

Molecular diagnosis of haemoglobin disorders

B. E. CLARK, *Department of Haematological Medicine, King's College Hospital and GKT School of Medicine,*
S. L. THEIN *Denmark Hill, London, UK*

Summary The haemoglobinopathies refer to a diverse group of inherited disorders characterized by a reduced synthesis of one or more globin chains (thalassaemias) or the synthesis of a structurally abnormal haemoglobin (Hb). In prevalent regions, the thalassaemias often coexist with a variety of structural Hb variants giving rise to complex genotypes and an extremely wide spectrum of clinical and haematological phenotypes. An appreciation of these phenotypes is needed to facilitate the definitive diagnosis of the causative mutations to inform management and counselling. Haematological and biochemical investigations, and family studies provide essential clues to the different interactions and are fundamental to DNA diagnostics of the Hb disorders. With the exception of a few rare deletions and rearrangements, the molecular lesions causing haemoglobinopathies are all identifiable by PCR-based techniques. Although a full spectrum of >1000 mutations causing haemoglobinopathies has been documented, in practice only a limited number are associated with disease states and clinical significance. Furthermore, each at-risk ethnic group has its own combination of common Hb variants and thalassaemia mutations. Prior identification of the ethnic origin is thus an important part of the diagnostic strategy which becomes less reliable in the UK because of the large ethnic mix. Although the current approach using a combination of different PCR-based techniques seems to work in most laboratories, practice pressures with the imminent implementation of universal antenatal screening for clinically significant Hb disorders in the UK will require a higher throughput approach for DNA diagnostics in the near future. The complex mutational spectrum and the compactness of the globin genes places them in an ideal position for the different non-gel based analytical platforms.

Keywords Haemoglobinopathies, DNA diagnostics, thalassaemia, Hb variants, sickle

Introduction

Haemoglobin (Hb) is a tetramer of two α -like and two β -like globin chains. Each of these globin chains is covalently linked to a prosthetic oxygen-binding heme group. In healthy adults, approximately 95% of the Hb is Hb A ($\alpha_2\beta_2$) with <3.2% being Hb A₂ ($\alpha_2\delta_2$) and <1% being foetal Hb (Hb F, $\alpha_2\gamma_2$). The α globin chain is encoded by two α globin genes (α_2 and α_1), located on the tip of chromosome 16p, while the γ , δ and β chains are

encoded by their respective genes ($^G\gamma$ and $^A\gamma$, δ and β) located in a cluster on chromosome 11p (Figure 1). Throughout human development, exquisitely co-ordinated expression of these α -like and β -like globin chains is required to generate a balanced and high concentration of haemoglobin within the red blood cells.

The haemoglobinopathies refer to a diverse group of disorders caused by a disruption of this normal pattern of globin gene expression (Weatherall & Clegg, 2001). The disorders are characterized by either a reduced synthesis of one or more of the globin chains (thalassaemias) or the synthesis of a structurally abnormal Hb variant. These characteristics, however, are not mutually exclusive. Some structural Hb variants are also ineffectively synthesized or the globin chain variants are so unstable that they are not able to form tetramers, resulting in a functional

Accepted for publication 31 March 2004

Correspondence: Prof. Swee Lay Thein, Department of Haematological Medicine, King's College Hospital and GKT School of Medicine, Denmark Hill, London SE5 9PJ, UK. Tel.: 020 7346 1682/1689; Fax: 020 7346 1681/5178; E-mail: sl.thein@kcl.ac.uk

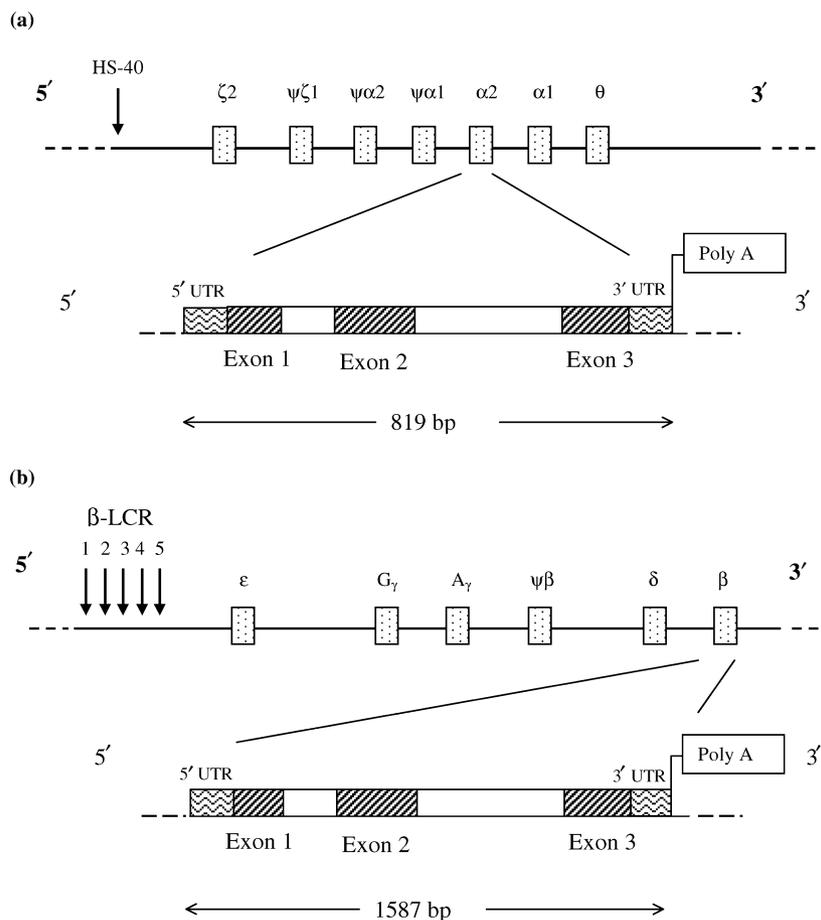


Figure 1. The α - and β -like globin gene clusters, with each globin gene having three exons and two introns. (a) The alpha locus on chromosome 16 p13.3. Of the seven genes shown only three genes are expressed at a clinically significant level, namely the $\zeta 2$ gene (expressed early in foetal development) and the $\alpha 1$ and $\alpha 2$ genes (the focus of most haemoglobinopathy tests). The other genes shown are pseudogenes ($\psi\zeta 1$, $\psi\alpha 1$ and $\psi\alpha 2$) or are not expressed at a significant level (θ). The vertical arrow indicates the location of the upstream hypersensitive site (HS-40), important for locus gene expression. The $\alpha 2$ gene structure is shown below indicating the three coding exons (striped boxes), the two introns (open boxes), and the untranslated regions (zigzag boxes) common to all globin genes. (b) The beta locus on chromosome 11 p15.4 with the ϵ , G_γ and A_γ , δ and β genes, arranged in the order of their developmental expression. During development two switches are made in beta like globin expression. The first occurs early in foetal development from ϵ globin to the γ globin genes (both G_γ and A_γ) at 6–8 weeks gestation, and the second switch occurs shortly prior to birth, from the γ genes to the δ and β genes. Again $\psi\beta$ denotes a nonexpressed pseudogene. Upstream of the beta globin cluster is the β locus control region (β LCR) comprised of the five hypersensitive sites indicated by vertical arrows. As in panel A, the β globin gene is expanded to show the gene structure covering 1.6 kb. This is bigger than the alpha genes because of a larger second intron.

deficiency of the globin chain and a thalassaemic phenotype. The former is sometimes referred to as thalassaemic haemoglobinopathies, and includes the $\delta\beta$ fusion variants (Hb Lepore) and Hb E, $\beta 26$ (Glu \rightarrow Lys), in which the substitution at β -codon 26 (GAG \rightarrow AAG) also causes alternative splicing of the β globin mRNA leading to a reduction of the normally spliced β message encoding the Hb E variant. The hyperunstable globin chains act in a dominant negative fashion causing a disease phenotype even when present in a single copy (Thein, 1999; Thein, 2001); for example Hb Showa Yakushiji caused by a single base substitution in β codon 110 (CTG \rightarrow CCG)

Leu \rightarrow Pro (Kobayashi *et al.*, 1987). Another group of haemoglobinopathies, referred to as hereditary persistence of foetal haemoglobin (HPFH), is characterized by variable increases of Hb F in otherwise normal adults (Wood, 2001). Because of their concomitant increase in Hb F levels, the $\delta\beta$ - and $\gamma\delta\beta$ -thalassaemias are often considered within the syndrome of raised Hb Fs, forming a continued spectrum within the HPFHs. For practical and clinical reasons however, the distinction between $\delta\beta$ -thalassaemias and HPFH should be retained. Homozygotes or compound heterozygotes of $\delta\beta$ -thalassaemia with β -thalassaemia have a clinical picture of thalassaemia

intermedia or major. In contrast, HPFH homozygotes or compound heterozygotes with β -thalassaemia tend to be clinically very mild (Rochette, Craig & Thein, 1994; Weatherall & Clegg, 2001).

As a group, the haemoglobinopathies are the commonest single gene disorder in the world and are found at high frequencies in tropical and sub-tropical regions where malaria is endemic (Flint *et al.*, 1998). A full spectrum of >1000 mutations has been documented, from rearrangements and deletions of gene(s) to single-base substitutions altering amino acid sequence. The globin gene server is a comprehensive database of these mutations and is accessible at <http://globin.cse.psu.edu/> (Hardison *et al.*, 2002). In many regions, α - and β -thalassaemia coexist with a variety of different structural Hb variants; these complex interactions give rise to an extremely wide spectrum of clinical phenotypes. There are four main categories of haemoglobin disorders that are associated with severe disease states and clinical significance.

- Sickle cell disease (major genotypes Hb S/S, Hb S/C and Hb S/ β -thal, and less common genotypes Hb S/D^{Punjab}, Hb S/O^{Arab} and Hb S/Lepore)
- β -thalassaemia syndrome; including $\delta\beta$ -thalassaemia and Hb E/ β -thalassaemia
- α -thalassaemia syndrome
- Hb variants resulting in haemolytic anaemias, polycythaemias and, more rarely, cyanosis

It is important to be able to make a definitive diagnosis of these causative mutations to inform management and counselling. Couples at risk of having an affected child can be offered genetic counselling and reproductive options including prenatal diagnosis which involves foetal sampling to determine the fetal genotype.

