenson CH, Warren TC, Boehm CD, Erlich HA. Diagnosis of sickle cell anemia and β-thalassemia with enzymatically amplified DNA and nonradioactive allele-specific oligonucleotide probes. *N Engl J Med* 1988b;**319**:537-41.

186. Saiki RK, Gyllensten UB, Erlich HA. The polymerase chain reaction. In: Davies KE, ed. *Genome analysis: a practical approach.* 1988c:141-52.

187. Saiki RK. The design and optimization of the PCR. In:

Erlich H, ed. *PCR Technology: Principles and Applications for DNA Amplification.* New York: Stockton Press, 1989:7-16.

188. Saiki RK, Walsh PS, Levenson CH, Erlich HA. Genetic analysis of amplified DNA with immobilized sequence-specific oligonucleotide probes. *Proc Natl Acad Sci USA* 1989b;**86**:6230-4

189. Sambrook J, Fritsch EF, Maniatis T. Sequencing amplified DNA by Sanger dideoxy-mediated chain-termination method. In: *Molecular cloning: A laboratory manual, 2nd ed.* Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press, 1989a;14.20-35.

190. Sambrook J, Fritsch EF, Maniatis T. Spun-column chromatography. In: *Molecular cloning: A laboratory manual, 2nd ed.* Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press, 1989b;E.37-E.38.

191. Sambrook J, Fritsch EF, Maniatis T. Extraction and purification of RNA In: *Molecular cloning: A laboratory manual, 2nd ed.* Cold Spring Harbor, New York: Cold Spring Harbor Laboratory, Press 1989c: 7.3-7.25.

192. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain terminating inhibitors. *Proc Natl Acad Sci* 1977;**74**:5463-7.

193. Sarkar G, Sommer SS. RNA amplification with transcript sequencing (RAWTS). *Nucleic Acids Res* 1988;**16**:5197.

194. Sarkar G, Sommer SS. Access to a messenger RNA sequence or its protein product is not limited by tissue or

species specificity. *Science* 1989;**244**:331-4.

195. Sarkar G, Kapelner S, Sommer SS. Formamide can dramatically improve the specificity of PCR. *Nucleic Acid Res* 1990;**18**:7465.

196. Sarkar G, Sommer SS. Shedding light on PCR contamination. *Nature* 1990;**343**:27.

197. Scharf SJ, Horn GT, Erlich HA. Direct cloning and sequence analysis of enzymatically amplified genomic sequences. *Science* 1986;**233**:1076-8.

198. Schofield JP, Vaudin M, Kettle S, Jones DSC. A rapid semi-automated plate method for analysis and sequencing by PCR from bacterial stocks. *Nucleic Acids Res* 1989;**17**:9498.

199. Scobie GA, Llewellyn DH, Urquhart AJ, Smyth SJ, Kalsheker NA, Harrison PR, Elder GH. Acute intermittent porphyria caused by a C to T mutation that produces a stop codon in the porphobilinogen deaminase gene. *Hum Genet* 1990a;**85**:631-4.

200. Scobie GA, Urquhart AJ, Elder GH, Kalsheker NA, Smyth J, Harrison PR. Linkage disequilibrium between DNA polymorphisms within the porphobilinogen deaminase gene. *Hum Genet* 1990b;**85**:157-9.

201. Sela M, Anfinsen CB, Harrington WF. The correlation of ribonuclease activity with specific aspects of tertiary structure. *Biochim Biophys Acta* 1957;26:502-12.

202. Shabasaki Y, Kawakami T, Kanazawa Y, Akanuma Y, Takaku F. Posttranslational cleavage of proinsulin is blocked by a

point mutation in familial hyperproinsulinemia. *J Clin Invest* 1985;**76**:378-80.

203. Sharif AL, Smith AG, Abell C. Isolation and characterization of a cDNA clone for chlorophyll synthesis enzyme in *Euglina gracilis. Eur J Biochem* 1989;**184**:353-9.

204. Sheffield VC, Cox DR, Lerman LS, Myers RM. Attachment of a 40-base-pair G+C-rich sequence (GC-clamp) to genomic DNA fragments by the polymerase chain reaction results in improved detection of single-base changes. *Proc Natl Acad Sci USA* 1989;**86**:232-6.

205. Sherman DR, Geliebter J, George AM. Rapid and simple amplification of a specific RNA sequence by polymerase chain reaction. *Trends Genet* 1989;**5**:137.

206. Shibata DK, Arnheim N, Martin WJ. Detection of human papilloma virus in paraffin-embedded tissue using the polymerase chain reaction. *J Exp Med* 1988;**167**:225-30.

