CHAPTER I

INTRODUCTION

1.1 Hemoglobin

Hemoglobin is the major protein in the red cell which represent more than 95 % of the soluble protein in cytoplasm of the erythrocyte. The major function of hemoglobin is transportation of oxygen from lung to body tissues. An associated function is the binding of carbon dioxide and proton by deoxyhemoglobin, thereby serving to buffer the blood on the venous side of the circulation. Hemoglobin synthesis requires one molecule of heme and four molecules of globin chains. The tetameric structure of normal hemoglobin contains two α-like globin chains and two β-like globin chains. (Figure 1.1) The α-like globin chain contains 141 amino acid residues and β-like globin chain contains 146 amino acid residues. Three types of hemoglobin are commonly found in adults comprising Hb A, Hb A_2 and Hb F (Weatherall and Clegg 1981). Hb F and Hb A are normally seen in fetal life and Hbs Gower 1 and 2 and Hb Portland are the embryonic hemoglobins.

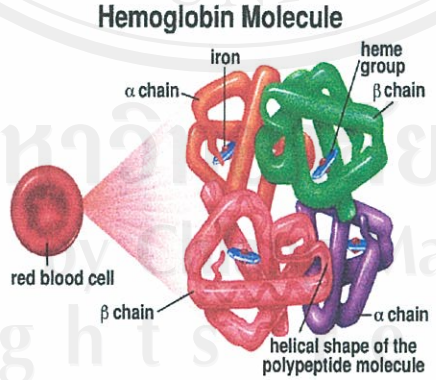

Figure 1.1 Model of hemoglobin A ($\alpha_2\beta_2$), a view of the subunit contacts and the heme pocket. (http://www.age.uci.edu/~bcrourke/blood)

1.2 Hemoglobin synthesis throughout human life

Hemoglobin synthesis requires the coordinate production of heme and globin. Heme is prosthetic group that mediates reversible binding of iron by hemoglobin. Globin is the protein that surrounds and protects the heme molecule. The developmental changes in hemoglobin production are brought about by differential activation of the globin genes, which is largely determined at the level of transcription. This series of events, called 'hemoglobin switching'. The globin genes are activated in sequence during development, moving from 5' to 3' on the chromosome. In embryos, the ζ gene is expressed during the first few weeks of gestation, hemoglobin synthesis is confined to the yolk sac, where Hb Gower 1 ($\zeta_2\varepsilon_2$) is produced. Thereafter, the α-globin gene take over. For the β-globin gene cluster, the ε gene is expressed initially during embryogenesis, Hbs Gower 2 ($\alpha_2\varepsilon_2$) and Portland ($\zeta_2\gamma_2$) are produced. At around 7-8 weeks' gestation, the γ gene is expressed during fetal development. The liver becomes the major site of erythropoiesis. The combination of two α-globin genes and two γ-globin genes forms fetal hemoglobin, or hemoglobin F ($\alpha_2\gamma_2$) (Leung, *et al* 1987). At mid-term, the bone marrow begins to take over as the major site of red-cell production, though erythropoiesis is also found in the spleen, as well as in other tissues (Keleman, *et al* 1981). The production of γ-globin chains declines in concert with an elevation in β-globin synthesis. The combination of two α-globin genes and two β-globin genes results in the normal adult hemoglobin or hemoglobin A ($\alpha_2\beta_2$). The δ- gene, which is located between the γ and β-globin genes on chromosome 11 produces a small amount of δ-globin in children and adults. Hemoglobin A_2 ($\alpha_2\delta_2$), which normally comprises less than 3 % of hemoglobin in adults, is composed of two α-globin chains and two δ-globin chains. At birth, cord blood normally contains~70% Hb F and this declines to ~ 20% by 3 months, 7.5% at 6 months, and less than 2% by the age of 1 year. Thus, in normal adult, three hemoglobins are generally found: Hb A (97%), Hb A_2 (2-3%) and Hb F (<1%) (Figure. 1.2 and Table 1.1 (Weatherall, *et al* 1981).

