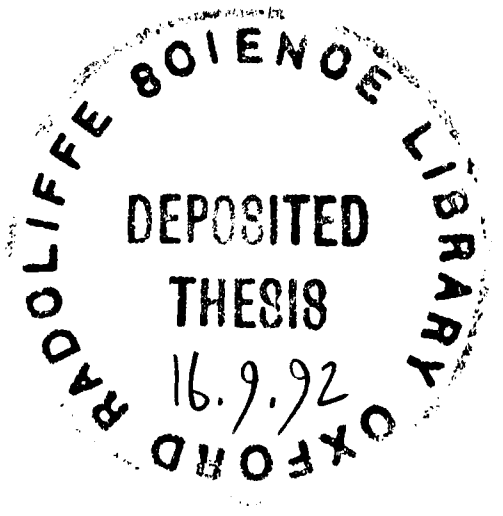
MOLECULAR GENETICS OF BETA THALASSAEMIA IN ASIAN INDIANS: BASIS FOR PRENATAL DIAGNOSIS

NERMEEN Y. VARAWALLA

Rhodes Visiting Research Fellow (1989-92), Somerville College.

A thesis submitted for the degree of Doctor of Philosophy at the University of Oxford.

Institute of Molecular Medicine, Oxford. Hilary, 1992.



ABSTRACT

Molecular genetics of beta thalassaemia in Asian Indians: basis for prenatal diagnosis

Nermeen Y Varawalla, Somerville College, Oxford

Submitted for degree of Doctor of Philosophy, Hilary, 1992

The primary aim of this thesis was to outline an approach for the prenatal diagnosis of ß-thalassaemia in the Asian Indian population by DNA analysis. A polymerase chain reaction (PCR) based, nonradioactive and rapid technique, allele specific PCR, was successfully developed for the detection of β-thalassaemia mutations. A large sample of 656 unrelated carriers from seven different regions of the Indian subcontinent was studied by allele specific PCR and DNA sequence analysis. Sixteen different β-thalassaemia mutations were identified, two of which were new mutations. Of these five common mutations accounted for 91.7% of β-thalassaemia alleles.

The β-globin gene haplotypes of 419 β-Th and 196 β-A chromosomes were constructed. On analysis of which it was inferred that β–thalassaemia mutations occurred relatively recently on existing chromosomal backgrounds and then they experienced positive selection. A strong but not invariant haplotype-mutation linkage was observed. A regional variation in the distribution of β-thalassaemia mutations was found.

α-Globin gene mapping studies identifed the single α-globin gene deletion in 24 out of 51 unrelated Asian Indians who were suspected to have α-thalassaemia. It is likely that the remaining carriers have nondeletional α-thalassaemia determinants.

To perform preimplantation diagnosis of ß-thalassaemia, by analysis of a 10-30 cell embryonic biopsy, a PCR protocol was developed. Using two rounds of PCR with nested primers, successful amplification of a 597 bp fragment of the ß-globin gene was achieved from as few as two embryonic cells. The problem of false positive amplification was encountered which appeared to be resolved by UV transillumination of the pre-amplification PCR mix. By allele specific PCR with nested primers it was possible to identify the presence or absence of five ß-thalassaemia mutations from 10 pg of template DNA (equivalent to approximately two diploid cells).

Thalassaemia control in India is a complex issue; the financial, social and demographic factors involved were considered and recommendations made.

to my country India

ACKNOWLEDGEMENTS

The studies presented in this thesis were performed at the Institute of Molecular Medicine, John Radcliffe Hospital, between April 1989 and Feburary 1992 under the supervision of Dr John Old and Professor Sir David Weatherall. I am deeply grateful to Professor Sir David Weatherall for giving an obstetrician like myself this precious opportunity to discover molecular medicine. I thank John for his technical guidance and for making available samples for analysis.

I am greatly indebted to the Rhodes Trustees for their generous financial support. The Overseas Research Student Award that I obtained was a valuable help, as on its basis the Rhodes Trustees kindly consented to fund my studies. The Peerbhoy fellowship for foreign studies awarded by the Bombay University has been of sentimental value to me.

I remain especially grateful to the Principals, Baroness Park of Monmouth (until September, 1989) and Mrs Catherine Hughes, and the Fellows, of Somerville College for their whole hearted support and valuable friendship. I remember with appreciation, the encouragement and kindness of late Professor Sir Alec Turnbull. I gratefully acknowledge the assistance of Dr PG Natrajan with sample collection in Bombay. Swee Choo Yeoh and Jennie Brown deserve a special mention for the friendly assistance they gave me, during my early days in Oxford. I wish to thank Bill Wood, Doug Higgs, Jenny Lynch, Ian Sargent and Adrian Hill for their helpful comments on this thesis. I am particularly thankful to Robin Olds for the valuable advice that he has given during the entire course of these studies and the writing of this thesis.

CONTENTS

Chapter 1	**Introduction**	**1**
1.1	Haemoglobin: structure and genetic control	1
1.2	The thalassaemia syndromes	7
1.3	The pathophysiology of thalassaemia	7
1.4	World distribution of thalassaemia	8
1.5	Molecular basis of ß-thalassaemia	10
1.6	Molecular basis of α-thalassaemia	15
1.7	Clinical and haematological findings in the thalassaemias	19
1.8	Diagnosis of the thalassaemia syndromes	22
1.9	Treatment of thalassaemia	25
1.10	Prevention of thalassaemia	27
1.11	Thalassaemia on theIndian subcontinent	34
1.12	Thesis plan	38
Chapter 2	**Materials and Methods**	**40**
2.1	Materials	40
2.2	Extraction of genomic DNA	40
2.3	Southern blot hybridisation	42
2.4	The polymerase chain reaction (PCR)	47
2.5	Oligonucleotide synthesis and purification	50
2.6	Dot-blot hybridisation	52
2.7	Direct genomic sequencing	56
Chapter 3	**Characterisation of ß-thalassaemia mutations**	**60**
3.1	Introduction	60
3.2	Subjects	62
3.3	Strategy	65
3.4	Methods	65
3.5	Results	82
3.6	Discussion and Conclusions	104

Chapter 4 Haplotype analysis **113**

4.1 Introduction 113
4.2 Subjects 115
4.3 Methods 115
4.4 Results 126
4.5 Discussion and Conclusions 136

Chapter 5 α-Globin gene mapping **146**

5.1 Introduction 146
5.2 Subjects 148
5.3 Methods 148
5.4 Results 149
5.5 Discussion and Conclusions 149

Chapter 6 An approach to preimplantation diagnosis of ß-thalassaemia **152**

6.1 Introduction 152
6.2 Experimental Development 155
6.3 Materials and Methods 159
6.4 Results 167
6.5 Discussion and Conclusions 177

Chapter 7 Outlook and Summary of Results **185**

7.1 Control of genetic disease 185
7.2 Background information for thalassaemia control in India 187
7.3 Indian thalassaemia control program: recommendations 191
7.4 Financial considerations 198
7.5 Thalassaemia control in India: summary 201
7.6 Summary of results 202

References 210

CHAPTER ONE

INTRODUCTION

The inherited disorders of haemoglobin (Hb) are the commonest monogenic disorders in humans. The World Health Organisation has estimated that at least 5% of the world's population are carriers for different haemoglobinopathies and that about 300,000 severely affected infants are born each year (WHO Working Group, 1982). The inherited disorders of Hb fall into three overlapping groups, structural variants, thalassaemias and hereditary persistence of fetal Hb (HPFH). As these different groups of haemoglobinopathies coexist at a high frequency in many populations, including the Asian Indians, individuals often inherit more than one type of disorder, resulting in widely varying clinical phenotypes.

This thesis describes a detailed study of the molecular genetics of ß-thalassaemia in the Asian Indian population. As the treatment of thalassaemia remains unsatisfactory, avoidance by prenatal diagnosis is an important option for at risk couples. For prenatal diagnosis by DNA analysis it is essential to know the molecular basis of the disease in that population. The primary aim of this thesis is to outline a strategy for the prenatal diagnosis of ß-thalassaemia in the Asian Indian population by DNA analysis.

1.1 Haemoglobin: structure and genetic control

1.1.1 Structure

Normal Hb has a tetrameric structure composed of two pairs of unlike peptide globin chains. Globins may be broadly divided into α-like, ζ and α, each consisting of 141 amino acids and ß-like, ε, γ, δ and ß, each consisting of 146 amino acids. In normal adults about 97% of the total Hb consists of Hb A ($\alpha_2ß_2$) with the remainder being Hb A_2 ($\alpha_2\delta_2$). Hb F ($\alpha_2\gamma_2$) is the main haemoglobin in fetal life, although approximately 0.5% is found in normal adults. The Hbs present in embryonic life are Hb Gower 1 ($\zeta_2\varepsilon_2$), Hb Gower 2

($\alpha_2\varepsilon_2$) and Hb Portland ($\zeta_2\gamma_2$). The amino acid sequences of all the globins have many common features and are related to each other.

1.1.2 Organisation of globin genes

The genes that encode the globin chains are arranged in two clusters, α and ß, which are located on different chromosomes. The ß-globin cluster is on the short arm of chromosome 11 distal to band p14 (Gusella et al, 1979) and the α-globin cluster has been localised to Giemsa negative band 16p13.3 at the tip of chromosome 16 (Buckle et al, 1988).

The ß-globin cluster has five functional genes, ε, $^{G}\gamma$, $^{A}\gamma$, δ and ß and one nonfunctional gene, ψß. The α-globin cluster has three expressed genes in the order ζ, α_2 and α_1 with three nonfunctional genes, $\psi\zeta_1$, $\psi\alpha_2$ and $\psi\alpha_1$ and a gene of undefined function, θ1. The nonfunctional genes, termed pseudogenes, have acquired mutations that have inactivated them and are considered to be relics of past evolutionary changes within the globin gene cluster. Within each cluster the genes are ordered in a 5'-3' direction according to their temporal expression, which is believed to have implications for globin gene regulation (figure 1.1).

1.1.3 Structure of globin genes

The globin genes from both clusters share a basic structure. Each gene is composed of three coding regions called exons and two noncoding regions, termed introns or intervening sequences (IVS). Primary protein and coding DNA sequences show a considerable degree of homology between the genes of the α and ß globin gene clusters. These observations suggest that the two gene families descended from a single ancestral gene by a series of gene duplication events followed by divergence (Efstratiadis et al, 1980, Slighthom et al, 1980).

In genes of the ß-globin gene cluster, the introns interrupt the coding sequence between codons 30 and 31 and between codons 104 and 105 (figure 1.2); in the α-globin gene family, the introns are between codons 31 and 32 and between codons 99 and 100

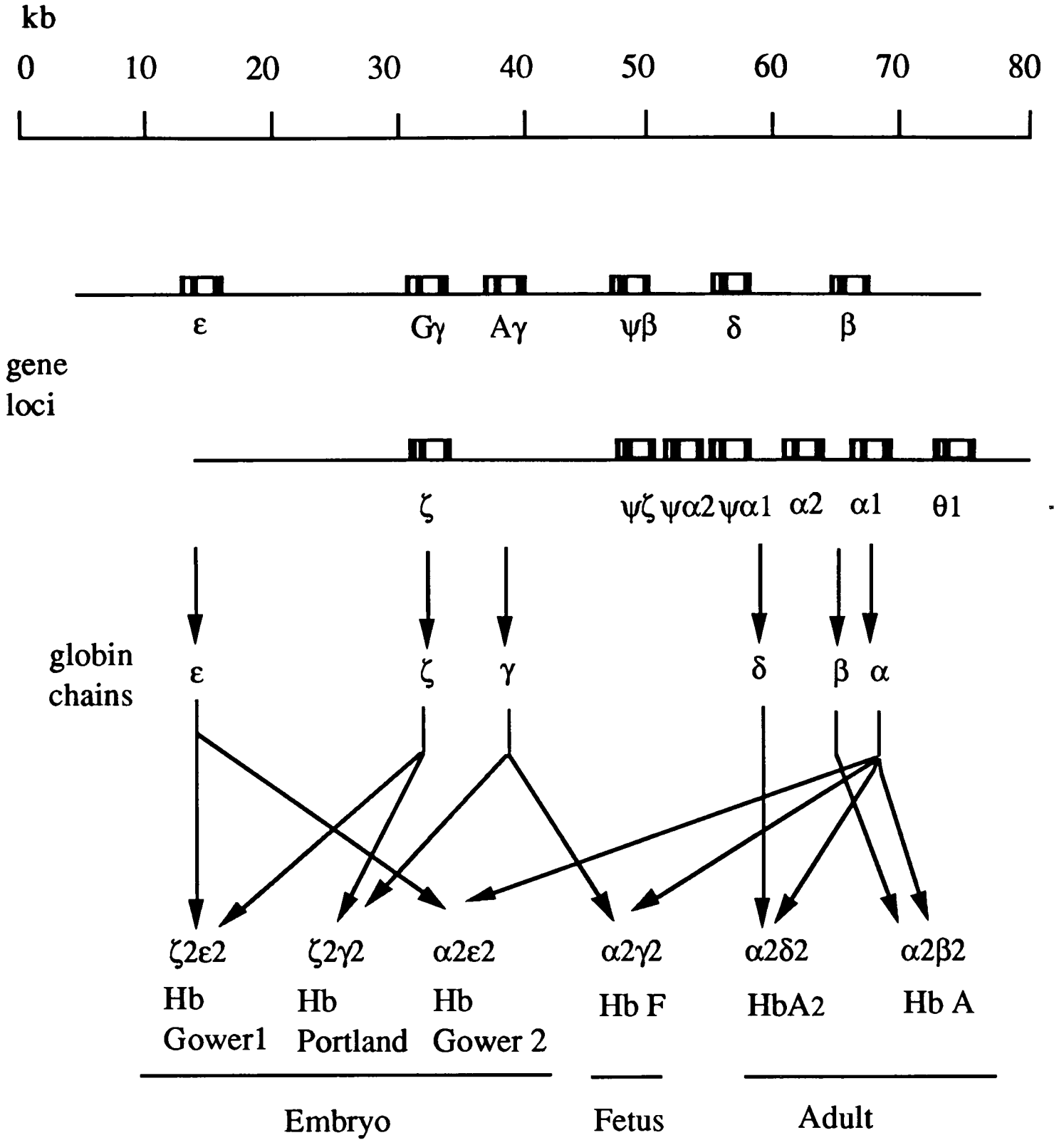


Figure 1.1 The genetic control of human Hb. The temporal expression of the globin gene loci is depicted.

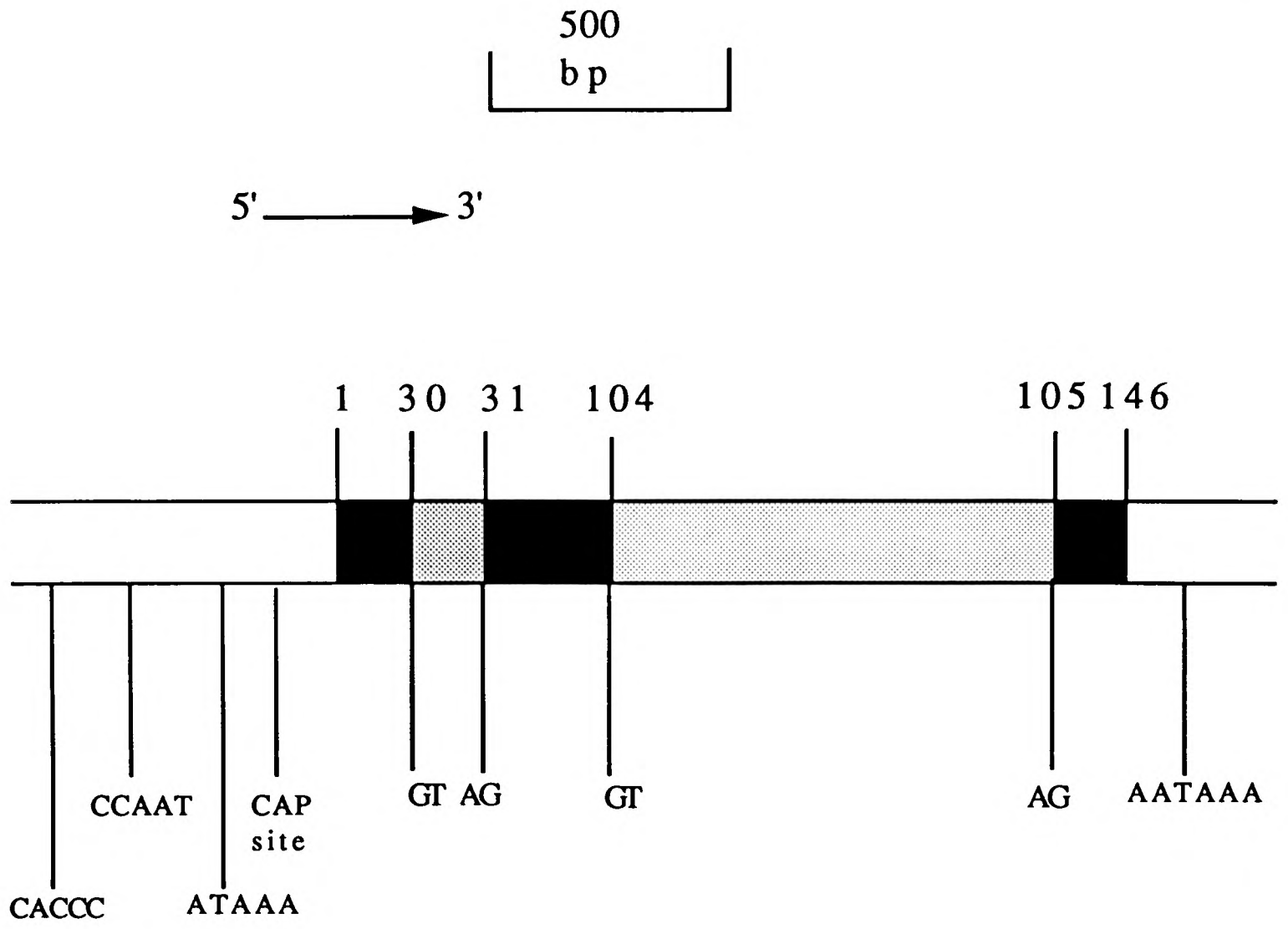


Figure 1.2 Fine structure of the human β–globin gene. The filled boxes represent the exons with the numbers above representing the codon positions. The stippled boxes represent the introns. The conserved sequences important for transcription and processing are shown below the gene.

The first intron is shorter than the second in both α and ß globin genes but the second intron of the α-globin gene is considerably shorter than that of the ß-globin gene. Sequences have been identified within the gene and in the 5' and 3' flanking regions which are conserved in globin genes of different species, as well as in other eukaryotic genes. Several of these sequences have been implicated in the control of gene function and expression. (Bunn and Forget, 1986)

1.1.4 ß-Globin gene expression

The expression of the ß-globin gene has been viewed as a model for all eukaryotic genes transcribed by RNA polymerase II. Three processes are involved, transcription, processing and translation. Transcription and processing occur within the cell nucleus and there is no clear demarcation between the end of transcription and the start of processing. Genomic DNA is transcribed, or copied, by RNA polymerase II into the complementary heterogeneous nuclear RNA (hnRNA) molecule. The 5' untranslated region of the ß-globin locus contains three sets of conserved sequences which are essential for normal gene transcription. They probably serve as binding sites or other regulatory signals for RNA polymerase II (Bunn and Forget, 1986). These conserved sequences are 1) ATAA, which constitutes the ATAA box, located 30 basepairs (bp) upstream to the cap site; 2) CCAAT, situated 80 bp upstream to the cap site and 3) CACCC, which is 90 bp upstream from the cap site (Efstratiadis et al, 1980; Collins and Weissman, 1984). Initiation of transcription occurs at the cap site, situated 50 bp upstream to the initiation codon (AUG). Transcription continues beyond the conserved AATAAA polyadenylation signal.

The large hnRNA molecule is rapidly processed to form mature messenger RNA (mRNA). The RNA transcript is chemically modified at its 5' end by the addition of a CAP structure, so called because of the formation of a 5'-5' pyrophosphate linkage which caps the 5' end of mRNA. This process probably stabilises the mRNA molecule and prevents attack by exonucleases. At the 3' end of the mRNA molecule a long string of adenylic acid residues (polyA) are attached which are believed to stabilise the molecule during its passage into the cytoplasm. The site of polyA addition is related to the highly conserved AATAAA sequence (Bunn and Forget, 1986).

The introns are removed from the precursor mRNA transcript by a two stage process called splicing. An important prerequisite for splicing is the presence of specific sequences at the exon-intron junctions. By comparison of these sequences in many different genes, two different "consensus" sequences that are universally found at the 5' and 3' ends have been derived (Mount, 1982). The

consensus sequence at the 5' donor site is (5')$^{C}/_{A}$AG**I**$\underline{GT}^{A}/_{G}$AGT and it is $(^{T}/_{C})_{n}N^{C}/_{T}\underline{AG}$**I**G(3') at the 3' acceptor site, where N represents any nucleotide, n a variable number of pyrimidine nucleotides equal to or greater than 11 and **I** indicates the site within the consensus sequence where cleavage actually occurs. The underlined dinucleotides GT and AG, located at the 5' and 3' ends of the intron, respectively, are almost invariant and thought to be essential for splicing (Breathnach et al, 1978). The precursor mRNA is first cut at the 5' splice site to generate a linear first exon and a branched lariat-type molecule containing the intron and second exon. In the second stage, the 3' splice site is cleaved, the lariat intron released, and the two exons joined (Green, 1986). Introns are removed in a sequential manner until the mature mRNA is produced.

Following processing the mRNA molecule is transported from the nucleus to the cytoplasm. Here the mRNA molecule forms the template on which amino acid residues are sequentially joined by peptide bonds to form polypeptide chains. This process is called translation. The initial step of mRNA binding to ribosomes is a complex enzymatic reaction termed initiation. Once initiated, the synthesis of a protein chain continues by the process of elongation. The initiator codon (AUG) sets the phase or reading frame of the mRNA, so that subsequent triplet nucleotide codons are read in a linear fashion and are translated into amino acids until a termination codon is encountered, which signals the end of polypeptide chain synthesis. Translation of mRNA proceeds in a 5' to 3' direction. (Lewin, 1983)

The ß-globin gene is expressed in a tissue and developmental specific manner. This is regulated primarily at the level of gene transcription by *trans* acting factors (Baron and Maniatis, 1986) that bind to a number of regulatory regions throughout the entire ß-globin gene cluster. In addition to those located in the immediate flanking regions of the gene, the entire gene cluster is regulated by a dominant control region at the 5' end of the gene cluster. Included in this region is a 20 kilobase (kb) fragment characterised by a set of tissue specific DNAase 1 hypersensitive sites (Grosveld et al, 1987).

1.2 The thalassaemia syndromes: classification

The thalassaemias are a group of recessive heterogeneous disorders characterised by disordered globin chain synthesis. The most common and clinically important forms of thalassaemia are the α, β and δβ thalassaemias. These disorders are named according to the globin chain whose synthesis is affected and can be subdivided on the basis of the quantity of globin chain synthesis. In α°, β° and $\delta\beta^{\circ}$ thalassaemia no globin chains are produced by the affected gene, while in α^{+}, β^{+} and $\delta\beta^{+}$ thalassaemia globin chains are synthesised albeit at a reduced rate (Weatherall et al, 1989). The diverse series of clinical disorders which result from the coinheritance of thalassaemia and a structural haemoglobin variant or from the coinheritance of more than one type of thalassaemia, constitute the thalassaemia syndromes.

1.3 The pathophysiology of thalassaemia

Thalassaemia typically results in reduced Hb synthesis, causing a hypochromic microcytic anaemia. The severity of the disorder is determined by the imbalanced globin chain synthesis. Absence or deficiency of the affected globin chain results in excess of the unaffected globin chain which continues to be synthesised at the normal rate, accumulating within red blood cells. The excess globin chains form aggregates and precipitate within the cell (Fessas, 1963). These precipitates cause red cell membrane damage and premature destruction of the red cells (Nathan et al, 1969). In ß-thalassaemia, the resulting α-globin chain aggregates, recognised as inclusion bodies, occur extensively in the erythroid precursor cells of the bone marrow resulting in marked ineffective erythropoiesis. In α-thalassaemia, the resulting ß-globin tetramers constitute Hb H, which is more soluble than α-globin chain aggregates. They precipitate more slowly, and so the ineffective erythropoiesis and red cell destruction is not as marked as that observed in homozygous ß-thalassaemia (Fessas and Yataganas, 1968). Hb Bart's, formed by γ-globin chain tetramers in α-thalassaemia, can be detected principally in

neonatal and fetal blood.

In ß-thalassaemia, Hb F synthesis persists at a variable level beyond infancy. Hb F is heterogeneously distributed amongst the red cells. Since the γ-globin chains combine with α-globin chains to form Hb F, cells with the most Hb F have the least excess of α-globin chains. Thus in these cells the magnitude of α-globin chain precipitation will be less, with an increase in red cell survival and more efficient erythropoiesis. The persistence of Hb F increases the oxygen affinity of red blood cells.

Red blood cells adapt poorly to thalassaemia, as reflected by inappropriately low levels of 2,3-diphosphoglycerate (2,3-DPG) and a high oxygen affinity. The profound anaemia of thalassaemia major combined with the high oxygen affinity of circulating red blood cells, results in severe tissue hypoxia. This causes erythropoietin production and expansion of the dyserythropoietic bone marrow, leading to skull deformities and porosity of the long bones. Also there is a diversion of calories for ineffective erythropoiesis causing growth retardation and wasting.

The large number of abnormal red cells in the circulation causes progressive splenomegaly and hypersplenism which exacerbates the anaemia by increasing the intravascular volume and causing haemodilution. Hypersplenic destruction of red blood cells also occurs, shortening their life span.

The ineffective erythropoiesis of thalassaemia results in increased intestinal absorption of iron with accumulation in the liver, spleen, endocrine glands, skin and myocardium. The regular blood transfusions most thalassaemia patients receive, further contributes to this iron accumulation. The clinical manifestations of iron loading include diabetes, hyperparathyroidism, hypogonadism and intractable cardiac failure (Weatherall et al, 1989). The pathophysiology of ß-thalassaemia is summarised in figure 1.3.

1.4 World distribution of thalassaemia

Thalassaemia primarily affects people of Mediterranean, African and Asian ancestry, however sporadic cases have been reported in

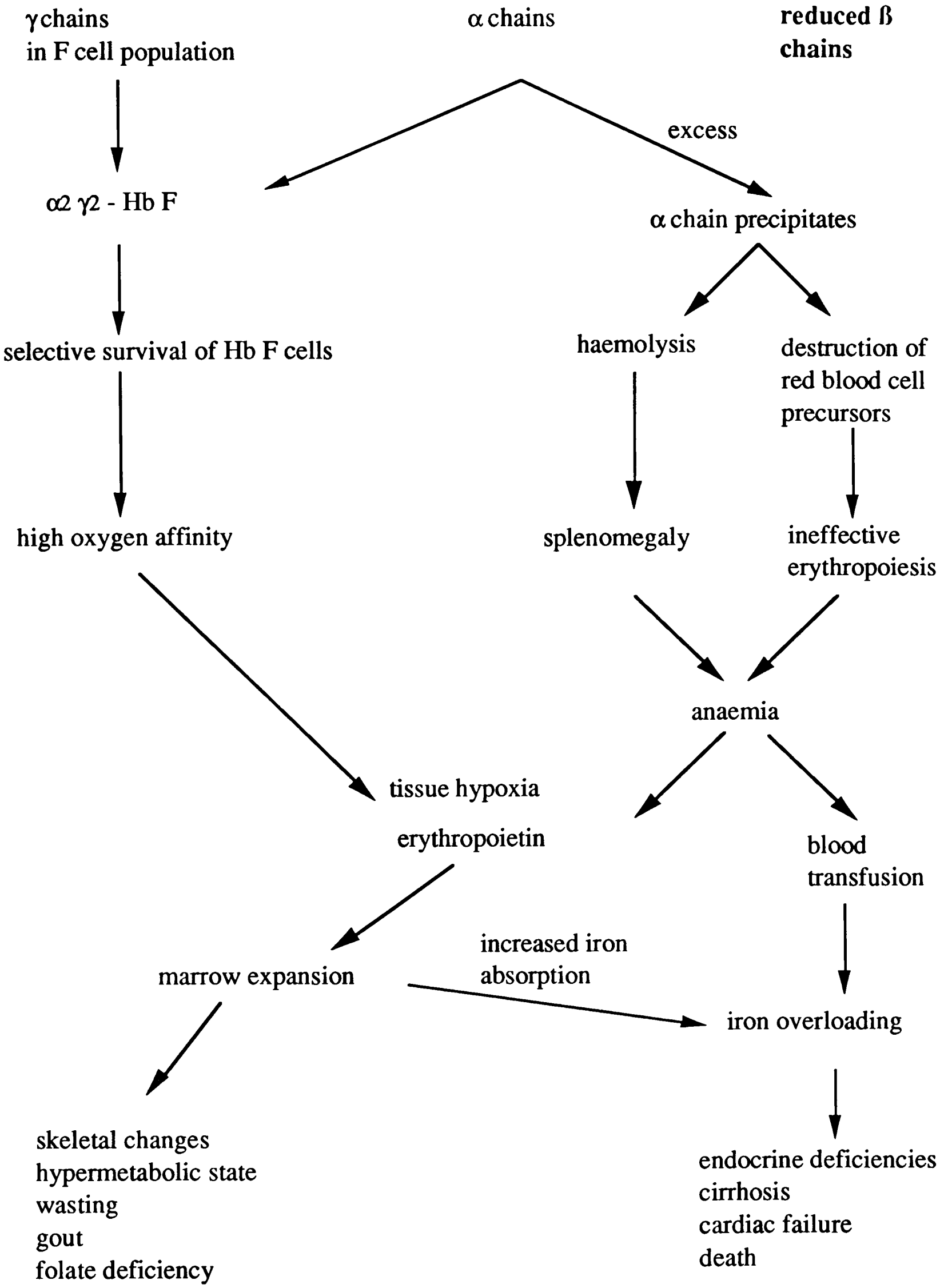


Figure 1.3 The pathophysiology of ß-thalassaemia
(adapted from Weatherall et al, 1989)

practically every racial group. It is believed that malaria has exerted selective pressure for the propagation of the thalassaemia gene. ß-Thalassaemia has a high prevalence in a broad tropical belt extending from the Mediterranean basin through the Middle East, the Indian subcontinent and Southeast Asia (Weatherall & Clegg, 1981). The prevalence of ß-thalassaemia in various population groups has been estimated. In southern Italian, Sicilian and Greek populations, approximately 10% of the population are heterozygous for ß-thalassaemia. In Southeast Asian populations, the prevalence of the ß-thalassaemia trait is about 5%, although a higher prevalence has been reported in certain regions. In African and American blacks, the prevalence of the ß-thalassaemia gene is 1.5%. (Bunn and Forget, 1986)

The α^+-thalassaemias are extremely common in parts of Africa, the Mediterranean region, the Middle East, throughout Southeast Asia and in the Pacific Islands. For example, the prevalence of α^+-thalassaemia in Negroes is 25% (Dozy et al, 1979) and about 10% in Thailand (Pootrakul et al, 1975). α°-thalassaemia is restricted to Mediterranean island populations and to Southeast Asia where it occurs frequently in China, Thailand, Laos and the Malay peninsula. (Weatherall et al, 1989)

1.5 The molecular basis of ß-thalassaemia

ß-Thalassaemia is mainly caused by point mutations and about a 100 different ß-thalassaemia mutations have been described. ß-Thalassaemia mutations may be divided into two categories, $ß^\circ$, in which no ß-globin chains are produced and $ß^+$, in which ß-globin chain synthesis is reduced. The majority of ß-thalassaemia mutations are single base substitutions situated in the coding regions of the ß-globin gene and its flanking DNA.

The spectrum of ß-thalassaemia mutations has been defined in a number of population groups in which the disease is prevalent. The populations studied include the Chinese (Kazazian et al, 1986; Chan et al, 1987), Mediterraneans (Kazazian et al, 1984a), Lebanese (Chehbab et al, 1987), Spanish (Amselem et al, 1988) and immigrant Asian Indians (Kazazian et al, 1984b; Thein et al, 1988). Each affected population was found to have its own set of

mutations, usually with a subset of a few mutations accounting for approximately 80-90% of ß-thalassaemia alleles.

1.5.1 Gene Deletions

A number of large deletions affecting regions of the ß-globin gene cluster have been described. The majority of (δß)° and (^{A}γδß)° thalassaemias, as well as different forms of (δß)°-HPFH are the result of these large deletions. These disorders are of considerable interest, as an understanding of the mechanisms by which these deletions cause the γ-globin genes to remain active in adult life may provide insights into the developmental regulation of globin genes. The (δß)°-thalassaemia deletions remove or inactivate only the δ and ß-globin genes, while those of (^{A}γδß)°-thalassaemia extend into or beyond the ^{A}γ-globin gene on the 5' side and also remove the δ and ß-globin genes. Deletions which cause (εγδß)°-thalassaemia begin about 55 kb upstream from the ε-globin gene and end within the ß-globin gene cluster. In the (εγδß)°-thalassaemia deletions which spare the ß-globin gene, no ß-globin chain production occurs.

Relatively few deletions affecting only the ß-globin gene have been described. Of these, the 619 bp deletion at the 3' end of the ß-globin gene which causes ß°-thalassaemia is present in Asian Indians (Orkin et al, 1979). Other rare deletions remove the 5' end of the ß-globin gene, leaving the δ-globin gene intact and result in a phenotype of unusually high Hb A_2 ß-thalassaemia (Padanilam et al, 1984; Popovich et al, 1986).

1.5.2 Transcriptional mutations

ß-Thalassaemia mutations that affect transcription are located either in the ATAA box, 30 bp upstream from the cap site, or in the proximal and distal CACACCC sequences situated 90 and 105 bp upstream from cap site of the ß-globin gene. Mutations of the ATAA residues have been described at positions -31, -30, -29 and -28; they cause reduced transcription and are generally associated with a mild clinical phenotype. However, ethnic variation in phenotype has been observed, probably due to the coexistence of

genetic determinants that increase Hb F production. For example, black homozygotes for the -29 A-G mutation have a very mild phenotype (Antonarakis et al, 1984), while a Chinese homozygote for the same mutation had transfusion dependent thalassaemia major (Huang et al, 1986).

Relatively mild mutations have been described in the proximal CACACCC sequence, located at positions -92, -88, -87 and -86. A "silent carrier" mutation has been found in the distal "CACACCC" sequence at position -101 (Gonzalez-Redondo et al, 1989).

1.5.3 RNA processing mutations

1.5.3.1 Cap site mutant

The cap site is where transcription begins and where the chemical modification of the 5' end of the mRNA transcript occurs. An A-C mutation at this site produces very mild symptoms due to its effect on either transcription or translation (Wong et al, 1987). The combination of this mild mutation with a severe ß-thalassaemia mutation produces clinically severe disease.

1.5.3.2 RNA Cleavage and Polyadenylation Mutants

Mutations of the polyadenylation signal, AATAAA, affect the efficiency of RNA cleavage at this site (Orkin et al, 1985). As a result only a small percentage of RNA transcripts are cleaved at the appropriate site, while the majority remain uncleaved. Transcription proceeds beyond the polyadenylation signal resulting in unstable, elongated transcripts and decreased ß-globin chain synthesis. These mutations cause $ß^{+}$-thalassaemia as some normal transcripts are produced and it is likely that normal ß-globin chains are synthesised from any elongated transcripts present *in vivo*. Some examples of polyadenylation mutants are AATAAA-AACAAA (Orkin et al, 1985) and AATAAA-AATGAA (Jankovic et al, 1989).

1.5.3.3 Mutations affecting RNA splicing

The complete absence of normal splicing of mRNA causes ß°-thalassaemia. When some normal splicing occurs, the clinical phenotype is less severe, producing ß$^+$-thalassaemia. Mutations at the exon-intron boundary affecting the invariant dinucleotides (GT/AG) which are essential for splicing, lead to ß°-thalassaemia. Mutations that occur in the remainder of the consensus splice site sequences lead to a reduction rather than a complete absence of splicing. Consensus changes occur both in the donor and acceptor sites producing diverse clinical phenotypes. For example, the IVS-1 nucleotide 6 (T-C) mutation produces almost normal levels of correctly spliced ß-globin mRNA, consistent with the associated mild clinical phenotype (Tamagnini et al, 1983), while the IVS-1 nucleotide 5 (G-C) mutation reduces the functional RNA output by about half, causing a more severe phenotype (Cheng et al, 1984). Two other mutations are found at IVS-1 nucleotide 5; the G-T substitution is associated with a severe phenotype (Atweh et al, 1985), while the G-A substitution causes a very mild disease (Lapoumeroulie et al, 1986). It remains unclear as to why different base substitutions at the same site, lead to different levels of splicing and such a variation in clinical phenotype.

Nucleotide substitutions can create a consensus splice site sequence within either an intron or an exon by activation of a cryptic splice site. One of these cryptic splice sites is present in codons 24 to 27 of exon 1 and contains a GT dinucleotide. A substitution in this region could result in a closer resemblance to the normal splice site causing the preferential use of the cryptic splice site even though the normal one was intact (Orkin and Kazazian, 1984). The codon 24 (GGT-GGA) mutation does not alter the amino acid sequence, but activates the cryptic splice site to cause moderately severe ß$^+$-thalassaemia (Goldsmith et al, 1983). Mutations at codons 26 and 27 produce two haemoglobin variants, Hb E (Glu-Lys) and Hb Knossos (Ala-Ser), respectively. Unlike other haemoglobin variants, which generally constitute 40-50% of the total haemoglobin in heterozygotes, both of these variants are present in reduced quantities because of the decreased efficiency of splicing (Orkin et al, 1982a; Orkin et al, 1984).

Mutations also activate cryptic splice sites within introns, for example, the IVS-1 nucleotide 110 (G-A) mutation (Spritz et al,

1981). This region of IVS-1 has a sequence similar to that of the 3' acceptor consensus sequence, but lacks the invariant AG dinucleotide. The G-A mutation at IVS-1 nucleotide 110 creates this AG dinucleotide, and about 90% of RNA transcripts splice at this site and less than 10% splice at the normal site, resulting in severe ß$^+$-thalassaemia. Of the other mutations that activate cryptic splice sites in introns, IVS-1 nucleotide 116 (T-G) and IVS-2 nucleotide 654 (C-T) result in ß°-thalassaemia (Metherall et al, 1986; Cheng et al, 1984).

1.5.4 Nonfunctional mRNA mutations

Nonfunctional mRNA mutations include nonsense mutants, frameshift mutants and initiator codon mutants. Nonsense mutations alter the coding signal for a particular amino acid to a termination signal resulting in premature termination of translation and ß°-thalassaemia. Codon 17 (AAG-TAG) was the first nonsense mutation to be described (Chang & Kan, 1979). Frameshift mutations are due to the insertion or deletion of nucleotides in the coding sequence that disrupt the normal reading frame. On translation of the resulting mRNA, anomalous amino acids are added until a termination codon is reached in the new reading frame. The abnormal RNA is found in very low levels in erythroid cells resulting in ß°-thalassaemia. An example of a frameshift ß-thalassaemia mutation is the insertion of a nucleotide at codons 8 and 9 which results in a premature termination signal at codon 22 (Kazazian et al, 1984b). Substitutions in the initiatation codon have been described which cause ß°-thalassaemia, an example is the ATG-ACG mutation (Jankovic et al, 1989).

1.5.5 Dominant ß-thalassaemia

Typically heterozygotes for ß-thalassaemia are asymptomatic with minimal haematological abnormalities. However there have been a number of reports of ß-thalassaemia heterozygotes presenting with dominant, dyserythropoietic anaemia associated with inclusion bodies in red cell precursors (Kazazian et al, 1989;

Thein et al, 1990). These cases of dominant ß-thalassaemia have been found to be caused by mutations in exon 3 of the ß-globin gene. The dominant ß-thalassaemia mutations cause unstable elongated ß-globin chains that produce aggregations, called inclusion bodies, which are resistant to intra-cellular proteolytic digestion. Dominant ß-thalassaemia alleles are rare and each mutation has been found in single families from widely dispersed ethnic groups. Presumably because of their severity they have not experienced positive selection pressure due to malaria, hence unlike recessive ß-thalassaemia have not attained high gene frequencies in parts of the world where malaria is endemic.

1.5.6 Silent ß-thalassaemia

Silent ß-thalassaemia is heterozygous ß-thalassaemia with a normal Hb A_2 level and normal red cell indices. The diagnosis of this condition is made by *in vitro* globin chain synthesis studies which reveal a chain imbalance. One of the few documented silent ß-thalassaemia mutations is a C-T substitution at position -101, in the distal CACCC box, and has been shown by *in vitro* expression studies to decrease effective translation (Gonzalez-Redondo et al, 1989).

1.6 Molecular basis of α-thalassaemia

Unlike ß-thalassaemia, the majority of α-thalassaemia determinants are deletions of one of the α-globin genes. The α-globin gene haplotype of an individual with unaffected α_2 and α_1 genes on both chromosomes is written as αα/αα. A single α-globin gene deletion is represented as -α and that of both α-globin genes *in cis* as --. A superscript is used to indicate the size of the deletion, for example, $-\alpha^{3.7}$ represents a 3.7 kb deletion involving one α-globin gene. When the size of the deletion has not been established, the superscript describes the geographical or individual origin of the gene. For the nondeletional α-thalassaemia determinants the nomenclature used is $\alpha\alpha^{T}$. When the precise molecular defect is known, for example Hb Constant Spring, $\alpha\alpha^{T}$ is replaced by the more informative $\alpha\alpha^{CS}$.

1.6.1 Deletional forms of α-thalassaemia

The most common of α-thalassaemia deletions are $-\alpha^{3.7}$ and $-\alpha^{4.2}$ which cause α^{+}-thalassaemia. The α-globin genes are embedded within two highly homologous 4 kb duplication units which are divided into homologous subsegments called X, Y and Z boxes by nonhomologous elements (I, II and III) (Higgs et al, 1989) (figure 1.4). The duplicated Z boxes are 3.7 kb apart and the X boxes are 4.2 kb apart. Misalignment and unequal crossing over between these homologous segments at meiosis gives rise to chromosomes with either single (Embury et al, 1980) or triplicated α-globin genes (Goossens et al, 1980). Such an occurrence between homologous Z boxes deletes 3.7 kb of DNA, termed a rightward deletion, $-\alpha^{3.7}$, whereas misalignment of two X boxes with an unequal cross over deletes 4.2 kb of DNA, termed a leftward deletion, $-\alpha^{4.2}$. The corresponding triplicated α-globin gene arrangements are $\alpha\alpha\alpha^{anti\ 3.7}$ and $\alpha\alpha\alpha^{anti\ 4.2}$ (figure 1.4). A $-\alpha^{3.5}$ α-thalassaemia deletion has been identified in an Asian Indian (Kulozik et al, 1988), but as the precise end points have not been defined it is not known whether this deletion has arisen by a homologous or an illegitimate recombination event.

Twelve large deletions of the α-globin gene cluster have been described which vary in size from 5.2 kb to approximately 80 kb. With the exception of the $\alpha\alpha^{RA}$, all these deletions either completely or partially delete both the α-globin genes, and there is no α-chain synthesis directed by these chromosomes (Higgs et al, 1989).

1.6.2 Nondeletional α-thalassaemia

Nondeletional α-thalassaemia alleles are much more infrequent than the deletion defects. Sixteen different nondeletional α-thalassaemia determinants have been characterised (Higgs et al, 1989). All but one of the nondeletional α-thalassaemia determinants occur in the dominant α_2-globin gene. There appears to be no associated increase in the expression of the unaffected α_1-globin gene, as observed in the single α-globin gene deletions (Liebhaber et al, 1985). Thus nondeletional α-thalassaemia alleles

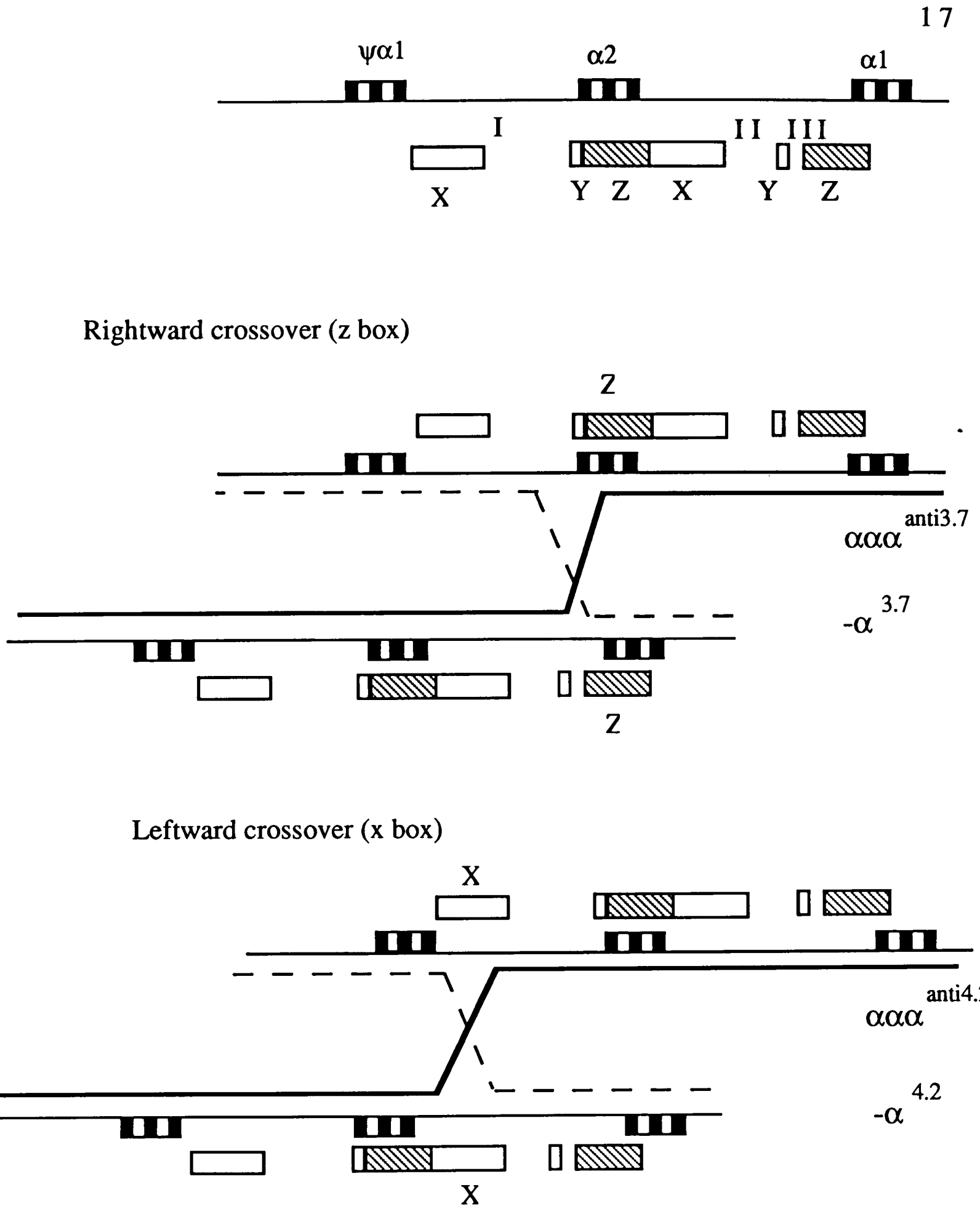


Figure 1.4 Mechanism for the common α-thalassaemia deletions. The duplicated XYZ box arrangement and then nonhomologous regions I, II and III containing the α- globin genes are shown in the upper half. Below misaligned chromosomes crossing over to produce the rightward and leftward α-globin gene deletions are shown. (adapted from: Higgs et al, 1989)

cause a more severe reduction in α-globin chain synthesis than the single α-globin gene deletions.

Nondeletional α-thalassaemia determinants may be classified according to the stage of gene expression that they affect. Two mutations affect RNA processing; one occurs in the polyadenylation site (Higgs et al, 1983), and the other is a pentanucleotide deletion which involves the donor splice site of IVS-1 (Orkin et al, 1981). Ten mutants affect mRNA translation. There are four mutations which disrupt the translation initiation sequences. Of these, three mutations occur in the initiation ATG codon (Pirastu et al, 1984; Olivieri et al, 1987; Moi et al, 1987) and the fourth mutation deletes 2 bp of the sequence preceding the initiation codon which results in a 30-50% reduction of mRNA translation (Morle et al, 1985; Morle et al, 1986). A further 2 nondeletional mutations result in premature termination of translation. The first of these is a base substitution in codon 116 (GAG-TAG) resulting in a termination codon (Liebhaber et al, 1987) and the second is frameshift mutation at codons 30/31 (-AG) (Safaya and Rieder, 1988). The final group of four translation mutations are substitutions in the termination codon (TAA) which cause mRNA translation to continue into the 3' noncoding region giving rise to elongated, 172 amino acid globin chains. Such chains appear to be stable and form variant Hb tetramers, namely, Hb Constant Spring (Clegg et al, 1971), Hb Koya Dora (De Jong et al, 1975), Hb Icaria (Clegg et al, 1974) and Hb Seal Rock (Bradley et al, 1975). However, the abnormal mRNA translation results in decreased mRNA stability and decreased α-globin chain synthesis.

Four types of nondeletional α-thalassaemia result from post-translational instability of the α-globin chains. Hb Quong Sze (Liebhaber and Kan, 1983) is an example of such an unstable α-globin variant. Soon after synthesis these globin chains are rapidly destroyed by proteolysis unless they are incorporated into a Hb tetramer, in which state they appear to be stable.

1.7 Clinical and Haematological findings in the thalassaemias

1.7.1 ß-thalassaemia

1.7.1.1 ß-thalassaemia major

ß-Thalassaemia major may result from either a homozygous state for a ß-thalassaemia mutation or from a compound heterozygous state for two different ß-thalassaemia mutations. The affected child presents with severe anaemia usually between 1-2 years of age, when the Hb F production has declined. If left untreated, the Hb level remains below 5 g/dl and the child suffers from marked growth retardation. The skin shows pallor, icterus and often brown pigmentation. Bone marrow expansion causes skeletal changes including the characteristic "thalassaemic" facies with frontal bossing and protrusion of the jaws and cheekbones. Long bones show considerable thinning and become prone to fractures. Progressive hepatosplenomegaly is a constant feature. Intercurrent infections occur frequently causing morbidity and mortality.

If regularly transfused so as to mantain Hb levels above 11 g/dl, growth and development are minimally affected. Splenectomy is often indicated to reduce transfusion requirements. By the age of 10 to 11 years, transfused patients suffer from the effects of iron overload, which manifest as hepatic, cardiac and endocrine disturbances. Unless the iron overloading is controlled by iron chelation, death results in the second or third decade, usually from cardiac failure.

The peripheral blood shows grossly abnormal red cell morphology, with marked anisocytosis, poikilocytosis and target cells, all associated with severe hypochromia. Nucleated red blood cells are present in the blood, often containing inclusion bodies. The bone marrow shows intense erythroid hyperplasia with a shift to less mature basophilic forms and inclusion bodies are seen in many normoblasts. Hb analysis in untransfused patients shows a raised level of Hb F. In patients with no ß-globin chain production Hb A is absent, while in those with reduced ß-globin

chain synthesis, a small amount of Hb A is present, usually less than 30%.

1.7.1.2 Thalassaemia Intermedia

The term thalassaemia intermedia describes ß-thalassaemia homozygotes who present later in life compared to those with thalassaemia major and can maintain their Hb at a level above 6 g/dl without transfusion. There is usually splenomegaly and the development of hypersplenism may make the patient transfusion dependent, which could be reversed by a splenectomy. With increasing age these patients tend to accumulate iron and manifest symptoms of iron overload. The main causes of the mild phenotype in thalassaemia intermedia are mild ß-thalassaemia mutations and the coinheritance of α-thalassaemia or of genetic determinants that increase Hb F production, but there are many cases which remain unexplained.

1.7.1.3 Heterozygous ß-thalassaemia

The carrier state for ß-thalassaemia is clinically asymptomatic and the diagnosis is based on a reduced mean corpuscular volume (MCV) and mean corpuscular Hb (MCH) of red blood cells along with a raised Hb A_2 level, between 3.5-6.5%. The Hb F level may be slightly increased to 1-3%. The peripheral blood film shows hypochromia and microcytosis with some anisocytosis, poikilocytosis and basophilic stippling. Although there is a globin chain imbalance there is little evidence of ineffective erythropoiesis. It appears that the proteolytic mechanisms within the red blood cell have the ability to deal with this degree of imbalance.

1.7.2 α-thalassaemia

A wide variety of α-thalassaemia haplotypes have been described with a potential for diverse interactions. However, there are three main α-thalassaemia phenotypes, namely, α-thalassaemia carrier states, Hb H disease and Hb Bart's hydrops

fetalis syndrome.

1.7.2.1 α-thalassaemia carrier states

The α-thalassaemia carrier state results from the interaction of either a normal haplotype αα with one of the α^+ or α° thalassaemia determinants or of two α^+-determinants with each other. Since the efficiency of α-globin chain synthesis varies with the type of α-thalassaemia determinant, these interactions produce a wide spectrum of phenotypes. The phenotypic variation is reflected in the degree of anaemia, MCV, MCH, α/ß-globin chain synthesis ratio and the level of Hb Bart's at birth.

Generally, chromosomes with a single gene deletion, -α, cause the mildest phenotype, the $-\alpha^{3.7}$ deletion being milder than the $-\alpha^{4.2}$ deletion. Nondeletional mutants affecting the dominant α_2-globin gene result in a greater reduction of α-globin chain synthesis and deletions involving both the α-globin genes *in cis* cause the most severe phenotype. It may be impossible to predict accurately the genotype from any given phenotype, thus, to perform accurate genetic counselling in families with α-thalassaemia, molecular studies are essential.

1.7.2.2 Haemoglobin H Disease

Hb H disease usually results from the interaction of α^+ and α° thalassaemia. It occurs in Southeast Asia and the Mediterranean basin where both these forms of α-thalassaemia exist. As the major mechanism for the anaemia is haemolysis, the predominant clinical features are a hypochromic, microcytic anaemia, with jaundice and splenomegaly. Wide ranging Hb levels have been reported with reticulocytosis and typical thalassaemic red cell indices. The Hb consists of Hb A with variable amounts of Hb H and occasionally Hb Bart's. Typical Hb H inclusion bodies are visualised on a stained peripheral blood smear.

1.7.2.3 Hb Bart's hydrops fetalis syndrome

The Hb Bart's hydrops fetalis syndrome is usually due to the interaction of two α°-thalassaemia determinants. This syndrome has been reported only in Southeast Asians and Mediterraneans. Affected infants die *in utero* usually between 30 to 40 weeks of gestation or soon after birth. The clinical picture is that of a pale, oedematous infant with cardiac failure, hepatosplenomegaly and prolonged intrauterine hypoxia. The Hb is between 3-10 g/dl, around 80% of which consists of Hb Bart's, the remainder being Hb H and Hb Portland. There is a high incidence of maternal complications, mainly toxemia and postpartum haemorrhage. (Weatherall et al, 1989)

1.8 Diagnosis of the thalassaemia syndromes

Patients homozygous for ß or δß thalassaemia or with Hb H disease usually present with symptoms of the disease whereas heterozygous carriers for α or ß thalassaemia are detected as part of a family study, a population screening program or during the investigation of a mild refractory hypochromic anaemia. Figure 1.5 represents a flow chart for the diagnosis of the thalassaemia syndromes. Routine haematological investigations and Hb electrophoresis studies will enable a diagnosis to be made in most cases. The haematological studies indicated are examination of a peripheral blood smear, measurement of the Hb level and haematocrit, and calculation of the red cell indices. There are numerous electrophoretic techniques for Hb analysis, some of the commonly used methods are cellulose acetate and starch gel electrophoresis at an alkaline pH. By these approaches separation of Hbs A, F, S, C and E can be achieved. Quantitative estimation of Hb A_2 may be done by electrophoresis on cellulose acetate or starch block, or by column chromatography. The estimation of the relative amounts of Hb F is important for the diagnosis of homozygous ß-thalassaemia, δß-thalassaemia and different types of HPFH. For this, a number of techniques are available, one of these, the acid elution technique, determines the relative amount of Hb F and its intracellular distribution (Kleihauer et al, 1957).

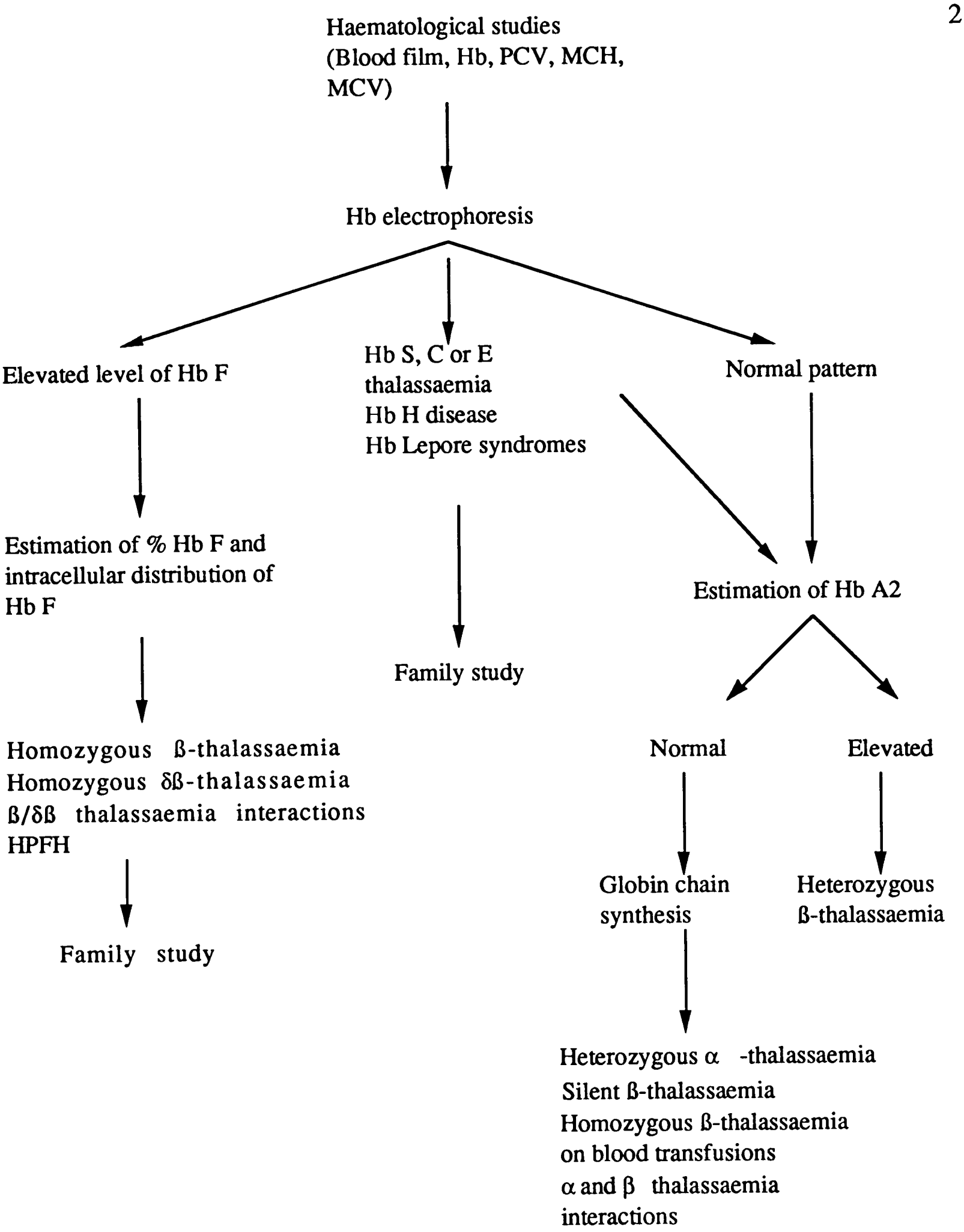


Figure 1.5 Flow diagram for the diagnosis of thalassaemia syndromes (adapted from Weatherall and Clegg, 1981)

For the characterisation of some homozygous ß-thalassaemia patients who have already received blood transfusions, some of the α-thalassaemia carrier states and some cases of interactions between α and ß thalassaemia, measurement of α/ß globin chain synthesis ratios are required. The ratio of newly synthesised globin chains can be measured in bone marrow erythroblasts, or more commonly in peripheral blood reticulocytes. For this, reticulocytes are incubated with radioactive amino acids for 1-2 hours. Globins are precipitated from the whole cell lysate and separated by column chromatography. The radioactivity incorporated into each globin chain is then measured by scintillation counting (Clegg et al, 1968).

1.8.1 Population screening

In setting up a population screening program it is important to consider the relative cost and accuracy of the available methods. The osmotic fragility screening techniques once widely used in some Mediterranean countries have now been replaced by electronic cell counters. Probably one of the simplest approaches to population screening is by estimating the MCH and MCV using electronic counters. Individuals with a MCH below 27 pg or a MCV below 75 fl should be analysed further. Their serum iron or ferritin levels should be estimated to rule out iron deficiency anaemia, though this may not be conclusive in some cases. In individuals without evidence of iron deficiency an Hb A_2 estimation should be done. Individuals with a raised Hb A_2 value are ß-thalassaemia carriers. Of the group of individuals with normal Hb A_2 values and thalassaemic red cell indices, the majority will be α-thalassaemia carriers, while some will have an elevated Hb F suggesting that they are δß-thalassaemia heterozygotes and some may have silent ß-thalassaemia. The α-thalassaemia carrier state and silent ß-thalassaemia can be confirmed by globin chain synthesis or molecular studies (Weatherall and Clegg, 1981).

In populations in which both α and ß thalassaemia are common, carriers who coinherit both disorders may have normal red cell indices, thus screening programmes in such populations should be

based on Hb A_2 estimation. If iron deficiency is common in the population, a few ß-thalassaemia heterozygotes with coexistent iron deficiency may be missed as their Hb A_2 values are reduced (Wasi et al, 1968). Thus before a survey for thalassaemia is conducted in areas where iron deficiency is widely prevalent, iron supplementation may be given for a few months prior to screening. It is also important to screen for the Hb variants known to be prevalent in the population.

Screening for α-thalassaemia is possible by the estimation of Hb Bart's in cord blood samples. α-Thalassaemia carriers with the αα/-- or -α/-α genotype would have a 5-10% Hb Bart's level in their cord blood and those with the αα/-α genotype would have a 1-2% Hb Bart's level, but some carriers from this latter group could have a normal Hb Bart's level (Weatherall and Clegg, 1981). Monoclonal antibodies against the human ζ-globin chains have also been successfully used to screen for the α-thalassaemia carrier state by a simple immunobinding assay (Luo et al, 1988).

1.9 Treatment of thalassaemia

Once the diagnosis of ß-thalassaemia major is made, the child and its family are destined for a life-long association with their physicians and hospitals. It would thus be preferable for treatment to be delivered by a specialist centre. Conventional management of ß-thalassaemia major entails regular blood transfusion, iron chelation and in some cases a splenectomy. When the disease is suspected in early infancy, care must be taken not to begin transfusion until it is evident that the patient needs it, so as to avoid the wrong categorisation of a child with thalassaemia intermedia. Once it is apparent that the child has transfusion dependent thalassaemia, the parents must be counselled in detail. To begin with, the genetic implications must be discussed. Efforts to remove any guilt feelings that the parents may have should be made and advice given about prevention in future pregnancies. The treatment available and prognosis for the affected child should be explained. It is extremely helpful to have social workers attached to clinics that look after thalassaemic children. Parent associations also play an important supportive role.

For normal growth and development of affected children, their Hb values should be maintained between 10-14 g/dl, by regular blood transfusions. Iron chelation is performed with desferrioxamine which is usually administered subcutaneously by a pump. Given complete co-operation on the part of the patients and their parents, with a regimen of regular blood transfusions and iron chelation, it may be possible to avoid growth retardation and the effects of iron loading including endocrine deficiency, liver and pancreatic damage, and heart disease (Modell et al, 1982; Weatherall et al, 1983). Thus there are many patients with thalassaemia major, in different parts of the world, now in their fourth decade of life, in good health and some even have their own children. But these patients are exceptions rather than the rule. Compliance rates for chelation therapy vary enormously, as iron chelation has complications and many children, particularly as they get older, find it increasingly difficult to comply (Hyman et al, 1990). An important aspect of the treatment of thalassaemia is general medical care, including early detection and treatment of infections, immunisation, and hormonal replacement. The conventional management of thalassaemia major is not satisfactory, even given the best facilities and clinical supervision. In developing countries where blood products, particularly those of acceptable quality, are difficult to obtain and desferrioxamine is unavailable because of its prohibitive cost, the outlook for a thalassaemia major child is still very poor.

The other therapeutic options include the use of oral iron chelators, for example, 1-2, dimethyl-3-hydroxypyrid-4-one (L1). Preliminary results of clinical trials with L1 appear promising and L1 might be a cheap and effective alternative to desferrioxamine (Agarwal et al, 1991). In patients who have an HLA-matched sibling donor, bone marrow transplantation offers a possible cure provided the procedure is carried out before the occurrence of hepatomegaly or portal fibrosis and the patient has had regular iron chelation therapy (Lucarelli, 1991). Looking into the future, the discovery of major regulatory regions flanking the globin genes (Grosveld et al, 1987; Higgs et al, 1990) and the development of apparently safe retroviral vectors may herald an approach for gene therapy of globin disorders. However, until

animal experiments demonstrate that it would be possible to achieve high level gene expression in the progeny of transfected haemopoietic stem cells, and that cells expressing inserted globin genes survive for long periods, this approach cannot be applied in a clinical situation. Moreover, nothing is known about the immunogenicity of the products of inserted globin genes or whether it will be necessary to combine gene therapy with some form of autologous marrow transplantation. Thus inserting genes to cure thalassaemia is not going to be feasible in the near future (Weatherall, 1991).

No treatment is required for α or ß thalassaemia heterozygotes. However they should be followed closely during pregnancy for the development of anaemia during the second and third trimesters. Most importantly these individuals must receive proper and timely genetic counselling. Most patients with Hb H disease maintain a reasonable Hb level, becoming anaemic only at times of stress. It is useful to administer regular folate supplements. Some patients who develop splenomegaly and hypersplenism may benefit from a splenectomy.

1.10 Prevention of thalassaemia major

Prevention of ß-thalassaemia major by genetic counselling and prenatal diagnosis is an important option for couples at risk of having an affected offspring. Amongst the α-thalassaemias, only the Hb Bart's hydrops fetalis syndrome requires prenatal diagnosis so as to protect the mother from pregnancy complications. For many years now thalassaemia prevention programmes have been actively pursued in Sardinia (Cao et al, 1990), Cyprus (Angastiniotis and Hadjiminas, 1981) and Greece (Loukopoulos et al, 1990) with notable success. Each of these programmes has educated the population, instituted widespread screening procedures to identify at risk couples and offered genetic counselling and prenatal diagnosis for all pregnancies of these couples. The incidence of live births with homozygous ß-thalassaemia in many parts of Greece and Italy is around 10% of the rate expected in the absence of such programmes (Loukopoulos et al, 1990; Cao et al, 1990).

Carrier screening has been done by targeting different groups in the population. The various groups screened have been school children, young adults, newly engaged couples, couples seeking contraception and women registering in antenatal clinics. Testing the family members of an affected child was also an important method of carrier detection. Following screening the unaffected individuals were reassured and the risks were explained to the carriers that were identified. An advantage of premarital counselling was that carriers had the additional option of selecting a non-carrier marriage partner. The partners of identified carriers were invited for investigation and if found to be normal, the couple were reassured. If the partner was affected, the couple were counselled about the choices available to them, which included abstaining from having children, adoption, artificial insemination with donor semen or oocytes or, more realistically, pregnancy with the option of prenatal diagnosis. Following the diagnosis of an affected fetus couples could choose to continue the pregnancy and have an accepted thalassaemia major child or terminate the pregnancy. By undergoing prenatal diagnosis in successive pregnancies couples could achieve a healthy family. Figure 1.6 summarises this approach for prevention of thalassaemia. One of the most important factors for the success of prevention programs has been education of the population, particularly the medical and paramedical professions.

Prevention of thalassaemia remains a particularly important problem in the developing world. In a number of developing countries, like India, the thalassaemias are prevalent enough to pose a serious national health problem. Further, the long-term and expensive treatment of affected individuals places a continual drain upon the limited resources available for health care.

1.10.1 Prenatal diagnosis of thalassaemia

There are two main approaches to the prenatal diagnosis of the thalassaemias. They are globin chain synthesis studies performed on a fetal blood sample obtained in the second trimester of pregnancy and DNA analysis performed on a chorion villus sample obtained in the first trimester or occasionally on fetal amniocytes

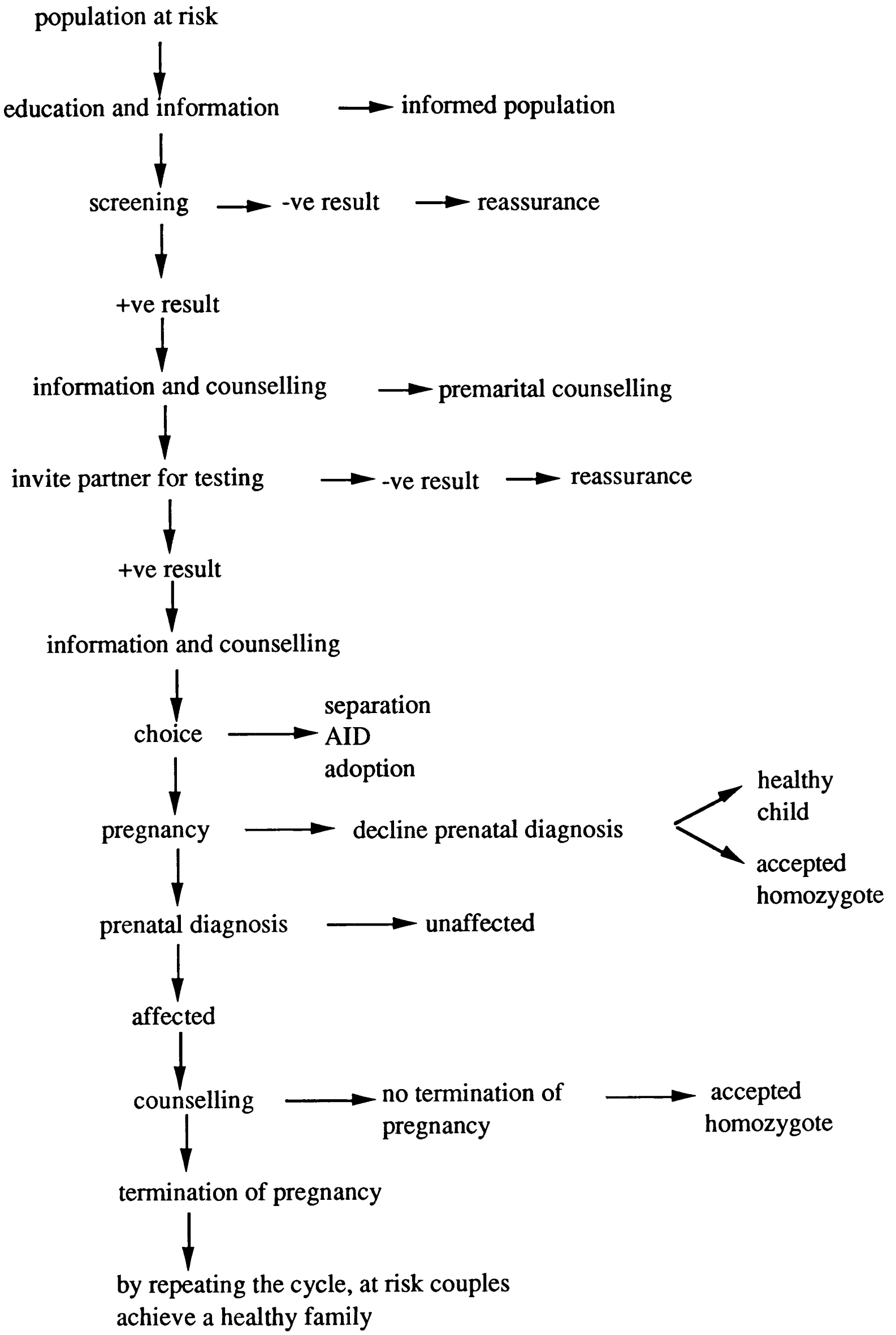


Figure 1.6 Flow diagram to show the strategy for prevention of thalassaemia (Adapted from Petrou et al, 1990)

obtained by amniocentesis in the second trimester. In 1975, the first case of prenatal diagnosis of ß-thalassaemia by globin chain synthesis studies was reported (Kan et al, 1975). In the following 15 years, about 20,000 fetuses at risk for haemoglobinopathies have had prenatal diagnosis, more than 13,000 of which have had fetal blood sampling in the second trimester, and more than 6000 were tested by DNA analysis, generally using fetal tissues obtained in the first trimester and occasionally using amniocytes obtained by second trimester amniocentesis (Alter, 1990).

1.10.1.1 Fetal blood sampling and globin chain synthesis studies

For prenatal diagnosis of thalassaemia by globin chain synthesis studies, fetal blood sampling is performed at 18-20 weeks of gestation. The fetal blood sample was initially obtained by aspiration of a placental blood vessel under direct vision, using a fetoscope. Almost all centres have now switched to cordocentesis, also known as percutaneous umbilical blood sampling (PUBS), which is ultrasound guided umbilical cord puncture. Fetoscopy carries a substantial risk to the fetus, around 5-10% (Alter, 1985), while cordocentesis is a safer procedure with an associated fetal loss rate of less than 2% (Alter, 1990). Contamination of the fetal blood sample with maternal blood was a problem with fetoscopic blood sampling, thus requiring the purity of the sample to be monitored by an electronic cell counter and particle size analyser. As fetal cells are much larger than adult red cells, contamination could be rapidly detected. Contamination is not a significant problem with cordocentesis.

Globin chain synthesis ratios are studied by isotopic labelling of fetal reticulocytes, followed by globin chain separation by carboxymethyl cellulose chromatography (CMC) (Alter et al, 1976). High pressure liquid chromatography (HPLC) can separate globin chains more efficiently and rapidly, (Alter and Stump, 1987) and has replaced CMC in a number of centres. Since cordocentesis provides a pure fetal blood, analysis by isoelectric focusing of Hb tetramers has been done, obviating the need for radioactivity and biosynthesis (Manca et al, 1986). For prenatal

diagnosis of α-thalassaemia, a quantitative or qualitative abnormality in α-globin chain synthesis is easily detectable, since the α-globin genes are fully expressed in the first trimester. The ß-globin gene is expressed at a low level in the first and second trimesters, but there is sufficient to permit detection of abnormal ß-globin chain synthesis *in utero* at this stage of gestation. From the study of a large number of cases, it was possible to define the range of normal fetal ß-globin chain synthesis, and determine the level below which ß-thalassaemia could be diagnosed (Alter et al, 1976).

Recently, prenatal diagnosis of ß-thalassaemia major by early cordocentesis has been described. This was performed at the 12th week of gestation, when 0.2-0.3 ml of fetal blood was aspirated from the site of the cord insertion and analysed by globin chain synthesis (Trapani et al, 1991). Fetal diagnosis by this approach could be useful in couples where the carrier state was discovered late in the first trimester and a DNA diagnosis was not possible. If following larger trials, prenatal diagnosis by early cordocentesis was proved to be safe and reliable it might be a suitable approach for a developing country lacking resources to establish a DNA diagnostic laboratory.

1.10.1.2 DNA analysis

There are two approaches for the prenatal diagnosis of ß-thalassaemia by DNA analysis. The first approach to be practised was linkage studies of DNA polymorphisms, which are DNA sequence variations that occur at a frequency of over 1 percent in the population. A number of DNA polymorphisms have been identified in the ß-globin gene cluster (Collins and Weissmann, 1984), which can serve as markers for ß-thalassaemia alleles. By performing a family study, it would generally be possible to establish a linkage between a polymorphism and the ß-thalassaemia gene. The inheritance of this ß-thalassaemia gene could then be tracked and the genotype of the fetus predicted. Although, prenatal diagnosis using linkage studies has been used in a number of populations with success (Little et al, 1980; Old et al, 1984), this approach is relatively tedious. Also, it requires the

study of a large number of family members and often diagnosis is not possible, particularly in the absence of an affected child.

With progress in DNA technology, many ß-thalassaemia mutations have been characterised and methods of mutation detection have been greatly simplified. Therefore, prenatal diagnosis by the second approach, direct mutation detection is more appropriate. However the feasibility of this approach would depend on the spectrum of ß-thalassaemia mutations in the population. Not surprisingly, in Sardinia, where a single mutation, codon 39 (C-T) accounts for more than 95% of the ß-thalassaemia alleles, prenatal diagnosis by mutation detection has been extremely effective (Rosatelli et al, 1987). However, laboratories that perform prenatal diagnosis for couples from a multi- ethnic community, have also adopted the mutation detection approach with success (Old et al, 1990; Kazazian, 1990).

Prenatal diagnosis of the Hb Bart's hydrops fetalis syndrome by DNA analysis involves α-globin gene mapping studies to demonstrate the deletion of all four α-globin genes. The application of DNA amplification by the polymerase chain reaction (PCR) to this analysis has greatly reduced the time and labour required (Lebo et al, 1990).

1.10.1.3 Sampling procedures for obtaining fetal tissues

When prenatal diagnosis by DNA analysis was first performed, fetal DNA was extracted from cultured amniocytes obtained by amniocentesis between 15-17 weeks of gestation. Later by chorion villus sampling (CVS) fetal tissues were obtained in the first trimester of pregnancy. CVS was introduced for first trimester prenatal diagnosis of genetic disease over 20 years ago (Hahneman and Mohr, 1968) and extensively used for fetal sexing in China (Anshan Department of Obstetrics and Gynaecology, 1975). Using this technique the developing chorion frondosum, which is representative of fetal tissue, is sampled by either the transcervical or the transabdominal route under ultrasound guidance, between 8-13 weeks of gestation. The tip of the sampling catheter is advanced into the main mass of the chorion frondosum and the villi are aspirated. The aspirated chorion villi

are examined under the dissecting microscope and contaminating maternal decidua is dissected away. It is usually possible to obtain 30-50 mg of villi for analysis.

The spontaneous pregnancy loss associated with second trimester amniocentesis is 0.5 to 1% (Tabor et al, 1986). The rate of fetal loss due to CVS has been estimated to be approximately 1% (Alter, 1990); however this rate cannot be calculated so readily as some of the pregnancies which had to be terminated following the diagnosis of a genetic abnormality might have aborted spontaneously. Further, it is not possible to control for the occurrence of a spontaneous abortion in women undergoing CVS. As a result of two large randomised clinical trials, the Canadian Randomised Trial (Canadian Collaborative CVS-Amniocentesis Clinical Trial Group, 1989) and the Medical Research Council (MRC) European Collaborative Trial (MRC Working Party on the Evaluation of Chorion Villus Sampling, 1991), there is information available on the comparative risks of CVS and amniocentesis. The European results show that a woman undergoing CVS has a 4.6% less chance of a successful pregnancy outcome than a woman undergoing second trimester amniocentesis and in the Canadian trial this difference was 1.7%. Making comparisons between these two procedures is complicated by the different time in gestation when these procedures are performed. CVS is performed in the first trimester when higher natural loss rates are expected and second trimester amniocentesis when natural loss has largely concluded. There have been a number of published reports of limb reduction defects observed in infants whose mothers had undergone CVS, following the first report from Oxford (Firth et al, 1991). The possibility of some risk of vascular disruption following CVS early in gestation cannot be ruled out, thus it is advisable to schedule the procedure after ten completed weeks of gestation.

CVS is potentially less safe than second trimester amniocentesis, but it has a major advantage in that prenatal diagnosis can be performed in the first trimester thereby avoiding the psychological and social burdens of a second trimester abortion. An alternative approach might be early pregnancy amniocentesis; high resolution sonographic imaging techniques permit 5-10 ml of

amniotic fluid to be obtained between 8-14 weeks of pregnancy with relative ease. DNA analysis by PCR of amniocytes obtained by early amniocentesis has been demonstrated (Rebello et al, 1991), however further studies are required before early amniocentesis may be used in clinical practice.

1.11 Thalassaemia on the Indian subcontinent

A varied assortment of haemoglobinopathies have been observed on the Indian subcontinent. This range of disorders probably arose from the interaction of the extreme genetic heterogeneity of Indian populations with the distinctive pattern of malarial infection that developed in the ecological setting of the Indian subcontinent (Brittenham, 1983). The thalassaemias represent the most common haemoglobinopathy on the Indian subcontinent, constituting an important public health problem. Further, thalassaemia is probably the commonest monogenic disorder in Asian Indians. Due to large Asian Indian immigrant communities in a number of other countries, thalassaemia amongst Asian Indians constitutes a significant proportion of genetic diseases in these countries.