207. Stein JA, Tschudy DP. Acute intermittent porphyria: a clinical and biochemical study of 46 patients. *Medicine* 1970;**49**:1-16.

208. Stoflet ES, Koebel DD, Sarkar G, Sommer SS. Genomic amplification with transcript sequencing. *Science* 1988;**239**:491-4.

209. Stubnicer AC, Picat C, Grandchamp B. Rat porphobilinogen deaminase cDNA: nucleotide sequence of the erythropoetic form. *Nucleic Acids Res* 1988;**16**:3102.

210. Studencki AB, Wallace RB. Allele-specific hybridization using oligonucleotide probes of very high specific activity: discrimination of the human β A- and β S-globin genes. *DNA* 1984;3:7-15.

211. Sudhof TC, Goldstein JL, Browm MS, Russell, DW. The LDL receptor gene: a mosaic of exons shared with different proteins. *Science* 1985;228:815-20.

212. Svensson E, Andersson B, Hagenfeldt L. Two mutations within the coding sequence of the phenylalanine hydroxylase gene. *Hum Genet* 85;300-4.

213. Tabor S, Richardson CC. Effect of manganese ions on the incorporation of dideoxynucleotides by bacteriophage T7 DNA polymerase and *Escherichia coli* DNA polymerase I. *Proc Natl Acad Sci USA* 1989;86:4076-80.

214. Tabor S, Richardson CC. DNA sequence analysis with a modified bacteriophage T7 DNA polymerase- Effect of pyrophosphorysis and metal ions. *J Biol Chem* 1990;265:8322-8.

215. Thomas SD, Jordan P. Nucleotide sequence of the *hem*C locus encoding porphobilinogen deaminase of *Escherichia coli* K 12. *Nucleic Acid Res* 1986;14:6215-26.

216. Tindall KR, Kunkel TA. Fidelity of DNA synthesis by the *Thermus aquaticus* DNA polymerase. *Biochemistry* 1988;27:6008-6013.

217. Traystman MD, Higuchi M, Kasper CK, Antonarakis SE, Kazazian HH. Use of denaturing gradient gel electrophoresis to detect point mutation in the factor

VIII gene. *Genomics* 1990;**6**:293-301.

218. Tschudy DP, Valsamis M, Magnussen CR. Acute intermittent porphyria: Clinical and selected research aspects. *Ann Intern Med* 1975;**83**:851-64.

219. Ullrich A, Shine J, Chirgwin J, Pictet R, Tischer E, Rutter WJ, Goodman HM. Rat insulin genes: construction of plasmids containing the coding sequences. *Science* 1977;**196**:1313-9.

220. Umanoff H, Russel CS, Cosloy SD. Availability of porphobilinogen controls appearance of porphobilinogen deaminase activity in *Escherichia coli* K-12. *J Bacteriol* 1988;**170**:4969-71.

221. Vallette F, Mege E, Reiss A, Adesnik M. Construction of mutant and chimeric genes using the polymerase chain reaction. *Nucleic Acids Res* 1989;**17**:723-33.

222. Vogelstein B, Gillespie D. Preparative and analytical purification of DNA from agarose. *Proc Natl Acad Sci, USA* 1979;**76**:615-9.

223. Vosberg H-P. The polymerase chain reaction: an improved method for analysis of nucleic acids. *Hum Genet* 1989;**83**:1-15.

224. Wang AL, Arredondo-Vega FX, Giampietro PF, Smith M, Anderson WF, Desnick RJ. Regional assignment of human porphobilinogen deaminase and esterase A4 to chromosome 11q23-11qter. *Proc Natl Acad Sci USA* 1981;**78**:5734-8.

225. Wang T, Okano Y, Eisensmith R, Huang S-Z, Zeng Y-T, Lo WHY, Woo SLC. Molecular genetics of phenylketonuria

in Orientals: linkage disequilibrium between a termination mutation and haplotype 4 of the phenylalanine hydroxylase gene. *Am J Hum Genet* 1989;**45**;675-80.

226. Wang T, Okano Y, Eisensmith R, Fekete G, Schuler D, Berensci G, Nasz I, Woo SLC. Molecular genetics of PKU in eastern Europe: A nonsense mutation associated with haplotype 4 of the phenylalanine hydroxylase gene. *Somat Cell Mol Genet* 1990;**16**:85-90.

227. Ward ES, Howe CJ. Trouble shooting in chain-termination DNA sequencing. In: Ward ES, Howe CJ, eds. *Nucleic acids sequencing: a practical approach.* Oxford: IRL Press, 1989:79-97.