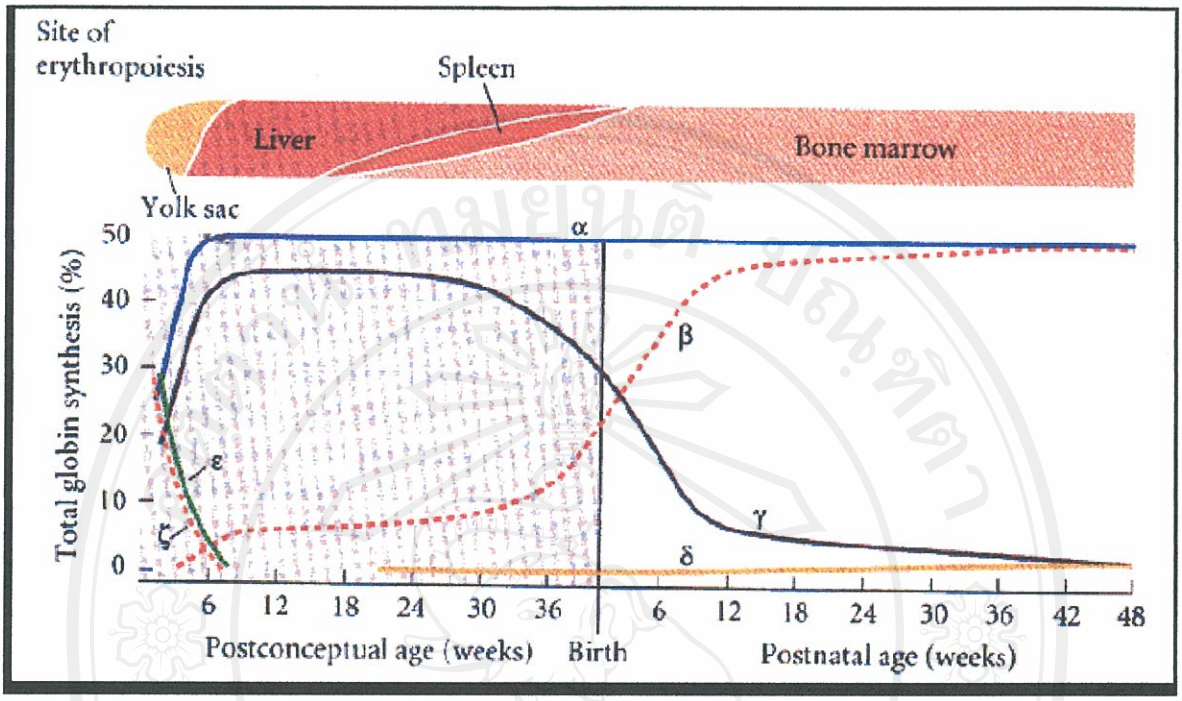

Figure1.2 Schematic representation of the pattern of synthesis of the different globin chains at various stages of development, showing the sites of erythropoiesis. (http://info.med.yale.edu/obgyn/reproimmuno/courses/class3/img002.gif)

Table 1.1 Hemoglobin types in human (Weatherall and Clegg 1981)

Embryonic hemoglobins	Fetal hemoglobin	Adult hemoglobins
Gower-1 ($\zeta_2\varepsilon_2$)	Hemoglobin F ($\alpha_2\gamma_2$)	Hemoglobin A ($\alpha_2\beta_2$)
Gower-2 ($\alpha_2\varepsilon_2$)		Hemoglobin A_2 ($\alpha_2\delta_2$)
Portland ($\zeta_2\gamma_2$)		

1.3 Globin genes

Human globin chains are encoded by globin genes on two separated globin gene clusters; the β-like globin gene cluster on chromosome 11p15 and the α-like globin gene cluster closed to the terminus of chromosome 16p (Deisseroth, *et al* 1977).