DNA-based methods do 'not' replace haematological and biochemical investigations

With few exceptions, the majority of the haemoglobinopathies are autosomal recessive; heterozygous individuals are generally fit and well. Identification of heterozygotes or carriers is possible and preferable, by haematological investigations (Cao, Rosatelli & Eckman, 2001). Furthermore, the haematological and biochemical investigations provide essential clues for the different genetic interactions. Family studies are often indispensable; analysis of the haematological parameters of parents, children and other family members provides clues to the segregation of the different genes and allows one to identify the single mutations. Carrier screening and mutation detection comprises a large part of the workload in haemoglobin-

opathy DNA diagnostics. Basic haematological investigations include full blood counts using an electronic counter, Hb analysis by electrophoresis, isoelectric focusing (IEF) and/or high-performance liquid chromatography (HPLC) and quantitation of Hb A₂ and Hb F levels. Supplementary investigations include evaluation of the iron status (serum zinc protoporphyrin, ferritin and transferrin saturation); functional tests for Hb variants such as sickle solubility test, heat stability, demonstration of inclusion bodies and oxygen dissociation studies, globin chain synthesis and intercellular distribution of Hb F.

Haematological phenotype of carriers – a guideline to DNA diagnostics of haemoglobinopathies

The primary screen is based on the full blood counts and Hb electrophoresis and estimation of Hb A₂ and F levels (Figure 2). The secondary screen involves further haematological studies to identify particular properties of any suspected variant such as sickle solubility and heat-instability tests, and biochemical evaluation of the iron status. The definitive diagnosis requires DNA analysis which is guided by the results of the primary and secondary screens. Cut-off values indicating possibility of heterozygosity for thalassaemia include: MCV < 78 fL and MCH < 27 pg. Evaluation of the blood counts in samples >24 h should be made with caution as the red cells increase in size leading to falsely raised MCV values. The majority of heterozygotes for β -thalassaemia, whether it be β^0 or β^+ , have hypochromic (MCH 19–23 pg) and microcytic (MCV 62–75 fL) red cells and an elevated Hb A₂ level in the range of 4.0–5.5%. About half of β -thalassaemia carriers have variably increased levels of Hb F, the majority within the range of 2.5–3.0% (Weatherall & Clegg, 2001). Deviation from the typical haematological phenotype of β -thalassaemia trait include:

- Reduced MCV and MCH with borderline or normal Hb A₂ levels when one should consider iron deficiency, α -thalassaemia, heterozygosity for mild β -thalassaemia mutations, heterozygosity for $\epsilon\gamma\delta\beta$ -thalassaemia.
- Borderline/normal MCV and MCH values with elevated Hb A₂ when one should consider co-inheritance of α and β -thalassaemia.
- Normal or reduced red cell indices with normal Hb A₂ but elevated Hb F values when one should consider heterozygous $\delta\beta$ -thalassaemia or HPFH.

A phenotype that is often encountered is that of borderline/normal Hb A₂ level with borderline reduced or normal MCV and MCH values. In some cases, this is

mutation, targeting the region for identification by other means.

The ultimate method of mutation identification is by direct sequence analysis of specifically amplified DNA. In the last decade, the use of automated sequencers has led to nonradioactive, more robust, and more rapid sequencing, making it suitable as a routine diagnostic tool. Direct sequencing analysis is particularly applicable to the globin genes which are compact and relatively small (1.2–1.6 kb) with the majority of the point mutations within the gene or its flanking sequences.

Southern blotting (Southern, 1975; Sambrook & Russell, 2001) is probably one of the few non-PCR based molecular techniques that still has a significant role to play in the molecular diagnosis of the haemoglobinopathies. It is very useful in the screening for large deletions or rearrangements and is essential in the characterization of novel deletions. The more common deletions can be detected by gap-PCR once the deletion break points have been defined and specific primers flanking the deletion designed.

Each of these techniques has its own limitations; the particular repertoire chosen by a laboratory for molecular diagnosis of the haemoglobinopathies depends on the spectrum of mutations encountered in their catchment area and the technical expertise available in the diagnostic laboratory. It is good practice for any DNA diagnostic laboratory to have at least two alternative methods for detecting each mutation.

Direct sequence analysis

For the β -globin gene, which is relatively small (covering 1.6 kb of genomic DNA including 5' and 3' untranslated region), the majority of mutations can be detected in two sequence reads (Figure 3), whereas for α -globin, each gene ($\alpha 1$ or $\alpha 2$) can be sequenced after specific amplification in a single read (Figure 4). This allows detection and identification of the mutations in a single procedure, as a consequence of which screening techniques such as SSCP and DGGE are not as cost-effective for haemoglobinopathy mutation analysis, as employed for mutation detection in large genes such as dystrophin. The ease with which the β -globin gene can be sequenced is taken advantage of in prenatal diagnosis cases where the mother has sickle trait and the father is unavailable for testing. If the foetus is also found to carry the sickle mutation then the foetal β -globin gene is sequenced to be sure that the fetus does not have SCD or coexisting beta-thalassaemia because of a compound heterozygous state. The only pitfalls of using sequencing as a routine diagnostic

technique are the cost and the time of analysis, which is increased compared with standard PCR assays.

Sequencing is a multistage procedure requiring PCR amplification, cycle sequencing and precipitation before the sequence can be determined. After this the sequence must be analysed and checked and any changes noted. Although sequence analysis software is available it is not 100% efficient at detecting heterozygotes. This is an important consequence for the haemoglobinopathies as most cases involve carrier testing, therefore sequence traces frequently require checking by eye. There are two different sequencing chemistries available based on the Sanger method (Sanger *et al.*, 1973; Sanger, Nicklen & Coulson, 1977), dye primer and dye terminator, they differ from each other in the way the fluorescent label is incorporated during linear cycling. Although the dye terminator chemistry is easier to set-up the signal from each nucleotide is less consistent making the dye primer chemistry more suitable for heterozygote identification and more compatible with sequence analysis software (Figure 3). The dye terminator chemistry would have more application in X-linked diseases such as G6PD mutations in which affected males are hemizygous and will show up as a homozygous change. In all, the process of acquiring a sequence from whole blood and the analysis may take 4–5 days, with confirmation using another PCR-based test before the mutation is reported.

Allele-specific priming – ARMS-PCR

Primer-specific amplification is based on the principle that a perfectly matched primer is much more efficient at annealing and directing primer extension than a mismatched primer. The most widely used method is the ARMS (Newton *et al.*, 1989; Old, Varawalla & Weatherall, 1990) in which allele-specific amplification relies on the specificity of the 3' terminal nucleotide. To enhance allele specificity, it is common practice to deliberately incorporate a second mismatch at position -2 or -3 from the 3' end. The target DNA is amplified in two separate reactions using a common forward primer and either one of two reverse allele-specific primers, one complementary to the mutant sequence, the other to the normal DNA sequence. Presence of the mutant allele will generate a PCR product in the tube containing the mutation-specific primer, and vice versa. To monitor false-negative results because of failure of the amplification reaction itself, an internal PCR control which amplifies another region of the genome should be included in the reaction (Figure 5). ARMS-PCR has the advantage that it is theoretically possible to detect any known mutation (Newton *et al.*, 1989). Recently the