228. Ward MA, Skandalis A, Glickman BW, Grosovsky AJ. Rapid generation of specific single-stranded template DNA from PCR amplified material. *Nucleic Acids Res* 1989;**17**:8394.

229. Watson JD, Hopkins NH, Roberts JW, Steitz JA, Weiner AM. The unexpected structure of eucaryotic genomes. In: *Molecular biology of the gene.* Menlo Park, California: Benjamine/Cummings Publishing Company, Inc. 1987:621-673.

230. Weatherall DJ. The molecular pathology of single gene disorders In: *The new genetics and clinical practice 2nd ed.* Oxford: Oxford University Press, 1985:66-90.

231. Welland FH, Hellmen ES, Collins A, Hunter GW, Tschudy DP. Factors affecting the excretion of porphyrin precursors by patients with acute intermittent porphyria. *Metabolism* 1964;**13**:251-8.

232. White TJ, Arnheim N, Erlich HA. The polymerase chain reaction. *Trends Genet* 1989;**5**:185-9.

233. White MB, Amos J, Hsu JMC, Gerrad B, Finn P, Dean M. A frame-shift mutation in the cystic fibrosis gene. *Nature* 1990;**344**:665-7.

234. Williams C, Weber L, Williamson R, Hjelm M. Guthrie spots for DNA-based carrier testing in cystic fibrosis. *Lancet* 1988;**II**:693.

235. Wilson JM, Stout JT, Pallela TD, Davidson BL, Kelley WN, Caskey CT. A molecular survey of hypoxanthine-guanine phosphoribosyltransferase deficiency in man. *J Clin Invest* 1986;**77**:188-95.

236. Winship PR, Rees DJG, Alkan M. Detection of polymorphisms at cytosine phosphoguanidine dinucleotides and diagnosis of haemophilia B carriers. *Lancet* 1989;**I**:631-4.

237. Winship PR. An improved method for directly sequencing PCR amplified material using dimethyl sulphoxide. *Nucleic Acids Res* 1989;**17**:1266.

238. Winter E, Yamamoto F, Almoguera C, Perucho M. *Proc Natl Acad Sci USA* 1985;**82**:7575-9.

239. Wolf H, Modrow S, Motz M, Jameson BA, Hermann G, Fortsch B. An integrated family of amino acid sequence analysis programs. *Comput Appl Biosci* 1988;**4**:187-91.

240. Wong C, Dowling CE, Saiki RK, Higuchui RG, Erlich HA, Kazazian Jr HH. Characterization of β-thalassemia mutations using direct genomic sequencing of amplified single copy DNA. *Nature* 1987;**330**:384-6.

241. Wood WI, Gitschier J, Lasky LA, Lawn RM. Base composition-independent hybridization in tetramethylammonium chloride: a method for oligonucleotide screening of highly complex gene libraries. *Proc Natl Acad Sci USA* 1985;**82**:1585-8.

242. Wrischnik LA, Higuchi RG, Stoneking M, Erlich HA, Arnheim N, Wilson AC. Length mutations in human mitochondrial DNA: direct sequencing of enzymatically amplified DNA. *Nucleic Acid Res* 1987;**15**:529-42.

243. Yang TP, Patel PI, Chinault AC, Stout JT, Jackson LG, Hildebrand BM, Caskey CT. Molecular evidence for new mutation at the *hprt* locus in Lesch-Nyhan patients. *Nature* 1984;**310**:412-4.

244. Youssoufian H, Kazazian HH Jr, Phillips DG, Aronis S, Tsiftism G, Brown VA, Antonarakis SE. Recurrent mutations in haemophilia A give evidence for CpG hotspots. *Nature* 1986;**324**:380-2.

245. Youssoufian H, Antonarakis SE, Aronis S, Tsiftis G, Philips DG, Kazazian HH Jr. Characterisation of five partial deletions of the factor five factor VIII gene. *Proc Natl Acad Sci USA* 1987;**84**:3772-6.

246. Youssoufian H, Antonorakis SE, Bell W, Griffin AM, Kazazian HH Jr. Nonsense and missenes mutations in hemophilia A: estimate of the relative mutation at CG dinucleotides. *Am J Hum Genet* 1988;**42**:718-25.

247. Zagursky RJ, Baumeister K, Lomax N, Berman ML. Rapid and easy sequencing of large linear double-stranded DNA and supercoiled plasmid DNA. *Gene Anal Techn* 1985;2:89-94.