1.3.1 α-globin gene cluster

In human, the α-globin gene cluster (Figure.1.3) occupies a region of about 30 kb closed to the tip of the short arm of chromosome 16, band 16p 13.3. The position of α-globin gene cluster arranged in the following order 5'-ζ-φζ-φα$_1$-φα$_2$-α$_2$-α$_1$-θ-3'. Thus, the α$_2$ gene, at +20 kb, lies 20 kb away from the ζ gene on the centromeric side, and the α$_1$ gene lies a further 3.7 kb away, at + 24 kb. Sequences which are essential for α-globin expression, notably the major α-globin gene regulatory element, HS –40, are centered on a region 40 kb upstream of the ζ-globin gene. In addition to the three functional α-like genes, the cluster contains three pseudogenes (-φζ-φα$_1$-φα$_2$) and another gene, θ, the later is transcribed at low level in erythroid cells, but so far no protein is detected. However, product has been identified and the protein predicted from its sequence is unlikely to be a functional globin (Clegg 1987). The region around the α-globin gene cluster is rich in Alu repetitive sequences. Large numbers of these repeats flank the cluster and several copies are interspersed within it. Their transcribed regions containe three exons, separated by two introns, or intervening sequences (IVS), of variable length. In the α-globin gene both introns are small, 117-149 bp, while in the ζ gene IVS 1 is ~ 886 and IVS 2 is ~ 239 bp (Weatherall, *et al* 1981).

1.3.2 β-globin gene cluster

The β-globin gene cluster encompasses approximately 70-kb region on the short arm of chromosome 11, band 11p15.4 (Figure1.3). The position of β-globin gene cluster arranges in the following order 5'-ε-γ^{G}- γ^{A}-φβ-δ-β-3'. The two fetal γ-globin genes lie 15 and 20 kb downstream from the embryonic ε-globin gene, while the δ-globin and β-globin genes are 35 and 43 kb further downstream. Upstream of the ε-globin gene, lies the locus-control region (LCR), the regulatory region that is essential for expression of all the genes in the complex. Their transcribed regions contain three exons, separated by two introns, or intervening sequences (IVS), of variable length. The first exon of the β-like globin chain contains codons for amino acids 1-30, the second exon for amino acids 31-104 and the third exon for amino acids 105-146 of the β globin chain (Bulger, *et al* 1999).

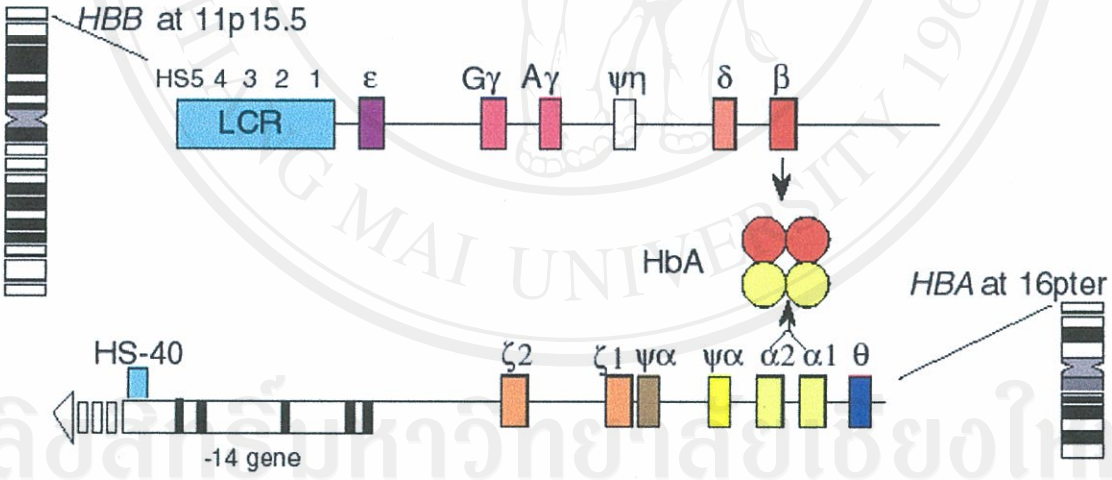

Figure 1.3 Localization and organization of α- and β-globin genes clusters.