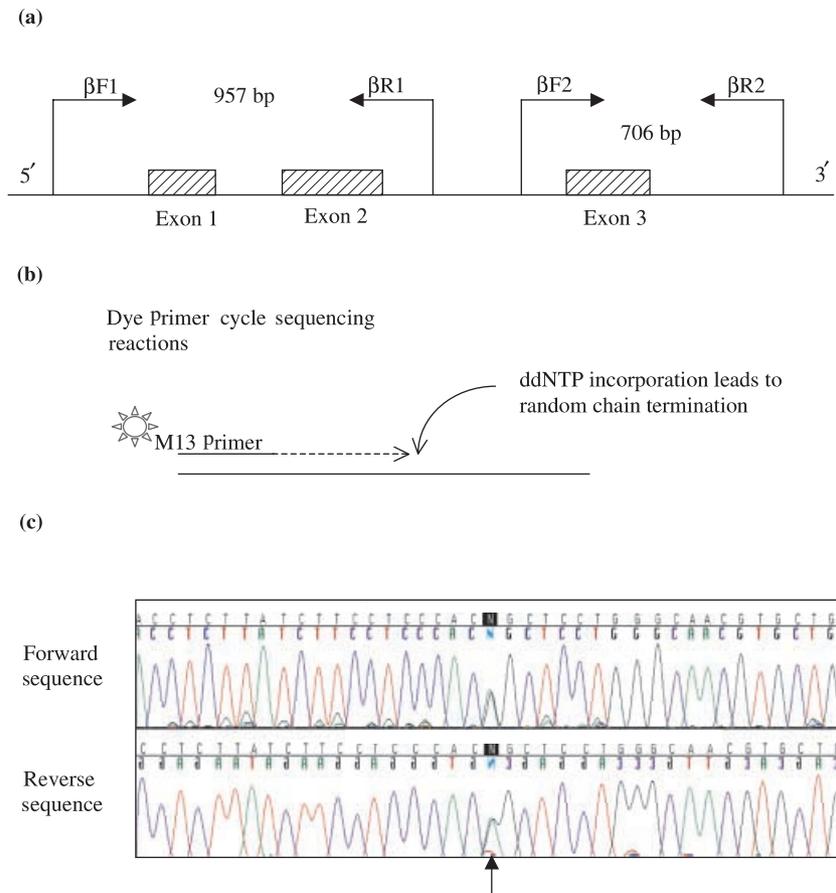


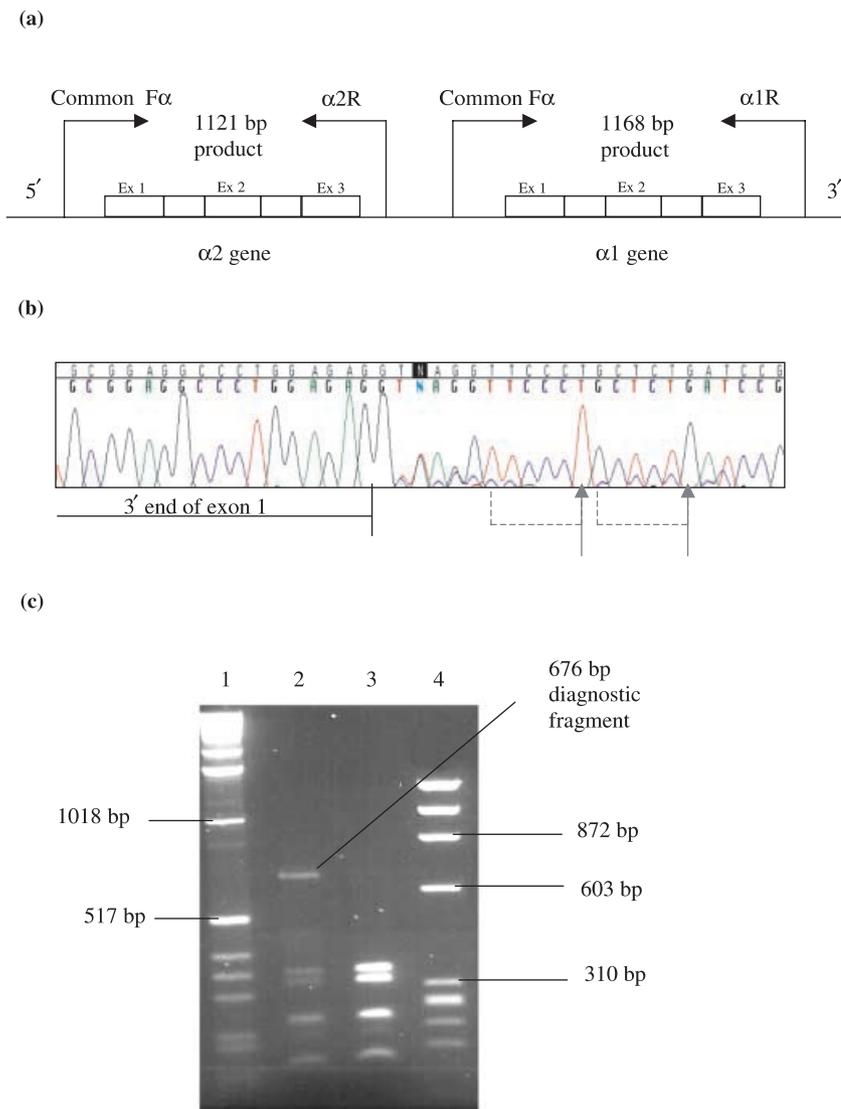
Figure 3. Direct sequence analysis of the β -globin gene using the dye primer chemistry. (a) The majority of the mutations causing β -thalassaemia are concentrated in two regions – (i) a region including the 5' promoter and 5' untranslated region (UTR), exon 1, intron 1, exon 2 and the flanking intron 2 sequences and (ii) 3' intron 2 sequences, exon 3 and the 3' UTR – and should be detectable in two sequence reads. Genomic structure of the β globin gene – exons are shown as striped boxes and introns as open boxes, and the UTRs as zigzag boxes. Primer locations are represented by horizontal arrows, the sequences of which are as follows but with 5' M13 tags, β F1-5'-CGATCTTCAATATGCTTACCAA-3', β R1-5'-CATTTCGTCTGTTTCCCATTCTA-3', β F2-5'-CAATGTATCATGCCTCTTTGCA-3', β R2-5'-TGCAGCCTCACCTTCTTTCAT-3'. The double-stranded PCR products are then used as templates in the cycle sequencing reactions (b) Dye primer sequencing using M13 fluorescently labelled primers to initiate elongation in a reaction mix containing one of the four dideoxynucleotides. Therefore four reactions are required per sequence which are pooled and analysed simultaneously using a capillary array and a laser to detect the fluorescent signal. (c) A sequence trace generated from the dye primer chemistry. The arrow indicates the location of a heterozygote where two signals are read at the same location with half the signal intensity of a homozygous signal. Above the chromatogram the computer reads the heterozygote as N, making it quick and easy to locate.

methodology has been improved by development of a single tube assay (Ye *et al.*, 2001), where both the mutant and the wild type alleles are detected simultaneously in the same reaction with an internal positive control, referred to as tetra primer ARMS-PCR. As the name implies, tetra ARMS-PCR uses two pairs of primers, one pair of flanking primers and one pair of internal primers that are complementary to different strands, and amplify the two different bases that occur in a single position. The different alleles can be distinguished on an agarose gel by designing the primers so that the mutant and wild type products are of different sizes. To aid with the design of

the primers for the Tetra ARMS-PCR methodology a computer programme has been developed and is freely available from the website at http://cedar.genetics.soton.ac.uk/public_html/primer1.html. Tetra ARMS-PCR has proved to be very effective for genotyping in our experience.

A single-tube allele-specific PCR can also be carried out in which the normal and mutant ARMS primers compete with each other to amplify the target. Allele-specific priming is detected by differential fluorescent labelling (Chehab & Kan, 1989; Chehab & Kan, 1990). A variation of this technique is to use ARMS primers that differ in

Figure 4. Detection of the $\alpha 2$ IVS1–5 bp del (α^{Hph}) thalassaemia variant. (a) Line diagram illustrating the $\alpha 2$ and $\alpha 1$ globin genes and positions of the amplification primers. The alpha globin genes are amplified using M13 tagged primers; common F (forward) primer (5'-GGAG GGTGGAGACGTCCTG-3') is used; $\alpha 2$ or $\alpha 1$ globin genes are amplified using the appropriate reverse primer, $\alpha 2$ R primer (5'-CCATTGTTGGCACATTCCGG-3') or the $\alpha 1$ R primer (5'-CGAGAGGTTCT AGCCATGTGTG-3'), respectively. (b) Sequence analysis of the $\alpha 2$ -globin gene at the exon 1/intron 1 junction. One base (G) after the end of exon 1 the sequencing peaks of the chromatogram are reduced in height with double peaks in some locations. This picture is typical of a frameshift due either to an insertion or deletion. Comparison of the frameshifted sequence with that of the wild type indicates a 5 bp deletion from IVS1-1 position ($\alpha 2$ IVS1-5 bp). The mutation is confirmed by *Hph* I restriction analysis. (c) Confirmation of the $\alpha 2$ IVS1-5 bp deletion by *Hph* I restriction analysis. The deletion destroys the *Hph* I site resulting in a larger restriction fragment of 676 bp in the affected individual. Lanes 1 and 4 are DNA size markers (markers X and IX respectively, Roche), lane 2 is a heterozygous individual and lane 3 is a wild type control.



length. This technique has been applied to the diagnosis of β -thalassaemia mutations in Taiwan and has been referred to as mutagenically-separated PCR (MS-PCR) (Chang *et al.*, 1995). More than one mutation can also be screened for at the same time in a single reaction by multiplexing the ARMS primers coupled with a common primer (Tan *et al.*, 1994). Multiplex ARMS requires stringent optimization of primer annealing conditions and primer concentrations.

Gap-PCR

PCR primer pairs are designed to flank a known deletion generating a unique amplicon that will be smaller in the mutant sequence compared with the wild type (Faa *et al.*, 1992). For small deletions such as the 619 bp deletion, a common cause of β -thalassaemia in Asian Indians,

differential amplification products are generated in the mutant and wild type. When the deletion is vast (>2 kb) it is technically difficult to generate a product in the wild type. As a control, a primer can be included that anneals within the deleted sequence to generate an additional product with one of the other primers flanking the deletion. In this way, false negatives are monitored in the wild type and indicates if an individual is heterozygous for the deletion. This principle has been used in the technique of gap-PCR for the common HPFH (Figure 6) and $\delta\beta$ -thalassaemias (Craig *et al.*, 1994). Unlike the β -thalassaemias, deletions are a common cause of α -thalassaemia. The common α -thalassaemia deletions and rearrangements can be routinely detected using gap-PCR (Chong *et al.*, 2000; Liu *et al.*, 2000; Tan *et al.*, 2001) (Figure 7). Gap-PCR is also routinely used to detect Hb Lepore, a variant created by deletion of 7 kb and

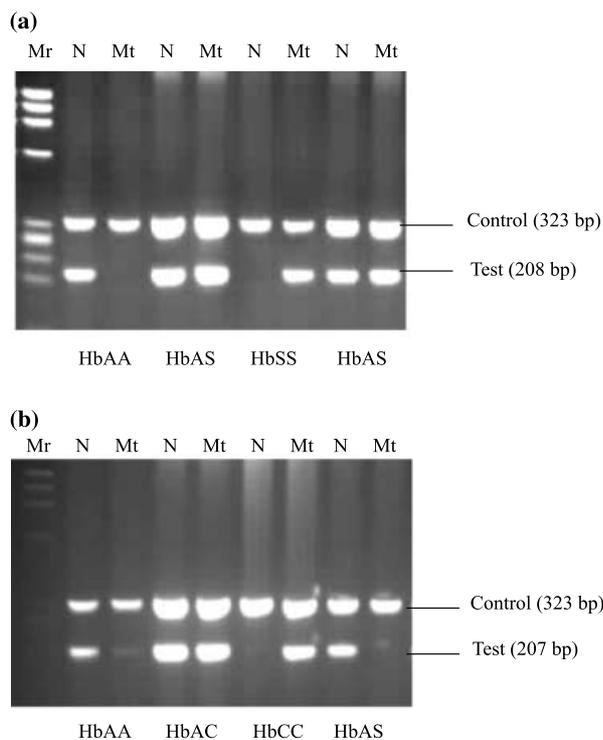


Figure 5. ARMS-PCR analysis for Hb S (a) and Hb C (b) mutations. Conventional ARMS-PCR is performed in two separate reactions using a common primer with a primer specific either for the wild type (N) or mutant (Mt) sequence. All PCRs are monitored by a control reaction (C) which amplifies a different part of the genome. Sequences of the primers are published in Weatherall and Clegg (2001). Lanes labelled Mr are DNA size markers (Marker IX, Roche).

results in a functional hybrid delta/beta globin product (Weatherall & Clegg, 2001).

Restriction enzyme analysis of PCR product

Restriction enzymes cut double-stranded DNA at specific recognition sequences; some mutations naturally create or abolish restriction enzyme sites. Computer programs are now available that allow sequences to be screened for putative recognition sites for restriction enzymes. Genomic DNA containing the mutation, the 'target', is amplified by PCR and the product is then digested by the diagnostic restriction enzyme and the resulting DNA fragments separated on gels. The presence or absence of the recognition site is determined from the pattern of the PCR digest (see Figure 8 for detection of Hb S using *Dde* I), hence an alternative term for this technique is restriction fragment length polymorphism (RFLP) analysis. Restriction enzyme analysis is simple, relatively cheap and robust leading to unequivocal results; the technique is an invaluable molecular diagnostic tool. However, it is limited

in its application as only a proportion of the haemoglobin variants, β -thalassaemia and α -thalassaemia mutations, naturally create or abolish restriction enzyme cutting sites (Pirastu, Ristaldi & Cao, 1989; Weatherall & Clegg, 2001). Incomplete or partial digests can be a problem for some restriction enzymes leading to false negative or positives – thus positive and negative controls should always be included. Restriction endonucleases are usually used when there is some other evidence suggesting the causative mutation, for example HPLC, Hb electrophoresis analysis and ethnic origin of the individual, or it is used as a secondary confirmation technique i.e. after sequence analysis. In some instances, although the mutations do not naturally create or abolish a restriction enzyme site, the sequence in the vicinity of the mutation allows one to artificially create restriction sites adjacent to the mutation sequence by introducing a single base mismatch in the amplification primer. This technique, referred to as amplification-created restriction site (ACRS) is routinely used to identify the Saudi nondeletional alpha thalassaemia mutation in the $\alpha 2$ -globin gene, AATAAA to AATAAG (Jassim *et al.*, 1999).