7. APPENDICES.

APPENDIX A: Regional assignment of PBG-D gene to chromosome 11. From the work of Meisner et al. (1980, 1981), Wang et al. (1981) and de Verneuil et al. (1982), PBG-D gene has been mapped to chromosome 11q23-qter.

APPENDIX B: The sequence PBG-D cDNA showing mutations that have been characterised to date.

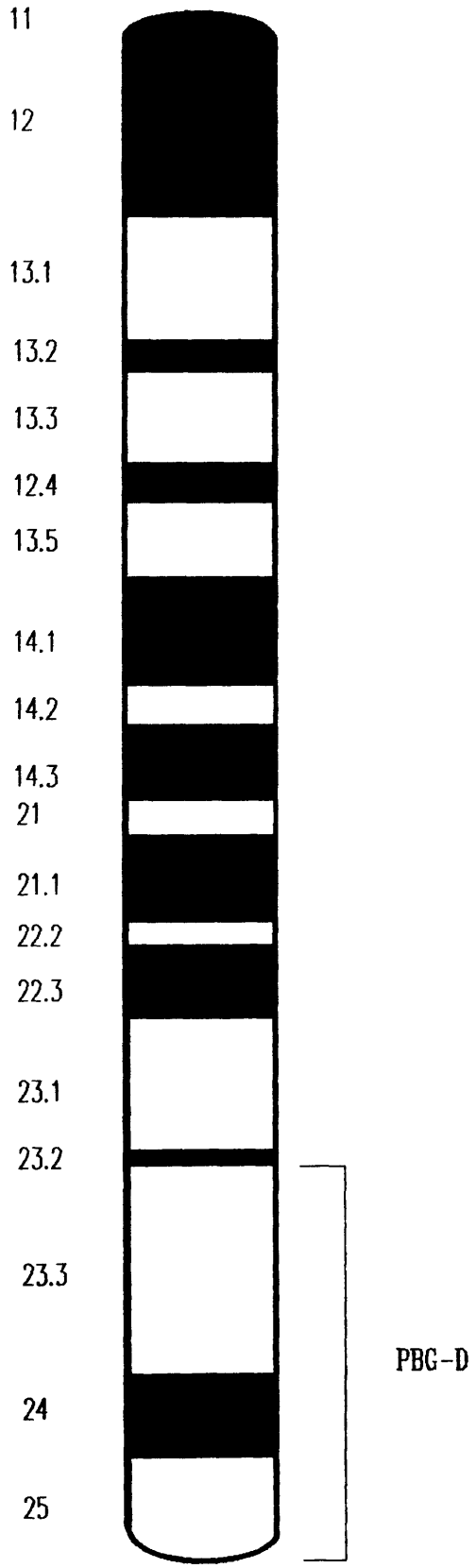
APPENDIX C: Amino acid sequence alignments of the PBG-D from various species. The open boxes denote the conserved residues. The underlined amino acids are some of those involved in mutations that have been reported in the literature, including the current study. The amino acid residue positions of the human PBG-D differ from that which has been described by Grandchamp et al. (1987), because of its alignment to other deaminases from different species. (Adapted from Prof. T. L. Blundell, University of London).

APPENDIX D: Three-dimensional structural diagram of the human non-erythroid PBG-D showing the positions of mutations detected in the current study after alignment with other deaminases. The amino acid positions differ to that which has been originally described by Grandchamp et al (1987). Thus Gln 19 is referred to in the text as Gln 34, Arg 149 as Arg 167, Leu 159 as Leu 177 and His 237 as His 256. (Drawn by Prof. T.L. Blundell).

APPENDIX E: Summarised protocol for PCR amplification and direct sequencing of the PBG-D cDNA.

APPENDIX A

POSITION OF THE PBG-D GENE ON CHROMOSOME 11.



APPENDIX B

SEQUENCE OF THE PORPHOBILINOGEN DEAMINASE cDNA.