(http://globin.cse.psu.edu)

1.4 Thalassemia and Hemoglobinopathies

Thalassemia is a type of anemic disease characterized by reduction or absence of globin chain synthesis. This results in imbalanced-globin chain synthesis; the major pathogenesis of the disease, leading to various fatal consequences. The erythocytes are characterized by decreased intracellular Hb content (hypochromia) and small size (microcytosis). In addition, the continued normal synthesis of unaffected globin chain leads to the accumulation of unstable aggregates of these unmatched globin chains. These unstable aggregates precipitate within the erythroid cells, leading to oxidative membrane damage and premature destruction of erythroid cells in peripheral circulation as well as at carlier stages of maturation in the bone marrow. This causes expansion of the ineffective marrow with severe effects on development, bone formation and growth. The major cause of morbidity and mortality is the effect of iron deposition in the endocrine organs, liver, and heart, which results from increased intestinal absorption and frequent blood transfusions (Weatherall and Clegg 2001). Thalassemia is considered as the most common autosomal single-gene disorder worldwide. It can be found in more than 150 countries with estimated carrier frequency of about 7 % (Weatherall and Clegg 2001). The Mediterranean region, certain parts of North and West Africa, Middle East, Indian subcontinent, Southern Far East and South East Asia have the highest prevalence of the disease .

Hemoglobinopathies is an inherited disorders of hemoglobin productions characterized by production of abnormal hemoglobins or hemoglobin structural variants occuring from genetic alterations including point mutations, deletions or insertion of the globin genes. 887 hemoglobin variants have been described to date (http://globin.cse.psu.edu/cgi-bin/hbvar/xounte: accessed 13/07/04), over 90 % of which are single amino-acid substitutions in the α, β, γ or δ-globin chains and over 60% involve the β-globin chain. The most important hemoglobin structural variants are hemoglobins Constant Spring (Cs), C, S, D and E. In particular, Hb E is the most

commonly found hemoglobin variant in Thailand accounting for approximately 13% to 70% of the population (Chernoff, *et al* 1954, Fucharoen and Winichagoon 1997)

1.5 Common types of thalassemia and hemoglobinopathies

Two types of thalassemia commonly found in Thailand, by genetic classification, are alpha (α) and beta (β) thalassemia.

1.5.1 α-thalassemia

The α-thalassemia is characterized by absent or decreased α-globin chain production. In humans, the α-globin gene cluster occupies two α-globin genes per haploid genome (α1 and α2 globin genes). A normal α-globin genotype can be represented as αα/αα. When both α-globin genes on a chromosome are deleted or otherwise inactivated, the condition is called α^0 thalassemia or α–thalassemia-1. The α–thalassemia-1 Southeast Asian (SEA) type is the most frequently found α-globin deletion in Thailand (Pressley, *et al* 1980). The SEA deletion removes about 19.304 kb of α-globin gene cluster including α1 and α2 globin genes (Figure1.4) (Kantamol 2003). When only one of the four α-globin genes is deleted or dysfunction, the condition is referred to as α^+ thalassemia or α – thalassemia-2. In Thailand, there are two types of deletions causing the α–thalassemia-2 including a deletion of 4.2 kb of DNA that removes α2 globin genes (leftward or 4.2-kb type) and 3.7 kb deletion removing the region between the duplicate α genes (rightward or 3.7-kb type) (Embury, *et al* 1980). The later is the most frequently observed in Thai population (Figure1.5). In Thailand, the overall frequency of α-thalassemia gene is 20-30 %. The frequency of α-thalassemia-1 is high in the North; 10% in Chiang Mai comparing to 3.5% in Bangkok whereas α-thalassemia-2 is between 16-20% (Fucharoen and Winichagoon 2002). Four main clinical types of α-thalassemia syndromes of progressively increasing severity were defined in this population and were thought to be caused by deletion or inactivating

mutation of one, two, three, or four of the α-globin gene loci. These four syndromes are namely called; α-thalassemia-2 or silent carrier state (one inactive α-globin gene); α-thalassemia-1 (two inactive α-globin genes); Hb H disease (three inactive α-globin genes), due to compound heterozygosity for α-thalassemia-2 plus α-thalassemia-1; and hydrops fetalis with Hb Bart's (four inactive α-globin genes), due to homozygosity for α-thalassemia-1.