Figure 1.7 shows the location of the places referred to in the text.

1.11.1 Prevalence of ß-thalassaemia on the Indian subcontinent

ß-Thalassaemia is widespread on the Indian subcontinent, but with an unequal distribution. The incidence of the ß-thalassaemia trait varies from 1-15% in different regions (Sukumaran, 1974), it is around 3% north of a line from Bombay to Calcutta and less than 1% below this line (Modell and Petrou, 1983). The ß-thalassaemia carrier rate is higher in some groups, for example, it is 13.6% in the Lohanas from Gujarat (Sharma et al, 1971), 10% in the Sindhi population investigated in Bombay (Sukumaran, 1974), 6% in the Sikhs from Punjab (Siddoo et al, 1956), 14.9% in the Bhanushali community from North India (Mehta et al, 1971) and 3.7% in Bengalis investigated in Calcutta (Chatterjea et al, 1957). There are few published gene frequency data for Pakistan, Bangladesh,

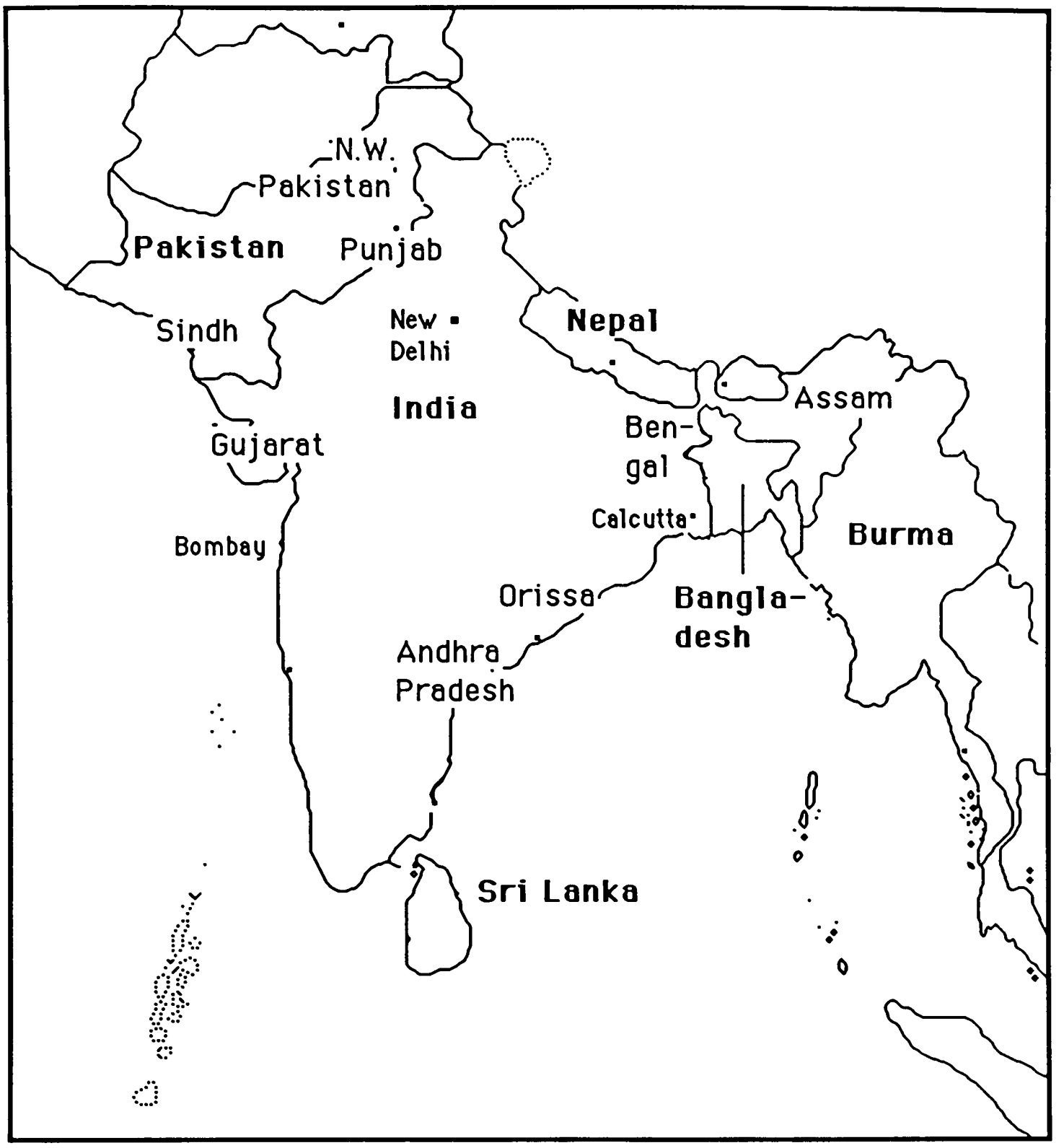


Figure 1.7 Map of the Indian subcontinent showing the countries, states and cities mentioned in the text.

Nepal and Sri Lanka but anecdotal reports suggest a high incidence, a 4% ß-thalassaemia carrier rate has been reported in the Pathans from Northwest Pakistan (Stern et al, 1968). A recent survey which screened a large number of school children in the three major Indian metropolitan cities found the prevalence of the ß-thalassaemia trait to be 5.5% in New Delhi, 2.6% in Bombay and 10.2% in Calcutta (ICMR collaborative study, 1988).

1.11.2 Prevalence of α-thalassaemia in India

There have been a number of published reports describing Hb Bart's analysis on cord blood, from different parts of India. Hb Bart's was observed in 4% of Bengali neonates (Swarup, 1965), in 4% of Punjabi newborns (Dash et al, 1988) and in about 1% of neonates screened in Bombay (Sukumaran, 1974). A recent report based on a survey of a large number of cord blood samples reported the prevalence of α-thalassaemia to be 0.5% in Bombay, 2.9% in New Delhi and 0.8% in Calcutta (ICMR collaborative study, 1988).

Apart from studies in the general population, there have been a number of published reports describing the prevalence of α-thalassaemia in some isolated tribal groups in India, which are almost totally endogamous. Amongst the families of sickle cell patients from a tribal group in Orissa, the frequency of the α-thalassaemia gene was estimated to be 29% (Kulozik et al, 1988) and in a tribal group from South India it was as high as 81% (Brittenham et al, 1981). A raised level Hb Bart's was observed in 23.15% of tribal neonates in Western India (Patra et al, 1983).

Hb Bart's hydrops fetalis has not been reported in Indians and there have been only anecdotal reports of Hb H disease in Indians (Chatterjea, 1961; Chouhan et al, 1970; Agarwal and Mehta, 1982). Hb electrophoresis of over 2754 blood samples did not reveal a single case of Hb H or Constant Spring (Misra, 1989). Hb Koya Dora, a nondeletional α-thalassaemia determinant has been described at a prevalence of 10% in the Koya Doras, a tribal group in Andhra Pradesh (DeJong et al, 1975).

1.11.3 Distribution of haemoglobin variants in India

A wide variety of Hb variants have been described on the Indian subcontinent. The most extensive data exists for Hb S; the highest sickle gene frequencies, varying between 19-35%, have been found in the hills of central and southern India amongst the aboriginal tribal populations (Brittenham et al, 1979). The highest prevalence of Hb E is found in Northeast India, where 20-60% of some Assamese tribal groups carry Hb E. The prevalence of Hb E in the general population of West Bengal is 3-4%. Hb D has been found mainly in Northern India with a prevalence of approximately 1% (Brittenham, 1981). Besides these common Hb variants, a number of rare variants have been described including Hb Hofu, Hb J, Hb J Meerut, Hb K, Hb L, Hb Lepore, Hb M, Hb M Iwate, Hb D, Hb Q India and Hb Rampa (Brittenham, 1983).

1.11.4 A current perspective of the problem of thalassaemia in India

It has been estimated that there are over 21 million ß-thalassaemia carriers in India (Agarwal, 1986). Seven thousand thalassaemia major children are estimated to be born each year and most of them in the absence of diagnosis and treatment, succumb to anaemia or infection before the age of three years (Sangani et al, 1990). Due to the prevalence of a variety of Hb variants, interactions with ß-thalassaemia do occur, often resulting in a severe phenotype. However it is the problem of ß-thalassaemia major that exists on a larger scale and imposes a greater drain on health services.

There are a number of medical centres in India which offer treatment to thalassaemia major children by providing transfusions at day care clinics. Desferrioxamine is marketed in India. However patient compliance is poor primarily due to socio-economic reasons. In order to highlight the experiences of families in India with children affected by thalassaemia major, 200 such families were interviewed in Bombay. The results of this survey showed that the treatment of thalassaemia major patients places a significant, unavoidable and increasing demand on the public

health services. Due to the potentially large number of patients and difficulties with long term management, there has been an evasion of the problem, failure of planning, no provisions for prevention, and inadequate treatment leading to premature death of affected children. The burden of chronic inherited disease has heavy economic and social demands on such families as, besides caring for the sick child their lives are dominated by the high cost of treatment, often 20-30% of their income. Seven mothers with no healthy children and 27 with only one healthy child had been sterilised. Ninety percent of couples in the reproductive age group felt that prenatal diagnosis was a necessity. Ignorance and prejudice in the community lead to social isolation for 20% of the families interviewed (Sangani et al, 1990).

There are two centres in Bombay which perform prenatal diagnosis in the second trimester by fetal blood sampling and globin biosynthesis studies. Both groups believe that prenatal diagnosis, although so far offered only after the birth of an affected child, is acceptable to most couples at risk. However almost all couples have stated a preference for first trimester prenatal diagnosis (Colah et al, 1990; Natrajan PG pers. comm.). Facilities for first trimester DNA diagnosis are nonexistent in India at present, and a fraction of the couples at risk have their blood and chorion villus samples referred to DNA diagnostic laboratories abroad.

1.12 Thesis plan

At present, mutation detection is the most appropriate strategy for prenatal diagnosis of ß-thalassaemia. In order to establish such a program in India it is mandatory to know the spectrum of ß-thalassaemia mutations in the Asian Indian population. Also it would be essential to develop mutation detection techniques that are applicable in a developing country. These results would also be valuable to centres in a number of countries which provide prenatal diagnosis of ß-thalassaemia for their Asian Indian immigrant communities. Chapter 3 describes the development of a PCR based nonradioactive mutation detection technique, allele specific PCR. This technique along with DNA sequence analysis was

used to characterise the ß-thalassaemia mutations of 656 unrelated disease carriers from seven different regions of the Indian subcontinent. The application of these results to prenatal diagnosis of ß-thalassaemia in Asian Indians is then discussed.

Chapter 4 describes the construction of ß-globin gene haplotypes of 196 normal chromosomes and 419 chromosomes carrying the ß-thalassaemia gene. These studies provide information on the origin and spread of ß-thalassaemia mutations on the Indian subcontinent. The regional variation in the distribution of mutations serve as useful anthropological markers, thus contributing to an understanding of the population genetics of the Indian subcontinent. The study of DNA polymorphisms within the ß-globin gene cluster would also be valuable for prenatal diagnosis by family linkage studies.

Chapter 5 describes preliminary α-globin gene mapping studies performed on 51 unrelated Asian Indians suspected to have α-thalassaemia. These studies aim to describe the molecular basis of α-thalassaemia in Asian Indians and would be useful for a thalassaemia control program in this population.

A number of couples who are at a recurrent risk for genetic disease find prenatal diagnosis and termination of pregnancy an unacceptable option. Preimplantation diagnosis would overcome the ethical objections that these couples may have to abortion following prenatal diagnosis. Chapter 6 describes the experiments that were done in order to develop an approach for preimplantation diagnosis of ß-thalassaemia in humans using DNA amplification by the PCR, to analyse 10-30 cell embryonic biopsies.

Establishing a thalassaemia control program in India would be a complex task, as a number of financial, social and demographic factors would have to be considered. The concluding chapter describes some of these considerations and makes recommendations for thalassaemia control in India.

CHAPTER TWO

MATERIALS AND METHODS

2.1 Materials

2.1.1 Chemical reagents

All chemical reagents (Analar R grade) were purchased from BDH, Sigma or Boehringer Mannheim Corporation (BCL) with the exception of the following: agarose (Ultrapure, Bethseda Research Laboratories; BRL), Sieving agarose (NuSieve, ICN Biomedicals Ltd) and urea (National Diagnostics).

2.1.2 Enzymes

Enzymes were purchased from a variety of sources. Restriction enzymes were obtained from New England Biolabs and BCL. Amersham International supplied the Klenow fragment of *Escherichia coli (E Coli)* polymerase I and T4 polynucleotide kinase. The modified T7 DNA polymerase was purchased from United States Biochemicals (USB) and *Taq* polymerase from Perkin-Elmer Cetus Corporation. Proteinase K was obtained from BCL.

2.1.3 Radioactive nucleotides

The following radioactive nucleotides were purchased from Amersham International: [α^{32}P]dATP, specific activity = >4000 Ci/mmol, [α^{32}P]dCTP, specific activity = 3000 Ci/mmol, [γ^{32}P]dATP, specific activity = >3000 Ci/mmol and [α^{35}S]dATP, specific activity = >1000 Ci/mmol.

2.2 Extraction of genomic DNA

Genomic DNA was extracted from anticoagulated whole blood and buffy coat samples using a variation of the procedure described by Old and Higgs (1983). Erythrocytes were haemolysed by dilution

with 10-20 ml of reticulocyte saline (0.13 M NaCl, 5 mM KCl and 7.4 mM $MgCl_2$). The erythrocyte ghosts and the white blood cells were isolated as a cell pellet by centrifugation at 3000 g, for 15 minutes (min) at 4°C. If DNA was to be extracted from a buffy coat sample this first step was omitted. The cells in the pellet were then lysed by gentle shaking for 10 min with approximately 25 ml of lysis mix (15 mM NH_4Cl and 1 mM $KHCO_3$) and the nuclei pelleted by centrifugation at 3000 g for 10 min. To the nuclei pellet, 10 ml of lysing solution (100 mM NaCl and 25 mM EDTA), 0.5 ml of 10% SDS (sodium dodecyl sulphate) and 20 µl of Proteinase K (20 mg/ml) was added. Digestion was carried out for a minimum of 4 hours at 37°C. The digested protein was extracted by shaking with an equal volume of a 1:1 mixture of phenol (chromatography grade, equilibrated with 1 M Tris pH 8.0) and chloroform. The phases were separated by centrifugation at 3000 g for 10 min at room temperature. The viscous aqueous phase was drawn off and transferred to a 50 ml Falcon tube using a wide mouthed Pasteur pipette to prevent shearing of the DNA. A second phenol/chloroform extraction was performed to remove any remaining protein, followed by a chloroform extraction to remove any contaminating phenol. DNA was precipitated by adding one half volume of 7.5 M NH_4OAc and two volumes of cold 99.7% ethanol. The DNA was pelleted by centrifugation at 3000 g, for 15 minutes at 4°C, washed twice with 70% ethanol and then resuspended in TE buffer (10 mM Tris pH 8 and 1 mM EDTA).

To further ensure DNA purity, the DNA was reprecipitated with a half volume of 7.5 M NH_4OAc and two volumes of 99.7% ethanol, stored overnight at -20°C, spun at 3000 g for 15 minutes, washed with 70% ethanol and resuspended in TE buffer. The quantity of DNA was determined by measuring the optical density of a 1:200 dilution using an ultra violet (UV) spectrophotometer, taking readings at 260 nm. A DNA solution of 1g/100 ml in a 1 cm light path has an absorbance of 200 units at 260 nm, accordingly the yield of DNA was calculated.

2.3 Southern Blot Hybridisation

Southern blot hybridisation includes the transfer of DNA fragments from an agarose gel to a filter (Southern, 1975) which is then hybridised with a radioactively labelled DNA probe. The DNA probes used were either genomic probes, prepared from cloned fragments of genomic DNA containing the sequence of interest, or synthetic oligonucleotide probes. Prior to transfer, genomic DNA was digested with an appropriate restriction enzyme and the DNA fragments were separated by agarose gel electrophoresis.

2.3.1 Restriction endonuclease digestion

Certain restriction endonucleases cleave double stranded DNA in a sequence specific manner, generating discrete blunt or sticky-ended fragments. Cleavage is accomplished by incubating the DNA with the enzyme in an appropriate buffer, using a 2-4 fold enzyme excess to promote complete digestion (Fuchs and Blakesey, 1983). Five µg of genomic DNA was used for each digestion and the reactions were incubated at the temperature recommended by the manufacturers, usually 37°C for a minimum of 4 hours. The digests were carried out in a 30 µl volume in the presence of 20 units of the restriction enzyme using reaction buffers supplied by the manufacturer. To promote efficient digestion, 1 mM spermidine was added to the digestion mixture. PCR products were subjected to restriction enzyme digestion in a similar manner.

2.3.2 Agarose gel electrophoresis

Separation of DNA fragments according to size was done by agarose gel electrophoresis, which is based upon the ability of nucleic acids to migrate in an electric field as they possess a negative charge. Gels act as molecular sieves and the extent to which a fragment travels in the electric field is inversely proportional to the logarithm of its molecular weight.

For Southern blotting of cleaved genomic DNA, electrophoresis was performed in a 0.8% agarose gel using horizontal gel tanks (BRL). Agarose was dissolved in TAE buffer (40 mM Tris-acetate

pH 8.3, 20 mM NaOAc and 0.2 mM EDTA) by boiling the mixture in a microwave oven. Digested DNA was mixed with a 1/6 th volume of loading dye (0.025% bromophenol blue, 0.025% xylene cyanol, 15% Ficoll type 400 in distilled water) and loaded into the wells. The gel was run in TAE buffer at 1.5 V/cm for about 16 hours at room temperature. Such gels efficiently separate DNA fragments in the range of 0.5 to 25 kb.

To determine the size of a DNA fragment, a comparison is made between the migration of DNA fragments of known and similar size, to the unknown fragment. The DNA fragments were stained by ethidium bromide, a flourescent DNA intercalating dye, at a concentration of 0.5 mg/ml, then visualised by transillumination with 260-320 nm UV light and photographed with a camera fitted with a red gelatin filter and Polaroid 677 fast film.

2.3.2.1 3'- End labelling of the marker fragment

To provide a radioactively labelled marker fragment for gel electrophoresis, the Klenow fragment of *E coli* DNA polymerase I (Jacobsen et al, 1974) was used to 3'-end label λ bacteriophage DNA cleaved with *Hind*III. Ten μg of digested DNA was suspended in 20 μl of buffer supplied by the manufacturer and incubated with 1 unit of the Klenow fragment in the presence of 2 μl of [α^{32}P]dATP. One unit of Klenow is defined as the amount of enzyme required to convert 10 nmol of deoxyribonucleotides to an acid insoluble form in 30 min at 37°C. The reaction was incubated at room temperature for 10-15 min and terminated by the addition of 2 μl of 500 mM EDTA prior to separation on a G50-80 Sephadex spun column (section 2.3.4.2). An aliquot of the labelled fragment mixed with a small amount of the unlabelled fragment was run together with the DNA samples to act as a marker fragment, and was visible both by transillumination of an ethidium bromide stained gel and following autoradiography.

2.3.3 Southern blotting

Following photography, the DNA was denatured by immersing the gel in a solution of 0.5 M NaOH and 1.5 M NaCl with gentle

shaking for 30 min. The DNA was then neutralised by immersing in 1 M Tris-HCl pH 7.5 and 3 M NaCl for two hours. A large tray was filled with 20xSSC (3 M NaCl, 300 mM trisodium citrate) and a glass plate placed across it. Three sheets of Whatman 3MM paper of size 36 x 24 cm were layered over the glass plates, with their ends dipping into the 20xSSC, to act as wicks. The gel was carefully placed on top of the Whatman sheets, taking care not to trap any air bubbles. A sheet of clingfilm was stretched loosely over the tray to cover the Whatman, but was cut away so that the gel was exposed. A sheet of nylon membrane (Hybond-N, Amersham) was cut to the size of the gel, soaked in water and placed upon the gel. Two sheets of Whatman 3MM paper, the same size as the gel, were wetted and placed on top of the filter; these were followed by a thick wad of paper towels, a glass plate and a weight. The gel was left overnight to allow DNA transfer, then the filter was retrieved and rinsed briefly in 2xSSC. The DNA was fixed to the filter by exposure to UV light for 2 min.

2.3.4 Hybridisation with genomic DNA probes

2.3.4.1 Multiprime radioactive labelling

Genomic DNA probes were radiolabelled using the Multiprime Labelling Kit (Amersham). In this reaction, random hexamers were used to prime DNA synthesis along template (Feinberg and Vogelstein, 1983). Approximately, 25 ng of the genomic DNA probe, was denatured at 93°C to produce a single-stranded template to which the hexamers annealed. DNA synthesis took place in a reaction volume of 50 µl in the presence of 5 units of Klenow, 50 mCi [α^{32}P]dCTP and excess dATP, dTTP and dGTP, suspended in a Tris-HCl buffer. The reaction was incubated at 37°C for 30 min and then the labelled probe was separated from unincorporated nucleotides and hexamers by G50-80 Sephadex spun column chromatography.

2.3.4.2 G50-80 Sephadex column chromatography

The G50-80 Sephadex columns were prepared by plugging the end of a 1 ml syringe with glass wool, then filling it with a slurry of G50-80 Sephadex suspended in 3xSSC (0.45 M NaCl, 45 mM trisodium citrate). The column was then equilibrated with 100 µl TE buffer by centrifugation, at 2000 g for 4 min. The labelled DNA probe was diluted to 100 µl with TE, applied to the top of an equilibrated column and spun at 2000 g for 4 min. The unincorporated nucleotides were retained on the column while the labelled probe passed through and was collected in an Eppendorf tube. The labelled probe was quantified by diluting 2 µl of the probe with 5 ml of scintillant and measuring in a scintillation counter.

2.3.4.3 Hybridisation of filters

In appropriate conditions the labelled DNA probes will specifically hybridise to their complementary DNA sequences bound to the nylon filter, permitting identification and analysis of the DNA fragments by autoradiography. Efficient probe hybridisation requires the presence of salt to stabilise DNA heteroduplex interactions between the probe and its target by masking the repulsive forces exerted by the two DNA strands. Heteroduplex formation is also influenced by the hybridisation temperature and is favoured if carried out at levels 10-20°C below the melting temperature (Tm) of the DNA.

The filters were pre-hybridised for one hour at 42°C in a sealed polythene bag containing 5-10 ml of pre-hybridisation buffer (50% formamide, 3xSSC, 10xDenhardt's [0.02% Ficoll 400, 0.02% BSA and 0.02% polyvinylpyrrolidene], 20 µg/ml sonicated denatured salmon sperm DNA [ssDNA] and 2% SDS). Pre-hybridisation reduces nonspecific binding of the labelled probe by blocking nonspecific binding sites on the filter with sheared carrier DNA and other high molecular weight molecules. Following pre-hybridisation, the pre-hybridisation buffer was squeezed out and approximately $4x10^6$ cpm of radiolabelled probe was added in 2 ml of hybridisation buffer, which was similar to the pre-hybridisation buffer except

that it also contained 10 µg poly(A) / 10% dextran sulphate. Dextran sulphate increases the hybridisation rate by increasing the concentration of the DNA probe because the DNA probe is excluded from the volume occupied by the polymers. Furthermore, dextran sulphate favours the formation of networks, collections of large numbers of probe molecules bound to the target DNA on the filter which increases the sensitivity of hybridisation. Sufficient probe was added to the hybridisation buffer so as to give a concentration of $2x10^6$ cpm/ml. The DNA probe was heated to 95°C for 5 min to denature it to single strands just before addition to the hybridisation buffer. Hybridisation was performed overnight at 42°C.

2.3.4.4 Post-hybridisation washing

Following hybridisation, the filters were removed from the plastic bags, immersed in 2xSSC and shaken gently for about 30 min at room temperature. This low stringency wash was done twice so as to remove most of the nonspecific radioactivity and unhybridised probe. The filters were then transferred to a more stringent wash, 0.1xSSC and 0.1% SDS at 65°C, and washed until a reading between 2-5 cpm was recorded by passing the Geiger counter over the filter. The stringency of the washing conditions determines the degree of nonspecific probe mismatch, thus a sufficiently stringent wash should be used so that only the target DNA sequence is identified. After post-hybridisation washing, the filters were blotted dry and wrapped in clingfilm.

2.3.4.5 Autoradiography

Autoradiography was used to detect the fragments containing the DNA sequence that had hybridised to the radioactively labelled probes. The filters were placed in a cassette with DuPont Lightening Plus intensifying screens, a sheet of X-ray film (Kodak XAR-5) was placed over the filters and left to expose overnight at -70°C. The length of exposure was determined by signal strength. The filters were developed in an automated developer (Kodak X-OMAT).

2.4 The polymerase chain reaction (PCR)

The PCR is an *in vitro* DNA amplification procedure that can rapidly isolate and exponentially amplify a specific fragment of DNA (Mullis and Faloona, 1987). Such a high degree of target enrichment enables a number of manipulations to be performed on the gene fragment of interest, like, mutation detection. The PCR has thus revolutionised approaches for gene analysis.

DNA amplification by the PCR classically requires two sequence specific oligonucleotide primers that flank the DNA fragment to be amplified. Repeated thermal cycling is also necessary. The cycling regimen comprises heat denaturation to produce two single-stranded templates from the double-stranded target DNA molecule, annealing of the primers to their complementary sequences and extension of the annealed primers with DNA polymerase. Using the deoxynucleotides present in the reaction mixture, the DNA polymerase enzyme, by virtue of its 5'-3' polymerase activity, synthesises a new strand of DNA starting from the 3' end of each primer. The primers are designed to anneal to opposite strands and oriented with their 3' ends towards each other, so that the newly synthesised DNA strand extends across the region between these primers, effectively copying that particular DNA segment. As the primers are incorporated into the end of the extension product the length of the fragment produced is the distance between the 5' termini of the primers. Since the extension products have primer binding sites, they serve as templates in successive cycles of the reaction. The result is an exponential accumulation of the fragment of interest, approximately 2^n-fold where n is the number of cycles.

2.4.1 DNA polymerase enzyme

The development of a thermostable DNA polymerase isolated from the bacterium *Thermus aquaticus (Taq)* transformed the PCR from a laborious research tool to a widely used, routine procedure (Saiki et al, 1988a). *Taq* polymerase is double-stranded DNA polymerase with a molecular weight of 94 KDa. The optimum enzyme concentration recommended is 1-4 units per 100 µl

reaction. At higher enzyme concentrations the yield of the target product decreases and nonspecific products tend to be amplified. AmpliTaq (Perkin-Elmer Cetus) was the DNA polymerase used in the studies described in this thesis, at a concentration of 2 units per 100 µl reaction volume. One enzyme unit is defined as the amount of enzyme that will incorporate 10 nmol of dNTPs into acid insoluble form in 30 min at 74°C.

2.4.2 Buffer and deoxynucleotides

Taq polymerase has an optimum activity at pH 8.0-8.3, in the presence of 50 mM KCl (Chein et al, 1976). The Mg^{2+} concentration alters the efficiency and specificity of *Taq* polymerase; the optimum concentration of $MgCl_2$ varies with different primer pairs, and should be determined for each primer pair by titration (Saiki, 1989). All the primer pairs used in this thesis amplified satisfactorily at a Mg^{2+} concentration of 1.5 mM of $MgCl_2$. Thus the buffer used for PCR contained 50 mM KCl, 10 mM Tris-HCl pH 8.3 and 1.5 mM $MgCl_2$.

The deoxynucleotide triphosphate (dNTP) mixture used for PCR consisted of equal amounts of 2'-deoxyguanosine-5'-triphosphate (dGTP), 2'-deoxyadenosine-5'-triphosphate (dATP), 2'-deoxythymidine-5'-triphosphate (dTTP) and 2'-deoxycytidine-5'-triphosphate (dCTP), purchased as neutralised 100 mM solutions (BCL). A dNTP stock solution was prepared containing 2 mM of each dNTP and was stored at -20°C. The total dNTP concentration used in the PCRs described in this thesis was 800 µM.

2.4.3 Primer design

The length of primers used varied from 19 to 30 nucleotides. When possible primers were designed such that they were approximately 50% G-C rich, with a random base distribution and no polypurine or polypyrimidine stretches. Substantial secondary structure which could result in self-annealing as well as complementarity between primer pairs was avoided. Twenty pmol of each primer was used in the reactions.

2.4.4 Template

A single molecule of target DNA sequence is theoretically enough to initiate DNA amplification by PCR. Too great a concentration of template produces a sub-optimal product, believed to be due to the increased likelihood of nonspecific amplification and because template molecules tend to reassociate with each other instead of the primers. For most reactions in this study the template used was 200-500 ng of genomic DNA.

2.4.5 Thermal cycling parameters

Denaturation of DNA was performed at 93°C for 1 min; use of a lower temperature could result in incomplete separation of double stranded DNA, thus impairing reaction efficiency. The annealing temperature used varied from 55°C to 66°C, higher annealing temperatures increase the stringency of the reaction. Annealing was allowed to occur for 1 min, though probably a shorter period would have sufficed. Extension by *Taq* polymerase is optimum at 72°C and the duration depends on the length of the target sequence. One min for synthesis of 1000 bp has been estimated to be adequate (Saiki, 1989). To ensure complete extension of the final PCR product, the extension time was prolonged in the final cycle. Typically 25-30 cycles of PCR were performed using Techne programmable Dri-Block PHC-1 and PREM (LEP scientific) thermal cyclers.

2.4.6 Limitations of the PCR

Taq polymerase has an overall nucleotide misincorporation rate of 0.25% for 30 amplification cycles. Errors occurring at an early stage will be magnified to a greater degree than those occurring later. The more serious drawback of PCR is its exquisite sensitivity, as minute amounts of contaminating DNA also get amplified, resulting in false positive results.

2.4.7 Visualisation of the PCR product

2.4.7.1 Minigels

Small gels (minigels) were used to visualise PCR products. As minigels are thinner and use narrower loading slots, less DNA is required to visualise a fragment. The running buffer was TBE (89 mM Tris HCl, 89 mM boric acid and 2.5 mM EDTA) and the concentration of the agarose was 1%. The DNA sample, a 10 µl aliquot of the PCR product, was mixed with a 1/6 th volume of loading dye (0.05% bromophenol blue, 15% Ficoll Type 400 in distilled water) and loaded into the wells. Minigels were run at high voltages (>10 V/cm) for approximately 45 min.

2.4.7.2 Sieving agarose gel electrophoresis

Sieving agarose can be used at high concentrations to separate much smaller DNA fragments than conventional agarose gels. The gel is prepared, loaded and run in the same way as an ordinary agarose gel and can resolve fragments in the range 50-1000 bp. Gels prepared with 2% sieving agarose (NuSieve, ICN Biomedicals, Ltd) and 2% ordinary agarose were used to resolve DNA fragments in the size range 50-550 bp.

2.5 Oligonucleotide synthesis and purification

The synthetic DNA molecules used were single-stranded oligonucleotides of between 19-30 bases in length. Oligonucleotides were used as PCR primers, DNA sequencing primers and as allele specific oligonucleotide (ASO) probes.

2.5.1 Synthesis of oligonucleotides

Oligonucleotides were synthesised using a Applied Biosystems 381 DNA synthesiser, by the ß-cyanoethyl phosphoramite method of DNA synthesis (Sinha et al, 1984). The individual DNA chains were synthesised on a solid support in a 3'-5' direction using phosphoramites which are chemically modified nucleosides. The

chemical groups protect the nucleoside from the actions of the reagents used in the synthesis procedure.

2.5.2 Purification of synthetic oligonucleotides

2.5.2.1 Deprotection and elution

The first stage in the purification process was to cleave the DNA molecule from its support, which was done by breaking the labile ester bond formed between the linker of the support and the 3' hydroxyl of the initial nucleoside with 100% NH_4OH. The protecting groups on the exocyclic amines were removed by ammoniolysis at 55°C during a 5-12 hours incubation. After deprotection the DNA-NH_4OH solution was cooled to room temperature.

The DNA was precipitated from the solution by adding 1/10th volume of 3M NaOAc pH 5.6 and 2 volumes of 99.7% ethanol. The DNA was pelleted by centrifugation at 9000 g, for 15 min, washed twice with 70% ethanol and resuspended in 200 µl of distilled water. DNA concentration was calculated by measuring the optical density at 260 nm in a UV spectrophotometer and considering the length of the oligonucleotide. The oligonucleotide solution was stored at -20°C.

2.5.2.2 Preparative polyacrylamide gel electrophoresis

Oligonucleotides to be used as sequencing primers were purified by polyacrylamide gel electrophoresis (PAGE) which separated the full length oligonucleotides from incompletely synthesised shorter chains. Polyacrylamide gels have a high resolving capacity and are capable of separating DNA fragments that differ in length by one base. Gels were formed by the polymerisation of acrylamide and N, N'-methylene bisacrylamide in the presence of TEMED (N,N,N',N'-tetramethylethylene diamine), and 0.1% $NH_4(SO_4)_2$. The acrylamide and N, N'-methylene bisacrylamide were stored at 4°C as a 40% (38:2) stock solution and diluted with TBE buffer when required. Gels were cast between two siliconised gel plates separated by 1.5 mm thick plastic spacers, using a 20% polyacrylamide solution with 7 M urea as the denaturing agent. A comb was placed in the

top of the gel to create the loading wells. Electrophoresis took place in a vertical gel tank (BRL). Before loading the samples, the gel was pre-run for 30 min at 6 W and then the wells were flushed out with TBE to remove any unpolymerised acrylamide or urea that may have diffused out of the gel.

Oligonucleotides to be purified by PAGE were prepared by lyophilisation of the deprotected DNA-NH_4OH solution, by vacuum centrifugation at 55°C. The powdered DNA was resuspended in 500 µl distilled water. A 250 µl volume of this solution was mixed with an equal volume of loading buffer containing bromophenol blue dye (20 mM Tris-HCl pH 7.5, 1 mM EDTA, 8 M urea and 0.05% bromophenol blue) prior to loading. Electrophoresis was performed at 6 W until the bromophenol blue dye front reached the bottom of the gel.

After electrophoresis, the apparatus was dismantled, the gel left on one glass plate and covered with clingfilm. This glass plate was then removed, with the gel left resting on the clingfilm, a thin layer chromatography (TLC) plate was then inserted beneath the gel and the DNA bands identified by long wave UV shadowing. The slowest migrating band on the ladder of DNA products representing, the full sized oligonucleotide, was excised from the gel. The slice of polyacrylamide gel was crushed, suspended in 2 ml elution buffer (500 mM NH_4OAc, 10 mM $Mg(OAc)_2$ and 1 mM EDTA) and eluted overnight at 37°C.

The eluate was passed down a Sep Pak C18 ion exchange column (Waters Associates), which had been activated with 99.7% ethanol and then rinsed with distilled water. To ensure that the oligonucleotide was retained on the column, the eluate was passed through twice. The purified DNA molecules were recovered from the column with 60% methanol and lyophilised in a Speed Vac concentrator (Savant). The purified oligonucleotide was resuspended in 200 µl distilled water.

2.6 Dot-blot hybridisation

ASOs were used for the direct detection of point mutations by dotblot hybridisation. Under appropriate stringent conditions the ASO probe will bind stably to only fully complementary DNA

templates. The stability of the target DNA:probe hybrid depends on several factors; target complexity, the length and base composition of the probe, base mismatches between the probe and the target, the ionic strength of the hybridisation buffer and the hybridisation temperature (Thein and Wallace, 1986). An index of this stability is the dissociation temperature (Td), the temperature at which half of the DNA:probe duplexes are dissociated which is derived by the empirical formula: Td (°C) = 4 (G+C) + 2 (A+T) where G, C, A and T indicate the number of the corresponding nucleotides in the ASO (Suggs et al, 1981). This equation is valid only for perfectly matched duplexes between 11-20 bp long in 1 M Na^+ and serves as a guide for determining an appropriate hybridisation temperature. As the diploid human genome is $3X10^9$ bp in length, an oligonucleotide nucleotide probe must be at least 17 nucleotides long in order to hybridise to a unique sequence (Thomas, 1966). The use of PCR products as the target sequence has greatly facilitated mutation detection by ASOs.

2.6.1 Preparation of dot-blot filters

The amplified DNA samples were diluted to a 177 µl volume with distilled water, denatured with 13 µl 6 M NaOH and made up to a final volume of 200 µl with 10 µl of 500 mM EDTA and placed on ice for 10 min. The perspex dot-blot apparatus (HybriDot, BRL) was assembled with two pieces of 3 MM Whatman paper and the nylon filter (Hybond N, Amersham, UK) which was cut to size and wetted with distilled water. The denatured DNA samples were pipetted into the individual wells of the dot-blot apparatus, in between two 200 µl applications of 2 M NH_4OAc to neutralise the DNA samples. Under vacuum the samples were transferred directly onto the filter. The apparatus was then dismantled, the filter rinsed in 2XSSC and the DNA bound to it by UV crosslinkage for 1 min. Besides the DNA samples a series of controls were dot-blotted on each filter. The positive controls were amplified genomic DNA from individuals known to carry the mutations to be sought and the negative control was DNA from an unaffected individual. It was important to ensure that an equal

amount of each sample was dot-blotted on the filter so that a valid comparison could be made between different samples.

2.6.2 Hybridisation with synthetic oligonucleotide probes

Radiolabelling of the ASO probes was done by 5'-end labelling with T4 polynucleotide kinase, followed by purification by either PAGE or column chromatography.

2.6.2.1 5'-End labelling of the oligonucleotides

T4 polynucleotide kinase, catalyses the transfer of the [γ^{32}P] label of [γ^{32}P]dATP to the 5'-hydroxyl terminus of the oligonucleotide probe (Richardson, 1981). One unit of the enzyme is defined as the amount catalysing the production of 1 nmol of acid insoluble ^{32}P in 30 min at 37°C. As only one moiety of ^{32}P may be incorporated into one molecule of oligonucleotide the specific activity of the isotope determines that of the oligoprobe; hence the isotope used was of high specific activity. Approximately 15 pmol of probe was labelled in a 10 µl reaction volume which contained kinase buffer (67 mM Tris-HCl pH 8.0, 10 mM $MgCl_2$, 10 mM DTT), 100 pmol [γ^{32}P]dATP and 4 units of T4 polynucleotide kinase. The mixture was incubated at 37°C for 30 min and the reaction stopped by either freezing or the addition of 10 µl of termination solution (20 mM Tris-HCl pH 7.5, 1 mM EDTA, 0.05% xylene cyanol, 0.05% bromophenol blue, 8 M urea), depending on the purification method to be used.

2.6.2.2. Separation of labelled oligoprobes by PAGE

Labelled and unlabelled oligonucleotides may be separated by PAGE as a result of the differential mobilities of the two species of molecules in an electric field. The presence of the negatively charged ^{32}P atom causes the labelled probe to migrate further than the unlabelled one. Ten µl of the termination solution was added to the labelled ASO, and the 20 µl sample was loaded onto a pre-run 15% polyacrylamide gel of 0.8 mm thickness, containing

7M urea. The gel was run at 25 W until the bromophenol blue dye front reached the bottom of the gel.

After electrophoresis the gel apparatus was dismantled and the gel left resting on the back glass plate. The labelled oligoprobes were located by autoradiography; a sheet of X-ray film was exposed over the gel, developed and then used as a template to cut out the fragments of gel containing the radiolabelled probe. This probe was eluted from the gel slice by overnight incubation in TE buffer at 37°C. The probe was quantitated by scintillation counting as described previously and stored at 4°C.

2.6.2.3 Separation of labelled probes by spun column chromatography

An alternative method for oligoprobe separation was spun column chromatography using a G25-50 Sephadex column. The labelling reaction was stopped by freezing the 10 µl aliquot of the labelling reaction, which was then diluted to a 100 µl volume with TE buffer. The diluted probe was spun down an equilibrated column as previously described.

2.6.3 Hybridisation of filters with radiolabelled ASOs and post-hybridisation washing

Nylon filters were pre-hybridised, in sealed plastic bags containing 5 ml of hybridisation buffer (0.1 M Tris HCl pH 8.0, 6 mM EDTA, 1 M NaCl, 5xDenhardts [1 mg/ml Ficoll, 1 mg/ml polyvinylpyrrolidone, 1 mg/ml BSA], 10% dextran sulphate, 0.5% NP-40 and 100 µg/ml sheared ssDNA) for one hour at the hybridisation temperature. The labelled probe was then added to the buffer so as to give a concentration of $2X10^6$ cpm per ml of buffer. Hybridisation was done for 2-3 hours at the hybridisation temperature, which was 10°C below the calcualted Td of the probe.

The filters were then washed in 6XSSC for 5 min at 5°C below the Td. Stringency of the washing conditions was increased by raising the washing temperature in 2-3°C increments. The stringency of the wash was assessed by autoradiography, at -70°C, with exposure periods of one hour, the hybridisation patterns of the

DNA samples being compared with known controls. Differential hybridisation was considered to have been achieved when only the positive control on the filter retained the probe together with any of the test samples and the negative control showed no sign of the probe.

2.7 Direct Genomic Sequencing

DNA sequencing was performed by the dideoxynucleotide chain termination method which is based on the ability of a DNA polymerase to extend a primer, annealed to the template that is to be sequenced, until a chain-terminating nucleotide is incorporated (Sanger et al, 1977). The chain terminating nucleotide analogues used were the 2',3'-dideoxynucleoside 5'-triphosphates (ddNTPs). When suitable mixtures of dNTPs and one of the four ddNTPs are used, enzyme-catalysed polymerisation was terminated in a fraction of the population of chains at each site where the ddNTP could be incorporated, thus four reactions each with a different ddNTP give complete sequence information. The resulting series of unique fragments are separated by polyacrylamide gel electrophoresis. Incorporated [α^{35}S]dATP radiolabels the strand being randomly extended permitting analysis by autoradiography. The modified bacteriophage T7 DNA polymerase, Sequenase, used has properties of high processivity, low 3' to 5' exonuclease activity, high speed and efficient usage of ddNTPs (Tabor and Richardson, 1987). The DNA templates were both double-stranded and single-stranded PCR products.

2.7.1 Template preparation

2.7.1.1 Electroelution into a high salt solution

Double-stranded PCR products were purified for DNA sequencing by electroelution into a high salt solution. The DNA sample was separated by gel electrophoresis using a 1.2% agarose gel in TBE buffer. After staining with ethidium bromide, the DNA fragment was visualised using a hand held, long wave 366 nm UV light. The gel slice containing the DNA was excised and placed in 2 ml

electroelution buffer (20 mM Tris-HCl pH 8.0, 5 mM NaCl and 0.2 mM EDTA) for 20 min. Following equilibration, the gel slice was placed in a unidirectional electroelutor (International Biotechnologies) and 75 µl of a high salt solution (7.5 M NH_4OAc and 0.01% bromophenol blue) was introduced in the well to form a high salt cushion, the blue dye enabling the integrity of the cushion to be monitored. Electrophoresis was conducted at 100 V for one hour, during which the DNA was eluted from the gel slice and migrated into the salt cushion where it was retained. Following electrophoresis, the high salt solution was removed using a drawn out Pasteur pipette and transferred to an Eppendorf tube. The gel slices were stained with ethidium bromide and visualised by UV transillumination to ensure that the electroelution was complete. The electroeluted DNA was precipitated in 2 volumes of 99.7% ethanol at -20°C, pelleted by centrifugation for 15 min at 9000 g, washed twice with 70% ethanol, dried and resuspended in distilled water.

2.7.1.2 Electroelution onto DEAE paper

The PCR product was isolated from unincorporated nucleotides and other nonspecific amplification products by electrophoresis in a 1.2% agarose gel at 100 V for 2-3 hours. The gel was stained with ethidium bromide and the amplified fragment was visualised with a long wave, 366 nm UV hand held lamp. An incision was made in the gel below the relevant band with a sterile scalpel blade and into this incision a piece of DEAE membrane (NA45 Schleicher & Schuell) was inserted. This membrane had been previously cut into strips, soaked in autoclaved double distilled water for 30 min with three changes of the water, and then stored in 1XTBE buffer at 4°C. The amplified fragment was run onto the DEAE membrane by continuing electrophoresis for 15 min. When there was danger of contamination from larger, nonspecific fragments, a second piece of DEAE membrane was inserted into the gel above the band of interest in order to trap the nonspecific fragments. After electrophoresis the DEAE paper was washed in distilled water to remove any traces of adherent gel and the PCR product was eluted from the membrane in 400 µl of elution

buffer (20 mM Tris HCl pH 7.6, 1.5 M NaCl and 1 mM EDTA) for 20 min at 65°C. DNA was extracted with phenol/chloroform, followed by ether and precipitated by adding 2 volumes of 99.7% ethanol, placed on dry ice for 20 min, and pelleted by centrifugation at 9000 g for 15 min. The DNA pellet was resuspended in 180 µl of distilled water, reprecipitated with 20 µl of 3 M NaOAc pH 5.6 and 400 µl 99.7% ethanol, placed on dry ice for 20 mins and respun; the pellet was washed in 70% ethanol, dried and resuspended in 21 µl of distilled water.

2.7.2 Sequencing with Sequenase

DNA sequencing with Sequenase is carried out in two steps, following primer-template annealing. The first is the extension of the annealed primer along the template in the presence of all four dNTPs, one of which is radioactively labelled and the second step is a chain termination reaction, using the appropriate dideoxynucleotide triphosphate (ddNTP). Sequencing reactions were performed using the reagents provided in the Sequenase sequencing kit of USB Corp. Seven µl of the purified sequencing template, 1-2 pmol, was denatured in a 10 µl volume, containing approximately 4 pmol of the sequencing primer and the Sequenase buffer (40 mM Tris HCl pH 7.5, 20 mM $MgCl_2$ and 50 mM NaCl) at over 95°C for 7 min. Template and primer were rapidly annealed by snap cooling in dry ice for 10 seconds followed by brief centrifugation and incubation on ice for 2 min. The labelling reaction was performed on ice for 3 min following the addition of 1 µl 0.1 M DTT, 2 µl of a 1:10 dilution of the labelling mix (1.5 µM dGTP, 1.5 µM dCTP and 1.5 µM dTTP), 0.5 µl [α^{35}S]dATP and 2 µl of a 1:8 dilution of Sequenase to the reaction mix. The chain termination reactions were performed in four separate Eppendorf tubes; 3.5 µl of the labelling reaction was aliquoted into each tube and mixed with 2.5 µl of one of the four ddNTP mixes, each mix contained 80 µM of each of the four dNTPs and 8 µM of one of the ddNTPs. The reactions were incubated at 37°C for 3 min, then stopped by addition of 4 µl of the stop solution (95% formamide, 20 mM EDTA, 0.05% bromophenol blue and 0.05% xylene cyanol FF).

2.7.3 Gel electrophoresis and autoradiography

The samples were subjected to electrophoresis in a 6% polyacrylamide gel in which 8M urea was incorporated as the denaturing agent. Wells were created using a "sharks tooth" comb (Biolabs), the comb and gel spacers were 0.4 mm thick, both glass plates were siliconised and a vertical gel apparatus (Tyler) was used for electrophoresis. Gels were pre-run for 30 min at 25-30 watts. The samples were heated at 94°C for 3 min, prior to loading onto a pre-run gel. The gel was run at 25-30 watts until the bromophenol blue dye front reached the bottom of the gel, which generally took about two hours. When electrophoresis was complete, the two glass gel plates were prised apart leaving the gel resting on the back one; this was immersed in a solution of 10% acetic acid and 10% methanol for 30 min. The acid fixes the DNA in the gel and the methanol prevents the gel swelling which would cause distortion of the running tracks. The gel was floated off its supporting plate, the fluid sucked away, the gel transferred to a piece of 3 MM Whatmann paper and dried down onto it using a vacuum gel dryer at 80°C for 30 min. The dried gel was set up for overnight autoradiography. The DNA sequence was read from the autoradiograph, typically 200-250 bases of sequence could be read off a single gel. Longer runs of sequence were obtained by sequentially loading aliquots of the sequencing reactions.

The methods used in this chapter were based on those described by Sambrook et al (1989).

CHAPTER THREE

CHARACTERISATION OF ß-THALASSAEMIA MUTATIONS

3.1 Introduction

Prenatal diagnosis of ß-thalassaemia by DNA analysis is best performed by direct detection of mutations. Despite there being approximately 100 described ß-thalassaemia mutations, prenatal diagnosis by mutation detection remains feasible as each ethnic group in which the disease is prevalent has its own spectrum of mutations around 10-15 in number of which only a handful account for roughly 90% of all ß-thalassaemia genes in that population (Kazazian and Boehm, 1988). However, in order to perform prenatal diagnosis by mutation detection, knowledge of the distribution and frequency of ß-thalassaemia mutations in the target population is imperative.

3.1.1 Methods of mutation detection

A number of methods are available for the detection of base changes in DNA, some of which have been successfully applied for the identification of ß-thalassaemia mutations. Cloning and sequencing of ß-thalassaemia genes (Spritz et al, 1981) was the approach used to characterise the first 40 or so mutations. A number of ß-thalassaemia mutations create or abolish cleavage sites for restriction enzymes, for example, codon 39 (C-T) creates a cleavage site for *Mae* I (Thein et al, 1985). Another means of detecting ß-thalassaemia mutations is hybridisation of genomic DNA with radiolabelled ASO probes (Pirastu et al, 1983). Denaturing gradient gel electrophoresis (DDGE) is a powerful method for the detection of a point mutation in a fragment of DNA. The basis of which is that heteroduplexes are formed when the test genomic DNA sample is denatured and mixed with excess of a labelled single-stranded DNA probe, a single base change will alter the mobility of the duplex when subjected to gel electrophoresis

under denaturing conditions, thus enabling identification of mutations within a specified region of the ß-globin gene (Myers et al, 1985).

The application of the PCR to rapidly amplify specific DNA fragments has revolutionised the detection of base changes in DNA. Point mutations in these amplified fragments can be detected by a number of methods including those previously used on genomic DNA. Some of the PCR based methods which have been successfully employed for the detection of ß-thalassaemia mutations are: direct DNA sequencing of the amplified product (Wong et al, 1987), restriction enzyme digestion of the amplified gene fragment for mutations which alter a restriction enzyme cutting site (Di Marzo et al, 1988), dot-blot hybridisation with ASOs which can be either radioactively (Saiki et al, 1986) or nonradioactively labelled (Saiki el al, 1988b), reverse dot-blot hybridisation where sequence specific oligonucleotides are immobilised on a membrane and the DNA sample to be tested is labelled and hybridised (Saiki et al, 1989), DDGE analysis of amplified regions of the ß-globin gene (Losekoot et al, 1990), chemical cleavage analysis of mismatched heteroduplexes formed by annealing amplified normal and mutant DNA sequences (Dianzani et al, 1991), allele specific PCR (Newton et al, 1989) and PCR with fluorescence-tagged primers (Chehab and Kan, 1989).

In January 1989, when this thesis was begun, the widely used method for the detection of ß-thalassaemia mutations was dot-blot hybridisation with radiolabelled ASOs. There was a need for the development of an alternative nonradioactive PCR based mutation detection method, which was achieved by the experiments described in this chapter.

3.1.2 Known Asian Indian ß-thalassaemia mutations

At the time this thesis was begun there were two reports on the spectrum of ß-thalassaemia mutations amongst Asian Indians, both of which described studies conducted in Asian Indian immigrant communities, there were no reports of surveys conducted on the Indian subcontinent itself. Kazazian et al (1984b) studied 44 ß-thalassaemia chromosomes from patients of Asian Indian origin living in the USA and found eight different ß-

thalassaemia mutations which in order of frequency, were IVS-1 nucleotide 5 (G-C), the 619 bp deletion, codons 41/42 (-CTTT), codons 8/9 (+G), codon 15 (G-A), codon 16 (-C), IVS-1 nucleotide 1 (G-T) and the 25 bp deletion at the 3' end of IVS-1. Subsequently, Thein et al (1988) reported a survey of 102 ß-thalassaemia chromosomes from affected individuals belonging to the Asian Indian immigrant community in the UK; the nine mutations that were identified included -88 (C-T) and cap site position +1 (A-C) along with those described in the previous study, the 25 bp deletion at the 3' end of IVS-1 was not found; of these mutations IVS-1 nucleotide 5 (G-C), the 619 bp deletion, codon 8/9 (+G), IVS-1 nucleotide 1 (G-T) and codon 41/42 (-CTTT) accounted for 89% of all alleles. The aims of experiments described in this chapter were to comprehensively describe the spectrum of ß-thalassaemia mutations prevalent on the Indian subcontinent and to look for any regional differences in their distribution. So that this data could form the basis for a prenatal diagnosis program in the Asian Indian population.

3.2. Subjects

The subjects were 656 unrelated individuals of Asian Indian origin who were diagnosed by their referring physicians to be heterozygous for ß-thalassaemia. Diagnosis of the carrier state for ß-thalassaemia was made by routine haematological studies including examination of a peripheral blood smear and determination of the MCV and MCH followed by estimation of the Hb A_2 level by quantitative cellulose acetate electrophoresis. Each of the ß-thalassaemia carriers included in this study had a reduced MCH and MCV with a normal serum ferritin along with an elevated Hb A_2. All the individuals studied were at risk of having a thalassaemia major offspring and were interested in prenatal diagnosis.

Amongst the subjects studied, 418 live in India, mainly in the areas around Bombay and New Delhi. In Bombay the subjects were personally interviewed and bled at the Nanavati Hospital Medical Research Centre with the kind assistance of Dr P.G. Natrajan and his colleagues. Subjects from New Delhi were interviewed and bled at the All India Institute of Medical Sciences by Professor I.C.

Verma and colleagues. The remaining 238 subjects were immigrants of Asian Indian origin residing in the UK who were referred to the National Haemoglobinopathy Referrence Laboratory, John Radcliffe Hospital, Oxford for prenatal diagnosis of ß-thalassaemia by DNA analysis. 10-20 ml of blood was collected from each of these individuals and, if possible, from their family members including their affected child, using either heparin or EDTA as the anticoagulant. The samples from Bombay were spun down, supernatant plasma discarded, the buffy coat stored at -20°C and transported on dry ice to Oxford. Professor I.C. Verma kindly sent the blood samples from New Delhi through a rapid courier service and samples from various centres in the UK were sent by post.

Of the 656 ß-thalassaemia carriers studied, there were 311 couples, in the remaining 34 individuals either the partner was not a disease carrier or it was not possible to collect the blood sample. Two hundred and nineteen of the 311 couples studied (70.4%) had already had an affected thalassaemia major offspring. Thus less than 30% of the couples studied were referred for genetic counselling and prenatal diagnosis before the birth of a thalassaemia major offspring.

The precise ethnic origins of each of the individuals studied were carefully ascertained by interview to be as follows: 201 (30.6%) were from Gujarat, 167 (25.5%) from Northwest Pakistan, 144 (22%) from Punjab, 114 (17.4%) from Sindh, 13 (2%) from Maharashtra, 10 (1.5%) from Bengal and Bangladesh and 7 (1%) from Tamil Nadu (figure 3.1). These regions are geographical ones not corresponding to political boundaries, Sindh is situated in Pakistan and Punjab lies on both sides of the Indo-Pakistan border. The unequal distribution of subjects reflects, in part, the varied prevalence of the ß-thalassaemia trait in the different regions of the Indian subcontinent and, in part, the bias of the sampling procedure that was employed. Population movements in 1947 (partition of India on the occasion of her Independence) brought substantial numbers of ß-thalassaemia carriers to Bombay from Sindh and into the area around New Delhi from parts of

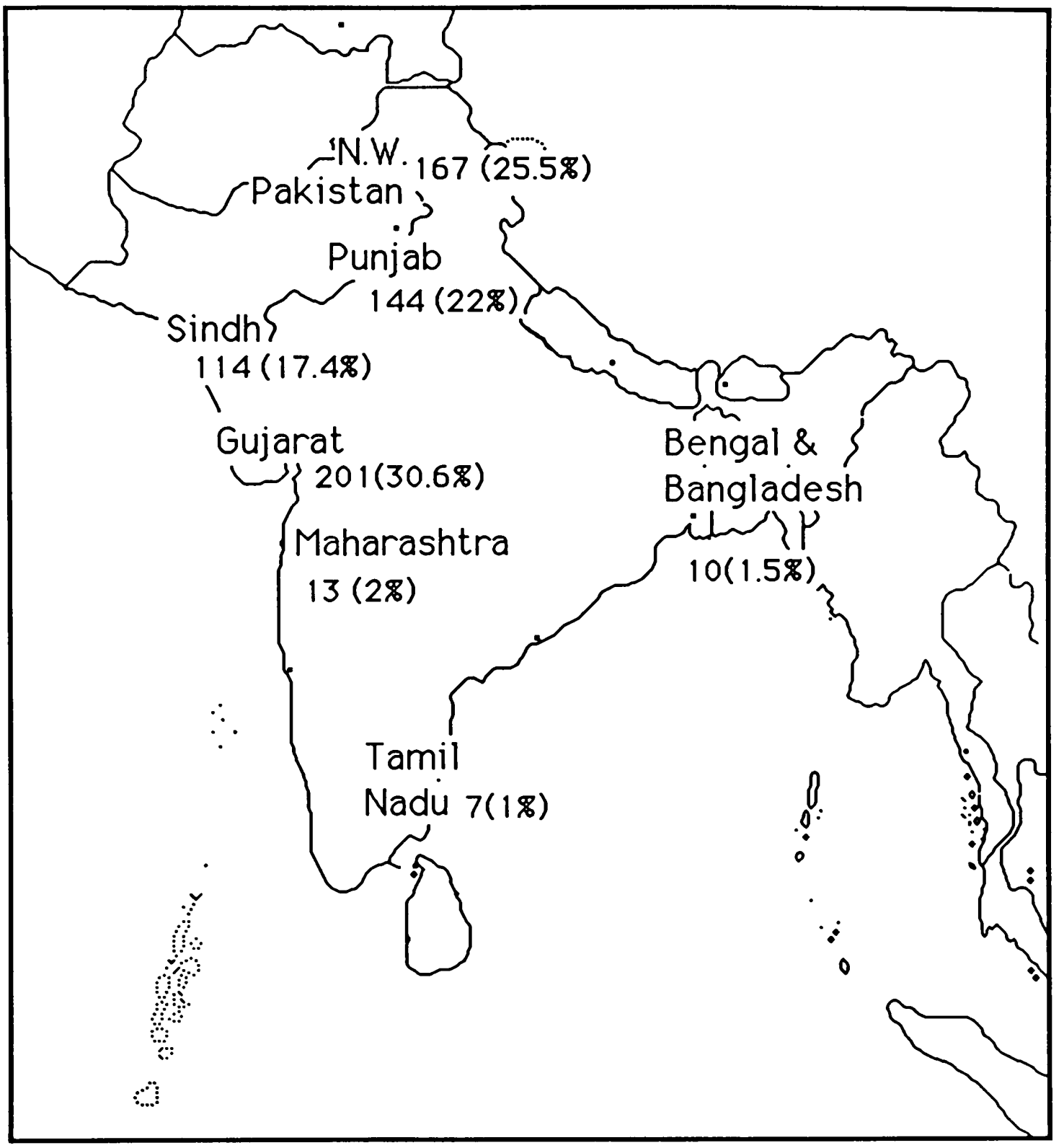


Figure 3.1 Map of the Indian subcontinent showing the number of ß-thalassaemia carriers that were studied from each region and the percentage in brackets is the proportion of the sample that originated from that region.

Punjab which then became Pakistan; as samples from India were collected in these cities the Sindhi and Punjabi ethnic groups are well represented. The immigrant Asian Indian community in the UK comprises mainly of individuals originating from Pakistan and Punjab along with the Gujaratis who had initially settled in East Africa before migrating to the UK

3.3 Strategy

The strategy devised for the characterisation of the ß-thalassaemia mutations in the population sample studied was as follows:

1. Screen for the presence of the five common Asian Indian mutations: IVS-1 nucleotide 5 (G-C), the 619 bp deletion, IVS-1 nucleotide 1 (G-T), codons 8/9 (+G) and codons 41/42 (-CTTT).
2. Screen the remaining uncharacterised subjects for the rare Asian Indian mutations: codon 15 (G-A), codon 16 (-C), -88 (C-T), cap site position +1 (A-C) and the 25 bp deletion at the 3' end of IVS-1.
3. Sequence the ß-globin genes of the individuals that remained uncharacterised.
4. On identification of a ß-thalassaemia mutation by DNA sequencing, screen the remaining uncharacterised subjects for the presence of that mutation.

3.4 Methods

DNA was extracted (section 2.2) from the anticoagulated blood samples and their ß-thalassaemia mutations were identified by the following methods.

3.4.1 Dot-blot hybridisation

For dot-blot hybridisation, the amplified target DNA was spotted on a nylon filter and probed with radioactively labelled ASOs. A 597 bp fragment of the ß-globin gene encompassing the first and second exon and the first intron was amplified using primers 5'-GGCCAATCTACTCCCAGGAG (5' end located at 68 nucleotides upstream from the cap site) and 5'-ACATCAAGGGTCCCATAGAC (5'

end located at IVS-2 nucleotide 24). The PCR was performed using 0.5 µg of genomic DNA and 2 units of *Taq* polymerase in a 100 µl volume containing 20 pmol of each primer, 800 µM total dNTPs, 10 mM Tris HCl pH 8.3, 50 mM KCl and 1.5 mM $MgCl_2$. The thermal cycling regimen consisted of 30 cycles of denaturation at 93°C for 1 min, annealing at 55°C for 1 min and extension at 72°C for 1.5 min; extension was prolonged for 3 min in the final cycle. A fraction of the PCR product, typically one twentieth of the total reaction volume, was dot-blotted on to the filter membrane as described in section 2.6.1.

These dot-blots were hybridised sequentially to four different ASO probes: IVS-1 nucleotide 5 (G-C), IVS-1 nucleotide 1 (G-T), codons 8/9 (+G) and codons 41/42 (-CTTT) (gift from Dr SL Thein) (table 3.1). As all samples to be analysed were known to be heterozygous for ß-thalassaemia, hybridisation with the normal ASO probe for each of these mutations was not done.

Table 3.1 Allele specific oligonucleotides used for dot-blot hybridisation

β-Thalassaemia mutation	Oligonucleotide sequence 5'-3'
IVS-1 nucleotide 5 (G-C)	CTTGATAGCAACCTGCCCA*
IVS-1 nucleotide 1 (G-T)	CTTGATACCAAACTGCCCA*
codons 8/9 (+G)	CTCCTGAGGAGAAGGTCTGC
codons 41/42 (-CTTT)	TGGACCCAGAGG/TTGAGTCCTTT

The positions of the single base changes are underlined and that of the 4 bp deletion marked by /. The ASO sequences, marked with *, are complementary to the coding strand while the others are complementary to the noncoding strand.

The ASO probes were labelled at their 5' end with [γ^{32}P]dATP using T4 polynucleotide kinase (section 2.6.2.1). After

hybridisation at 55°C for 2-3 hours, filters were subjected to washes of increasing stringency, so as to obtain differential hybridisation (section 2.6.3), indicated by the positive control retaining the probe and the negative control showing no sign of the probe. Tetra-alkylammonium ions reduce the dependence of Td on the G-C content of the ASO, thus when included in the post-hybridisation wash, the stringency of the wash would be a function of only probe length and not its composition (Wood et al, 1985). In order to have uniform post-hybridisation washing conditions for each of the four probes used, the filters were washed twice for 20 min each time with 3 M tetramethylammonium chloride (Me_4NCl), 50 mM Tris HCl (pH 8), 2 mM EDTA and 0.1% SDS at 50°C.

Results obtained by dot-blot hybridisation were not uniformly satisfactory. Nonspecific hybridisation resulted in equivocal signals which made the assignment of the genotype impossible. The other serious drawback was the use of [γ^{32}P]dATP with its high specific activity, particularly as it became apparent that a number of different probes would be required to characterise all the ß-thalassaemia mutations in this population entailing repeated radiolabelling procedures. Thus there was a need to develop a nonradioactive PCR based method of mutation detection to overcome these problems. The technique that was developed for the detection of ß-thalassaemia mutations was the Amplification Refractory Mutation System (ARMS) (Newton et al, 1989) or allele specific PCR which was nonradioactive and conveniently combined the amplification and mutation detection in one step making the procedure rapid and less labour intensive.

3.4.2 Allele specific PCR

3.4.2.1 Principles

The basis of allele specific PCR is that oligonucleotides with a mismatched 3' end will not function as amplimers in the PCR under appropriate specific reaction conditions (Newton et al, 1989). Thus an oligonucleotide primer could be designed such that its 3' end is complementary to the sequence of the mutation that is being screened for so that when the PCR is performed under

stringent conditions the presence of an amplified product will indicate the presence of the mutation while its absence will indicate the presence of normal DNA sequence at the site of the mutation. The allele specific primer will anneal to the target DNA only when there is a perfect match at the 3' nucleotide, thus permitting extension with the DNA polymerase enzyme and successful amplification. In the absence of a perfect match primer-template annealing would fail to occur and there would be no amplification. Allele specific primers may be complementary to the coding or the noncoding strands and are accordingly coupled with an upstream or downstream primer, respectively. To ensure that absence of the amplified fragment was due to absence of the mutation being sought and not due to inefficacy of the PCR, an internal control, in the form of a pair of primers which co-amplify another fragment of the human genome, was always included. To ensure that the reaction conditions were of the correct stringency negative and positive controls were also included.

3.4.2.2 Primer Design

Allele specific primers were designed such that the nucleotide at the 3' end of each primer was complementary to the DNA sequence alteration caused by the ß-thalassaemia mutation being sought. These oligonucleotide primers were 30 bp in length so as to permit the use of a high annealing temperature which would make the PCR conditions stringent. To enhance their specificity a deliberate mismatch was also introduced at position four nucleotides from the 3' end. The sequence of the allele specific primer was almost entirely determined by the position and nature of the mutation being sought. However when this deliberate mismatch was selected, care was taken to ensure an approximately 50% G-C content in the primer sequence and to avoid creating regions which could result in self-annealing or complementarity with the other PCR primers.

Newton et al (1989) demonstrated that the degree of specificity of the allele specific PCR correlated with the 3' nucleotide mismatch type; C/T, A/A and T/T mismatches were found to be considerably more refractory to extension by *Taq* polymerase than G/T, T/G, A/C or C/A mismatches. It was also shown that further,

deliberate mismatches, which destabilise the primer/template complexes, render the primers increasingly refractory, particularly if the additional mismatch is moved progressively closer to the 3' end. Subsequently Kwok et al (1990) demonstrated that of all possible 3'-terminal mismatches, A/G, G/A and C/C mismatches reduced overall product yield almost 100-fold, A/A mismatch about 20-fold while all other mismatches, particularly those with T, had a minimal effect on PCR product yield. While designing allele specific primers for this study, the terminal mismatch was determined entirely by the mutation being sought. However, when determining the deliberate mismatch at position four nucleotides from the 3' end, A/G, G/A and C/C mismatches were favoured.

For each of the ß-thalassaemia mutations being sought both normal and mutant allele specific primers were synthesised; the 3' nucleotide of the normal primer being complementary to the wild-type nucleotide, while that in the mutant primer was complementary to the point mutation. A DNA sample from an individual homozygous for a particular mutation would be amplified only with the mutant allele specific primer and not with the normal one due to the absence of the wild type nucleotide at that site. Similarly DNA from a normal individual would be amplified only with the normal allele specific primer and not with the mutant one; DNA from a heterozygote would be amplified with both allele specific primers. Thus this technique permitted identification of all states.

The problem of nonspecific amplification was encountered with two allele specific primers designed for this thesis: the IVS-1 nucleotide 1 (G-T) normal primer and the codon 15 (G-A) mutant primer. The former when used with a mutant template had a C/T 3' nucleotide primer-template mismatch and the latter when used with a normal template had a T/G mismatch. Altering the type and position of the deliberate mismatch at position four nucleotides from the 3' end did not resolve this problem. Finally, the direction of these two primers was reversed such that they now were complementary to the noncoding strand and the resulting 3' mismatches were now G/A and A/C respectively. This completely resolved the problem of nonspecific amplification and is in concurrence with the observation reported by Kwok et al (1990) that mismatches of T with either G or C have a minimal

destabilising effect. As the 3' nucleotide of the allele specific primer is determined by the mutation being sought, the problem of nonspecific amplification can be resolved by altering the deliberate mismatch, failing which the direction of the primer should be reversed, so as to create an unstable terminal mismatch.

3.4.2.3 Establishment of specific reaction conditions

To determine the efficacy of allele specific PCR for the detection of ß-thalassaemia mutations, allele specific primers were first synthesised for the detection of four common Asian Indian mutations, IVS-1 nucleotide 5 (G-C), IVS-1 nucleotide 1 (G-T), codons 8/9 (+G) and codons 41/42 (-CTTT). Both normal and mutant allele specific primers were synthesised and tested with genomic DNA from individuals known to be homozygous and normal for these mutations, according to the dot-blot hybridisation studies. The test reactions were performed using 0.5 µg genomic DNA and 2 units *Taq* DNA polymerase in a 100 µl reaction volume containing 20 pmol of each primer, 800 µM total dNTPs, 10 mM Tris pH 8.3, 50 mM KCl and 1.5 mM $MgCl_2$. The thermal cycling regimen consisted of 30 cycles, with denaturation at 93°C for 1 min, annealing at 60°C for 1 min and extension at 72°C for 1.5 min; in the last cycle extension was prolonged for 3 min. Using this protocol nonspecific amplification was observed i.e. amplification when the mutant primer was used with a normal DNA control sample and when the normal primer was used with a sample homozygous for that particular mutation. The various approaches that were then tried in order to achieve specificity were Mg^{+2} concentration titrations, decreasing the amount of template DNA, primer concentration titrations, decreasing dNTP concentration, increasing annealing temperatures and reducing the number of cycles. Each of these parameters was found to affect the amplification efficiency of the PCR.

In order to perform simultaneous screening for all four common mutations using the same thermal cycler, albeit different reaction tubes, it was important to establish specific reaction conditions that worked satisfactorily with each of the eight allele specific primers being tested. Furthermore, it was hoped that these uniform reaction conditions could also be applied to screen for

various ß-thalassaemia mutations in the future, using a number of different allele specific primers. It was found by trial and error, that by increasing the annealing temperature to 66°C and reducing the number of cycles to 25, but maintaining all other reaction variables as described above, specific amplification was achieved. The reaction volume was scaled down to 25 µl without any effect on efficacy but improving economy. The time taken for the PCR was reduced by performing a combined annealing and extension at 66°C for 2 min. These reaction conditions were established as the uniform specific ones at which all allele specific PCRs described in this thesis were conducted.

3.4.2.4 Detection of ß-thalassaemia mutations

Allele specific primers were coupled with either upstream primer A or one of the downstream primers B, C or D (figure 3.2), depending on the location of the mutation in the ß-globin gene and the direction of the allele specific primer i.e. whether it was complementary to the coding or the noncoding strand. The PCR was performed in 25 µl reaction volumes with 20 pmol of each of the four primers i.e. the allele specific primer coupled with the appropriate upstream or downstream primer and the pair of internal control primers, using the established uniform specific reaction conditions. The thermal cycling regimen consisted of 25 cycles with denaturation at 93°C for 1 minute followed by combined annealing and extension at 66°C for 2 minutes. In the last cycle extension was prolonged for 3 minutes. A 10 µl aliquot of the PCR product was visualised after electrophoresis on a gel {1.5% NuSieve (I.C.N. Biomedicals, Ltd.) and 1.5% agarose} by ethidium bromide staining and UV transillumination. As the population sample studied comprised individuals known to be heterozygous for ß-thalassaemia, studies with only the mutant allele specific primers were necessary. Screening experiments included a positive and negative control, the former being a genomic DNA sample from an individual known to carry the mutation that was being sought and the later was from an unaffected individual. The presence of the amplified fragment of expected size in the positive control sample and its absence in the

negative control were confirmed before considering the results of any screening experiment.

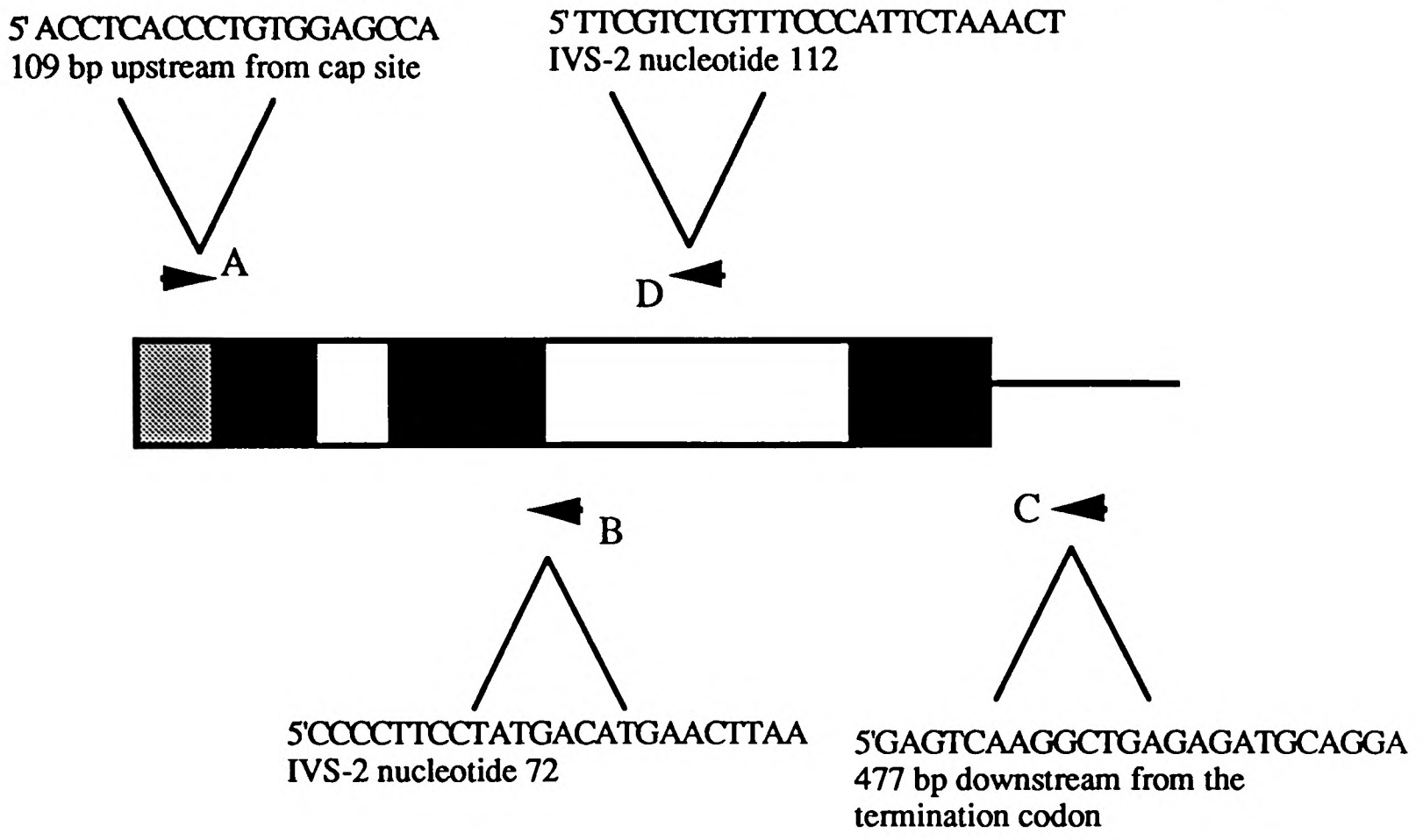


Figure 3.2 Schematic representation of the ß-globin gene showing the sequences and positions of the 5' nucleotides of the common primers that were coupled with allele specific primers. Shaded areas represent exons, unshaded areas, introns and the stippled area represents the 5' untranslated region.

Normal and mutant allele specific primers used for the detection of the four common Asian Indian ß-thalassaemia mutations are shown in table 3.2 along with the primers they were coupled with and the expected product size. Screening for the rare Asian Indian ß-thalassaemia mutations was performed using allele specific primers shown in table 3.3. Normal allele specific primers were synthesised only for the codon 15 (G-A) and codon 16 (-C) mutations as DNA samples from individuals homozygous for the other mutations were not available for testing.

Table 3.2 Allele specific primers for the detection of common Asian Indian β-thalassaemia mutations

Mutation	Oligonucleotide sequence 5'-3'	Product size
IVS-1 nt 5 (G-C)	N-CTCCTTAAACCTGTCTTGTAACCTTGTTAC	285 bp
	M-CTCCTTAAACCTGTCTTGTAACCTTGTTAG	285bp
IVS-1 nt 1 (G-T)	N-GATGAAGTTGGTGGTGAGGCCCTGGGTAGG*	454 bp
	M-TTAAACCTGTCTTGTAACCTTGATACGAAA	281 bp
codons 8/9 (+G)	N-CCTTGCCCCACACGGCAGTAACGGCACACT	214 bp
	M-CCTTGCCCCACACGGCAGTAACGGCACACC	214 bp
codons 41/42	N-GAGTGGACAGATCCCCAAAGGACTCAAAGA	443 bp
(-CTTT)	M-GAGTGGACAGATCCCCAAAGGACTCAACCT	443 bp

Table 3.3 Allele specific primers for the detection of rare Asian Indian β-thalassaemia mutations

Mutation	Oligonucleotide sequence 5'-3'	Product size
codon 15 (G-A)	N-TGAGGAGAAGTCTGCCGTTACTGCCCAGTG*	500 bp
	M-TGAGGAGAAGTCTGCCGTTACTGCCCAGTA*	500 bp
codon 16 (-C)	N-TCACCACCAACTTCATCCACGTTCACGTTG	238 bp
	M-TCACCACCAACTTCATCCACGTTCACGTTC	238 bp
-88 (C-T)	M-TCACTTAGACCTCACCCTGTGGAGCCTCAT*	684 bp
cap site +1 (A-C)	M-ATAAGTCAGGGCAGAGCCATCTATTGGTTC*	596 bp
IVS-1 3'end -25 bp	M-CTCTGGGTCCAAGGGTAGACCACCAGCATA	386 bp

N =normal and M =mutant allele specific primers with their mismatched sequences underlined, primers marked with * were complementary to the noncoding strand and were coupled with primer B and the others which were complementary to the coding strand were coupled with primer A. (nt=nucleotide)

Besides the allele specific primers used for the detection of Asian Indian ß-thalassaemia mutations, some were also synthesised for the detection of mutations known to be prevalent in other population groups (table 3.4), so as to enable prenatal diagnosis in members of these ethnic groups. Care was taken to ensure that all the allele specific primers functioned efficiently at the uniform reaction conditions. This greatly facilitated the day to day working of the prenatal diagnosis service, as mutation detection of a large number of ß-thalassaemia mutations, occurring in different ethnic groups could be conveniently performed at uniform reaction conditions.

3.4.2.5 Selection of internal controls

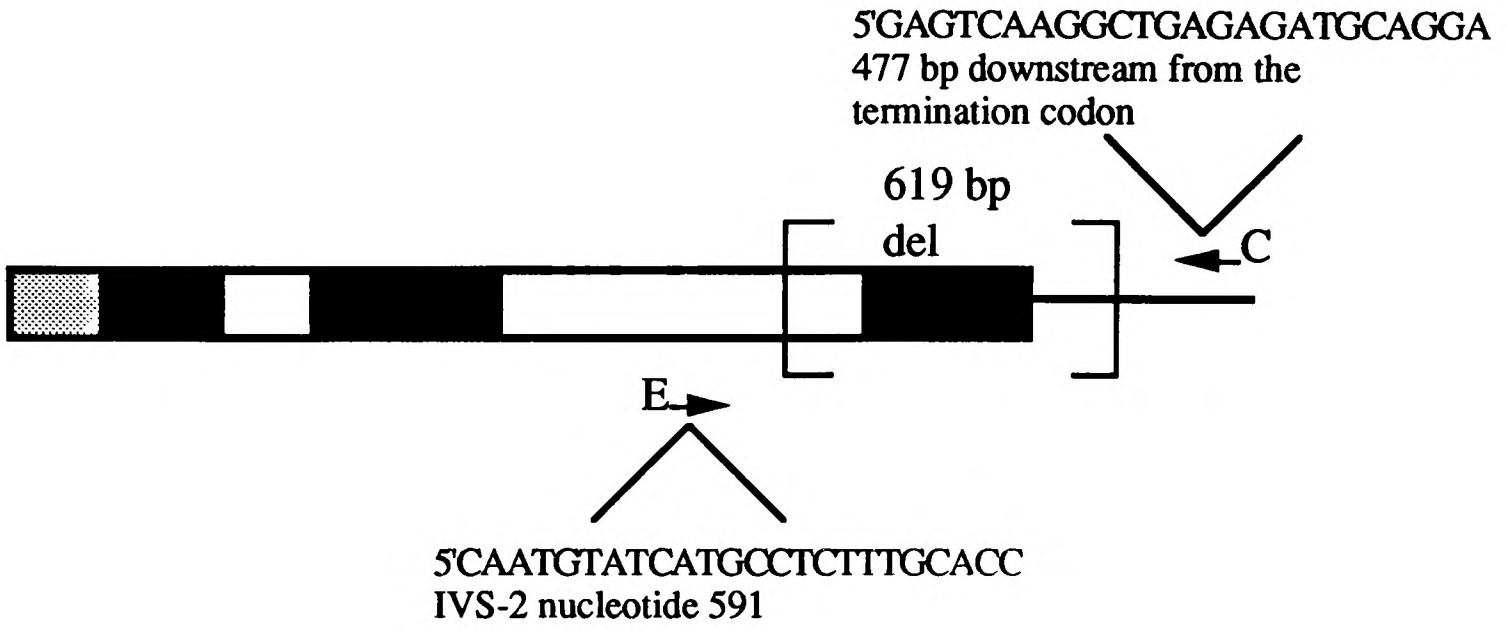
Each allele specific PCR must include an internal control, so that successful amplification of this internal control fragment would indicate the efficacy of the reaction. Thus the absence of the allele specific PCR product, could be inferred to be due to the absence of the mutation being sought and not due to failure of the PCR. On establishment of the uniform reaction conditions for the allele specific PCR experiments, it was necessary to select internal control primer pairs which could be efficiently co-amplified at these conditions.