5'----------tca agactgtagg acgacctcgg

gtcccacgtg tccccggtac tcgccggccg gagcctccgg cttcccgggg

t

ccgggggacc ttagcggcac ccacacacag cctactttcc aagcggagcc

10 20 30 E1/E2 40 F1-> 50

ATGTCTGGTA ACGGCAATGC GGCTGCAACG GCGGAAGAAA ACAGCCCAAA

F2-> 60 70 80 TA E2/E3 100

GATGAGAGTG ATTCGCGTGG GTACCCGCAA GAGCCAGCTT GCTCGCATAC

A

110 F3-> 120 130 140 A

AGACGGACAG TGTGGTGGCA ACATTGAAAG CCTCGTACCC TGGCCTGCAG

E4/E5 A <-R6 180 A 190 200

TTTGAAATCA TTGCTATGTC CACCACAGGG GACAAGATTC TTGATACTGC

E5/E6 220 230 240 250

ACTCTCTAAG ATTGGAGAGA AAAGCCTGTT TACCAAGGAG CTTGAACATG

260 E6/E7 280 290 300

CCCTGGAGAA GAATGAAGTG GACCTGGTTG TTCACTCCTT GAAGGACCTG

F4-> 310 320 330 340 E7/E8

CCCACTGTGC TTCCTCCTGG CTTCACCATC GGAGCCATCT GCAAGCGGGA

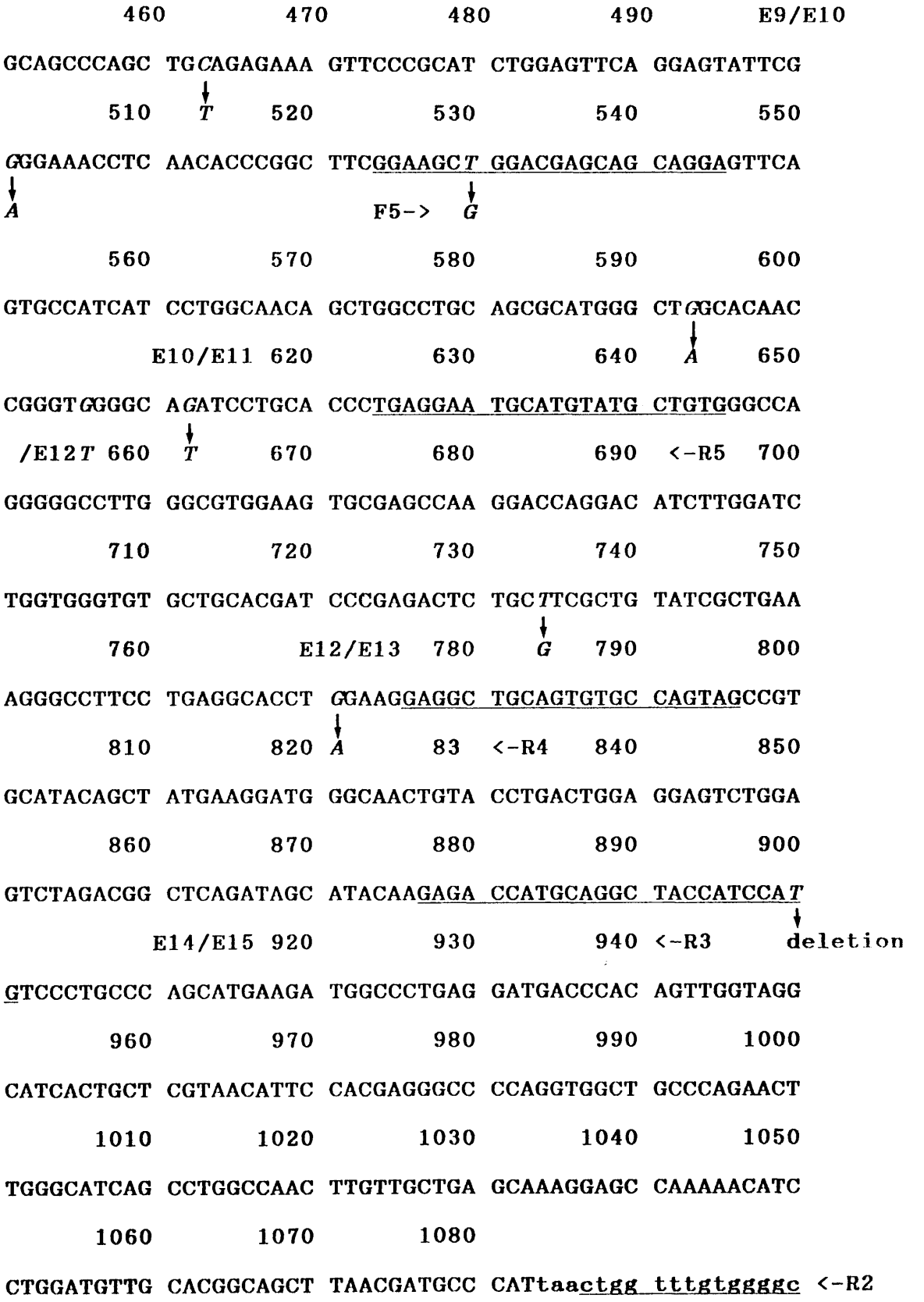
T

360 370 380 390 T 400

AAACCCTCAT GATGCTGTTG TCTTTCACCC AAAATTTGTT GGGAAGACCC

410 420 E8/E9 430 440 450

TAGAAACCCT GCCAGAGAAG AGTGTGGTGG GAACCAGCTC CCTGCGAAGA


460 470 480 490 E9/E10
GCAGCCCAGC TGCAGAGAAA GTTCCCGCAT CTGGAGTTCA GGAGTATTCG
T
510 520 530 540 550
GGGAAACCTC AACACCCGGC TTCGGAAGCT GGACGAGCAG CAGGAGTTCA
A
F5-> G
560 570 580 590 600
GTGCCATCAT CCTGGCAACA GCTGGCCTGC AGCGCATGGG CTGGCACAAC
E10/E11 620 630 640 A 650
CGGGTGGGGC AGATCCTGCA CCCTGAGGAA TGCATGTATG CTGTGGGCCA
/E12T 660 T 670 680 690 <-R5 700
GGGGGCCTTG GGCGTGGAAG TGCGAGCCAA GGACCAGGAC ATCTTGGATC
710 720 730 740 750
TGGTGGGTGT GCTGCACGAT CCCGAGACTC TGCTTCGCTG TATCGCTGAA
G
760 E12/E13 780 790 800
AGGGCCTTCC TGAGGCACCT GGAAGGAGGC TGCAGTGTGC CAGTAGCCGT
A
810 820 83 <-R4 840 850
GCATACAGCT ATGAAGGATG GGCAACTGTA CCTGACTGGA GGAGTCTGGA
860 870 880 890 900
GTCTAGACGG CTCAGATAGC ATACAAGAGA CCATGCAGGC TACCATCCAT
deletion
E14/E15 920 930 940 <-R3
GTCCCTGCCC AGCATGAAGA TGGCCCTGAG GATGACCCAC AGTTGGTAGG
960 970 980 990 1000
CATCACTGCT CGTAACATTC CACGAGGGCC CCAGGTGGCT GCCCAGAACT
1010 1020 1030 1040 1050
TGGGCATCAG CCTGGCCAAC TTGTTGCTGA GCAAAGGAGC CAAAAACATC
1060 1070 1080
CTGGATGTTG CACGGCAGCT TAACGATGCC CATtaactgg tttgtggggc <-R2