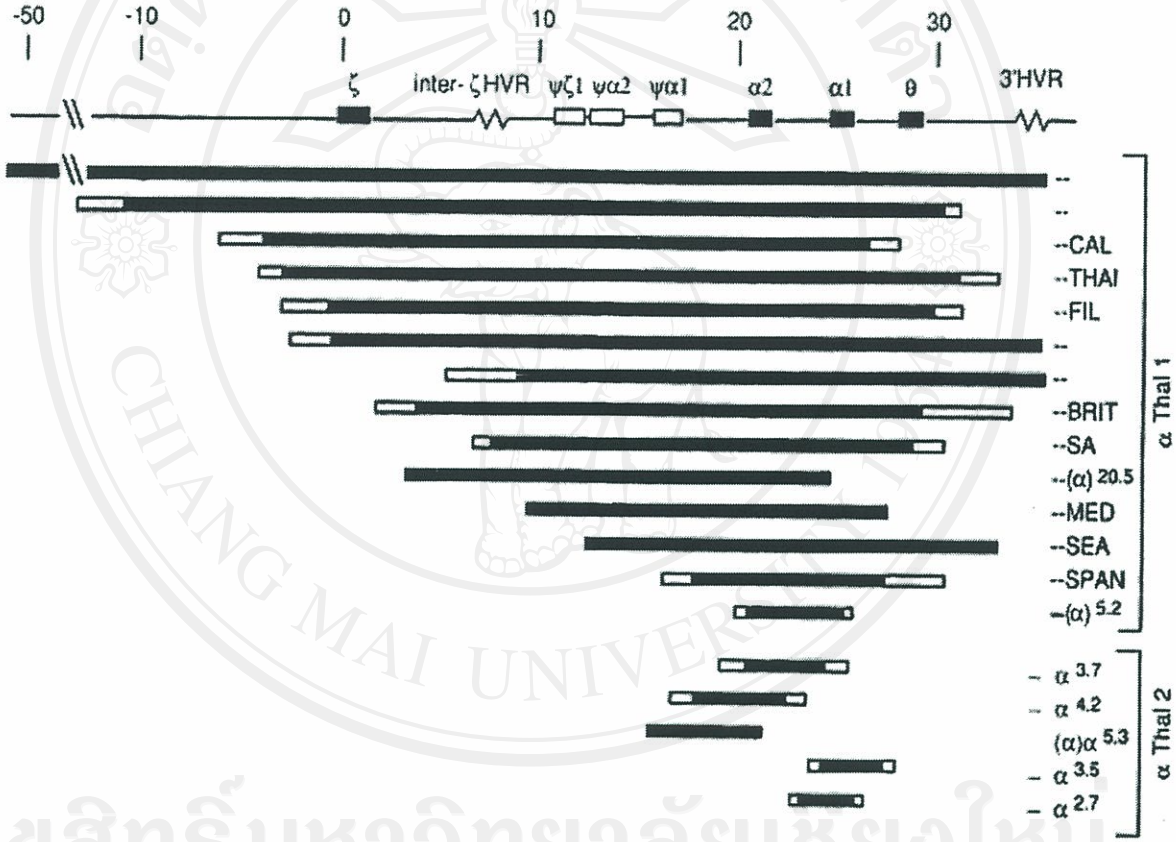

Figure 1.4. Extent of various deletions in the α-globin gene cluster associated with α-thalassemia (α-thal 1 and α-thal 2). The letters at the right part of the figure demonstrate types of deletion from which each type of α-thalassemia is resulted. They also indicate the ethnic background where each deletiion is commonly found. BRIT, British; CAL, Calabrian; FIL, Filipino; HVR, hypervariable region; MED, Mediterranean; SA, South African; SEA, Southeast Asian; SPAN, Spanish. (Higgs, *et al* 1989).