The primer pair, C and E, which amplify a 861 bp fragment spanning the site of the 619 bp deletion at the 3' end of the ß-globin gene (figure 3.3), were fortunately found to work efficiently at these uniform reaction conditions. The 619 bp deletion can be easily detected by PCR using primers C and E, as in the presence of the deletion, the amplified product would be reduced in size by 619 bps. Also as almost all the ß-globin gene mutations that were being sought were located at the 5' end of the gene, the 3' end could be conveniently co-amplified without any interference with

Table 3.4 Allele specific primers for the detection of some β-thalassaemia mutations present in other ethnic groups

Mutation	Ethnic group	Reference	Oligonucleotide sequence 5'-3'	Product size
IVS-1 nt 6 (T-C)	Mediterraneans	Orkin et al, 1982b	TCTCCTTAAACCTGTCTTGTAACCTTCATG	286 bp
IVS-1 nt 1 (G-A)	Mediterraneans	Orkin et al, 1982b	TTAAACCTGTCTTGTAACCTTGATACGAAT	281 bp
codon 39 (C-T)	Mediterraneans	Trecartin et al, 1981	CAGATCCCCAAAGGACTCAAAGAACCTGTA	436 bp
IVS-2 nt 745 (C-G)	Mediterraneans	Orkin et al, 1982b	TCATATTGCTAATAGCAGCTACAATCGAGG**	737 bp
codon 24 (T-A)	American Black	Goldsmith et al, 1983	ACCAACCTGCCCAGGGCCTCACCTCCT	257 bp
-29 (A-G)	American Black, Chinese	Antonarkis et al, 1984	GGGCAGGAGCCAGGGCTGGGTATG*	619 bp
IVS-2 3' end (CAG-AAG)	Iranian, Egyptian & Black	Wong et al, 1989	TCATGTTCATACCTCTTATCTTCCTCTCAA**	635 bp
IVS-2 nt 654 (C-T)	Chinese	Cheng et al, 1984	GAATAACAGTGATAATTTCTGGGTTATGGT**	829 bp
codon 17 (A-T)	Chinese	Chang and Kan, 1979	CTCACCACCAACTTCATCCACGTTCAGCTA	240 bp
-28 (A-G)	Chinese	Orkin et al, 1983	AGGGAGGGCAGGAGCCAGGGCTGGGCTTAG*	624 bp
codon 71/72 (+A)	Chinese	Cheng et al, 1984	CAAGAAAGTGCTCGGTGCCTCTAA***	233 bp

All these are mutant allele specific primers with their mismatched sequences underlined; those marked with asteriks are complementary to the noncoding strand and were coupled with one of the downstream primers, * with primer B, ** with primer C and *** with primer D; and the others which are complementary to the coding strand were coupled with primer A. Internal control primers F and G were used with the allele specific primers coupled with C.



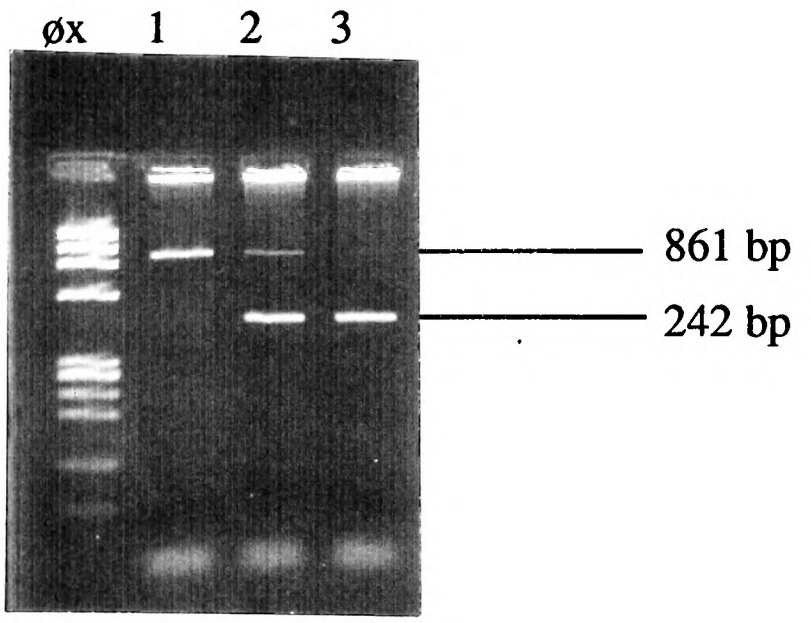


Figure 3.3 Detection of the 619 bp deletion by PCR. The upper half is a schematic representation of the ß-globin gene showing primers C and E which encompass a 861 bp fragment from the 3' end of the gene that includes the site of the deletion. Primer sequences and the position of the 5' nucleotide are shown. Ethidium bromide stained agarose gel electrophoresis of the PCR products: lane 1, normal, lane 2, heterozygous and lane 3 homozygous for the 619 bp deletion. The marker is øx 174 DNA digested with *Hae*III (øx).

mutation detection; making it possible for this primer pair to function as an internal control as well as simultaneously screen for the 619 bp deletion. It could be difficult to differentiate the 242 bp band that would be observed in a sample carrying the 619 bp deletion, from a 200-300 bp band that would occur if the sample carried a mutation which produced a fragment of that size with its allele specific primer. However if the 242 bp fragment was observed in several allele specific PCRs, the 619 bp deletion would be a probable diagnosis which could be quickly confirmed by repeating the reaction with only the internal control primers (figure 3.3).

The other primer pair which was found to amplify efficiently at these uniform reaction conditions was F - 5'-AGTGCTGCAAGAAGAACAACTACC (5' end at IVS-2 nucleotide 685) and G - 5'-CTCTGCATCATGGGCAGTGAGCTC (5' end at 22 bp downstream from the termination site of the $^{G}\gamma$-globin gene) which amplify a 328 bp fragment from the $^{G}\gamma$-globin gene. This pair was used as an internal control when allele specific PCR was performed to screen for ß-globin gene mutations situated at the 3' end of the gene.

3.4.3. Direct Genomic Sequencing

Direct genomic sequencing of amplified DNA fragments by the dideoxynucleotide chain termination method was performed to identify the uncharacterised ß-thalassaemia mutations (section 2.7). Both double-stranded and single-stranded PCR products were used as sequencing templates.

3.4.3.1 Sequencing of double-stranded templates

A 916 base pair fragment which included the 5' untranslated region, the first and second exons and the first intron of the ß-globin gene was amplified using 0.5 µg of genomic DNA and 2 units of *Taq* polymerase in a 100 µl reaction volume containing 20 pmol of each of the primers P1 and P2 (table 3.5), 800 µM total dNTPs, 10 mM Tris HCl pH 8.3, 50 mM KCl and 1.5 mM $MgCl_2$. The thermal cycling regimen consisted of 30 cycles of denaturation at 93°C for 1 min, annealing at 55°C for 1 min and extension at 72°C

for 1.5 min; in the last cycle extension was prolonged to 3 min. Following amplification, the PCR products of four separate reactions were pooled and the resulting 400 µl sample was purified for sequence analysis by electroelution into a high salt solution as described in section 2.7.1.1. An approximate quantification of the amount of isolated product was made by comparison with a known amount of DNA in a molecular weight marker and accordingly an aliquot was used as the sequencing template using an internal sequencing primer S1 or S2 (table 3.5).

Sequencing reactions were performed using the reagents provided in the Sequenase sequencing kit of USB Corp. as described in section 2.7.2. Following the sequencing reaction, samples were subjected to electrophoresis in a 6% polyacrylamide gel in which 8M urea was incorporated. After electrophoresis, the gel was dried down and set up for overnight autoradiography.

The results of sequencing double-stranded amplified DNA were not consistently satisfactory. The short, linear double-stranded template can rapidly reanneal during the sequencing reaction, blocking or displacing the primer from the template strand, thereby reducing the amount of specific termination products formed and the reaction efficiency. To overcome these problems single-stranded templates were prepared.

3.4.3.2 Sequencing of single-stranded templates

Single-stranded DNA templates were generated by a modification of the asymmetric PCR method in which unequal molar amounts of the two amplification primers are used in order to produce an excess of single-stranded DNA of a chosen strand (Gyllensten and Erlich, 1988). A 2.1 kb fragment containing the entire ß-globin gene along with its 5' and 3' flanking sequences was amplified from approximately 0.5 µg of genomic DNA with 20 pmol of each primer P3 and P4 (table 3.5), other reaction conditions were similar to those described above. The thermal cycling regimen was altered to enable amplification of the large 2.1 kb product; 30 cycles of denaturation at 94°C for 2 min, annealing at 60°C for 2 min and extension at 72°C for 5 min was performed. In the first cycle denaturation was prolonged for 5 min and in the last cycle extension for 10 min. This amplified fragment was purified by

Table 3.5 Oligonucleotide primers used for amplification and sequencing of the β-globin gene 79

Name	Position of 5' nucleotide	Sequence 5'-3'
P1	307 nt upstream from the cap site	CGATCTTCAATATGCTTAC
P2	IVS-2 nt 114	CATTCGTCTGTTTCCCATTCTA
P3	307 nt upstream from the cap site	CGATCTTCAATATGCTTACCAAG
P4	320 nt downstream from the termination site	GCATAGGCATCAGGGCTGTTG
S1	codon 22, exon 1	TTCATCCACGTTCACCTT
S2	codon 46, exon 2	CAAAGGACTCAAAGAACCTC
S3	IVS-2 nt 114	CATTCGTCTGTTTCCCATTCTA
S4	IVS-2 nt 293	CTTAATGTATCTCAGAGATA
S5	IVS-2 nt 784	CCAGCCTTATCCCAACCAT
S6	67 nt downstream from the termination site	CAGTTTAGTAGTTGGACTTA
S7	279 nt downstream from the termination site	GCAGCCTCACCTTCTTTCATGG

PCR primers P1 and P3 are complementary to the noncoding strand while P2, P4 and sequencing primers S1-S7 are complementary to the coding strand.

electrophoresis onto a DEAE membrane (section 2.7.1.2). Approximately 5 ng of this purified PCR product was used as the template for a second PCR, in which all reaction conditions were the same except 40 pmol of only primer P3 was used in order to generate a single-stranded product of the coding strand. This single-stranded product was identified in an ethidium bromide stained 1% agarose gel as a faster moving and less intensely stained fragment in comparison to the double-stranded template. The single-stranded fragment was isolated and purified in the same manner as the product of the first PCR and resuspended in 21 µl of distilled water for use as a sequencing template. It was possible to sequence this 2.1 kb fragment with a series of seven internal oligonucleotide primers which are listed in table 3.5. The sequencing reaction and gel electrophoresis was performed as described for the double-stranded template with the exception of the initial denaturing step, which was unnecessary with the single-stranded template. Template-primer annealing was directly performed at 65°C for 2 min, which was followed by cooling on ice for 2 min.

As the DNA samples sequenced were heterozygous for ß-thalassaemia both normal and mutant alleles were equally amplified by PCR. Subsequent extension from the sequencing primer occurred on products from both alleles resulting in a superimposition of sequencing ladders derived from each allele. This became apparent when heterozygosity at a nucleotide residue occurred, because two extension products of equal intensity were produced in different lanes of the sequencing autoradiograph. Thus a substitution appeared as two bands at the same position in a sequencing ladder and a frameshift mutation resulted in two sequencing ladders diverging from the common ladder at the site of the frameshift mutation.

3.4.4. Allele specific PCR to screen for mutations identified by sequencing

Once a mutation was identified by sequencing, an allele specific primer for each mutation was synthesised as shown below.

Table 3.6 Allele specific primers for the detection of β-thalassaemia mutations identified by sequencing

Mutation	Oligonucleotide sequence 5'-3'	Product size
codon 5 (-CT)	ACAGGGCAGTAACGGCAGACTTCTCCGCGA	205 bp
IVS-1 minus 1 (G-C)	TAAACCTGTCTTGTAACCTTGATACCTACG	280 bp
IVS-2 nt 837 (T-G)	CCTTTTGCTAATCATGTTCATACCTCGTAG	646 bp
IVS-2 nt 1 (G-A)	AAGAAAACATCAAGGGTCCCATAGACTGAT	635 bp
IVS-1 nt 110 (G-A)	ACCAGCAGCCTAAGGGTGGGAAAATACACT	419 bp
IVS-1 minus 1 (G-A)	TAAACCTGTCTTGTAACCTTGATACCTACT	280 bp
codon 88 (+T)	GTGCAGCTTGTCACAGTGCAGCTCACACAA	613 bp

These are all mutant allele specific primers with their mismatches underlined, and are complementary to the coding strand thus coupled with primer A; the single exception being the allele specific primer for IVS-2 nt 837 (T-G) which is complementary to the noncoding strand and coupled with downstream primer C and used with internal control primers F and G.

Each of the allele specific primers was tested at the uniform PCR conditions previously described, using the sequenced DNA sample as the positive control. Their efficacy was ascertained by observation of the expected size fragment only with the positive control and not with the negative control, the internal control amplified fragment being present in each reaction. All the allele specific primers worked satisfactorily at the uniform reaction conditions. They were then used to perform family studies for each index case to look for Mendelian inheritance of the mutation and also to screen the remaining uncharacterised individuals for the presence of that mutation.

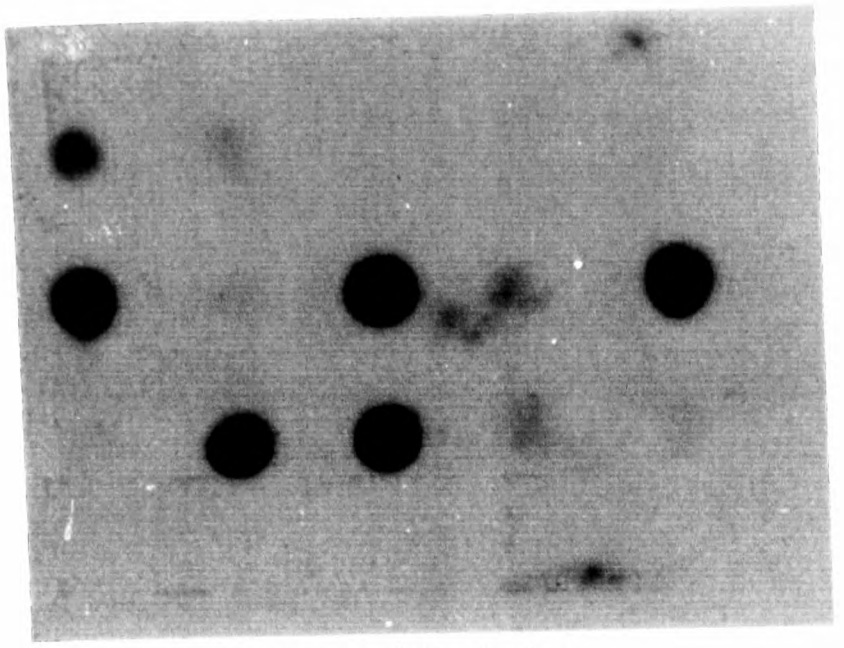
3.5. Results

3.5.1 Comparison of dot-blot hybridisation and allele specific PCR

After the exclusion of the 619 bp deletion, 200 DNA samples were screened by dot-blot hybridisation for the following ß-thalassaemia mutations: IVS-1 nucleotide 5 (G-T), codons 8/9 (+G), IVS-1 nucleotide 1 (G-T) and codons 41/42 (-CTTT). Figure 3.4 illustrates dot-blot filters hybridised with the IVS-1 nucleotide 5 (G-T) and codons 8/9 (+G) radiolabelled ASO probes. The results of dot-blot hybridisation with different probes were not uniformly satisfactory and clear positive signals, indicating the presence of a mutation, and clear negative signals, indicating its absence, with appropriate hybridisation of the controls were not obtained with all filters and probes. There were thus a number of samples with ambiguous grey signals from which a diagnosis of the genotype could not be made. All these 200 samples were re-analysed by allele specific PCR. Figure 3.5 shows the identification of the homozygous and normal state for the codons 8/9 (+G) mutation. Here the absence or presence of the amplified fragment using the mutant and normal allele specific primers can be clearly seen, enabling accurate genotype diagnosis. All the four mutations, whose analysis was being attempted by dot-blot hybridisation, could be satisfactorily studied by allele specific PCR.

The accuracy of gene diagnosis achieved by dot-blot hybridisation in the 200 heterozygotes analysed was assessed by comparison with the results obtained by allele specific PCR (table 3.7). Following screening by allele specific PCR for the IVS-1 nucleotide 5 (G-T), codons 8/9 (+G), IVS-1 nucleotide 1 (G-T) and codons 41/42 (-CTTT) mutations, gene diagnosis was achieved in 178 (89%) of the 200 carriers analysed; in 167 (83.5%) cases the dot-blot hybridisation results were confirmed, in 30 (16.5%) cases the results failed to be confirmed. Analysing the dot-blot hybridisation results obtained with each of the ASOs, it was found that the accuracy of dot-blot hybridisation was not uniform and varied with different probes. The IVS-1 nucleotide 5 (G-T) probe was found to be most accurate with 93% of cases being correctly diagnosed and three false positive results. The accuracy of the

IVS-1 nucleotide 5 (G-C)



codons 8/9 (+G)

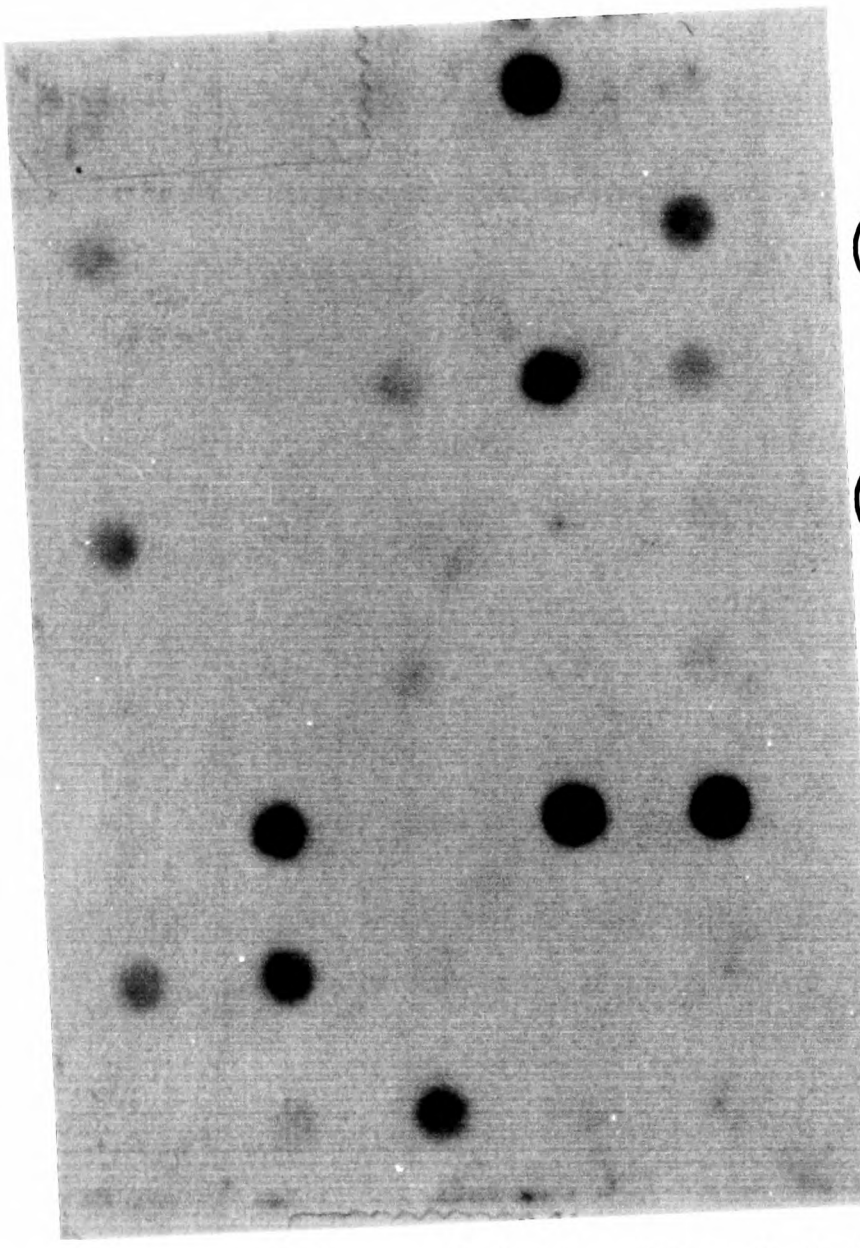


Figure 3.4 Dot-blot hybridisation of amplified DNA from ß-thalassaemia carriers with the radiolabelled IVS-1 nucleotide 5 (G-C) and codons 8/9 ASO probes. + and - represent the positive and negative controls. Numbered circles represent samples carrying the mutation and those marked with ? are ambiguous not permitting accurate genotype diagnosis.

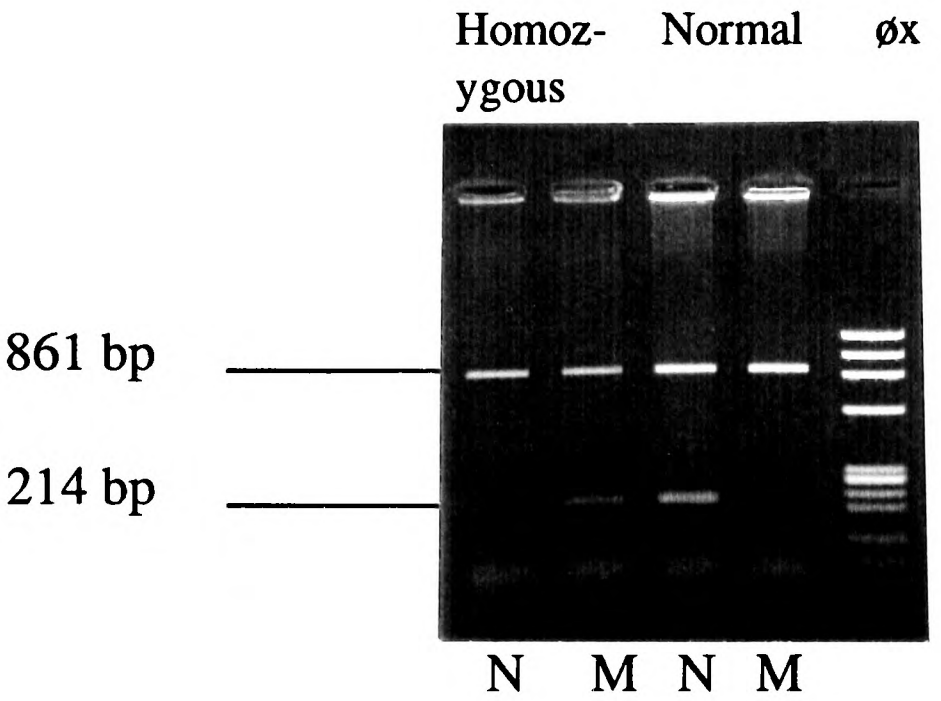


Figure 3.5 Determination of the homozygous and normal state for codons 8/9 (+G) mutation by allele specific PCR. Ethidium bromide stained gel electrophoresis of the PCR products using the normal (N) and mutant (M) allele specific primers with homozygous and normal DNA samples respectively. The 214 bp fragment is the product of the allele specific primers and the 861 bp fragment is the internal control. The marker is øx 174 DNA digested with *Hae* III (øx).

three other probes varied from 65% to 78%. Twenty eight individuals remained uncharacterised following screening by dot-blot hybridisation, of these six individuals had their mutations determined by the allele specific PCR screening, and the remaining were later found to carry one of the rare ß-thalassaemia mutations.

Table 3.7 Comparison of dot-blot hybridisation results with those of allele specific PCR

Mutation	IVS-1 nt 5 (G-C)	c 8/9 (+G)	IVS-1 nt 1 (G-T)	c 41/ 42 (-CTTT)	Unkno-wn	Total
No. of dot-blot hybridisation results confirmed by allele specific PCR	62 (93%)	51 (78%)	17 (73%)	15 (65%)	22	167
False negative results	5	14	6	8	--	33
False positive results	3	9	7	8	6	33
No. allele specific PCR results	67	65	23	23	22	200

(c=codon)

Following these experiences with dot-blot hybridisation and allele specific PCR, it was decided that of these two screening methods the latter was clearly preferable. Therefore allele specific PCR was used to screen for various ß-thalassaemia mutations in the 656 subjects that were analysed.

3.5.2 Spectrum of ß-thalassaemia mutations on the Indian subcontinent

ß-Thalassaemia was found to be heterogeneous at the molecular level in the Asian Indian population with a total of 16 different ß-thalassaemia mutations identified in the 656 carriers studied. Their incidence and regional distribution maybe conveniently presented by grouping the results according to the strategy followed for their detection.

3.5.2.1 Common ß-thalassaemia mutations

On screening for the IVS-1 nucleotide 5 (G-C), 619 bp deletion, IVS-1 nucleotide 1 (G-T), codons 8/9 (+G) and codons 41/42 (-CTTT) mutations, which are the commonest mutations described in previous studies on the immigrant population, it was possible to characterise the ß-thalassaemia mutations in 602 of the 656 (91.7%) carriers studied. Figure 3.6 shows screening for the IVS-1 nucleotide 5 (G-C) mutation. The frequency and regional distribution of the common ß-thalassaemia mutations is shown in Table 3.8. The IVS-1 position 5 (G-C) mutation was found to be the most frequent and present in 226 individuals (34.4%). Noteworthy regional differences in the frequencies of the five common ß-thalassaemia mutations were found. The IVS-1 nucleotide 5 (G-C) mutation was observed in carriers from each of the seven regions studied. However, its frequency varied greatly from 71% of all prevalent mutations in Tamil Nadu to 12% in Sindh; it was found to be more frequent in the Southeastern regions of the Indian subcontinent as compared to the Northwestern regions. The 619 bp deletion was observed to have a relatively restricted distribution being present in only the neighbouring regions of Sindh, Gujarat and Punjab; its frequency being highest in Sindh (54%) where it is by far the commonest ß-thalassaemia mutation. Of the 113 individuals with the codons 8/9 (+G) mutation, 73 (65%) originated from Northwest Pakistan, where at a frequency of 44% it was the commonest mutation present. The codons 8/9 (+G) mutation is present at a higher frequency in the Northern regions, its frequency declining as a transition is made from Northern to Southern regions of the subcontinent. The IVS-1 nucleotide 1 (G-T)

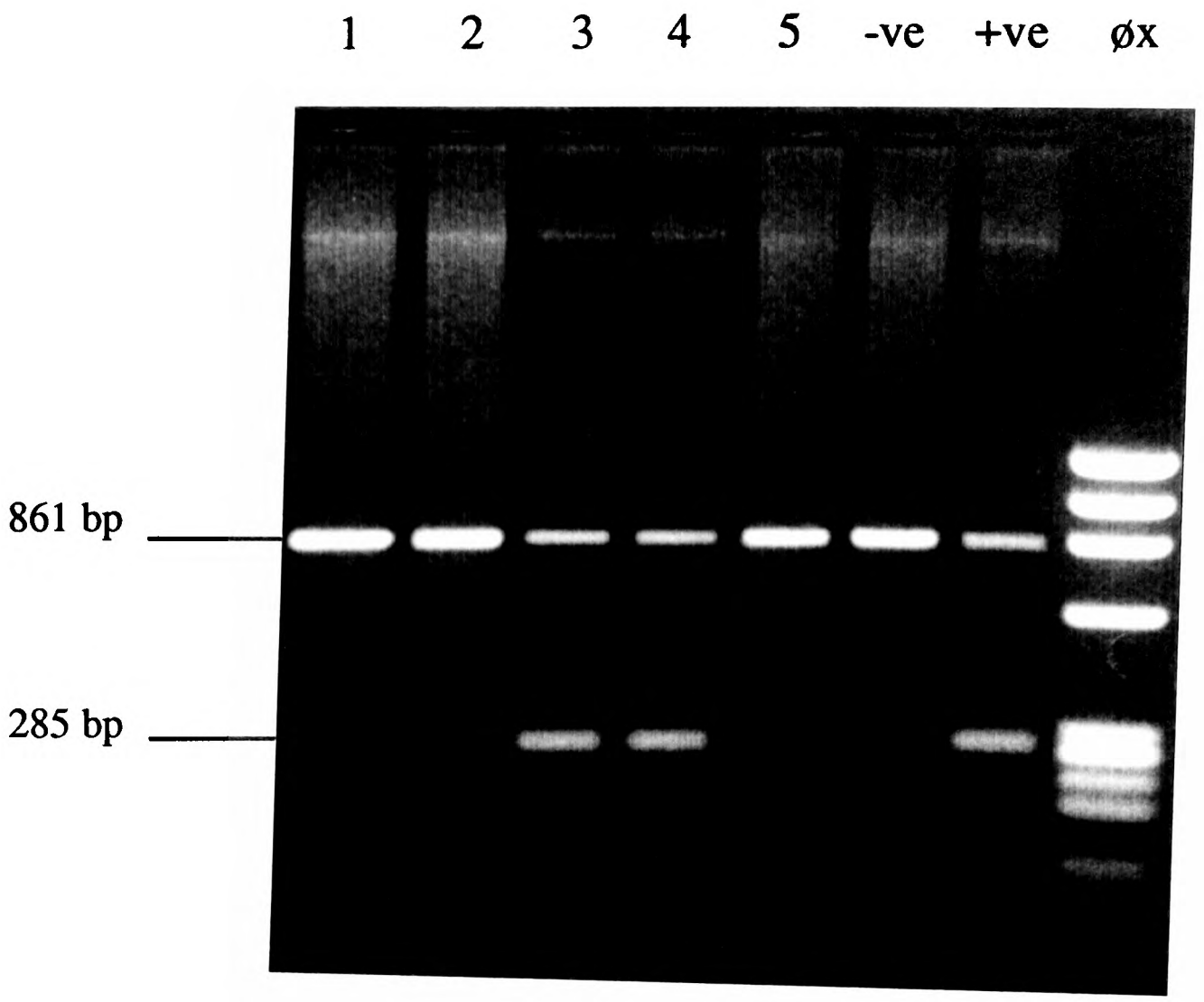


Figure 3.6 Screening for the IVS-1 nucleotide 5 (G-C) mutation. Ethidium bromide stained gel electrophoresis of the PCR products using the mutant allele specific primer: samples in lanes 3, 4 and positive control carry the mutation as indicated by the 285 bp band, the other samples and the negative control do not carry the mutation. The 861 bp internal control product indicates the efficacy of the reaction itself. The marker is øx 174 DNA digested with *Hae* III (øx).

Table 3.8 The frequency and regional distribution of the common Asian Indian β-thalassaemia mutations

REGION MUTATION	Gujarat	Northwest Pakistan	Punjab	Sindh	Maharash-tra	Bengal & Bangladesh	Tamil Nadu	Total
IVS-1 nt 5 (G-C)	81 (41%)	59 (35%)	54 (38%)	14 (12%)	7 (54%)	6 (60%)	5 (71%)	226 (34.4%)
619 bp deletion	51 (26%)	--	22 (16%)	61 (54%)	--	--	--	134 (20.4%)
codons 8/9 (+G)	8 (4%)	73 (44%)	19 (13%)	12 (11%)	--	1 (10%)	--	113 (17.2%)
IVS-1 nt 1 (G-T)	33 (17%)	2 (1%)	15 (11%)	20 (18%)	--	--	--	70 (10.6%)
codons 41/42 (-CTTT)	17 (9%)	17 (10%)	20 (14%)	5 (4%)	1 (8%)	2 (20%)	--	62 (9.5%)

Figures in brackets represent the percentage of ß-thalassaemia mutations in that region accounted for by that particular mutation.

mutation was almost entirely restricted to the neighbouring states of Sindh, Gujarat and Punjab, where it was observed at similar frequencies. The codons 41/42 (-CTTT) mutation was found to be present in all regions except Tamil Nadu. Figure 3.7 summarises the regional distribution of the common ß-thalassaemia mutations on the Indian subcontinent.

3.5.2.2. Rare Asian Indian ß-thalassaemia mutations

By screening for the five rare ß-thalassaemia mutations previously described in the immigrant Asian Indian population: codon 15 (G-A), codon 16 (-C), -88 (C-T), cap site position +1 (A-C) and the 25 bp deletion from the 3' end of IVS-1, it was possible to characterise the ß-thalassaemia mutations in 24 of the 656 (3.7%) individuals studied. Figure 3.8 shows a screening experiment for the codon 16 (-C) mutation. Codon 15 (G-A) was the commonest mutation, identified in 16/656 (2%) individuals, of these 11/16 (69%) were from Northwest Pakistan. All four carriers of the codon 16 (-C) mutation originated from Punjab as did the three carriers of the cap site position +1 (A-C) mutation. The -88 (C-T) mutation was detected in a single individual from Punjab and the IVS-1 3' end -25 bp mutation was not found in this study. The regional distribution and frequencies of each of these mutations is shown in Table 3.9.

3.5.2.3 Mutations identified by direct DNA sequencing

Seven different ß-thalassaemia mutations were identified by DNA sequencing of the ß-globin genes of the carriers that remained uncharacterised following the screening experiments; of these two have not been described before and five have been described in other populations in which ß-thalassaemia is prevalent. The five ß-thalassaemia mutations identified by DNA sequencing which have been previously described in other populations were codon 5 (-CT), IVS-1 minus 1 (G-C), IVS-2 nucleotide 1 (G-A), IVS-1 nucleotide 110 (G-A) and IVS-1 minus 1 (G-A).

The codon 5 (-CT) mutation has been described in Mediterranean patients with transfusion dependent thalassaemia (Kollia et al, 1989). The absence of the CT dinucleotide from codon 5 results in

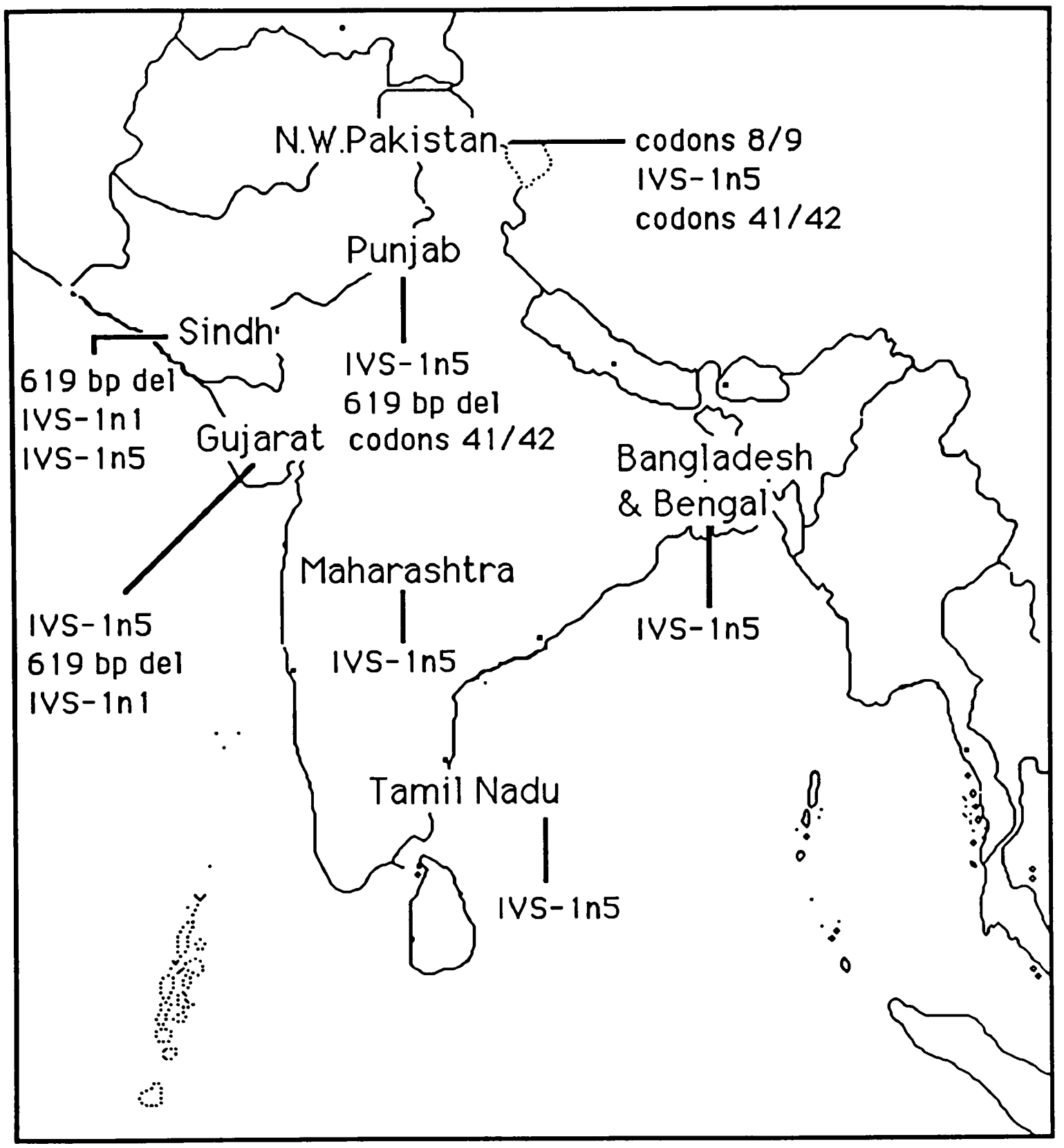


Figure 3.7 Map of the Indian subcontinent showing the distribution of the common Asian Indian ß-thalassaemia mutations. (n = nucleotide)

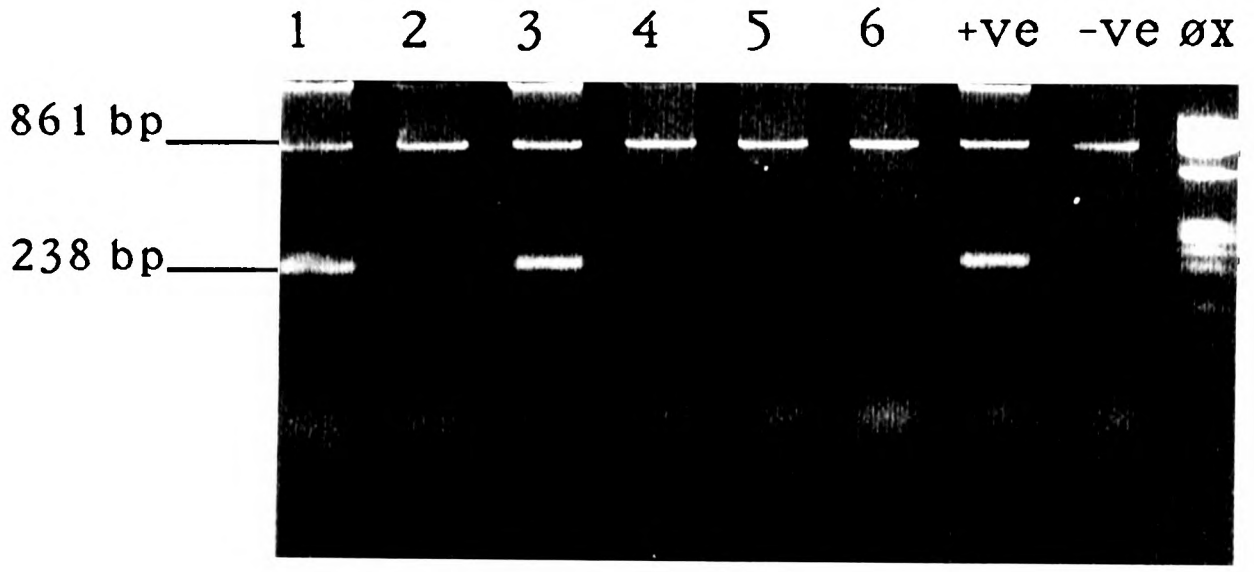


Figure 3.8 Screening for the codon 16 (-C) mutation by allele specific PCR. Ethidium bromide stained gel electrophoresis of the PCR products using the mutant allele specific primer: samples in lanes 1, 3 and positive control carry the mutation as indicated by the 238 bp band, the other samples and the negative control do not carry the mutation. The 861 bp internal control product indicates the efficacy of the reaction itself. The marker is øx 174 DNA digested with *Hae* III (øx).

Table 3.9 The frequency and regional distribution of the rare Asian Indian β-thalassaemia mutations

REGION MUTATION	Gujarat	Northwest Pakistan	Punjab	Sindh	Maharas-htra	Bengal & Bangladesh	Tamil Nadu	Total
codon 15 (G-A)	3 (1.5%)	11 (7%)	1 (0.7%)	--	--	1 (10%)	--	16 (2.3%)
codon 5 (-CT)	4 (2%)	4 (2%)	1 (0.7%)	1 (0.9%)	--	--	--	10 (1.4%)
IVS-1 minus 1 (G-C)	4 (2%)	--	1 (0.7%)	1 (0.9%)	--	--	--	6 (0.9%)
codon 16 (-C)	--	--	4 (3%)	--	--	--	--	4 (0.6%)
IVS-2 nt 837 (T-G)	--	--	--	--	2 (15%)	--	2 (29%)	4 (0.6%)
cap site +1 (A-C)	--	--	3 (2%)	--	--	--	--	3 (0.4%)
IVS-2 nt 1 (G-A)	--	--	1 (0.7%)	--	1 (7.5%)	--	--	2 (0.3%)
IVS-1 nt 110 (G-A)	--	--	1 (0.7%)	--	1 (7.5%)	--	--	2 (0.3%)
IVS-1 minus 1 (G-A)	--	1 (0.6%)	--	--	1 (7.5%)	--	--	2 (0.3%)
-88 (C-T)	--	--	1 (0.7%)	--	--	--	--	1 (0.1%)
codon 88 (+T)	1 (0.5%)	--	--	--	--	--	--	1 (0.1%)

Figures in brackets represent the percentage of ß-thalassaemia mutations in that region accounted for by that particular mutation.

a shift in the reading frame with a premature termination signal at the new codon 21 causing complete inhibition of ß-globin chain synthesis (figure 3.9). This mutation was identified in 10 (1.5%) unrelated carriers originating from the Northern regions of Gujarat, Northwest Pakistan, Punjab and Sindh.

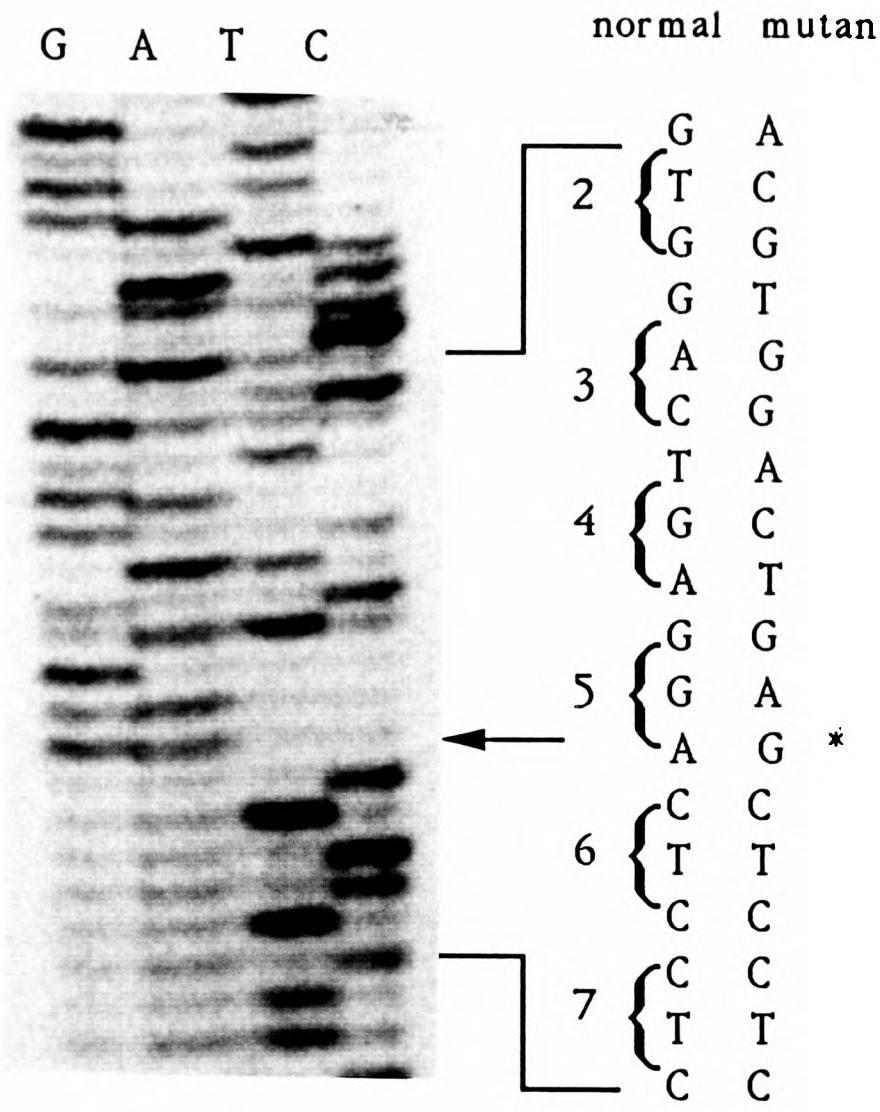
The IVS-1 minus 1 (G-C) and IVS-1 minus 1 (G-A) mutations, being two different substitutions at the same site, have identical mechanisms of action (figure 3.10). Both result in changes in the consensus sequence of the donor site of the first intron, thereby interfering with efficient splicing and have been described in individuals from Tunisia and Bulgaria, respectively; whether they result in a $ß^{\circ}$ or $ß^{+}$ phenotype has not been determined (Kazazian, 1990). The IVS-1 minus 1 (G-C) mutation was identified in six (0.9%) carriers, four of whom originated from Gujarat and one each from neighbouring Punjab and Sindh; the IVS-1 minus 1 (G-A) mutation was detected in two (0.3%) individuals from Northwest Pakistan and Maharashtra. As all the individuals found to carry these two consensus change mutations were heterozygous it was not possible to determine whether they would cause a $ß^{\circ}$ or $ß^{+}$ phenotype.

The G-A transition at IVS-2 nucleotide 1 site destroys the GT dinucleotide invariant at the 5' splice site, inactivating it to result in abnormal RNA splicing and $ß^{\circ}$-thalassaemia (Treisman et al, 1982); this mutation has been identified in Mediterraneans (Kazazian et al, 1984a) and in an American black (Wong et al, 1986), it was found in two (0.3%) individuals from Punjab and Maharashtra.

The G-A substitution at IVS-1 position 110 creates a sequence that is homologous to that at the authentic 3' splice junction at six out of seven bases resulting in an anomalous splicing signal, that causes abnormal splicing of some ß-globin mRNA precursor molecules at this site and $ß^{+}$-thalassaemia (Spritz et al, 1981). The IVS-1 position 110 (G-A) mutation is widely prevalent in the Mediterranean basin (Huisman, 1990) and is being described for this first time outside of this area in 2 (0.3%) individuals originating from Punjab and Maharashtra.

IVS-2 nucleotide 837 (T-G) and codon 88 (+T) were the two new mutations identified by DNA sequencing. The T-G substitution at

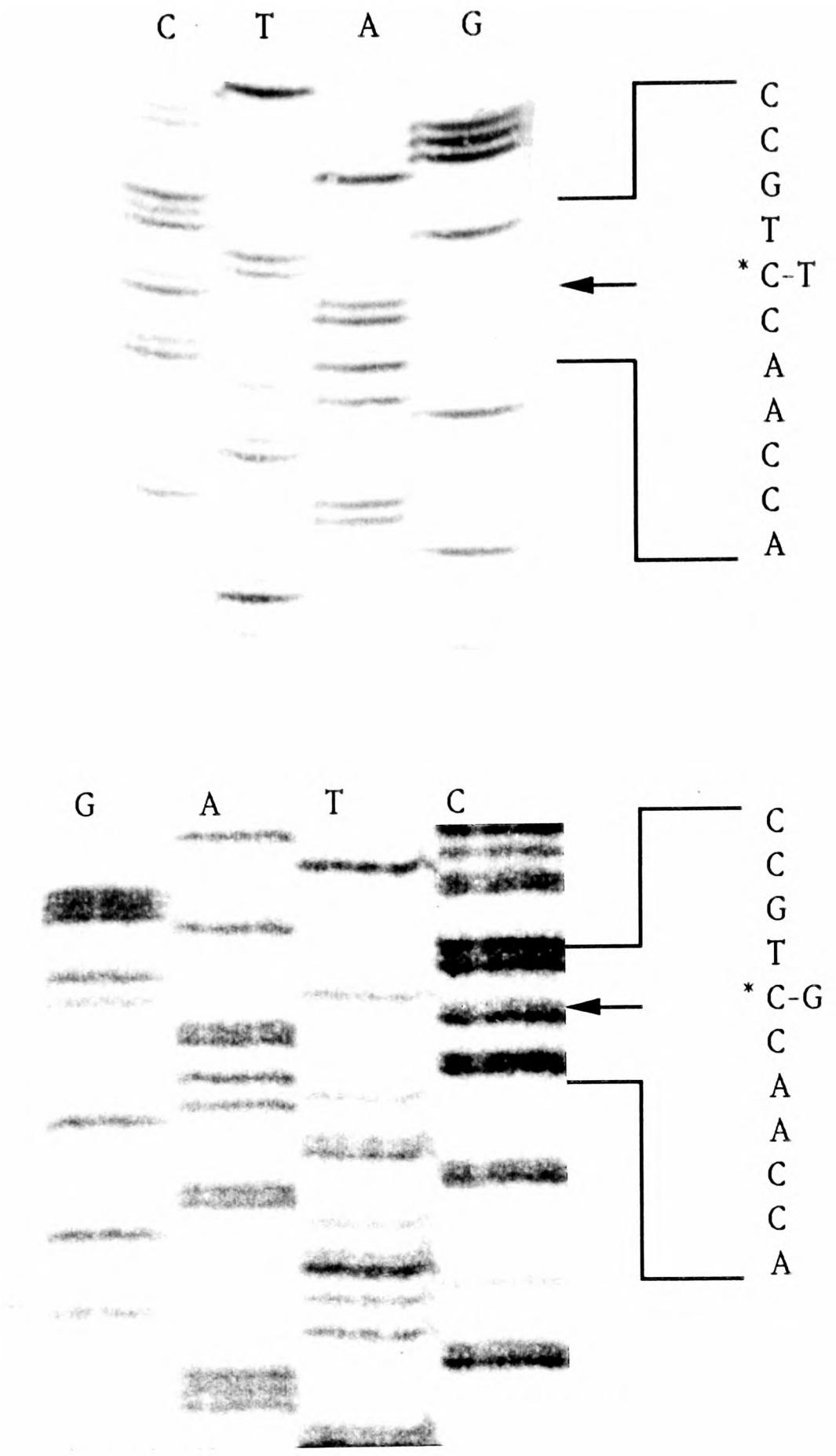
noncoding strand



coding strand

codons	4	5	6	7	8----	20	21
normal	ACT	C<u>CT</u>	GAG	GAG	AAG	GTG	GAT
mutant	ACT	CGA	GGA	GAA	GTC	GGA	(stop)

Figure 3.9 Sequence analysis of the ß-globin gene in an individual heterozygous for the codon 5 (-CT) mutation. The noncoding strand can be read from the autoradiograph where the position of the deletion and the resulting frameshift is marked with an arrow and *. The coding strand in which the -CT deletion is underlined is shown in the lower half. The frameshift mutation results in a premature stop signal TGA at codon 21.



coding strand- IVS-1 splice junction
C29 C30 IVS-1

GGG CAG GTTGGTT
▼
A/C

Figure 3.10 Sequence analysis of the ß-globin gene in heterozygotes. Both individuals have a substitution at IVS-1 position minus 1; G-A and G-C, respectively. It is the noncoding strand that is seen on the autoradiograph where the mutation is marked by an arrow and *. The coding strand sequence is shown in the lower half.

nucleotide 837 of the second intron of the ß-globin gene results in an AG dinucleotide, 14 nucleotides upstream from the invariant AG dinucleotide at the normal acceptor splice site, thus creating a sequence (AGCTTC) homologous to the authentic 3' splice junction of the second intron (AGCTCC) (figure 3.11). Possibly this results in a functional alternative acceptor splice site which causes abnormal splicing of ß-globin mRNA. This mutation was identified in 4 (0.6%) individuals all of whom originated from the Southern states of Tamil Nadu and Maharashtra. As all the individuals in whom the IVS-2 nucleotide 837 (T-G) mutation was identified were heterozygous for ß-thalassaemia, it is not possible to comment whether this mutation would cause $ß^{\circ}$ or $ß^{+}$ thalassaemia.

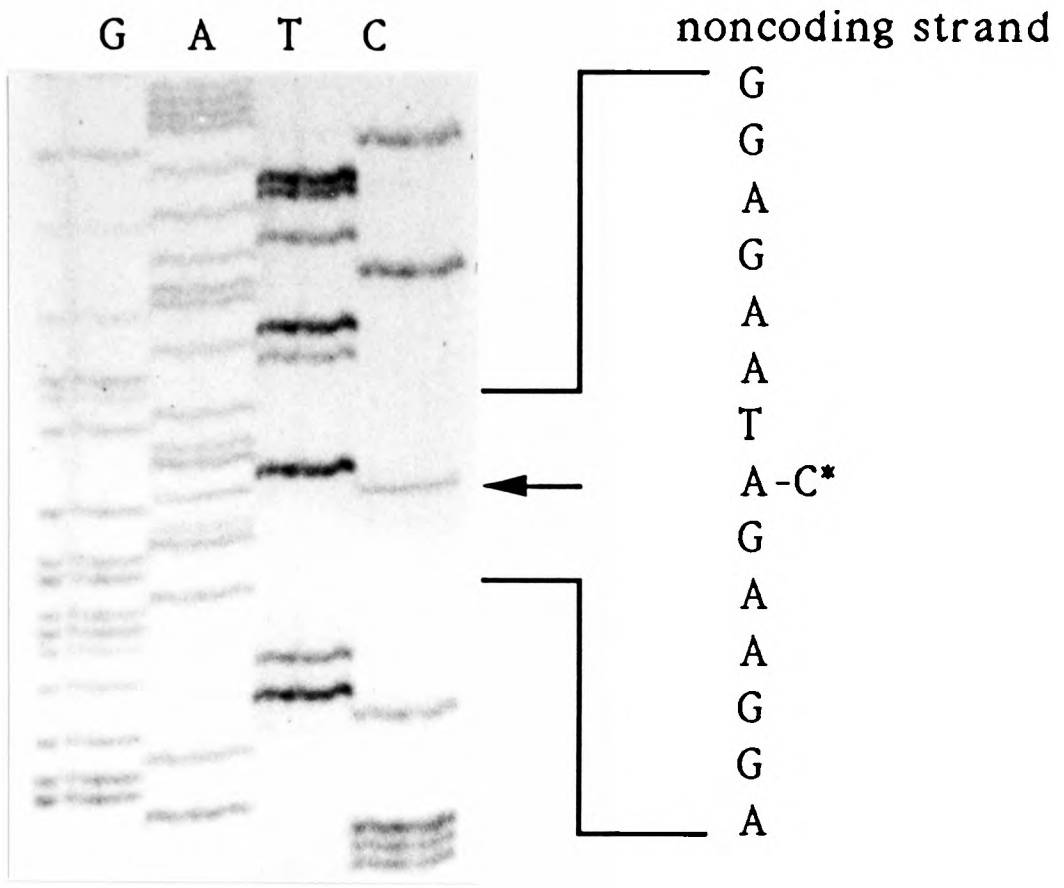
The second new mutation identified in a single Gujarati individual was a single nucleotide insertion (+T) in codon 88 which creates a shift in the reading frame leading to a premature stop codon, TGA, at codon 90 (figure 3.12). This chain termination codon in the middle of the transcribed mRNA molecule would result in premature termination of translation with the production of a shortened and physiologically useless peptide fragment and a clinical phenotype of $ß^{\circ}$-thalassaemia.

Both the ß-globin genes in which these two new mutations were identified were completely sequenced to rule out the presence of any other deviation from the known ß-globin gene sequence and polymorphisms expected in this population. Family studies of the index cases with allele specific primers for each of these two newly described mutations demonstrated their Mendelian inheritance (figure 3.13).

Once a ß-thalassaemia mutation was identified by DNA sequencing, allele specific primers were used to screen the remaining uncharacterised individuals enabling the detection of other individuals with that particular mutation (figure 3.14).

3.5.3 Regional distribution of ß-thalassaemia mutations on the Indian subcontinent

The ß-thalassaemia mutations identified in each of the seven regions studied are shown in table 3.10. Of the ß-thalassaemia mutations present in each of these regions, 4-6 mutations accounted for over 90% of ß-thalassaemia alleles in that region. As



coding strand

IVS-2 nt 836-841	A T C T T C	normal
	A G C T T C	mutant
normal splice junction	A G C T C C	

Figure 3.11 Sequence analysis of the ß-globin gene in an individual heterozygous for the IVS-2 nt 837 (T-G) mutation. It is the noncoding strand that can be read from the autoradiograph where the mutation is marked by an arrow and *. The coding strand sequence is compared with that of the IVS-2 normal splice junction in the lower half of the figure.

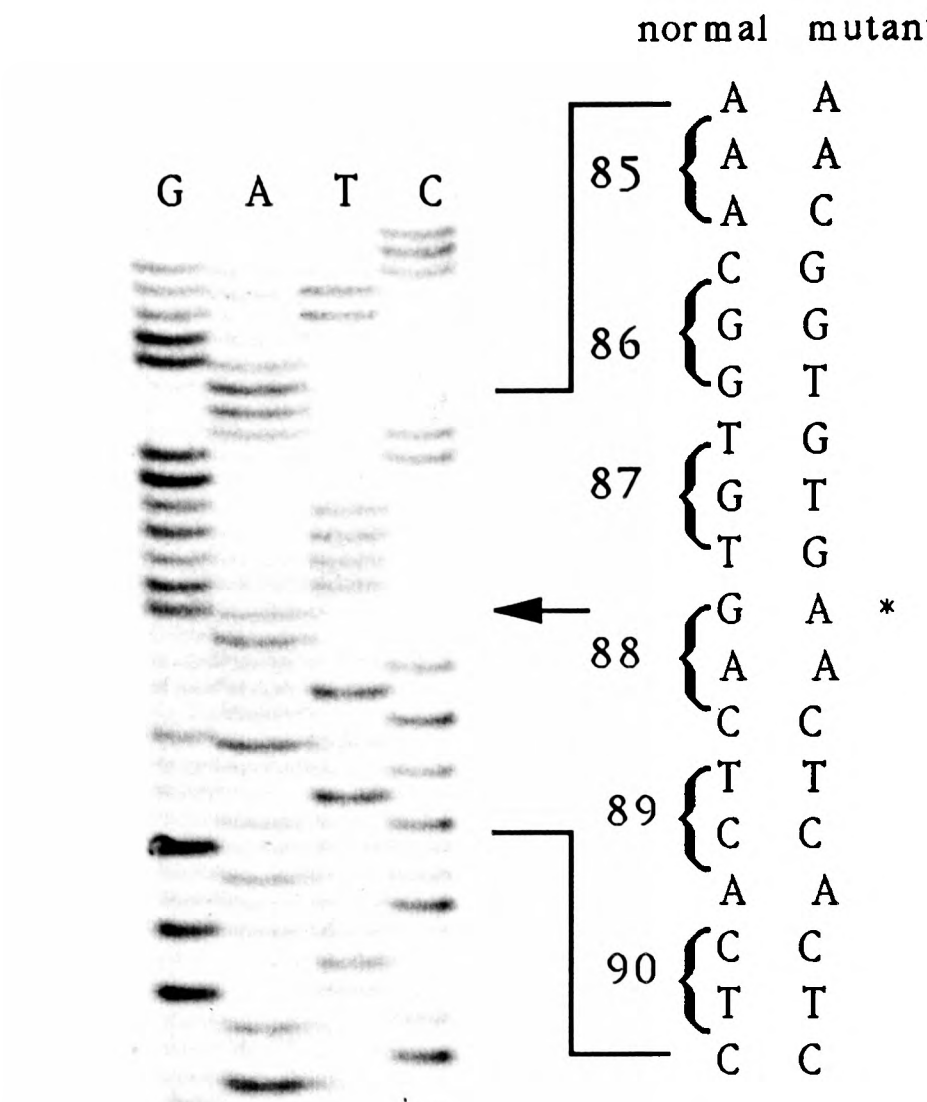
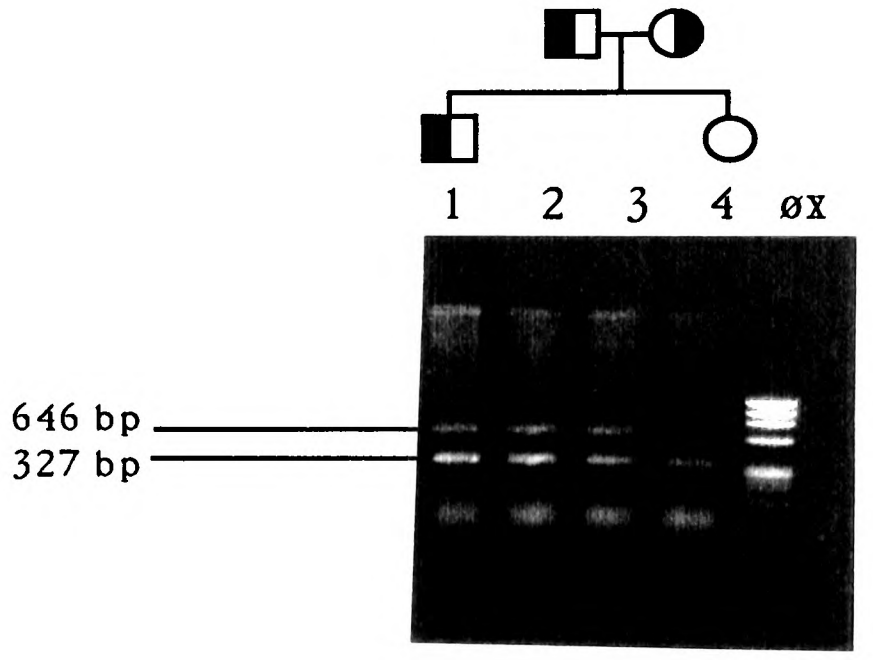


Figure 3.12 Sequence analysis of the ß-globin gene in an individual heterozygous for the codon 88 (+T) mutation. The noncoding strand can be read from the autoradiograph where the position of the insertion and the resulting frameshift is marked with an arrow and *. The coding strand showing the +T insertion and resulting in a premature stop codon is depicted in the lower half.

IVS-2 nucleotide 837 (T-G)



codon 88 (+T)

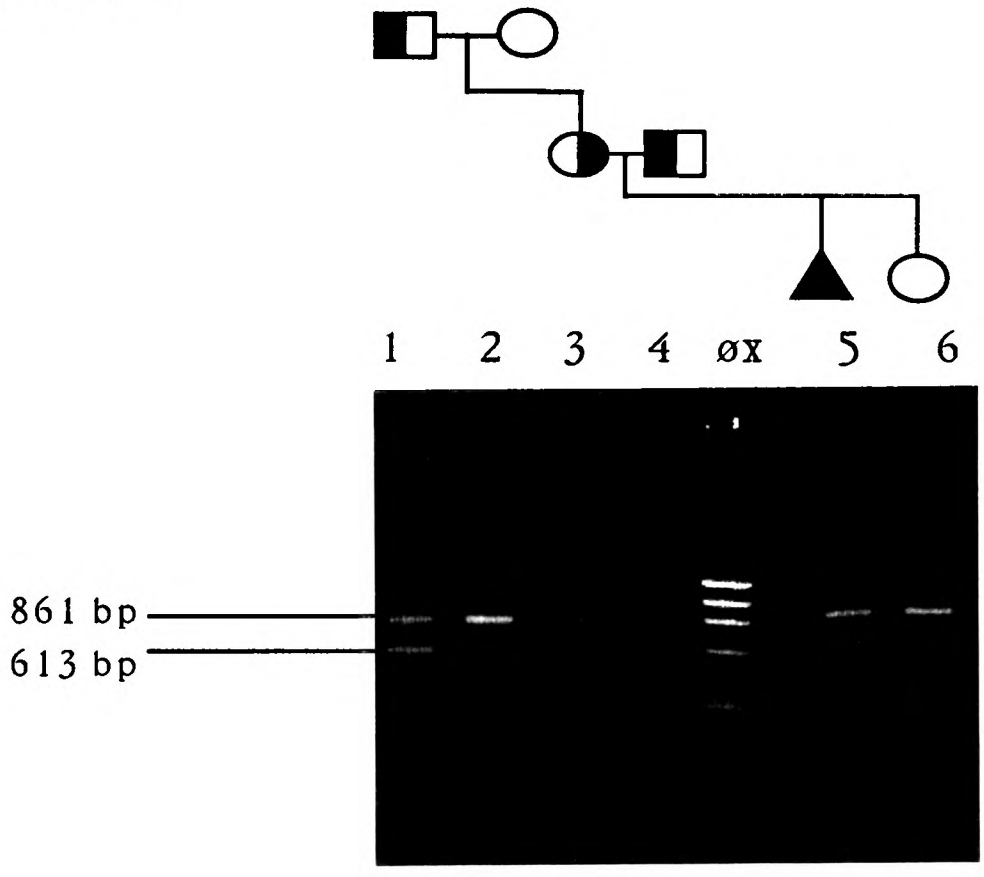


Figure 3.13 Family studies for the newly identified ß-thalassaemia mutations using allele specific PCR. Half shaded symbols represent ß-thalassaemia carriers, unshaded symbols represent normal individuals and the shaded symbol represents a thalassaemia major fetus diagnosed by RFLP analysis. The ethidium bromide stained agarose gels show electrophoresis of the PCR products with the respective mutant allele specific primers. The 646 bp and 613 bp band indicates the presence of the mutation and the 327 bp and 861 bp fragments are the internal controls in the top and bottom gels, respectively. øx 174 DNA digested with *Hae*III is the marker.

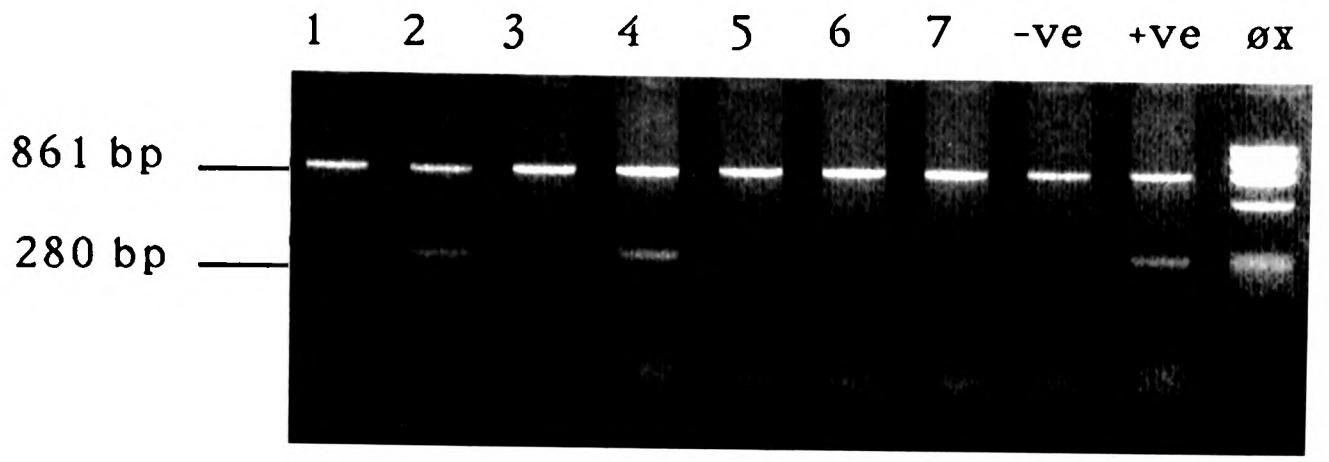


Figure 3.14 Screening for the IVS-1-1 (G-A) mutation identified by DNA sequencing using allele specific PCR. Ethidium bromide stained agarose gel electrophoresis of the PCR products with the mutant allele specific primer: visualisation of the 280 bp band in lanes 2, 4 and positive control indicates the presence of the mutation and the absence of this band in the negative control sample indicates the specificity of the reaction. The 861 bp internal control band proves the efficacy of the PCR itself and the marker is øx/174 DNA digested with *Hae*III.

Table 3.10 Spectrum of ß-thalassaemia mutations in different regions of the Indian subcontinent

Region	Spectrum of mutations
Gujarat	**I**-IVS-1 nt 5 (G-C), 619 bp del, IVS-1 nt 1 (G-T) & c 41/42 (-CTTT) **II**-c 8/9 (+G), c 5 (-CT), IVS-1-1 (G-C), c 15 (G-A) & c 88 (+T)
Northwest Pakistan	**I**-c 8/9 (+G), IVS-1 nt 5 (G-C), c 41/42 (-CTTT) & c 15 (G-A) **II**-c 5 (-CT) & IVS-1-1 (G-A)
Punjab	**I**-IVS-1 nt 5 (G-C), 619 bp del, c 41/42 (-CTTT), c 8/9 (+G), IVS-1 nt 1 (G-T) & c 16 (-C) **II**-cap site +1 (A-C), c 15 (G-A), c 5 (-CT), IVS-1-1 (G-C), IVS-2 nt 1 (G-A), IVS-1 nt 110 (G-A), IVS-1-1 (G-A) & -88 (C-T)
Sindh	**I**-619 bp del, IVS-1 nt 1 (G-T), IVS-1 nt 5 (G-C) & c 8/9 (+G) **II**-c 41/42 (-CTTT), c 5 (-CT) & IVS-1-1 (G-C)
Maharash-tra	**I**-IVS-1 nt 5 (G-C), IVS-2 nt 837 (T-G), c 8/9 (+G), IVS-2 nt 1 (G-A) & IVS-1 nt 110 (G-A) **II**-IVS-1-1 (G-A)
Bengal & Bangladesh	**I**-IVS-1 nt 5 (G-C), c 41/42 (-CTTT) & c 8/9 (+G) **II**-c 15 (G-A)
Tamil Nadu	**I**-IVS-1 nt 5 (G-C) & IVS-2 nt 837 (T-G)

I-represents the panel of ß-thalassaemia mutations which account for over 90% of mutations in that region and **II**-represents the panel of remaining mutations
(IVS-1-1= IVS-1 minus 1, del= deletion)

an adequate population sample was analysed from Northwest Pakistan, Gujarat, Punjab and Sindh it was possible to accurately comment on the spectrum of ß-thalassaemia mutations in these regions only. The disease was observed to be heterogeneous in all these regions; in Punjab as many as 14 different mutations were identified, nine in Gujarat, seven in Sindh and six in Northwest Pakistan. Figure 3.15 shows the proportional representation of the ß-thalassaemia mutations which account for around 90% of the alleles in each of these regions.

The small number of subjects studied from Maharashtra, Bengal and Bangladesh and Tamil Nadu make it difficult to comment on the distribution of ß-thalassaemia alleles in these regions. However, IVS-1 nucleotide 5 (G-C) is the commonest mutation in each of these regions accounting for 54%, 60% and 71% of their ß-thalassaemia alleles, respectively. On the basis of the regional differences in the distribution and frequency of ß-thalassaemia on the Indian subcontinent, three distinct areas have been defined: Northwest Pakistan, the Northwestern regions of Gujarat, Punjab and Sindh; and the Southeastern regions of Maharashtra, Bengal and Bangladesh and Tamil Nadu.

3.5.4 Summary of results

The 16 different ß-thalassaemia mutations which were identified in the population sample studied could be divided into three groups according to their frequencies (table 3.12). The first group comprising the common mutations which accounted for 91.7% of ß-thalassaemia alleles, the second group comprising the five mutations which together accounted for 6.1 % of all alleles and the last group comprising the six rare muations which accounted for 1.7% of alleles. Of the 311 couples that were studied, 186 (60%) carried identical mutations. This proportion was highest amongst couples originating from Northwest Pakistan where 70 of the 82 (85.3%) couples studied had identical mutations.

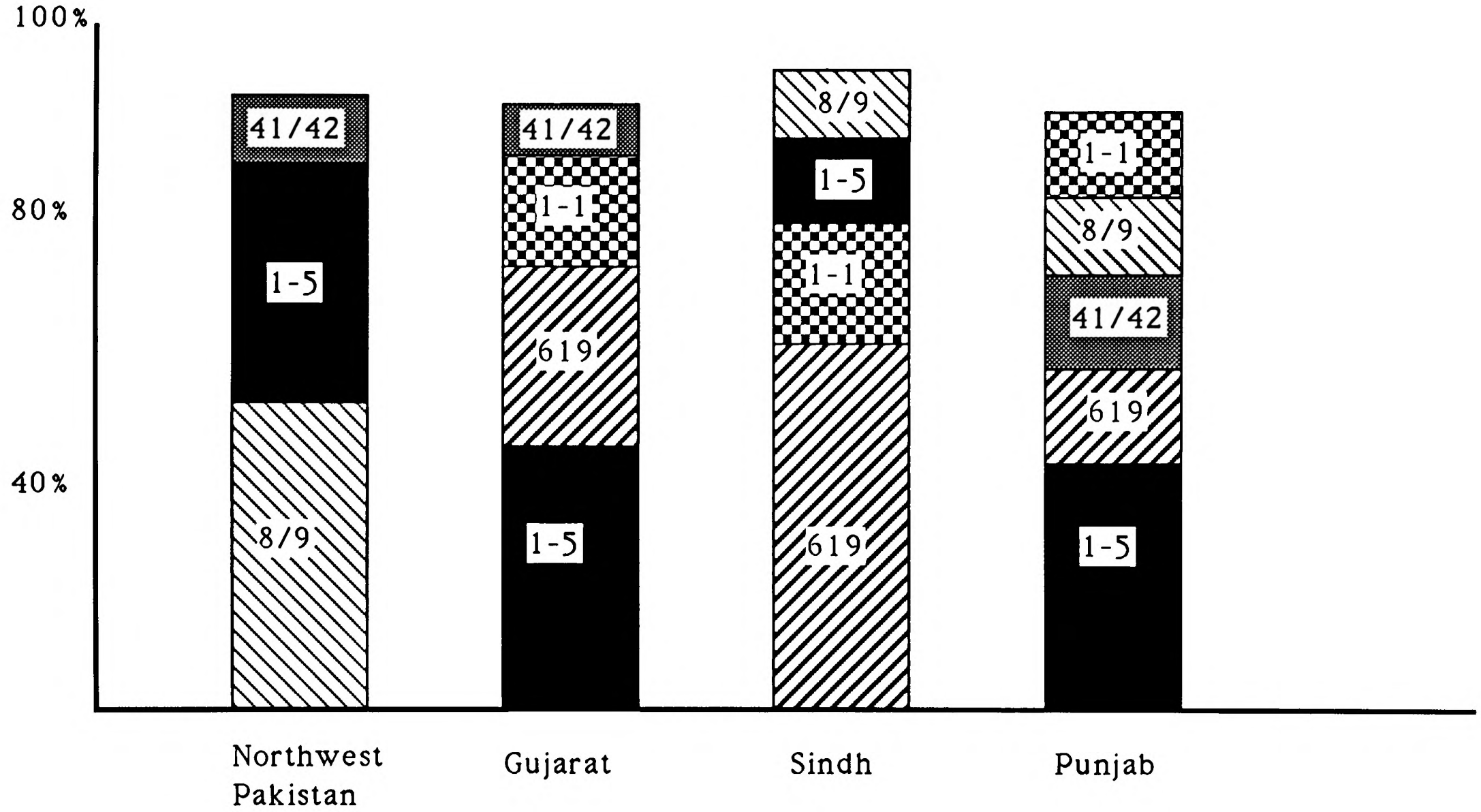


Figure 3.15 Schematic representation of the ß-thalassaemia mutations that account for about 90% of mutations in four different regions of the Indian subcontinent. The mutations are arranged in ascending order of frequency. [1-5= IVS-1 nucleotide 5 (G-A), 619 = 619 bp deletion, 1-1= IVS-1 nucleotide 1 (G-T), 8/9= codons 8/9 (+G), 41/42= codons 41/42 (-CTTT)].

Table 3.11 Frequency of ß-thalassaemia mutations in Asian Indians: summary of results

Mutations	Frequency
IVS-1 nt 5 (G-C), 619 bp deletion, codons 8/9 (+G), IVS-1 nt 1 (G-T) & codons 41/42 (-CTTT)	91.7%
codon 15 (G-A), codon 5 (-CT), IVS-1 minus 1 (G-C), codon 16 (-C) & IVS-2 nt 837 (G-T)	6.1%
cap site +1 (A-C), IVS-2 nt 1 (G-A), IVS-1 nt 110 (G-A), IVS-1 minus 1 (G-A), -88 (C-T) & codon 88 (+T)	1.7%

3.6 Discussion and Conclusions

The strategy and experiments described in this chapter were successful in characterising the ß-thalassaemia mutations in 656 carriers from seven different regions of the Indian subcontinent. A total of sixteen different mutations were identified, including seven that have not been previously described in the immigrant Asian Indian populations, two of which were completely new mutations. The comprehensive description of the frequencies and regional distribution of the ß-thalassaemia mutations will form the basis for a prenatal diagnosis program in this population by DNA analysis. Based on the observations of this large mutation detection study it was possible to recommend an appropriate methodology and strategy for prenatal diagnosis of ß-thalassaemia in the Asian Indian population, both in India itself and in any other country with an Asian Indian immigrant community.

3.6.1 Strategy for mutation detection

DNA analysis should be performed on couples at risk of having a thalassaemia major offspring and are interested in prenatal diagnosis in order to identify their ß-thalassaemia mutations. The

approach would be to first seek the five common ß-thalassaemia mutations which account for 92% of all alleles; namely, IVS-1 nucleotide 5 (G-C), the 619 bp deletion, codons 8/9 (+G), IVS-1 nucleotide 1 (G-T) and codons 41/42 (-CTTT). The panel of common mutations primarily sought may be advantageously altered if the precise regional origins of the couple were known. Furthermore a diagnostic laboratory in India could have its own specific panel of common mutations according to the regional origins of the population it served. This panel of common mutations should include those that account for around 90% of all the ß-thalassaemia alleles in that region, as described previously. It was not possible to accurately define such a panel of common mutations for Maharashtra, Bengal and Bangladesh, and Tamil Nadu because of the small number of carriers that were studied from these regions. However, on characterisation of the ß-thalassaemia mutations of 52 unrelated carriers from Tamil Nadu, it was found that 43 (83%) had the IVS-1 nucleotide 5 (G-C) mutation, six (12%) carried the codons 41/42 (-CTTT) mutation, two individuals had the codons 8/9 (+G) mutation and a single carrier was found to carry the 619 bp deletion (R. Sarkar, pers. comm.). These results when combined with those of the seven carriers from Tamil Nadu described previously indicate that IVS-1 nucleotide 5 (G-C) and codons 41/42 (-CTTT) account for 91% of the ß-thalassaemia mutations in this region. ß-Thlassaemia was found to be most homogenous in Tamil Nadu; where 48/59 (81.3%) of carriers had the IVS-1 nucleotide 5 (G-C) mutation. Further, 28/29 (96.5%) of couples carried identical mutations.