Allele-specific oligonucleotide (ASO) hybridization and reverse dot-blot

The method is based on the principle of specific hybridization of two oligonucleotide probes, one complementary to the mutant sequence and the other to the normal sequence (Wallace *et al.*, 1981; Conner *et al.*, 1983). The ASO's differ from each other by a single base change designed to be in the centre of the ASO to maximize instability of any mismatch. Genomic DNA is specifically amplified using specific primers encompassing the mutation and the PCR product bound to a nylon membrane in the form of dots (Saiki *et al.*, 1988). The ASO's are 5'-end-labelled with 32 P-dATP, biotin or horseradish peroxidase and the PCR dot sequentially hybridized with the mutant and wild type ASO's. The genotype of the DNA sample is read by the presence or absence of hybridization signal from the mutant and wild type probes. The technique is reliable and has been used with great success specifically in populations where there are one or two predominant mutations (Ristaldi *et al.*, 1989). However, the method becomes limited when screening for multiple mutations because of the repeated hybridization and washing steps. To overcome this, a method of reverse dot-blotting has been developed in which pairs of cold mutant and wild type ASO's are fixed as dots or slots to nylon membrane strips (Saiki *et al.*, 1989). Amplified genomic DNA, encompassing the region containing the putative

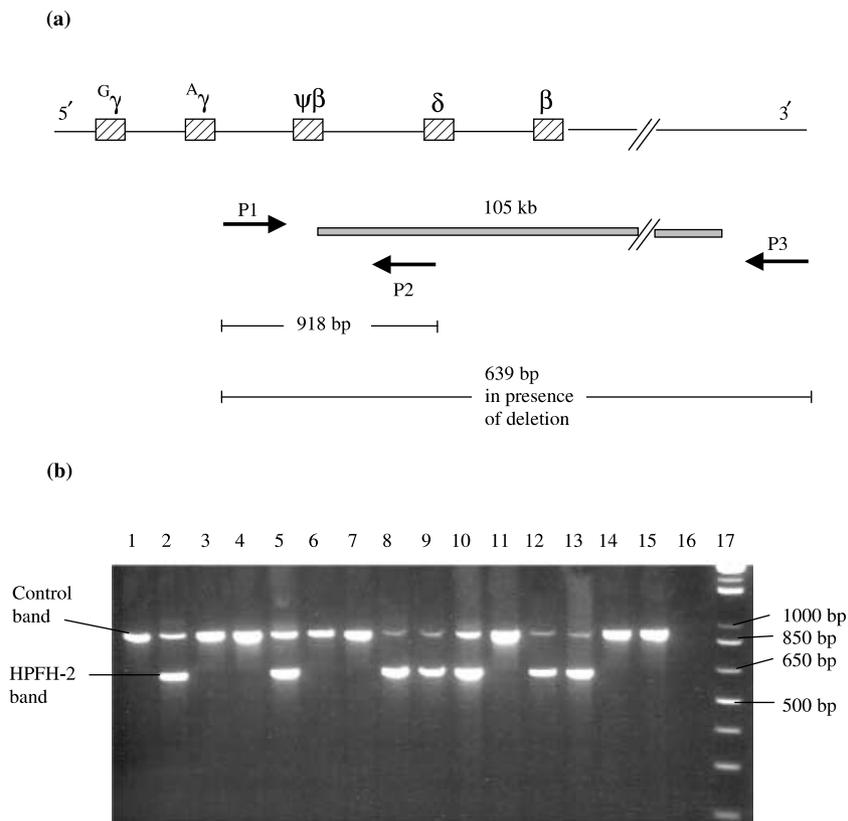


Figure 6. Gap-PCR for the African HPFH-2 deletion. (a) Diagram showing part of the beta globin locus containing the β -like globin genes. The striped boxes indicate genes separated by genomic DNA (black line). The location of the HPFH-2 deletion is indicated by the shaded box and primers are represented as horizontal arrows. Primer sequences are as follows, P1-5'-TGCTCTGTGATTATGACTATCCC-3', P2-5'-TATCCTTCTTACTTGTGATGCC-3', and P3-5'-ATGCTACTGATTTTGTATTTTATTTC-3'. The deletion is detected by the presence of the 639 bp product while the normal allele is identified by the presence of the 918 bp product (adapted from Craig *et al.*, 1994). (b) Electrophoresis of HPFH-2 PCR products. Lanes 1–11 are patient samples, lanes 12 and 13 are heterozygous controls and 14 and 15 are normals, and lane 17, the DNA size marker (1 kb ladder, Invitrogen). Individuals in lanes 1, 3, 4, 6, 7, and 11 do not carry the deletion, while individuals in lanes 2, 5, 8, 9 and 10 are heterozygotes as in the controls in lanes 12 and 13. Lane 16 is the blank control.

mutations, is labelled by use of end-label primers or the internal incorporation of biotinylated dUTP and then hybridized to the strips. The genotype of the DNA sample is then diagnosed from the presence or absence of hybridization signals to the panel of ASOs. The reverse dot-blot allows the simultaneous detection of multiple β -thalassaemia mutations in Mediterraneans (Maggio *et al.*, 1993), Thais (Sutcharitchan *et al.*, 1995a) and African-Americans (Sutcharitchan *et al.*, 1995b), but this requires careful optimization of the hybridisation efficiency of the different ASOs. Commercial kits based on reverse dot-blotting for some of the common mutations are available – one utilizes a nylon strip with ASOs for the common Mediterranean β -thalassaemia mutations (Vienna Lab, Vienna, Austria), while the other utilises ASOs immobilised on the walls of a microtitre plate developed by BioRad (Hercules, CA, USA). Both have limited applications as the

panel detects a limited number of mutations and the kits are expensive and demonstrate batch-to-batch variation. Hence, ASO and reverse dot-blotting have been largely superseded by other PCR-based techniques such as ARMS-PCR and direct sequencing.

DGGE and SSCP

Denaturing gradient gel electrophoresis and single-stranded conformation polymorphism (SSCP) (Orita *et al.*, 1989) are two techniques that have been used to screen for unknown mutations within a PCR product. Both techniques rely on the mutant sequence altering the mobility of the PCR product on a gel compared with the wild type, and require further analysis to resolve the causative mutation, usually sequencing. Single-base substitutions can be detected by both methods but the detection

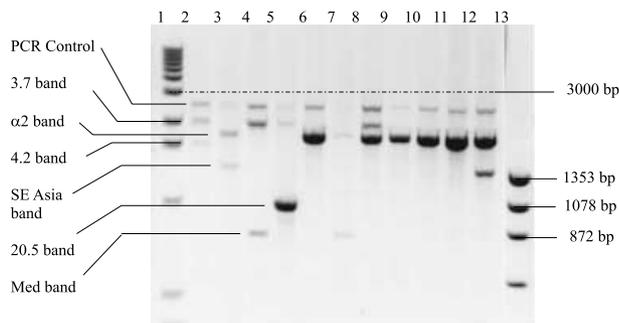


Figure 7. Multiplex gap-PCR for detection of the common α -thalassaemia deletions. The primer sequences are as published in Chong *et al.* (2000). In the alpha multiplex an internal PCR control is included which amplifies another region of the genome to monitor false negatives. Lanes 1 and 13 are DNA size markers, markers X and IX respectively; lane 2 is a control for the 3.7 kb and the 4.2 kb deletions ($-\alpha^{3.7}/-\alpha^{4.2}$), lane 3 is a control for South-east Asia deletion ($-\alpha^{SEA}/\alpha\alpha$), lane 4 is a compound heterozygous control for the Mediterranean and the 3.7 kb deletion ($-\alpha^{MED}/-\alpha^{3.7}$), lane 5 is a compound heterozygous control for 20.5 kb and the 3.7 kb deletions ($-\alpha^{20.5}/-\alpha^{3.7}$), lane 6 is the wild type control ($\alpha\alpha/\alpha\alpha$), lane 7 is a heterozygous control for the Mediterranean deletion ($-\alpha^{MED}/\alpha\alpha$) and lane 8 is a control for the 3.7 kb deletion ($-\alpha^{3.7}/\alpha\alpha$). Lanes 9–11 are individuals found not to contain any of the common deletions while lane 12 is an individual found to be heterozygous for the South-east Asian deletion. In view of the numerous controls needed for the multiplex, it is efficient to perform these multiplex reactions for α -thalassaemia deletions in batches in microtitre plates.

efficiency is between 70 and 90%, with SSCP requiring various gel conditions to achieve this higher range. These mutation screening approaches have application for genes that have many exons spread over large genomic regions with a uniform distribution of mutations because the methodology allows multiplexing and three to five PCR products can be run in a single lane.

DGGE has been useful in some laboratories for large-scale screening of β -thalassaemia mutations (Losekoot *et al.*, 1990). The technique involves the electrophoresis of double-stranded DNA fragments through a linearly increasing denaturing gradient until the lowest melting temperature domain of the fragment denatures creating a heteroduplex that becomes stationary in the gel matrix. The β -globin gene is amplified in segments using five to seven pairs of primers, a GC-rich sequence is added to one of each pair to create a high melting domain thus increasing the differential mobility of the fragments. Clues to the nature and location of the mutations are provided by characteristic patterns of the heteroduplexes. These screening techniques simply pin-point the presence of a mutation which has to be definitively identified by another PCR-based technique, usually direct sequence analysis.

The set-up requires optimisation, and interpretation of the heteroduplex patterns can be difficult without experience.

Diagnostic strategy

The spectrum and frequency of the different haemoglobinopathy mutations within the populations in which the haemoglobinopathies are prevalent, can vary substantially. Positive selection by malaria for the sickle mutation and α - and β -thalassaemias has led to an increased frequency of some of these mutations in the malarious regions, such that each population has its own combination of a limited number of common Hb variants and thalassaemia mutations together with a larger number of rarer ones. Prior information on the ethnic origin of the patient is therefore an important part of the diagnostic strategy, enabling a quick identification of the mutations most likely to be present. However, this strategy becomes less efficient and reliable in populations where there is a large ethnic mix, such as in the UK and among the Afro-Caribbean group.