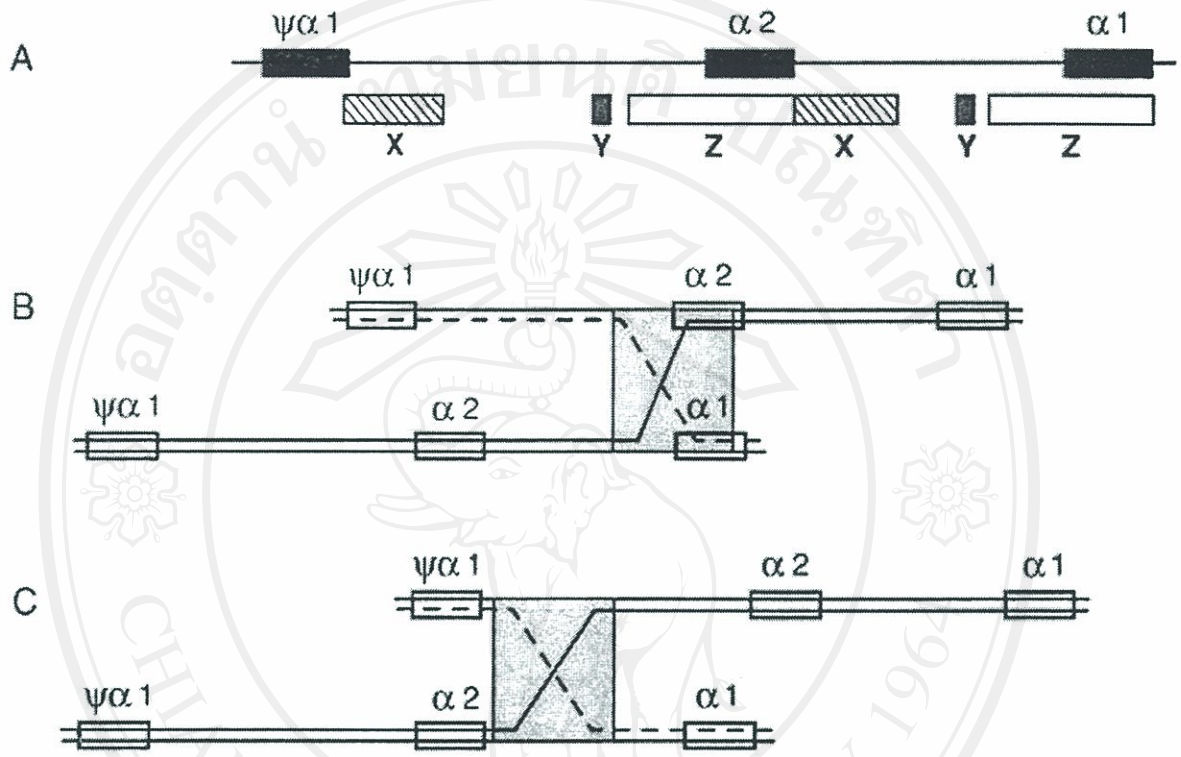

Figure 1.5. Model of unequal crossover events in the α-globin gene cluster resulting in α-thalassemia-2 deletions and reciprocal triple α gene clusters. A: The X, Y, and Z homology blocks in the region of the α-globin genes are shown. B: Mispairings and crossing over in the Z homology block produce the $-\alpha^{3.7}$ chromosome containing a 3.7-kb deletion (*dashed line*) (Rightward deletion). The reciprocal of this event (*solid line*) is a triple α chromosome. C: Mispairing and crossing over in the X homology block produces a single chromosome with a 4.2-kb deletion (*dashed line*)(Leftward deletion). The reciprocal of this event (*solid line*) is also a triple α chromosome (Kazazian 1990).

1.5.2 α - hemoglobinopathies

α-hemoglobinopathy is characterized by production of abnormal α- globin chains. In Thailand, α-hemoglobinopathies are varities, including Hb Anantharaj, Hb Suan Dok, Hb Mahidol (Hb Q), Hb Siam, Hb Pakse and Hb Constant Spring (Hb Cs). Hb Cs is common in Thailand, about 1% heterozygous carrier, 5 to 7 % in homozygotes and 3 to 5 % in hemoglobin H disease from Hb Cs *in-trans* to deletion α-thalassemia. Hb Cs is an α-globin chain variant which is elongated by 31 amino acid residues at its C-terminal end. Heterozygotes of Hb Cs and two normal *trans* α-globin genes are hematologically normal. Homozygotes for Hb Cs have a mild anemia and may have splenomegaly. The patients with the forms of Hb H associated with Hb Cs are even more anemia than those of patients with other forms of Hb H disease (Schrier, *et al* 1997).