In order to identify the ß-thalassaemia mutations in the 8% of individuals who remain uncharacterised after the primary screening procedure, two approaches are available. Firstly, a second panel of five rare mutations: codon 15 (G-A), codon 5 (-CT), IVS-1 minus 1 (G-C), codon 16 (-C) and IVS-2 nucleotide 837 (G-T) which account for 6% of the alleles present in the population as a whole could be sought. Following this secondary screening procedure, it should be possible to characterise the mutations in 98% of individuals, irrespective of their particular regional origins. On consideration of the regional distribution of the rare mutations, modifications of the secondary panel could be made. For example, codon 15 (G-A) and codon 16 (-C) could be a part of the primary

panel for Northwest Pakistan and Punjab respectively or cap site position +1 (A-C) could be included in the secondary panel for Punjab. With these regional modifications it would be possible to increase the proportion of cases characterised by primary and secondary screening to above 98%. The exception to this would be Punjab where as many as seven rare mutations were identified in single individuals. The 2% of individuals whose ß-thalassaemia mutations remain uncharacterised following this secondary screening procedure would then have to be studied by DNA sequencing of their amplified ß-globin genes. The second approach for identifying the ß-thalassaemia mutations in individuals who remain uncharacterised following the primary screening procedure would be to omit the secondary screening procedure and directly study the DNA sample by sequencing.

A DNA diagnostic laboratory in India should perform both the primary and secondary screening procedure prior to sequence analysis; this should reduce the number of cases requiring DNA sequencing to a minimum. The scenario would be rather different in a DNA diagnostic laboratory in a developed country with a multi-ethnic population, which was providing ß-thalassaemia prenatal diagnosis to its immigrant Asian Indian community. In such a situation it would be easier to perform a primary screen for the five common mutations followed by mutation detection in the uncharacterised individuals by DNA sequencing. This would obviate the necessity of having an array of different panels of mutations for secondary screening, for the various ethnic groups requiring prenatal diagnosis. Further, as discussed in the next section, advances in DNA sequencing technology make it relatively rapid and less labour intensive.

Once the ß-thalassaemia mutations are detected in both parents prenatal diagnosis becomes straightforward, requiring the detection of the presence or absence of the maternal and paternal mutations in the fetal tissues obtained for analysis. If both parents are found to carry the same mutation as would occur in about 60% of Asian Indians it would be vital to distinguish the homozygous from the heterozygous state in the fetus.

In some rare individuals, whose haematological investigations are suggestive of a ß-thalassaemia carrier state or who have an affected thalassaemia major child, it is not possible to identify the

causative mutation, inspite of sequencing the entire ß-globin gene along with its flanking sequences (Semenza et al, 1984). In such individuals the molecular pathology is unknown and the causative factor might be in the dominant control region at the 5' end of the ß-globin gene cluster. Prenatal diagnosis in these cases cannot be carried out by mutation detection.

3.6.2. Appropriate methodology

Mutation detection is performed most efficiently by employing the PCR to selectively amplify the DNA fragment of interest. This will reduce the time, labour and amount of DNA required for mutation detection. Detection of the 619 bp deletion is an excellent example of the PCR greatly simplifying gene diagnosis, as the labour intensive steps of restriction enzyme digestion, Southern blotting and hybridisation are replaced by a single amplification reaction followed by analysis of the abundant amounts of amplified product by gel electrophoresis. Detecting a point mutation is only slightly more complicated. To follow the prenatal diagnosis strategy outlined above an appropriate method would be required to rapidly screen for 4-5 ß-thalassaemia mutations. The mutation detection method successfully developed in this chapter, allele specific PCR, has the advantages of being accurate and nonradioactive. Once specific reaction conditions have been determined screening experiments can be performed simply and rapidly. Screening for the five common Asian Indian ß-thalassaemia mutations by allele specific PCR would entail four reactions performed simultaneously in a thermal cycler though in separate tubes, with the primers flanking the 619 bp deletion serving as internal controls. Examining these four PCR products by electrophoresis on a sieving agarose gel would characterise the ß-thalassaemia mutation in 92% of cases.

Similarly, secondary screening for the five mutations which account for 6% of ß-thalassaemia alleles could also be performed efficiently by allele specific PCR. It was possible to detect 28 different ß-thalassaemia mutations by allele specific PCR, using the uniform reaction conditions, proving the efficacy and robustness of this technique. Allele specific PCR will be particularly useful in a laboratory performing prenatal diagnosis for a multi-ethnic

population, as a variety of ß-thalassaemia mutations can be studied without the inconvenience of repeated radiolabelling of different oligoprobes.

As the ß-thalassaemia mutations are located close to each other in a relatively small gene, multiplex allele specific PCR, whereby mutations situated in different regions of the gene are simultaneously analysed in a single reaction, would be difficult to perform. The need for PCR product analysis by gel electrophoresis could be obviated by a colour complementation assay where flourescent dyes are conjugated to the oligonucleotide primers; the normal and mutant allele specific primers would be labelled with a different dyes, so that by analysis of the colour of the PCR product under UV light the normal, homozygous or heterozygous state would be immediately diagnosed (Chehbab and Kan, 1989); making the procedure amenable to automation.

Gene diagnosis by dot-blot hybridisation was found to be unsatisfactory due to the nonspecificity of signals obtained and only 70% of the results were confirmed by allele specific PCR. However the accuracy, sensitivity and specificity was found to vary with the four different ASOs used possibly implying that accuracy could have been improved by manipulating various parameters like the sequence of the ASOs, the mismatch, hybridisation and washing conditions. The use of P^{32} as the radioactive label was a particular disadvantage especially as it was necessary to screen for a number of different mutations entailing frequent radiolabelling procedures for the different ASOs. This problem would be further compounded in a laboratory serving a multi-ethnic population. Subsequent developments in dot-blot hybridisation technology have resolved some of these drawbacks. Reverse dot-bots where the ASOs are fixed to the filter, the amplified DNA covalently attached to horseradish peroxidase and hybridised to the filter, so as to enable mutation detection by a nonradioactive colorimetric reaction involving the enzymatic oxidation of a colourless soluble chromogen to a blue insoluble product, could be an alternative method for screening for ß-thalassaemia mutations (Saiki et al, 1989). A viable alternative approach for screening for the common Asian Indian ß-thalassaemia mutations, excluding the 619 bp deletion, could thus be reverse dot-blots with nonradioactive hybridisation. Nylon

filters with immobilised ASOs for 4-5 common mutations could be prepared on a large scale, maybe commercially, and DNA from individuals to be screened or from fetal tissues could be amplified and hybridised in a single step to this filter. Further, the ASOs on the filters could be modified for different regional groups.

DNA sequencing could be facilitated by rapid screening of different regions of the ß-globin gene by DGGE or by chemical cleavage of mismatch in order to identify the gene fragment containing the mutation. The use of magnetic beads could facilitate the template preparation step, the amplified fragment to be sequenced is end-labelled with biotin permitting separation by streptavidin coated magnetic beads to produce a single-stranded purified sequencing template (Hultman et al, 1989). Further each of the four terminating dideoxynucleotides could be fluorescent labelled, then fluorescence detectors could "read" the DNA sequence at the bottom of the gel as each fragment appears and a computer prints out the DNA sequence. With these developments DNA sequencing is becoming increasingly simple, rapid and amenable to automation enabling it to become applicable to a DNA diagnostic laboratory. Thus it would be more efficient to sequence all samples that remain uncharacterised after the primary screen for the commonest mutations without the intermediary secondary screening procedure; particularly so in a developed country with a multi-ethnic population.

3.6.3. Time frame for prenatal diagnosis by mutation detection

DNA could be extracted from a blood sample or from the fetal tissues within a single day, the primary screen for the five common mutations could be carried out in about four hours by allele specific PCR. For the less than 10% of uncharacterised individuals a secondary screen would require a similar amount of time. Identification of the fragment carrying the mutation by either DGGE or by chemical cleavage would require a working day and DNA sequence analysis of the fragment could be done in two to three days. Theoretically it should be possible to determine the ß-thalassaemia mutations in both parents in less than two days in the majority of cases and in a maximum of six working days in all

cases. Ideally the parents should be screened at the pre-conception stage or during early pregnancy, making it straightforward to detect the presence or absence of paternal and maternal mutations in the fetal sample and in couples who carry identical mutations to determine whether the fetus is homozygous for that mutation. In the occasional cases in which the parents carry an unusual mutation, allele specific primers could be designed and tested before the pregnancy is undertaken or the fetal sample received.

3.6.4 Causes for error

Errors in the genotype diagnosis by allele specific PCR could occur because of various reasons. An important reason is maternal contamination of the fetal sample, emphasising the need to examine the chorion villus sample under the dissecting microscope and to dissect out any contaminating maternal decidua. Minute traces of maternal tissue might result in a false diagnosis of the maternal mutation in the fetus, as the allele specific primer would anneal to the small amounts of maternal sequence present in the test sample, more so in the absence of complementary fetal DNA sequences, with resulting amplification inspite of the mutation being absent in the fetus. To reduce this risk to a minimum it is is important to use stringent reaction conditions like a reduced number of PCR cycles and a high annealing temperature. When a band of lower intensity as compared to the internal control fragment is visualised on the ethidium bromide stained gel, the possibility of maternal contamination must be considered. Maternal contamination of the fetal sample could be detected by co-amplification of a highly polymorphic, variable number of tandem repeats (VNTR) segment (Horn et al, 1989) which would show the contribution of two maternal chromosomes, indicating maternal contamination. Indeed a pair of primers which amplify such a polymorphic VNTR segment could be used as the internal control primers for allele specific PCR. This system would also address the other major cause of error: nonpaternity, it would be meaningless to look for the putative fathers mutation in the fetal sample in these cases.

The exquisite sensitivity of the PCR is its major drawback and to avoid false positive results, the importance of appropriate controls cannot be emphasised enough. Negative and positive controls must be included in all screening experiments to ensure specificity and sensitivity, respectively. False positive results due to contamination of the PCR mix with external DNA did not seem to be a problem probably because the number of cycles was restricted to 25-30. However the potential risk of contamination of the PCR mix must be kept in mind; the most important cause of contamination is PCR carry-over product, thus precautions must be taken to try and segregate work areas and equipment used to set up reactions from those used to analyse the amplified product. False negative results were occasionally obtained with allele specific PCR. Thus before an uncharacterised DNA sample was studied by DNA sequencing it was re-screened by allele specific PCR and a mutation previously missed was detected. It might be time and labour saving for a prenatal diagnosis service, if a such a policy of re-screening the DNA sample before sequencing was adopted.

3.6.5 Asian Indian ß-thalassaemia mutations: Other studies

After the commencement of studies described in this chapter there have been other studies describing the molecular basis of ß-thalassaemia in Asian Indians. Parikh et al (1990) screened 110 ß-thalassaemia carriers from six different ethnic groups for the 10 ß-thalassaemia mutations previously described in the immigrant population, by dot-blot hybridisation. Five common mutations were found to account for 89% of alleles and 48% of couples carried identical mutations. IVS-1 nucleotide 5 (G-C) and the 619 bp deletion were found to be the two commonest mutations. Jain et al, (1991) studied 135 disease carriers originating from Northern India by allele specific PCR to find the following mutations: the 619 bp deletion in 30% of individuals, IVS-1 nucleotide 1 (G-T) in 25%, IVS-1 nucleotide 5 (G-C) in 20%, codons 8/9 (+G) in 13%, codons 41/42 (-CTTT) in 12% and 1.5% remained uncharacterised. C. Mahadik (pers. comm.) characterised the ß-thalassaemia mutations in a small number of carriers from

Maharashtra to find the IVS-1 nucleotide 5 (G-C), codon 15 (G-A), 619 bp deletion, IVS-1 minus 1 (G-A) and a new frameshift mutation at codon 55. All these studies confirm the spectrum of ß-thalassaemia mutations reported here.

CHAPTER FOUR

HAPLOTYPE ANALYSIS

4.1 Introduction

A polymorphism has been defined as the existence of two or more alleles for a particular locus, where at least two alleles appear with frequencies of more than 1 percent (Bodmer and Cavalli-Sforza, 1976). The blood groups and Hb variants are typical examples of polymorphic systems with multiple alleles. Such a variation in the structure of DNA gives rise to DNA polymorphisms. Some of these DNA polymorphisms may introduce or remove cleavage sites for specific restriction endonucleases, bacterial enzymes which recognise specific DNA sequences, and cleave DNA at particular positions (Nathan and Smith, 1975). The differences in the size of DNA fragments produced when DNA is cut with specific restriction enzymes are called restriction fragment length polymorphisms (RFLPs), which provide a considerable degree of individual variation and are inherited according to Mendelian laws. Over 20 such restriction enzyme cutting site polymorphisms have been found in the ß-globin gene cluster (Collins and Weissman, 1984; Antonarakis et al, 1985). Each of these polymorphic restriction enzyme cutting sites will either be present (+) or absent (-) in a particular chromosome.

The pattern or combination of polymorphic restriction sites for any chromosome in relation to a particular locus represents the haplotype. Given the considerable number of restriction enzyme site polymorphisms present in the ß-globin gene cluster, potentially a very large number of haplotypes could exist. However many of these possible haplotypes have not been observed and in all populations a few particular haplotypes predominate (Antonarakis et al, 1982a). It has been shown that, in the ß-globin gene cluster, there are two regions of such non-random association or linkage disequilibrium, the 5' and 3' haplotypes, with an intervening region of randomisation, situated 5' to the ß-globin gene (Antonarakis et al, 1982a; Chakravarti et al, 1984). This

region of randomisation has been shown both indirectly (Chakravarti et al, 1984) and directly (Gerhard et al, 1984; Old et al, 1986a) to be a hotspot for recombination.

Following the examination of numerous ß-globin genes, the existence of different normal ß-globin gene sequences, termed frameworks, were observed. The ß-globin gene corresponding to that sequenced by Lawn et al, (1980) was designated framework 1. Framework 2 genes differ by a single nucleotide substitution, G-T at IVS-2 nucleotide 74; while framework 3 genes in Mediterraneans, have this substitution and four others, codon 2 (CAC-CAT), IVS-2 nucleotides 16 (C-G), 81 (C-T) and 666 (T-C). In Asian and black genes of the framework 3 variety, the substitution at IVS-2 nucleotide 81 is absent. Identification of the ß-globin gene frameworks is possible by restriction mapping, because of the linkage of the framework 2 gene to an absent *Bam*HI restriction site 3' to the ß-globin gene and the absence of the *Ava*II restriction site at IVS-2 nucleotide 16 in framework 3 genes. Both these restriction sites are present in framework 1 genes (Orkin et al, 1982b).

Analysis of DNA polymorphisms has a number of clinical and research applications, including mapping disease genes (Reeders et al, 1985) and performing prenatal diagnosis by RFLP linkage analysis (Old et al, 1984). Haplotype analysis has been most useful for the identification of a number of ß-thalassaemia mutations, as each mutation tends to be associated with a particular haplotype. Haplotype-mutation associations were first demonstrated by studies in Mediterraneans (Orkin et al, 1982b; Kazazian et al, 1984a) and later in studies conducted in the Asian Indian (Kazazian et al, 1984b), Chinese (Cheng et al, 1984) and Black populations (Antonarakis et al, 1984). By cloning and sequencing ß-thalassaemia genes associated with different haplotypes in each population group, the number of mutations identified was maximised. With the discovery of an increasing number of restriction site polymorphisms the potential of genetic anthroplogy has increased enormously (Cooper and Schmidtke, 1984). Using multiple polymorphisms genetic distance analysis may be performed (Wainscoat et al, 1986) and also certain population specific polymorphisms serve as valuable population markers to

track down population migration (Hill et al, 1985; Kulozik et al, 1986).

The ß-globin gene haplotypes of a large number of normal chromosomes (ß-A) and chromosomes carrying the ß-thalassaemia gene (ß-Th) were analysed in order to determine the origin and subsequent spread of the disease on the Indian subcontinent. Further, regional variations in the distribution of ß-thalassaemia mutations could serve as useful markers for anthropological studies on the Indian subcontinent.

4.2 Subjects

The population sample studied consisted of the 656 unrelated ß-thalassaemia carriers whose ß-thalassaemia mutations were characterised in the previous chapter. These individuals and their famliy members were interviewed and bled as described previously.

4.3 Methods

DNA was extracted from the anticoagulated blood samples as described in section 2.2. The seven restriction enzyme cutting sites analysed for ß-globin gene haplotype construction were *Hind*II-ε, *Hind*III-$^{G}\gamma$, *Hind*III-$^{A}\gamma$, *Hind*II-5'ψβ, *Hind*II-3'ψβ, *Ava*II-ß and *Bam*HI-ß (figure 4.1).These seven polymorphic restriction sites are clustered in two groups, a 34 kb region 5' to the δ gene contains the first five restriction sites and the remaining are present within and 3' to the ß-gene. The combination of restriction sites in the two groups constitute the 5' and 3' ß-globin gene haplotypes, respectively. RFLP analysis for each of these restriction sites was performed by either Southern blot hybridisation or by digestion of PCR products.

4.3.1 Southern blot hybridisation

Five µg of DNA was digested with the appropriate restriction enzyme using the conditions recommended by the manufacturer. Following digestion the DNA fragments were separated according to size by electrophoresis in a 0.8% agarose gel and then

transferred onto a filter by Southern blotting. The filters were hybridised with one of the appropriate genomic DNA fragment probes; 1.3 kb ε *Bam*HI/*Eco*RI fragment in pBR322, 3.2 kb γ *Hind*III fragment in pBR322, 3.9 kb ψβ *Pst*I fragment in pBR322 and 4.4 kb ß *Pst*I fragment in pUC 13. These genomic DNA probes were radiolabelled with [α^{32}P]dCTP, using the Multiprime Labelling Kit (Amersham) as described in section 2.3.4.1. Hybridisation and washing of filters was done as described in section 2.3.4. Filters were wrapped in cling film and exposed overnight to X-ray film at -80°C, using cassettes with intensifying screens. From the sizes of the bands visualised by autoradiography, the presence or absence of the RFLP was determined. This entire procedure took 5-6 days to complete.

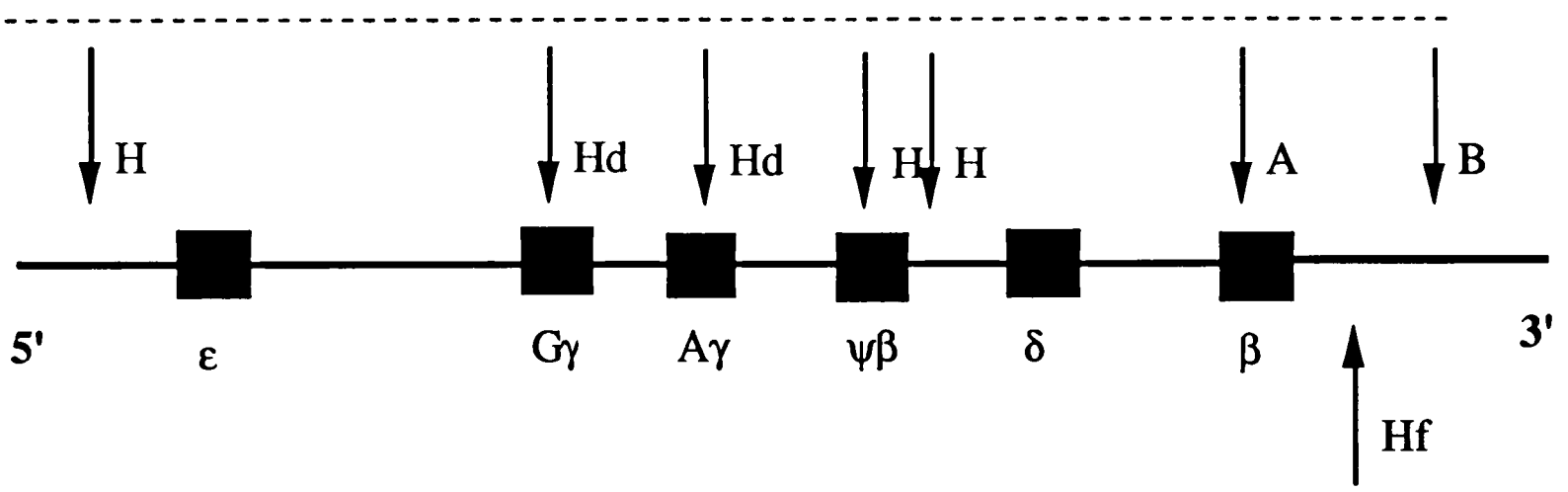


Figure 4.1 A schematic representation of the ß-globin gene cluster showing the restriction enzyme cutting sites used for haplotype construction (H=*Hind*II, Hd=*Hind*III, A=*Ava*II and B=*Bam*HI. Hf=*Hinf*I).The *Hinf*I site was examined instead of the *Bam*HI, when the haplotypes were constructed by PCR.

During the period 1982-88, prenatal diagnosis of ß-thalassaemia was performed by RFLP analysis, using Southern blot hybridisation as outlined above, at the National Haemoglobinopathy Reference Laboratory, John Radcliffe Hospital. One hundred and fifty four of the 656 subjects studied had been referred for prenatal diagnosis

during this period and were studied by RFLP analysis. Dr John Old kindly made available the results of these RFLP studies.

4.3.2 RFLP analysis by PCR

In 502 ß-thalassaemia carriers RFLP analysis was performed more simply by restriction enzyme digestion of PCR product. The seven restriction sites listed above were determined by this approach, with the exception of the *Bam*HI-ß site. It is difficult to amplify the region around the *Bam*HI polymorphic site as it lies within a repetitive sequence. Hence the *Hinf*I site, located 3' to the ß-globin gene which has been described to be in complete linkage disequilibrium with the *Bam*HI (Semenza et al, 1989) was studied instead. The *Hinf*I site located 3' to the ß-globin gene was analysed in the 154 subjects in whom the *Bam*HI-ß site had been determined by Southern blot hybridisation. In all these individuals both polymorphic sites were always either present or absent, confirming the total association of the *Hinf*I and *Bam*HI sites; thus the *Hinf*I site was analysed instead of the *Bam*HI site for haplotype construction.

DNA amplification was performed using approximately 0.5 µg of genomic DNA and 0.5 units of *Taq* polymerase in a 25 µl reaction volume containing 20 pmol of each primer, 800 µM total dNTPs, 10 mM Tris HCl pH 8.3, 50 mM KCl amd 1.5 mM $MgCl_2$. The thermal cycling regimen consisted of 30 cycles of denaturation at 93°C for 1 min, annealing at 55°C for 1 min and extension at 72°C for 1.5 min. In the final cycle extension was prolonged to 3 min. Amplification of the DNA fragment encompassing the *Hind*III-$^{G}\gamma$ site, was achieved using an annealing temperature of 65°C. The amplified products were subjected to restriction enzyme digestion according to the manufacturers instructions. All the digests performed in this study were done using 25 µl of the amplified product, 20 units of the restriction enzyme and 3 µl of the appropriate buffer, supplied by the manufacturer. Digestion was done at 37°C for a minimum of four hours. Following digestion a 10 µl aliquot of the PCR product was examined by electrophoresis in a 1.5% agarose and 1.5% NuSieve (ICN Biomedicals, Ltd) gel, which was then stained with ethidium bromide and visualised by UV transillumination.

Presence or absence of the polymorphic restriction site was determined by the sizes of the digested PCR products.

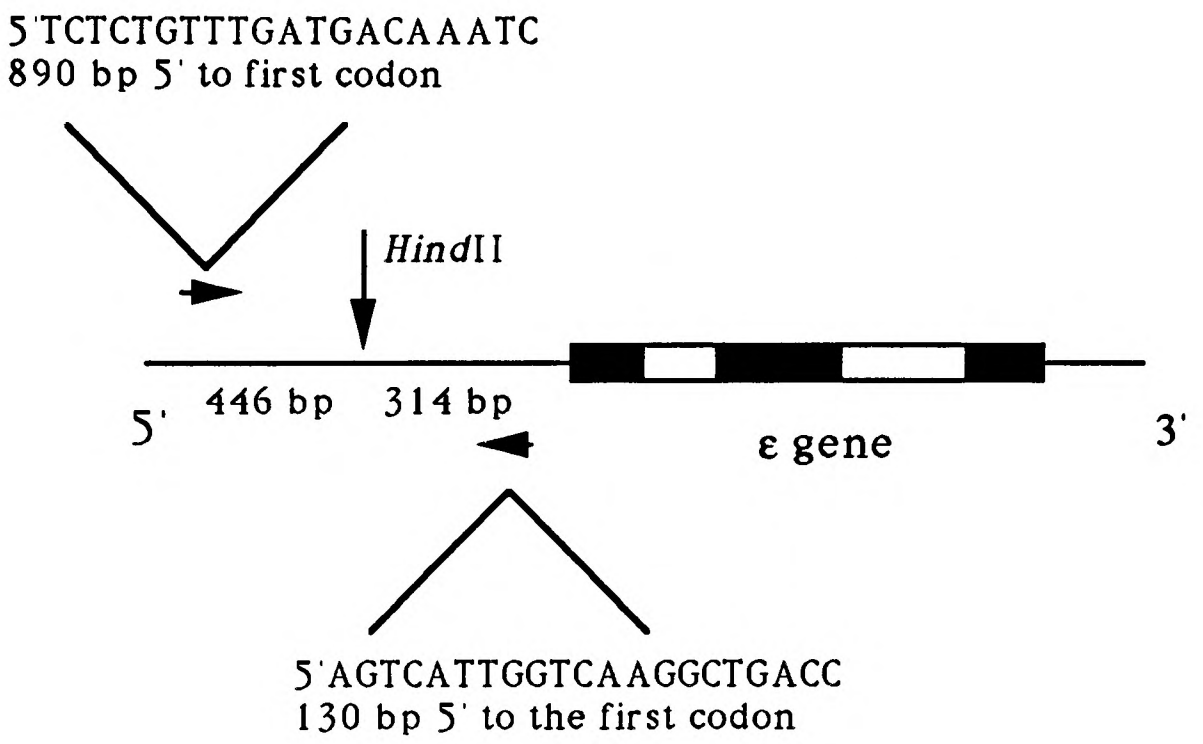
The oligonucleotide primers used for the amplification of the gene fragments containing the restriction sites analysed are shown in figures 4.2 to 4.8. Also shown, are the PCR products obtained, following restriction enzyme digestion, in the absence and presence of the polymorphic restriction site.

4.3.3 Haplotype Construction

Every individual possesses two ß-globin gene clusters and hence two haplotypes. As all the subjects analysed were ß-thalassaemia carriers, they all had a ß-A and a ß-Th haplotype. Attempts to construct the ß-Th and ß-A haplotypes were made in 656 and 210 subjects, respectively.

Each restriction site is dimorphic, being recorded as either present (+) or absent (-) according to the sizes of the fragments visualised. To construct the ß-globin gene haplotype, the restriction sites were aligned according to their position in the ß-globin gene cluster in a 5'-3' direction. The 5' haplotype was derived from the restriction sites in the ε, $^{G}\gamma$, $^{A}\gamma$ and $\psi\beta$ genes, and the 3' haplotype was derived from the sites within and 3' to the ß gene.

In order to construct the ß-A and ß-Th haplotype, it was necessary to determine which RFLP was associated with the ß-Th chromosome and which one with the ß-A chromosome. If an individual was homozygous for the RFLP, distinction between the ß-A and ß-Th chromosomes was not necessary. However if a carrier was found to be heterozygous for an RFLP, DNA from a thalassaemia major offspring, was studied for that particular RFLP. If this RFLP occurred in the homozygous state in the affected child, it could be concluded that the RFLP was linked to the ß-thalassaemia gene, thus assignment of the RFLP to the ß-Th chromosome of the parent could be made. But, if this RFLP occurred in the heterozygous state in the affected child, RFLP assignment in the parent under question could be done only if the RFLP inherited by the child from its other parent was known. If DNA from a thalassaemia major child was unavailable for study, DNA from a normal offspring, if available, was analysed and RFLP assignment to ß-A and ß-Th haplotypes done in a similar manner.



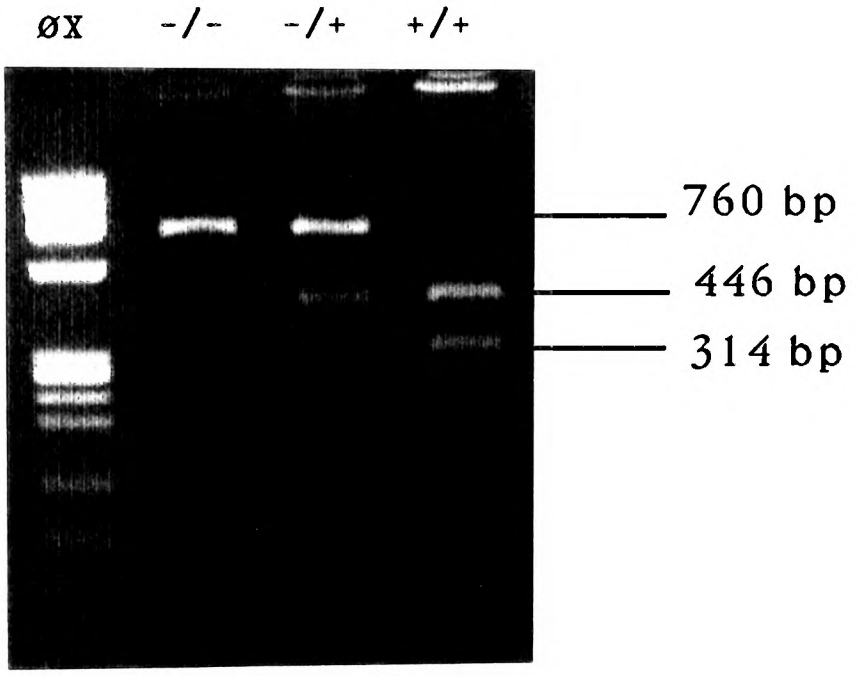
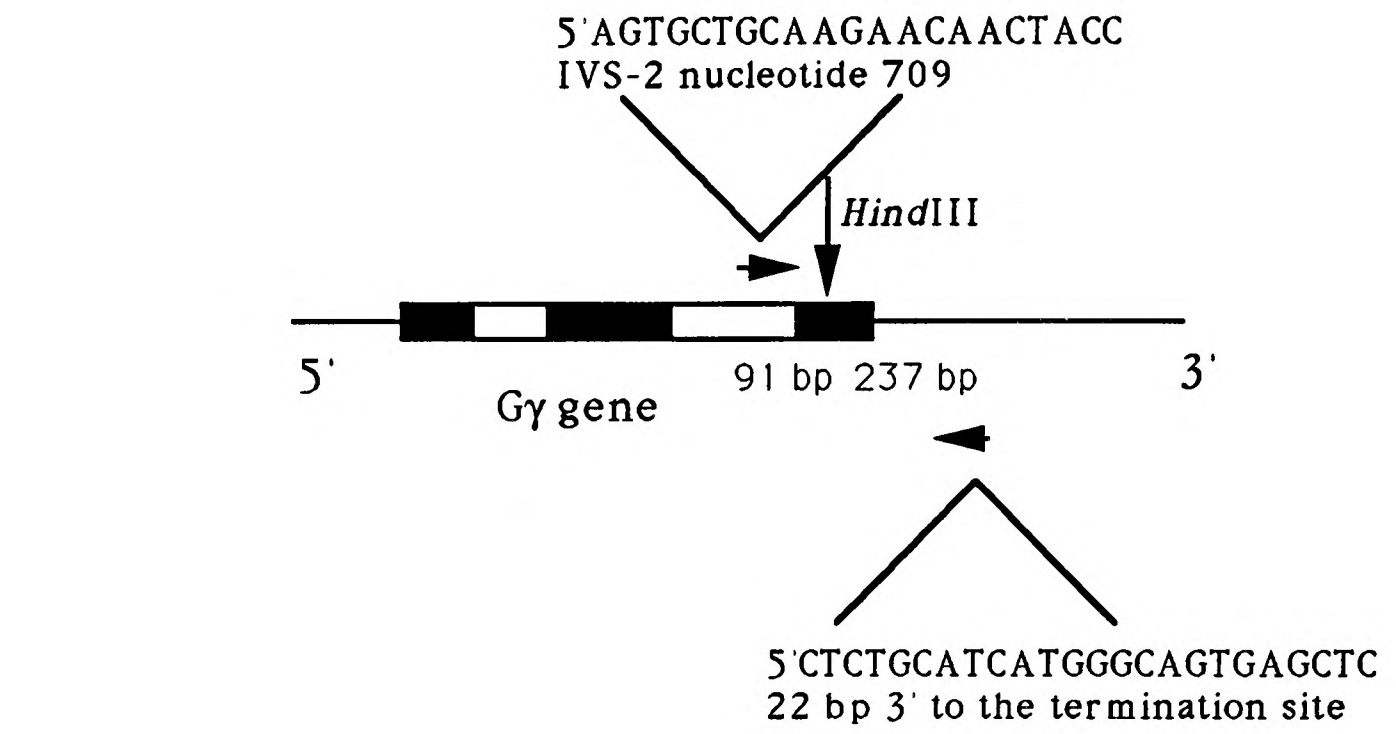


Figure 4.2 Determination of the ε-*Hind*II RFLP by restriction enzyme digestion of PCR product. The top half of the figure shows a schematic representation of the gene with the primers used for amplification and the location of the restriction enzyme cutting site. The position of the 5' nucleotide of each primer is indicated below the sequence. The lower half shows an ethidium bromide stained agarose gel with the amplified products following restiction enzyme digestion. The marker is øx 174 DNA digested with *Hae*III.



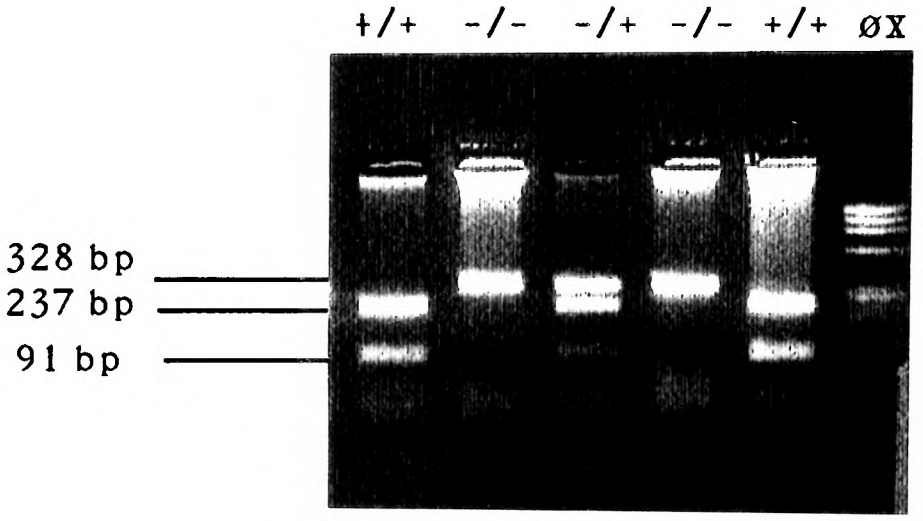
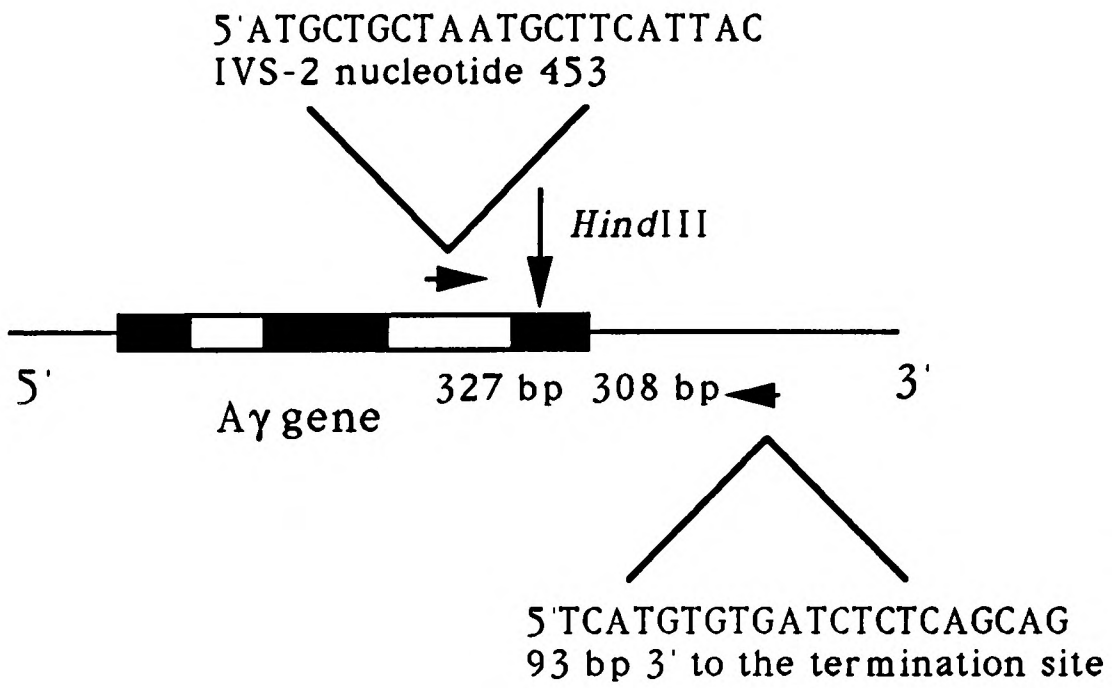


Figure 4.3 Determination of the Gγ-*Hind*III RFLP by restriction enzyme digestion of PCR product. The top half of the figure shows a schematic representation of the gene with the primers used for amplification and the location of the restriction enzyme cutting site. The position of the 5' nucleotide of each primer is indicated below the sequence. The lower half shows an ethidium bromide stained agarose gel with the amplified products following restiction enzyme digestion. The marker is øx 174 DNA digested with *Hae*III.



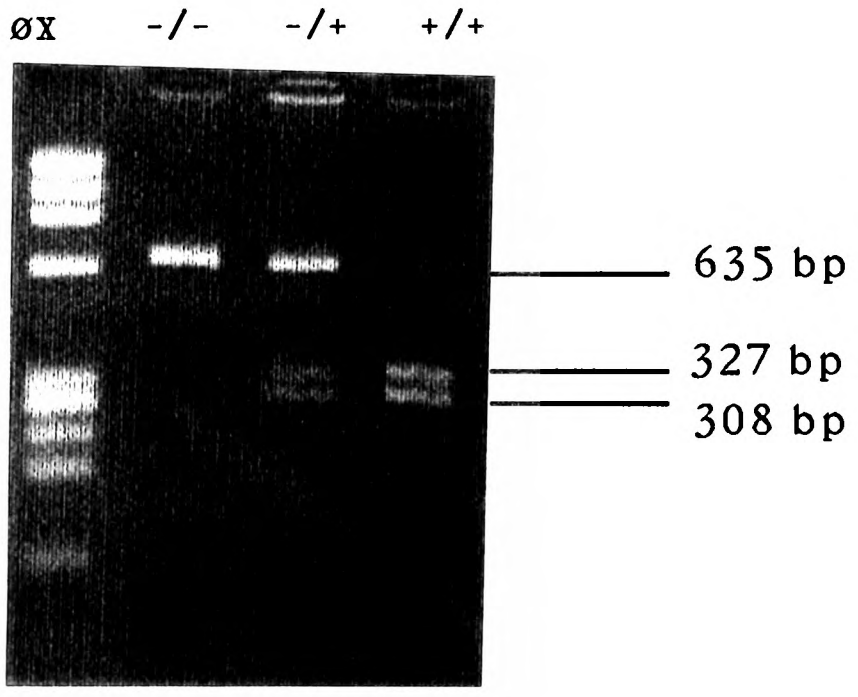
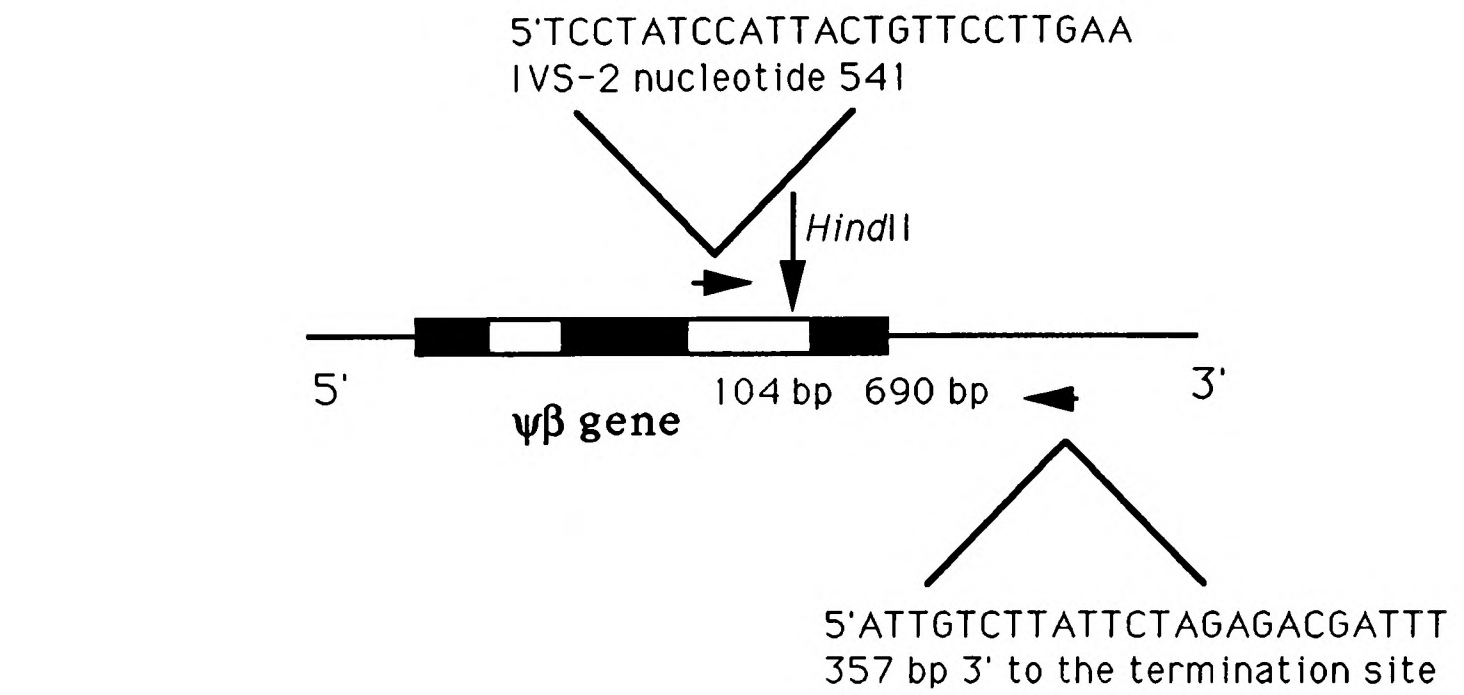


Figure 4.4 Determination of the Aγ-*Hind*III RFLP by restriction enzyme digestion of PCR product. The top half of the figure shows a schematic representation of the gene with the primers used for amplification and the location of the restriction enzyme cutting site. The position of the 5' nucleotide of each primer is indicated below the sequence. The lower half shows an ethidium bromide stained agarose gel with the amplified products following restiction enzyme digestion. The marker is øx 174 DNA digested with *Hae*III.



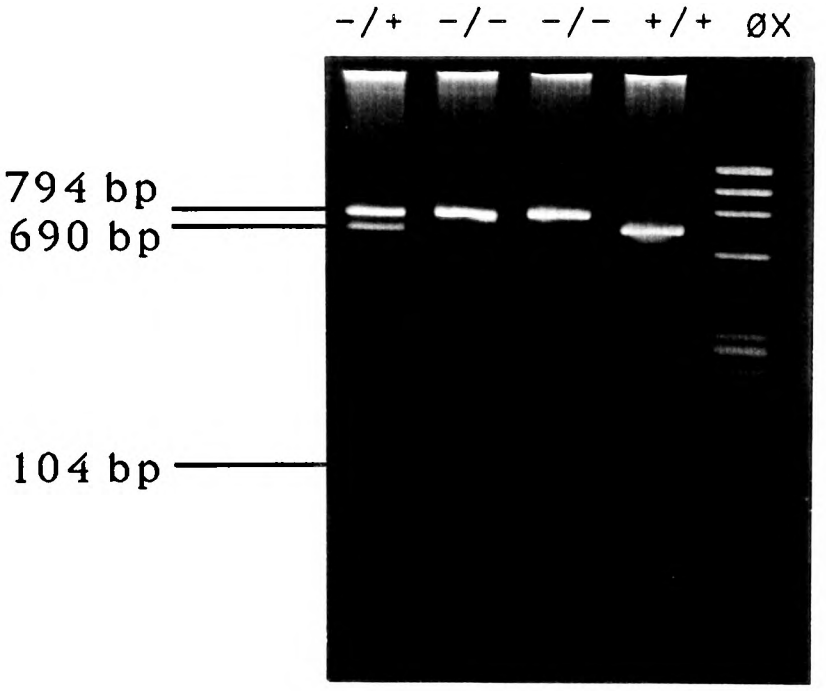
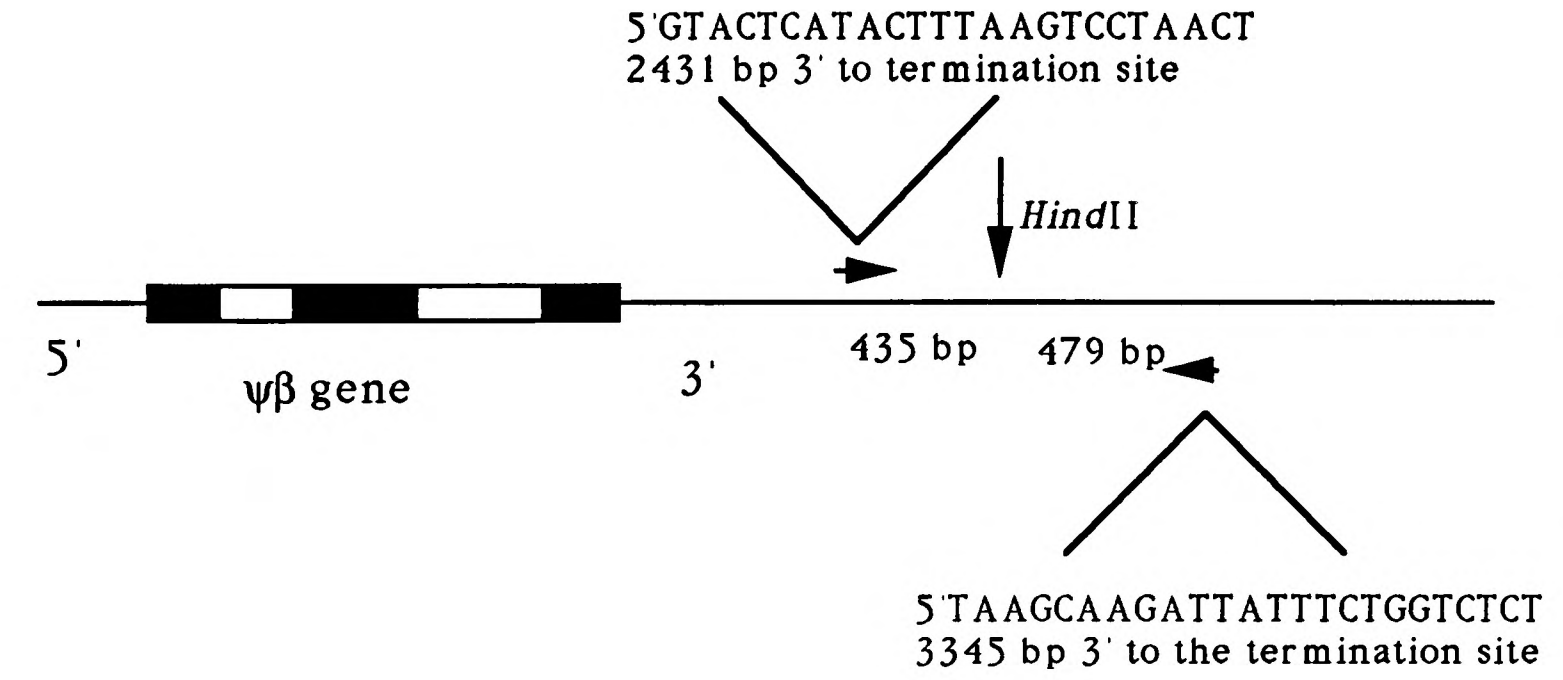


Figure 4.5 Determination of the 5'ψβ-*Hind*II RFLP by restriction enzyme digestion of PCR product. The top half of the figure shows a schematic representation of the gene with the primers used for amplification and the location of the restriction enzyme cutting site. The position of the 5' nucleotide of each primer is indicated below the sequence. The lower half shows an ethidium bromide stained agarose gel with the amplified products following restiction enzyme digestion. The marker is øx 174 DNA digested with *Hae*III.



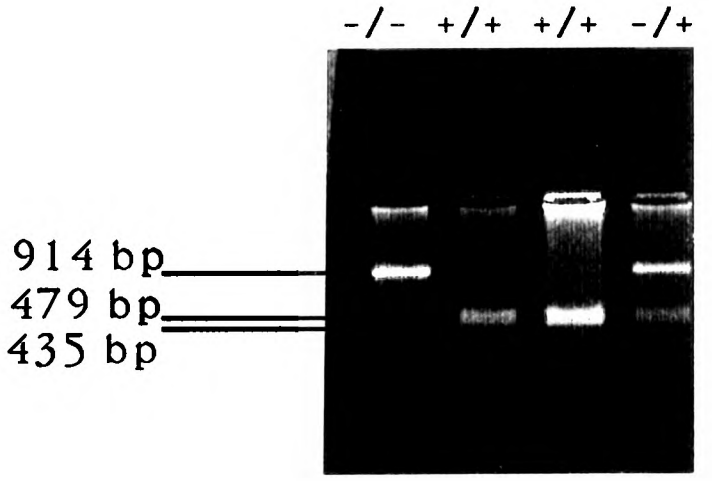


Figure 4.6 Determination of the 3'ψβ-*Hind*II RFLP by restriction enzyme digestion of PCR product. The top half of the figure shows a schematic representation of the gene with the primers used for amplification and the location of the restriction enzyme cutting site. The position of the 5' nucleotide of each primer is indicated below the sequence. The lower half shows an ethidium bromide stained agarose gel with the amplified products following restiction enzyme digestion. The marker is øx 174 DNA digested with *Hae*III (not shown).

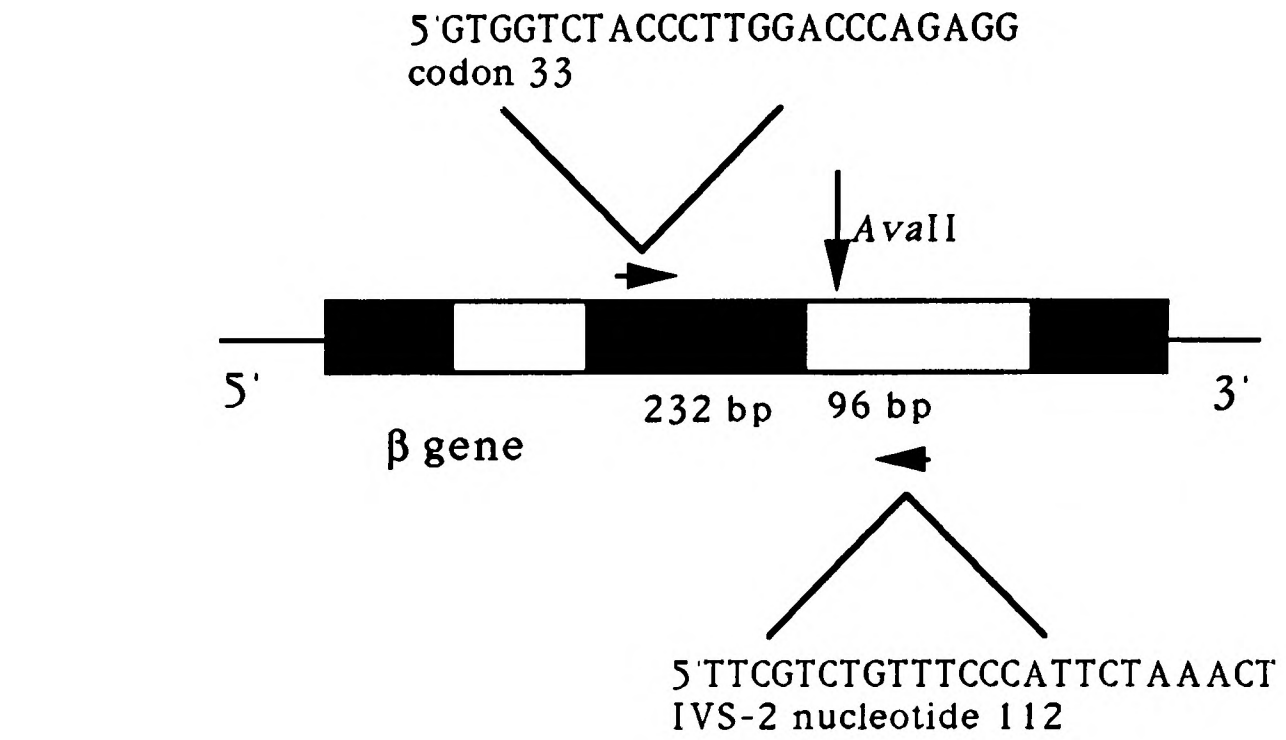
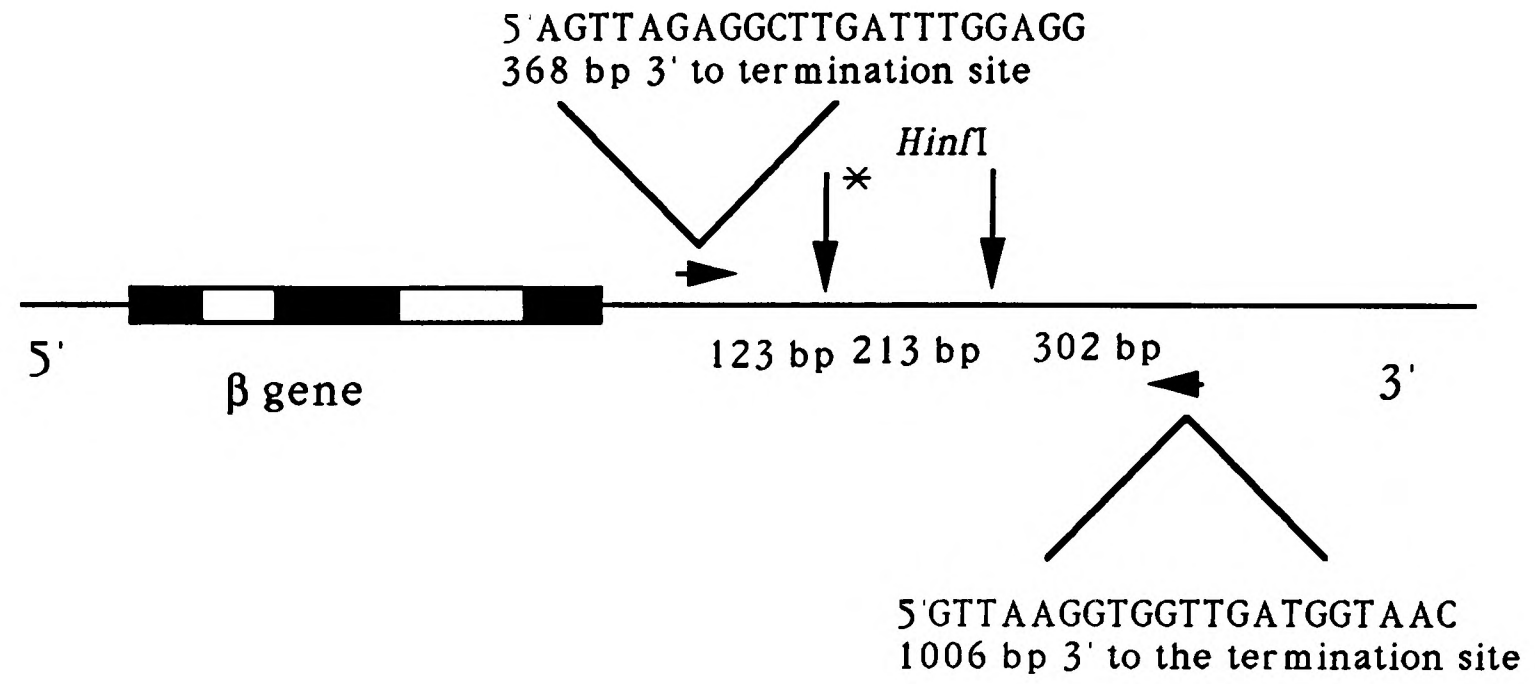


Figure 4.7 Determination of the β-*Ava*II RFLP by restriction enzyme digestion of PCR product. The top half of the figure shows a schematic representation of the gene with the primers used for amplification and the location of the restriction enzyme cutting site. The position of the 5' nucleotide of each primer is indicated below the sequence. The lower half shows an ethidium bromide stained agarose gel with the amplified products following restiction enzyme digestion. The marker is øx 174 DNA digested with *Hae*III.



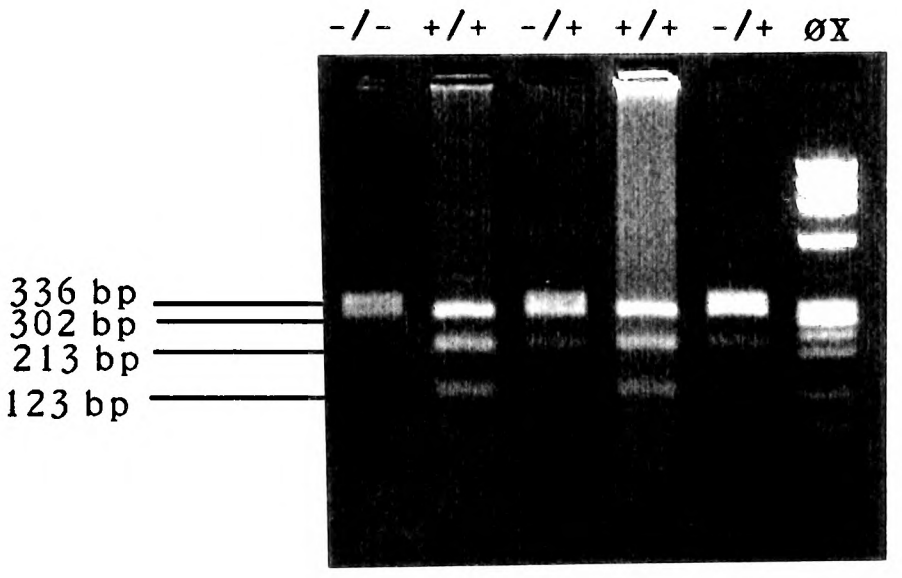


Figure 4.8 Determination of the β-*Hinf*I RFLP by restriction enzyme digestion of PCR product. The top half of the figure shows a schematic representation of the gene with the primers used for amplification and the location of the restriction enzyme cutting sites. The site marked with a * is the polymorphic site. The position of the 5' nucleotide of each primer is indicated below the sequence. The lower half shows an ethidium bromide stained agarose gel with the amplified products following restiction enzyme digestion. The marker is øx 174 DNA digested with *Hae*III.

In a situation in which DNA from an affected or normal offspring was not available for analysis it became necessary to study the RFLP in other family members like parents or siblings of the subject. Figure 4.9 shows two examples of family studies.

4.4 Results

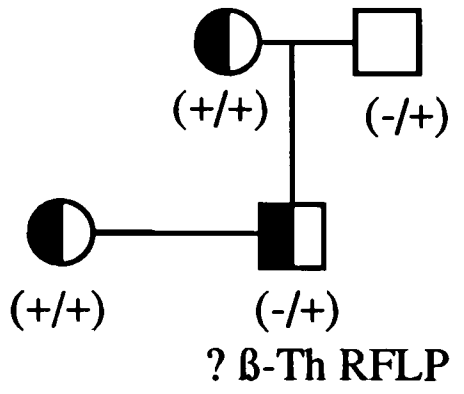
Of the 656 subjects studied, a certain assignment of ß-Th haplotype was possible for 419 ß-Th chromosomes. In remaining 237 subjects it was not possible to assign the ß-Th haplotype with certainty, because of heterozygosity for a particular RFLP and nonavailability of family members for analysis; these subjects were excluded from subsequent analysis. Of the 210 carriers in whom construction of the ß-A haplotype was attempted, certain haplotype assignment was possible in 196 cases.

4.4.1 ß-A and ß-Th haplotypes

Amongst 196 normal Asian Indian chromosomes in which a certain assignment of the ß-globin gene haplotype was possible, 19 different ß-gene haplotypes were identified (table 4.1). Seven of these ß-A haplotypes were present at a frequency of over 3%. Of the 32 (2^5) possible 5' haplotypes, eight were observed. Three 5' haplotypes +---- (51%), -+-++ (20%) and -++-+ (19%) were found in 90% of chromosomes. Of the 3' haplotypes, both the *Ava*II-ß and *Bam*HI-ß sites were present (framework 1) in 97 (50%) chromosomes, in 31 (16%) the *Bam*HI-ß site was absent (framework 2) and in 68 (35%) the *Ava*II-ß site was absent (framework 3).

A total of twelve different ß-globin gene haplotypes were identified in the 419 ß-Th chromosomes studied (table 4.1), although only five of these ß-Th haplotypes occurred at a frequency of over 3%. Three common haplotypes +---- -+, +---- ++ and -+-++ ++ were associated with 80% of the ß-Th chromosomes. With the exception of five chromosomes the 5' haplotype observed was either +---- (78%), -+-++ (13%) or -++-+ (8%). The distribution of the 3' haplotypes and ß-globin gene frameworks was different from that of the normal chromosomes (table 4.2).

(i) *Ava*II-β



(ii) *Hind*II-ε

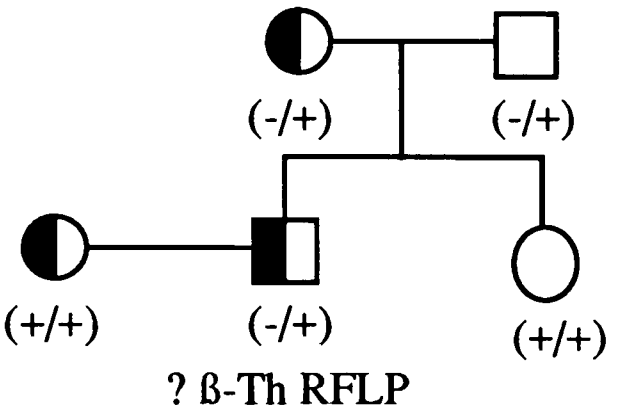


Figure 4.9 RFLP studies of family members in order to construct ß-globin gene haplotypes. In the first example, the male partner must have inherited his ß-Th chromosome from his carrier mother as his father is not a disease carrier. As the mother is homozygous for the *Ava*II-ß RFLP, it can be inferred that presence of this RFLP is linked to the ß-Th chromosome. In the second example, the male partner's family is studied for the *Hind*II-ε RFLP. In this family, the presence of the RFLP is linked to the ß-A chromosome, as the unaffected sister is homozygous for the presence of the RFLP. It may be inferred that the ß-Th chromosome inherited from the carrier mother is associated with the absence of the RFLP. Half shaded symbols represent ß-thalassaemia carriers and unshaded symbols represent unaffected individuals.

Table 4.1 β-A and β-Th haplotypes in Asian Indians

Haplotype	β-A	β-Th
+---- ++	41 (21%)	95 (23%)
-+-++ ++	25 (13%)	44 (10%)
-++-+ ++	21 (11%)	9 (2%)
-++-- ++	3 (2%)	--
-+--+ ++	1 (0.5%)	--
++-++ ++	4 (2%)	--
----- ++	1 (0.5%)	1 (0.5%)
+++-+ ++	1 (0.5%)	--
+---- +-	15 (8%)	38 (9%)
-+-++ +-	4 (2%)	5 (1%)
-++-+ +-	5 (3%)	3 (1%)
-++-- +-	2 (1%)	--
-+--+ +-	1 (0.5%)	--
++-++ +-	4 (2%)	--
+---- -+	**44 (22%)**	**194 (46%)**
-+-++ -+	10 (5%)	4 (1%)
-++-+ -+	12 (6%)	22 (5%)
-++-- -+	1 (0.5%)	--
-+--+ -+	1 (0.5%)	3 (1%)
+-+-+ -+	--	1 (0.5%)
TOTAL	196	419

The percentage in brackets indicates the frequency of each haplotype amongst ß-A and ß-Th haplotypes. Highlighted in bold print is the most frequent β-globin gene haplotype.

Table 4.2 Frequency of β-globin gene frameworks amongst β-A and β-Th chromosomes

Framework	ß-A chromosomes	ß-Th chromosomes
framework 1	97 (50%)	148 (35%)
framework 2	31 (16%)	46 (11%)
framework 3	68 (35%)	225 (54%)
Total	196	419

ß-Th haplotypes were found to be less diverse than the ß-A ones. With the exception of +-+-+ -+, all the ß-Th haplotypes were represented amongst the ß-A haplotypes. Further, the frequency of each ß-Th haplotype was similar to its ß-A counterpart. However, the commonest haplotype, +---- -+, occurred at twice the frequency amongst ß-thalassaemia chromosomes (46%) as compared to normal ones (22%).

4.4.2 Linkage disequilibrium between mutations and haplotypes

A strong linkage disequilibrium between the ß-thalassaemia mutations and the ß-globin gene haplotypes was observed. Table 4.3 shows the number of chromosomes carrying each of the ß-thalassaemia mutations linked to a particular haplotype. The association between the mutations and haplotypes was not absolute and with the exception of some of the rare mutations, all were found to be associated with more than one haplotype. The association between the common mutations and their major haplotype is shown in table 4.4a. The linkage was strongest for the 619 bp deletion, being almost complete. The IVS-1 nucleotide 5 (G-C) and codons 41/42 (-CTTT) mutations were observed to have a lesser degree of linkage, 78% and 77%, respectively. Table 4.4b shows the extent of association between each of the five common ß-Th haplotypes and their predominant mutation, which was found to vary from 60% to 86%. The most frequent ß-Th haplotype, +---- -+, was associated with the IVS-1 nucleotide 5 (G-C) mutation in 60% of chromosomes and with the 619 bp deletion in 37%.

Table 4.3 Linkage disequlilbrium between β-thalassaemia mutations and β-globin gene haplotypes

Mutation Haplotype	IVS-1n5 (G-C)	c 8/9 (+G)	619 bp del	IVS-1n1 (G-T)	c 41/42 (-CTTT)	c 15 (G-A)	c 5 (-CT)	IVS-1-1 (G-C)	c 16 (-C)	IVS-2n837 (T-G)	cap site +1 (A-C)	IVS-1-1(G-A)	IVS-2n1 (G-A)	-88 (C-T)	c 88 (+T)
+---- ++	4	81		1	2	2		3				2			
-+-++ ++	1	9		33				1							
-++-+ ++					1	6		2							
+---- +-					27		4		3	3				1	
-+-++ +-					3		1						1		
-++-+ +-					2										1
+---- -+	117	2	71	2		1					1				
-+-++ -+	3					1									
-++-+ -+	19					1	1				1				
-+--+ -+	2		1												
+-+-+ -+	1														
----- -+	1														
Total	148	92	72	36	35	11	6	6	3	3	2	2	1	1	1

Thus by haplotype analysis it would be possible to predict the ß-thalassaemia mutation with 60% to 86% accuracy.

Table 4.4 Extent of mutation-haplotype linkage

Table 4.4a Linkage with main haplotype

Mutation	Haplotype	Linkage
IVS-1 nucleotide 5 (G-C)	+---- -+	78% (117/148)
codons 8/9 (+G)	+---- ++	88% (81/92)
619 bp deletion	+---- -+	99% (71/72)
IVS-1 nucleotide 1 (G-T)	-+-++ ++	92% (33/36)
codons 41/42 (-CTTT)	+---- +-	77% (27/35)

Table 4.4b Linkage with main mutation

Haplotype	Mutation	Linkage
+---- -+	IVS-1 nucleotide 5 (G-C)	60% (117/194)
+---- ++	codons 8/9 (+G)	85% (81/95)
-+-++ ++	IVS-1 nucleotide 1 (G-T)	75% (33/44)
+---- +-	codons 41/42 (-CTTT)	71% (27/38)
-++-+ -+	IVS-1 nucleotide 5 (G-C)	86% (19/22)

Figures in brackets represent the actual proportion of chromosomes.

Six ß-thalassaemia mutations, IVS-1 nucleotide 5 (G-C), codons 8/9 (+G), IVS-1 nucleotide 1 (G-T), codons 41/42 (-CTTT), codon 15 (G-A) and codon 5 (-CT) were found to be linked to two different ß-globin gene frameworks. Amongst the common mutations, only the 619 bp deletion was associated with a single framework, framework 3.

The -+-++ and the ++-++ 5' haplotypes have been found to be strongly linked with a *Xmn*I restriction site at -158 $^{G}\gamma$, which is associated with increased Hb F production (Thein et al, 1987). The -+-++ 5' haplotype was found in 53 out of 419 (12.7%) ß-Th chromosomes, including 33 out of 36 (92%) chromosomes carrying the IVS-1 nucleotide 1 (G-T) mutation. Linkage of the IVS-1 nucleotide 1 (G-T) mutation with a haplotype associated with increased Hb F production might suggest that this mutation could be associated with a milder form of the disease. The ++-++ 5' haplotype was not found in a single ß-Th chromosome. In contrast, the -+-++ and ++-++ 5' haplotypes were present in proportionately twice as many ß-A chromosomes, 39 (20%) and 8 (4%) of the 196 ß-A chromosomes analysed, respectively.

4.4.3 Haplotype analysis in different regions

Of the 419 subjects, whose ß-Th haplotypes were constructed, 128 (31%) originated from Northwest Pakistan, 124 (30%) from Gujarat, 85 (20%) from Punjab, 67 (16%) from Sindh and the remaining 15 (3%) individuals originated from Maharashtra, Tamil Nadu, and Bengal and Bangladesh. Of the 196 subjects, whose ß-A haplotypes were constructed, 81 (41%) originated from Northwest Pakistan, 48 (24%) from Gujarat, 36 (18%) from Punjab, 22 (11%) from Sindh and the remaining 9 (5%) from Maharashtra, Tamil Nadu, and Bengal and Bangladesh. As the ß-globin gene haplotypes could be constructed in only a small number of subjects from Maharashtra, Tamil Nadu and Bangladesh these regions were excluded from further analysis. Seven of the 20 haplotypes observed in the population sample occurred at a frequency of over 3%. The distribution of these seven haplotypes amongst the ß-A and ß-Th chromosomes of carriers from Northwest Pakistan, Gujarat, Punjab and Sindh is shown in table 4.5.

The proportion of the total number of ß-A haplotypes in each region that were accounted for by the seven haplotypes, present in the population sample at a frequency of over 3% were as follows: Northwest Pakistan, 65 out of 81 (80%), Gujarat, 43 out of 48 (90%), Punjab, 33 out of 36 (92%) and Sindh, 20 out of 22 (91%). Regional differences in the distribution of ß-A haplotypes were observed. For example, the +---- ++ haplotype accounted for a

Table 4.5 Regional distribution of the common β-A and β-Th haplotypes on the Indian subcontinent

Region	N.-W. Pakistan		Gujarat		Punjab		Sindh	
Haplotype	β-A	β-Th	β-A	β-Th	β-A	β-Th	β-A	β-Th
+---- ++	18 (25%)	58 (45%)	8 (17%)	11 (9%)	7 (19%)	14 (17%)	8 (36%)	10 (15%)
+---- +-	8 (10%)	14 (11%)	2 (4%)	8 (6%)	2 (5%)	12 (14%)	2 (9%)	1 (1%)
+---- -+	15 (19%)	36 (28%)	11 (23%)	77 (62%)	13 (36%)	32 (37%)	2 (9%)	40 (60%)
-+-++ ++	9 (10%)	10 (8%)	6 (13%)	13 (11%)	5 (14%)	8 (9%)	4 (18%)	13 (19%)
-+-++ -+	- -	2 (2%)	6 (13%)	- -	1 (3%)	1 (1%)	3 (14%)	- -
-++-+ ++	9 (12%)	4 (3%)	6 (13%)	4 (3%)	3 (6%)	1 (1%)	1 (5%)	- -
-++-+ -+	6 (7%)	4 (3%)	4 (8%)	6 (5%)	2 (5%)	12 (14%)	- -	- -
Total no. studied	81	128	48	124	36	85	22	67

Percentages in brackets represent the proportion of chromosomes in each region with that particular haplotype.

greater proportion of ß-A haplotypes in carriers from Sindh, as compared to the other regions. Similarly, the +---- -+ haplotype was most frequent in Punjab and the -+-++ -+ haplotype in Gujarat and Sindh.

The proportion of ß-Th haplotypes in each region accounted for by the seven haplotypes present in the population sample at a frequency of over 3% was as follows: 100% in Northwest Pakistan, 119 out of 124 (96%) in Gujarat, 80 out of 85 (94%) in Punjab and 64 out of 67 (95%) in Sindh. All of the ß-Th haplotypes observed in each of the regions were also represented amongst the ß-A haplotypes in that region, the single exception being the -+-++ -+ haplotype in Northwest Pakistan. Within each region, a particular ß-Th haplotype accounted for the majority of ß-thalassaemia chromosomes, the exception being Punjab. The predominant regional ß-Th haplotypes were as follows: +---- ++ constituting 45% of ß-Th haplotypes in Northwest Pakistan, +---- -+ constituting 62% of ß-Th haplotypes in Gujarat and 60% of ß-Th haplotypes in Sindh. Each of these predominant ß-Th haplotypes was also present amongst the ß-A chromosomes in that region, but were represented at a substantially greater frequency amongst ß-Th chromosomes as compared to the ß-A ones.

The regional distribution of the five common ß-thalassaemia mutations carried by the ß-Th chromosomes that were haplotyped is shown in table 4.6 and it is these frequencies that are considered in the following analysis. It was observed that the predominant ß-Th haplotype in each region was the one that was most closely associated with the regions predominant mutations. In Northwest Pakistan, the +---- ++ haplotype that accounted for 45% of the ß-Th chromosomes was associated with the codons 8/9 (+G) mutation, which accounted for 48% of ß-thalassaemia mutations in this region. +----- -+ was the haplotype that constituted 62% of ß-Th haplotypes in Gujarat and was associated with the IVS-1 nucleotide 5 (G-C) mutation and the 619 bp deletion in 97% of cases; it is these two mutations that constituted 66% of Gujarati ß-thalassaemia alleles. In Sindh, the +---- -+ haplotype constituted 60% of ß-Th haplotypes, the 619 bp deletion and IVS-1 nuceotide 5 (G-C) associated with this haplotype, accounted for 48% and 18% of the mutations present, respectively.

Table 4.6 Regional distribution of common ß-thalassaemia mutations

Region Mutation	N.W. Pakistan	Gujarat	Punjab	Sindh	Whole population
IVS-1 nt 5 (G-C)	40 (31%)	53 (43%)	33 (40%)	12 (18%)	148 (35%)
codons 8/9 (+G)	62 (48%)	7 (6%)	15 (18%)	7 (10%)	92 (22%)
619 bp deletion	--	29 (23%)	11 (13%)	32 (48%)	72 (17%)
IVS-1 nt 1 (G-T)	2 (2%)	16 (13%)	7 (8%)	11 (16%)	36 (9%)
codons 41/42 (-CTTT)	11 (9%)	11 (9%)	10 (12%)	3 (4%)	35 (8%)
Total no. studied	128	124	85	67	419

Percentages in brackets represent the proportion of carriers from each region carrying that particular mutation.

In Punjab the most common haplotype, +---- -+ was found to be present at similar frequencies amongst both the ß-Th and ß-A chromosomes, 37% and 36% respectively. However two of the ß-Th haplotypes, +---- +- and -++-+ -+, were present at a much higher frequency amongst the ß-Th chromosomes as compared to the normal ones. The +---- +- is in linkage disequilibrium with the codons 41/42 (-CTTT) mutation, which accounts for 12% of the mutations present in Punjab. Further, amongst the regions studied, the codons 41/42 (-CTTT) mutation was most frequent in Punjab. The -++-+ -+ haplotype was in linkage disequilibrium with the IVS-

1 nucleotide 5 (G-C) mutation as 19 of the 22 (86%) chromosomes with this haplotype carried this mutation. Ten of the 19 ß-Th haplotypes associated with the IVS-1 nucleotide 5 (G-C) mutation were from Punjabi carriers.

In summary, in Northwest Pakistan, Gujarat, Punjab and Sindh, the ß-A haplotypes were found to be more diverse than the ß-Th ones, similar to that found in the population as a whole. In each of these regions, the distribution of ß-Th haplotypes reflected that of its prevalent ß-thalassaemia mutations. The haplotypes associated with the common mutations in each region occurred at a substantially greater frequency amongst ß-Th chromosomes as compared to ß-A chromosomes. No regional differences in mutation-haplotype associations were observed. A difference in the regional distribution of mutations was present, but these mutations remained linked with the same haplotypes irrespective of the regional origin.

4.5 Discussion and Conclusions

The analysis of 615 ß-globin gene haplotypes and mutations of ß-thalassaemia carriers from different regions of the Indian subcontinent revealed clear evidence of local genetic differentiation, allowed inferences to be made about the origin and spread of particular ß-thalassaemia mutations, and suggested that certain globin gene variants may be useful anthropological markers in this region.

4.5.1 Origin of ß-thalassaemia mutations on the Indian subcontinent

The similarity of ß-Th and ß-A haplotypes on the Indian subcontinent and in each of the regional groups suggests that ß-thalassaemia mutations may have arisen relatively recently on chromosomal backgrounds already existing in this population. As described in other populations where ß-thalassaemia is prevalent, a close but not invariant association of each ß-thalassaemia mutation with a haplotype (Kazazian et al, 1984a; Cheng et al, 1984 and Antonarakis et al, 1984) was observed. The haplotype-mutation associations described in this chapter are in agreement

with those reported in previous studies on the Asian Indian immigrant population (Kazazian et al, 1984b; Thein et al, 1988).

The association of all mutations, present at a frequency of over 1%, with more than one haplotype and the association of six mutations with more than one ß-globin gene framework is of interest in relation to the origins of these mutations. Linkage of a particular mutation with different 5' haplotypes but the same 3' haplotype or framework, is consistent with a single origin followed by spread to the other chromosomes by recombination within a crossover site located in the region of the recombination hotspot lying between the 5' end of the δ-globin gene and the ß-globin gene (Chakravarti et al, 1984). The association of a particular mutation with two different frameworks could imply that the mutation has arisen independently at least twice, as has been argued for the sickle cell mutation in Africa (Chebloune et al, 1988). Further, hotspots for mutational events have been identified, this could explain the multiple occurrence of a mutation, each time on different ß-globin gene frameworks, as has been suggested for the codon 39 (C-T) mutation (Pirastu et al, 1987). Alternatively, the association of a mutation with different ß-globin gene frameworks may be explained by interallelic gene conversion events which have been well documented in globin genes (Slightom et al, 1980). However, when more than one site is different in the haplotypes associated with a mutation, it has been proposed that multiple cross over events in short stretches of DNA are statistically less probable than the independent recurrence of that mutation at the same position (Antonarakis et al, 1982b).

In spite of significant differences in the regional distribution of ß-thalassaemia mutations, no regional differences in haplotype-mutation associations were found; this argues against multiple independent origins of these mutations in the different regions of the Indian subcontinent. Further, the number of chromosomes found to be associated with the rare framework were very few, therefore gene conversion events would be a more likely explanation, rather than multiple independent origins. However, there could be an alternative explanation for the IVS-1 nucleotide 5 (G-C) and codons 41/42 (-CTTT) mutations. IVS-1 nucleotide 5 (G-C) was found to be predominantly linked with framework 3 on the Indian subcontinent and only five (3.5%) of the chromosomes

carrying this mutation were associated with framework 1. Similarly, codons 41/42 (-CTTT) was found to be mainly associated with framework 2 and three (9.4%) chromosomes were associated with framework 1. IVS-1 nucleotide 5 (G-C) is common in Melanesia in association with framework 1 (Hill et al, 1988) and codons 41/42 (-CTTT) is common in South China in association with framework 1 (Kazazian et al, 1986; Chan et al, 1987). It may thus be argued that these two mutations have had more than one origin and the chromosomes carrying IVS-1 nucleotide 5 (G-C) and codons 41/42 (-CTTT) in association with framework 1 were introduced on the Indian subcontinent by population migration from Melanesia and South China. Of the five individuals who carried the IVS-1 nucleotide 5 (G-C) mutation in association with framework 1, two were from Punjab, two from Sindh and one from Northwest Pakistan. Of the three individuals with codons 41/42 (-CTTT) in association with framework 3, two were from Gujarat and one from Northwest Pakistan. As these carriers were not confined to a single region, gene conversion events seem unlikely, but cannot be ruled out.