β -thalassaemia syndromes, including $\delta\beta$ -thalassaemia and HPFH

About 200 mutations account for β -thalassaemia worldwide but only about 30 of these reach a frequency of 1% or more in the at-risk groups (Flint *et al.*, 1998). With the exception of the 619 bp deletion, which accounts for 20% of the β -thalassaemias in Asian Indians (Varawalla *et al.*, 1991), the common mutations are all point mutations because of single base substitutions, insertion or deletion of a few bases (Thein, 1998; Weatherall & Clegg, 2001). These mutations are all found within the β -globin gene itself or its immediate 5' and 3' flanking sequences and all are amenable to detection by PCR-based techniques including direct sequence analysis.

The strategy for identifying β -thalassaemia mutations in most diagnostic laboratories is to screen for the most common ones, based on the ethnic origin of the individual, by using a PCR-based technique. The most commonly used screening techniques for known mutations are allele-specific priming (ARMS) and the reverse dot-blot analysis. RFLP analysis is useful for a limited number of mutations. This approach should identify the mutations in about 90% of the cases. Laboratories which have access to one of the screening set-ups such as DGGE or SSCP may resort to these approaches to localize the putative mutation within the β -globin gene either as an initial screening approach or after failure to define the mutation by one of the direct approaches. This is then followed by direct sequence

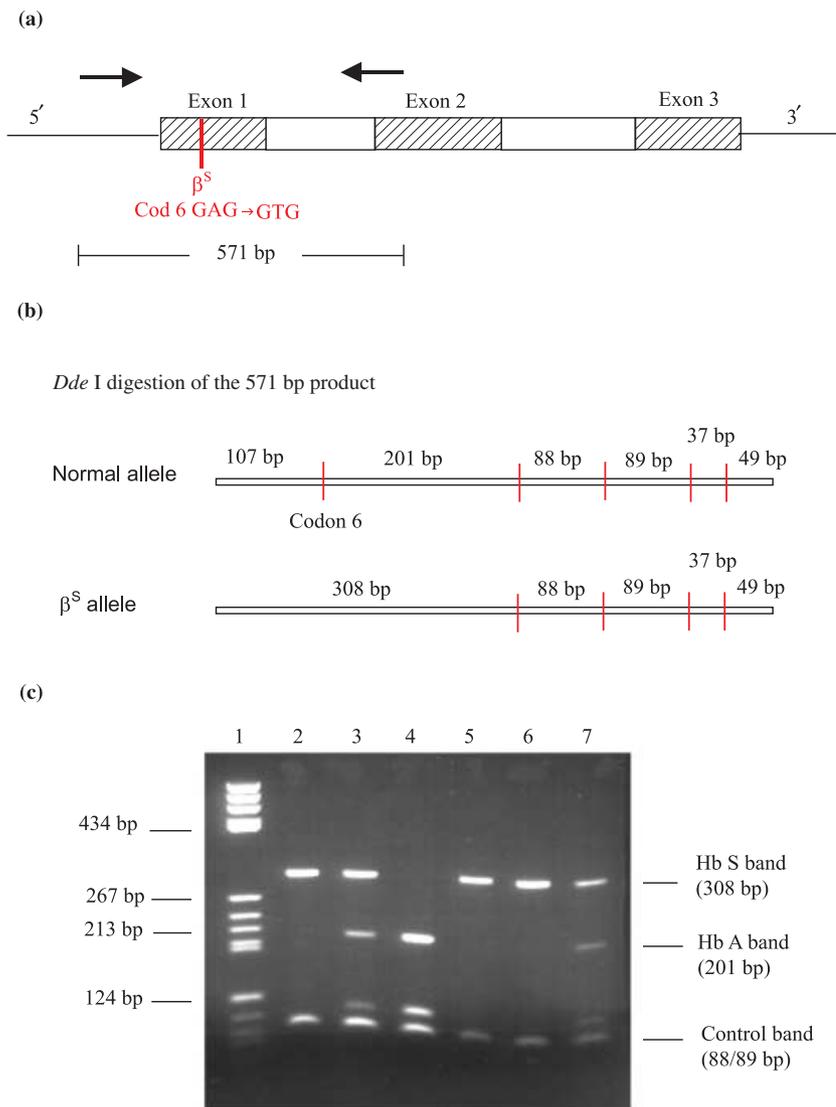


Figure 8. Detection of β^S mutation by *Dde* I restriction analysis. (a) Line diagram of the β globin gene showing locations of the PCR primers (horizontal arrows) and the β^S mutation (in red) in codon 6, exon 1. Exons are shown in striped boxes and introns in open boxes. (b) Several *Dde* I restriction sites are found in this PCR amplicon, the *Dde* I site in codon 6 is removed by the β^S mutation, creating the largest restriction fragment of 308 bp. The additional and constant *Dde* I sites act as internal control for complete *Dde* I digestion. (c) Electrophoresis of the *Dde* I restriction digest. Lane 1 is the DNA size marker, marker V Roche; lane 2 is a homozygous affected control (Hb SS), lane 3 is the heterozygous control (Hb AS) and lane 4, the wild type control (Hb AA). Lanes 5, 6 and 7, are patient samples found to be Hb SS, Hb SS and Hb AS, respectively.

analysis of PCR-amplified β -gene fragment. In practice, most DNA diagnostic laboratories now have access to automated DNA sequencing facilities, and it is more efficient to proceed to direct sequence analysis of the β -globin gene, after excluding one of the common mutations by ARMS-PCR, restriction enzyme analysis or reverse dot blot. Arguably, in populations where there is a large ethnic mix, direct sequence analysis may be more cost-effective and efficient as a first-line approach as the vast majority of the mutations can be defined in just two sequence reads of the β -globin gene (see Figure 3).

Hb E/ β -thalassaemia is an important cause of β -thalassaemia major and thalassaemia intermedia in South-east Asians and Asian Indians. The Hb E mutation (β^{26} GAG \rightarrow AAG, Glu \rightarrow Lys) removes an *Mnl* I site (Thein *et al.*, 1987) and thus may be identified by restriction enzyme analysis of specifically amplified DNA. The β^E

mutation is also easily detectable by ARMS-PCR or ASO hybridization. Some laboratories proceed to direct sequence analysis of the Hb E/ β -thalassaemia as this approach would confirm the β^E and define the β -thalassaemia mutation.

Apart from the 619 bp deletion mutation, β -thalassaemia is rarely caused by deletions. The 619 bp deletion removes the 3' part of intron 2 and exon 3 but leaves the 5' end of the β -gene intact. It is routinely detected by the technique of gap-PCR generating differential size products. The other deletions which range from 290 bp to approximately 65 kb remove, in common, a region of the 5' β promoter. These deletions are not common, a clue to their presence is the unusually high levels of A_2 ($>6.5\%$) in heterozygotes (Thein, 1998). These deletions are also amenable to detection by gap-PCR but it may be more efficient to screen for their presence by Southern blot

hybridization using a panel of restriction enzymes and the β -globin gene probe.

Gap-PCR is also used to identify Hb Lepore, $\delta\beta$ -thalassaemia and the HPFH deletion mutations. However, gap-PCR is only possible for those deletions in which the breakpoints have been characterized. Primers have been designed for six common $\delta\beta$ -thalassaemias (Chinese, Sicilian, Spanish, Vietnamese/SE Asian, and the Turkish and Indian inversion/deletion mutations) and three common HPFH deletions – African HPFH-1 and 2, and the Indian HPFH-3 (Craig *et al.*, 1993a; Craig *et al.*, 1994) (See Figure 6 as an example of the gap-PCR technique for detection of the HPFH-2 deletion). Southern blot hybridization is a useful technique to screen for the less-common deletions in the absence of which a nondeletion mutation should be suspected.

The nondeletion HPFH mutations are all clustered in three regions of the $G\gamma$ or $A\gamma$ globin gene promoter (Wood, 2001). The most efficient and cost-effective approach to identify these mutations is direct sequence analysis of specifically amplified $G\gamma$ - or $A\gamma$ -globin promoter regions (Craig *et al.*, 1993b). Some of these nondeletion HPFH mutations alter restriction enzyme cleavage sites.

α -thalassaemia

In contrast to β -thalassaemia, in which mutations are predominantly caused by point mutations in the structural genes, α -thalassaemia is more often caused by deletions involving one or both of the α -globin genes (Higgs, 2001). A small number of point mutations, usually within the $\alpha 2$ gene, have been characterized. Rare deletions removing the upstream α -globin regulatory element have also been described.

α^0 -thalassaemia, because of the loss of both $\alpha 1$ and $\alpha 2$ genes on the same chromosome ($/-$), is found in the Mediterranean ($/-$ $^{-MED}$ and $/-(\alpha)^{20.5}$) and in South-east Asia ($/-$ $^{-SEA}$, $/-$ $^{-FIL}$ and $/-$ $^{-THAI}$). Although α^0 -thalassaemia has also been described in individuals of Asian Indian and South African origin, it is extremely uncommon; the phenotype of two α gene deletions is usually because of a homozygous deletion of single α globin genes, i.e. $-\alpha/-\alpha$. α^+ -thalassaemia ($/-\alpha$) is prevalent in parts of Africa and Asia; the $/-\alpha^{3.7}$ predominantly found in Africa, the Mediterranean and Asia while the $/-\alpha^{4.2}$ deletion is found in South-east Asia and the Pacific Islands.

The gap-PCR technique is extremely useful in the identification of these common α^0 and α^+ -thalassaemias (Figure 7). The primers for these deletions can be multiplexed in various combinations to capture the deletions that are most likely to be encountered in the different

catchment areas (Chong *et al.*, 2000; Tan *et al.*, 2001). The multiplex set encompassing gap-PCR primers for $/-$ $^{-SEA}$, $/-$ $^{-MED}$, $/-$ $-(\alpha)^{20.5}$, $/-$ $^{-FIL}$, $/-\alpha^{3.7}$ and $/-\alpha^{4.2}$ is used by many laboratories supplemented by another set for identification of triple α complex ($/\alpha\alpha\alpha$) (Dodé *et al.*, 1992; Liu *et al.*, 2000) useful in the work-up of thalassaemia intermedia. Gap-PCR primers are also available for the Thai α^0 deletion ($/-$ $^{-THAI}$) and can be included in the multiplex panels (Tan *et al.*, 2001). Multiplex gap-PCR requires careful optimization but once optimized, is highly efficient and rapid for the detection of these common α^0 and α^+ -thalassaemia deletions, particularly when performed in large batches in microtitre plates.