acagatgcct gggttgcctg tgtccagtgc ctacatcccg ggcctcagtg

<-R1

ccccattctc actgctatct ggggagtgat taccccggga gactgaactg

cagggttcaa gccttccagg gatttgcctc accttggggc cttgatgact

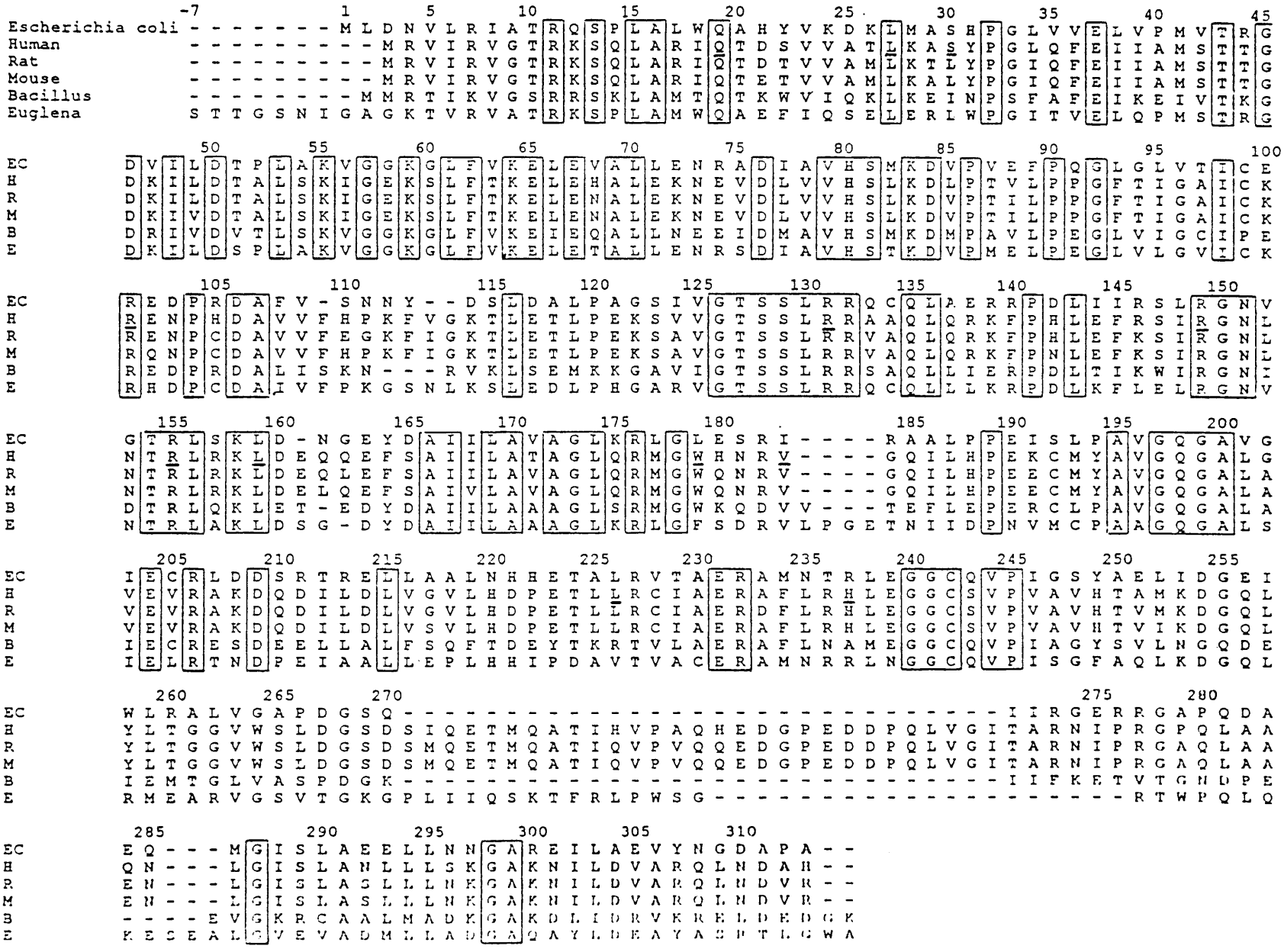
gccttgcctc ctcagtatgt gggggcttca tctctttaga gaagtccaag

caacagcctt tgaatgtaac caatcctact aataaaccag ttctgaaggt----3'

the lower case letters denote the untranslated sequences and the upper case letters the open reading frame of the PBG-D