The strong linkage disequilibrium between the 619 bp deletion, IVS-1 nucleotide 1 (G-T) and codons 8/9 (+G) mutations and their ß-globin gene haplotypes suggests a relatively recent origin of these mutations. Furthermore, their limited spread is suggestive of a local origin in certain regions. The 619 bp deletion was observed to have the strongest linkage with its ß-globin gene haplotype, 71 out of 72 chromosomes were found to be associated with the +----- -+ haplotype. The single exception was associated with the -+--+ -+ haplotype, which was probably due to a recombination event occurring in the hotspot.

On the other hand, the IVS-1 nucleotide 5 (G-C) and codons 41/42 (-CTTT) mutations were less closely associated with their major haplotype and more widely distributed on the Indian subcontinent. The great diversity of haplotypes associated with the IVS-1 nucleotide 5 (G-C) mutation, its high frequency and widespread distribution, all suggest that it may be the oldest mutation on the Indian subcontinent.

4.5.2 Positive selection experienced by ß-thalassaemia carriers

On the Indian subcontinent as a whole and within each of the regions studied, the predominant ß-Th haplotype occurred at a substantially greater frequency than its normal counterpart. The +---- -+ haplotype, associated with the commonest Asian Indian ß-thalassaemia mutation, IVS-1 nucleotide 5 (G-C), occurred in 46% of ß-Th chromosomes and in only 22% of normal ones. The same phenomeon was observed in each regional "micro-population" studied, an outstanding example being the presence of the +---- -+ haplotype in 60% of ß-Th chromosomes in Sindh and in only 9% of its ß-A chromosomes. A greater frequency of the predominant ß-Th haplotype as compared to its ß-A counterpart occurred irrespective of the associated ß-thalassaemia mutation. The codons 8/9 (+G) mutation in Northwest Pakistan associated with the +---- ++ haplotype, the 619 bp deletion in Sindh associated with the +---- -+ haplotype, the codons 41/42 (-CTTT) mutation in Punjab associated with the +---- +- haplotype and the IVS-1 nucleotide 5 (G-C) mutation in Gujarat and Sindh associated with the +---- -+ haplotype and in Punjab with the -++-+ -+ haplotype, were all observed to have the same effect.

These observations are consistent with the hypothesis that ß-thalassaemia mutations arose on chromosomal backgrounds already existing in the population that inhabited the Indian subcontinent and, having arisen, caused their chromosomes to experience positive selection pressure, both in the population as a whole and within each regional group. Positive selection occurred irrespective of the type of ß-thalassaemia mutation. Although endogamy is common in the Indian population, genetic drift is unlikely to have contributed substantially to the predominant ß-Th haplotype occurring at a greater frequency than its normal counterpart, in continental populations of this size (e.g. Gujarat has a population of about 60 million). It was Haldane (1949) who first suggested that high frequencies of thalassaemia could be due to selection by malaria and evidence for a geographic relationship between ß-thalassaemia and endemic malaria was obtained in Sardinia (Siniscalco et al, 1966). More direct evidence for protection of ß-thalassaemia heterozygotes from malaria was

provided by Willcox et al (1983) in West Africa. Thus it is likely that protection against malaria in the ß-thalassaemia carrier was responsible for positive selection of ß-Th chromosomes on the Indian subcontinent where malaria has been endemic.

4.5.3 ß-thalassaemia mutations as population markers

The ethnic composition of the Asian Indian population is varied with six main races described in prehistoric India (Hutton, 1931) before the arrival of the Aryan-speaking tribes from Iran into the Northwestern region of the Indian subcontinent in 1500 BC (Thapar, 1966). This is reflected in the ethnographic complexity of the population that now inhabits the Indian subcontinent. The strict endogamy required by the caste system has perpetuated local genetic diversity, so that the Indian subcontinent has not been a "melting pot" but remained a composite of a number of distinct genetic units. Each regional group therefore represents a genetically isolated "micropopulation". This extreme and long standing genetic heterogeneity is a distinctive feature of the Asian Indian population accounting to a certain extent for the uneven and variable distribution of various genetic markers such as the blood group antigens (Mourant, 1983), the haemoglobinopathies (Brittenham, 1983) and also for the regional variation at the molecular level of the ß-globin gene. The mutations characteristic of this regional variation could serve as useful anthropological markers for local populations.

Examples of the considerable variation in regional distribution of ß-thalassaemia mutations on the Indian subcontinent are the predominance of the codons 8/9 (+G) mutation in Northwest Pakistan and that of the 619 bp deletion in Sindh. Another example is the restriction of the IVS-1 nucleotide 1 (G-T) mutation to the neighbouring regions of Gujarat, Sindh and Punjab. The strong linkage disequilibrium between the 619 bp deletion and +-- -- -+ haplotype and the geographical distribution of this mutation suggests that the 619 bp deletion arose in Sindh and subsequent spread has been by gene flow through population migration; this ß-thalassaemia allele could thus serve as a useful anthropological marker for population movements on the Indian subcontinent. The codons 8/9 (+G) and the IVS-1 nucleotide 1 (G-T) mutations are

similarly useful. Two ß-Th chromosomes with the -+-++ -+ haplotype and the IVS-1 nucleotide 1 (G-T) mutation were found in Northwest Pakistan, where this haplotype is not represented amongst the ß-A chromosomes and where this mutation is rare. It may be inferred that this mutation was introduced in Northwest Pakistan by population movements from Gujarat, Punjab or Sindh.

Besides being useful population markers for anthropological studies on the Indian subcontinent, ß-thalassaemia mutations also serve as useful indicators of gene flow both into and from the subcontinent. Throughout its recorded history the Indian subcontinent has had political, military and commercial interactions with Central, Western and Southeast Asia and later Europe (Thapar, 1966). This is likely to have resulted in gene flow causing the spread of some ß-thalassaemia alleles. The IVS-2 nucleotide 1 (G-A) mutation was identified in an individual from Punjab on the same haplotype as has been described in Turkey and Northern Cyprus (Diaz-Chico et al, 1988) which could be explained by the Turkish invasion of Northern India in the 11th century (Thapar, 1966). The codon 5 (-CT) mutation has been described in Greeks on a haplotype identical to that found in the Asian Indian population (Kollia et al, 1989) also suggesting a common origin with subsequent gene flow. Table 4.7 shows the haplotype associations of ß-thalassaemia mutations present in Southeast Asia which could have spread from India by population migration and figure 4.10 shows the geography of this area. Analysis of the haplotype-mutation associations in this region suggests that the IVS-1 nucleotide 5 (G-C), and IVS-1 nucleotide 1 (G-T) mutations spread to Burma, Malaysia and Indonesia from India. Similarly, the codon 15 (G-A) and IVS-1 minus 1 (G-C) mutations spread to Indonesia. However the codons 41/42 (-CTTT) mutation in Burma, Thailand, Indonesia and South China probably had an independent origin. Although some chromosomes with this mutation in Malaysia could have arisen on the Indian subcontinent. As early as the 6th century BC, Indian sea-traders were exploring the coasts of Burma, the Malay peninsula and western Indonesia; this continued for centuries, with some of these traders later marrying local women (Ma, 1985), offering possible routes for the spread of these alleles.

Table 4.7 Haplotype associations of some Southeast Asian ß-thalassaemia mutations

Mutation	India	Burma[a]	Thailand[b]	Malaysia[c]	Indonesia[d]
IVS-1 n 5 (G-C)	+---- -+*	+---- -+		+---- -+*	+---- -+
	+---- ++			+---- ++	+---- ++
c 41/42 (-CTTT)	+---- +-*	+---- -+	+---- -+	+---- +-*	+---- ++
	+---- ++	+---- ++	+---- ++	+---- ++	
IVS-1 n1 (G-T)	-++-+ ++*	+---- ++		+---- ++	-++-+ ++
	+---- ++				
c 15 (G-A)	-++-+ ++*				+---- -+
	+---- ++				+---- ++
	+---- -+				
IVS-1-1 (G-C)	+---- ++*				+---- ++

* indicates the predominant haplotype
a = Brown, 1990; b = Lynch, 1987; c = Yang et al, 1989; d = Lie-Injo et al, 1989. In South China, the codons 41/42 (-CTTT) mutation is predominantly associated with the +---- ++ haplotype (Kazazian et al, 1986; Chan et al, 1987)

Further detailed microepidemiological analysis of ß-thalassaemia on the Indian subcontinent including the isolated tribal groups should be of value in providing evidence of historical gene flow in this region and in defining local population relationships.

4.5.4 RFLP studies: application to prenatal diagnosis

Establishment of linkage between a haplotype or an RFLP and the ß-thalassaemia gene in a particular family can be used for prenatal diagnosis. In order to perform prenatal diagnosis by RFLP analysis, DNA polymorphisms have to be assigned to the normal and ß-Th

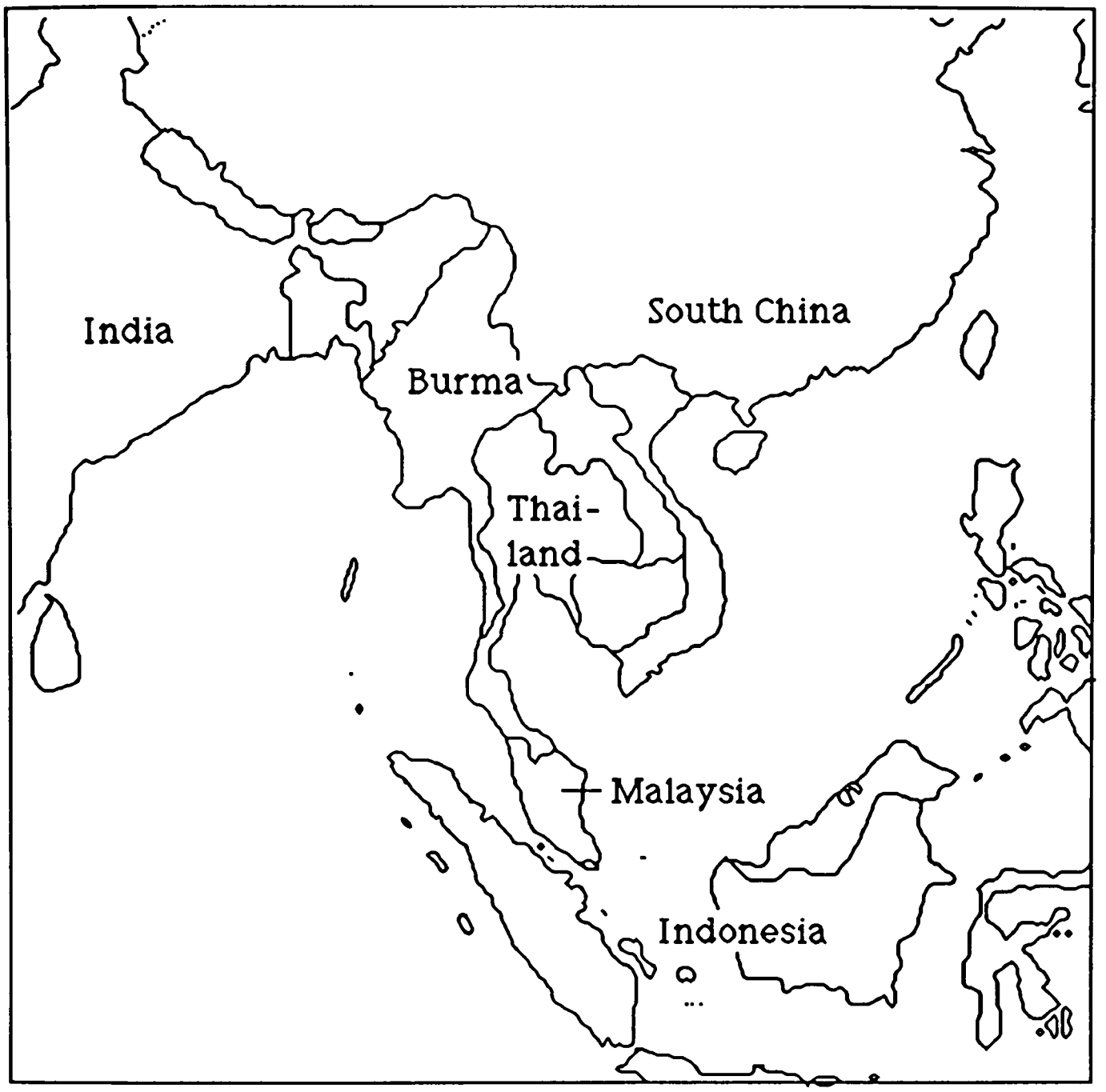


Figure 4.10 Map of Southeast Asia showing the regions discussed in Table 4.7.

chromosomes in a couple at risk by family studies. RFLP analysis by restriction enzyme digestion of the PCR product has greatly simplified such linkage studies. Using the primer pairs described in this chapter it would be possible to rapidly study the RFLPs and construct the ß-globin gene haplotypes. The heterogeneity of ß-globin gene haplotypes in the Asian Indian population would facilitate linkage of an RFLP to the ß-thalassaemia gene. This would assist in tracking the inheritance of the disease gene in a given family, thus being advantageous for prenatal diagnosis.

The main disadvantage of performing prenatal diagnosis by RFLP linkage is that the assignment of RFLPs to ß-Th and ß-A chromosomes may not be possible in the first pregnancy or in couples who have only children with the thalassaemia trait. In these cases linkage could be established by studying both sets of grandparents and other family members. Availability of samples for analysis from family members could be difficult or impossible. Another disadvantage of using RFLP linkage for prenatal diagnosis is that nonpaternity maybe a cause of error; if the true paternal chromosome has a common haplotype similar to that of the husband, nonpaternity could remain undetected leading to misdiagnosis. Lastly, a cause of error could be meiotic recombination between the polymorphism and the ß-thalassaemia mutation. It is estimated that crossing over could occur in the hotspot, 5' to the ß-globin gene once in every 350-400 meioses (Chakravarti et al, 1984).

Prenatal diagnosis of ß-thalassaemia by mutation detection remains the most appropriate approach, however, RFLP linkage studies have an important role in confirming the results of mutation detection studies. The risks of misdiagnosis would be minimised if corroborative results of two independent tests were obtained in all cases. It would be ideal to perform prenatal diagnosis by both approaches simultaneously, particularly as the PCR based method of RFLP analysis is rapid and less labour intensive than Southern blot hybridisation studies. Furthermore, in the occasional case where the ß-thalassaemia mutation cannot be identified, RFLP linkage studies would be valuable for prenatal diagnosis. A knowledge of the haplotype-mutation associations described in this chapter will be useful to a ß-thalassaemia

prenatal diagnosis program in the Asian Indian population, which combines both mutation detection and RFLP linkage analysis.

CHAPTER FIVE

α GLOBIN GENE MAPPING

5.1 Introduction

For ß-thalassaemia control in the Asian Indian population it is necessary to know the prevalence and molecular basis of α-thalassaemia in this population. There are two important reasons for this. Firstly, ß-thalassaemia carriers who also have α-thalassaemia show a reduction in amount of chain imbalance and have milder haematological abnormalities (Kanavakis et al, 1982); their Hb levels, MCV and MCH are all higher compared to other ß-thalassaemia carriers. In a population with both α and ß thalassaemia, a screening programme based on MCV and MCH measurements using electronic cell counters, would fail to detect some ß-thalassaemia heterozygotes, as in ß-thalassaemia carriers with the αα/-α genotype there is some overlap with normal values. Thus in these populations it is essential to use a relatively high MCH cut off point. Secondly, in a population in which both α and ß thalassaemia are common it is possible for a ß-thalassaemia homozygote to co-inherit α-thalassaemia. α-Thalassaemia is an important factor in reducing the severity of homozygous ß-thalassaemia resulting in a milder clinical phenotype (Kan and Nathan, 1970). Much of the pathology of ß-thalassaemia is due to the pool of excess α-globin chains; as α-globin chain synthesis is reduced in α-thalassaemia, an interaction of the two disorders ameliorates the severity of the clinical phenotype. Gene mapping studies have established that α-thalassaemia interacts with homozygous ß-thalassaemia to result in a clinical picture of thalassaemia intermedia (Wainscoat et al, 1983). However, the extent of amelioration of the clinical phenotype depends on the severity of the ß-thalassaemia mutation, the nature of the α-thalassaemia determinant and the coinheritance of genetic determinants that increase Hb F production. While performing prenatal diagnosis it is important to be able to identify α and ß

thalassaemia interactions in the fetus and inform parents that their child would possibly have a mild clinical disorder.

Analysis of Hb Bart's in cord blood samples from various parts of India suggest that in the general population, α-thalassaemia is prevalent at frequencies varying from 0.5 to 4% (Swarup et al, 1965; ICMR collaborative study, 1988). These figures would be a underestimate, as a number of individuals with the αα/-α genotype will not have a raised Hb Bart's level at birth (Weatherall and Clegg, 1981). In some of the isolated, endogamous, tribal populations in India, the reported frequency of α-thalassaemia is even higher (Brittenham, 1981). However as yet the prevalence of α-thalassaemia on the Indian subcontinent remains largely unknown. Hb Bart's hydrops fetalis has not been reported in India and there have been only anecdotal reports of Hb H disease in the Indian population (Chatterjea, 1961; Chouhan et al, 1970; Agarwal and Mehta, 1982).

The molecular basis of α-thalassaemia in Asian Indians has not been studied in detail. Molecular studies on 102 chromosomes of Kachari subjects from Assam, found nine α^{+}-thalassaemia heterozygotes, eight of which had the rightward deletion and one had the leftward one. The frequency of the single α-globin gene deletion was 8.8% in this group (Hundreiser et al, 1987). The α-globin genotype was determined by gene mapping studies in 282 individuals from families with sickle cell disease belonging mainly to tribal groups in Orissa. The overall frequency of α-thalassaemia was 29% which was mainly caused by the $-\alpha^{3.7}$ and the $-\alpha^{4.2}$ deletions, the $-\alpha^{3.5}$ deletion was found in a single family and α°-thalassaemia deletions were not detected (Kulozik et al, 1988). In a small group of apparently healthy individuals from the tribal communities of Koya Dora and Konda Reddi, in Andhra Pradesh, a high frequency of the $-\alpha^{3.7}$ and the $-\alpha^{4.2}$ deletions, and a nondeletional α-thalassaemia determinant, Hb Koya Dora were found, but the -- α-thalassaemia haplotype was not observed (Fodde et al, 1988). A deletion of approximately 23 kb from the α-globin gene cluster, including both the α-globin genes, termed as $--^{SA}$, has been described in an individual from Vayra, in Gujarat (Drysdale and Higgs, 1989). The apparently low frequency of the -

-, α-thalassaemia haplotype in the Asian Indian population would explain the absence of Hb Bart's hydrops fetalis.

The preliminary studies of the molecular basis of α-thalassaemia in the Asian Indian population are described in this chapter.

5.2 Subjects

The population sample studied comprised 1240 unrelated adults who all belong to the Asian Indian immigrant community in East London. They were referred for routine haematological investigations by their general practioners and from antenatal clinics. An additional 10 ml of blood was collected from each of these individuals for the purposes of this study. Sample collection and the haematological investigations were performed by Dr Louise Tillyer and the staff at the Department of Haematology, Newham General Hospital, London.

The experiments described here, contribute to one aspect of a large study conducted by Dr Louise Tillyer on the prevalence of Hb disorders in the Asian Indian immigrant population in the UK. Following haematological investigations, α-thalassaemia was suspected in 51 individuals. These individuals had a reduced MCH, less than 27 pg (normal range, 27-34 pg), a normal Hb A_2 (normal range, 1.5-3.5%) and a serum ferritin level above 20 μg/l (normal range, 20-300 μg/l). Buffy coat samples of these 51 suspected cases of α-thalassaemia were transported to Oxford for α-globin gene mapping studies.

5.3 Methods

DNA was extracted from the buffy coat sample of 51 subjects in whom α-thalassaemia was suspected, as described in section 2.2. The presence of the single α-globin gene deletion was sought in each subject by restriction enzyme digestion followed by Southern blot hybridisation.

5.3.1 Southern blot hybridisation

Five µg of genomic DNA from each subject was digested with *Bam*HI. Digestion was carried out at 37°C for a minimum of four hours using 20 units of the enzyme and 2 µl of the appropriate buffer, supplied by the manufacturer. The digested DNA fragments were separated according to size by electrophoresis in a 0.8% agarose gel and then transferred onto a nylon filter (Hybond N+, Amersham) by Southern blotting (section 2.3). The filters were hybridised with an α-globin gene probe, 1.5 kb α *Pst*I fragment in pBR322 (gift from Dr DR Higgs) which was radiolabelled with [α^{32}P]dCTP using a Multiprime Labelling Kit (Amersham). Hybridisation and washing was done as described in section 2.3.4. Filters were wrapped in cling film and exposed overnight to X-ray film at -80°C using cassettes with intensifying screens.

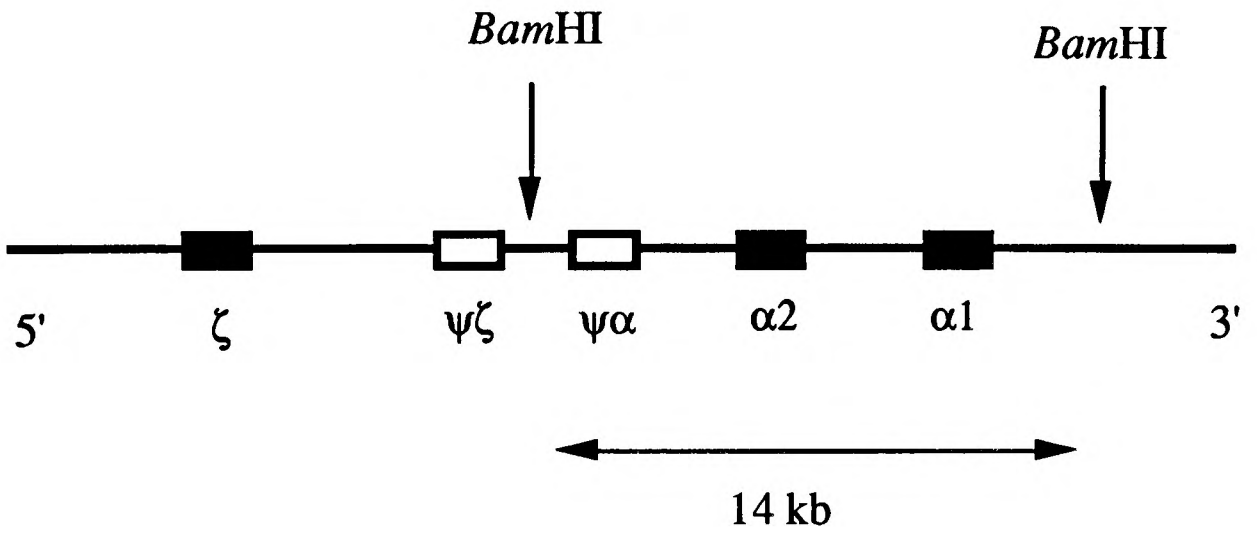
5.4 Results

Normal DNA when digested with *Bam*HI and hybridised with the α-globin gene specific probe, produced a single 14 kb fragment containing both α-globin gene loci, whereas a 10.5 kb fragment is generated when one of the two α-globin loci has been deleted. Visualisation of a 14 kb and 10.5 kb fragment indicates a deletion of one of the α-globin genes on one chromosome i.e. the αα/-α genotype. A single 10.5 kb fragment would be consistent with -α/-α genotype, deletion of two α-globin genes in *trans* (figure 5.1).

Of the 51 individuals in whom α-globin gene mapping studies were performed, 15 individuals had a single α-globin gene deleted on one chromosome, αα/-α, and nine individuals had the -α/-α genotype. The remaining 27 individuals with suspected α-thalassaemia did not have the single α–globin gene deletion.

5.5 Discussion and Conclusions

The studies on the Asian Indian immigrants in the UK described here are amongst the first that attempt to comprehensively describe the nature of α-thalassaemia in this population. The



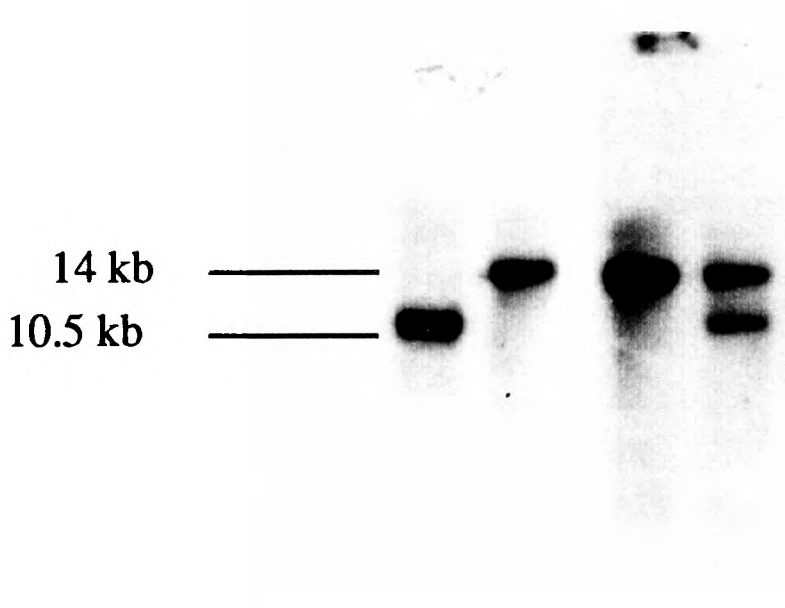


Figure 5.1 α-globin gene mapping with *Bam*HI. The upper half of the figure is a schematic representation of the α-globin gene cluster showing the position of the *Bam*HI restriction sites. The lower half is an autoradiograph showing the DNA fragments obtained following *Bam*HI digestion and Southern blot hybridisation with a *Pst*I α-globin gene probe. The marker is λ bacteriophage DNA digested with *Hind*III (not shown).

preliminary DNA studies show that 24 of the suspected cases of α-thalassaemia had the single α-globin gene deletion. As Hb H disease and Hb Bart's hydrops fetalis are very rare it is unlikely that the α-globin gene deletion *in cis* (--) would be frequent. Thus it appears that the remaining 27 suspected α-thalassaemia carriers have a nondeletional α-thalassaemia determinant.

In order to estimate the frequency of the single α-globin gene deletion in this population it would be necessary to look for this deletion in the remaining population sample in whom α-thalassaemia was not suspected on the basis of haematological investigations. As it is very likely that a number of individuals with normal MCH values, the ß-thalassaemia carrier state or iron defeciency have coexistent α-thalassaemia. The detailed molecular studies which will determine the nature of the nondeletional α-thalassaemia determinants in these carriers will prove to be significant as they will contribute to an understanding of the molecular pathology of the condition. Further it will be interesting to look for regional differences in the distribution of these determinants.

Knowledge of the frequency and molecular basis of α-thalassaemia in the Asian Indian population will be valuable in planning strategies for thalassaemia control in this population.

CHAPTER SIX

AN APPROACH TO PREIMPLANTATION DIAGNOSIS OF ß-THALASSAEMIA

6.1 Introduction

In the past decade tremendous advances have been made in the antenatal diagnosis of genetic disorders, so that many couples who had no option but to limit their family following the birth of an affected child, can now have their desired healthy family. However to achieve this, genetic counselling and prenatal diagnosis must be performed in every pregnancy. Prenatal diagnosis of ß-thalassaemia involves chorion villus biopsy or fetal blood sampling in the first or second trimester of pregnancy respectively, both of which are invasive procedures associated with an increased risk of miscarriage; further, on diagnosis of a thalassaemia major fetus the pregnancy is usually terminated. These procedures could be replaced by preimplantation diagnosis which involves diagnostic testing for a genetic disorder performed on embryonic biopsies with subsequent transfer of only unaffected embryos into the uterus.

6.1.1 Indications for preimplantation diagnosis

Couples requiring prenatal diagnosis fall into two main categories, the first being those with a relatively low and sporadic risk of having a child affected by a genetic disorder, for example, Down's syndrome. The second group consists of couples with a high and recurrent risk of transmitting a genetic disease to their offspring; autosomal recessive disorders like ß-thalassaemia, cystic fibrosis and sickle cell anaemia, X-linked disorders like Duchenne muscular dystrophy and haemophilia or dominantly inherited diseases like Huntington's chorea all come under this category. In order to achieve a healthy family, couples with such a high and recurrent risk of having an affected offspring need to undergo prenatal diagnosis in every pregnancy. Most couples manage to have their

families with minimal trauma, however some are unlucky enough to have repeated prenatal diagnosis with unhappy outcomes, often remaining childless or managing to have only one healthy child. The problem is further compounded in the 10-15% of couples with reduced fertility. Individuals who have religious and ethical objections to termination of pregnancy find the prospect of prenatal diagnosis with the possibility of abortion unacceptable and choose to remain childless or risk having an affected child. There are a number of women who underwent sterilisation after the birth of an affected child, who now wish to take advantage of recent developments in gene diagnosis and have more children, this is particularly true for cystic fibrosis carriers in the UK (Coutelle et al, 1989). Couples with the special circumstances described above would greatly benefit from preimplantation genetic diagnosis which would help to overcome some of the dilemmas associated with prenatal diagnosis, reduced fertility and the termination of a wanted pregnancy.

6.1.2 Embryo biopsy

Micromanipulation of human oocytes and embryos has provided new opportunities both for the treatment of infertility and the preimplantation diagnosis of genetic disease. The ability of biopsied embryos to implant and develop into live offspring has been reported in mice (Monk et al, 1988) and marmosets (Summers et al, 1988). Established human pregnancies have also been reported following the transfer of biopsied embryos obtained by *in vitro* fertilisation (IVF) (Handyside et al, 1990). Biopsies of human embryos are performed either at the 6-10 cell stage (day 3) or at the blastocyst stage (day 5-6). In the former, a hole is drilled through the zona pellucida with acid Tyrodes (pH 2.4) applied locally from a fine micropipette, a larger micropipette is then introduced through the hole and a single eight-cell stage cell (blastomere) aspirated for analysis (Handyside et al, 1989). Trophectoderm biopsy is performed at the blastocyst stage; a slit is made in the zona pellucida opposite the inner cell mass, through which herniation of trophectoderm cells occurs 18-24 hours later. These herniated cells, usually about 10-30 in number, are separated by gently rubbing the end of a siliconised glass needle across the

waist of the herniation and against the bottom of the dish (Dokras et al, 1990).

Gene diagnosis performed at the preconceptional stage could circumvent some of the ethical difficulties of dealing with human embryos. By genetic analysis of the first polar body it would be possible to infer whether or not the oocyte carries the defective gene, thus only normal oocytes could be fertilised and replaced in the mother. To obtain the first polar body from unfertilised oocytes, the cumulus cells and the zona pellucida are removed by incubation in hyaluronidase and acid Tyrodes solution respectively, following which the first polar body can be separated from the oocyte by gentle pipetting with a micropipette (Monk and Holding, 1990). As such an analysis of spermatozoa is not possible, this approach to preconceptional diagnosis is applicable for only maternally inherited genetic defects.

6.1.3 Genetic analysis of the embryonic biopsy

A number of tests have been used for the genetic analysis of embryonic biopsies. Preimplantation sexing was first performed by the detection of sex chromatin in trophectoderm cells obtained by rabbit blastocyst biopsy (Gardner and Edwards, 1968). Since then a number of genetic disorders have been detected by the analysis of embryonic biopsies. Hypoxanthine phosphoribosyl transferase (HPRT) deficiency was diagnosed by a biochemical microassay of HPRT activity, performed in a single mouse blastomere providing a model for preimplantation diagnosis of the Lesch-Nyhan syndrome (Monk et al, 1987). Prerequisites for accurate diagnosis by a biochemical assay are that the embryonic gene must be expressed at the stage of biopsy. The success of preimplantation diagnosis by DNA analysis, however, does not depend on expression of the genes. *In situ* hybridisation with labelled DNA probes has been applied for preimplantation genetic diagnosis, for example sexing with Y-specific DNA probes (Jones et al, 1987). The rapid amplification of specific gene fragments from a single cell is possible with the PCR (Li et al, 1988) and this technique has been applied for the preimplantation diagnosis of a number of genetic diseases. Sexing of human embryos by the amplification of a repeated Y-specific sequence has been described (Handyside et al, 1989). Successful

amplification of the single copy ß-globin gene from mouse blastomeres (Holding and Monk, 1989), the first polar body and oocyte in humans (Monk and Holding, 1990) has been reported and the genes associated with Duchenne muscular dystrophy and cystic fibrosis have also been amplified from single human oocytes (Coutelle et al, 1989).

The experiments described in this chapter were aimed at developing an approach for the preimplantation diagnosis of ß-thalassaemia in humans by DNA amplification using the PCR. At the time these experiments were begun, there were no published reports describing preimplantation diagnosis by DNA amplification.

6.2 Experimental Development

6.2.1 Materials

In order to achieve the goal of amplification of a fragment of the single copy ß-globin gene from an embryonic biopsy, an initial series of experiments were done using experimental models. Firstly, diluted genomic DNA solutions were used; genomic DNA was prepared at 1 µg/µl concentration, this "stock solution" was serially diluted with sterile double distilled water by a series of tenths ranging from 1:10 to 1:100,000. One µl was used as the template in experiments to determine a suitable PCR protocol. Secondly, as an experimental model for whole human diploid cells, anticoagulated whole blood and peripheral blood mononuclear cell suspensions were used. One µl of whole blood was estimated to have 10,000 white blood cells, thus the 1:10 and 1:100 dilutions prepared had approximately 1000 and 100 nucleated cells per µl, respectively. A white blood cell suspension was prepared in normal saline from 8 ml of whole fresh blood by density gradient centrifugation on Ficoll-Paque and washed three times in normal saline before use. This suspension had approximately 1250 white blood cells per µl and was serially diluted by a factor of 1:10 and 1:100 to prepare suspensions for experimental use, each of which contained approximately 125 and 12.5 cells per µl respectively.

6.2.2 Extraction of DNA

Two methods were employed to extract DNA from small numbers of cells, using whole blood and cell suspensions as the experimental models. In the first method 50 µl of sterile double distilled water was added to the cell sample, which was then overlaid with 100 µl of light liquid paraffin (BDH) and boiled for 20 min to release the nuclear DNA. The lysate was then separated by centrifugation at 9000 g for 15 min and a 10 µl aliquot of the supernatant was used as a template for PCR (Kogan et al, 1987). In the second method, the cell sample was suspended in a final volume of 20 µl containing 0.05 mg/ml Proteinase K, 20 mM DTT (dithiothreitol) and 1.7 µM SDS, incubated for one hour at 37°C and then heated to 85°C for 1 min. The PCR was then performed by adding the various reaction components to this sample (Li et al, 1988). Both methods were successful in extracting the DNA but as boiling was relatively simple and had less potential for contamination, it was the method adopted.

6.2.3 PCR conditions

A fragment of the single copy ß-globin gene encompassing its first and second exons and first intron was selected for amplification, as this region of the gene is the site for most of the ß-thalassaemia mutations. Primers ß-1 and ß-2 (figure 6.1) were used in the PCR which was performed in a 100 µl reaction volume containing 20 pmol of each primer, 800 µM total dNTPs, 10 mM Tris HCl pH 8.3, 50 mM KCl and 1.5 mM $MgCl_2$ with 2 units *Taq* polymerase. Five reactions were performed with different amounts of template DNA: approximately 100 ng, 10 ng, 1 ng, 100 pg and 10 pg. The thermal cycling regimen consisted of 35 cycles of denaturation at 93°C for 1 min, annealing at 55°C for 1 min and extension at 72°C for 1.5 min, in the final cycle extension was prolonged for 3 min. A 10 µl aliqout of the PCR product was examined by electrophoresis in a 1% agarose gel which was stained by ethidium bromide and visualised by UV transillumination. The effect of increasing the number of cycles to 40, 45 and 50 was investigated. Nonspecific amplification, which was visualised as smearing on the ethidium bromide stained agarose gel, resulted, without satisfactory amplification of the fragment of interest (figure 6.2). The reason for this probably was

the loss of efficiency of *Taq* polymerase after 25-30 cycles. To overcome these problems amplification was attempted with two rounds of PCR each consisting of 30 cycles. A 2 µl aliquot of the first round PCR product was used as the template for the second round, which was performed using reaction conditions and primers identical to those used in the first round. With this approach it was possible to amplify the fragment of interest from approximately 100 pg of template DNA but there was still some evidence of nonspecific amplification. The protocol was further modified to allow nested PCR. In the second round of PCR nested primers ß-3 and ß-4, situated internally to the primers employed in the first round, were used (figure 6.1). All reaction conditions remaining similar to those described above.

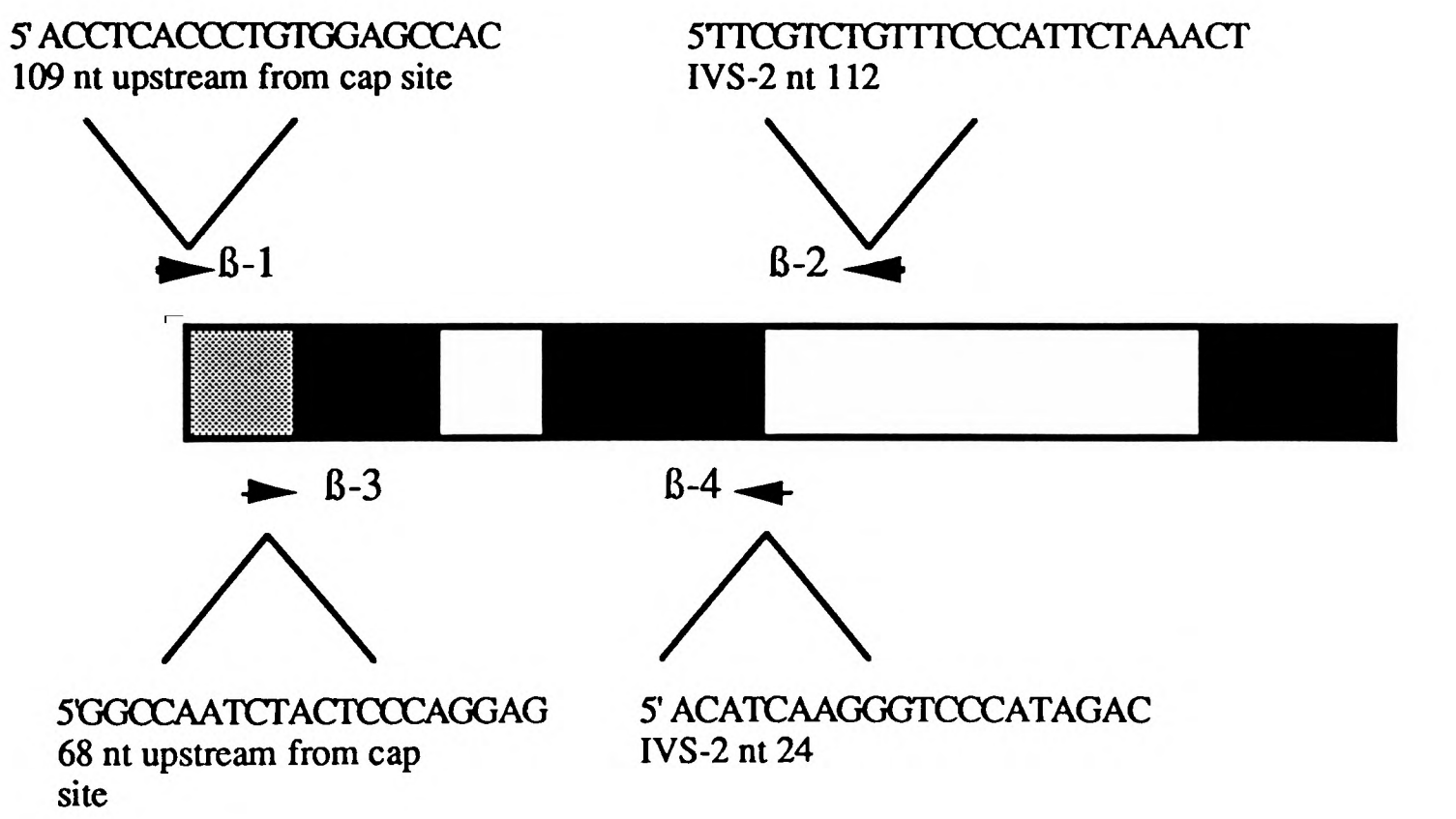


Figure 6.1 Schematic representation of the ß-globin gene showing the primers used for amplification of a 597 bp gene fragment by nested PCR. The position of the 5' nucleotide of each primer is indicated below its sequence. The shaded areas represent exons, the clear areas, introns and the 5' flanking region is cross hatched.

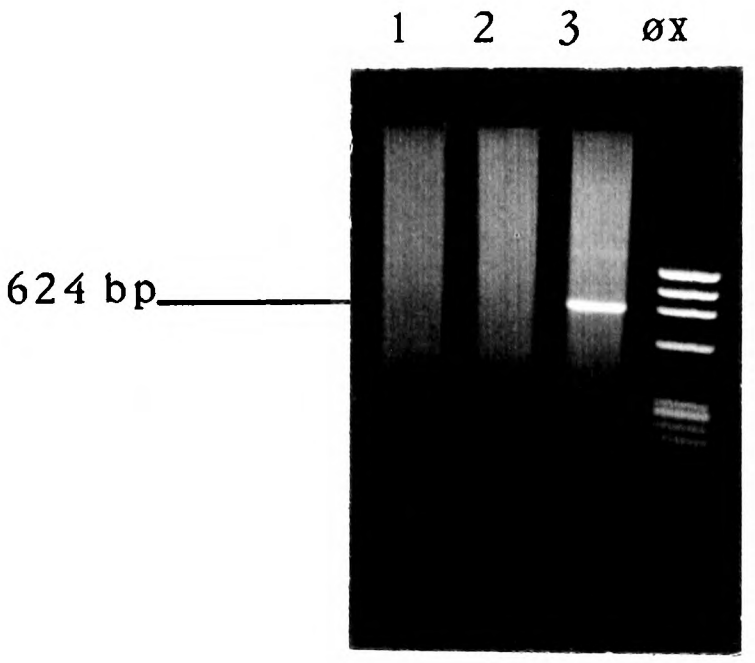


Figure 6.2 Amplification of a ß-globin gene fragment using 50 cycles of PCR. Ethidium bromide stained agarose gel showing the PCR products; the amount of template DNA in lane 1 was 100 pg, lane 2, 1 ng and lane 3, 10 ng. The expected 624 bp fragment is present in only lane 3. Nonspecific amplification visualised as smearing is observed. The marker is øx 174 DNA digested with *Hae*III.

Two rounds of PCR using nested primers successfully amplified a 597 bp fragment of the ß-globin gene from as little as approximately 100 pg of template DNA with almost no evidence of nonspecific amplification, which would be equivalent to about 17 diploid human cells, as each diploid cell contains 6 pg of DNA.

The PCR protocol defined above was then tested on whole diploid cells using aliqouts of whole blood and the cell suspension such that the template for PCR consisted of DNA extracted from approximately 1000, 100 and 10 cells. Successful amplification was achieved from template derived from approximately 1000 and 100 cells.

The ß-globin gene primers used to amplify the 597 bp fragment in a nested manner were very suitable for the preimplantation diagnosis of ß-thalassaemia, as the gene fragment amplified is the site for 77% of the ß-thalassaemia mutations in Asian Indians, 94% of those present in the Mediterranean peoples and 84% of mutations prevalent the Chinese (Kazazian and Boehm, 1988).

Having demonstrated the feasibility of using the PCR to amplify a fragment of the ß-globin gene from approximately 100 pg of target DNA and from DNA derived from approximately 100 whole diploid cells, experiments were performed on embryonic samples in order to describe an approach for the preimplantation diagnosis of ß-thalassaemia in humans.

6.3 Materials and Methods

6.3.1 IVF procedures and preparation of embryonic samples

The human preimplantation embryos used in this study were obtained from patients attending the IVF Unit at the John Radcliffe Hospital, Oxford. This project was approved by the Central Oxford Research Ethics Committee and the Interim Licensing Authority, with informed consent of all couples. None of these couples were known to be at risk of having an offspring with a genetic disorder.

The IVF superovulation regime involved pituitary down regulation with buserelin (Suprefact, Hoescht UK Ltd) administered from the luteal phase of the previous cycle. Once down regulation had been achieved, 150 IU daily of gonadotrophins (Pergonal, Serono Lab. UK Ltd) were used for stimulation for 9-13 days. The ovarian response

was monitored both biochemically and by vaginal ultrasound, and 10,000 IU of human chorionic gonadotrophin (hCG) (Profasi, Serono Lab. UK Ltd) was administered when adequate follicular development was achieved. Oocyte retrieval was performed 35 hours later by vaginal ultrasound guided aspiration. All oocytes were incubated for 5-6 hours prior to insemination (day 0). They were checked 12-18 hours later for fertilisation by the presence of two pronuclei (day 1). The oocytes and embryos were cultured in 2 ml Tyrode's 6 (T6) medium, supplemented with 10% heat inactivated maternal serum, at 37°C, in 5% CO_2 in air. On day 2, the embryos were assessed and the best three selected for transfer, on the basis of morphology, cleavage rate and quality of oocyte and granulosa. Embryos not selected for transfer were assigned for research purposes and assessed daily by microscopy. Twenty three embryos in which development had arrested between the 2-32 cell stage were used in this study, 15 of which were analysed as whole embryos while the remaining eight were split into two approximately equal halves using a pair of micromanipulators (Leitz, Luton) under a phase contrast microscope (M2, Microinstruments, Oxford). Trophectoderm biopsies each consisting of 10-30 cells were performed on 5-6 day old blastocysts as described in section 6.1.2; seven such biopsies were analysed.

Each of these embryonic samples was prepared for analysis by PCR in a class I laminar flow cabinet. The embryonic sample was transferred into tissue culture dishes containing sterile double distilled water and washed, using sterile Pasteur pipettes. The embryonic sample was then transferred in 10 µl of fresh sterile doubled distilled water into an autoclaved Eppendorf tube. A new pipette was used for each transfer and on each occasion the pipette was viewed under the microscope to confirm that no cells remained in the tip. The samples were stored at -20°C.

Collection and preparation of the embryonic samples was done by Dr Anuja Dokras under the guidance of Professor David Barlow and Dr Ian Sargent at the IVF Unit, Nuffield Department of Obstetrics and Gynaecology.

6.3.2 Amplification of the ß-globin gene

Amplification of a fragment of the ß-globin gene was achieved using two rounds of PCR with nested primers. Prior to amplification the embryonic cell sample was thawed, spun down, overlaid with 100 µl of liquid paraffin and heated at 97°C for 20 minutes to ensure cell lysis and release of nuclear DNA. The PCR was performed in a 100 µl reaction volume containing 20 pmol of each of the oligonucleotide primers ß-1 and ß-2, 2 units of *Taq* polymerase, 800 µM total dNTPs, 10 mM Tris HCl pH 8.3, 50 mM KCl and 1.5 mM $MgCl_2$. This reaction mix was stored on ice and added to the boiled cells just prior to amplification. The amplification regimen consisted of 30 cycles of denaturation at 93°C for 1 min, annealing at 55°C for 1 min and extension at 72°C for 1.5 min. In the final cycle extension was prolonged for 3 min. A 2 µl aliquot of the PCR product of the first round was used as the template for the second round, which was performed in an identical fashion using nested primers ß-3 and ß-4.

A number of different negative and positive controls were set up in each experiment. One ng of genomic DNA was used as the PCR template for the positive controls. Negative controls included both media and water blanks; the former consisted of 10 µl of the medium in which the embryos were cultured, pipetted into a sterile Eppendorf tube in exactly the same manner and at the same time as the embryonic cell samples. The water blanks were 10 µl aliquots of sterile double distilled water which were set up both in the area where the embryonic cell samples were prepared and in the area where the PCR was set up. Both the samples and the controls were treated in an identical fashion.

A 10 µl aliquot of the second round PCR product was subjected to electrophoresis in a 1% agarose gel, which was then stained with ethidium bromide and visualised on a UV transilluminator. The time taken for the PCR analysis was eight hours. To confirm the identity of the amplified product, following agarose gel electrophoresis of the PCR products from one of the experiments, the amplified DNA was transferred from the gel onto a nylon filter (Hybond-N+, Amersham, UK) by Southern blotting (section 2.3). This filter was hybridised with a ß-globin gene specific internal ASO probe ß-5, 5'-CTTGATACCAACCTGCCCA, complementary to the coding strand with

its 5' nucleotide at IVS-1 nucleotide 12 (section 2.6.2-3). The probe was 5' end labelled with [γ^{32}P]dATP, then added to the hybridisation buffer so as to give a concentration of $2X10^6$ cpm per ml of buffer and the filter was hybridised for two hours at 53°C. Post hybridisation washing was done in 6XSSC, first at 50°C for 5 min. Stringency of the washing conditions was then increased by raising the washing temperature in 2-3°C increments. Hybridisation patterns were assessed by autoradiography with exposure for one hour at -70°C. Differential hybridisation was considered to have been achieved when only the positive control retained the probe and the negative control showed no sign of the probe.

6.3.3 Detection of ß-thalassaemia mutations

Allele specific PCR (section 3.4.2) was modified so as to make the technique sensitive enough to accurately detect the presence or absence of ß-thalassaemia mutations from less than 30 cells. The quantity of DNA present in a 10-30 cell trophectoderm biopsy would be 60-180 pg, as 6 pg of DNA are contained in a single diploid cell. To make allele specific PCR suitable for preimplantation diagnosis, it was necessary to use two rounds of PCR with nested primers, the allele specific primer being employed as one of the internal primers in the second round. To be applicable to preimplantation diagnosis, the internal control fragment would also have to be amplified in a nested fashion with two rounds of PCR. It was thus necessary to find such a set of internal control primers which could amplify efficiently at the stringent reaction conditions necessary for allele specific PCR. A 377 bp fragment encompassing exon 3A and flanking intron sequences of the antithrombin gene on chromosome 1 was selected as the internal control (primers were a gift from Dr Robin Olds). In order to test the efficacy of allele specific PCR for mutation detection from less than 30 cells, five ß-thalassaemia point mutations that are situated within the 597 bp fragment of the ß-globin gene amplified in previous experiments were analysed. They were IVS-1 nucleotide 5 (G-C), IVS-1 nucleotide 1 (G-T), codons 8/9 (+G), codons 41/42 (-CTTT) and codon 15 (G-A) (figure 6.3).

As it was not possible to obtain embryonic cell samples from couples carrying ß-thalassaemia mutations, diluted DNA solutions

from various individuals whose ß-thalassaemia mutations had already been characterised were used as experimental models. Dilute DNA solutions were prepared from individuals known to be homozygous, heterozygous and normal for each of the five mutations selected for study. Approximately 1 µg of DNA from each of these individuals was serially diluted by a factor of 1:100,000 with sterile double distilled water, such that 1 µl would contain approximately 10 pg of DNA.

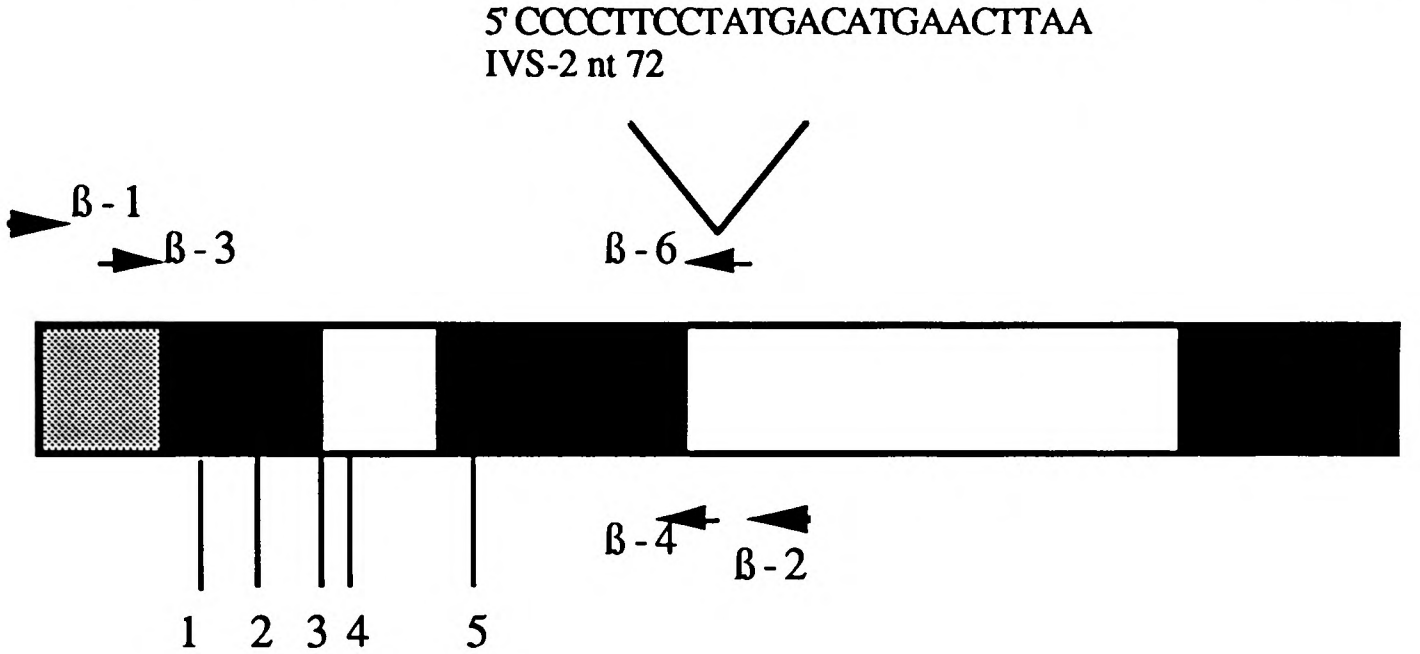


Figure 6.3 Schematic representation of the ß-globin gene showing the positions of the 5 ß-thalassaemia mutations analysed by nested allele specific PCR. 1: codons 8/9 (+G), 2: codon 15 (G-A), 3: IVS-1 nt 1 (G-T), 4: IVS-1 nt 5 (G-C), 5: codons 41/42 (-CTTT). Primer ß-6 was used as the common reverse primer in the nested allele specific PCR, the position of its 5' nucleotide is indicated below the sequence. Sequences of the other primers is shown in figure 6.1.

The first round of PCR was performed using 10 pg of target DNA with ß-globin primers ß-1 and ß-2 and antithrombin primers AT-1 and AT-2, using the reaction conditions described above, the only modification being that the reaction mix was exposed to 260 nm UV light for 5 minutes before the addition of *Taq* polymerase and the target DNA. For the second round of PCR the nested oligonucleotides

were the ß-globin gene allele specific primers coupled with either primer ß-3 or ß-6 (figure 6.3) and the antithrombin gene primers AT-3 and AT-4 (figure 6.4).

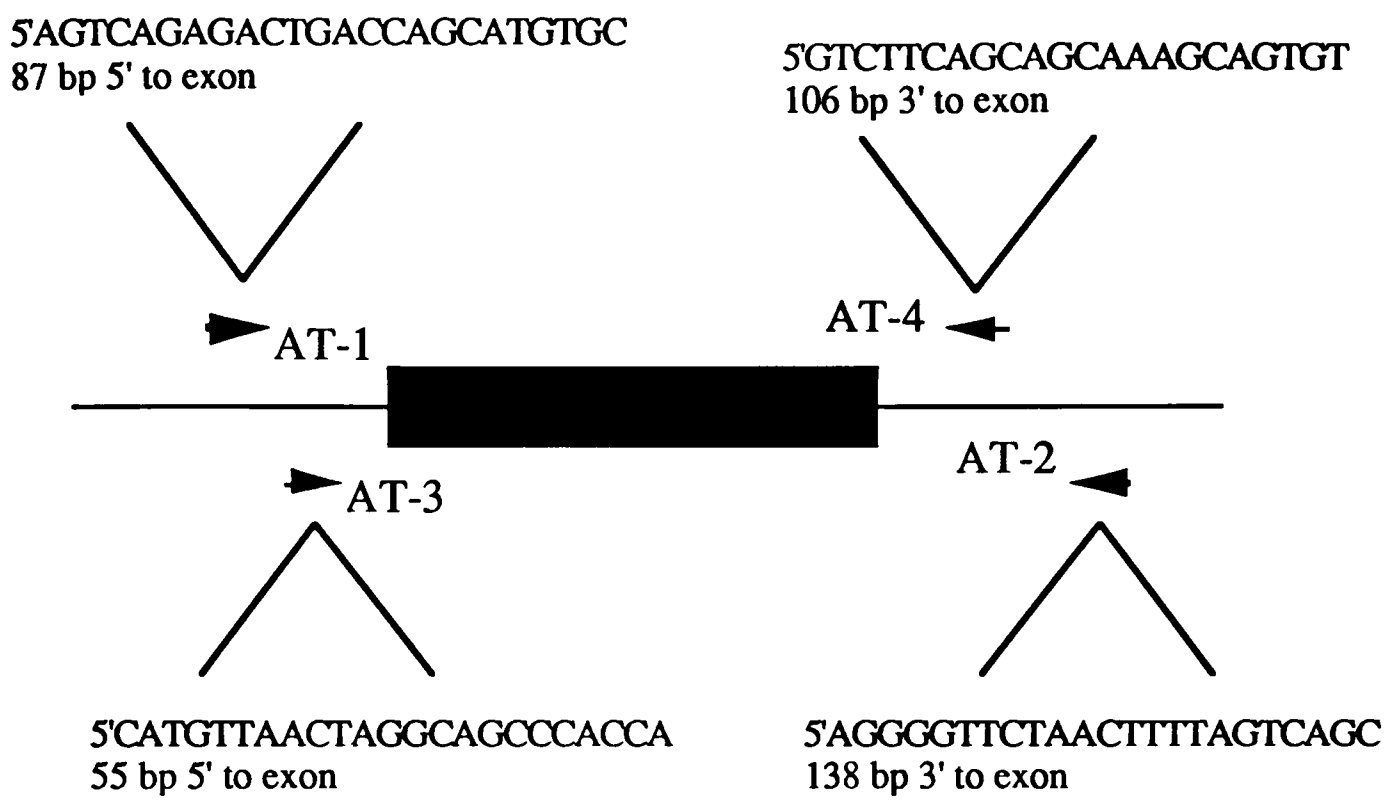


Figure 6.4 Schematic representation of exon 3A and the flanking intron regions of the antithrombin gene, with the internal control primers used to co-amplify a 377 bp fragment of the gene in a nested manner. The position of the 5' nucleotide of each primer is indicated below its sequence.

To ensure specificity of the allele specific PCR, the annealing temperature was increased to 66°C in the second round in which a combined annealing and extension step was performed at 66°C for 2 minutes. With the exception of this modification all other reaction conditions were identical to those used for amplification of the 597 bp ß-globin gene fragment. Both the mutant and normal allele specific primers were tested for each of the mutations analysed (table 6.1). Negative controls which were included in all experiments were 10 µl of sterile double distilled water added to the reaction mix. For these negative controls, the ß-globin gene

primers used in the second round were ß-3 and ß-4 along with the internal control antithrombin gene primers. A 10 µl aliquot of the final PCR product was run out on a 2% NuSieve and 2% agarose gel to facilitate separation of the bands, stained with ethidium bromide and visualised by UV transillumination. The time taken for this analysis was approximately eight hours.

Table 6.1 Allele specific primers for the detection β-thalassaemia mutations by nested PCR.

Mutation	Oligonucleotide sequence 5'-3'	Product size
IVS-1 nt 5 (G-C)	N-CTCCTTAAACCTGTCTTGTAACCTTGTTAC	254 bp
	M-CTCCTTAAACCTGTCTTGTAACCTTGTTAG	254bp
IVS-1 nt 1 (G-T)	N-GATGAAGTTGGTGGTGAGGCCCTGGGTAGG*	454 bp
	M-TTAAACCTGTCTTGTAACCTTGATACGAAA	250 bp
codons 8/9 (+G)	N-CCTTGCCCCACACGGCAGTAACGGCACACT	183 bp
	M-CCTTGCCCCACACGGCAGTAACGGCACACC	183 bp
codons 41/42	N-GAGTGGACAGATCCCCAAAGGACTCAAAGA	412 bp
(-CTTT)	M-GAGTGGACAGATCCCCAAAGGACTCAACCT	412 bp
codon 15 (G-A)	N-TGAGGAGAAGTCTGCCGTTACTGCCCAGTG*	500 bp
	M-TGAGGAGAAGTCTGCCGTTACTGCCCAGTA*	500 bp

N=normal and **M**=mutant allele specific primer and the positions of the mismatches are underlined. All the allele specific primers were complementary to the coding strand and coupled with common primer β-3, with the exception primers marked with *, which were complementary to the noncoding strand and coupled with common primer β-6.

6.3.4 Precautions against contamination

The embryonic cell samples for PCR were prepared in a sterile class I laminar flow cabinet, using sterilised Pasteur pipettes which were washed in sterile tissue culture medium and autoclaved pipette tips and Eppendorf tubes. The operator wore sterile gloves and a gown. This was done in a laboratory in the IVF unit where the embryonic samples were cultured, inspected and biopsied.

The PCR reactions were set up in a physically distant class II laminar flow cabinet using autoclaved pipette tips and Eppendorf tubes. All reaction components were filtered through a 0.2 μm filter (Acrodisc) and autoclaved. To prevent the spread of contamination all reagents were stored in small aliquots and to minimise repeated samplings, the PCR buffer and the dNTP mix were combined to constitute a "pre-mix". Gloves were worn and changed prior to every entry into the set up area. In order to avoid contamination with prior amplified products, the areas where the PCR reactions were performed and amplified products analysed were totally segregated from the areas where the biopsy was done and the reaction set up. Further, all equipment like micropipettes (Gilson, Anachem), pipette tips, Eppendorf tubes, tube racks and ice buckets were designated to the separate areas. Pipettes designated for the setting up of the PCR were regularly rinsed in dilute nitric acid, double distilled water and absolute alcohol and then exposed to 260 nm UV light for 20 min in order to remove any adherent DNA. Four separate areas were designated for these experiments; the IVF unit laboratory where the embryonic samples were prepared, the PCR set up area for the first round of PCR, an area where an aliquot of the first round PCR product was removed and the second round PCR set up and the main laboratory where the thermal cyclers were kept and the final PCR product analysed by electrophoresis. Every precaution was taken to avoid the transfer of equipment or reagents between these areas.

A reduction in false positive amplification by irradiating the PCR mix with 254-300 nm UV light prior to the addition of *Taq* polymerase and the target DNA was reported (Sarkar and Sommer, 1990; Cimino et al, 1990). Pyrimidine dimers contribute to UV induced sterilisation by functioning as termination sites during the extension reactions of the PCR. Only a fraction of all pyrimidines

within a DNA strand form dimers and as dimers are both made and broken by UV exposure, this fraction establishes a steady state level which varies with the irradiation wavelength, the type of pyrimidine dimer and the nucleotide sequences next to the dimer site (Setlow and Setlow, 1962). Following the description of UV irradiation of the PCR mix to avoid false positive amplification, experiments were performed to assess the efficacy of UV irradiation in overcoming false positive amplification and also to determine whether UV irradiation would affect the efficiency of PCR. A number of experimental modifications were tried, these included varying the length of UV exposure time from 5 to 20 min, irradiation of the reaction mix before and after the addition of *Taq* polymerase and the oligonucleotide primers. Subsequently, in the nested allele specific PCR experiments, the first round PCR mix was exposed to 260 nm UV light for 5 min prior to addition of *Taq* polymerase and the template DNA. Irradiation was done by placing the PCR samples in 0.5 ml clear Eppendorf tubes in direct contact with the glass platform of the transilluminator (Fotodyne 1000).

6.4 Results

6.4.1 Amplification of the ß-globin gene

6.4.1.1 Arrested embryos

Successful amplification of a 597 bp fragment of the single copy ß-globin gene was achieved by two rounds of PCR with nested primers, from an arrested human embryo at the two cell stage. The amplified product was clearly visualised following electrophoresis in an ethidium bromide stained agarose gel by UV transillumination. There was no evidence of false positive amplification and the identity of the amplified fragment was confirmed by Southern blot hybridisation with a radiolabelled ß-globin gene specific internal ASO probe (figure 6.5).

The seven experiments performed to amplify the ß-globin gene fragment from a total of 15 arrested human embryos at the 2-32 cell stage are listed in table 6.2. Experiment number 1 was described above. Experiment number 2 included four embryonic cell

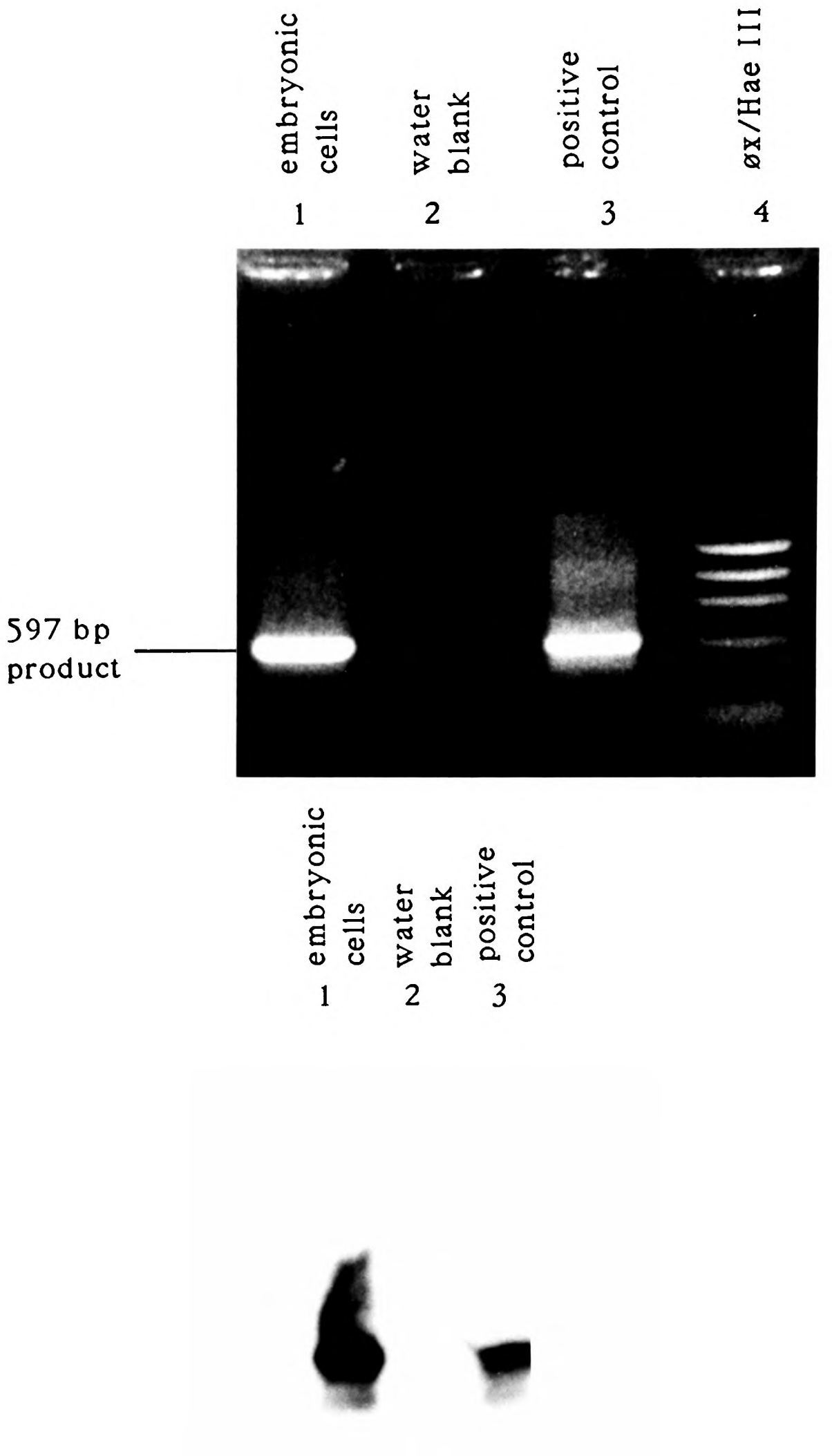


Figure 6.5 Amplification of a ß-globin gene fragment from a two cell arrested embryo. Ethidium bromide stained agarose gel showing a 10 µl aliquot of the PCR product: the negative control was 10 µl of distilled water and the positive control was 1 ng of genomic DNA. øx 174 DNA digested with *Hae*III is the marker. To confirm the identity of the amplified product, Southern blot hybridisation with a radiolabelled ß-globin gene specific internal oligonucleotide probe was performed, the autoradiograph is shown below the gel.

samples, all of which showed successful amplification and the positive and negative controls showed the presence and absence of amplification respectively, as expected. As both the positive controls of the first two experiments were successfully amplified, demonstrating the the sensitivity of the PCR protocol, positive controls were omitted in the next five experiments but the number of negative controls was increased. The PCR products of experiment number 7 are shown in figure 6.6. Successful amplification was achieved in 14 out of the 15 (93%) embryonic samples analysed. Amongst the negative controls, there was false positive amplification in 2 of the 7 media blanks, however none of the 5 and 7 water blanks prepared in the biopsy and PCR set up area respectively, showed any false amplification. The restriction of false positive amplification to some of the media blank negative controls suggests that contamination may have been a sporadic event occurring in the medium used to culture the embryos.

Table 6.2 Amplification of a ß-globin gene fragment from arrested human embryos

Expt. No.	Embryonic cell sample		Medium blank		Water blank IVF unit		Water blank PCR area		Positive control	
	+	-	+	-	+	-	+	-	+	-
1.	1	0	--	--	--	--	0	1	1	0
2.	4	0	--	--	--	--	0	1	1	0
3.	2	0	0	2	0	1	0	1	--	--
4.	1	1	1	1	0	1	0	1	--	--
5.	2	0	1	0	0	1	0	1	--	--
6.	2	0	0	1	0	1	0	1	--	--
7.	2	0	0	1	0	1	0	1	--	--
Total	14	1	2	5	0	5	0	7	2	0



Figure 6.6 Amplification of a 597 bp fragment of the ß-globin gene from less than 30 embryonic cells. Ethidium bromide stained agarose gel showing the PCR products of experiment number seven (table 6.2). There are three negative controls: a media blank consisting of 10 µl of the medium in which the embryos were cultured and two water blanks each consisting of a 10 µl aliquot of sterile double distilled water prepared in the IVF unit lab and in the area where the PCR was set up, respectively. øx 174 DNA digested with *Hae*III is the marker.

+ denotes the presence of amplification and - denotes its absence. False positive and false negative results are underlined. -- indicates that the control was not included in the experiment.

6.4.1.2 Trophectoderm biopsies

Two experiments were performed in which amplification of the ß-globin gene fragment was attempted from a total of seven trophectoderm biopsies (table 6.3). Each experiment included three negative controls, two media blanks and one water blank, all set up in the area where the embryonic samples were prepared. A positive control with 1 ng of template DNA was also included in each experiment. Figure 6.7 shows the gel electrophoresis results of the first experiment. Successful amplification was achieved with 5 out of the 7 trophectoderm biopsies. One of the 5 media blanks and 1 of the 2 distilled water blanks showed false positive amplification. Successful amplification was observed with both the positive controls.

6.4.1.3 Split arrested embryos

Four experiments were performed in which the arrested embryos were split into approximate halves, both of which were subjected to PCR in order to investigate the possibility of being able to analyse each trophectoderm biopsy in duplicate (table 6.4). Negative controls included two medium and two water blanks set up in the laboratory where the embryonic samples were prepared, so that each half of the embryonic sample was accompanied by its own set of negative controls. A single water blank was also prepared in the PCR set up area. Figure 6.8 shows the gel electrophoresis results of experiment number 4. Of the eight split embryos analysed successful amplification was observed in only four samples. In this series of experiments, false positive amplification was observed in 2 out of 8 medium blanks and in 1 out of 4 distilled water blanks prepared in the PCR set up area. None of the eight distilled water blanks set up in the biopsy area showed any false positive amplification.

Table 6.3 Amplification of a ß-globin gene fragment from trophectoderm biopsies

Expt. No.	Embryonic biopsy		Medium blank		Water blank IVF unit		Positive control	
	+	-	+	-	+	-	+	-
1.	4	1	0	2	1	0	1	0
2.	1	1	1	1	0	1	1	0
Total	5	2	1	3	1	1	2	0

+ denotes the presence of amplification and - denotes its absence, false positive and false negative results are underlined.

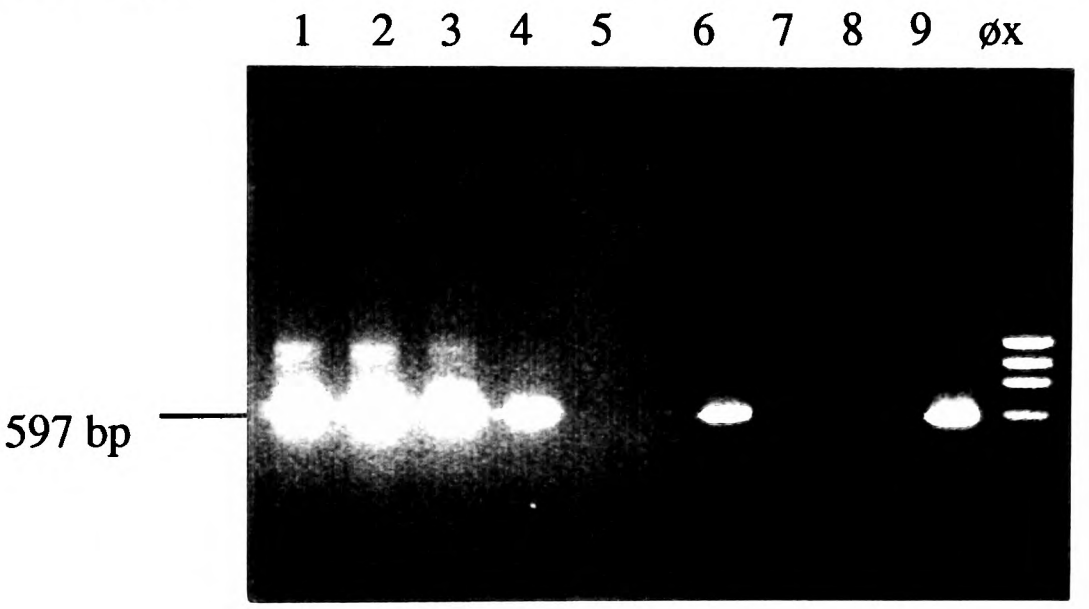


Figure 6.7 Amplification of a 597 bp ß-globin gene fragment from trophectoderm biopsies. Ethidium bromide stained agarose gel showing the PCR products from: 10-30 cell trophectoderm biopsies (lanes 1-5), water blank (lane 6), medium blanks (lanes 7-8) and positive control (lane 9). The marker is øx 174 DNA digested with *Hae*III.

Table 6.4 Amplification of the ß-globin gene from split arrested embryos

Expt. No.	Embryonic cell sample		Medium blank		Water blank IVF unit		Water blank PCR area	
	+	-	+	-	+	-	+	-
1.	2	0	0	2	0	2	1	0
2.	0	2	1	1	0	2	0	1
3.	0	2	1	1	0	2	0	1
4.	2	0	0	2	0	2	0	1
Total	4	4	2	6	0	8	1	4

+ denotes the presence of amplification and - denotes its absence, false positive and false negative results are underlined.

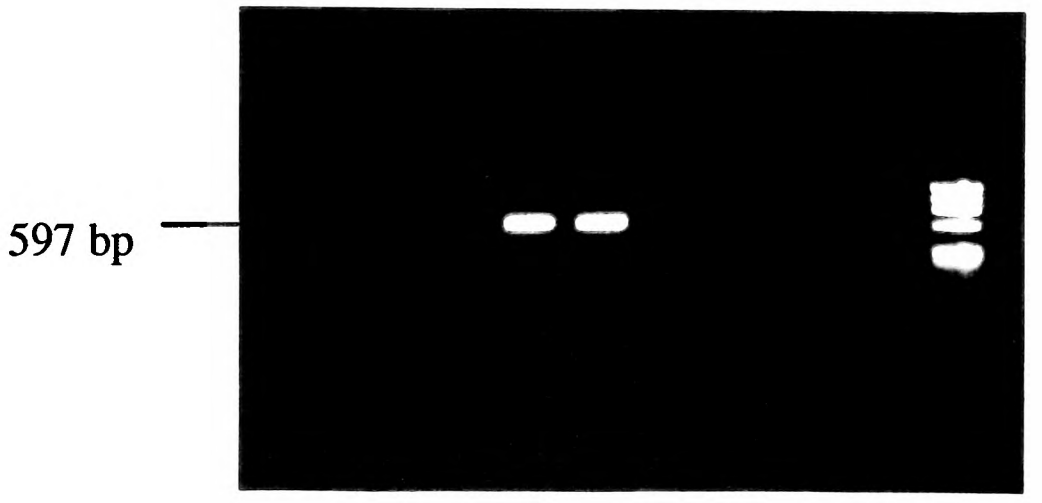


Figure 6.8 Amplification of a 597 bp ß-globin gene fragment from split arrested embryos. Ethidium bromide stained agarose gel showing the PCR products from: medium blanks (lanes 1-2), split embryo samples (lanes 3-4), water blanks set up in the IVF unit laboratory (lanes 5-6) and water blank prepared in the PCR set up area (lane 7). The marker is øx 174 DNA digested with *Hae*III.

6.4.2 UV irradiation

In order to overcome the problem of false positive amplification the reaction mix was exposed to 260 nm UV light; it was found that an exposure period of 20 min decreased the amplification efficiency as determined by visualisation of the intensity of the amplified band on a UV transilluminated ethidium bromide stained gel. Exposure of *Taq* polymerase to UV irradiation greatly reduced the amplification efficiency but exposure of the oligonucleotide primers appeared to have no detrimental effect. It was concluded that UV irradiation for 5 min prior to the addition of *Taq* polymerase and the template DNA, but after addition of the oligonucleotide primers, was effective in preventing false positive amplification without any obvious reduction in amplification efficiency (figure 6.9).

6.4.3 Detection of ß-thalassaemia mutations

Using nested allele specific PCR, with the normal and mutant allele specific primers for the IVS-1 nucleotide 5 (G-C), IVS-1 nucleotide 1 (G-T), codons 8/9 (+G), codons 41/42 (-CTTT) and codon 15 (G-A) ß-thalassaemia mutations, it was possible to correctly determine the genotype from approximately 10 pg of target DNA from individuals known to be homozygous, heterozygous and normal for that mutation. The experiment with primers for the IVS-1 nucleotide 5 (G-C) mutation is shown in figure 6.10; an amplified fragment was present only with the mutant allele specific primer and not with the normal one in a sample homozygous for the mutation. Conversely, amplification with only the normal allele specific primer and not with the mutant one occurred with a normal DNA sample and amplification with both the normal and mutant allele specific primers was observed in the heterozygous sample although the intensity of the amplified band was lower than that obtained with the homozygous and normal samples, because only one allele was being amplified. The presence of the 377 bp internal control band in all the reactions proved the efficacy of the PCR. Similar results were obtained with both the normal and mutant allele specific primers for each of the other four mutations analysed. None of the seven water blanks included in the nested allele specific PCR experiments as negative controls showed any evidence of false positive

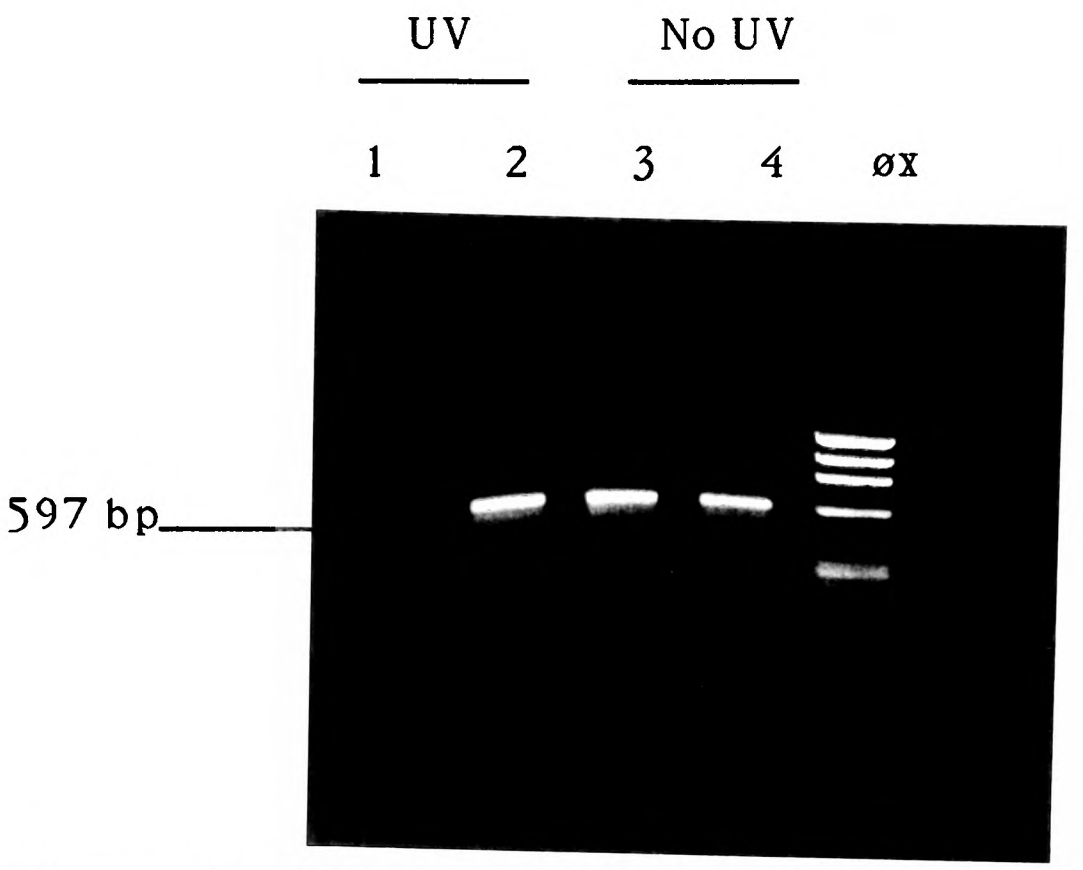


Figure 6.9 Efficacy of UV irradiation in eliminating false positive amplification. Ethidium bromide stained agarose gel showing the PCR products of an experiment in which samples 1 and 2 were UV irradiated for 5 min prior to addition of *Taq* polymerase and the template DNA; samples 3 and 4 were not UV irradiated. Samples 1 and 3 were negative controls in which 10 µl of sterile distilled water was added to the reaction mix and samples 2 and 4 contained approximately 100 pg of genomic DNA. UV irradiation appears to have eliminated the false positive amplication seen in sample 3 without an apparent reduction in amplification efficiency. The marker is øx 174 DNA digested with *Hae*III.