α -thalassaemia is encountered occasionally in British individuals of Anglo-Saxon origin; in most cases this is caused by the British α^0 -thalassaemia deletion ($/-$ $^{-BRIT}$). The breakpoints of the $/-$ $^{-BRIT}$ deletion have not been characterized. Southern blot analysis using a combination of *Bam*H I and *Bgl* II digestion and hybridization to α -globin and ζ -globin probes is used to identify the $/-$ $^{-BRIT}$ and the less common α^0 and α^+ -thalassaemia deletions (Weatherall & Clegg, 2001).

α^+ -thalassaemia may also be caused by a point mutation in one of the α -globin genes although those of clinical significance usually affect the dominant $\alpha 2$ -globin gene. Although the $\alpha 2$ and $\alpha 1$ -globin genes are highly homologous, there is sufficient sequence difference in their 3' UTRs to permit design of primers for differential amplification of the α -globin genes (Dodé, Rochette & Krishnamoorthy, 1990). Again, these nondeletion α -thalassaemia variants show regional specificity; for example, the $\alpha 2$ poly (A) signal mutation ($\alpha^T SAUDI/\alpha$) is specific for the Middle East and Mediterranean, Hb Constant Spring (α^{CS}/α) in South-east Asia, $\alpha 2$ IVS1-5 bp del (α^{Hph}/α) in Mediterranean, and Hb Quong Sze (α^{QS}/α) in South-east Asia (Higgs, 2001). Several of the common nondeletion mutations alter a restriction enzyme site and may be identified by restriction enzyme analysis of specifically amplified α -globin product, for example, the α^{Hph}/α mutation (*Hph* I) (see Figure 4) and Hb Constant Spring (*Mse* I). For mutations that are not amenable to restriction enzyme analysis, but are common enough, some DNA laboratories have invested in the technique of ACRS or utilized ASO hybridization on specifically amplified α -globin genes.

In theory, any of the PCR-based techniques are applicable for the diagnosis of the nondeletion α -thalassaemia variants but no simple strategy equivalent to that for diagnosing point β -thalassaemia mutations has been developed for the following reasons: nondeletion α -thalassaemia is relatively uncommon; any approaches will have to differentiate the $\alpha 2$ and $\alpha 1$ gene and specific

amplification and screening strategies (DGGE and SSCP) are complicated by the GC-richness of these gene sequences. Thus, most DNA laboratories opt for direct sequence analysis of specifically amplified $\alpha 2$ or $\alpha 1$ globin genes, either as a first line detection strategy or after failure to identify the mutation by restriction enzyme analysis.

Sickle cell disease

Seventy per cent of SCD is caused by homozygosity for the β^S mutation (β codon 6 GAG \rightarrow GTG, Glu \rightarrow Val) and 29% by compound heterozygosity for the β^S and β^C (β codon 6 GAG \rightarrow AAG, Glu \rightarrow Lys). Compound heterozygosity for Hb S and β -thalassaemia accounts for <1% followed by the rarer genotypes of Hb S/Hb D Punjab ($\beta 121$ GAA \rightarrow CAA, Glu \rightarrow Gln), Hb S/Hb Lepore and Hb S/O Arab ($\beta 121$ GAA \rightarrow AAA, Glu \rightarrow Lys). All of these variants are highly amenable to direct detection by PCR-based techniques.

The β^S mutation removes cleavage sites for *Mnl* I, *Mst* II and *Dde* I. *Dde* I restriction analysis is the favoured restriction enzyme as it is a frequent cutter and primers can be designed to include constant sites in the amplicon to act as controls for complete digestion (Figure 8). A note of caution – two other β -thalassaemia mutations [codon 6 (-C) and codon 5 (-CT)] also remove the *Dde* I cutting site in this region, which may produce a false-positive result of β^S homozygosity in Hb S/ β -thalassaemia. Although the β^C mutation occurs in codon 6 it does not affect the *Dde* I or *Mst* II cutting sites in this vicinity; identification of β^C is usually by ARMS-PCR (Figure 5) or dot blot analysis. The Hb D Punjab mutation destroys an *Apo* I site and an *EcoR* I site in this region. Although Hb O Arab mutation occurs in the same codon, the GAA \rightarrow AAA mutation does not affect the *Apo* I cleavage site but it does remove the *EcoR* I cutting site. All of these Hb variants can also be identified by alternative PCR-based techniques such as ARMS-PCR, ASO hybridization and direct sequence analysis.

Other Hb variants

The majority of Hb variants are clinically benign and have been discovered in population surveys using electrophoretic analysis of haemoglobin. Because only variants that alter the charge of the Hb molecule are detectable in routine Hb electrophoresis, the current number of approximately 800 Hb variants is probably an underestimate (Huisman, Carver & Efremov, 1998). The most common Hb variants are those that cause SCD (Hb S and Hb C) and Hb E. Clinical disorders caused by Hb variants are shown in Table 1.

Clues to the presence of these structural Hb variants are often provided by the clinical and family history and haematology with evidence of peripheral haemolysis, etc. Examination of the peripheral blood smear is usually instructive. Definitive identification by DNA analysis is guided by supplementary investigations such as demonstration of instability of the Hb by presence of flocculated precipitates on heating a dilute solution of Hb to 50 °C or by the addition of isopropanol. Oxygen dissociation studies to confirm Hb variants with altered oxygen affinity require more technical expertise. Unlike the thalassaemias, Hb variants with altered oxygen affinity, the Hbs M and the majority of the unstable Hb variants follow an autosomal dominant pattern of inheritance; affected individuals are almost exclusively heterozygotes. Further, in many families, only the proband is affected suggesting a *de novo* mutation.

All of these variants should be identifiable by specific amplification of the suspected globin gene by direct sequence analysis. Some of the variants alter cleavage sites of restriction enzyme; and thus can be confirmed by restriction analysis of PCR product.

Table 2 lists the clinically significant mutations and indicates the commonly used techniques of detection. The sequence of the PCR primers and the cycling parameters that are used in the analysis are well documented (Weatherall & Clegg, 2001; Old, 2003).

Table 1. Clinical disorders caused by Hb variants

1.	Sickle syndrome – Hb S and interaction of Hb S with other Hb variants (Hb S/S, S/C, S/D Punjab, S/O Arab, S/Lepore) or β -thalassaemia Hb S/ β thal
2.	Chronic haemolysis (congenital Heinz body anaemia, CHBA) e.g. Hb Köln ($\beta 98$ GTG \rightarrow ATG, Val \rightarrow Met), Hb Hammersmith ($\beta 42$ TTT \rightarrow TCT, Phe \rightarrow Ser)
3.	Congenital polycythaemia – high oxygen affinity Hb variants, e.g. Hb Olympia ($\beta 20$ GTG \rightarrow ATG, Val \rightarrow Met)
4.	Congenital cyanosis – low oxygen affinity Hb variants and Hbs M e.g. Hb M-Iwate ($\alpha 87$ CAC \rightarrow TAC, His \rightarrow Tyr)
5.	Hypochromic microcytic anaemia (thalassaemic haemoglobinopathy) e.g. Hb E, Hb Lepore, and hyperunstable Hbs e.g. Hb Geneva ($\beta 114$; – CT, +G)
6.	Drug-induced haemolysis, e.g. Hb Zurich ($\beta 63$ CAT \rightarrow CGT, His \rightarrow Arg)

Table 2. Commonly used methods for the identification of clinically significant haemoglobinopathies

Haemoglobinopathy	Mutation type	Techniques available
β -thalassaemia	Point mutation	ARMS-PCR ASO Restriction digest Direct sequence analysis
α -thalassaemia	Deletions	Gap-PCR
	Deletions	Gap-PCR Southern blotting
	Nondeletion	On specifically amplified $\alpha 1$ or $\alpha 2$ genes Restriction digest ARMS-PCR ASO Direct sequence analysis
Hb S variant	Point mutation	ARMS-PCR Restriction digest Direct sequence analysis
Hb C variant	Point mutation	ARMS-PCR Direct sequence analysis
Hb E variant	Point mutation	ARMS-PCR Restriction digest Direct sequence analysis
Hb D ^{Punjab} variant	Point mutation	ARMS-PCR Restriction digest Direct sequence analysis
Hb O ^{Arab} variant	Point mutation	ARMS-PCR Restriction digest Direct sequence analysis
Hb lepore variant	Deletion	Gap-PCR
HPFH and $\delta\beta$ -thalassaemia	Deletions	Gap-PCR Southern blotting
$\epsilon\gamma\delta\beta$ -thalassaemia	Deletions	Southern blotting

Foetal DNA analysis

Many laboratories offering DNA diagnostics of the Hb disorders are also involved in analysis of foetal DNA for the prenatal diagnosis of these disorders. Foetal DNA is usually, and preferably, obtained through chorionic villus sampling in the first trimester of pregnancy (10–12 weeks) (Old, 1986). Chorionic villus sampling provides a good yield of DNA which is isolated using conventional methods of phenol chloroform extraction after careful microscopic dissection to remove any contaminating maternal decidua. Occasionally, if the sample is too small, it may need to be cultured increasing the risk of maternal contamination. Amniocytes obtained in second trimester can also be used as a source of foetal DNA but the foetal cells are often contaminated with maternal cells and the results have

to be interpreted with caution after analysis for maternal contamination.

Every prenatal diagnosis should be accompanied by copies of haematology results of the parents and prior confirmation of the parental phenotypes and genotypes. PCR-based techniques that best suit the expertise of the laboratory are then used to screen for the presence of the parental mutations in the foetal DNA. Parental, appropriate positive and negative controls must always be included in the investigations. A limited number of PCR cycles (25–28) should be performed to avoid amplification of any minor DNA species and to minimize amplification of any contaminating maternal DNA. As an added precaution, maternal DNA contamination should be checked for by using polymorphic DNA markers including the variable tandem repeats (VNTRs) such as ApoB, and the short tandem repeats or micro satellites (Decorte *et al.*, 1990). This is particularly important when the foetal genotype is the same as the maternal genotype. Fetal DNA analysis should also be performed in duplicate and confirmed by an independent PCR-based technique.

In cases where the father is unavailable for testing, and the heterozygous woman requests prenatal diagnosis, the foetal DNA is analysed to exclude clinically significant disorders depending on the maternal genotype; e.g. exclusion of Hb SS, SC and S/ β thal in an AS woman; and homozygosity or compound heterozygosity for β -thalassaemia in a β -thalassaemia carrier. In most instances, direct sequence analysis of the PCR-amplified β -globin gene would provide a comprehensive genotype followed by confirmation of the genotype using another PCR-based technique.

Future approaches in DNA diagnostics

With the imminent implementation of universal antenatal screening for the clinically significant Hb disorders in the UK, there is a drive for a higher throughput method for DNA diagnostics. This will necessitate the use of automated platforms, not only to perform the analyses steps but also to prepare samples and reactions prior to the actual detection step.