gene. The letters in *italics* are the known mutations with arrows pointing to the nature of the change. The underlined letters are the sequences of the oligonucleotide used in PCR amplifications and sequencing in the case of forward (F) primers and complementary sequences in the case of reverse (R) primers. E stands for exon, with the slash symbol / indicating the exact location of the splicing sites.

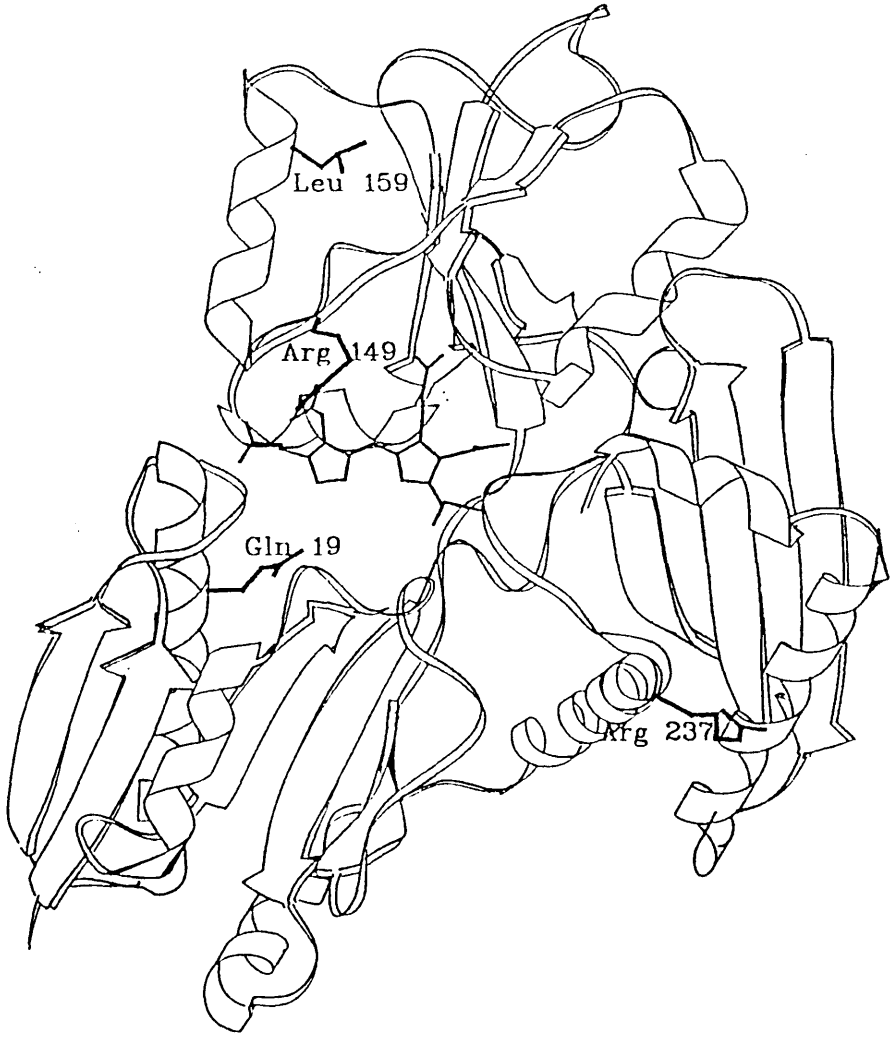
APPENDIX C.