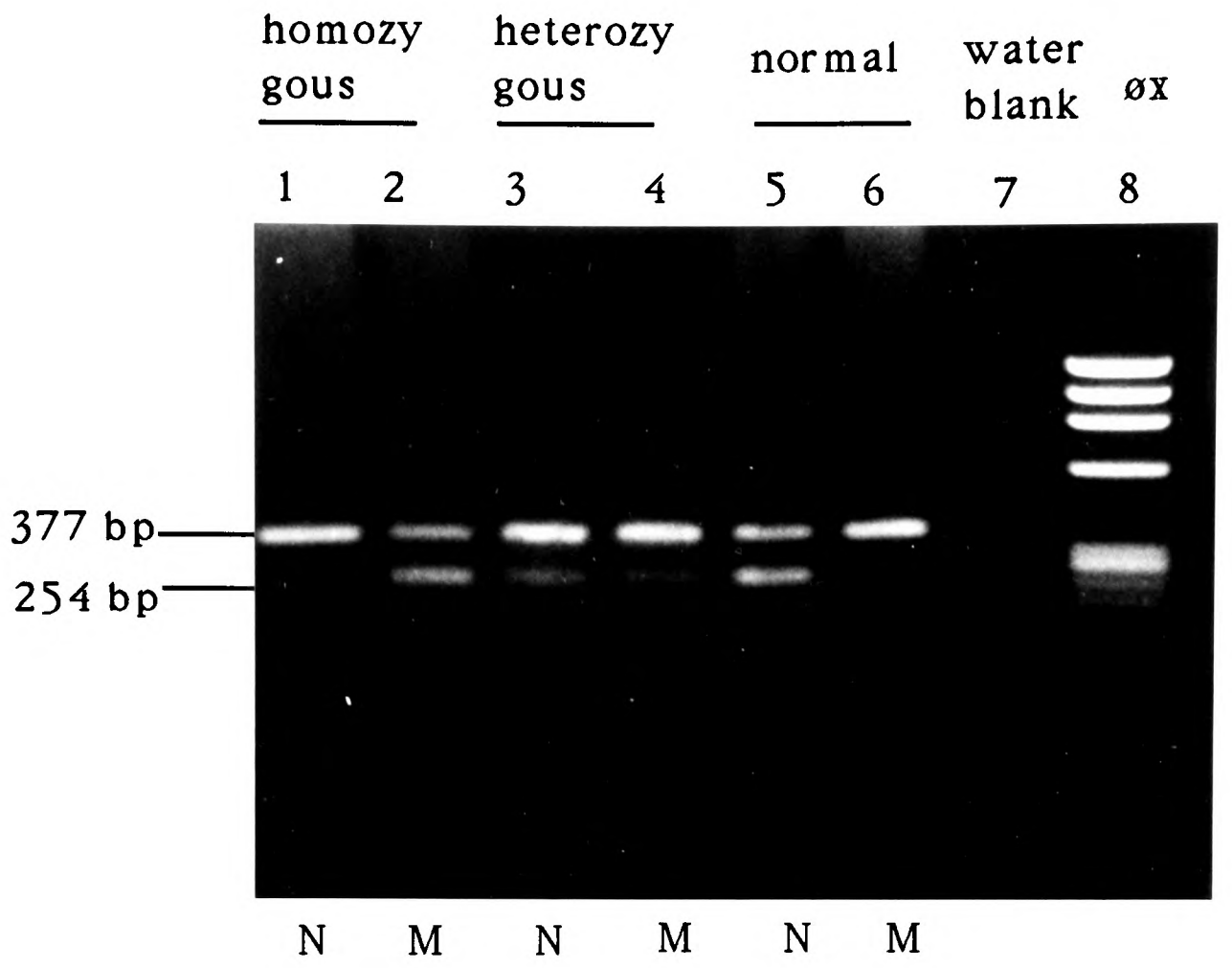


Figure 6.10 Detection of the IVS-1 nucleotide 5 (G-C) mutation by nested allele specific PCR from approximately 10 pg of template DNA. Ethidium bromide stained agarose gel showing the PCR products of allele specific PCR with normal (N) and mutant (M) allele specific primers using template DNA from individuals known to be homozygous, heterozygous and normal for the mutation. The 254 bp amplified fragment is the product of allele specific PCR; it is seen with the homozygous DNA sample only with the mutant primer, while the converse occurs with the normal DNA sample. The heterozygous sample shows amplification with both the normal and mutant allele specific primers but the intensity of the band is lower as only one allele is being amplified. A 377 bp fragment encompassing exon 3A of the antithrombin gene is co-amplified in a nested fashion as the internal control. The negative control consists of 10 µl of sterile distilled water and øx 174 DNA digested with *Hae*III is the marker.

amplification with either the ß-globin or the antithrombin gene primers.

6.5 Discussion and Conclusions

The experiments described in this chapter were designed to evaluate the feasibility of performing preimplantation diagnosis for ß-thalassaemia in humans using DNA amplification. Successful amplification of a fragment of the single copy ß-globin gene from 2-32 cell arrested human embryos and 10-30 cell trophectoderm biopsies was demonstrated, indicating the potential of this approach for preimplantation diagnosis of monogenic disorders like ß-thalassaemia. However before preimplantation genetic diagnosis can be offered to suitable at risk couples the following issues must be addressed: accuracy of diagnosis, possibility of damage to the embryo by the biopsy procedure and the logistic and financial implications of this procedure.

6.5.1 Accuracy of preimplantation diagnosis by PCR

The experiments described in this chapter along with other published reports (Holding and Monk, 1989; Coutelle et al, 1989; Monk and Holding, 1990) demonstrate that a single copy gene can be amplified from a few or even a single cell. Whether this approach could be successfully applied to preimplantation genetic diagnosis in clinical practice remains to be determined.

6.5.1.1 Sensitivity of PCR for preimplantation diagnosis

Successful amplification was obtained in 14 of the 15 whole embryos studied and in 5 out of the 7 trophectoderm biopsies studied which is comparable to the results of larger studies. The fragment of the cystic fibrosis transmembrane conductance regulator (CFTR) gene around the Δ508 mutation could be successfully amplified in 60/63 (95%) of single cells studied (Lesko et al, 1991) and the ß-globin gene was successfully amplified in 79% of the 102 single buccal cells analysed (Monk, 1991). In the latter study, in 71% of cells it was possible to analyse both alleles. This degree of sensitivity is adequate because for preimplantation

diagnosis it is not necessary to be able to obtain a diagnosis in 100% of available embryos, but it is imperative that diagnosis be accurate in some embryos.

For preimplantation diagnosis by DNA analysis, mutation detection at present represents the best strategy. Linkage studies by RFLP analysis have minimal application primarily because a diagnosis must be made following the analysis of a single sample. Allele specific PCR with nested primers was shown to be able to make an accurate diagnosis from approximately 10 pg of DNA. In theory this is equivalent to less than two diploid cells, making the technique suitable for analysis of DNA from 10-30 cell trophectoderm biopsies, single blastomeres obtained from 8 cell embryos and for preconceptional diagnosis by analysis of the first polar body. The ability of this technique to detect a mutation in the heterozygous state suggests that it is probably sensitive enough to analyse a single allele. With careful design of primers, allele specific PCR with nested primers could also be applied to the direct detection of mutations in any other region of the ß-globin gene or in any other gene of interest. It should also be possible to analyse several different distant gene fragments simultaneously by co-amplifying the various fragments.

In practise, if this technique of mutation detection was adopted, the mutations which the parents carry could be characterised by allele specific PCR and the available embryos could then be biopsied and analysed for the presence or absence of these mutations. Since just 2 µl of the PCR product from the first round is required for the second round of PCR, it should be possible to perform a number of second round reactions with different allele specific primers. It is not uncommon for the parents to have identical mutations, therefore, it would be necessary to detect the homozygous state, using normal allele specific primers which would fail to amplify a DNA sample homozygous for the mutation. The PCR product could be easily visualised as an ethidium bromide stained band on UV transillumination of an electrophoretic gel, obviating the need for radioactive probes and permitting the diagnosis to be made in the course of a single working day and the transfer of unaffected embryos in the same ovarian cycle.

As it was not possible to analyse embryonic samples at risk of carrying ß-thalassaemia mutations, the approach for

preimplantation diagnosis of ß-thalassaemia outlined here has not been tested in clinical practice. The results of these pilot experiments however suggest that the strategy outlined could be subjected to clinical trials.

6.5.1.2 Specificity of PCR

The sensitivity of the PCR, which enables DNA amplification from a few cells, is also responsible for false positive amplification as was found in the experiments performed in this study, in spite of precautions against contamination. This emphasises the importance of the multiple negative controls prepared in the different areas that were included in each experiment. The experience of the larger studies quoted above has been rather different: Lesko et al, (1991) had no false positive amplification amongst 76 negative controls while 2% false positive amplification was observed by Monk et al, (1991); the details of the nature and number of negative controls set up in each experiment were not described. Before preimplantation genetic diagnosis by DNA analysis using the PCR can be offered to at risk couples, the problem of contamination must be overcome. False positive amplification is usually due to "carry-over" of amplified products generated by previous PCR reactions (Kwok and Higuchi, 1989) and strategies to prevent contamination should be targeted at preventing this "carry-over". One approach is the complete segregation of areas where the reaction is set up from those where the amplified product is analysed, as has been described in section 6.3.4. The PCR reactions described in this chapter were set up in a laboratory adjacent to one in which the ß-globin gene was being constantly amplified for the purpose of prenatal diagnosis. In retrospect this was probably a serious disadvantage due to the presence of substantial amounts of ß-globin gene PCR products in the vicinity despite segregating work areas and equipment. The high incidence of false positive amplification that was observed could be attributed to this factor. Furthermore, no false positive amplification was encountered with the antithrombin gene primers suggesting that once amplification of a particular gene fragment was performed a number of times the risk of contamination greatly increases. It therefore appears that a

completely separate laboratory is necessary for preimplantation diagnosis by PCR.

Another means of tackling the problem of PCR contamination would be to subject the reaction mixture to a pre-amplification procedure to destroy any contaminating DNA sequences and sterilise it. A number of such procedures have been described, including gamma irradiation (Deragan et al, 1990) and psoralin treatment (Jinno et al, 1990). UV irradiation of the reaction mix for 5 min prior to the addition of *Taq* polymerase and the template DNA was shown to be potentially useful in overcoming the problems of contamination. However, UV irradiation has its limitations as the efficiency of sterilisation is dependent on the size and specific sequence of the contaminating DNA fragment, and, the possibility of UV irradiation having a detrimental effect on the amplification efficiency of the oligonucleotide primers cannot be excluded (Cimino et al, 1990). The hazard of PCR carry-over has been addressed by generating PCR products which, if carried over to subsequent PCR reactions, can be prevented from amplifying. In this method, deoxyuracil triphosphate (dUTP) is substituted for dTTP during template amplification resulting in PCR products containing uracil residues. Prior to subsequent PCR reactions with other templates, treatment of the reaction mix with uracil N-glycosylase results in the cleavage of any dUTP containing DNA, rendering it incapable of serving as a template or primer (Longo et al, 1990). The major drawback of this approach is that it cannot be applied to any procedure requiring two rounds of PCR like those employed in this chapter. None of the techniques currently available to overcome the problem of PCR contamination are totally satisfactory; a method which can effectively sterilise the reaction mix without interfering with its efficiency must be developed before preimplantation genetic diagnosis by DNA amplification can be confidently used in clinical practice.

Due to the difficulties with both the sensitivity and the specificity of PCR based preimplantation diagnosis, it would be advantageous if the embryonic biopsy could be analysed more than once. It has been demonstrated that mutation detection could be performed from less than ten cells; thus a 10-30 cell trophectoderm biopsy could be split into 2-3 groups of approximately 10 cells each and be analysed in parallel. Such a duplicate or triplicate analysis of the

embryonic biopsy would enable the gene diagnosis to be made with greater confidence. However preliminary experiments to assess the feasibility of this approach were unsatisfactory with successful amplification being observed in only half of the eight split embryos analysed. Loss of the embryonic cells during the process of division and transfer into the Eppendorf tube may be a possible explanation for the poor success. With further experience and refinement of the splitting technique it is likely that the amplification success rate will improve.

6.5.2 Embryonic biopsy

Besides the problems associated with PCR analysis of the embryonic biopsy there are a number of important embryological aspects of preimplantation diagnosis which need to be resolved.

6.5.2.1 Safety of procedure

One of the most important questions that must be satisfactorily answered is whether the biopsy procedure itself has a detrimental effect on the growth and development of the embryo. Following biopsy of a totipotential blastomere at the 6-10 cell stage, normal development to the blastocyst stage occurred in the same proportion as that observed in unmanipulated embryos (Handyside et al, 1989). However, good morphological development does not necessarily indicate the potential to implant and develop successfully. As the loss of blastomeres is commonly observed in IVF embryos which seem to develop normally (Trounson and Mohr, 1983), it is argued that removal of a single blastomere would not cause fetal abnormality, although there is evidence that the potential for implantation may be affected (Puissant et al, 1987). Trophectoderm biopsies performed on day 5-6 human blastocysts does not impair further development of the manipulated embryos as judged by their rates of hatching (Dokras et al, 1990) and their ability to adhere to the culture dish and produce outgrowths (Dokras et al, 1991a). A decrease in the hCG secretion was found in the biopsied embryos which correlated with the number of trophectoderm cells removed. When over 10 trophectoderm cells were removed the hCG secretion was significantly reduced, however

when less than 10 cells were biopsied this was not so, suggesting that a trophectoderm biopsy consisting of less than 10 cells would not have a deleterious effect on the embryo (Dokras et al, 1991a). In spite of these reports many questions about the safety of embryonic biopsy remain unresolved. There have been descriptions of its harmful effects in experiments on rabbit and mouse embryos (Jarmulowicz, 1989) and as yet there is insufficient experimental data available on primate embryos.

6.5.2.2 Timing of embryonic biopsy

There is some controversy about the ideal stage at which embryonic biopsy should be performed: the 6-10 cell or the blastocyst stage. Advocates of the former, i.e. a single cell biopsy on the day 3 embryo, argue that as a single totipotential cell is removed there is less likelihood of abnormal subsequent embryonic development (Handyside et al, 1989). As the cells removed by trophectoderm biopsy are extra-embryonic this biopsy technique has been likened to a chorion villus biopsy at a very early stage of development and it is believed that removal of 10-30 trophectoderm cells would not cause embryonic damage. Besides safety of the procedure, it has been argued that embryos surviving to the blastocyst stage *in vitro* have a much better chance of implantation, thus efforts at analysis of embryonic samples obtained at this later stage would be less wasteful. Blastocysts for trophectoderm biopsy need not be obtained by IVF but could also be obtained by uterine lavage (Buster and Carson, 1989) following natural conception in couples with unimpaired fertility which would considerably reduce the expense and strains of these procedures. The possibility of splitting the trophectoderm biopsy and performing the analysis in duplicate or triplicate would be another advantage. Biopsying the extra-embryonic trophectoderm cells helps to resolve some of the ethical dilemmas associated with the removal of a totipotential cell which itself could develop into an embryo. Only 19.1% of embryos cultured *in vitro* were observed to reach the blastocyst stage while 59.8% developed to the eight cell stage (Dokras et al, 1991b); the present inability to maintain embryos *in vitro* until they develop into blastocysts is an important limitation of trophectoderm biopsy. Thus before trophectoderm

biopsy becomes feasible it will be essential to develop the expertise required to successfully mantain viable embryos *in vitro* until the 5-6 day stage.

6.5.3 Feasibility and patient perspective

As presently proposed, preimplantation diagnosis would be labour intensive and costly, requiring expertise in reproductive endocrinology, embryology and molecular genetics. To date the chance of completing an IVF pregnancy, i.e. the "take home baby rate", would be as low as 15% (Medical Research International and the Society for Assisted Reproductive Technology, 1990). A major improvement in IVF success rates must therefore occur before preimplantation diagnosis becomes a feasible option. Uterine lavage could be an alternative noninvasive and less expensive method for obtaining human blastocysts for biopsy (Buster and Carson, 1989), but the blastocyst yield was found to be disappointingly low in spite of superovulation (Carson et al, 1991). Blastocyst recovery by uterine lavage carries the risk of an abnormal embryo being left behind in the uterine cavity which may develop into a viable pregnancy and also the risk of ectopic pregnancy. On the other hand, 86% of women undergoing first trimester chorion villus sampling for prenatal diagnosis achieve a liveborn infant (MRC Trial of CVS, 1991). Couples therefore have a choice between natural conception followed by chorion villus biopsy or IVF with preimplantation diagnosis. Those opting for the former would have to be prepared for a termination of pregnancy in the first trimester while the other group would have to undergo follicle stimulation, monitoring of the ovarian cycle, egg retrieval and embryo transfer with a considerably lower chance of success. The option of IVF would circumvent some, but not all, of the qualms of those who find abortion a difficult choice.

Recently Pergament (1991) conducted a survey amongst 58 females, prospective patients for preimplantation diagnosis, in order to assess their impressions and attitudes. When asked for their first impressions, half the women believed that preimplantation diagnosis offered a moderate improvement or an excellent alternative to first trimester prenatal diagnosis. For women who had had a previous abnormal genetic outcome, this favourable response

increased to 72%. When questioned about their overall impression, for 86% of women not being subjected to the trauma of a pregnancy termination was a major advantage. The three main perceived disadvantages were low success rate, risk to the embryo and cost of the procedure. In their final assessment 45% of women considered the disadvantages outweighed the advantages, 31% believed that the advantages and disadvantages offset each other and 24% felt that the advantage of undertaking a normal pregnancy without further testing outweighed all the disadvantages. It may be concluded that the majority of at risk couples would not consider preimplantation genetic diagnosis, but there are certainly some couples who would like to have the option, moreso if some of the present disadvantages can be overcome.

In this chapter experiments to explore the feasibility of preimplantation diagnosis of monogenic disorders like ß-thalassaemia using DNA amplification by PCR have been described. The results indicate that it is possible to amplify single copy genes and even detect point mutations from cells obtained by embryonic biopsy. However there remain a number of important problems that need to be resolved before these procedures can be confidently employed in clinical practice.

CHAPTER SEVEN

OUTLOOK and SUMMARY OF RESULTS

The primary aim of experiments described in this thesis was to develop a strategy for the prenatal diagnosis of ß-thalassaemia in the Asian Indian population. This aim has been successfully accomplished by the development of a suitable mutation detection technique, allele specific PCR, and by the detailed molecular studies performed in this thesis. In this chapter the feasibility of preventing thalassaemia in India is assessed and a protocol for a national thalassaemia control program is outlined. The final section of this chapter summarises and evaluates the important results obtained in this thesis.

7.1 Control of genetic disease

On a world wide basis, as infant deaths from infection and malnutrition are prevented, congenital disorders like thalassaemia emerge as a major cause of paediatric morbidity and mortality. The treatment of thalassaemia is unsatisfactory, prolonged and expensive. Thus prevention remains the only way to reduce childhood morbidity and mortality, spare the emotional and financial strain on affected families and prevent the enormous drain on health services caused by thalassaemia. To control thalassaemia in an entire community, a well planned program is required. This program will have to establish the epidemiology of the disease, build awareness of genetic risk amongst the medical profession and the population, provide diagnosis and advice to carriers, and integrate the best possible management for affected patients with the prevention of further births by the offer of fetal diagnosis.

7.1.1 Thalassaemia control programs

The feasibility of community based thalassaemia control programs has been clearly demonstrated. In Ferrara, Northeast Italy, between 1970-1, when contraception was not freely available, genetic counselling had no effect on the thalassaemia major birth rate. Between 1972-9 with the availability of contraception there was a 39% drop in the general birth rate, with a further 34% fall in the thalassaemia major birth rate. With the introduction of fetal diagnosis the birth rate of affected children decreased by 50%, and after the establishment of population screening and prenatal diagnosis a 90% decrease in the thalassaemia major birth rate was achieved (Barrai and Vullo, 1980).

Cyprus which has a population of 600,000 and a 17% ß-thalassaemia carrier rate has an extremely successful thalassaemia control program (Angastiniotis and Hadjiminas, 1981) which has reduced the thalassaemia major birth rate to less than 5% of that expected on the basis of the carrier rate. Since the thalassaemia major control program was established in 1977, in Sardinia, an island with a population of around 1.4 million and a 12.5% ß-thalassaemia carrier rate, 90% of thalassaemia major births predicted on the basis of carrier rate have been prevented (Cao et al, 1990). The thalassaemia control program in Greece has been equally successful (Loukopoulos et al, 1990). Potentially 60 infants with thalassaemia major could be born annually in the UK mainly to couples belonging to the Cypriot and Asian Indian immigrant communities (Petrou et al, 1990). Even in the UK which has a relatively low incidence of the disease, the prevention program has been extremely cost effective. A DNA diagnostic laboratory has been recently established in Guangzhou, the capital of the Guangdong province of South China, for the prenatal diagnosis of thalassaemia. In this province the thalassaemia carrier rate is 3-5% and the population is over 50 million (Zhang et al, 1990). Prenatal diagnosis of haemoglobinopathies has been described in the first 100 cases in Thailand, where around 6000 severely affected infants are born each year. Both fetal blood

sampling and globin chain synthesis and CVS and DNA analysis are being performed (Fucharoen et al, 1991).

7.2 Background information for thalassaemia control in India

7.2.1 General considerations

The magnitude of the task of establishing a thalassaemia control program in India can be appreciated by considering some background information. The population is 844 million (Government of India Census, 1991). Illiteracy is high, the overall literacy rate has been estimated to be 36.2%; 46.9% of the male population is literate as compared to 24.8% of the female population. The majority of the population, 76.7%, live in rural areas. In 1987, the birth rate was 32/1000 population and the infant mortality rate was 95/1000 live births (Central Statistical Organisation, 1989). Some of the priority health targets in the seventh five year plan for 1985-90 have been setting up of Primary Health Centres, family planning, immunisation and prophylaxsis against nutritional anaemias. The Medical Termination of Pregnancy Act (1971) made abortion legal for a number of indications including the diagnosis of severe genetic disorders, provided it was performed before 20 completed weeks of gestation, by a registered medical practitioner, in a certified medical centre.

The decline in childhood deaths caused by infection and malnutrition has unfortunately not yet occurred in India. Therefore, congenital diseases are not a major paediatric health issue at the national level. In urban India, however, the situation is different; immunisation, nutrition and family planning programs have caused the infant mortality rate to decline. At the All India Institute of Medical Sciences, New Delhi, congenital diseases comprise 31% of paediatric medical ward admissions and 61% of those in paediatric surgical wards (Professor IC Verma, pers. comm.). Thalassaemia is one of the commonest congenital disorders and is being diagnosed in increasing numbers of children. The prevalence of ß-thalassaemia in India varies in

different ethnic groups ranging from 1-15% (Sukumaran, 1974), with the overall prevalence being approximately 2-3%. There are over 21 million carriers in India (Agarwal, 1986) and around 7000 affected children are born each year. Currently, it is not possible to organise a service adequate to cope with all the new cases diagnosed, moreso as with improved treatment the life expectancy of many patients is increasing. In India there exists a mixed public-private health system, with most services provided by the private sector to those who can afford them. As the public health services in India do not meet the expenses of blood transfusions and iron chelation, the treatment of patients is determined by the financial capabilities of their parents. Thus at present thalassaemia is treated in an uncoordinated manner.

Due to the large scale of the problem and scarce resources available, to date long term planning has been inadequate. The success achieved by thalassaemia control programs in other countries coupled with the advances in obstetric sampling procedures and molecular biology has generated a great deal of interest in similar programs in India. As 40% of the Indian population is less than 15 years old (Central Statistical Organisation, 1989), the aggregate of thalassaemia major births will double in the next two decades. Thus it is imperative that a control program be established immediately. To model this program on those which have proven to be so successful in Italy, Greece and Cyprus would be inappropriate as the large population, scarce financial resources and a comparatively low carrier rate make the Indian scenario totally different. The situation in China and Thailand would be more akin to that existing in India and future progress in these countries should be carefully considered. However, India will have to develop a prevention program suited to her own needs.

7.2.2 Social considerations

No community based genetic disease control program can succeed without careful consideration of the distinct social characteristics of that community. In Indian society it is of particular importance for a couple to prove their fertility and

have a family, which should make prenatal diagnosis acceptable to at risk couples, as in its absence they would be forced to abstain from normal reproductive behaviour.

Indians, in general, are a religious people. The main religions are Hinduism and Islam; Christianity, Jainism, Buddhism and Zoroastrianism are also practised. According to Hindu Vedic laws *(Manusmriti)*, abortion is not permitted under any circumstances. Islam prohibits abortion as a family planning method. However, one school of Moslem Law *(Hanafi)* has permitted pregnancy termination for valid medical reasons, albeit considering it to be an undesirable act, provided it is performed before the lapse of 120 days after the beginning of the pregnancy. Preliminary experiences in India suggest that at risk couples desire genetic information, a reproductive choice and prenatal diagnosis irrespective of their religious beliefs (Sangani et al, 1990; Colah et al, 1990). Second trimester prenatal diagnosis and termination has not been widely accepted by the predominantly Moslem British Pakistanis (Modell et al, 1984), however they do accept first trimester diagnosis and termination (Old et al, 1986b).

At present there are 929 females for every 1000 males in India. The reasons for this uneven sex ratio include a preference for male children resulting in the neglect of female babies, as well as a relative gap in health conditions between the two sexes. Selective abortion of females after fetal sexing is also being practised. Demographers believe that this latter practise is not so widespread as to affect the sex ratio. Prenatal determination of sex for social reasons is illegal in public hospitals all over India, however in some states it is permitted in private clinics. When planning for genetic services in India these issues must be considered and it would be advisable for the thalassaemia control program to discourage any requests for fetal sexing.

7.2.2.1 Consanguinity

Consanguinity is widespread, involving over 17% of the world's population (WHO Advisory Group, 1985) and has important social functions, particularly protection of women. The frequency of consanguineous marriage is believed to be falling in most

populations as a result of social change and increased mobility. However consanguineous marriages continue to be practised in a number of Indian ethnic groups, particularly, the Moslem population. The practise of consanguinity will increase the birth rate of homozygotes for recessively inherited genetic disorders. As this will lead to an accelerated loss of lethal recessives from the gene pool, the custom is believed to correct its pathological genetic effects with the passage of time. There has been evidence for this in South India where close consanguineous marriages have been practised for several centuries (Rao et al, 1972). An increased contribution of recessive inheritance to mental retardation (Bundey et al, 1985) and congenital heart disease (Gatrad et al, 1984) has been proposed. However, there are no convincing genetic reasons for discouraging consanguineous marriages, and attempts to do so by genetic counsellors should be restrained. As consanguinity distributes the risk of having offspring with recessive disorders, like thalassaemia, unequally in the population, the index case marks a high risk family cluster. Testing all extended family members of the index case, termed inductive screening, and offering genetic counselling and prenatal diagnosis to the carriers detected would facilitate the control of thalassaemia in the community (Darr and Modell, 1988).

The social practise of marriage within the community has probably resulted in the regional variation in the distribution of ß-thalassaemia mutations on the Indian subcontinent that was documented in this thesis. Endogamy and consanguinity have also been responsible for 60% of couples carrying identical mutations. The proportion of couples from Northwest Pakistan with identical mutations was 85.3% (70/82); 52 (63%) couples were married to their first cousins and all but one of them carried identical mutations. As all these individuals belonged to the British Pakistani community which has been estimated to have a 55% incidence of first cousin marriages (Darr and Modell, 1988), this was not surprising. Similarly, 96.5% (28/29) of couples from Tamil Nadu carried identical ß-thalassaemia mutations (R. Sarkar, pers. comm.). Thus in both ethnic groups which have a high incidence of consanguineous marriage, the proportion of couples with identical mutations was particularly high. This factor would be worth

considering while performing mutation detection studies in couples from these communities.

7.3 Indian thalassaemia control program: recommendations

7.3.1 Organisation of the thalassaemia control program

Ideally a thalassaemia control program should be organised as a national public health service available equally to all, regardless of income or social class. However at present the Government of India does not have the resources to set up such a service. With an infant mortality rate of 95/1000 live births (Central Statistical Organisation, 1989), there are many more pressing demands on the scarce health resources available. Thus a thalassaemia control program would have to be established in phases; the first phase directed towards urban communities.

To establish and coordinate the control program a central thalassaemia group should be formed. Its members would be responsible for defining and solving problems, and making recommendations to the government and international health organisations. This group should comprise haematologists, paediatricians, obstetricians, community physicians, epidemiologists, representatives of parent's associations, geneticists, nurses, social workers and health education specialists who would meet regularly to define problems and protocols, and evaluate data and progress. It would be useful to have international collaboration with experts who have had experience in this field. A thalassaemia control program depends on the introduction of advanced technology into developing countries, requiring the staff to be highly trained, with regular opportunities to update their knowledge. A link with other countries with similar socio-economic problems, whose prevention programs are also in their infancy, like China or Thailand would be particularly fruitful.

Initially there should be two or three centres participating in thalassaemia control. These centres should be based in university hospitals in some of the large cities and consist of a team of

paediatricians, haematologists and obstetricians who cooperate to provide optimal diagnosis, treatment and engage in research. The central thalassaemia group would coordinate the functioning of these centres. A parents association is a valuable community resource and each centre should encourage the formation of such a body, which could play an important role by raising financial support, exerting political pressure and assisting with community education programs. As the thalassaemia carrier rate in India is not uniform, it will be more cost effective to target communities with a high thalassaemia prevalence rate. This could be done by satellite units attached to the urban centre. For example, the township of Ulhasnagar near Bombay where 150 thalassaemia major births occur annually, could be affiliated to the thalassaemia centre in Bombay (Dr SM Merchant, pers. comm.).

7.3.2 Population information

Public awareness is vital for the success of a thalassaemia control program. The task of public education will be facilitated by employing existing public health education programs, to provide genetic information to the population. For example, the national family planning program could be used to convey the message of reproductive choice. The public could be educated with the help of parents associations, community and religious leaders, school curicula, television, radio, the lay press and distribution of information leaflets. To avoid the potential hazards of stigmatisation and personality disturbances that a genetic disease control program might cause, proper education of the public is crucial. It is also important to ensure that all medical and paramedical personnel receive information about genetic disease in the course of their training. Until adequate facilities are available to meet the demand for screening and antenatal diagnosis, it would be wise to refrain from wide publicity, as these efforts would be wasted in the absence of prenatal diagnosis facilities and may cause undue anxiety in the community.

7.3.3 Heterozygote screening

Carrier screening would include both retrospective and prospective screening. Testing for the carrier state after the birth of an affected child is termed retrospective screening and testing before the birth of an affected child, prospective screening. Retrospective screening and prenatal diagnosis will be the initial approach to disease prevention. As compared to other countries, retrospective prenatal diagnosis will probably have a greater impact in India, as families are large and homozygote life expectancy is comparatively short. The value of inductive screening in communities which practise endogamy and consanguinity should be stressed. Carrier screening should be performed without the violation of any individual's freedom of choice.

Prevention programs in other countries have targeted different sections of the population for prospective screening. Their experiences indicate that the most suitable target group would be couples of reproductive age and women have been found to be more receptive to genetic information than men. In Greece and Italy, couples are encouraged to be tested before marriage or when they seek contraception. In Northern Cyprus, engaged couples are required by law to provide a certificate of testing before they can be married. However, no screening program has ever been shown to modify marriage patterns. Antenatal screening as performed in the UK, has a number of disadvantages. They include, at risk couples are often identified too late for either first or second trimester diagnosis, the emotional involvement makes the choice unnecessarily painful and the short time available for prenatal diagnosis forces couples to make a hurried decision. Further, other options like abstaining from having children, adoption or separation are not available to at risk couples if prospective screening is performed in the antenatal clinic. Considering the social pattern of Indian life, the population size and scarcity of resources, the most appropriate approach to prospective carrier screening would be screening all women at the time of antenatal registration. Adequate prenatal diagnosis

facilities are essential prior to a prospective population screening exercise.

As the results of carrier testing have important implications for an individual's reproductive life, ideally concordant results from two independent test methods are desirable. However other factors like population size, financial resources and the structure of the public health service must be considered when recommending appropriate testing methods. In the Indian situation, heterozygote screening would be best performed by a three tier screening strategy. The first tier consisting of screening by osmotic fragility tests, which are simple, inexpensive and have been shown to be appropriate for developing countries (Kattamis et al, 1981). One such test using 36% buffered saline has been widely used in India with apparent success. Testing of 2000 pregnant women at the time of antenatal registration by this method demonstrated this test to have 100% sensitivity and 84.2% specificity. Around 2% of the subjects tested had a positive result and 0.7% were found to be ß-thalassaemia carriers. The cost of each test was estimated to be Re (rupee) 1 {1£ = approximately Rs (rupees) 50} (Mehta et al, 1989). All positive and dubious results should be investigated further by the second tier screening methods. High risk groups targeted by retrospective and inductive screening and the partners of all individuals with positive results should be tested by the second tier screening methods. These precautions would minimise the unfavourable outcomes that could result from failing to detect the carrier state. However large multi-centre studies will have to be done in India to assess the suitability of osmotic fragility screening tests and if the false negative rate was found to be unacceptably high, all individuals would have to be screened by the second tier methods.

The second tier of screening should include determination of MCH, MCV and Hb A_2 which would identify almost all ß-thalassaemia carriers, including those with coexisting α-thalassaemia interactions. Although carriers of silent ß-thalassaemia mutations would be missed. The widespread prevalence of iron deficiency anaemia in the population will interfere with carrier screening as microcytosis, hypochromia and decreased red cell osmotic fragility are observed in both iron

deficiency anaemia and in thalassaemia. Thus iron deficiency must be ruled out by serum ferritin estimation before a diagnosis of thalassaemia is made, but this is not conclusive in all cases. Hb A_2 levels may be also decreased in severe iron deficiency anaemia, resulting in the failure of detection of some carriers with coexistent iron deficiency (Wasi et al, 1968). Heterozygote screening must be associated with screening and treatment of nutritional and parasitic anaemias. In certain ethnic groups it would be essential to also screen for the prevalent Hb variants, for example, Hb E in Bengalis and Hb D Punjab in Punjabis. Although Hb variants are prevalent in some regional groups, it is the problem of ß-thalassaemia that is of prime importance and should be the focus of the thalassaemia control program. The third tier of testing would be globin chain synthesis and molecular studies which would be indicated in occasional cases.

7.3.4 Genetic counselling

To perform carrier screening without genetic counselling is useless and can be positively harmful. On identification of the carrier state the partner must be invited for testing. If found to be negative the couple must be reassured, and if found to be positive the various options must be discussed in a nondirective manner. There will be a need to train more genetic counsellors in India. It is important to ensure that counselling is uniformly given and that the genetic information is really understood by the individual counselled, so that it can be used appropriately without unnecessary anxiety. The implications of genetic information are influenced by cultural and religious factors, so it will be necessary to adapt the counselling technique.

7.3.5 Prenatal diagnosis

Prenatal diagnosis is an indispensable component of the thalassaemia control program. Facilities for first trimester fetal diagnosis by CVS and DNA analysis should be established immediately. The obstetric expertise to perform CVS is available in India and more obstetricians could be trained. The hurdle that

remains is establishment of DNA diagnostic laboratories. Provisions for DNA analysis by mutation detection must be made and the development of an appropriate methodology and strategy has been a major component of this thesis. Allele specific PCR is considered to be the most suitable mutation detection technique as it is nonradioactive, rapid and straightforward, now that the reaction conditions have been established. The most effective strategy for mutation characterisation would be that described in Chapter 3. A primary screen for the five mutations which account for 92% of all ß-thalassaemia alleles should be performed first. In the 8% of carriers in whom the mutation remains uncharacterised, a secondary screen for mutations that account for 6% of the ß-thalassaemia alleles should be performed. There is no urgency to establish DNA sequencing facilities as in the remaining 2% of cases in whom the mutation cannot be characterised, prenatal diagnosis could be performed by RFLP linkage analysis or by cordocentesis and globin chain synthesis studies. As termination of pregnancy is the usual outcome of the diagnosis of thalassaemia major in the fetus, corroborative results of two independent test methods would be ideal. RFLP linkage analysis would be a suitable confirmatory method. However due to scarce financial resources and technical inexperience, it is recommended that facilities for RFLP studies should be established only after DNA diagnosis by mutation detection has been set up.

As DNA technologies are relatively expensive and local expertise is yet to be developed, it is important that the implementation of this technology is done in coordinated manner. A uniform adoption of techniques and strategies by the different centres, with sharing of resources will be most helpful. If prospective carrier screening is conducted in the antenatal period, there will be a number of cases requiring prenatal diagnosis in the second trimester. Thus facilities for second trimester cordocentesis and globin chain synthesis will have to be also established.

7.3.6 Follow up

It is important to ensure that the thalassaemia control program meets the highest standards of safety and accuracy. For carrier

screening in such a large community, a number of laboratories would have to be involved. To decrease the possibility of error, all of them should participate in regular quality control and have a referral centre for difficult cases. The simplification of DNA diagnostics might encourage the establishment of relatively inexperienced private DNA diagnostic laboratories. A misdiagnosis in the early phase of the program, with adverse publicity, could result in a serious setback for the entire program. Thus it is essential for the central thalassaemia group to maintain rigorous standards. The potential causes for misdiagnosis using PCR mutation detection techniques are maternal contamination, nonpaternity and false positive amplification. To reduce the potential error rate due to sensitivity of the PCR, rigorous quality control procedures will have to be undertaken. It would be important to follow up all children born following prenatal diagnosis and reconfirm the diagnosis in affected pregnancies by analysis of aborted tissues. The fetal loss rate related to the obstetric sampling procedure must also be carefully monitored.

7.3.7 Research

An important function of the thalassaemia centres will be research. The available epidemiological data is insufficient as it consists of limited surveys confined to particular groups. The extrapolations that have been made to the whole community may be found to be inaccurate. Detailed epidemiological surveys would facilitate efficient carrier screening. It would also be interesting to have more data on the natural history, pathology and treatment of thalassaemia in India. The molecular analysis of thalassaemia intermedia in India is an important area that has not received sufficient attention as yet. Characterisation of the molecular lesions that cause a mild phenotype would facilitate more accurate genetic counselling. For this thesis, a number of attempts were made to collect blood samples of thalassaemia intermedia children and their families in India. These attempts failed, partly because haematologists and paediatricians in India believed that soon it would be possible for the molecular studies to be done in laboratories in India itself.

7.3.8 Benefits

The success of the thalassaemia control program will depend on the enthusiasm with which it is implemented and on the degree of acceptance by the population. If successful, there would be a number of wide ranging benefits. To begin with the program would contribute to the diagnosis and treatment of all anaemias, including the widely prevalent nutritional anaemias. The rewards of a public health education program will extend to other important health issues like infant nutrition and antenatal care. A decline in the number of thalassaemia major patients will improve the treatment that can be provided to existing patients. It would also increase the efficiency of blood transfusion services, making more blood available for other conditions. A thalassaemia control program will serve as a starting point for the implementation of other genetic services. Prenatal diagnosis by DNA analysis of other diseases like Duchenne muscular dystrophy could follow and India could then benefit from other advances in molecular medicine. An effective program in India will also be of help to other countries in South Asia, where thalassaemia is common.

At present preimplantation diagnosis is of limited significance in India. However with the establishment of prenatal diagnostic services, it might be possible for centres in India to benefit from further developments in this area and offer preimplantation diagnosis to certain at risk couples. As thalassaemia is so widely prevalent on the Indian subcontinent, molecular studies will contribute to the understanding of the population genetics of the region. As demonstrated in this thesis globin gene variations could serve as useful anthropological markers and for defining population relationships on the Indian subcontinent.

7.4 Financial considerations

Thalassaemia control programs have been conclusively shown to be highly cost beneficial. Only in preventive medicine can moderate investments produce such wide ranging benefits (WHO memorandum, 1983). Public education, carrier screening, prenatal diagnosis and followup are the program components requiring

financial investment. Public education would not involve a large expenditure if done through the existing infrastructure, with help of voluntary organisations. Blood testing for carrier screening would initially require a large financial outlay, but after 5-10 years when the backlog is completed, the number of tests required should be related to the birth rate. Genetic counselling and fetal diagnosis would also involve a large financial investment. Follow up, evaluation and quality control, will be relatively inexpensive.

7.4.1 Cost benefit analysis: the Indian situation

In order to calculate the cost benefits of a thalassaemia control program in India, New Delhi was taken as an example, with the kind assistance of Professor N. Sood, University College of Medical Sciences, New Delhi. In New Delhi, the population is 9.37 million, the birth rate is 28.2/1000 and the ß-thalassaemia carrier rate is 5.5%. The annual number of ß-thalassaemia major births is 200. On an average, for these affected children blood transfusion is required once every three weeks, each costing Rs 600. As seventeen blood transfusions are required each year, the annual cost of transfusions, per patient is Rs 10,000. Laboratory investigations cost another Rs 3,500 and iron chelation therapy Rs 65,000, per year. Thus the total yearly expenditure per affected child is calculated to be around Rs 80,000. Other expenses like running a day care transfusion clinic, medical consultations and loss of man hours of the patient's family are not even considered. The expense of providing treatment to the 200 children born each year would be Rs 16 million and if treatment was provided for at least 10 years of life, the expense would be Rs 160 million, without considering inflation.

Each year there are about 264,000 pregnancies in New Delhi. Screening by osmotic fragility tests has been estimated to cost Re 1 per test, when combined with the labour costs, it would rise to Rs 2 per test. Thus the cost of first tier screening of all expecting mothers in New Delhi by osmotic fragility tests would be Rs 528,000. Second tier testing would be indicated in around twice the number of carriers present, 11% of all women screened. The

costs of MCV, MCH, Hb A_2 and serum ferritin estimation would be Rs 25 per individual. If 11% of women and their partners were tested, the second tier testing would cost approximately Rs 1.45 million. The total cost of carrier screening by this approach would be around Rs 2 million per year. If it was decided to replace osmotic fragility tests by MCV, MCH and Hb A_2 estimation in all pregnant women, at Rs 20 per case this would cost Rs 5.28 million. Serum ferritin estimation and partner testing will cost an additional Rs 1 million, making the total cost of carrier screening by the second approach to be Rs 6.28 million.

Prenatal diagnosis would have to be performed in around 800 cases, each year. The cost of consumables to perform a family study by mutation detection using allele specific PCR as described in Chapter 3, would be approximately Rs 1000. It is difficult to estimate the cost of personnel, equipment, rent and other such factors. However, if existing infrastructure such as university hospitals were to be used the cost per case would be Rs 2000, including the obstetric procedures. Thus the yearly cost of prenatal diagnosis would be Rs 1.6 million. The annual cost of prevention of all thalassaemia major births in New Delhi would be approximately Rs 3.6 million or Rs 7.88 million depending on the carrier screening method used. Both these amounts are less than one half of the cost of annual treatment of the affected children that would have been born that year. The cost benefit ratio of thalassaemia control is indisputable.

7.4.2 Financing the thalassaemia control program

In spite of the cost effectiveness of thalassaemia control, raising finances for this program will be difficult in India. At present the state does not make a significant contribution to financing the treatment of thalassaemia major children. Since the health budget makes no provision for the treatment of thalassaemia, establishment of a disease control program will not result in savings in real terms for the state. Further, the more urgent demands on the limited health budget makes it difficult for the government to fund genetic services, which are considered to be luxury medical care. Furthermore, the recent devaluation of the

Indian rupee has increased by 20% the cost of equipment and consumables which have to be imported. The private sector provides a number of other expensive treatments, like IVF, however, a private sector based prevention program would not be successful. As although the private sector may be able to provide prenatal diagnosis by CVS and DNA analysis to a limited number of individuals, this would have to be integrated with a public program for education and screening of the majority of the population. A thalassaemia control program available to only financially privileged sections of society will not achieve a significant reduction in the thalassaemia major birth rate. Funding will have to be obtained from international organisations and private agencies. With these funds and the infrastructure of university and public hospitals it should be possible to establish prenatal diagnosis facilities in a few thalassaemia centres. The costs of consumables for DNA diagnosis might be met by levying a small charge for prenatal diagnosis.

7.5 Thalassaemia control in India: summary

The control of thalassaemia is complex encompassing a number of wide ranging issues. Some of the important social, economic and demographic factors that would affect a thalassaemia control program in India have been assessed in this chapter. On the basis of this assessment, the experiences of other countries and the results obtained in this thesis, recommendations for thalassaemia control in India have been made.

The thalassaemia control program in India should be administered by a central body and consist of a few urban centres located in university hospitals, with facilities for patient management and prenatal diagnosis. This program should be primarily directed towards the urban population. Prenatal diagnosis facilities are indispensable and unless they are established, thalassaemia control in India cannot be contemplated. Technological advances in molecular biology have greatly facilitated gene diagnosis, such that mutation detection techniques developed in this thesis are applicable in a developing country. Further, the detailed molecular studies performed in this thesis

have permitted the delineation of a mutation detection strategy which will be valuable for prenatal diagnosis. On the basis of these contributions a first trimester prenatal diagnosis program based on CVS and DNA analysis by mutation detection should be established. Once prenatal diagnosis is available, it should be offered to couples identified by retrospective and inductive carrier screening.

The next phase would be prospective carrier screening of all antenatal women in urban communities. This should be done by a three tier testing strategy comprising, 1) osmotic fragility tests, 2) MCV, MCH and Hb A_2 estimation and 3) globin chain synthesis and molecular studies. It might be decided to omit osmotic fragility screening tests after further evaluation. Genetic information must be made available to the urban population with the help of public education campaigns. Couples detected by prospective screening must be offered antenatal diagnosis.

After the second phase has been successful, strategies for the future should be carefully planned. It will have to be determined whether thalassaemia control should be extended into rural communities and if so how this could be done. These are complex issues and would depend on the success of the initial phases, the progress made in other public health areas and the economic situation in India.

7.6 Summary of results

7.6.1 Characterisation of ß-thalassaemia mutations

In order to comprehensively describe the spectrum of ß-thalassaemia mutations in the Asian Indian population, 656 unrelated disease carriers originating from seven different regions of the Indian subcontinent were studied by DNA analysis. Initial attempts at mutation detection by dot-blot hybridisation using radioactive ASOs highlighted the problems with this method and stressed the need for an alternative method of mutation detection. Allele specific PCR was then developed for the detection of ß-thalassaemia mutations. After the initial experiments to establish the appropriate reaction conditions, this technique was found to

be most satisfactory and was successfully used for the detection of 28 different ß-thalassaemia mutations. Having the advantages of being nonradioactive, rapid and relatively simple, allele specific PCR would be applicable to a developing country. As obtaining synthetic oligonucleotide primers is difficult in a developing country like India, to implement allele specific PCR assistance from a more developed country will be required. The risk of misdiagnosis due to false positive amplification could be minimised by precautions to prevent PCR contamination and the inclusion of appropriate controls. Co-amplification of a polymorphic VNTR segment (Horn et al, 1989) could serve as an internal control, as well as detect maternal contamination and nonpaternity, two important causes of error.

Using allele specific PCR to screen the 656 unrelated carriers for the ten mutations previously described in the immigrant Asian Indian population it was possible to characterise the mutations in 626 (95.4%) carriers. By direct DNA sequencing of the ß-globin genes of the remaining 30 subjects seven additional ß-thalassaemia mutations were identified, two of these, IVS-2 nucleotide 837 (G-T) and codon 88 (+T) were newly described mutations. Thus the strategy employed for characterisation of the ß-thalassaemia mutations of this large population sample proved to be most effective. It was found that ß-thalassaemia is heterogeneous in the Asian Indian population with sixteen different mutations present. Further molecular studies will probably detect some more rare mutations in this population. Five common mutations, IVS-1 nucleotide 5 (G-C), the 619 bp deletion, codons 8/9 (+G), IVS-1 nucleotide 1 (G-T) and codons 41/42 (-CTTT) accounted for 92% of all ß-thalassaemia alleles. IVS-1 nucleotide 5 (G-C) was found to be the commonest and most widespread ß-thalassaemia mutation, with a frequency of 34.4%. The remaining 11 mutations could be divided into two groups on the basis of their frequencies. Codon 15 (G-A), codon 5 (-CT), IVS-1 minus 1 (G-C), codon 16 (-C) and IVS-2 nucleotide 837 (G-T) comprised 6% of the ß-thalassaemia alleles. The last group of mutations included cap site +1 (A-C), IVS-2 nucleotide 1 (G-A), IVS-1 nucleotide 110 (G-A), IVS-1 minus 1 (G-A), -88 (C-T) and codon 88 (+T) which accounted for 1.7% of ß-thalassaemia alleles.

Nine of the 16 ß-thalassaemia mutations observed were ß°-thalassaemia mutations and the IVS-1 nucleotide 5 (G-C) mutation is a severe ß+-thalassaemia mutation. It would be important to study the molecular basis of Asian Indian patients with thalassaemia intermedia so as to determine the ß-globin gene mutations and other genetic factors that cause a mild clinical phenotype. These results would facilitate accurate genetic counselling.

It was possible to study an adequate number of carriers from Gujarat, Punjab, Sindh and Northwest Pakistan. In these regions, a variation in the distribution of ß-thalassaemia alleles were observed. For example, the codons 8/9 (+G) and the 619 bp deletion were predominant in Northwest Pakistan and Sindh, respectively. Further studies of disease carriers from other regions of the Indian subcontinent are likely to describe other examples of such regional variation. Knowledge of these variations will be valuable while screening for mutations in couples interested in prenatal diagnosis. Sixty percent of couples were observed to carry identical mutations, this will facilitate mutation detection. However on exclusion of the 82 British Pakistani couples who are known to marry their first cousins frequently, 50.6% of couples were found to carry identical mutations. The latter proportion would be more representative of the situation in India.

On the basis of the spectrum of ß-thalassaemia mutations observed, a detailed strategy for prenatal diagnosis by mutation detection was outlined. To characterise the ß-thalassaemia mutations in couples requiring prenatal diagnosis, the five common mutations should first be sought. On consideration of the carrier's regional origins the mutations sought might be altered. In less than 10% of carriers the mutation will remain unidentified following this primary screen, then a secondary screen for the five mutations that account for 6% of alleles should be done. DNA sequencing could be done for the remaining 2% of carriers. This strategy has been successfully used for the prenatal diagnosis of ß-thalassaemia in over 90 Asian Indian families in the UK (Dr JM Old, pers. comm.). Thus, the mutation detection strategy outlined will be applicable both in India itself, as well as in any other

centre which provides prenatal diagnosis of ß-thalassaemia to an immigrant Asian Indian community.

7.6.2 Haplotype analysis

By the analysis of seven polymorphic restriction enzyme cutting sites in the ß-globin gene cluster, construction of the ß-globin gene haplotypes of the 656 carriers whose mutations had been characterised was attempted. Haplotype construction was possible in 419 ß-Th and 196 ß-A carriers who originated mainly from Northwest Pakistan, Gujarat, Punjab and Sindh. Analysis of the ß-globin gene haplotypes and mutations of this large number of chromosomes provided information on the origin and spread of ß-thalassaemia mutations on the Indian subcontinent. The ß-A haplotypes were found to be more diverse than the ß-Th ones. With a single exception, each of the ß-Th haplotypes was represented amongst the ß-A haplotypes and the frequency of each ß-Th haplotype was similar to that of its ß-A counterpart. However the commonest haplotype, +---- -+, which was associated with the most frequent mutation, IVS-1 nucleotide 5 (G-C), occurred at twice the frequency amongst ß-Th chromosomes (46%) as compared to normal ones (22%). These observations suggest that ß-thalassaemia mutations occurred relatively recently on existing chromosomal backgrounds. Having arisen, the ß-thalassaemia mutations experienced positive selection probably because of protection against malaria conferred on the carriers.

On analysis of the ß-globin gene haplotypes and mutations in carriers from Northwest Pakistan, Gujarat, Punjab and Sindh it was found that in each of the regional "micropopulations", the situation was identical to that of the population as a whole. In each region the ß-A haplotypes were more diverse than the ß-Th ones and generally each of the ß-Th haplotypes was represented amongst the ß-A ones. The ß-Th haplotype associated with the common mutations present in that region, occurred at a substantially greater frequency amongst the ß-Th chromosomes as compared to the ß-A chromosomes, suggesting positive selection of the ß-Th chromosomes associated with the predominant mutation of that region. Irrespective of the ß-thalassaemia

mutation, this positive selection was observed, for example, codons 8/9 (+G) associated with the +---- ++ haplotype in Northwest Pakistan and the 619 bp deletion associated with the +---- -+ haplotype in Sindh.

A strong linkage disequilibrium between the ß-thalassaemia mutations and ß-globin gene haplotyes was observed, although this linkage was not absolute and with the exception of some rare mutations, all were found to be associated with more than one haplotype. On the basis of the extent of linkage disequilibrium between mutations and their haplotypes and the geographical distribution of mutations it was deduced that the IVS-1 nucleotide 5 (G-C) and the codons 41/42 (-CTTT) mutations were probably the oldest mutations on the Indian subcontinent. The association of mutations with more than one 5' haplotype but the same framework indicates recombination events occurring in the hotspot, 5' to the ß-globin gene. Six mutations were associated with more than one framework, the likely explanation for this was believed to be gene conversion events. Although an alternative explanation, population migration from South China and Melanesia, was proposed for the codons 41/42 (-CTTT) and IVS-1 nucleotide 5 (G-C) mutations, respectively. Though the haplotype-mutation association studies provided clues about the origins of the ß-thalassaemia mutations, it was not possible to resolve with certainty all the questions raised. May be further studies with markers which are more polymorphic in this population would be more informative.

The regional variation in the distribution of mutations, caused by the strict endogamy practised by the population, made these mutations useful anthropological markers. Considering the haplotype-mutation associations and the history of the region, population movements on the Indian subcontinent itself and between the neighbouring regions of Asia could be studied. Further detailed microepidemiological analysis of globin gene variants on the Indian subcontinent, particularly studies on some isolated tribal groups, will be of value in providing further evidence of historical gene flow in this region and defining population relationships. It will be interesting to note whether future studies corroborate the conclusions made here and to see

how more data from other regions that were not analysed contributes to the picture presented here.

RFLP analysis for the haplotype construction was done by restriction enzyme digestion of PCR product; this rapid method of RFLP analysis will also be useful for prenatal diagnosis. RFLP analysis is a valuable method for confirming the results of mutation detection studies and for prenatal diagnosis in cases where the ß-thalassaemia mutation cannot be identified. The results of haplotype-mutation linkage obtained by this large study will be valuable to a prenatal diagnosis program combining mutation detection and haplotype analysis in the Asian Indian population.

7.6.3 α-Globin gene mapping studies

A knowledge of the prevalence and molecular basis of α-thalassaemia in the Asian Indian population is necessary for a thalassaemia control program. Following haematological investigations of 1240 apparently normal Asian Indian immigrants residing in East London, α-thalassaemia was suspected in 51 individuals. α-Globin gene mapping studies detected the common, single α-globin gene deletion in 24 individuals. As Hb H disease and Hb Bart's hydrops fetalis are very rare the α-globin deletion *in cis* would be infrequent, thus it appears that the remaining carriers have a nondeletional α-thalassaemia determinant. These studies are amongst the first that attempt a comprehensive description of the molecular basis of α-thalassaemia in the Asian Indian population. Future detailed molecular studies that will characterise these nondeletional α-thalassaemia determinants will contribute to the understanding of the molecular pathology of this condition.

7.6.4 Preimplantation diagnosis of ß-thalassaemia

Some couples find prenatal diagnosis and selective abortion to be an unacceptable method of achieving a healthy family; preimplantation diagnosis would be a suitable alternative for these couples. An approach to preimplantation diagnosis of ß-

thalassaemia in humans by PCR analysis, of a 10-30 cell trophectoderm biopsy was developed. After initial experimentation using different templates and PCR protocols, the amplification of a 597 bp fragment of the single copy ß-globin gene from 100 pg of DNA and about 100 diploid cells was achieved by two rounds of PCR with nested primers. Using this amplification protocol, successful amplification of a ß-globin gene fragment was achieved in 14 out of 15 embryonic cell samples, each comprising 2-30 cells and in five out of seven trophectoderm biopsies, each comprising 10-30 cells. The identity of the PCR product was confirmed by hybridisation with an internal ASO probe. Thus the PCR protocol developed was sensitive enough to analyse the single copy ß-globin gene from as few as two embryonic cells.

However, in spite of meticulous precautions to prevent contamination, false positive amplification was observed in four out of 25 negative controls. To ensure diagnostic accuracy embryonic cell samples were split so that each sample could be studied in duplicate. Successful amplification was obtained in four out of eight split embryonic samples, the reason for this low success rate was probably the failure to transfer the split sample into the Eppendorf tube whose contents were subjected to PCR, because of technical inexperience. To overcome the persistent problem of false positive amplification, UV transillumination of the reaction mix for 5 min prior to the addition of *Taq* polymerase was attempted, with success. Allele specific PCR performed in a nested manner was then demonstrated to be sensitive enough to detect the presence or absence of a ß-thalassaemia mutation from approximately 10 pg of DNA, which is equivalent to two diploid cells. Diagnosis of the normal, homozygous and heterozygous state for the five common Asian Indian ß-thalassaemia mutations was shown to be possible by this method.

Thus preimplantation diagnosis by embryo biopsy and PCR analysis is theoretically possible. Larger studies analysing different gene fragments, including the polymorphic VNTR segments need to be done. Also the various methods now available to sterilise the pre-amplification PCR mix must be assessed so that appropriate PCR protocols could be defined for

preimplantation diagnosis of genetic disease. Also the problem of low IVF success rates and maintaining human embryos *in vitro* will have to be overcome before preimplantation diagnosis of ß-thalassaemia may be used in clinical practice.

In keeping with the primary aim of this thesis: prenatal diagnosis of ß-thalassaemia in India, the final chapter considered the many factors affecting thalassaemia prevention in India and made recommendations for a thalassaemia control program. The strategy described in this thesis for prenatal diagnosis of ß-thalassaemia in the Asian Indian population by mutation detection has proved to be successful in the UK. The next challenge will be its application and success in India.

REFERENCES

Agarwal MB, Mehta BC (1982) Genotypic analysis of symptomatic Thalassaemia syndromes. *J Postgrad Med* 28: 1

Agarwal MB (1986) Living with Thalassaemia. Prabhat Agency, Bombay

Agarwal MB, Gupte SS, Vasandani D, Viswanathan C, Puniyani RR, Ramanathan J, Massil DE, Shah S, Rajadhyaksha GC, Bhave AA (1991) Efficacy and safety of 1-2, dimethyl-3-hydroxypyrid-4-one (LI) as an oral iron chelator in patients of beta thalassaemia major with iron overload. *J Association of Physicians of India* 39: 669

Alter BP, Modell BC, Fairweather D, Hobbins J, Mahoney MJ, Frigoletto FD, Sherman AS, Nathan DG (1976) Prenatal diagnosis of hemoglobinopathies: A review of 15 cases. *N Engl J Med* 295: 1437

Alter BP for the WHO International Registry (1985) Antenatal diagnosis of thalassemia: A review. *Ann NY Acad Sci* 445: 393

Alter BP, Stump DD (1987) Prenatal diagnosis of hemoglobinopathies: A potential application of high preformance liquid chromatography. *Hemoglobin* 11: 341

Alter BP (1990) Antenatal diagnosis: Summary of results. *Ann NY Acad Sci* 612: 237

Amselem S, Nunes V, Vidaud M, Estivill X, Wong C, d'Auriol L, Vidaud D, Galibert F, Baiget M, Goosens M (1988) Determination of the spectrum of β–thalassemia in Spain by use of dot-blot analysis of amplified β-globin DNA. *Am J Hum Genet* 43: 95

Angastiniotis MA, Hadjiminas MG (1981) Prevention of thalassaemia in Cyprus. *Lancet* i: 369

Anshan Department of Obstetrics and Gynaecology (1975) Fetal sex prediction by sex chromatin of chorionic villi cells during early pregnancy. *Clin Med J* 1: 117

Antonarakis SE, Boehm CD, Giardina PJV, Kazazian HH (1982a) Non-random association of polymorphic restriction sites in the ß-globin gene cluster. *Proc Natl Acad Sci USA* 79: 137

Antonarakis SE, Orkin SH, Kazazian HH, Goff SC, Boehm SC, Waber PG, Sexton JP, Ostrer H, Fairbanks VK, Chakravarti A (1982b)

Evidence for multiple origins of the ßE globin gene in Southeast Asia. *Proc Natl Acad Sci USA* 79: 6608

Antonarakis SE, Orkin SH, Cheng T-C, Scott AF, Sexton JP, Trusho SP, Charache S, Kazazian HH (1984) ß-thalassemia in american blacks: Novel mutations in the TATA box and IVS-2 acceptor splice site. *Proc Natl Acad Sci USA* 81: 1154

Antonarakis SE, Kazazian HH, Orkin SH (1985) DNA polymorphism and molecular pathology of the human globin gene clusters. *Hum Genet* 69: 1

Atweh GF, Wong C, Reed R, Antonarakis SE, Zhu D, Ghosh PK, Maniatis T, Forget BG, Kazazian HH (1987) A new mutation in IVS-1 of the human ß-globin gene causing ß-thalassemia due to abnormal splicing. *Blood* 70: 147

Baron MH, Maniatis T (1986) Rapid reprogramming of globin gene expression in transient heterokaryons. *Cell* 46: 591

Barrai I, Vullo C (1980) Population screening for carriers of recessively inherited disorders. *Lancet* ii: 638

Bodmer WF, Cavalli-Sforza LL (1976) Genetics, Evolution, and Man. WH Freeman and Company

Bradley TB, Wohl RC, Smith GJ (1975) Elongation of the α-globin chain in a black family: Interaction with Hb G Philadelphia. *Clin Res* 23:1314

Breathnach R, Benoist C, O'Hare K, Gannon F, Chambon P (1978) Ovalbumin gene: Evidence for a leader sequence in in mRNA and DNA sequences at the exon-intron boundaries. *Proc Natl Acad Sci USA* 75: 4853

Brittenham GM, Lozoff B, Harris JW, Mayson SM, Miller A, Huisman THJ (1979) Sickle cell anemia and trait in Southern India: Further studies. *Am J Hematol* 6: 107

Brittenham GM (1981) Hemoglobinopathies in India. *Hemoglobin* 5: 751

Brittenham GM (1983) The geographic and ethnographic distribution of haemoglobinopathies in India. In: "Distribution and evolution of haemoglobin and globin loci." (ed Bowman JE) Elsevier Scientific Publishing Co, Amsterdam p 169

Brown JM (1990) The molecular genetics of thalassaemia in Burma. DPhil thesis, University of Oxford

Buckle VJ, Higgs DR, Wilkie APM, Super M, Weatherall DJ (1988) Localisation of the human α-globin to 16p13.3 pter. *J Med Genet* 25: 847

Bundey S, Webb TP, Thake A, Todd J (1985) A community study of severe mental retardation in the West Midlands and the importance of the fragile X chromosome in its aetiology. *J Med Genet* 22: 258

Bunn HF, Forget BG (1986) Haemoglobin: molecular, genetic and clinical aspects. WB Saunders, Philadelphia

Buster JE, Carson SA (1989) Genetic diagnosis of the preimplantation embryo. *Am J Med Genet* 34: 211

Canadian Collaborative CVS-Amniocentesis Clinical Trial Group (1989) Multicentre randomised clinical trial of chorion villus sampling and amniocentesis. *Lancet* i: 1

Cao A, Rosatelli MC, Leoni GB, Tuveri T, Scalas MT, Monni G, Olla G, Galanello R (1990) Antenatal diagnosis of ß-thalassemia in Sardinia. *Ann NY Acad Sci* 612: 215

Carson SA, Smith AL, Scoggan JL, Buster JE (1991) Superovulation fails to increase human blastocyst yield after uterine lavage. *Prenatal Diagnosis* 11: 513

Central Statisitical Organisation, Government of India (1989) Statistical Pocket Book. Government of India Press

Chakravarti A, Breton KH, Antonarakis SE, Waber PG, Boehm C, Kazazian HH (1984) Non-uniform recombination within the human ß-globin gene cluster. *Am J Hum Genet* 36: 1239

Chan V, Chan TK, Chehab FF, Todd D (1987) Distribution of ß-thalassaemia mutations in South China and their association with haplotypes. *Am J Hum Genet* 41: 678

Chang JC, Kan YW (1979) ß° thalassaemia, a nonsense mutation in man. *Proc Natl Acad Sci USA* 76: 2886

Chatterjea JB, Swarup S, Ghosh SK, Ray RN (1957) Incidence of haemoglobin E and thalassaemia trait in Bengalees. *Bull Calcutta Sch Trop Med* 5: 159

Chatterjea JB (1961) Haemoglobin-H Thalassaemia. *Bull Calcutta Sch Trop Med* 9: 93

Chebab FF, Der Kaloustian V, Khouri FP, Deeb SS, Kan YW (1987) The molecular basis of β-thalassemia in Lebanon: Application to prenatal diagnosis. *Blood* 69: 1141

Chehab FF, Kan YW (1989) Detection of specific DNA sequences by fluorescence amplification: a color complementation assay. *Proc Natl Acad Sci USA* 86: 9178

Chebloune Y, Pagnier J, Trabuchet G, Faure C, Verdier G, Labie D, Nignon V (1988) Structural analysis of the 5' flanking region of the ß-globin gene in African sickle cell anemia patients: Further evidence for three origins of the sickle cell mutation in Africa. *Proc Natl Acad Sci USA* 85: 4431

Chein A, Edgar DB, Trela JM (1976) Deoxyribonucleic acid polymerase from the extreme thermophile *Thermus aquaticus*. *J Bacteriol* 127: 1550

Cheng T, Orkin SH, Antonarakis SE, Potter MJ, Sexton JP, Markam AF, Giardina PJV, Li A, Kazazian HH (1984) ß-thalassaemia in Chinese: use of *in vivo* RNA analysis and oligonucleotide hybridisation in systematic characterisation of molecular defects. *Proc Natl Acad Sci USA* 81: 2821

Chouhan DM, Sharma RS, Parekh JG (1970) Alpha-thalassaemia in India. *J Indian Med Ass* 54: 364

Cimino GD, Metchette K, Isaacs ST, Zhu YS (1990) More false-positive problems. *Nature* 345: 773

Clegg JB, Naughton MA, Weatherall DJ (1968) Separation of the alpha and beta chains of human haemoglobin. *Nature* 219: 69

Clegg JB, Weatherall DJ, Milner PG (1971) Haemoglobin Constant Spring-α chain termination mutant? *Nature* 234: 337

Clegg JB, Weatherall DJ, Contopoulos-Griva I, Caroutsos K, Poungouras P, Tsevrenis H (1974) Haemoglobin Icaria, a new chain termination mutant which causes α-thalassaemia. *Nature* 251: 245

Colah R, Dastur AE, Nadkarni A, Pawar A, Mudera VC, Gorakshakar AC, Malapur RP (1990) Prenatal Diagnosis of Thalassaemia in Bombay- a five year experience. Proceedings of the First National Congress on Prenatal Diagnosis and Therapy, Bombay, p 5

Collins FS, Weissman SM (1984) The molecular genetics of human hemoglobin. *Progress in Nucleic Acids Research and Molecular Biology*. (ed Cohn WE, Moldave K) New York Academic Press p 315

Cooper DN, Schmidtke J (1984) DNA restriction fragment length plymorphisms and heterozygosity in the human genome. *Hum Genet* 66: 1

Coutelle C, Williams C, Handyside A, Hardy K, Winston R, Williamson R (1989) Genetic analysis of DNA from single human oocytes: a model for preimplantaion diagnosis of cystic fibrosis. *Br Med J* 299: 22

Darr A, Modell B (1988) The frequency of consanguineous marriage among British Pakistanis. *J Med Genet* 25: 186

Dash S, Vashist K, Sehgal S, Devi S, Verma A (1988) Isoelectric focusing for measurement of haemoglobin Bart's levels in cord blood & detection of alpha thalassaemia. *Indian J Med Res* 87: 62

De Jong WW, Khan PM, Bernini LF (1975) Hemoglobin Koya Doya: High frequency of a chain termination mutant. *Am J Hum Genet* 27: 81

Deragan JM, Sinnett D, Mitchell G, Potier M, Labuda D (1990) Use of Gamma Irradiation to Eliminate DNA Contamination for PCR. *Nucl Acids Res* 18: 6149

Dianzani I, Camaschella C, Saglio G, Forrest S, Ramus S, Cotton RGH (1991) Simultaneous Screening for ß-Thalassemia Mutations by Chemical Cleavage of Mismatch. *Genomics* 11: 48

Diaz-Chico JC, Yang KG, Stoming TA, Efremov DG, Kutlar A, Lutlar F, Aksoy M, Altay C, Gurgey A, Kininc Y, Huisman THJ (1988) Mild and severe ß-thalassemia among homozygotes from Turkey: Identification of the types by hybridisation of amplified DNA with synthetic probes. *Blood* 71: 248

Di Marzo R, Maggio A, D'Agostino S, Dowling CE, Wong C, Kazazian HH (1988) A rapid DNA method for first-trimester prenatal diagnosis. *Brit J Haematol* 70: 504

Dokras A, Sargent IL, Ross C, Gardner RL, Barlow DH (1990). Trophectoderm biopsy in human blastocysts. *Hum Reprod* 5: 821

Dokras A, Sargent IL, Gardner RL, Barlow DH (1991a) Human trophectoderm biopsy and chorionic gonadotrophin (HCG) secretion. *Hum Reprod* 6: 1453

Dokras A, Sargent IL, Ross C, Gardner RL, Barlow DH (1991b) The human blastocyst: morphology and human chorionic gonadotrophin secretion *in vitro*. *Hum Reprod* 6: 1143

Dozy AM, Kan YW, Embury SH, Mentzer WC, Wang WC, Lubin B, Davis JR, Koenig HM (1979) α-globin gene organization in blacks precludes the severe form of α-thalassemia. *Nature* 280:605

Drysdale HC, Higgs DR (1989) α-Thalassaemia in an Asian Indian. *Brit J Haematol* 72: 264

Efstratiadis A, Posakony JW, Maniatis T, Lawn RM, O'Connell C, Spritz RA, deRiel JK, Forget BG, Weissman SM, Slightom JL, Blechl AE, Smithies O, Baralle FE, Shoulders CC, Proudfoot NJ (1980) The structure and evolution of the human ß-globin gene family. *Cell* 21: 653

Embury SH, Miller JA, Dozy AM, Kan YW, Chan V, Todd D (1980) Two different molecular organizations account for the single α-globin gene of the α–thalassaemia-2 genotype. *J Clin Invest* 66: 1319

Feinberg AP, Vogelstein B (1983) A technique for radiolabelling DNA restriction endonuclease fragments to a high specific activity. *Anal Biochem* 132: 6

Fessas P (1963) Inclusions of hemoglobin in erythroblasts and erythrocytes of thalassemia. *Blood* 21: 21

Fessas P, Yataganas X (1968) Intraerythroblastic instability of hemoglobin $ß_4$ (Hgb H). *Blood* 31: 323

Firth HV, Boyd PA, Chamberlain P, MacKenzie IZ, Lindenbaum RH, Huson SM (1991) Severe limb abnormalities after chorion villus sampling at 56-66 days gestation. *Lancet* 337: 762

Fodde R, Losekoot M, van den Broek MH, Oldenburg M, Rashida N, Schreuder A, Wijnen JT, Giordano PC, Nayudu NVS, Meera Khan P, Bernini LF (1988) Prevalence and molecular heterogeneity of alfa+ thalassemia in two tribal populations from Andhra Pradesh, India. *Hum Genet* 80: 157

Fucharoen S, Winichagoon P, Thonglairoam V, Siriboon W, Siritanaratkul N, Kanokpongsakdi S, Vantanasiri C (1991) Prenatal diagnosis of thalassemia and hemoglobinopathies in Thailand: experience from 100 pregnancies *Southeast Asian J Trop Med Public Health* 22: 16

Fuchs R, Blakesley R (1983) Use of type II restriction endonucleases. *Method Enzymol* 100: 3

Gardner RL, Edwards RG (1968) Control of the sex ratio at term in the rabbit by transferring sexed blastocytes. *Nature* 218: 346

Gatrad AR, Read AP, Watson GH (1984) Consanguinity and complex cardiac anomalies with situs ambiguus. *Arch Dis Child* 59: 242

Gerhard DS, Kidd KK, Kidd JR, Egeland JA, Housman DE (1984) Identification of a recent recombination event within the human ß-globin gene cluster. *Proc Natl Acad Sci USA* 81: 7875

Goldsmith ME, Humphries RK, Ley T, Cline A, Kantor JA, Nienhuis AW (1983) "Silent" nucleotide substitution in a β^+ thalassemia globin gene activates splice site in coding sequence RNA. *Proc Natl Acad Sci USA* 80: 2318

Gonzalez-Redondo JM, Stoming TA, Kutlar F, Kutlar A, Lanclos KD, Howard EF, Fei YJ, Ahsoy M, Altay C, Gurgey A, Basak AN, Efremov GD, Petkov G, Huisman THJ (1989) A C-T substitution at -101 in a conserved DNA sequence of the promotor region of the ß-globin gene is associated with "silent" ß-thalassaemia. *Blood* 73: 1705

Goosens M, Dozy AM, Embury SH, Zachariades Z, Hadjiminas MG, Stamatoyannopoulos G, Kan YW (1980) Triplicated α-globin gene loci in humans. *Proc Natl Acad Sci USA* 77: 518

Government of India Census (1991) (in preparation)

Green MR (1986) Pre-mRNA splicing. *Annu Rev Genet* 20: 671

Grosveld FG, van Assendelft B, Greaves DR, Kollias G (1987) Position-independent, high-level expression of the human ß-globin gene in transgenic mice. *Cell* 51: 975

Gusella JF, Varsanyi-Breiner A, Kao FT, Jones C, Puck TT, Keys C, Orkin S, Housman D (1979) Precise localization of human ß-globin gene complex on chromosome 11. *Proc Natl Acad Sci USA* 76: 5239

Gyllensten UB, Erlich HA (1988) Generation of single-stranded DNA by the polymerase chain reaction and its application to direct sequencing of the HLA-DQA locus. *Proc Natl Acad Sci USA* 85: 7652

Hahneman N, MohrJ (1968) Genetic diagnosis in the embryo by means of biopsy from extra-embryonic membrane. *Bull Europ Soc Hum Genet* 2: 23

Haldane JBS (1949) Disease and Evolution. *Ricera Sci* 19 (Suppl): 2

Handyside AH, Penketh RJA, Winston RML, Pattison JK, Delhanty JDA, Tuddenham EGD (1989) Biopsy of preimplantation embryos and sexing by DNA amplification. *Lancet* i: 347

Handyside AH, Kontogianni EH, Hardy K, Winston RML (1990) Pregnancies from biopsied human preimplantation embryos sexed by Y-specific DNA amplification. *Nature* 344: 768

Higgs DR, Goodbourn SEY, Lamb J, Clegg JB, Weatherall DJ, Proudfoot NJ (1983) α-Thalassaemia caused by a polyadenylation signal mutation. *Nature* 306: 398

Higgs DR, Vickers MA, Wilkie AOM, Pretorius I-M, Jarman AP, Weatherall DJ (1989) A Review of the Molecular Genetics of the Human α-Globin Gene Cluster. *Blood* 73: 1081

Higgs DR, Wood WG, Jarman AP, Sharpe J, Lida J, Pretorius I-M, Ayyub H (1990) A major positive regulatory region located far upstream of the human α-globin gene locus. *Genes and Development* 4: 1588

Hill AVS, Bowden DK, Trent RJ, Higgs DR, Oppenheimer SJ, Thein SL, Mickleson KNP, Weatherall DJ, Clegg JB (1985) Melanesians and Polynesians share a unique α-thalassaemia mutation. *Am J Hum Genet* 37: 571

Holding C, Monk B (1989) Diagnosis of ß-thalassaemia by DNA amplification in single blastomeres from mouse preimplantation embryos. *Lancet* ii: 532

Horn GT, Richards B, Klinger KW (1989) Amplification of a highly polymorphic VNTR segment by the polymerase chain reaction. *Nucl Acid Res* 17: 2140

Huang S-Z, Wong C, Antonarakis SE, Ro-lein T, Lo WHY, Kazazian HH (1986) The same TATA box ß-thalassaemia mutation in Chinese and US Blacks: Another example of independent origins of mutation. *Hum Genet* 74: 152

Huisman THJ (1990) Frequencies of common ß-thalassaemia alleles among different populations: Variability in clinical severity. *Brit J Haematol* 75: 454

Hultman T, Stahl S, Hornes E, Uhlen M (1989) Direct solid phase sequencing of genomic and plasmid DNA using magnetic beads as a support. *Nucl Acids Res* 17: 4937

Hundrieser J, Deka R, Gogoi BC (1987) α-Thalassemia in the Kachari population of Assam (India). *Hemoglobin* 11: 517

Hutton JH (1931) Ethnographic notes by various hands. In: "Census of India, vol 1 India, Part III B" Government of India Press, Simla

Hyman CB, Chapman CL, Neufeld N, Santulli T (1990) Is there a place for hypertransfusion/desferal when bone marrow transplant is an option? *Blood* 76: 63

Indian Council for Medical Research (1988) Collaborative study on the prevalence of haemoglobinopathies in India. (manuscript in preparation)

Jacobsen H, Klenow H, Ovargaard H (1974) The N-terminal amino acid sequences from DNA polymerase 1 from *Escherichia coli* and of the large and small fragments obtained by limited proteolysis. *Eur J Biochem* 45: 623

Jain PK, Verma IC, Petrou M, Modell B, Old JM (1991) Molecular characterization of beta thalassemia in North India. Proceedings of the 8th International Congress of Human Genetics, Washington D.C. (abstr) 1024, p 192

Jankovic L, Efremov GD, Petkov G (1989) Three novel mutations leading to ß-thalassemia. *Blood* 74: 226 (abstr)

Jarmulowicz M (1989) Embryo Biopsy. *Lancet* i: 547

Jinno Y, Yoshiura K, Niikawa N (1990) Use of Psoralin as Extinguisher of Contaminated DNA in PCR. *Nucl Acids Res* 18: 6739

Jones KW, Singh L, Edwards RG (1987) The use of probes for the Y chromosome in preimplantation embryo cells. *Hum Reprod* 2: 439

Kan YW, Nathan DG (1970) Mild thalassemia: The result of interactions of alpha and beta thalassemia genes. *J Clin Invest* 49: 635

Kan YW, Golbus MS, Klein P, Dozy AM (1975) Successful application of prenatal diagnosis in a pregnancy at risk for homozygous ß-thalassaemia. *New Engl J Med* 292: 1096

Kanavakis E, Wainscoat JS, Wood WG, Weatherall DJ, Cao A, Furbeta M, Galanello R, Georgiou D, Sophocleous T (1982) The interaction of α-thalassaemia with heterozygous ß thalassaemia. *Brit J Haematol* 52: 465

Kattamis C, Efremove G, Pootrakul S (1981) Effectiveness of one tube osmotic fragility screening in detecting ß-thalassemia trait. *J Med Genet* 18: 266

Kazazian HH, Orkin SH, Markham AF, Chapman CR, Youssoufian H, Waber PG (1984a) Quantification of the close association between DNA haplotypes and specific ß-thalassaemia mutations in Mediterraneans. *Nature* 310: 152

Kazazian HH, Orkin SH, Antonarakis SE, Sexton JP, Boehm CD, Goff SC, Waber PG (1984b) Molecular characterization of seven ß-thalassaemia mutations in Asian Indians. *EMBO J* 3: 593

Kazazian HH, Dowling CE, Waber PG, Huang S, Wilson HYLo (1986) The sprectrum of ß-thalassaemia genes in China and Southeast Asia. *Blood* 68: 964

Kazazian HH, Boehm CD (1988) Molecular basis and prenatal diagnosis of ß-thalassaemia. *Blood* 72: 1107

Kazazian HH, Dowling CE, Hurwitz RL, Coleman M, Adams JG (1989) Thalassemia mutations in exon 3 of the ß-globin gene often cause a dominant form of thalassemia and show no predilection for malarial-endemic regions of the world. *Am J Hum Genet* 45:242 (abstr)

Kazazian HH (1990) The Thalassemia Syndromes: Molecular Basis and Prenatal Diagnosis in 1990. Seminars in Haematology, 27: 209

Kleihauer E, Braun H, Betke K (1957) Demonstration von fetalem haemoglobin in den Erythrocyten eines Blutausstriche. *Klin Wochenschr* 35: 635

Kogan SC, Marie Doherty M, Gitschier J (1987) An improved method for prenatal diagnosis of genetic diseases by analysis of amplified DNA sequences. *New Engl J Med* 317: 985

Kollia P, Gonzalez-Redondo JM, Stoming TA, Loukopoulos D, Politis C, Huisman THJ (1989) Frameshift codon 5 [FSC-5(-CT)] Thalassemia: A novel mutation detected in a Greek patient. *Hemoglobin* 13: 597

Kulozik AE, Wainscoat JS, Serjeant GR, Al-Awamy B, Essan F, Falusi Y, Hague SK, Hilali AM, Kate S, Ranasinghe WACP, Weatherall DJ (1986) Geographical survey of β^S globin gene haplotypes: evidence for an independent Asian origin of the sickle-cell mutation. *Am J Hum Genet* 39: 239

Kulozik A, Kar BC, Serjeant BE, Serjeant GR, Weatherall DJ (1988) α–thalassaemia in India: Its interaction with sickle cell disease. *Blood* 71: 467

Kwok S, Higuchi R (1989) Avoiding false positives with PCR. *Nature* 339: 237

Kwok S, Kellog DE, McKinney N, Spasic D, Goda L, Levenson C, Sninsky JJ (1990) Effects of primer-template mismatches on the polymerase chain reaction: Human immunodeficiency virus type 1 model studies. *Nucl Acids Res* 18: 999

Lapoumeroulie C, Pagnier J, Bank A, Labie D, Krishnamoorthy R (1986) ß-thalassemia due to a novel mutation in IVS-1 sequence donor site consensus sequence creating a restriction site. *Biochem Biophy Res Comm* 139: 709

Lawn RM, Efstratiadis A, O'Connell C, Maniatis T (1980) The nucleotide sequence of the human ß-globin gene. *Cell* 21: 647

Lebo RV, Saiki RK, Swanson K, Montano MA, Erlich HA, Golbus MS (1990) Prenatal diagnosis of α-thalassemia by polymerase chain reaction and dual restriction enzyme analysis. *Hum Genet* 85: 293

Lesko J, Snabes M, Handyside A, Hughes M (1991) Amplification of the Ctystic Fibrosis Δ508 mutation from single cells: Applications toward genetic diagnosis of the preimplantation embryo. Proceedings of the 8th International Congress of Human Genetics, Washington D.C., (abstr) 1210 p 223

Lewin B (1983) Genes. John Wiley & Sons, Inc., New York

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

Liebhaber CA, Kan YW (1983) α-Thalassaemia caused by an unstable α-globin mutant. *J Clin Invest* 71: 461

Liebhaber SA, Cash FE, Main DM (1985) Compensatory increase in α1-globin gene expression in individuals heterozygous for the α-thalassaemia-2 deletion. *J Clin Invest* 76: 1057

Liebhaber CA, Coleman MB, Adams JG, Cash FE, Steinberg MH (1987) Molecular basis for non-deletion α-thalssemia in American blacks $\alpha 2^{116GAG\text{-}UAG}$. *J Clin Invest* 80: 154

Lie-Injo LE, Cai S-P, Wahidijat I, Moeslischan S, Lim ML, Evangelista L, Doherty M, Kan YW (1989) ß-thalassaemia mutations in Indonesia and their linkage to ß haplotypes. *Am J Hum Genet* 45: 971

Little PFR, Annison G, Darling S, Williamson R, Cariba L, Modell B (1980) Model for antenatal diagnosis of ß-thalassaemia and other monogenic disorders by molecular analysis of linked DNA polymorphisms. *Nature* 285: 144

Longo MC, Berninger MS, Hartley JL (1990) Use of Uracil DNA Glycosylase to Control carry-over Contamination in Polymerase Chain Reactions. *Gene* 93: 125

Losekoot M, Fodde R, Harteveld CL, Van Heeren H, Giordano PC, Bernini LF (1990) Denaturing gradient gel electrophoresis and direct sequencing of PCR amplified genomic DNA: a rapid and reliable diagnostic approach to beta thalassaemia. *Brit J Haematol* 76: 269

Loukopoulos D, Hadji A, Papadakis M, Karababa P, Sinopoulou K, Boussiou M, Kollia P, Xenakis M, Antsaklis A, Mesoghitis S, Loutradi A, Fessas P (1990) Prenatal diagnosis of thalassemia and of the sickle cell syndromes in Greece. *Ann NY Acad Sci* 612: 226

Luo H-Y, Clarke BJ, Gauldie J, Patterson M, Liao S-K, Chui DHK (1988) A novel monoclonal antibody based diagnostic test for α-thalassaemia carriers due to the $(^{-SEA}/)$ deletion. *Blood* 72: 1589

Lucarelli G (1991) For debate: Bone marrow transplantation for severe thalassaemia, The view from Pesaro. *Brit J Haematol* 78: 300

Lynch JR (1987) The molecular basis of ß-thalassaemia in Thailand. D.Phil. thesis, University of Oxford

Ma LJ (1985) Cultural Diversity. In: "South East Asia: Realm of contrasts" (Ed Dutt AK) 3rd edition, West View Press, London.