A promising approach which allows rapid and cost-effective DNA diagnostics is to use arrays of oligonucleotides (DNA chips) as solid support in hybridization assays (Brown & Botstein, 1999; Southern, Mir & Shchepinov, 1999). High-density arrays of oligonucleotides and sophisticated systems for fluorescence detection and data analysis for this assay format are under development. Hybridization with ASO probes have been the most frequently applied principle for detecting single nucleotide

polymorphisms (SNPs) and point mutations in a DNA chip format. However, imperfect hybridization specificity and difficulty in discriminating a large number of sequence variants simultaneously because of inherent properties of nucleic acid hybridization is still a limiting factor in these methods (Hacia & Collins, 1999).

DNA microarrays as a means of resequencing genes, for which a reference sequence is known, to scan for all possible sequence variation or mutation is another possible development (Hacia, 1999). The 'mini-sequencing' method of Syvänen (Syvänen, Sajantila & Lukka, 1993; Pastinen *et al.*, 1997) has been applied for the identification of a limited number of known mutations. Here, unlabelled target is hybridized to an oligonucleotide primer which is attached to an array. Solid-phase mini-sequencing is then applied by extending the immobilized primers specifically from the 3'-end with single labelled dideoxynucleotide triphosphates using a DNA polymerase. An advantage of mini-sequencing, compared with hybridization with ASOs, is that all variable nucleotides are identified with optimal conditions in the same reaction conditions. Sequencing by hybridization to immobilized oligonucleotides on chips has been applied to DNA diagnostics for a limited number of β -thalassaemia mutations (Dubiley, Kirillov & Mirzabekov, 1999). It may be possible to scale up the mini-sequencing method to scan all of the β -globin gene which is particularly compact and well studied, with the majority of all the known mutations contained within the 1600 bp sequence.

The mini-sequencing approach has also been applied on a format using capillary electrophoresis on the Genetic Analyzer (ABI 3100, Applied Biosystems, Foster City, CA, USA). The principle is similar to mini-sequencing on DNA chips. After PCR amplification of the genomic region of interest a cycle sequencing reaction is carried out using a primer that directly abuts the nucleotide base of interest. A fluorescently labelled dideoxy chain terminator can then be incorporated that extends the primer by a single nucleotide. The reaction is then electrophoresed with a size marker in an automatic sequencer. Individuals that are heterozygous will be detected by the different coloured nucleotides that are incorporated and multiplex reactions can be distinguished by making the cycling primers of different lengths. This approach has been investigated for the sickle mutation and some β -thalassaemia mutations (Fiorentino *et al.*, 2003). The ease with which the method can be multiplexed lends itself very well to diseases like the β -haemoglobinopathies where there are a number of different mutations to test for within a relatively small region. It is easy to envisage the mini-sequencing approach being used as a first line screen for the haemoglobinopa-

thies but the technique must prove to be accurate, robust and cost effective for it to replace the current methods.

Pyrosequencing is rapidly gaining popularity as a resequencing methodology for detecting known mutations and polymorphisms (Nordfors *et al.*, 2002) in small well-defined regions. It relies on individual nucleotides being sequentially added to an extended primer that releases pyrophosphate during nucleotide incorporation. This pyrophosphate is converted to a light signal using ATP sulphurylase and luciferase and is proportional to the quantity of nucleotides incorporated (Ronaghi *et al.*, 1996; Ronaghi, 2001). The sequence is read in real time and is semi-quantitative making it quick and highly suitable for heterozygote identification. The pyrosequencing technique is most suitable for identifying known mutations and polymorphisms; the appropriate nucleotides can be added in sequence order reducing the overall sequencing time. In a similar way pyrosequencing can be fine-tuned to detect known mutations in well-characterized sequences. For the haemoglobinopathies, pyrosequencing would be useful for analysing the alpha genes especially in the presence of a triplicated alpha globin gene complex. Pyrosequencing has also been used to screen for mutations within the N-ras oncogene (Sivertsson *et al.*, 2002) that contains mutation hotspots. By focusing the attention on these regions pyrosequencing was found to be just as accurate as SSCP in detecting mutations but had the added advantage of providing a definitive diagnosis. Further, pyrosequencing is accurate and flexible; it allows parallel processing, and can be easily automated. The technique dispenses with labelled primers, labelled nucleotides and gel electrophoresis.

Conclusion

The compactness of the globin genes means that haemoglobinopathy detection is largely a PCR-based approach that can utilise direct sequencing analysis. Almost all haemoglobinopathies can be detected with the current PCR-based assays with the exception of a few rare deletions. However, the molecular diagnostic service is still under development to try and meet the demands of the population it serves. A higher throughput approach will be required to meet practice pressures and the increasing needs set by the antenatal screening programme. This increasing workload dictates increasing automation, which may necessitate the use of automated robotic platforms to prepare samples and reactions, and the use of automated platforms to perform the actual detection. In most populations the β -thalassaemias (and related haemoglobinopathies) are clinically more relevant

than the α -thalassaemias. The complex mutational spectrum of the haemoglobinopathies, especially relevant in a multi-ethnic community, requires a method with the capacity to scan the β (and/or α) globin genes rapidly and accurately for all mutations. This aspect is being addressed by the development of arrays (Chan *et al.*, 2004). Although in their infancy, the arrays hold great promise and are amenable to scaling up and automation. In the short term, the current generation of instruments such as the capillary electrophoresis systems, has greatly simplified DNA sequence analysis. The capillary electrophoresis system also lends itself to the multiplexed mini-sequencing methodology which is highly suitable for screening for the common globin gene mutations.

Acknowledgements

We thank Claire Steward for help in preparing the manuscript; Alice Gallienne, Chris Lambert, Lisa Thompson and Tania Senior for providing the illustrations used in some of the figures.

References

- Brown P.O. & Botstein D. (1999) Exploring the new world of the genome with DNA microarrays. *Nature Genetics* **21**(Suppl.), 33–37.
- Cao A., Rosatelli M.C. & Eckman J.R. (2001) Prenatal diagnosis and screening for thalassaemia and sickle cell disease. In: *Disorders of Hemoglobin; Genetics, Pathophysiology and Clinical Management* (eds M.H. Steinberg, B.G. Forget, D.R. Higgs & R.L. Nagel), pp. 958–978. Cambridge University Press, Cambridge.
- Chan K., Wong M.S., Chan T.K. & Chan V. (2004) A thalassaemia array for Southeast Asia. *British Journal of Haematology* **124**, 232–239.
- Chang J.-G., Lu J.-M., Huang J.-M., Chen J.-T., Liu H.-J. & Chang C.-P. (1995) Rapid diagnosis of β -thalassaemia by mutagenically separated polymerase chain reaction (MS-PCR) and its application to prenatal diagnosis. *British Journal of Haematology* **91**, 602–607.
- Chehab F.F. & Kan Y.W. (1989) Detection of specific DNA sequences by fluorescence amplification: a color complementation assay. *Proceedings of the National Academy of Sciences of the United States of America* **86**, 9178–9182.
- Chehab F.F. & Kan Y.W. (1990) Detection of sickle cell anaemia mutation by colour DNA amplification. *Lancet* **335**, 15–17.
- Cheung M.-C., Goldberg J.D. & Kan Y.W. (1996) Prenatal diagnosis of sickle cell anaemia and thalassaemia by analysis of fetal cells in maternal blood. *Nature Genetics* **14**, 264–268.
- Chiu R.W., Lau T.K., Leung T.N., Chow K.C., Chui D.H. & Lo Y.M. (2002) Prenatal exclusion of beta thalassaemia major by examination of maternal plasma. *Lancet* **360**, 998–1000.
- Chong S.S., Boehm C.D., Higgs D.R. & Cutting G.R. (2000) Single-tube multiplex-PCR screen for common deletional determinants of alpha-thalassaemia. *Blood* **95**, 360–362.
- Conner B.J., Reyes A.A., Morin C., Itakura K., Teplitz R.L. & Wallace R.B. (1983) Detection of sickle cell β^S -globin allele by hybridization with synthetic oligonucleotides. *Proceedings of the National Academy of Sciences of the United States of America* **80**, 278–282.
- Craig J.E., Barnetson R., Weatherall D.J. & Thein S.L. (1993a) Rapid detection of a 13.4 kb deletion causing delta beta thalassaemia in an Egyptian family by polymerase chain reaction. *Blood* **81**, 861–870.
- Craig J.E., Sheerin S.M., Barnetson R. & Thein S.L. (1993b) The molecular basis of HPFH in a British family identified by heteroduplex formation. *British Journal of Haematology* **84**, 106–110.
- Craig J.E., Barnetson R.A., Prior J., Raven J.L. & Thein S.L. (1994) Rapid detection of deletions causing delta beta thalassaemia and hereditary persistence of fetal hemoglobin by enzymatic amplification. *Blood* **83**, 1673–1682.
- Decorte R., Cuppens H., Marynen P. & Cassiman J.J. (1990) Rapid detection of hypervariable regions by the polymerase chain reaction technique. *DNA and Cell Biology* **9**, 461–469.
- Dodé C., Rochette J. & Krishnamoorthy R. (1990) Locus assignment of human alpha-globin mutations by selective amplification and direct sequencing. *British Journal of Haematology* **76**, 275–281.
- Dodé C., Krishnamoorthy R., Lamb J. & Rochette J. (1992) Rapid analysis of $-\alpha^{3.7}$ thalassaemia and α - α - $\alpha^{\text{anti } 3.7}$ triplication by enzymatic amplification analysis. *British Journal of Haematology* **82**, 105–111.
- Dubiley S., Kirillov E. & Mirzabekov A. (1999) Polymorphism analysis and gene detection by minisequencing on an array of gel-immobilized primers. *Nucleic Acids Research* **27**, e19.
- Faa V., Rosatelli M.C., Sardu R., Meloni A., Toffoli C. & Cao A. (1992) A simple electrophoretic procedure for fetal diagnosis of beta-thalassaemia due to short deletions. *Prenatal Diagnosis* **12**, 903–908.
- Fiorentino F., Magli M.C., Podini D., Ferraretti A.P., Nuccitelli A., Vitale N., Baldi M. & Gianaroli L. (2003) The minisequencing method: an alternative strategy for preimplantation genetic diagnosis of single gene disorders. *Molecular Human Reproduction* **9**, 399–410.
- Flint J., Harding R.M., Boyce A.J. & Clegg J.B. (1998) The population genetics of the haemoglobinopathies. In: *Baillière's Clinical Haematology* (ed. G.P. Rodgers), Vol. 11, pp. 1–52. Baillière Tindall, London.
- Galanello R., Barella S., Ideo A., Gasperini D., Rosatelli C., Paderi L., Paglietti E., Sollaino C., Perseu L., Loi D. (1994) Genotype of subjects with borderline hemoglobin A2 levels: implication for beta-thalassaemia carrier screening. *American Journal of Hematology* **46**, 79–81.
- Hacia J.G. (1999) Resequencing and mutational analysis using oligonucleotide microarrays. *Nature Genetics* **21**(Suppl.), 42–47.
- Hacia J.G. & Collins F.S. (1999) Mutational analysis using oligonucleotide microarrays. *Journal of Medical Genetics* **36**, 730–736.
- Hardison R.C., Chui D.H., Giardine B., Riemer C., Patrinos G.P., Anagnou N., Miller W. & Wajcman H. (2002) HbVar: a relational database of human hemoglobin variants and thalassaemia mutations at the globin gene server. *Human Mutation* **19**, 225–233.