Porphobilinogen deaminase (EC 4.3.1.8) - sequence alignments



APPENDIX D.

THREE-DIMENSIONAL STRUCTURE OF PORPHOBILINOGEN DEAMINASE.



APPENDIX E.

PROTOCOL FOR RAPID PCR AMPLIFICATION AND DIRECT SEQUENCING OF PBG-D cDNA.

1. To a cell pellet containing approximately 4 X 10^6 lymphocytes, add 50 ul of 0.1% DEPC in water and boil in a water bath at 100^0C for 5 min to lyse the cells. Remove the cell debris by centrifugation and transfer 50 ul of the cell lysate to a 500 ul microfuge tube.
2. Prepare PCR amplification mixture as follows:

Cell lysate	50.0 ul
dH_2O	2.5 ul
10X PCR buffer	10.0 ul
4 mixed dNTPs (1.25 mM each dNTP)	16.0 ul
Oligo $(dT)_{12-18}$ (100 ug/ml)	10.0 ul
Forward primer (10 uM)	5.0 ul
Reverse primer (10 uM)	5.0 ul

Vortex thoroughly and add:

MMLV-reverse transcriptase(200 u)	1.0 ul

Vortex and spin briefly, then incubate at 42^0C for 30 min. At the end of incubation add:

Taq DNA Polymerase (2.5 units)	0.5 ul

Mix gently, overlay with approximately 100 ul of light mineral oil and spin briefly. Perform a 30 cycle PCR amplification as follows:

Denature at 95^{0}C for 3 min in the first cycle and thereafter for 1 min in the subsequent cycles. Anneal at 60^{0}C for 1 min and perform the extension reaction at 72°C for 1.5 min with the last extension lasting 7 min.

3. At the end of PCR amplification, remove the reaction mix below the mineral oil and resolve in 1.4% agarose. Excise the appropriate cDNA band and elute for 4 h in 100 ul TE buffer. The eluted cDNA can be stored for future use at –20^{0}C.

4. Prepare a second PCR mix with unequal concentrations of primers (asymmetric PCR) as follows:

dH_2O	70.5 ul
10X PCR buffer	10.0 ul
4 mixed dNTPs (1.25 mM each dNTP)	4.0 ul
Forward primer (10.0 uM)	5.0 ul
Reverse primer (0.1 uM)	5.0 ul

Also prepare a second set of reaction mix with a reversed ratio of the primers. Treat with U.V. light for 10 min and add:

cDNA	5.0 ul

Vortex mix thoroughly and add:

Taq DNA polymerase (2.5 units)	5.0 ul

Mix gently, overlay with 100 ul of mineral oil and spin briefly. Perform 41 cycles of PCR using the same thermal parameters as in the first PCR.

5. At the end of the second PCR remove 5 ul of the reaction mix and resolve in 1.4% agarose. If sufficient amount of ssDNA is noted, remove the rest of the mixture below the oil and purify.

6. To the remaining 95 ul of the PCR product, add 95 ul 4 M ammonium acetate and 190 ul propan-2-ol. Mix thoroughly and store for 10 min at room temperature. Recover the cDNA pellet by centrifuging at 12,000 *g* for 10 min.

7. Wash the cDNA pellet once with 70% ethanol, dry and resuspend in 10 ul of TE buffer.

8. Perform sequencing reactions using either Sequenase™ or *Taq* DNA polymerase from sequencing kits (USB) following the manufacturer's instructions.