Manca M, Cossu G, Angioni G, Gigliotti B, Bosisio AB, Gianazza E, Righetti PG (1986) Antenatal diagnosis of ß-thalassemia by

isoelectric focusing in immobilized pH gradients. *Am J Hematol* 22: 285

Medical Research International and the Society for Assisted Reproductive Technology (1990) *In vitro* fertilization-embryo transfer in the United States: 1988 results from the IVF-ET Registry. *Fertil Steril* 53: 13

Mehta BC, Dave VB, Joshi SR, Baxi AJ, Bhatia HM, Patel JC (1971) Study of haematological and genetical characteristics of Bhanushali community from North India. Proceedings of 2nd meeting of the Asian-Pacific Division of International Society of Haematology, Melbourne, p 16

Mehta BC, Gandhi SH, Singhal SB, Rao PS (1989) Naked eye single tube red cell osmotic fragility test for beta thalassemia trait: experience in antenatal clinic. *J Obstet Gynaecol India* 39: 450

Metherall JE, Collins FS, Pan J, Weissman SM, Forget BG (1986) ß°-thalassaemia caused by a base substitution that creates an alternative splice acceptor site in an intron. *EMBO J* 5: 2551

Misra RC (1989) Detection of alpha thalassaemia in sickle cell trait patients by Hb-Bart's screening & quantitation of Hb-A & Hb-S. *Indian J Med Res* 90: 459

Modell B, Letsky EA, Flynn DM, Peto R, Weatherall DJ (1982) Survival and desferrioxamine in thalassaemia major. *Brit Med J* 284: 1081

Modell BC, Petrou M (1983) The problem of Haemoglobinopathies in India. *Indian J Haematol* 1: 5

Modell BC, Petrou M, Ward RHT, Fairweather DVI, Rodeck C, Varnavides LA, White JM (1984) Effect of fetal diagnostic testing on the birth-rate of thalassaemia in Britain. *Lancet* ii: 1383

Moi P, Cash FE, Liebhaber SA, Cao A, Piratsu M (1987) An initiation codon mutation (AUG-GUG) of the human α1-globin gene: Structural characterisation and evidence for a mild thalassemic phenotype. *J Clin Invest* 80: 1416

Monk M, Handyside A, Hardy K, Whittingham DG (1987) Preimplantation diagnosis of deficiency of hypoxanthine phophoribosyl transferase in a mouse model for Lesch-Nyhan syndrome. *Lancet* i: 423

Monk M, Muggleton-Harris A, Rawlings E, Whittingham DG (1988) Preimplantaion diagnosis of HPRT-deficient male, and carrier female mouse embryos by trophectoderm biopsy. *Hum Reprod* 3: 377

Monk M, Holding C (1990) Amplification of a ß-haemoglobin sequence in individual human oocytes and polar bodies. *Lancet* i: 985

Monk M (1991) Preimplantation diagnosis of genetic disease. Paper presented at the 8th International Congress of Human Genetics, Washington D.C.

Morle F, Lopez B, Henni T, Godet J (1985) α-Thalassaemia associated with the deletion of two nucleotides at position -2 and -3 preceding the AUC codon. *EMBO J* 4: 1245

Morle F, Starck J, Godet J (1986) α-thalassaemia due to the deletion of nucleotides at -2 and -3 preceding the AUG initiation codon affects translation efficiency both in vitro and in vivo. *Nucl Acids Res* 14: 3279

Mount SM (1982) A catalogue of splice junction sequences. *Nucl Acids Res* 10: 459

Mourant AE (1983) Asia. In: "Blood Relations Blood Groups and anthroplogy" (ed Mourant AE) Oxford Scientific Publications p 57

MRC Working party on the Evaluation of Chorion Villus Sampling (1991) Medical Research Council European Trial of chorion villus sampling. *Lancet* 337: 1491

Mullis KB, Faloona FA (1987) Specific synthesis of DNA *in vitro* via a polymerase-catalyzed chain reaction. *Method Enzymol* 155: 335

Myers RM, Lumelsky N, Lerman LS, Maniatis T (1985) Detection of single base substitutions in total genomic DNA. *Nature* 313: 495

Nathan DG, Stossel TB, Gunn RB, Zarkowsky HS, Laforet MT (1969) Influence of hemoglobin precipitation on erythrocyte metabolism in in alpha and beta thalassemia. *J Clin Invest* 48: 33

Nathan DG, Smith HO (1975) Restriction endonucleases in the analysis and restructuring of DNA molecules. *Annu Rev Biochem* 44: 273

Newton CP, Graham A, Heptinstall LE, Powell SJ, Summers C, Kalsheker N, Smith JC, Markham AF (1989) Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS). *Nucl Acids Res* 17: 2503

Old JM, Higgs DR (1983) Gene analysis. In: "Methods in Haematology. Vol 6 The thalassaemias" (ed Weatherall DJ) Churchill Livingstone, Edinburgh p74

Old JM, Petrou M, Modell B, Weatherall DJ (1984) Feasibility of antenatal diagnosis of ß-thalassaemia by DNA polymorphisms in Asian Indian and Cypriot populations. *Brit J Haematol* 57: 255

Old JM, Heath C, Fitches A, Thein SL, Jeffreys AJ, Petrou M, Modell B, Weatherall DJ (1986a) Meiotic recombination between two polymorphic restriction sites within the ß-globin gene cluster. *J Med Genet* 23: 14

Old JM, Heath C, Fitches A, Thein SL, Weatherall DJ, Warren R, McKenzie C, Rodeck CH, Modell B, Petrou M, Ward RHT (1986b) First-trimester fetal diagnosis for haemoglobinopathies: report on 200 cases. *Lancet* ii: 763

Old JM, Varawalla NY, Weatherall DJ (1990) Rapid detection and prenatal diagnosis of ß-thalassaemia: studies in Indian and Cypriot populations in the U.K. *Lancet* 336: 834

Olivieri NF, Chang LS, Poon AO, Michelson AM, Orkin SH (1987) An α-globin gene initiation codon mutation in a Black family with Hb H disease. *Blood* 70: 729

Orkin SH, Old JM, Weatherall DJ, Nathan DG (1979) Partial deletion of ß-globin gene DNA in certain patients with ß° thalassaemia. *Proc Natl Acad Sci USA* 76: 2400

Orkin SH, Goff SC, Hechtman RL (1981) Mutation in an intervening sequence splice junction in man. *Proc Natl Acad Sci USA* 78: 5041

Orkin SH, Kazazian HH, Antonarakis SE, Ostrer H, Goff SD, Sexton JP (1982a) Abnormal RNA processing due to an exon mutation of the $ß^{E}$-globin gene. *Nature* 300: 768

Orkin SH, Kazazian HH, Antonarakis SE, Goff SC, Boehm CD, Sexton JP, Waber PG, Giardina PJV (1982b) Linkage of ß-thalassaemia mutations and ß-globin gene polymorphisms with DNA

polymorphisms in human ß-globin gene cluster. *Nature* 296: 627

Orkin SH, Sexton JP, Cheng T-C, Goff SC, Giardina PJV, Lee JI, Kazazian HH (1983) ATA box transcription mutation in ß-thalassemia. *Nucl Acids Res* 11: 4727

Orkin SH, Kazazian HH (1984) The mutation and polymorphism of the human ß-globin gene and its surrounding DNA. *Annu Rev Genet* 18: 131

Orkin SH, Antonarakis SE, Loukopoulos D (1984) Abnormal processing of ßKnossos RNA. *Blood* 64: 311

Orkin SH, Cheng TC, Antonarakis SE, Kazazian HH (1985) Thalassemia due to a mutation in the cleavage-polyadenylation signal of the human ß-globin gene. *EMBO J* 4: 453

Padanilam BJ, Felice AE, Huisman THJ (1984) Partial deletion of the 5' ß globin gene region causes ß° thalassaemia in members of an American black family. *Blood* 64: 941

Parikh P, Cotton M, Boehm C, Kazazian HH (1990) Ethnic distribution of ß-thalassaemia in Indian subcontinent. *Lancet* ii: 1006

Patra SB, Patel SR, Giri DD (1983) Hb Bart's in tribals of Western India. *Indian J Haematol* 1: 46

Pergament E (1991) Preimplantation diagnosis: a patient perspective. *Prenatal Diagnosis* 11: 493

Petrou M, Modell B, Darr A, Old J, Kin E, Weatherall DJ (1990) Antenatal Diagnosis: How to Deliver a Comprehensive Service in the United Kingdom. *Ann NY Acad Sci* 612: 251

Pirastu M, Kan YW, Cao A, Conner BJ, Teplitz RL, Wallace RB (1983) Prenatal diagnosis of ß-thalassemia: detection of a single nucleotide mutation in DNA. *New Engl J Med* 309: 284

Pirastu M, Saglio G, Chang JC, Cao A, Kan YW (1984) Initiation codon mutation as a cause of α-thalassemia. *J Biol Chem* 259: 12315

Piratsu M, Galanello R, Doherty MA, Tuveri T, Cao A, Kan YW (1987) The same ß-globin gene mutation is present on nine different ß-thalassaemia chromsomes in a Sardinian population. *Proc Natl Acad Sci USA* 84: 2882

Pootrakul S, Wasi P, Na-Nakorn S, Pravatmuang P (1975) Hemoglobin Bart's and hemoglobin Constant Spring in the cord

blood. Istambul Symposium on Abnormal Hemoglobins and Thalassemia (ed by M Aksoy) p 111

Popovich BW, Rosenblatt DS, Kendall AG, Nishioka Y (1986) Molecular characterization of an atypical ß-thalassaemia caused by a large deletion in the 5' ß-globin gene region. *Am J Hum Genet* 39: 797

Puissant F, van Rysselberge M, Barlow P, Deweze J, Leroy F (1987) Embryo scoring as a prognostic tool in IVF treatment. *Hum Reprod* 2: 705

Rao PSS, Inbaraj SG, Jesudian G (1972) Rural and urban differentials in consanguinity. *J Med Genet* 9: 174

Rebello MT, Hackett G, Smith J, Loeffler FE, Robson S, Maclachlan N, Beard RW, Rodeck CH, Williamson R, Coleman DV (1991) Extraction of DNA from amniotic fluid cells for the early prenatal diagnosis of genetic disease. *Prenatal Diagnosis* 11: 41

Reeders ST, Breuning MH, Davies KE, Nicholls RD, Jarman AP, Higgs DR, Pearson PL, Weatherall DJ (1985) A highly polymorphic DNA marker linked to adult polycystic kidney disease on chromosome 16. *Nature* 317: 542

Richardson CC (1981) Bacteriophage T4 polynucleotide kinase. In: " The enzymes" 3rd edition (ed Boyer PD) vol 14 Academic Press, New York p 229

Rosatelli C, Leoni GB, Tuveri T, Scalas MT, DiTucci A, Cao A (1987) ß-thalassaemia mutations in Sardinians: implications for prenatal diagnosis. *J Med Genet* 24: 97

Safaya S, Rieder RF (1988) Dysfuctional α-globin gene in hemoglobin H disease in blacks. *J Biol Chem* 263: 4328

Saiki RK, Bugawan TL, Horn GT, Mullis KB, Erlich HA (1986) Analysis of enzymatically amplified ß-globin gene and HLA-DQ alpha DNA with allele-specific oligonucleotide probes. *Nature* 324: 163

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

Saiki RK, Chu-An Chang BS, Levenson CH, Warren TC, Boehm CD, Kazazian HH, Erlich HA (1988b) Diagnosis of sickle cell anemia and ß-thalassemia with enzymatically amplified DNA and

nonradioactive allele-specific oligonucleotide probes. *New Engl J Med* 319: 537

Saiki RK (1989) The design and Optimization of the PCR. In: "PCR Technology: Principles and Applications for DNA Amplification" (ed Erlich HA) Stockton Press p 7

Saiki RK, Walsh PS, Levenson CH, Erlich HA (1989) Genetic analysis of amplified DNA with immobilised sequence specific oligonucleotide probes. *Proc Natl Acad Sci USA* 86: 6230

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual. 2nd edition, Cold Spring Harbor Laboratory Press, New York

Sangani B, Sukumaran PK, Mahadik C, Yagnik H, Telang S, Vas F, Oberroi RA, Modell B, Merchant SM (1990) Thalassaemia in Bombay: the role of medical genetics in developing countries. *Bull WHO* 68: 75

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

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

Sharma RS, Kabir SMS, Baxi AJ, Shanbaug SR, Bhatia HM (1971) Haematological and genetical studies in the Saraswat and Lohana communities. Proceedings of the Annual Meeting of the Indian Society of Haematology and Blood Transfusion, Mangalore, p 27

Semenza GL, Delgrosso K, Poncz M, Malladi P, Schwartz E, Surrey S (1984) The silent carrier allele: ß thalassaemia without a mutation in the ß-globin gene or its immediate flanking regions. *Cell* 39: 123

Semenza GL, Dowling CE, Kazazian HH (1989) *Hinf* I polymorphism 3' to human ß-globin gene detected by the polymerase chain reaction (PCR). *Nucl Acids Res* 17: 2376

Setlow RB, Setlow JK (1962) ?? *Proc Natl Sci Acad USA* 48: 1250

Siddoo JK, Siddoo SK, Chase WH, Morgan-Dean L, Perry WH (1956) Thalassaemia in Sikhs. *Blood* 11: 197

Sinha ND, Biernat J, McManus J, Koster H (1984) Polymer support oligonucleotide synthesis XVIII: use of ß-cyanoethyl-N,N-dialkylamino-/N- morpholino phosphoramidite of deoxynucleosides for the synthesis of DNA fragments

simplifying deprotection and isolation of the final product. *Nucl Acids Res* 12: 4539

Siniscalco M, Bernini L, Filippi G, Latte B, Khan M, Piomelli S, Rattazzi M (1966) Population genetics of haemoglobin variants, thalassaemia and glucose-6-phosphate dehydrogenase deficiency, with particular reference to the Malaria hypothesis. *Bull WHO* 34: 379

Slightom JL, Blechl AE, Smithies O (1980) Human fetal $^{G}\gamma$– and $^{A}\gamma$–globin genes: Complete nucleotide sequence suggest that DNA can be exchanged between these duplicated genes. *Cell* 21: 627

Southern EM (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J Mol Biol* 98: 503

Spritz RA, Jagadeeswaran P, Choudary PV, Biro PA, Elder JT, deRiel JK, Manley JL, Gefter ML, Forget BG, Weismann SM (1981) Base substitution in an intervening sequence of a ß$^{+}$-thalassemic human globin gene. *Proc Natl Acad Sci USA* 78: 2455

Stern MA, Kynoch PA, Lehmann H (1968) ß-thalassaemia, glucose-6-phosphate dehydrogenase deficiency, and haemoglobin D-Punjab in Pathans. *Lancet* 1: 1284

Suggs SV, Wallace RB, Hirose T, Kawashima EH, Itahura K (1981) Use of synthetic oligonucleotides as hybridisation probes: isolation of cloned cDNA sequences for human ß^{2}-microglobulin *Proc Natl Acad Sci USA* 67: 6613

Sukumaran PK (1974) Abnormal haemoglobins in India. In: "Trends in Haematology" (ed by NN Sen) p 225

Summers PM, Campbell JM, Miller MW (1988) Normal in-vivo development of marmoset monkey embryos after trophectoderm biopsy. *Hum Reprod* 3: 389

Swarup S, Bannerji PG, Ghosh SK, Chatterjea JB (1965) Haemoglobin Bart's in Bengalee blood. *Bull Calcutte Sch Trop Med* 13: 47

Tabor A, Philip J, Madsen M, Bang J, Obe EB, Norgaard-Pedersen B (1986) Randomised control trial of genetic amniocenetsis in 4606 low-risk women. *Lancet* i: 1287

Tabor S, Richardson CC (1987) DNA synthesis analysis with a modified bacteriophage T7 DNA polymerase. *Proc Natl Acad Sci USA* 84: 4767

Tamagnini GP, Lopes MC, Castanheira ME, Wainscoat JS, Wood WG (1983) ß+ thalassaemia-Portugese type: Clinical, haematological and molecular studies of a newly defined form of ß thalassaemia. *Br J Haematol* 54: 189

Thapar R (1966) A History of India Vol 1, Penguin Books Ltd, Harmondsworth.

Thein SL, Wainscoat JS, Lynch JR, Weatherall DJ, Sampietro M, Fiorelli G (1985) Direct detection of ß° 39 thalassaemic mutation with *Mae*I. *Lancet* i: 1095

Thein SL, Wallace RB (1986) The use of synthetic oligonucleotides as specific hybridization probes in the diagnosis of genetic disorders. In: "Human genetic diseases: A practical approach" (ed Davies KE) IRL Press, Oxford p 33

Thein SL, Wainscoat JS, Sampietro M, Old JM, Cappellini D, Florelli G, Modell B, Weatherall DJ (1987) Association of thalassaemia intermedia with a beta-globin gene haplotype. *Br J Haematol* 65: 367

Thein SL, Hesketh C, Wallace RB, Weatherall DJ (1988) The molecular basis of thalassaemia major and thalassaemia intermedia in Asian Indians: applications to prenatal diagnosis. *Brit J Haematol* 70: 225

Thein SL, Hesketh C, Taylor P, Temperley J, Hutchinson RM, Old JM, Wood WG, Clegg JB, Weatherall DJ (1990) Molecular basis for dominantly inherited inclusion body ß-thalassaemia *Proc Natl Acad Sci USA* 87: 3924

Thomas CA (1966) Recombination of DNA molecules. *Prog Nucleic Acid Research and Molecular Biol* 5: 315

Trapani FD, Marino M, D'Alcamo E, Abate I, D'Aostino S, Lauricella S, Musico M, Orlandi F, Sammarco P, Maggio A, Modell B (1991) Prenatal diagnosis of haemoglobin disorders by cordocentesis at 12 weeks' gestation. *Prenatal Diagnosis* 11: 899

Trecartin RF, Liebhaber SA, Chang JC, Lee YW, Kan YW (1981) ß°-thalassemia in Sardinia is caused by a nonsense mutation. *J Clin Invest* 68: 1013

Treisman R, Proudfoot NJ, Shander M, Maniatis T (1982) A single base change at a splice site in a ß°-thalassemic gene causes abnormal RNA splicing. *Cell* 79: 903

Trounson A, Mohr L (1983) Human pregnancy following cryopreservation, thawing and trasfer of an 8 cell embryo. *Nature* 305: 707

Wainscoat JS, Kanavakis E, Wood WG, Letsky EA, Huehns ER, Marsh GW, Higgs DR, Clegg JB, Weatherall DJ (1983) Thalassaemia intermedia in Cyprus-The interaction of α- and ß-thalassaemia. *Brit J Haematol* 53: 411

Wainscoat JS, Hill AVS, Boyce AL, Flint J, Hernandez M, Thein SL, Old JM, Lynch JR, Fahisi AG, Weatherall DJ, Clegg JB (1986) Evolutionary relationships of human populations from an analysis of nuclear DNA polymorphisms. *Nature* 319: 491

Wasi P, Disthasongchan P, Na-Nakorn S (1968) The effect of iron deficiency on the levels of hemoglobins A_2 and E. *J Lab Clin Med* 71: 85

Weatherall DJ, Clegg JB (1981) The thalassaemia syndromes. 3rd edition, Blackwell Scientific Publications, Oxford

Weatherall DJ, Pippard MJ, Callender ST (1983) Iron loading in thalassemia-five years with the pump. *New Engl J Med* 308: 456

Weatherall DJ, Clegg JB, Higgs DR, Wood WG (1989) The Hemoglobinopathies. In: "The Metabolic Basis of Inherited Disease" 6th edition (eds Scriver CR, Beaudet AL, Sly WS) NewYork, NY, McGraw-Hill Book Co p 2281

Weatherall DJ (1991) For debate: Bone marrow transplantation for severe thalassaemia, to be or not to be. *Brit J Haematol* 78: 301

WHO Working Group (1982) Hereditary anaemias: Genetic basis, clinical features, diagnosis and treatment. *Bull WHO* 60: 643

WHO Memorandum (1983) Community control of hereditary anaemias. *Bull WHO* 61: 63

WHO Advisory Group (1985) Community Approaches to the Control of Hereditary Diseases. *WHO Unpublished Document* HDP/WG/85.10

Willcox M, Bjorkman A, Brohult J, Pehrson PO, Rombo L, Bengtsson E (1983) A case control study in northern Liberia of Plasmodium falciparum malaria in haemoglobin S and beta-thalassaemia traits. *Ann Trop Med Parasitol* 77: 239

Wong C, Antonarakis SE, Goff SC, Orkin SH, Boehm CD, Kazazian HH (1986) On the origin and spread of ß-thalassaemia: Recurrent observations of four mutations in different ethnic groups. *Proc Natl Acad Sci USA* 85: 6529

Wong C, Dowling CE, Saiki RK, Higuchi RG, Erlich HA, Kazazian HH (1987) Characterization of ß thalassaemia mutations using direct genomic sequencing of amplified single copy DNA. *Nature* 330: 384

Wong C, Antonarakis SE, Goff SC, Orkin SH, Forget BG, Nathan DG, Giardina PJV (1989) ß-thalassemia due to two novel nucleotide substitutions in consensus acceptor splice sequences of the ß-globin gene. *Blood* 173: 914

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

Yang KG, Kutlar F, George E, Wilson JB, Kutlar A, Stoming TA, Gonzalez-Redondo JM, Huisman THJ (1989) Molecular characterisation of ß-globin gene mutations in Malay patients with Hb E-ß-thalassaemia and thalassaemia major. *Brit J Haematol* 72: 73

Zhang J, Cai S, Kan YW (1990) Prenatal diagnosis of Thalassemia in South China. *Ann NY Acad Sci* 612: 264

Expression of Vi antigen-like reactivity

When 16 randomly selected *E coli* O111:B4 outbreak strains were grown on agar plates at 22°C for two weeks, small colonies appeared on top of the original larger ones. At the same time, the margins of the original colonies acquired a laced appearance. Colonies derived from the original large ones by isolation and re-culture reacted with O111 antiserum. By contrast, colonies derived from the small colonies or the margin area no longer reacted with this antiserum nor with B4 antiserum. However, they expressed a vigorous positive reaction with two commercial (Wellcome; Behringwerke) and one laboratory-prepared anti-Vi serum. O111 reactivity could not be detected in the colonies that had Vi antigen-like reactivity, even after treatment at 100°C. As determined by 'API 20E' identification kits (API, API Systems S.A., France), these colonies were identical to the original strains; the outbreak strains were also identical to the original strains with respect to invasiveness and toxin production. 3 *E coli* O111:B4 isolates and their Vi antigen-positive convertants were investigated by the FAS test. The convertants were more adherent (+++) to Hep-2 cells than were the parent strains (+ to ++). Vi antigen-like reactivity was temperature-dependent; the outbreak strains remained reactive with O111 antisera when incubated for two weeks at 35°C. Colonies with Vi antigen-like reactivity were not recovered from any of the primary stool cultures on standard incubation at 35°C. Neither the reference *E coli* O111:B4 strain nor 75 sequential *E coli* strains (various serotypes) isolated from clinical specimens in our laboratory converted to colonies that expressed Vi antigen-like reactivity when grown at 22°C.

All but 2 outbreak strains were ampicillin-resistant. Colonies around ampicillin disks (10 μg ampicillin/disc, Ab Bio-Disk, Solna, Sweden) on culture plates converted to Vi antigen-like reactivity within two days. However, conversion in the presence of ampicillin was not as temperature-limited as before, since it occurred even at 35°C. Bacterial growth was uniform over the entire agar plate, with no visible differences in the size or number of the colonies closer or farther from the disc. All tested colonies near (diameter of the area 29–30 mm) the disc reacted with Vi antisera and lost their O111 reactivity. Elsewhere, the colonies remained reactive with antiserum to O111 antigen; no Vi conversion was detected. Additionally, to see whether the emergence of Vi antigen-like reactivity could be accounted for by positive selection, conventional pure culturing of outbreak strains was repeated ten times before incubation at 22°C or exposure to ampicillin; the conversion was still subsequently detected. Also, a group of colonies that had been identically positioned on replicate plates (by a velvet inoculator) became Vi-reactive and were O111-negative within ten days at 22°C; at 35°C, colonies remained unchanged. The Vi-reactive colonies maintained their reactivity even when passaged several times at 22°C or 35°C. So far, we have not been able to achieve reversion of any Vi-positive strain to the original prototype.

Discussion

This study shows that *E coli* O111:B4 can cause an extensive outbreak of diarrhoea, not only among newborn babies and infants, but also among school children and healthy adults. We have also found that these outbreak strains can express the Vi antigen, a property which has previously been described only in *Salmonella* and *Citrobacter* species.[13] Furthermore, Vi antigen-like reactivity enhanced the adherence of bacteria to Hep-2 cells. An unexpected finding was that the youngest patients of this outbreak were 2 years old, although there were children of the usual risk age in the households surveyed. Furthermore, that morbidity in adults was moderately high indicates that *E coli* O111:B4 and perhaps other EPEC may indeed be more common causes of adult diarrhoea than has been generally realised. Since EPEC are usually associated with infant diarrhoea, these organisms are not sought in adult diarrhoea.

In healthy adult volunteers, the infectious doses needed have been high—ie, 10^8–10^{10} organisms.[5] It is unlikely that bacterial contamination of this magnitude could occur without an unpleasant taste in the food. However, no subjective signs of spoilage were observed by consumers of the food before the present outbreak. Thus, the source of the *E coli* O111:B4 remains unknown.

Nearly 40 years ago, Ferguson and June[5] suggested that the virulence of *E coli* O111:B4 from different sources may vary substantially. Our observations seem to support the possibility of such virulence variability as an explanation for the high morbidity and exceptional age of the patients in the present outbreak. High virulence was also indicated by the spread of the outbreak to the households. 137 diseased household members from 85 households, without personal contact with the school complex, were not likely to have ingested large amounts of the organisms: it is noteworthy that earlier studies emphasised the importance of a contact mode of spread.[3,14]

The findings obtained by repeated pure culturing of outbreak strains and the replicate plate method suggest that reactivity with the Vi antisera was based on conversion and was not merely selection and growth of bacteria with Vi antigen-like reactivity as a minority bacterial population contaminating the major flora. However, the strongest evidence for conversion was the appearance of Vi antigen-like reactivity in the presence of ampicillin with no reduction in the number of colonies. Since the Vi antigen-positive convertants were clearly more adhesive to Hep-2 cells than were the parent strains, the expression of Vi antigen may have given a definitive advantage to the outbreak strain in colonising the intestine of the patients. However, we cannot confirm this hypothesis since no Vi antigen-positive bacteria were observed in the primary cultures.

S typhi and *S paratyphi* C, as well as a few atypical (but genetically related) *Citrobacter* and *Salmonella* strains, can synthesise the Vi antigen.[13] In mice, capsular Vi antigen is associated with *S typhi* virulence.[15] That highly purified Vi polysaccharide (either alone or conjugated to a protein) induces a good antibody response and enhances protection against typhoid in man[16,17] is indirect evidence that the Vi antigen is also important in the pathogenesis of human enteric infections. However, the mechanisms by which the Vi antigen increases intestinal virulence are unknown.

In our study, *E coli* O111:B4 was recovered in pure growth in almost all stool cultures; this has also been reported in volunteer experiments. The adhesiveness of *E coli* O111:B4 and other EPEC gives these organisms a distinct colonisation advantage and would account for the predominant growth. Thus, Vi antigen expression may have a role in virulence by enhancing adherence to the intestinal epithelium.

Two non-linked chromosomal gene loci, *Via*A and *Via*B, determine Vi antigen expression in strains of *Salmonella* and *Citrobacter*,[18-20] although functional *Via*A genes in *E coli* do

not normally express the Vi antigen.[21] Johnson and Baron[21] transferred the *Via*B locus from *S typhi* into *E coli*, where the Vi antigen was expressed. The same mechanism of *Via*B transfer may have occurred naturally in our strains with Vi antigen-like reactivity, but to confirm this notion the presence of both gene loci must be demonstrated in these strains. Final proof for conversion would be the demonstration of these gene loci in the outbreak *E coli* O111:B4 strain.

We thank Dr S. Knutton, Institute of Child Health, University of Birmingham, for the FAS tests, and Dr Anja Siitonen, National Public Health Laboratory, Helsinki, for assessing toxin production of strains. Reagents for toxin tests were kindly provided by Dr Ann-Mari Svennerholm, Department of Medical Microbiology, University of Göteborg, Göteborg, Sweden.

This work was supported by a grant from the Sigrid Jusélius Foundation.

REFERENCES

1. Bray J. Isolation of antigenically homogeneous strains of Bact. coli neapolitanum from summer diarrhoea in infants. *J Pathol* 1945; **57:** 239–47.
2. Kauffmann F, Dupont A. Escherichia strains from infantile epidemic gastro-enteritis. *Acta Path Microbiol Scand* 1950; **27:** 552–64.
3. Kessner DM, Shaughnessy HJ, Googins J, et al. An extensive community outbreak of diarrhea due to enteropathogenic *Escherichia coli* O111:B4. I. Epidemiologic studies. *Am J Hyg* 1962; **76:** 27–43.
4. Kirby AC, Hall EG, Coackley W. Neonatal diarrhoea and vomiting. Outbreaks in the same maternity unit. *Lancet* 1950; ii: 201–07.
5. Ferguson WW, June RC. Experiments on feeding adult volunteers with *Escherichia coli* 111, B4, a coliform organism associated with infant diarrhoea. *Am J Hyg* 1952; **55:** 155–69.
6. Stevensson JS. Bact. coli D 433 in cases of diarrhoea in adults. *Br Med J* 1950; **2:** 195–96.
7. Kaufmann F. The bacteriology of Enterobacteriaceae. Copenhagen: Munsgaard, 1966.
8. Sarkkinen HK, Tuokko H, Halonen P. Comparison of enzyme-immunoassay and radioimmunoassay for detection of human rotaviruses and adenoviruses from stool specimens. *J Virol Methods* 1980; **1:** 331–41.
9. Knutton S, Baldwin T, Williams PH, McNeish AS. Actin accumulation at sites of bacterial adhesion of tissue culture cells: basis of a new diagnostic test for enteropathogenic and enterohemorrhagic *Escherichia coli*. *Infect Immun* 1989; **57:** 1290–98.
10. Svennerholm A-M, Wiklund G. Rapid GM1-enzyme-linked immunosorbent assay with visual reading for identification of *Escherichia coli* heat-labile enterotoxin. *J Clin Microbiol* 1983; **17:** 596–600.
11. Svennerholm A-M, Wikstsröm M, Lindblad M, Holmgren J. Monoclonal antibodies against *Escherichia coli* heat-stable toxin (STa) and their use in a diagnostic ST ganglioside GM1-enzyme-linked immunosorbent assay. *J Clin Microbiol* 1986; **24:** 585–90.
12. Granfors K, Viljanen MK, Toivanen A. Measurement of immunoglobulin M (IgM, IgG and IgA antibodies against *Yersinia enterocolitica* by enzyme-linked immunosorbent assay: Comparison of lipopolysaccharide and whole bacterium as antigen. *J Clin Microbiol* 1981; **14:** 6–14.
13. Ewing WH. Edwards and Ewing's identification of Enterobacteriaceae. 4th ed. New York: Elsevier, 1986: 181–245.
14. Mian KA. Isolation of enteropathogenic *Escherichia coli* from household pets. Relation to infantile diarrhea. *J Am Med Assoc* 1959; **171:** 1957–61.
15. Felix A, Pitt RM. A new antigen of *B typhosus*. Its relation to virulence and to active and passive immunization. *Lancet* 1934; ii: 186–91.
16. Acharya IL, Lowe CU, Thapa R, et al. Prevention of typhoid fever in Nepal with the Vi capsular polysaccharide of *Salmonella typhi*. *N Engl J Med* 1987; **317:** 1101–04.
17. Szu SC, Stone AL, Robbins JD, Schneerson R, Robbins JB. Vi capsular polysaccharide-protein conjugates for prevention of typhoid fever. Preparation, characterization, and immunogenicity in laboratory animals. *J Exp Med* 1987; **166:** 1510–24.
18. Johnson EM, Krauskopf B, Baron LS. Genetic mapping of Vi and somatic antigenic determinants in Salmonella. *J Bacteriol* 1985; **90:** 302–08.
19. Johnson EM, Kraukskopf B, Baron LS. Genetic analysis of the ViA-his chromosomal region in Salmonella. *J Bacteriol* 1966; **92:** 1457–63.
20. Snellings NJ, Johnson EM, Kopecko DJ, Collins HH, Baron LS. Genetic regulation of variable Vi antigen expression in a strain of *Citrobacter freundii*. *J Bacteriol* 1981; **145:** 1010–17.
21. Johnson EM, Baron LS. Genetic transfer of the Vi antigen from *Salmonella typhosa* to *Escherichia coli*. *J Bacteriol* 1969; **99:** 358–59.

Rapid detection and prenatal diagnosis of β-thalassaemia: studies in Indian and Cypriot populations in the UK

J. M. OLD N. Y. VARAWALLA D. J. WEATHERALL

The application of the amplification refractory mutation system (ARMS) to the detection of individual β-thalassaemia mutations in heterozygous parents and "at risk" fetuses has been assessed in Indian and Cypriot immigrant populations in the UK. 100 first trimester prenatal diagnoses have been done, entailing the detection of 17 different mutations. The method, which allows the determination of the mutations in both parental and fetal DNA on the same day, should have wide application to the carrier detection and prenatal diagnosis of monogenic diseases with heterogeneous molecular defects.

Lancet 1990; **336**: 834–37.

Introduction

The molecular pathology of single gene disorders is very heterogeneous. For example, over 90 different mutations are known to produce the β-thalassaemia phenotype, and studies suggest that cystic fibrosis, monogenic hypercholesterolaemia, Duchenne muscular dystrophy, haemophilia A and B, and other common genetic diseases have an equally varied molecular basis.

These observations have major implications for screening and prenatal diagnosis programmes aimed at secondary prevention of common genetic diseases. Prenatal diagnosis of β-thalassaemia in the UK offers a good example of the difficulties that will be encountered because of the racial heterogeneity of the immigrant population. In each affected ethnic group there are a few common mutations and a variable number of rare ones. The existence of so many different mutations means that fetuses at risk for β-thalassaemia may be compound heterozygotes, that is they will have inherited two different β-globin gene mutations. Therefore, to provide a comprehensive prenatal

ADDRESS: **National Haemoglobinopathy Reference Laboratory, Institute of Molecular Medicine, John Radcliffe Hospital, Oxford OX3 9DU, UK** (J. M. Old, PhD, N. Y. Varawalla, MD, Prof D. J. Weatherall, FRCP, FRS). Correspondence to Sir David Weatherall.

diagnosis programme for β-thalassaemia or similarly heterogeneous recessive conditions individual mutations in the DNA of parents and fetuses must be identified rapidly. Until recently this entailed the expensive and time-consuming process of preparing individual oligonucleotide probes for each mutation, labelling them with ^{32}P, and identifying each mutation by "dot blot" analysis of amplified DNA in decreasing order of frequency for each mutation for a particular racial group.[1] A modification of the polymerase chain reaction (PCR) called the amplification refractory mutation system (ARMS) has been developed, based on specific priming of the PCR.[2,3] To diagnose a specific mutation the technique requires two oligonucleotide primers identical in sequence except for the terminal 3′ nucleotides, one of which is complementary to the normal DNA sequence and the other to the changed nucleotide in the mutant DNA. For a primer to act as a template for the DNA polymerase enzyme the terminal 3′ nucleotide has to be perfectly matched to the target DNA sequence. Under carefully controlled conditions a primer with its terminal 3′ nucleotide mismatched (for example the normal primer hybridised to the mutant genomic DNA) will not function properly and no amplification occurs. We have evaluated this method for the prenatal diagnosis of β-thalassaemia in populations in the UK originating from Cyprus and the Indian subcontinent.

Materials and methods

PCR conditions

Genomic DNA was isolated from peripheral blood cells and chorionic villus samples.[4] PCR was done in a mixture of 10 mmol/l "tris" (pH 8·3), 50 mmol/l KCl, 1·5 mmol/l $MgCl_2$, and 0·01% gelatin. 0·5 to 1 μg of genomic DNA was added last to the PCR mixture in a total volume of 25 μl containing 0·2 μmol/l of each of four primers (see below for details) and 0·5 units of 'Amplitaq' enzyme (Cetus). The mixture was overlaid with 25 μl of light paraffin oil and subjected to 25 cycles of 93°C for 1 min, 65°C for 1 min, and 72°C for 1·5 min, with a final extension period of 3 min at 72°C, in a Techne PHC1 programmable heating block. 20 μl was then removed, mixed with 5 μl of 15% 'Ficoll' and 0·05% bromopheonol-blue, and loaded on a minigel of 1·5% agarose and 1·5% 'Nusieve' agarose (FMC BioProducts). After electrophoresis at 100V for 30 min the gel was stained with ethidium bromide and photographed under ultraviolet illumination.

Design of ARMS primers

Oligonucleotide primers were synthesised by the Oswell DNA Service, at the department of chemistry, Edinburgh University, and purified by high pressure liquid chromatography. ARMS primers were designed to detect normal and mutant DNA. The 3′ nucleotide of the mutant primer was complementary to the mutant sequence, and that of the normal primer was complementary to the normal DNA sequence. A single mismatched base (the most suitable mismatch and its position found by trial and error) was introduced three or four nucleotides from the 3′ end of both primers to enhance their specificity. A control pair of primers were included in each assay. The control primers amplify a region of DNA a short distance from that amplified by the ARMS primers, and must produce a fragment of a size distinguishable from any produced by the ARMS primers. Initially, 24-nucleotide primers were used for both the control and ARMS amplification; however, increasing the length to 30 nucleotides permitted a 65°C annealing temperature, thus preventing non-specific bands and improving the specificity of the ARMS primers.

Primers for Indian β-thalassaemia mutations

Primers* were developed to detect 10 mutations in the Indian population. The control primers A and B amplify an 861 bp fragment from the 3′ end of the β-globin gene. Each primer lies on either side of a 619 bp deletion, which is one of the five most common Indian mutations.[5] Therefore, they directly detect this thalassaemia mutation by amplifying a characteristic fragment of 242 bp. The other four mutations, which with the 619 bp deletion comprise 90% of Indian β-thalassaemia, are intervening sequence nucleotide number 5 (IVSI nt5) (G to C), frameshift codon (Fr) 8–9 (+G), IVSI nt1 (G to T), and Fr 41–42 (−CTTT). DNA samples were screened for these mutations using four separate reactions, each containing the two control primers, the common primer C (the primer coupled to either the normal or mutant ARMS primer), and one of the mutant ARMS primers for each of the four mutations. After amplification and electrophoresis, the presence of an amplified band in addition to the control band of 861 bp in one of the reactions signified a particular β-thalassaemia mutation. If a band of 242 bp was observed in all four of the tracks, the mutation present was the 619 bp deletion gene. If no band apart from the 861 bp control band was observed in all four reactions, it could be concluded that none of the five common mutations was present, and the DNA sample was screened in a similar fashion for the five less frequent Indian mutations.

To diagnose individuals homozygous for a β-thalassaemia mutation ARMS primers for normal DNA are required so that homozygous and heterozygous DNA can be distinguished. Initially, the normal primer for IVSI nt1 repeatedly gave a band with DNA samples from patients homozygous for the IVSI nt1 mutation as determined by allele-specific oligonucleotide dot blot analysis. The primer was re-made in the reverse direction with its 5′ end upstream of the mutation site. On retesting in combination with common primer D it did not produce an amplified fragment with DNA samples homozygous for the IVSI nt1 mutation (data not shown).

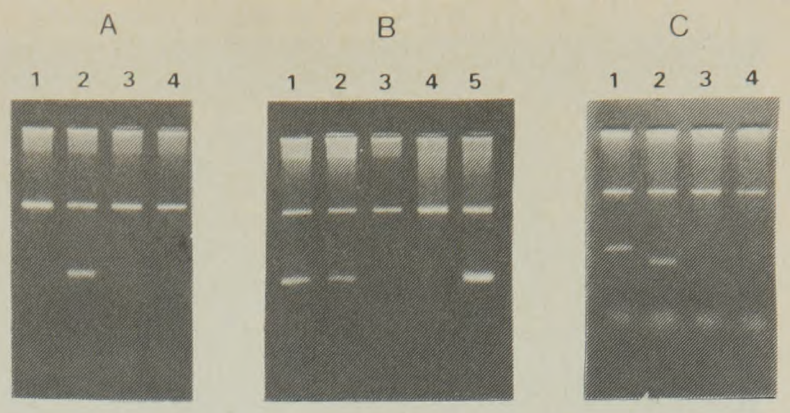


Fig 1—ARMS analysis of Indian mutations.

Gel A shows analysis of an unknown DNA sample with the primer for the IVSI nt5 mutation (lane 1), IVSI nt1 (2) Fr 41–42 (3), and Fr 8–9 (4). Gel B shows a prenatal diagnosis for the IVSI nt5 mutation: lanes 1 and 2 = maternal and paternal DNA, respectively, with IVSI nt5 mutant primer; lane 3 = normal DNA with IVSI nt5 normal primer; lane 4 = chorionic villus DNA with IVSI nt5 normal primer; lane 5 = chorionic villus DNA with the IVSI nt5 mutant primer. Gel C shows chorionic villus DNA with IVSI nt5 mutant primer (lane 1) and Fr 8–9 mutant primer (2), and normal DNA with IVSI nt5 mutant primer (3) and Fr 8–9 mutant primer (4). The upper bands in each gel are the 861 bp control bands.

Primers for Cypriot β-thalassaemia mutations

ARMS primer sequences* were developed by trial and error to detect each of seven mutations found in the UK Cypriot population. All the mutant ARMS primers were used in combination with common primer C, except for the primer for the IVSII nt745 mutation which is too far away. The control primer B was used in this case as a common primer. Primers A and B could not be used as controls with the mutant IVSII nt745 primer, so two primers which amplify a 323 bp fragment across the *Hind*III polymorphic site at the $^{G}\gamma$-globin gene were used instead (primers E and F). The IVSI

*Details of oligonucleotide sequences of primers for Indian and Cypriot mutations and RFLP analysis are available from *The Lancet*.

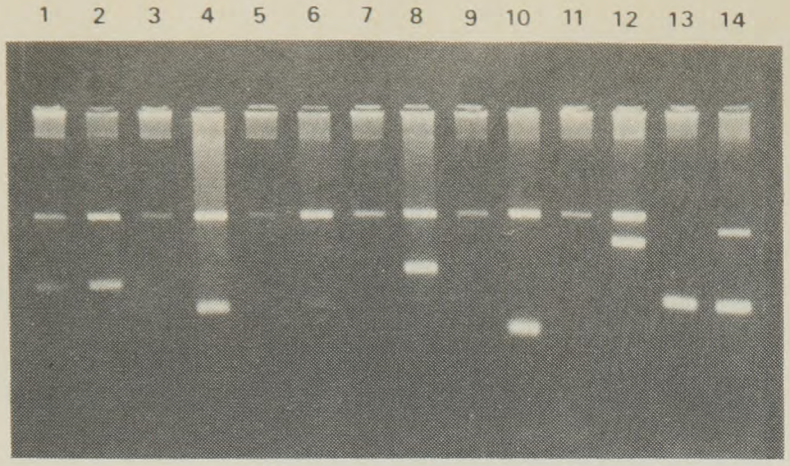


Fig 2—ARMS analysis of Cypriot mutations.

Odd-numbered lanes contain an unknown DNA sample; even-numbered lanes contain a control DNA with the appropriate mutation. Mutant-specific ARMS primers used were: IVSI nt110 (lanes 1 and 2), IVSI nt1 (3,4), IVSI nt6 (5,6), codon 39 (7,8), codon 6 (9,10), IVSI nt1 (11,12), and IVSII nt745 (13,14). The upper bands in lanes 1–12 are the 861 bp control bands. The lower bands in lanes 13 and 14 are the 323 bp control bands from primers E and F, while the upper band in lane 14 is produced by the ARMS primer.

nt110 mutation is the most common found, and all parental DNA samples were screened for this mutation first. If it was not found, each DNA was screened for the other six mutations.

Primers for haplotype analysis

Primers were used to determine β-globin gene haplotypes. Primers for the *Hind*III $^{G}\gamma$ gene restriction fragment length polymorphism (RFLP) and the *Hinf*I β gene RFLP are as described previously.[6,7] The same PCR conditions as described above were used for 30 cycles, and the products visualised on a 3% agarose gel following digestion with 10 units of the appropriate enzyme. Direct sequencing[8] of the PCR products was done after electrophoresis in low-melting temperature agarose.[9] Every prenatal diagnosis result by ARMS was checked by RFLP linkage analysis or direct DNA sequencing.

Results

Figure 1 shows the detection of various Indian mutations. A typical screening for the five most common mutations in an uncharacterised DNA sample from a β-thalassaemia heterozygote is shown in fig 1A. An amplified fragment was produced by the IVSI nt1 mutation primer in lane 2. No bands were observed with the other three ARMS mutation primers, indicating that this individual carries the IVSI nt1 mutation.

A prenatal diagnosis for IVSI nt5 mutation is shown in fig 1B. Lanes 1 and 2 show the presence in both parents of an amplified band with the mutant IVSI nt5 ARMS primer. Lane 3 shows DNA from their normal child which did not produce an amplified fragment of the mutant primer. Fetal chorionic villus DNA was tested with the normal IVSI nt5 primer (lane 4) and the mutant primer (lane 5). An amplified fragment was seen only with the mutant primer, indicating there was no normal DNA at the IVSI nt5 mutation position, and a diagnosis of β-thalassaemia major was made. This was confirmed by RFLP linkage analysis.

For couples whose child is at risk from more than one β-thalassaemia mutation the normal ARMS probes are not required. An example of a prenatal diagnosis for a compound heterozygote of IVSI nt5 and Fr 8–9 mutations is shown in fig 1C. The first two lanes show an amplified fragment for each mutant primer, indicating an affected fetus. The controls in this case were normal DNA samples tested with the same mutant primers. Again

PRENATAL DIAGNOSIS USING ARMS PRIMERS

At risk genotypes	Number of cases
Indian mutations	
IVSI nt5/IVSI nt5	11
IVSI nt1/IVSI nt1	2
Fr 41–42/Fr 41–42	1
Fr 8–9/Fr 8–9	3
IVSI nt5/Fr 41–42	5
IVSI nt5/Fr 8–9	1
IVSI nt5/-88	1
IVSI nt1/Fr 41–42	2
Fr 41–42/Fr 8–9	1
IVSI nt5/619 bp Del	2
IVSI nt1/619 bp Del	2
Fr 41–42/619 bp Del	1
Total	32
Cypriot mutations	
IVSI nt110/IVSI nt110	55
IVSI nt110/IVSI nt1	8
IVSI nt110/IVSI nt6	3
IVSI nt110/c39	4
IVSI nt110/IVSII nt745	3
Total	73

the diagnosis was confirmed by RFLP linkage analysis.

Fig 2 shows the results from an uncharacterised DNA sample from a Cypriot β-thalassaemia trait tested with each of the seven ARMS primers (odd numbered lanes), and the results produced by DNA samples from individuals carrying each of the seven mutations (even numbered lanes). The uncharacterised DNA sample produced a band with the IVSI nt110 mutation primer. In all the at-risk couples screened, either each partner had the IVSI nt110 mutation or, less frequently, one partner had the IVSI nt110 mutation and the other one of the other six mutations. Although normal ARMS primers have been synthesised for each of the seven mutations, only the IVSI nt110 and IVSI nt1 normal primers have been tested because of lack of DNA from individuals homozygous for the other five mutations. However, for prenatal diagnosis in 73 cases to date (table), only the normal IVSI nt110 primer has been required.

In all but one family the result obtained with ARMS primers have confirmed those predicted by RFLP linkage analysis. The exception was a case in which the ARMS primers predicted a normal fetus but RFLP analysis indicated a heterozygote for the IVSI nt110 mutation. Further testing for the IVSI nt110 mutation with ARMS primers showed that two samples had been mislabelled and the wrong haplotypes assigned. This was confirmed by analysing fresh DNA samples and obtaining haplotypes which predicted a normal fetus.

Discussion

The results show that ARMS could make possible a comprehensive prenatal diagnosis programme for β-thalassaemia. The method was extremely quick and simple: DNA samples could be screened for the five common mutations found in the UK Indian population in just 3 h, and the remaining 10% of samples could be rescreened for the rarer mutations later the same day. This compares well with other PCR-based non-radioactive techniques, such as oligonucleotide hybridisation[10] or colour complementation assays with fluorescent primers.[11]

We have now used ARMS primers for 32 first-trimester prenatal diagnoses of Indian β-thalassaemia mutations. In each case, the diagnosis made by the ARMS approach was confirmed either by the PCR-based methods of linkage analysis of RFLPs or, in one case when no informative

markers were available, by direct DNA sequencing. Therefore, our current strategy for prenatal diagnosis is to screen the parents first to identify the mutations, and to confirm the diagnosis by RFLP linkage analysis whenever possible. Occasionally, novel mutations will be encountered, and RFLP linkage or, if time permits, direct DNA sequencing provide an alternative approach. This strategy can be adopted for other ethnic groups at risk for β-thalassaemia, as indeed we have done for the UK Cypriot population. Seven ARMS primers were sufficient to detect the mutation in every case of 73 at-risk couples. In couples where both partners carried the IVSI nt110 mutation, it was possible to screen for the mutation in the parents first and then carry out a prenatal diagnosis on chorionic villus DNA later the same day. The approach has proved as reliable as haplotype analysis, and in one case more so with the identification of mislabelled samples.

We wish to thank Alison Fitches and Debra Valler for their technical assistance in haplotype analysis and direct DNA sequencing. We also thank all clinicians who referred samples for prenatal diagnosis.

REFERENCES

1. Saiki RK, Bugawan TL, Horn GT, Mullis KB, Erlich HA. Analysis of enzymatically amplified β-globin and HLA-DQ α DNA with allelic-specific oligonucleotide probes. *Nature* 1986; **324:** 163–66.
2. Newton CR, Graham A, Heptinstall LE, et al. Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS). *Nucl Acids Res* 1989; **17:** 2503–16.
3. Newton CR, Heptinstall LE, Summers C, et al. Amplification refractory mutation system for prenatal diagnosis and carrier assessment in cystic fibrosis. *Lancet* 1989; ii: 1481–82.
4. Old JM, Higgs DR. Gene analysis. In: Weatherall DJ, ed. Methods in haematology, vol 6. The thalassaemias. Edinburgh: Churchill Livingstone, 1982: 74–102.
5. Thein SL, Hesketh C, Wallace RB, Weatherall DJ. The molecular basis of thalassaemia major and thalassaemia intermedia in Asian Indians: application to prenatal diagnosis. *Br J Haematol* 1988; **70:** 225–31.
6. Kulozik AE, Lyons J, Kohne E, Geliger H, Bartram CR, Kleihauer E. Rapid and non-radioactive prenatal diagnosis of β-thalassaemia and sickle cell disease: application of the polymerase chain reaction (PCR). *Br J Haematol* 1988; **70:** 455–58.
7. Semenza GL, Dowling CE, Kazazian HH Jr. Hinf I polymorphism 3′ to the human β-globin gene detected by the polymerase chain reaction (PCR). *Nucl Acids Res* 1989; **17:** 2376.
8. Thein SL, Hesketh C, Brown JM, Anstey AV, Weatherall DJ. Molecular characterisation of a high A_2 β-thalassaemia by direct sequencing of single strand enriched amplification genomic DNA. *Blood* 1989; **73:** 924–30.
9. Kretz KA, Carson GS, O'Brien JS. Direct sequencing from low-melt agarose with Sequenase. *Nucl Acids Res* 1989; **17:** 5864.
10. Saiki RK, Chang CA, Levenson CH, et al. Diagnosis of sickle cell anaemia and β-thalassaemia with enzymatically amplified DNA and non-reactive allele-specific oligonucleotide probes. *N Engl J Med* 1988; **319:** 537–41.
11. Chehab FF, Kan YW. Detection of specific DNA sequences by fluorescent amplification: A colour cocomplementation assay. *Proc Natl Acad Sci USA* 1989; **86:** 9178–82.

Arterial responses during migraine headache

H. K. IVERSEN T. H. NIELSEN J. OLESEN P. TFELT-HANSEN

The superficial temporal artery has been thought to be the main focus of pain during migraine attacks, but its diameter has never been measured directly. The use of a new, high-resolution ultrasound machine to measure arterial size in 25 migraine patients with unilateral head pain showed that the lumen was wider on the painful than on the non-painful side during a migraine attack. The diameters of both radial arteries and the temporal artery on the non-painful side were smaller during than between attacks. The generalised vasoconstriction was not shared by the temporal artery on the affected side, which suggests a local vasodilatory response. The findings suggest that cephalic arteries may play a role in migraine pathogenesis.

Lancet 1990; **336**: 837–39.

Introduction

Studies on the mechanism of migraine have focused on the dilatation and increased pulsations of the external carotid artery and its branches.[1–3] So far the diameter of the temporal artery has been measured by indirect methods. A new, high-resolution ultrasound machine has now made it possible to measure directly, yet non-invasively, the luminal diameter of superficial small arteries such as the temporal artery.[4] The purpose of our investigation was to examine whether the temporal artery dilates on the symptomatic side during a migraine headache.

Patients and methods

25 women with migraine who had unilateral pain—3 with aura (classic migraine) and 22 without aura (common migraine)—entered the study. Their median age was 49 years (range 24–84). The diagnosis was made according to the criteria of the International Headache Society.[5] Patients were asked to go to the hospital during a spontaneous, unilateral migraine attack. Patients with daily headache were not included. The median monthly frequency of migrainous attacks was 3 (range 1–8) and of tension-type headaches 2 (range 0–20). The study was approved by the local ethical committee and patients gave their informed consent. The luminal diameter of the temporal artery was measured with a high-resolution ultrasound unit ('Dermascan C', Cortex technology, 20 MHz centre frequency and 15 MHz bandwidth). When the sound beam is directed perpendicular to the skin superficially located arteries such as the temporal artery can be located, since the high amplitudes reflect the interfaces between blood and vessel wall that are seen pulsating (fig 1). The mean of four measurements randomly distributed within the shortest possible interval (within 1 minute) was used to augment precision. The measurements were taken at sound velocity 1605 m/s, reported to be the sound velocity in human blood.[4]

Patients arrived at the clinic as soon as possible after onset of a migraine attack. No ergotamine or morphine-like agent was allowed for 48 h, and no analgesics for 6 h, before the study. Patients were carefully instructed not to reveal laterality of migraine headache to the investigator. With the patient supine the right and left frontal

ADDRESSES: **Department of Neurology, Gentofte Hospital** (H. K. Iversen, T. H. Nielsen, Prof J. Olesen) **and Bispebjerg Hospital** (P. Tfelt-Hansen), **University of Copenhagen, Denmark.** Correspondence to Dr H. K. Iversen, Department of Neurology, Gentofte Hospital, 2900 Hellerup, Denmark.

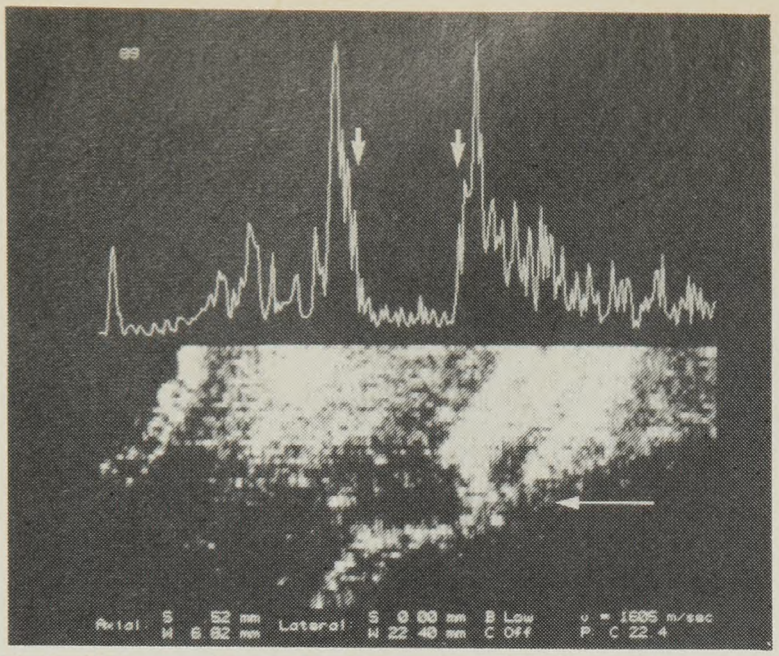


Fig 1—Ultrasound B-mode presentation of the frontal branch of the temporal artery (lower part) and luminal diameter as the peak to peak distance on A-mode (upper part).

The peaks represent the tissue interfaces.

branches of the superfical temporal arteries were gently palpated without being compressed. Any palpable difference in size between the sides was noted. The sites at which measurements were made were defined in terms of the angle and distance relative to the orbitomeatal line to ensure that repeat measurements were made at exactly the same position. After the patient had rested for 30 min, the luminal diameters of both temporal arteries were measured. The two radial arteries were measured at a point 2 cm proximal to the distal volar crest of the wrist. After the measurements had been made, the patients were asked to palpate the frontal branch of their temporal arteries and report any difference in size. When patients had been free from migraine for at least a week, the measurements were repeated at the same positions. The room temperature was kept at 21–23°C.

Statistics

Wilcoxon's rank sum test for paired data was used to evaluate differences in diameters of temporal and radial arteries. The sign test was used for palpation and pain data.

Results

Clinical characteristics of migraine patients and of the attacks studied are given in table I. The investigations were done at a median of 5 h (range 2–58) after onset of attacks. Headache was right sided in 14 and left sided in 11 patients. During a migraine headache the median luminal diameter of the frontal branch of the temporal artery of the affected side was greater than that on the other side, but during intervals between attacks the diameters were similar (table II). The

TABLE I—CHARACTERISTICS OF MIGRAINE ATTACKS

Characteristic	Findings
History	
Duration of migraine (yr)	Median 24, range 1–55
Frequency of attacks (1 mo)	Median 3, range 1–8
Duration of attacks (h)	Median 24, range 1–55
Features of attacks	
Laterality	14 right, 11 left
Quality	21 pulsating, 4 pressing
Intensity	15 moderate, 10 severe
Aggravation by physical activity	22 yes, 3 no
Accompaning symptoms	
Nausea	22 yes, 3 no
Vomiting	5 yes, 0 no
Photophobia	25 yes, 0 no
Phonophobia	23 yes, 2 no

TABLE II—LUMINAL DIAMETER OF THE FRONTAL BRANCH OF THE TEMPORAL AND RADIAL ARTERIES DURING AND BETWEEN MIGRAINE ATTACKS

Arteries	Luminal diameter (mm)		Luminal diameter during attack as percent of diameter between attacks
	During attack	Between attacks	
Temporal (n = 25)			
Symptomatic	1·10 (0·78–1·61)*	1·14 (0·70–1·48)	100·0 (67·2–149·5)
Non-symptomatic	1·01 (0·62–1·61)	1·13 (0·83–1·55)	90·6 (57·4–113·3)
Radial (n = 20)			
Symptomatic	2·42 (1·87–2·88)	2·45 (2·16–3·11)	95·7 (76·0–113·0)
Non-symptomatic	2·34 (1·88–2·64)	2·48 (2·07–2·97)	96·4 (70·2–107·8)

*$p < 0.01$ for difference from non-symptomatic temporal artery.

diameter of the temporal artery on the symptomatic side during an attack did not differ from that between attacks ($p = 0.75$), but on the other side the diameter decreased by 9·4% during attacks (median value) ($p < 0.003$) (fig 2).

The diameter of the radial arteries was measured in 20 of the 25 patients. There was no difference between sides during or between attacks ($p = 0.4$) (table II), but on both sides the diameters were smaller during than between attacks by 4·3% on the symptomatic side and by 4·6% on the non-symptomatic side.

Clinically the investigator judged the temporal artery to be larger on the symptomatic side in 15, and on the non-symptomatic side in 3 of 22 patients. In 4 patients no difference was found ($p < 0.01$) (table III). Palpation reflects the outer diameter of the artery, as opposed to the luminal diameter measured by ultrasound. Of 21 patients, 14 reported the symptomatic side and 1 the non-symptomatic side to be greater, and 6 patients found no difference between the two sides ($p < 0.01$). Neither the investigator nor the patients found any significant difference between sides in the attack-free intervals (table III). After the investigation, 23 patients were asked if the gentle palpation caused local pain—15 did not, 7 did, and 1 could not tell ($p > 0.05$). Furthermore, they were asked if the migraine headache changed during palpation—20 patients reported no change, 2 worsening, and 1 relief.

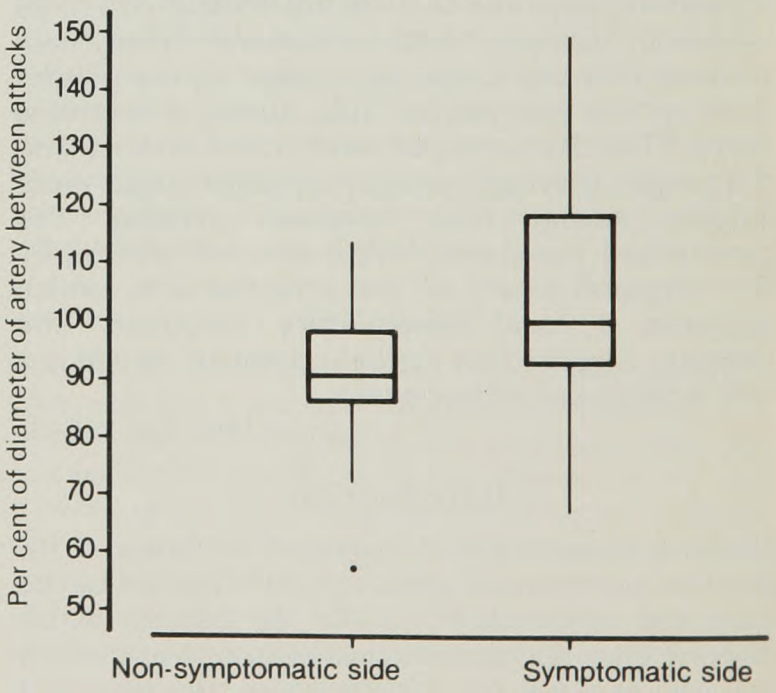


Fig 2—Luminal diameter of the frontal branch of the temporal artery during migraine attack.

The box covers the middle 50% of the values between the lower and upper quartiles; the central line is at the median; the "whiskers" extend out to the extremes, but only to those points that are within 1·5 times the interquartiles range.

PRENATAL DIAGNOSIS, VOL. 11, 775–785 (1991)

AN APPROACH TO PREIMPLANTATION DIAGNOSIS OF β-THALASSAEMIA

N. Y. VARAWALLA*, A. DOKRAS†, J. M. OLD*, I. L. SARGENT† AND D. H. BARLOW†

*Institute of Molecular Medicine, John Radcliffe Hospital, Oxford, OX3 9DU, U.K.;
†Nuffield Department of Obstetrics and Gynaecology, John Radcliffe Hospital, Oxford, OX3 9DU, U.K.

SUMMARY

Using the polymerase chain reaction (PCR), it was possible to amplify a single copy fragment of the β-globin gene from 2–32 human embryonic cells obtained from arrested preimplantation embryos. For the detection of β-thalassaemia mutations, allele specific priming of the PCR using nested primers was employed using approximately 10 pg of DNA from individuals known to carry these mutations. This approach was successful in detecting the presence or absence of five Asian Indian β-thalassaemia mutations that were selected for this study. In spite of meticulous precautions against contamination, false-positive amplification was observed, a problem that will have to be overcome before this approach can be used in clinical practice.

KEY WORDS Preimplantation diagnosis Polymerase chain reaction β-thalassaemia

INTRODUCTION

β-Thalassaemia is a heterogeneous recessive disorder, the molecular genetics of which are well defined. As the treatment of children with β-thalassaemia major is unsatisfactory and expensive, genetic counselling and prenatal diagnosis are an important option for couples at risk of having an affected offspring. Hitherto, this has involved DNA analysis of fetal material obtained by chorionic villus sampling, with abortion in the first trimester of pregnancy. While most couples achieve a healthy family, some repeatedly have affected pregnancies with abortions and remain childless. Furthermore, 10–15 per cent of these couples have reduced fertility, while others have religious and ethical objections to termination of pregnancy.

With recent advances in *in vitro* fertilization (IVF) and techniques for the micromanipulation of embryos, preimplantation genetic diagnosis is now possible (Handyside *et al.*, 1990). Recently a technique for obtaining 10–30 trophectoderm cells by biopsy of the human blastocyst has been described (Dokras *et al.*, 1990). Here, we outline an approach for the preimplantation diagnosis of β-thalassaemia with DNA amplification by the polymerase chain reaction (PCR) which could be used on such embryonic biopsies. By allele specific priming of the PCR with nested primers, we have developed a method for the direct detection of β-thalassaemia mutations which is applicable to the analysis of a few cells. This approach could also be applied for the diagnosis of other genetic disorders caused by point mutations, such as cystic fibrosis.

Addressee for correspondence: Dr Nermeen Y. Varawalla, Institute of Molecular Medicine, John Radcliffe Hospital, Oxford, OX3 9DU, U.K.

0197–3851/91/100775–11$05.50
© 1991 by John Wiley & Sons, Ltd.
Received 14 February 1991
Accepted 25 April 1991

MATERIALS AND METHODS

IVF procedures

The human preimplantation embryos used in this study were obtained from patients attending the IVF Unit at the John Radcliffe Hospital, Oxford. This project was approved by the Central Oxford Research Ethics Committee and the Interim Licensing Authority, with informed consent of all couples. None of these couples was known to be at risk of having an offspring with a genetic disorder.

The details of the IVF superovulation regime, oocyte retrieval, and embryology were as previously described (Dokras *et al.*, 1990). Briefly, down regulation was achieved with Suprefact (Hoechst U.K. Ltd) and gonadotrophins (Pergonal, Serono Lab. U.K. Ltd) were used for stimulation for 9–13 days. The ovarian response was monitored both biochemically and by vaginal ultrasound, and hCG (10 000 IU) was administered when adequate follicular development was achieved. Oocyte retrieval was performed 35 h later by vaginal ultrasound-guided aspiration. All oocytes were incubated for 5–6 h prior to insemination (day 0). They were checked 12–18 h later for fertilization by the presence of two pronuclei (day 1). The oocytes and embryos were cultured in 2 ml of Tyrode's 6 (T6) medium supplemented with 10 per cent heat-inactivated maternal serum at 37°C in 5 per cent CO_2 in air. On day 2, the embryos were assessed and the best three selected for transfer, on the basis of morphology, cleavage rate and quality of oocyte and granulosa. Embryos not selected for transfer were assigned for research purposes and assessed daily by microscopy. Fifteen embryos in which development had arrested between the 2 and 32 cell stage were used in this study.

The embryonic sample was prepared for analysis by PCR in a class I laminar flow cabinet. The arrested embryos were transferred to tissue culture dishes containing sterile twice-distilled water with sterile tissue culture-washed Pasteur pipettes and washed to remove traces of the culture medium. They were then individually transferred in 10 μl of fresh sterile twice-distilled water into an autoclaved Eppendorf tube. A new pipette was used for each transfer and on each occasion the pipette was viewed under the microscope to confirm that no cells remained in the tip. The samples were stored at −20°C.

Amplification of the β-globin gene

Amplification of a single copy 597 bp region encompassing the first and second exons and the first intron of the β-globin gene was achieved using two rounds of PCR with nested primers. Prior to amplification, the embryonic cell sample was thawed, spun down, overlaid with 100 μl of liquid paraffin, and heated at 97°C for 30 min to ensure cell lysis and release of nuclear DNA. The PCR was performed in a 100 μl reaction volume with 20 pmol of each of the oligonucleotide primers A and B (Table 1), 2 units of Taq polymerase (AmpliTaq, Cetus) and 200 μM each dNTP in 10 mM Tris–HCl (pH 8·4), 50 mM KCl and 1·5 mM $MgCl_2$. This reaction mix was stored on ice and added to the boiled cells just prior to amplification. The amplification regime consisted of 30 cycles of denaturation at 93°C for 1 min, annealing at 55°C for 1 min, and extension at 72°C for 1·5 min, followed by a final extension step at 72°C for 3 min. 2 μl of the PCR product of the first round was used as the template for the

Table 1. Sequences of oligonucleotide primers used for preimplantation diagnosis of β-thalassaemia

β-Globin primers
A—5′ ACCTCACCCTGTGGAGCCAC
B—5′ TTCGTCTGTTTCCCATTCTAAACT
C—5′ GGCCAATCTACTCCCAGGAG
D—5′ ACATCAAGGGTCCCATAGAC
E—5′ CTTGATACCCAACCTGCCCA
F—5′ CCCCTTCCTATGACATGAACTTAA (coupled with allele specific primers marked with an*)
Allele specific primers†
1. IVS-1 position 5 (G–C)
N 5′ CTCCTTAAACCTGTCTTGTAACCTTGTTAC (254 bp)
M 5′ CTCCTTAAACCTGTCTTGTAACCTTGTTAG (254 bp)
2. IVS-1 position 1 (G–T)
N* 5′ GATGAAGTTGGTGGTGAGGCCCTGGGTAGG (454 bp)
M 5′ TTAAACCTGTCTTGTAACCTTGATACGAAA (250 bp)
3. Codons 8/9 (+G)
N 5′ CCTTGCCCCACAGGGCAGTAACGGCACACT (183 bp)
M 5′ CCTTGCCCCACAGGGCAGTAACGGCACACC (183 bp)
4. Codons 41/42 (−CTTT)
N 5′ GAGTGGACAGATCCCCAAAGGACTCAAAGA (412 bp)
M 5′ GAGTGGACAGATCCCCAAAGGACTCAACCT (412 bp)
5. Codon 15 (G–A)
N* 5′ TGAGGAGAAGTCTGCCGTTACTGCCCAGTG (500 bp)
M* 5′ TGAGGAGAAGTCTGCCGTTACTGCCCAGTA (500 bp)
Antithrombin primers
AT 1—5′ AGTCAGAGACTGACCAGCATGTGC
AT 2—5′ AGGGGTTCTAACTTTTAGTCAGC
AT 3—5′ CATGTTAACTAGGCAGCCCACCA
AT 4—5′ GTCTTCAGCAGCAAAGCAGTGT

† = Normal; M = mutant. The size of the PCR product in base pairs is given in parentheses alongside the respective primers.

second round, which was performed in an identical fashion using nested primers C and D (Table 1). The location of these primers with respect to the β-globin gene is shown in Figure 1.

Negative and positive controls were set up by adding the reaction mix to the following constituents. In the first two experiments, the negative control was a water blank consisting of 10 μl of sterile twice-distilled water which was prepared in the PCR set-up area and for the positive control, 1 ng of DNA was used. In the remaining five experiments, positive controls were omitted and the negative controls included both media blanks and water blanks. The media blank consisted of 10 μl of the medium in which the embryos were cultured, pipetted into a sterile Eppendorf tube in exactly the same manner and at the same time as the embryonic cell samples. The water blanks were 10 μl aliquots of sterile twice-distilled water which were set up both in the area where the embryonic cell samples were prepared and in the area where the PCR was set up. Both the samples and the controls were treated in an

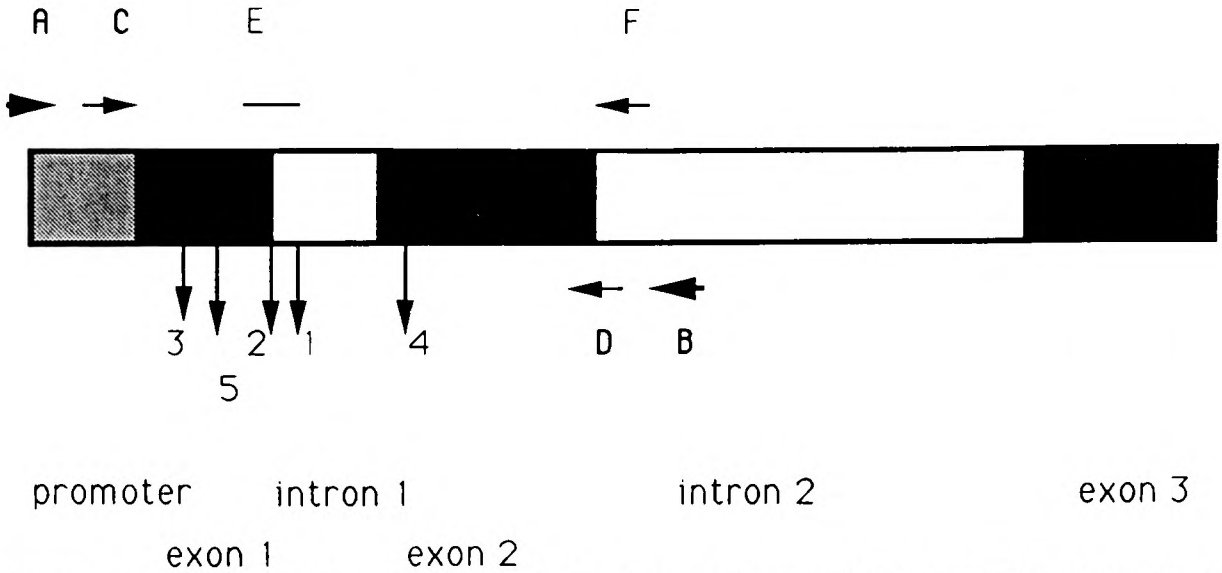


Figure 1. Map of the β-globin gene showing the location of primers A, B, C, D, and F, and the internal oligonucleotide probe E. The sites for the β-thalassaemia mutations detected by allele specific priming are also shown: 1—IVS-1 position 5 (G–C); 2—IVS-1 position 1 (G–T); 3—codons 8/9 (+G); 4—codons 41/42 (−CTTT) and 5—codon 15 (G–A)

identical fashion. The experiments were set up in a double blind fashion such that the identity of each of the tubes was not known until each experiment was complete.

A 10 μl aliquot of the second round PCR product was electrophoresed on a 1 per cent agarose gel, stained with ethidium bromide, and visualized on a UV transilluminator. The time taken for the PCR analysis was 8 h. To confirm the identity of the amplified product, one of the agarose gels was blotted using a nylon filter (Hybond-N+, Amersham, U.K.) which was hybridized with a ^{32}P-labelled β-globin gene-specific oligonucleotide probe E (Table 1).

Direct detection of β-thalassaemia mutations

The mutations were detected by a PCR method based on allele specific priming called the amplification refractory mutation system (ARMS) (Newton *et al.*, 1989). The underlying principle of this system is that oligonucleotides with a mismatched 3′ end will not function as amplimers in the PCR under appropriately stringent conditions. Thus, an oligonucleotide primer can be designed with its 3′ end complementary to the sequence of the mutation that is being looked for, such that when the PCR is performed under stringent conditions an amplified product will indicate the presence of that mutation, while its absence will indicate normal DNA sequence at the site of the mutation. This technique has been successfully used for first-trimester prenatal diagnosis (Old *et al.*, 1990).