- Higgs D.R. (2001) Molecular mechanisms of alphas-thalassaemia. In: *Disorders of Hemoglobin: Genetics, Pathophysiology, and Clinical Management* (eds M.H. Steinberg, B.G. Forget, D.R. Higgs & R.L. Nagel), pp. 405–430. Cambridge University Press, Cambridge, UK.
- Huisman T.H.J., Carver M.F.H. & Efremov G.D. (1998) *A Syllabus of Human Hemoglobin Variants*, 2nd edn. The Sickle Cell Anemia Foundation, Augusta, GA, USA.
- Jassim N., Al-Arrayed S., Gerard N., Al-Mukharraq H., Al-Ajami A., Ramasawmy R. & Krishnamoorthy R. (1999) A mismatched-primer polymerase chain reaction-restriction fragment length polymorphism strategy for rapid screening of the polyadenylation signal mutation alpha(T-Saudi) (AATAAA → AATAAG) in the alpha2-globin gene. *Hemoglobin* **23**, 213–220.
- Kobayashi Y., Fukumaki Y., Komatsu N., Ohba Y., Miyaji T. & Miura Y. (1987) A novel globin structural mutant, Showa-Yakushiji (Beta110 Leu-Pro) causing a beta-thalassemia phenotype. *Blood* **70**, 1688–1691.
- Liu Y.T., Old J.M., Miles K., Fisher C.A., Weatherall D.J. & Clegg J.B. (2000) Rapid detection of alpha-thalassaemia deletions and alpha-globin gene triplication by multiplex polymerase chain reactions. *British Journal of Haematology* **108**, 295–299.
- Losekoot M., Fodde R., Hartevelde C.L., van Heeren H., Giordano P.C. & Bernini L.F. (1990) Denaturing gradient gel electrophoresis and direct sequencing of PCR amplified genomic DNA: a rapid and reliable diagnostic approach to beta thalassaemia. *British Journal of Haematology* **76**, 269–274.
- Maggio A., Giambona A., Cai S.P., Wall J., Kan Y.W. & Chehab F.F. (1993) Rapid and simultaneous typing of hemoglobin S, hemoglobin C, and seven mediterranean beta-thalassaemia mutations by covalent reverse dot-blot analysis: application to prenatal diagnosis in Sicily. *Blood* **81**, 239–242.
- 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 Research* **17**, 2503–2516.
- Nordfors L., Jansson M., Sandberg G., Lavebratt C., Sengul S., Schalling M. & Arner P. (2002) Large-scale genotyping of single nucleotide polymorphisms by PyrosequencingTM and validation against the 5' nuclease (Taqman®) assay. *Human Mutation* **19**, 395–401.
- Old J.M. (1986) Fetal DNA analysis. In: *Human Genetic Diseases – A Practical Approach* (ed. K.E. Davies), pp. 1–17. IRL Press, Oxford.
- Old J.M. (2003) Screening and genetic diagnosis of haemoglobin disorders. *Blood Reviews* **17**, 43–53.
- Old J.M., Varawalla N.Y. & Weatherall D.J. (1990) Rapid detection and prenatal diagnosis of beta-thalassaemia: studies in Indian and Cypriot populations in the UK. *Lancet* **336**, 834–837.
- Orita M., Iwahana H., Kanazawa H., Hayashi K. & Sekiya T. (1989) Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms. *Proceedings of the National Academy of Sciences of the United States of America* **86**, 2766–2770.
- Pastinen T., Kurg A., Metspalu A., Peltonen L. & Syvänen A.-C. (1997) Minisequencing: a specific tool for DNA analysis and diagnostics on oligonucleotide arrays. *Genome Research* **7**, 606–614.
- Pirastu M., Ristaldi M.S. & Cao A. (1989) Prenatal diagnosis of beta thalassaemia based on restriction endonuclease analysis of amplified fetal DNA. *Journal of Medical Genetics* **26**, 363–367.
- Ristaldi M.S., Pirastu M., Rosatelli C., Monni G., Erlich H., Saiki R. & Cao A. (1989) Prenatal diagnosis of beta-thalassaemia in Mediterranean populations by dot blot analysis with DNA amplification and allele specific oligonucleotide probes. *Prenatal Diagnosis* **9**, 629–638.
- Rochette J., Craig J.E. & Thein S.L. (1994) Fetal hemoglobin levels in adults. *Blood Reviews* **8**, 213–224.
- Ronaghi M. (2001) Pyrosequencing sheds light on DNA sequencing. *Genome Research* **11**, 3–11.
- Ronaghi M., Karamohamed S., Pettersson B., Uhlen M. & Nyren P. (1996) Real-time DNA sequencing using detection of pyrophosphate release. *Analytical Biochemistry* **242**, 84–89.
- Saiki R.K., Scharf S., Faloona F., Mullis K.B., Horn G.T., Erlich H.A. & Arnheim N. (1985) Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anaemia. *Science* **230**, 1350–1354.
- Saiki R.K., Chang C.-A., Levenson C.H., Warren T.C., Boehm C.D., Kazazian H.H.J. & Erlich H.A. (1988) Diagnosis of sickle cell anemia and beta-thalassaemia with enzymatically amplified DNA and nonradioactive allele-specific oligonucleotide probes. *New England Journal of Medicine* **319**, 537–541.
- Saiki R.K., Walsh P.S., Levenson C.H. & Erlich H.A. (1989) Genetic analysis of amplified DNA with immobilized sequence-specific oligonucleotide probes. *Proceedings of the National Academy of Sciences of the United States of America* **86**, 6230–6234.
- Sambrook J. & Russell D.W. (2001) *Molecular Cloning: A Laboratory Manual*, 3rd edn. Cold Spring Harbor Laboratory Press, New York.
- Sanger F., Donelson J.E., Coulson A.R., Kossel H. & Fischer D. (1973) Use of DNA polymerase I primed by a synthetic oligonucleotide to determine a nucleotide sequence in phage phi DNA. *Proceedings of the National Academy of Sciences of the United States of America* **70**, 1209–1213.
- 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.
- Sivertsson A., Platz A., Hansson J. & Lundberg J. (2002) Pyrosequencing as an alternative to single-strand conformation polymorphism analysis for detection of N-ras mutations in human melanoma metastases. *Clinical Chemistry* **48**, 2164–2170.
- Southern E.M. (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. *Journal of Molecular Biology* **98**, 503–517.
- Southern E., Mir K. & Shchepinov M. (1999) Molecular interactions on microarrays. *Nature Genetics* **21**(Suppl.), 5–9.
- Sutcharitchan P., Saiki R., Fucharoen S., Winichagoon P., Erlich H. & Embury S.H. (1995a) Reverse dot-blot detection of Thai beta-thalassaemia mutations. *British Journal of Haematology* **90**, 809–816.
- Sutcharitchan P., Saiki R., Huisman T.H., Kutlar A., McKie V., Erlich H. & Embury S.H. (1995b) Reverse dot-blot detection of the African-American beta-thalassaemia mutations. *Blood* **86**, 1580–1585.

- Syvänen A.-C., Sajantila A. & Lukka M. (1993) Identification of individuals by analysis of biallelic DNA markers, using PCR and solid-phase minisequencing. *American Journal of Human Genetics* **52**, 46–59.
- Tan J.A., Tay J.S., Lin L.I., Kham S.K., Chia J.N., Chin T.M., Aziz N.B. & Wong H.B. (1994) The amplification refractory mutation system (ARMS): a rapid and direct prenatal diagnostic technique for beta-thalassaemia in Singapore. *Prenatal Diagnosis* **14**, 1077–1082.
- Tan A.S., Quah T.C., Low P.S. & Chong S.S. (2001) A rapid and reliable 7-deletion multiplex polymerase chain reaction assay for alpha-thalassaemia. *Blood* **98**, 250–251.
- Thein S.L. (1998) Beta thalassaemia. In: *Baillière's Clinical Haematology, Sickle Cell Disease and Thalassaemia* (ed. G.P. Rodgers), Vol. 11:1, pp. 91–126. Baillière Tindall, London.
- Thein S.L. (1999) Is it dominantly inherited beta thalassaemia or just a beta-chain variant that is highly unstable? *British Journal of Haematology* **107**, 12–21.
- Thein S.L. (2001) Structural variants with a beta-thalassaemia phenotype. In: *Disorders of Hemoglobin: Genetics, Pathophysiology, and Clinical Management* (eds M.H. Steinberg, B.G. Forget, D.R. Higgs & R.L. Nagel), pp. 342–355. Cambridge University Press, Cambridge, UK.
- Thein S.L., Lynch J.R., Old J.M. & Weatherall D.J. (1987) Direct detection of haemoglobin E with MnlI. *Journal of Medical Genetics* **24**, 110–111.
- 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.
- Wallace R.B., Johnson M.J., Hirose T., Miyake T., Kawashima E.H. & Itakura K. (1981) The use of synthetic oligonucleotides as hybridization probes. II: Hybridisation of oligonucleotides of mixed sequence to rabbit β -globin DNA. *Nucleic Acids Research* **9**, 879–894.
- Weatherall D.J. & Clegg J.B. (2001) *The Thalassaemia Syndromes*, 4th edn. Blackwell Science, Oxford.
- Wood W.G. (2001) Hereditary persistence of fetal hemoglobin and delta beta thalassaemia. In: *Disorders of Hemoglobin: Genetics, Pathophysiology, and Clinical Management* (eds M.H. Steinberg, B.G. Forget, D.R. Higgs & R.L. Nagel), pp. 356–388. Cambridge University Press, Cambridge, UK.
- Ye S., Dhillon S., Ke X., Collins A.R. & Day I.N. (2001) An efficient procedure for genotyping single nucleotide polymorphisms. *Nucleic Acids Research* **29**, E88.