To make this technique suitable for preimplantation diagnosis, it was necessary to use two rounds of PCR with nested primers. The ARMS technique requires an internally amplified DNA fragment in each reaction acting as a control, to prove that the absence of amplification is due to the relevant mutation not being present and not because of the failure of the PCR. For this purpose a 377 bp fragment encompassing exon 3A and flanking intron sequences of the antithrombin gene on chromosome 1 was co-amplified in a nested fashion.

As it was not possible to obtain embryonic cell samples from couples carrying β-thalassaemia mutations, diluted DNA solutions from various individuals whose

β-thalassaemia mutations had already been characterized were used as experimental models. Dilute DNA solutions were prepared from individuals known to be homozygous, heterozygous and normal for each of the five mutations selected for study. The quantity of DNA present in a 10–30 cell trophectoderm biopsy would be 60–180 pg, considering that 6 pg of DNA is contained in a single diploid cell. Approximately 1 μg of DNA from each of these individuals was serially diluted by a factor of 1:100 000 with sterile twice-distilled water, such that 1 μl of the dilute solution would contain approximately 10 pg of DNA.

The first round of PCR was performed using 10 pg of DNA with β-globin primers A and B and antithrombin primers AT 1 and AT 2, using the reaction conditions described above except that the reaction mix was exposed to 260 nm UV light for 5 min before the addition of AmpliTaq and the DNA. For the second round of PCR, the β-globin primers either C or F in combination with an allele specific primer for one of the β-thalassaemia mutations selected along with the antithrombin gene primers AT 3 and AT 4 located internally to AT 1 and AT 2 (see Table 1 for sequences) were used. In order for the allele specific primers to work specifically, an annealing temperature of 66°C was required; all the other conditions were as described for the first round of PCR. Besides the mutant allele specific primers for each of the mutations selected, the normal ones were also tested. The sequence of the ten allele specific primers are shown in Table 1 along with the sizes of their PCR products; the location of each of these mutations is shown in Figure 1. Each of these primers was tested with DNA known to be homozygous, heterozygous and normal for each of these mutations. Negative controls in which 10 μl of sterile twice-distilled water was added to the reaction mix were included in all experiments. For these negative controls the β-globin primers used were C and D along with the internal control antithrombin gene primers. A 10 μl aliquot of the final PCR product was run out on a 2 per cent NuSieve (FMC Bioproducts, U.S.A.) and 2 per cent agarose gel to facilitate separation of the bands, stained with ethidium bromide, and visualized by UV transillumination. The time taken for this analysis was 8 h.

Precautions against contamination

The embryonic cell samples for PCR were prepared in a sterile class I laminar flow cabinet, using sterilized tissue culture-washed Pasteur pipettes and autoclaved pipette tips and Eppendorf tubes. The operator wore sterile gloves and a gown.

The PCRs were set up in a physically distant class II laminar flow cabinet using autoclaved pipette tips and Eppendorf tubes. All reaction components were filtered through a 0·2 μm filter (Acrodisc) and autoclaved. Gloves were worn which were changed prior to every entry into the set-up area. To avoid the serious hazard of contamination with the final amplification product, the areas where the PCRs were performed and where amplified product was analysed were totally segregated from the areas where the biopsy was done and the reaction set up. Further, all equipment such as Gilson micropipettes, pipette tips, Eppendorf tubes, tube racks, and ice buckets were designated to the separate areas.

During the course of this study, a reduction of false-positive amplification by UV-induced sterilization which created termination sites for Taq DNA polymerase was described (Cimino *et al.*, 1990). Therefore, in the experiments carried out to detect the β-thalassaemia mutations directly, together with the precautions described

Table 2. Amplification of the β-globin gene by PCR from arrested human embryos

Expt. No.	Embryonic cell samples		Negative controls: Media blanks		Negative controls: Water blanks: Biopsy area		Negative controls: Water blanks: PCR set-up area		Positive controls	
	+	−	+	−	+	−	+	−	+	−
1	1	0	—	—	—	—	0	1	1	0
2	4	0	—	—	—	—	0	1	1	0
3	2	0	0	2	0	1	0	1	—	—
4	1	1	1	1	0	1	0	1	—	—
5	2	0	1	0	0	1	0	1	—	—
6	2	0	0	1	0	1	0	1	—	—
7	2	0	0	1	0	1	0	1	—	—
Total	14	1	2	5	0	5	0	7	2	0

+ Denotes the presence of amplification and − denotes its absence.

above, the first round PCR mix was exposed to 260 nm UV light for 5 min prior to addition of the AmpliTaq enzyme and the template DNA.

RESULTS

Seven experiments were performed to amplify a 597 bp fragment of the β-globin gene from arrested human embryos at the 2–32 cell stage. The results of these experiments are listed in Table 2. Successful amplification was obtained in 14 out of the 15 embryonic samples analysed. Figure 2 depicts an ethidium bromide-stained agarose gel showing the 597 bp β-globin amplified fragment obtained from the embryonic cell sample of experiment 1. Also shown is the autoradiograph of the Southern blot of this gel, which was hybridized to a ^{32}P-labelled β-globin gene specific oligonucleotide. A band was observed in the lane with the embryonic cell sample, identical to that seen in the lane with the positive control, confirming that the PCR product is a fragment of the β-globin gene. The absence of any band in the negative control (water blank) lane indicated the absence of contamination.

Experiment 2 was set up in a similar manner except that there were four embryonic cell samples, all of which showed successful amplification, and the positive and negative controls showed the presence and absence of amplification, respectively, as expected. As both the positive controls included in the first two experiments showed successful amplification indicating that the nested primer approach had worked satisfactorily, we omitted these controls in the next five experiments while increasing the negative controls. There was false-positive amplification of two of the seven media blanks; however, none of the five and seven water blanks prepared in the biopsy and PCR set-up area, respectively, showed any amplification. The ethidium bromide-stained agarose gel of experiment 7 is shown in

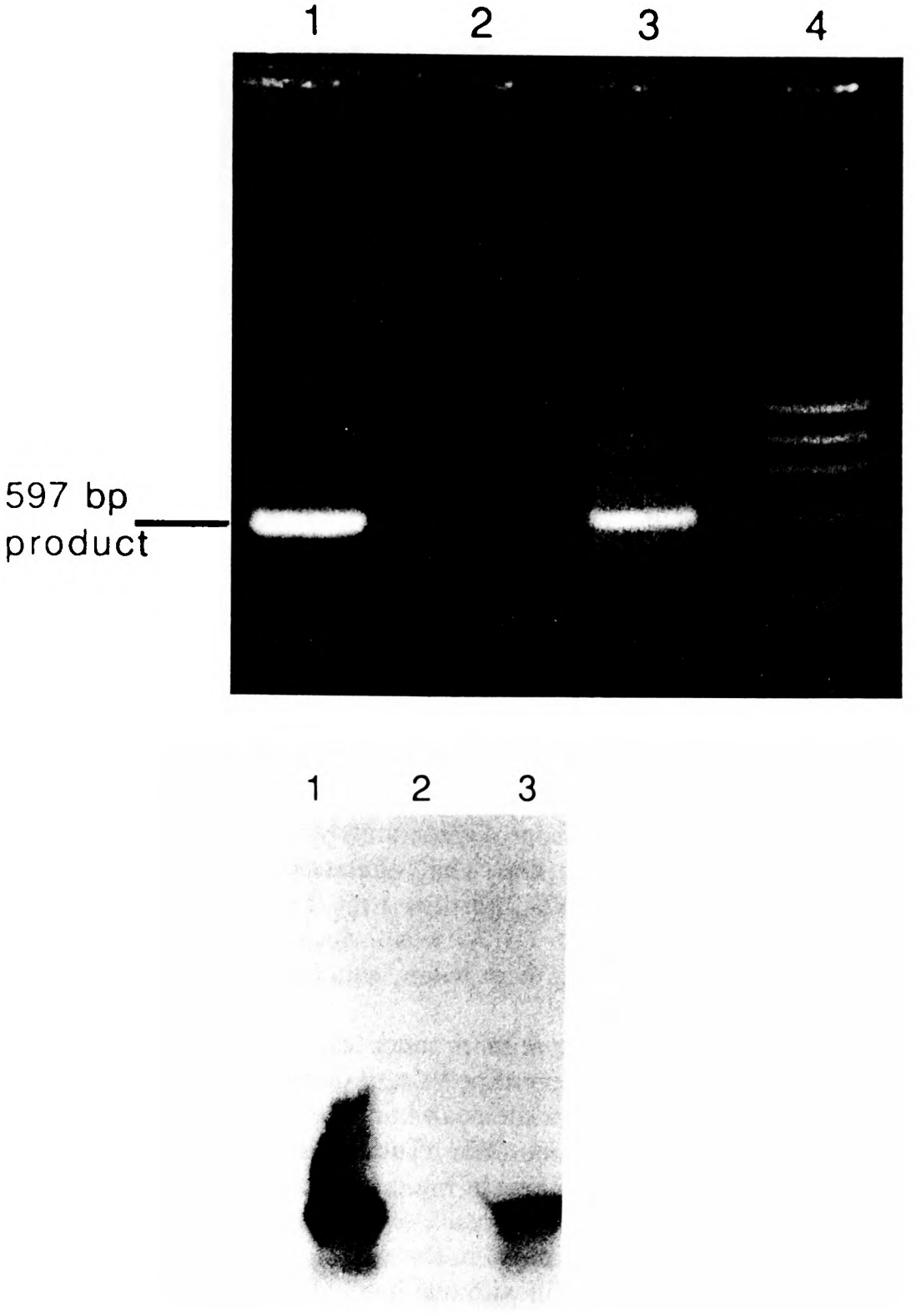


Figure 2. Southern blot of the PCR products from experiment 1 hybridized to a β-globin gene specific oligonucleotide probe, sequence E—Table 1. The upper half is the ethidium bromide-stained agarose gel and the lower half is the autoradiograph. Lane 1 contains the embryonic cell sample, lane 2 the negative control water blank and lane 3 the positive control. The marker in lane 4 is ox digested with *Hae* III

Figure 3. Successful amplification was obtained from both embryonic samples and no false-positive amplification from any of the three negative controls was observed.

Having been able to successfully amplify a region of the β-globin gene which is known to be the site of many of the β-thalassaemia mutations from 2-32 human



Figure 3. Amplification of the β-globin gene from human embryonic cells of experiment 7. Ethidium bromide-stained agarose gel showing the presence of a 597 bp PCR product obtained from embryonic cells in lanes 1 and 2. Lane 3 contains a media blank and lanes 4 and 5 sterile twice-distilled water blanks. The marker in lane 6 is øx 174 digested with *Hae* III

embryonic cells, we tested a method for the direct detection of some of the β-thalassaemia mutations using approximately 10 pg of DNA from individuals known to carry these mutations. The β-thalassaemia mutations selected for study were IVS-1 position 5 (G–C), IVS-1 position 1 (G–T), codons 8/9 (+G), codons 41/42 (−CTTT), and codon 15 (G–A) which account for 77 per cent of the β-thalassaemia mutations present in an Asian Indian population (Varawalla *et al.*, 1991).

Each of the ten allele specific primers tested was able to correctly detect the presence or absence of the corresponding β-thalassaemia mutations using approximately 10 pg of DNA from individuals known to be homozygous, heterozygous and normal for the particular mutation. The results using the primers for the IVS-1 position 5 (G–C) mutation are shown in Figure 4. The DNA sample homozygous for this mutation showed the presence of a 254 bp amplified fragment with only the mutant allele specific primer and not with the normal one. Conversely, the normal DNA sample showed amplification with only the normal allele specific primer and not with the mutant one. The heterozygous DNA sample showed amplification with both the normal and the mutant allele specific primers, although the intensity of the 254 bp band was lower than that obtained with the homozygous and normal samples. This is because only one allele was being amplified instead of two, as in normal and homozygous samples. The presence of the 377 bp internal control band in all the reactions proved that the PCR had worked satisfactorily. Similar results were obtained with both the normal and the mutant allele specific primers for each of the other four mutations (data not shown). Also shown in Figure 4 is the negative control water blank, which does not show any amplification with either the β-globin or the antithrombin gene primers. None of the seven water blanks which were

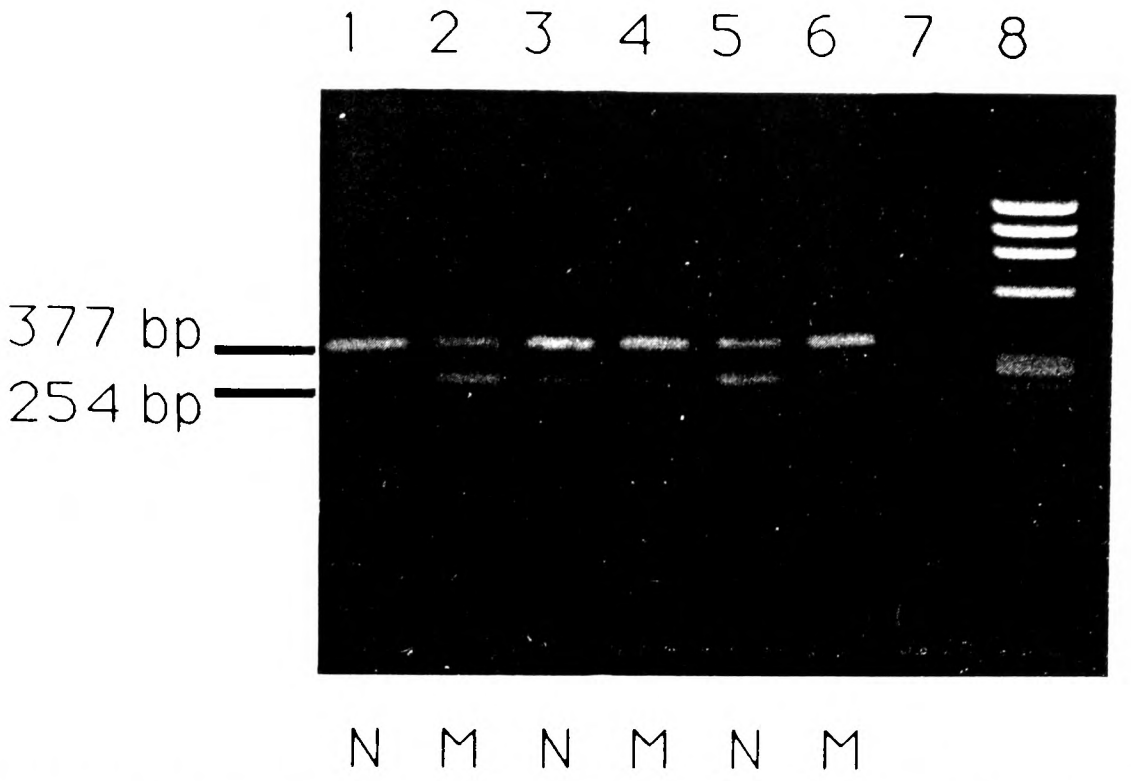


Figure 4. Ethidium bromide-stained agarose gel showing the detection of the IVS-1 position 5 (G C) β-thalassaemia mutation from 10 pg of DNA using nested allele specific primers. Lanes 1 and 2 contain DNA homozygous for the mutation; lanes 3 and 4 contain heterozygous DNA and lanes 5 and 6 contain normal DNA. Each of these samples was tested with the normal (N) and mutant (M) allele specific primer, respectively. Lane 7 contains the negative control water blank and lane 8 has øx 174 digested with *Hae* III as a marker. The upper 377 bp band is the internal control product of the antithrombin gene primers and the lower 254 bp band is the product of amplification with the allele specific primers

included in the experiments for the direct detection of the β-thalassaemia mutations showed any false-positive amplification. Exposing the reaction mix to UV light for 5 min did not seem to affect the efficiency of the primers.

DISCUSSION

Successful pregnancies following the transfer of human embryos which had been sexed by amplification of a Y-specific repeat sequence have been reported (Handyside *et al.*, 1990). Amplification of the single copy β-globin gene from mouse blastomeres (Holding and Monk, 1989) and from human oocytes and polar bodies (Monk and Holding, 1990) had also been described. Thus, it should be possible to perform preimplantation diagnosis for hereditary disorders of the β-globin gene such as β-thalassaemia by PCR analysis of the genetic defect using embryonic cells. A technique for obtaining a 10–30 cell trophectoderm biopsy from the human blastocyst has been described which would provide an embryonic cell sample for genetic analysis. Evaluating the feasibility of this approach, we have demonstrated that a fragment of the β-globin gene can be amplified from 2–32 cell arrested human embryos using two rounds of amplification with nested primers. Successful amplification was possible in 14 of the 15 cases attempted, indicating the potential of this approach for genetic diagnosis.

For preimplantation diagnosis of genetic disorders by PCR analysis, a method that directly detects the mutation is more desirable than an indirect one like linkage analysis of DNA polymorphisms. One such method is the ARMS technique, which when used with nested primers can successfully detect the presence or absence of a β-thalassaemia mutation from approximately 10 pg of DNA, making it suitable for analysis of a 10–30 cell trophectoderm biopsy. This approach would also be sensitive enough for the PCR analysis of a single blastomere obtained by a single-cell biopsy from the eight-cell embryo. The ability of this approach to detect a mutation in the heterozygous state suggests that it is sensitive enough to analyse a single allele. Thus, it could also be used for the analysis of haploid oocytes for preconceptional diagnosis. Besides being the site for 77 per cent of the β-thalassaemia mutations in Asian Indians, the region of the β-globin gene that we have studied is also the site for 94 per cent of the β-thalassaemia mutations present in the Mediterranean peoples and 84 per cent of the mutations prevalent in the Chinese (Kazazian and Boehm, 1988), making these primers particularly suitable for the preimplantation diagnosis of β-thalassaemia. However, with careful design of primers it could also be applied to another region of the β-globin gene or indeed to any other gene known to be associated with genetic disease.

In cases under study, the mutations which the parents carry could be characterized and the available embryos could then be biopsied and analysed for the presence or absence of these mutations. As only 2 μl of the PCR product of the first round is required for the second round of PCR, it would be possible to perform a number of second round reactions with different allele specific primers. It is not uncommon for the parents to have identical mutations; thus, it would be necessary to be able to detect the homozygous state, using normal allele specific primers. As the DNA analysis can be carried out in the course of a single working day, transfer of unaffected embryos would be possible in the same cycle. A number of unaffected embryos could be transferred to the uterus and the extra ones stored by cryopreservation for subsequent pregnancies. Blastocysts for trophectoderm biopsy need not be obtained by IVF, but could also be obtained by uterine lavage (Buster *et al.*, 1985) following natural conception in couples with unimpaired fertility; this would considerably reduce the expense and strains of these procedures.

However, before preimplantation genetic diagnosis by DNA analysis using the PCR can be offered to at-risk couples, the serious problem of contamination must be overcome. The exquisite sensitivity of the PCR which enables DNA amplification from a few cells is also responsible for false-positive amplification, as we found in two of our negative control media blanks, in spite of meticulous precautions against contamination. This emphasizes the importance of appropriate negative contols. A way of overcoming this problem would be to analyse the embryonic samples in duplicate, allowing a countercheck in the result; this would be possible with trophectoderm biopsies where the 10–30 cells available for study could be split into two samples. It appears that in a laboratory where a particular gene is being constantly amplifed for other purposes such as prenatal diagnosis of the genetic disease, the problem of contamination may be more persistent owing to the presence of substantial amounts of the PCR product in the vicinity, in spite of segregating work areas and equipment. Thus, a completely separate laboratory is necessary for preimplantation diagnosis by PCR.

ACKNOWLEDGEMENTS

We wish to acknowledge Professor Sir David Weatherall and the late Professor Sir Alexander Turnbull for their support. We would also like to thank Dr Robin Olds for the antithrombin oligonucleotide primers. Dr Nermeen Varawalla is in receipt of a Rhodes research fellowship.

REFERENCES

Buster, J.E., Bustillo, M., Rodi, I.A., Cohen, S.W., Hamilton, M., Simon, J.A., Thorneycroft, I.H., Marshall, J.R. (1985). Biologic and morphologic development of donated human ova recovered by nonsurgical uterine lavage, *Am. J. Obstet. Gynecol.*, **153,** 211–217.

Cimino, G.D., Metchette, K., Isaacs, S.T., Zhu, Y.S. (1990). More false-positive problems, *Nature*, **345,** 773–774.

Dokras, A., Sargent, I.L., Ross, C., Gardner, R.L., Barlow, D.H. (1990). Trophectoderm biopsy in human blastocysts, *Hum. Reprod.*, **5,** 821–825.

Handyside, A.H., Kontogianni, E.H., Hardy, K., Winston, R.M.L. (1990). Pregnancies from biopsied human preimplantation embryos sexed by Y-specific DNA amplification, *Nature*, **344,** 768–770.

Holding, C., Monk, M. (1989). Diagnosis of beta-thalassaemia by DNA amplification in single blastomeres from mouse preimplantation embryos, *Lancet*, **ii,** 532–353.

Kazazian, H.H., Boehm, C.D. (1988). Molecular basis and prenatal diagnosis of β-thalassemia, *Blood*, **72,** 1107–1116.

Monk, M., Holding, C. (1990). Amplification of a β-haemoglobin sequence in individual human oocytes and polar bodies, *Lancet*, **335,** 985–988.

Newton, C.R., Graham, A., Heptinstall, L.E., Powell, S.J., Summers, C., Kalsheker, N., Smith, J.C., Markham, A.F. (1989). Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS), *Nucleic Acids Res.*, **17,** 2503–2516.

Old, J.M., Varawalla, N.Y., Weatherall, D.J. (1990). Rapid detection and prenatal diagnosis of β-thalassaemia: studies in Indian and Cypriot populations in the U.K., *Lancet*, **336,** 834–837.

Varawalla, N.Y., Old, J.M., Sarkar, R., Venkatesan, R., Weatherall, D.J. (1991). The spectrum of β-thalassaemia mutations on the Indian subcontinent: the basis for prenatal diagnosis, *Br. J. Haematol.*, **78,** 242–247.

British Journal of Haematology, 1991, 78, 242–247

ADONIS 000710489100131U

The spectrum of β-thalassaemia mutations on the Indian subcontinent: the basis for prenatal diagnosis

N. Y. VARAWALLA, J. M. OLD, R. SARKAR,* R. VENKATESAN* AND D. J. WEATHERALL
*Institute of Molecular Medicine, John Radcliffe Hospital, Oxford, and *Department of Genetics, Postgraduate Institute of Basic Medical Sciences, Taramani, Madras, India*

Received 12 November 1990; accepted for publication 18 January 1991

Summary. The β-thalassaemia mutations in 702 unrelated carriers originating from seven different regions of the Indian subcontinent have been characterized using allele specific priming of the polymerase chain reaction (PCR). It was possible to identify the mutations in 688 (98%) of the individuals studied. Eleven different mutations were identified, of which five common ones accounted for 93·6%: namely the ones at IVS-1 position 5 (G-C), codons 8/9 (+G), IVS-1 position 1 (G-T), codons 41/42 (-CTTT) and the 619 bp deletion at the 3′ end of the gene. The mutations at IVS-2 position 1 (G-A) and codon 30 (G-C), previously undescribed in Asian Indians, were found in two and six individuals respectively. Some regional variation in the distribution of β-thalassaemia alleles was noted. These findings should prove useful for the development of a first trimester prenatal diagnosis programme based on direct detection of mutations.

β-Thalassaemia is an heterogeneous inherited disorder characterized by reduced or absent β-globin gene expression. This disease has a high frequency in the Mediterranean Basin, Africa, South-East Asia and the Indian subcontinent (Weatherall & Clegg, 1981). The treatment of individuals with β-thalassaemia major, which entails regular blood transfusions and expensive iron chelation regimes, is not yet satisfactory. Thus the disease causes significant morbidity and mortality in affected individuals, making prenatal diagnosis an important option for couples at risk of having a thalassaemia major offspring.

β-Thalassaemia is an important genetic disease on the Indian subcontinent. The average incidence of the β-thalassaemia trait in India is 3·3%, with 1–2 per 1000 couples being at risk of having an affected offspring, each year. In India, 6000 children with thalassaemia major are born annually constituting about 10% of the total number born in the world each year (Modell & Petrou, 1983). Prevention of the disease by genetic counselling and prenatal diagnosis has a particularly important role in this part of the world where there are limited resources for the medical care of affected individuals.

In Sardinia, Cyprus and Greece successful β-thalassaemia prevention programmes have been established (Angastiniotis & Hadjiminas, 1981; Cao *et al.* 1984; Hadji *et al.* 1987). Recent reports show that the incidence of live births with β-thalassaemia major in many parts of Greece and Italy has been reduced to 10–20% of that expected (Kazazian & Boehm, 1988). Advances in techniques of chorion villus sampling along with those of DNA analysis, particularly methods that allow direct detection of the relevant mutations using DNA amplified by the polymerase chain reaction (PCR) make first trimester diagnosis possible for most couples at risk. For this purpose it is essential to understand the distribution and frequency of the relevant mutations in that region. More than 90 mutations are known to be associated with β-thalassaemia. However, in each affected ethnic group there are just a few common mutations together with a variable number of rarer ones (Kazazian *et al.* 1986; Chehab *et al.* 1987; Chan *et al.* 1987; Amselem *et al.* 1988). To date, 10 mutations have been described in Asian Indians of which five have been found to be most common (Kazazian *et al.* 1984; Wong *et al.* 1986, 1987; Thein *et al.* 1988). In order to form a basis for a prenatal diagnosis programme on the Indian subcontinent, the β-globin gene mutations in 702 unrelated β-thalassaemia carriers from seven different regions have been characterized by a nonradioactive PCR-based method using allele specific primers.

Correspondence: Dr N. Y. Varawalla, Institute of Molecular Medicine, John Radcliffe Hospital, Oxford OX3 9DU.

SUBJECTS AND METHODS

Subjects

A total of 702 unrelated β-thalassaemia carriers from seven different regions of the Indian subcontinent were studied. They were all couples at risk of having β-thalassaemia major

Table I. Primers used for the detection of β-thalassaemia mutations by allele specific priming of the PCR

Mutation	Oligonucleotide sequence
Allele specific primers	
1. IVS-1 position 5 (G-C)	5′CTCCTTAAACCTGTCTTGTAACCTTGTTAG
2. IVS-1 position 1 (G-T)	5′TTAAACCTGTCTTGTAACCTTGATACGAAA
3. Codons 8/9 (+G)	5′CCTTGCCCCACACGGCAGTAACGGCACACC
4. Codons 41/42 (-CTTT)	5′GAGTGGACAGATCCCCAAAGGACTCAACCT
5. Codon 15 (G-A)*	5′TGAGGAGAAGTCTGCCGTTACTGCCCAGTA
6. Codon 16 (-C)	5′TCACCACCAACTTCATCCACGTTCACGTTC
7. -88 (C-T)*	5′TCACTTAGACCTCACCCTGTGGAGCCTCAT
8. Cap site +1 (A-C)*	5′ATAAGTCAGGGCAGAGCCATCTATTGGTTC
9. IVS-1 3′end -25 bp	5′CTCTGGGTCCAAGGGTAGACCACCAGCATA
10. IVS-2 position 654 (C-T)**	5′GAATAACAGTGATAATTTCTGGGTTATGGT
11. Codon 17 (A-T)	5′CTCACCACCAACTTCATCCACGTTCAGCTA
12. -28 (A-G)*	5′AGGGAGGGCAGGAGCCAGGGCTGGGCTTAG
13. IVS-1 position 110 (G-A)	5′ACCAGCAGCCTAAGGGTGGGAAAATAGTCT
14. IVS-1 position 6 (T-C)	5′TCTCCTTAAACCTGTCTTGTAACCTTCATG
15. IVS-1 position 1 (G-A)	5′TTAAACCTGTCTTGTAACCTTGATACGAAT
16. IVS-2 position 1 (G-A)	5′AAGAAAACATCAAGGGTCCCATAGACTGAT
17. Codon 39 (C-T)	5′CAGATCCCCAAAGGACTCAAAGAACCTGTA
18. IVS-2 3′-end (CAG-AAG)**	5′TCATGTTCATACCTCTTATCTTCCTCCCAA
19. Codon 30 (G-C)	5′TAAACCTGTCTTGTAACCTTGATACCTACG

Other primers used

A. 5′ACCTCACCCTGTGGAGCCAC
B. 5′CCCCTTCCTATGACATGAACTTAA (used with allele specific primers marked with *)
C. 5′GAGTCAAGGCTGAGAGATGCAGGA (used with allele specific primers marked with **)
D. 5′CAATGTATCATGCCTCTTTGCACC
E. 5′AGTGCTGCAAGAAGAACAACTACC
F. 5′CTCTGCATCATGGGCAGTGAGCTC

offspring, 238 of which are immigrants in the United Kingdom, while the remaining 464 live in various parts of India. Some of these couples had children with β-thalassaemia major who were receiving treatment at the Nanavati Hospital Medical Research Centre, Bombay, and the Government Children's Hospital, Madras. The Indian subcontinent is geographically a large region including three different nations namely, India, Pakistan and Bangladesh, with a number of distinct regional and ethnic groups each with their own cultural, religious and linguistic traditions. Thus the precise regional and ethnic origins of each of the studied individuals was ascertained.

Methods

DNA was extracted from the buffy coat cells of EDTA anticoagulated blood samples as described (Old & Higgs, 1982). The mutations were characterized by a PCR method based on allele specific priming which is called the amplification refractory mutation system (ARMS) (Newton *et al*, 1989). The basis of this system is that oligonucleotides with a mismatched 3′ end will not function as amplimers in the PCR under appropriate specific conditions. Thus an oligonucleotide primer could be designed such that its 3′ end is complementary to the sequence of the mutation that is being screened for so that when the PCR is performed under stringent conditions the presence of an amplified product will suggest the presence of the mutation while its absence will suggest the presence of normal DNA sequence at the site of the mutation. Allele specific primers were designed such that the nucleotide at the 3′ end of each primer was complementary to the change of DNA sequence caused by the mutation that was being looked for. To enhance their specificity a deliberate additional mismatch was introduced at position minus 4 from the 3′ end. Each of these primers was tested with appropriate positive and negative DNA samples under uniform stringent PCR conditions to ensure that successful amplification occurred only in the presence of the mutation that was being looked for. In some instances despite the additional deliberate mismatch near the 3′ end false amplification may occur (Newton *et al*, 1989), it is then necessary to alter the additional mismatch or even reverse the direction of the primer to obtain specificity of action.

Mutations selected for study. The 10 β-thalassaemia mutations that have been previously described in Asian Indians, namely the mutations at IVS-1 position 5 (G-C), IVS-1 position 1 (G-T), codons 8/9 (+G), codons 41/42 (-CTTT), codon 15 (G-A), codon 16 (-C), −88 (C-T), cap site +1 (A-C), IVS-1 3′-end -25 bp and the 619 bp deletion at the 3′ end of the gene (Kazazian *et al*, 1984; Wong *et al*, 1986, 1987; Thein *et al*, 1988) were studied. Screening was also done for β-thalassaemia mutations which are frequent in other

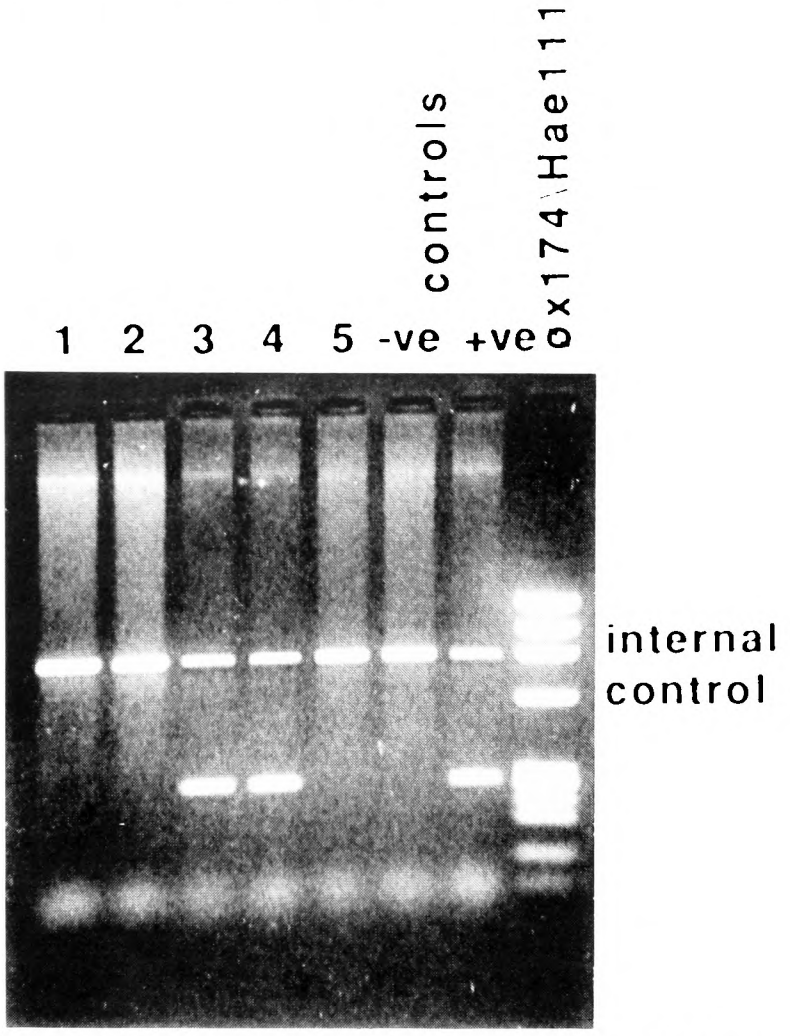


Fig 1. Screening for the mutation at IVS-1 position 5 (G-C). Ethidium bromide stained gel electrophoresis of the PCR products with the allele specific primer for the mutation IVS-1 position 5 (G-C) coupled with primer A, along with the internal control primers C and D (Table I). The 285 bp product seen in lanes 3, 4 and positive control indicates the presence of the mutation, while the absence of this band in lanes 1, 2, 5 and negative control indicates its absence. The 861 bp internal control product in all lanes indicates the efficacy of the PCR reaction. The marker is ox 174 digested with *Hae* III.

population groups. The mutations at IVS-2 position 654 (C-T), codon 17 (A-T) and −28 (A-G) which are frequent in the South-East Asians and the Chinese (Kazazian *et al*, 1986) and those at IVS-1 position 110 (G-A), IVS-1 position 6 (T-C) and IVS-1 position 1 (G-A) which are frequent in the Mediterranean peoples (Kazazian *et al*, 1984; Di Marzo *et al*, 1988; Diaz-Chico *et al*, 1988) were screened for. Further, some of the mutations described in more than one ethnic group were looked for, namely those at IVS-2 position 1 (G-A) (Treisman *et al*, 1982; Wong *et al*, 1986; Chibani *et al*, 1988), codon 39 (C-T) (Orkin & Goff, 1981; Trecartin *et al*, 1981; Moschonas *et al*, 1981) and IVS-2 3′-end (CAG-AAG) (Wong *et al*, 1989).

Allele specific primers. Oligonucleotide primers to detect the 18 mutations selected for this study were synthesized according to the sequences shown in Table I. These primers were coupled with one of the common primers A, B or C (Table I) depending on the location of the mutation in the β-globin gene and the direction of the allele specific primer. To demonstrate the efficacy of the PCR, a set of internal control primers was included. For most reactions the internal control primers were C and D, which amplify a 861 bp fragment from the 3′ end of the β-globin gene spanning the site of the 619 bp deletion described in Asian Indians (Orkin *et al*, 1979), the presence of which results in a small 242 bp fragment. In the reactions to detect the mutation at IVS-2 position 654 (C-T) and IVS-2 3′-end (CAG-AAG), the internal control primers used were E and F (Table I) which amplify a 323 bp fragment from the G_γ globin gene. When each of these allele specific primers were tested with appropriate positive and negative DNA samples under the uniform stringent PCR conditions, the primer for the codon 15 (G-A) mutation showed false positive amplification and it was necessary to reverse its direction and couple it with primer B (Table I) before specificity of action was achieved.

PCR conditions. A 25 μl reaction volume was used, containing approximately 0·5 μg of genomic DNA, 5 pmol of each of the four primers, 0·5 units of Taq polymerase (Amplitaq, Cetus), 30 μM of each dNTP in 10 mM Tris HCl (pH 8·4), 50 mM KCl and 1·5 mM $MgCl_2$. The thermal cycling regimen consisted of 25 cycles, with denaturation at 93°C for 1 min and combined annealing and extension at 66°C for 2 min. In the last cycle extension was prolonged for 3 min. All of the allele specific primers were used at these uniform stringent conditions. The PCR product was visualized after electrophoresis on a 1·5% NuSieve and 1·5% agarose gel by ethidium bromide staining and ultraviolet light illumination. All experiments included appropriate positive and negative controls. However, we did not have a positive control for the mutation at S-2 3′-end (CAG-AAG).

Screening strategy. Samples were first screened for the five mutations which have been found to be most common amongst the Asian Indians namely the ones at IVS-1 position 5 (G-C), codons 8/9 (+G), IVS-1 position 1 (G-T), codons 41/42 (-CTTT) and the 619 bp deletion at the 3′ end of the β-globin gene (Thein *et al*, 1988). Fig 1 shows the screening for the mutation at IVS-1 position 5 (G-C). Uncharacterized samples were then screened for the less common mutations namely the ones at codon 15 (G-A), codon 16 (-C), cap site +1 (A-C), −88 (C-A) and IVS-1 3′-end-25 bp. The samples which still remained uncharacterized were screened for the other mutations selected for this study.

Direct genomic sequencing. The promoter region, the first and second exons and the first intervening sequence of the β-globin gene from one of the individuals in whom the β-thalassaemia mutation remained uncharacterized was enzymatically amplified by the PCR using 20 pmol of each of the primers 5′-CGATCTTCAATATGCTTAC and 5′-CATTCGTCTGTTTCCCATTCTA and a thermal cycling regimen consisting of 30 cycles, with denaturation at 93°C for 1 min, annealing at 55°C for 1 min and extension at 72°C for 1·5 min. In the last cycle extension was prolonged to 3 min. The reaction was performed in a 100 μl volume with other conditions similar to those described for allele specific priming. Following amplification the 916 bp product was isolated by electroelution using the IBI unidirectional electroelutor (International Biotechnologies), ethanol precipitated and resuspended in sterile distilled water. This amplified fragment was directly sequenced by the dideoxynucleotide chain termination method (Sanger *et al*, 1977) using 4 pmol of sequencing primer 5′-CAAAGGACTCAAAGAACCTC, ^{35}S-dATP and Sequenase (United States Biochemical).

Table II. The frequency and regional distribution of the β-thalassaemia mutations on the Indian subcontinent

	Region							
Mutation	Gujarat 197 (28·1%)	N.W. Pakistan 167 (23·8%)	Punjab 142 (20·2%)	Sindh 114 (16·2%)	Tamil Nadu 59 (8·4%)	Maharashtra 13 (1·9%)	Bangladesh and Bengal 10 (1·5%)	Indian Subcontinent 702
IVS-1 position 5 (G-C)	81	59	54	14	48	7	6	269 (38·3%)
619 bp deletion	51		22	61	1	—	—	135 (19·2%)
Codons 8/9 (+G)	8	73	19	12	2	—	1	115 (16·4%)
IVS-1 position 1 (G-T)	33	2	15	20	—	—		70 (10%)
Codons 41/42 (-CTTT)	17	17	20	5	6	1	2	68 (9·7%)
Codon 15 (G-A)	3	11	1	—	—	—	1	16 (2·3%)
Codon 30 (G-C)	4	—	1	1	—	—	—	6 (0·9%)
Codon 16 (-C)		—	4	—	—	—	—	4 (0·6%)
Cap site +1 (A-C)		—	2	—		—	—	2 (0·3%)
IVS-2 position 1 (G-A)	—	—	1	—	—	1	—	2 (0·3%)
−88 (C-T)	—	—	1	—	—	—	—	1 (0·2%)
Uncharacterized	—	5	2	1	2	4	—	14 (2%)

RESULTS

Using allele specific priming of the PCR it was possible to characterize the β-thalassaemia mutations in 688 of the 702 (98%) individuals studied. The β-thalassaemia alleles were heterogeneous; 11 different mutations were defined. However, five common mutations namely the ones at IVS-1 position 5 (G-C), codons 8/9 (+G), IVS-1 position 1 (G-T), codons 41/42 (-CTTT) and the 619 bp deletion at the 3′ end of the gene accounted for 93·6% of the alleles studied. The frequency of the various β-thalassaemia alleles and their regional distribution is depicted in Table II.

Of the 702 β-thalassaemia carriers studied, 197 (28·1%) originated from Gujarat, 167 (23·8%) from North-West Pakistan, 142 (20·2%) from Punjab, 114 (16·2%) from Sindh, 59 (8·4%) from Tamil Nadu, 13 (1·9%) from Maharashtra and 10 (1·4%) from Bengal and Bangladesh. These regions along with their predominant mutations are shown in Fig 2. The mutation at IVS-1 position 5 (G-C) was found to be the commonest one on the Indian subcontinent and was present along with the one at codons 41/42 (-CTTT) in individuals from all of the regions studied. 99% of individuals with the 619 bp deletion and 97% of those with the mutation at IVS-1 position 1 (G-T) were from the neighbouring regions of Sindh, Gujarat and Punjab. In North-West Pakistan the distribution of β-thalassaemia alleles was different from the rest of the subcontinent with the mutation at codons 8/9 (+G) being the commonest one and 69% of individuals with the mutation at codon 15 (G-A) originated from here. The Punjabis were the most heterogeneous group, with at least one representative of every β-thalassaemia mutation that we detected on the subcontinent. While carriers from Tamil Nadu were the most homogeneous group with 81% having the mutation at IVS-1 position 5 (G-C).

The mutation at IVS-2 position 1 (G-A) hitherto described in Mediterraneans (Treisman *et al*, 1982), American blacks (Wong *et al*, 1986) and Tunisians (Chibani *et al*, 1988) was detected in two individuals in this study. We did not find the mutations at IVS-1 3′-end −25 bp, IVS-2 position 654 (C-T), codon 17 (A-T), −28 (A-G), IVS-1 position 110 (G-A), IVS-1 position 6 (T-C), IVS-1 position 1 (G-A), codon 39 (C-T) and IVS-2 3′-end (CAG-AAG) in any of the individuals we studied.

Direct sequencing of amplified DNA from one of the remaining uncharacterized carriers revealed the presence of a mutation at codon 30 (G-C) which has been described in Tunisians (Chibani *et al*, 1988). An allele specific primer was then synthesized (Table I) in order to screen for the presence of this mutation in the other remaining uncharacterized carriers. The codon 30 (G-C) mutation was found in six individuals all of whom originated from the neighbouring regions of Gujarat, Sindh and Punjab. This approach is being currently applied to characterize the β-thalassaemia mutations of the 14 remaining unknown cases.

DISCUSSION

We have investigated the molecular basis of β-thalassaemia in a large number of individuals from the Indian subcontinent and detected 11 different β-thalassaemia mutations in 98% of the individuals investigated, with five common ones accounting for 93·6%. This is in agreement with previous studies in the immigrant Asian Indian population which have reported eight mutations to account for 82% of the 44 β-thalassaemia alleles studied with three common mutations accounting for 70% of cases (Kazazian *et al*, 1984); and nine mutations to account for 98% of the 102 β-thalassaemia alleles studied with five mutations accounting for 88% of cases (Thein *et al*, 1988).

Prevention of β-thalassaemia by genetic counselling and prenatal diagnosis is an important health issue in India. Techniques for prenatal diagnosis of monogenic disorders by DNA analysis have progressed, such that today following DNA amplification by the PCR prenatal diagnosis by the direct detection of the relevant mutations is possible in most

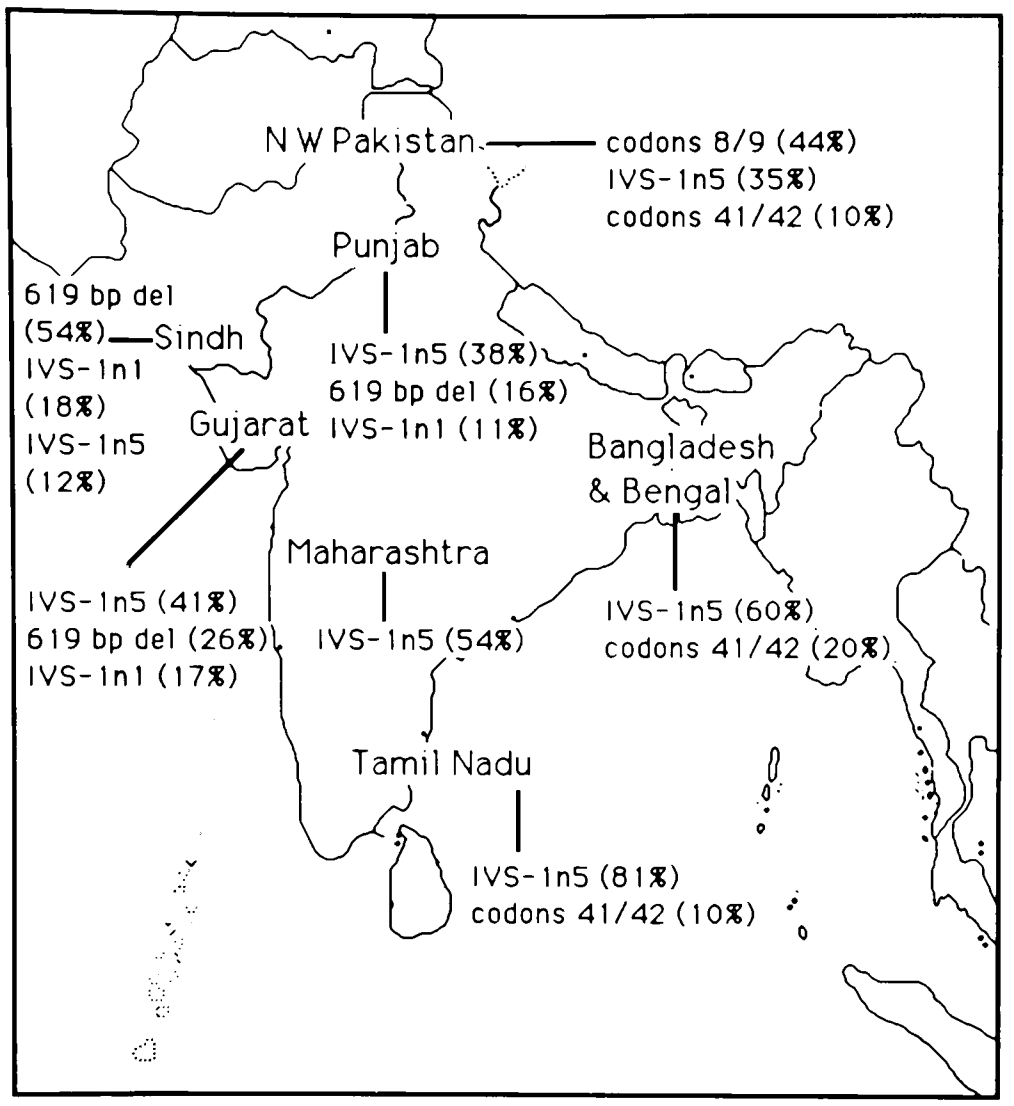


Fig 2. Distribution of β-thalassaemia mutations in different regions of the Indian subcontinent. The frequency of the common mutations in the seven regions studied is shown. (n = nucleotide.)

cases in the first trimester of pregnancy. The technique of allele specific priming of the PCR used for carrier screening in this study, has also been successfully used for over 100 cases of prenatal diagnosis (Old *et al*, 1990). As it detects directly the mutations this method overcomes the labour intensive process of studying the DNA polymorphisms in family members. Furthermore, unlike described techniques of dot–blot analysis of PCR products with labelled radioactive oligoprobes (Saiki *et al*, 1986), this one is nonradioactive, rapid and relatively simple once the reaction conditions have been determined, more so if all the allele specific primers in use are designed to work at uniform PCR conditions. Thus this method is particularly relevant for the prenatal diagnosis of β-thalassaemia on the Indian subcontinent considering the heterogeneous nature of the disease in this population and the scarce resources available for more expensive and tedious methods.

The distribution of the β-thalassaemia gene is not uniform on the Indian subcontinent. The incidence of the β-thalassaemia trait in Gujarat is 10–15%, in Sindh it is 10%, in Punjab 6·5%, 3·5% in Bengal and 4·3% in South India (Sukumaran & Master, 1974). There is little information available about the carrier rate in Pakistan; 4% has been reported in the North-West Frontier region (Stern *et al*, 1968). This is reflected in the regional distribution of the individuals we studied. The distribution and frequency of the various β-thalassaemia mutations that we have described in this large study should assist in a national prenatal diagnosis programme. Furthermore, some of the regional variations that we have been able to show like the predominance of the mutation at IVS-1 position 5 (G-C) in Tamil Nadu and the relative predominance of the one at codons 8/9 (+G) and at codon 15 (G-A) in North-West Pakistan, should provide guidelines for attempts to characterize the mutations in candidate couples, depending on their ethnic and regional origins.

Throughout its history the Indian subcontinent has had political, military and commercial interactions with the peoples of Central, Western and South-East Asia, the Mediterranean region and later Western Europe (Thapar, 1966), which may account for the heterogeneous nature of β-thalassaemia in this population. However, the tradition of marriage within each ethnic group are responsible for the regional variation in the distribution of the β-thalassaemia mutations that we have found. Indeed, 209 of the 335 (62%) couples we studied had identical mutations also reflecting the social practices of marriage within each community. Further in the Muslim communities of North-West Pakistan where consanguinity is favoured, 85·2% of couples had the same mutation.

This has been the first study to describe the mutations at IVS-2 position 1 (G-A) and codon 30 (G-C) in Asian Indians. 14 alleles yet remain uncharacterized in spite of screening for 20 different mutations. In terms of a prenatal diagnosis programme this is not a significant proportion; however, studies to try and characterize these alleles by DNA sequencing are being currently undertaken.

ACKNOWLEDGMENTS

We wish to thank Dr P. G. Natrajan, Nanavati Hospital

Medical Research Centre, Bombay, and Professor I. C. Verma, All India Institute of Medical Sciences, New Delhi, for their assistance in providing the subjects for this study. We also wish to thank Dr S. L. Thein for the oligonucleotide primers used for direct genomic sequencing, the Rhodes Trust for supporting Dr N. Y. Varawalla and the British Council, Madras Division, for supporting Dr R. Sarkar.

REFERENCES

Amselem, S., Nunes, V., Vidaud, M., Estivill, X., Wong, C., d'Auriol, L., Vidaud, D., *et al* (1988) Determination of the spectrum of β-thalassaemia in Spain by use of dot-blot analysis of amplified β globin DNA. *American Journal of Human Genetics*, **43**, 95–100.

Angastiniotis, M.A. & Hadjiminas, M.G. (1981) Prevention of thalassaemia in Cyprus. *Lancet*, i, 369–371.

Cao, A., Pintus, L., Lecca, U., Olla, G., Cossu, P., Rosatelli, C. & Galanello, R. (1984) Control of homozygous β-thalassaemia by carrier screening and antenatal diagnosis in Sardinians. *Clinical Genetics*, **26**, 12–22.

Chan, V., Chan, T.K., Chebab, F.F & Todd, D. (1987) Distribution of β-thalassaemia mutations in South China and their association with haplotypes. *American Journal of Human Genetics*, **41**, 678–685.

Chebab, F.F., Der Kaloustian, V., Khouri, F.P., Deeb, S.S. & Kan, Y.W. (1987) The molecular basis of β-thalassemia in Lebanon: application to prenatal diagnosis. *Blood*, **69**, 1141–1145.

Chibani, J., Vidaud, M., Duquesnoy, O., Berge-Lefran, C., Pirastu, M., Ellouze, F., Rosa, J., *et al* (1988) The peculiar spectrum of β-thalassaemia genes in Tunisia. *Human Genetics*, **78**, 190–192.

Diaz-Chico, J.C., Yang, K.G., Stoming, T.A., Efremov, D.G., Kutlar, A., Lutlar, F., Aksoy, M., *et al* (1988) Mild and severe β-thalassemia among homozygotes from Turkey: identification of the types by hybridization of amplified DNA with synthetic probes. *Blood*, **71**, 248–251.

Di Marzo, R., Dowling, C.E., Wong, C., Maggio, A. & Kazazian, H.H. (1988) The spectrum of β-thalassaemia in Sicily. *British Journal of Haematology*, **69**, 393–397.

Hadji, A., Papadakis, M., Antsaklis, A., Mesoghitis, S. & Loukopoulos, D. & Fessas, F. (1987) Prevention of thalassaemia in Greece. II. Prenatal diagnosis. Paper presented at the Second International Conference on Thalassaemia and the Hemoglobinopathies, 37 (abstract RT7-56).

Kazazian, H.H., Orkin, S.H., Antonarakis, S.E., Sexton, J.P., Boehm, C.D., Goff, S.C. & Waber, P.G. (1984) Molecular characterization of seven β-thalassaemia mutations in Asian Indians. *EMBO Journal*, **3**, 593–596.

Kazazian, H.H., Jr, Orkin, S.H., Markham, A.F., Chapman, C.R., Youssoufian, H. & Waber, P.G. (1984) Quantification of the close association between DNA haplotypes and specific β-thalassaemia mutations in Mediterraneans. *Nature*, **310**, 152–154.

Kazazian, H.H., Dowling, C.E., Waber, P.G., Huang, S. & Lo, W.H.Y. (1986) The spectrum of β-thalassaemia genes in China and South East Asia. *Blood*, **68**, 964–966.

Kazazian, H. & Boehm, C.D. (1988) Molecular basis and prenatal diagnosis of β-thalassemia. *Blood*, **72**, 1107–1116.

Modell, B.C. & Petrou, M. (1983) The problem of haemoglobinopathies in India. *Indian Journal of Haematology*, **1**, 5–16.

Moschonas, N., deBoer, E., Grosveld, F.G., Dahl, H.H.M., Shewmaker, C.K. & Flavell, R.A. (1981) Structure and expression of a cloned β-thalassaemia globin gene. *Nucleic Acids Research*, **9**, 4391–4401.

Newton, C.R., Graham, A., Heptinstall, L.E., Powell, S.J., Summers, C., Kalshekar, N. & Smith, J.C., *et al* (1989) Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS). *Nucleic Acids Research*, **17**, 2503–2516.

Old, J.M. & Higgs, D.R. (1982) Gene analysis. *Methods in Haematology*, Vol. 6. *The Thalassaemias*. Churchill Livingstone, Edinburgh.

Old, J.M., Varawalla, N.Y. & Weatherall, D.J. (1990) Rapid detection and prenatal diagnosis of β-thalassaemia: studies in Indian and Cypriot populations in the U.K. *Lancet*, **336**, 834–837.

Orkin, S.H., Old, J.M., Weatherall, D.J. & Nathan, D.G. (1979) Partial deletion of β-globin gene DNA in certain patients with β° thalassemia. *Proceedings of the National Academy of Sciences of the United States of America*, **76**, 2400–2404.

Orkin, S.H. & Goff, S.C. (1981) Nonsense and frameshift mutants in β-thalassaemia detected in cloned β-globin genes. *Journal of Biological Chemistry*, **256**, 9782–9784.

Saiki, R.K., Bugawan, T.L., Horn, G.T., Mullis, K.B. & Erlich, H.A. (1986) Analysis of enzymatically amplified β-globin gene and HLA-DQ alpha DNA with allele-specific oligonucleotide probes. *Nature*, **324**, 163–166.

Sanger, F., Nicklen, S. & Coulson, A.R. (1977) DNA sequencing with chain termination inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*, **74**, 5463–5469.

Stern, M.A., Kynoch, P.A. & Lehmann, H. (1968) β-thalassaemia, glucose-6- phosphate dehydrogenase deficiency, and haemoglobin D-Punjab in Pathans. *Lancet*, i, 1284–1285.

Sukumaran, P.K. & Master, H.R. (1974) The distribution of abnormal haemoglobin in the Indian population. *Proceedings of the First Conference of the Indian Society for Human Genetics*, 1, 91–94.

Thapar, R. (1966) *A History of India*, Vol. 1. Penguin Books Ltd, Harmondsworth.

Thein, S.L., Hesketh, C., Wallace, R.B. & Weatherall, D.J. (1988) The molecular basis of thalassaemia major and thalassaemia intermedia in Asian Indians: application to prenatal diagnosis. *British Journal of Haematology*, **70**, 225–231.

Trecartin, R.F., Liebhaber, S.A., Chang, J.C., Lee, Y.W. & Kan, Y.W. (1981) β°-thalassemia in Sardinia is caused by a nonsense mutation. *Journal of Clinical Investigation*, **68**, 1013–1017.

Treisman, R., Proudfoot, N.J., Shander, M. & Maniatis, T. (1982) A single base change at a splice site in a β°-thalassemic gene causes abnormal RNA splicing. *Cell*, **79**, 903–911.

Weatherall, D.J. & Clegg, J.B. (1981) *The Thalassaemia Syndromes*, 3rd edn. Blackwell Scientific Publications, Oxford.

Wong, C., Antonarakis, S.E., Goff, S.C., Orkin, S.H., Boehm, C.D. & Kazazian, H.H. (1986) On the origin and spread of β-thalassemia: recurrent observation of four mutations in different ethnic groups. *Proceedings of the National Academy of Sciences of the United States of America*, **83**, 6529–6532.

Wong, C., Dowling, C.E., Saiki, R.K., Higuchi, R.G., Erlich, H.A. & Kazazian, H.H. (1987) Characterization of β-thalassaemia mutations using direct genomic sequencing of amplified single copy DNA. *Nature*, **330**, 384–386.

Wong, C., Antonarakis, S.E., Goff, S.C., Orkin, S.H., Forget, B.G., Nathan, D.G. & Giardina, P.J.V. (1989) β-thalassemia due to two novel nucleotide substitutions in consensus acceptor splice sequences of the β-globin gene. *Blood*, **173**, 914–918.

British Journal of Haematology, 1991, **79**, 640–644

RAPID PAPER

Rare β-thalassaemia mutations in Asian Indians

N. Y. VARAWALLA, J. M. OLD AND D. J. WEATHERALL *Institute of Molecular Medicine, John Radcliffe Hospital, Oxford*

Received 5 July 1991; accepted for publication 20 August 1991

Summary. Five β-thalassaemia mutations hitherto undescribed in Asian Indians were identified in β-thalassaemia carriers originating from the Indian subcontinent by direct sequencing of their β-globin genes which were amplified by the polymerase chain reaction (PCR). A T–G substitution at IVS 2 position 837, which probably creates an alternative acceptor splice site and a T insertion in codon 88, resulting in a shift in the reading frame with a premature stop codon, are new β-thalassaemia mutations. The others were framshift codon 5 (−CT), IVS 1 position 110 (G–A) and IVS-1 minus 1 (G–A) which have been described previously in other populations. These results complete the characterization of the β-thalassaemia mutations in 708 carriers of Asian Indian origin and will enable a comprehensive programme of carrier screening and prenatal diagnosis of β-thalassaemia in this population.

β-Thalassaemia is probably the most important monogenic disorder amongst Asian Indians and prevention of the disease by genetic counselling and prenatal diagnosis is an important health issue. Detection of β-globin gene mutations in carrier parents and from fetal tissue obtained in the first trimester of pregnancy is now relatively straightforward. The success of this approach depends on the knowledge of the spectrum of β-thalassaemia mutations in the population.

The heterogeneous nature of β-thalassaemia is well known with 91 different mutations now described (Kazazian, 1990). Each population group in which the disease is prevalent has a different spectrum of β-globin gene mutations and 11 mutations have been identified in the Asian Indian population, with five common ones accounting for 93·6% of the β-thalassaemia alleles (Varawalla *et al.* 1991). Here we report five rare β-thalassaemia mutations that have been identified in carriers of Asian Indian origin by direct sequencing of their amplified β-globin genes.

MATERIALS AND METHODS

Subjects. The subjects were 19 unrelated individuals from a panel of 708 carriers of β-thalassaemia originating from various parts of the Indian subcontinent and were previously found not to carry any of the 11 Asian Indian β-thalassaemia mutations (Varawalla *et al.* 1991).

DNA amplification and sequencing. DNA was prepared from the buffy coat cells of EDTA anticoagulated blood samples by phenol extraction (Old & Higgs, 1982). A 2·1 kilobase (kb) fragment containing the β-globin gene along with its 5′ and 3′ flanking sequences was amplified using standard conditions (Saiki *et al.* 1988) from 1 μg of genomic DNA with primers P1 and P2 (Table I). The thermal cycling regimen consisted of 30 cycles, with denaturation at 94°C for 2 min, annealing at 60°C for 2 min and extension at 72°C for 5 min. In the first cycle denaturation was prolonged for 5 min and in the last cycle extension was prolonged for 10 min. The PCR product was isolated by electrophoresis onto a DEAE membrane (NA45 Schleicher & Schuell), eluted in 400 μl of buffer (1·5 M NaCl, 20 mM Tris HCl pH 7·6 and 1 mM EDTA) at 65°C for 20 min and purified by two phenol/chloroform extractions and ethanol precipitation.

Approximately 5 ng of this purified fragment was used as a template for a second PCR in which 40 pmol of the primer P1 (which is complementary to the noncoding strand) was used without the addition of primer P2 in order to produce a single-stranded PCR product. The single-stranded product was identified in an ethidium bromide stained 1% agarose gel as a faster moving and less intensely stained fragment in comparison to the double-stranded template. The single-stranded fragment was isolated as described above and resuspended in 21 μl of distilled water for use as a sequencing template.

DNA sequencing was performed by the dideoxy chain termination method (Sanger *et al.* 1977) using 7 μl of template, 4 pmol of the appropriate internal sequencing primer (Table I), 5 μCi ^{35}S-dATP (Amersham UK) and 2 units of Sequenase (United States Biochemical). As the DNA samples were from heterozygous individuals both the normal and mutant alleles were sequenced. Thus a substitution

Correspondence: Dr N. Y. Varawalla, Institute of Molecular Medicine, John Radcliffe Hospital, Oxford OX3 9DU.

Table I. Oligonucleotide primers used for DNA amplification and sequencing of the β-globin gene

Name	Position of 5′-nucleotide	Sequence
PCR primers		
P1	307 nt upstream from the Cap site	5′-CGATCTTCAATATGCTTACCAAG
P2	320 nt downstream from the termination site	5′-GCATAGGCATCAGGGCTGTTG
Sequencing primers		
S1	Codon 22, exon 1	5′-TTCATCCACGTTCACCTT
S2	Codon 46, exon 2	5′-CAAAGGACTCAAAGAACCTC
S3	IVS-2 nt 114	5′-CATTCGTCTGTTTCCCATTCTA
S4	IVS-2 nt 293	5′-CTTAATGTATCTCAGAGATA
S5	IVS-2 nt 784	5′-CCAGCCTTATCCCAACCAT
S6	67 nt downstream from the termination site	5′-CAGTTTAGTAGTTGGACTTA
S7	279 nt downstream from the termination site	5′-GCAGCCTCACCTTCTTTCATGG

PCR primer P1 is complementary to the noncoding strand while P2 and all the sequencing primers are complementary to the coding strand. (nt=nucleotide.)

Table II. Allele specific primers for the detection of β-thalassaemia mutations

Mutation	Sequence of the allele specific primer	Product size	No. of cases
Frameshift codon 5 (−CT)	5′-ACAGGGCAGTAACGGCAGACTTCTCCGCGA	205 bp	10 (1·4)*
IVS-2 nt 837 (T–G)	5′-CCTTTTGCTAATCATGTTCATACCTCGTAG	646 bp	4 (0·6)
IVS-1 nt 110 (G–A)	5′-ACCAGCAGCCTAAGGGTGGGAAAATACACT	419 bp	2 (0·3)
IVS-1 minus 1 (G–A)	5′-TAAACCTGTCTTGTAACCTTGATACCTACT	280 bp	2 (0·3)
Frameshift codon 88 (+T)	5′-GTGCAGCTTGTCACAGTGCAGCTCACACAA	613 bp	1 (0·1)

Other primers used: common upstream primer A: 5′-ACCTCACCCTGTGGAGCCAC; internal control primers B: 5′-CAATGTATCATGCCTCTTTGCACC and C: 5′-GAGTCAAGGCTGAGAGATGCAGGA (these amplify a 861 bp fragment from the 3′ end of the β-globin gene).

The allele specific primer for IVS-2 nt 837 (T–G) was coupled with downstream primer C and the internal control primers were used D: 5′-AGTGCTGCAAGAAGAACAACTACC and E: 5′-CTCTGCATCATGGGCAGT GAGCTC which amplify a 323 bp fragment from the Gγ-globin gene. (nt=nucleotide).

* Numbers in parentheses are percentages.

appeared as two bands at that position and a frameshift mutation resulted in two sequencing ladders being present from the site of the mutation.

Allele specific priming of PCR. Once a mutation was identified, an allele specific primer was synthesized such that its 3′ nucleotide was complementary to the mutation. To enhance specificity of action a mismatch was introduced at position minus 4 from the 3′ nucleotide. The primer was used to confirm the presence of the mutation in the sequenced DNA sample after it had been observed to function specifically by failing to amplify appropriate negative controls. The primer was also used to identify the mutation in other family members and lastly to screen for the mutation in the remaining uncharacterized individuals. Each allele specific primer was coupled with one of the common primers A or C and internal control primer pairs B and C or D and E were also included as listed in Table II. The PCR conditions and methods used were as previously reported (Old *et al*, 1990; Varawalla *et al*, 1991), except that the annealing temperature for the IVS-2 nucleotide 837 (G–T) primer was 64°C.

RESULTS

Five different β-globin gene mutations were identified following DNA sequence analysis of the amplified β-globin gene fragment spanning positions −307 to 320 nucleotides downstream from the termination site. Two of these have not been previously described. The first was a T–G substitution at position 837 of the second intron of the β-globin gene which results in an AG dinucleotide and a sequence (AGCTTC) in which five out of six nucleotides match that of the authentic 3′ splice junction of the second intron (AGCTCC). The second new mutation was a single nucleotide insertion (+T) at codon 88 which creates a shift in the reading frame leading to a premature stop codon TGA at codon 90 (Fig 1). The other three mutations have not been previously described in individuals of Asian Indian origin. These were the framshift mutation at codon 5 (−CT), the IVS-1 position 110 (G–A) mutation and a consensus change mutation at IVS-1 position minus 1 (codon 30) (G–A).

Studies with each of the allele specific primers confirmed

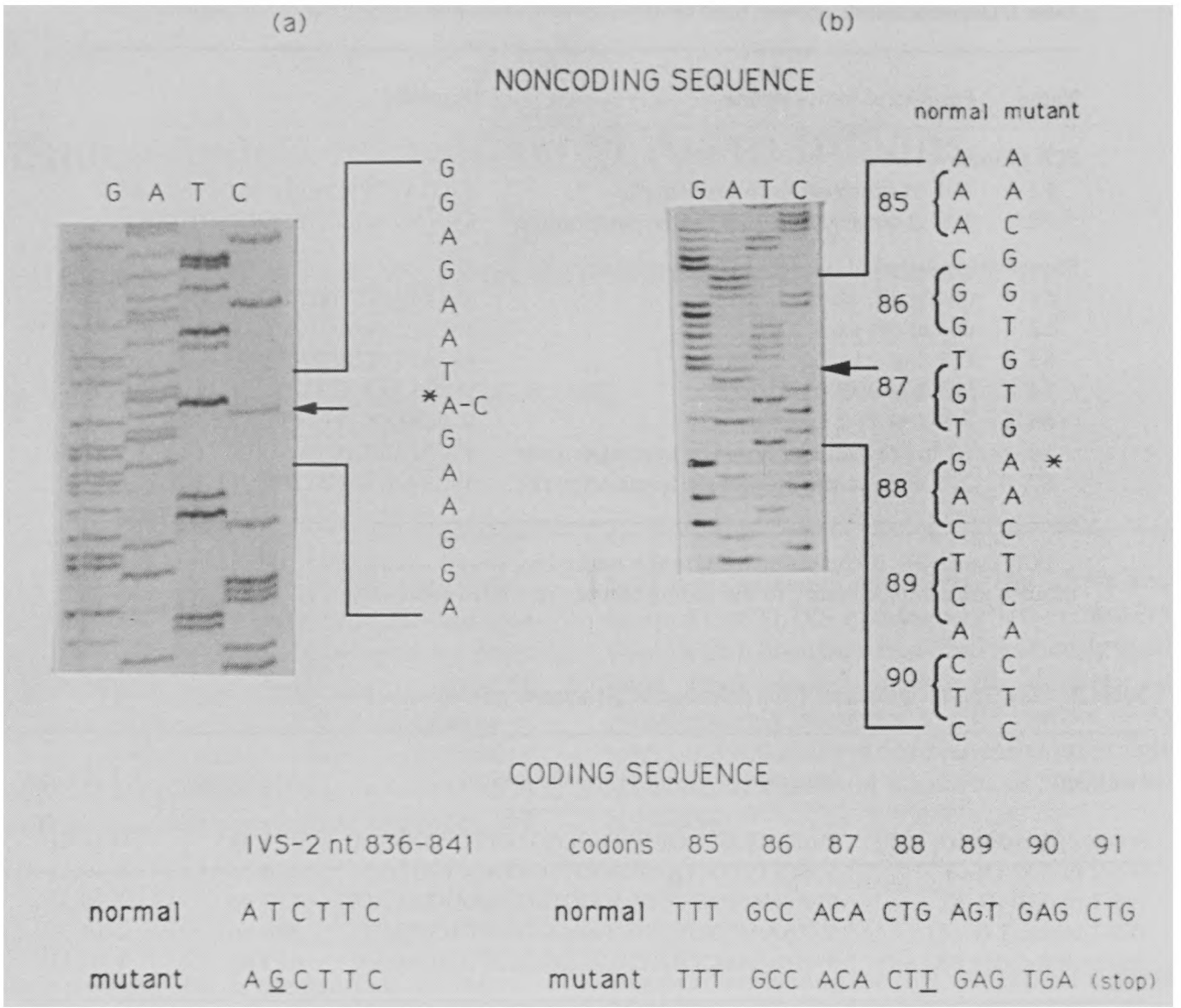


Fig 1. Sequence analysis of the amplified β-globin gene DNA from individuals heterozygous for the β-thalassaemia mutations: (a) a T–G substitution at IVS-2 position 837 (b) a frameshift mutation (+T) at codon 88. The sequencing primers used were S6 and S3 respectively (Table I) which were complementary to the coding strand thus the noncoding sequence is shown. The position of the mutation is marked by * and an arrow. In Fig 1(b) the codons are indicated by brackets. The coding sequence of the respective regions is shown below with the mutation underlined.

the results of DNA sequence analysis by the presence of an appropriate size PCR product visualized on an ethidium bromide stained agarose gel with the DNA sample carrying the mutation along with its absence with the normal negative control sample. To further confirm these findings family studies were performed for each of the individuals. Fig 2 shows such a study for the codon 5 (−CT) mutation where its transmission through three generations in a particular kindred can be seen. This approach was also used to screen the remaining uncharacterized β-thalassaemia carriers. Fig 2 also shows the results of the screening for the IVS-1 minus 1 (G–A) mutation amongst seven uncharacterized disease carriers. In this manner it was possible to identify the β-globin gene mutation in all the subjects. The frequency of each of these mutations amongst Asian Indians which is listed in Table II was calculated using the total sample size of 708 subjects.

DISCUSSION

We have described five rare β-globin gene mutations amongst Asian Indian β-thalassaemia carriers, two of which are newly described mutations. The T–G nucleotide substitution at IVS-2 position 837 creates a sequence that is homologous in five out of six nucleotides to that at the authentic 3′ splice junction. This subsitution results in an AG dinucleotide 14 base pairs upstream from the invariant AG dinucleotide found at the normal acceptor site and thus could possibly create an alternative acceptor splice site which gets preferentially utilized causing abnormal splicing of β-globin mRNA, as has been described for the G–A and T–G substitutions at IVS-1 position 110 and 116 respectively (Spritz *et al*, 1981; Metherall *et al*, 1986). As all the individuals in whom the IVS-2 position 837 (T–G) mutation was identified were heterozygous for β-thalassaemia, it is not possible to comment whether this mutation would cause β° or β^{+} thalassaemia. The second new mutation is most likely to be of a β° type as the insertion of a single T nucleotide in codon 88 of the second exon causes a shift in the reading frame such that premature termination of translation would occur at codon 90, as seen with other nonsense and frameshift β-globin gene mutations (Cheng *et al*, 1984).

Of the β-thalassaemia mutations we identified that have been previously reported in other populations, the frameshift

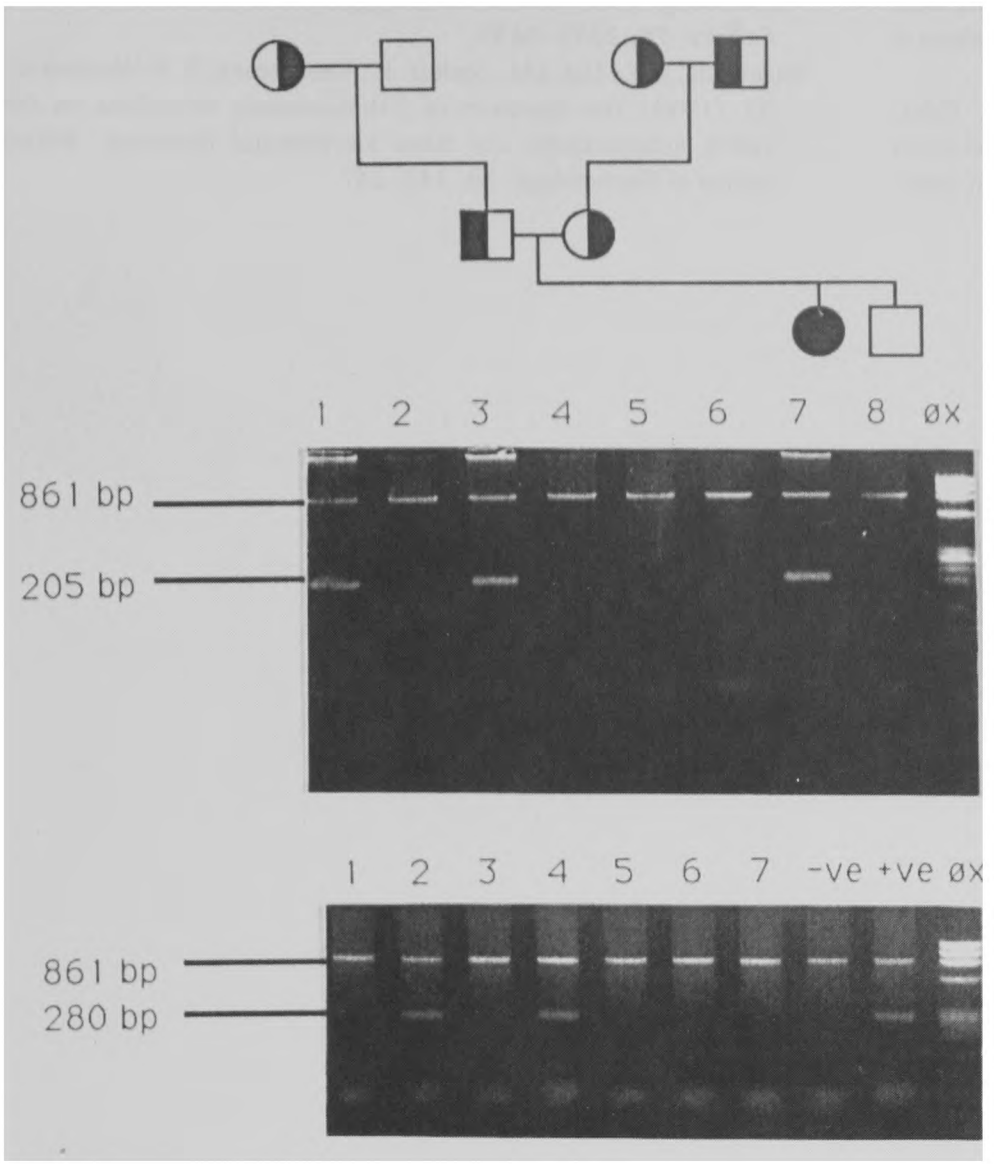


Fig 2. Ethidium bromide stained agarose gels showing allele specific priming of the PCR for the detection of β-thalassaemia mutations. The upper gel shows the results of a family study for the codon 5 (−CT) mutation. In the pedigree, half-shaded symbols represent carriers of β-thalassaemia, totally shaded symbols represent individuals with thalassaemia major and unshaded symbols represent normal individuals. The lower gel shows the results of screening for the IVS-1 minus 1 (G–A) mutation in seven uncharacterized carriers together with the negative and positive controls. The presence of the mutation was indicated by successful amplification with the respective allele specific primer to produce an appropriately sized amplification product (Table II) as is seen in lanes 1, 3 and 7 in the family study and in lanes 2, 4 and the positive control in the screening experiment. The 861 bp fragment represents the amplification product of internal control primers B and C (Table II) and the marker is Øx 174 DNA digested with *Hae*III (Øx).

codon 5 (−CT) mutation has been described in a Greek individual (Kollia *et al.* 1989) and the consensus change mutation at IVS-1 minus 1 (G–A) in the Bulgarian population (Kalaydjieva *et al.* 1989). Interestingly the IVS-1 position 110 mutation which is widely prevalent in the Mediterranean basin (Kazazian & Boehm, 1988) was found to account for 0·3% of β-thalassaemia alleles on the Indian subcontinent.

A total of 16 different β-thalassaemia mutations have now been described in Asian Indians and using the allele specific priming technique it is now possible to detect them all rapidly. The strategy presented here of direct sequencing followed by detection with allele specific primers should enable characterization of the β-thalassaemia mutation in any future rare individual in whom the mutation remains unknown after screening for the 16 known ones.

ACKNOWLEDGMENTS

We wish to thank Dr S. L. Thein for the oligonucleotide primers used for DNA sequencing. Dr Nermeen Y. Varawalla holds a Rhodes research fellowship.

REFERENCES

Cheng, T., Orkin, S.H., Antonarkis, S.E., Potter, M.J., Sexton, J.P., Markham, A.F., Giardina, P.J., *et al* (1984) β-Thalassemia in Chinese: use of *in vivo* RNA analysis and oligonucleotide hybridization in systematic characterization of molecular defects. *Proceedings of the National Academy of Sciences of the United States of America*, **81**, 2821–2825.

Kalaydjieva, L., Eigel, A. & Argiris, A. (1989) The molecular basis of β-thalassaemia in Bulgaria. *3rd International Conference on Thalassaemia and the Haemoglobinopathies, Sardinia*, p. 43 (abstract).

Kazazian, H.H. (1990) The thalassemia syndromes: molecular basis and prenatal diagnosis in 1990. *Seminars in Haematology*, **27**, (3), 209–228.

Kazazian, H.H. & Boehm, C.D. (1988) Molecular basis and prenatal diagnosis of β-thalassemia. *Blood*, **72**, 1107–1116.

Kollia, P., Gonzalez-Redondo, J.M., Stoming, T.A., Loukopoulos, D., Politis, C. & Huisman, T.H. (1989) Frameshift codon 5 [FSC-5 (−CT)] thalassaemia: a novel mutation detected in a Greek patient. *Haemoglobin*, **13**, (6), 597–604.

Metherall, J.E., Collins, F.S., Pan, J., Weissman, S.M. & Forget, B.G. (1986) β°-thalassaemia caused by a base substitution that creates an alternative splice acceptor site in an intron. *EMBO Journal*, **5**, 2551–2557.

Old, J.M. & Higgs, D.R. (1982) Gene analysis. *Methods in Haematology*, Vol. 6. *The Thalassaemias*. Churchill Livingstone, Edinburgh.

Old, J.M., Varawalla, N.Y. & Weatherall, D.J. (1990) Rapid detection and prenatal diagnosis of β-thalassaemia: studies in Indian and Cypriot populations in the U.K. *Lancet*, **336**, 834–837

Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B. & Erlich, H. (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science*, **239**, 487–91.

Sanger, F., Nicklen, S. & Coulson, A.R. (1977) DNA sequencing with chain termination inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*, 74, 5463–5469.

Spritz, R.A., Jagadeeswaran, P., Choudary, P.V., Biro, P.A., Elder, J.T., deRiel, J.K., Manley, J.L., *et al* (1981) Base substitution in an intervening sequence of a β^+-thalassemic human globin gene. *Proceedings of the National Academy of Sciences of the United States of America*, **78**, 2455–2459.

Varawalla, N.Y., Old, J.M., Sarkar, R., Venkatesan, R. & Weatherall, D.J. (1991) The spectrum of β-thalassaemia mutations on the Indian subcontinent: the basis for prenatal diagnosis. *British Journal of Haematology*, **78**, 242–247