1.5.3 β - thalassemia

The β – thalassemia are a diverse group of disorders of hemoglobin synthesis which is characterized by reduced or absent β–globin chain synthesis. There are two main varieties of β thalassemia, β^0- thalassemia, in which no β -globin chain is produced, and β^+- thalassemia, in which some β-globin chain is produced, but less than normal. The majority of β -thalassemia are caused by point mutations and small deletions or insertion within the β-globin genes. A few β -thalassemia are caused by large gene deletion in the nucleotide sequences. In generally, β^+-thalassemia mutations are resulted from the defects at RNA processing or promoter region of the gene, and some cases are caused by a mutation within the introns of β-globin gene. β^0- thalassemia mutations are resulted from the several causes such as a complete blocking of the β-globin mRNA production and some cases are caused by point mutations in the neucleotide sequence that result in termination codon (nosense mutations). In Thailand, β-thalassemia is common with the frequency varying from 1-9 % (Fucharoen and Winichagoon 2002).

1.5.4 β-hemoglobinopathies

β-hemoglobinopathy is characterized by the production of abnormal β-globin chains. To date, there are more than 660 β-globin variants reported across the world (http://globin.cse.psu.edu/cgi-bin/hbvar/xounte: accessed 13/07/04). In Thailand, Hb E is the most common β-globin structural variant accounting for approximately 13 to 70 % of population (Fucharoen and Winichagoon 1997). In Hb E, the G-A substitution at codon 26 partially activates a cryptic splice site towards the 3' end of exon 1, resulting in a proportion of abnormally spliced mRNA (Orkin, *et al* 1982). Thus, less β^E globin is synthesized and a mild thalassemia phenotype results. Hb E heterozygotes are clinically normal and have only minor hematological changes. Homozygotes have a very mild anemia, but are otherwise well; their hematological changes are similar to those of hetrozygous β thalassemia. It is the compound heteozygous state between Hb E and β thalassemia that gives rise to really serious clinical disease, with a phenotype ranging from mild anemia to the most severe form like β thalassemia major (Weatherall and Clegg 1981).

1.6 The molecular defect of β-thalassemia

Most mutation that cause of β-thalassemia are point mutations and small deletion in the part of the β-globin genes.

1.6.1 Point mutations

Point mutations is the cause of β-thalassemia, which results in ether absence of the synthesis of β-globin chains; β^0 thalassemia, or reduction of synthesis ; β^+ thalassemia. Todate, more than 200 different point mutation have thus far been identified as causes of β-thalassemia (Weatherall and Clegg 2001). These point mutations can be classified into 3 group; transcription, RNA processing and RNA translation mutations.

First, Transcription mutations involve the promoter mutations and 5' UTR mutations. The promoter mutations are single base subsubsitutions in the β-globin promoter including the binding site of transcriptions factor such as the Erythroid Kruppel-Like Factor (EKLF); a zinc-finger transcription nuclear protein. Transient expression system show that they result in an output of β-globin mRNA ranging from 10 to 25% of normal, a level compatible with the relatively mild phenotype of these β^+ thalassemia (Treisman, *et al* 1983). Several mutations in the 5' UTR include singe base substitutions and minor deletions distributed along this stretch of 50 nucleotides. Compound heterozygote of mutations in the 5'UTR and the more severe β-thalassemia alleles tend to have a mild disease. **Second**, RNA processing mutations. The sequences critical in this process include the invariant dinucleotides, GT at the 5' (donor) and AG at the 3'(acceptor) splice junctions between exons and introns (Breathnach and Chambon 1981, Mount 1982). The regains flanking these invariant dinucleotides are fairly well conserved and a consensus sequence can be recognized at exon-intron boundaries. They encompass the last three nucleotides of the exon and the first six nucleotieds of the intron for the 5' donor site, and the last 10 nucleotides of the intron and the first nucleotied of the exon for the 3' acceptor site. Both exons and introns also contain 'cryptic' splice sites, sequences which mimic the consensus sequence for a splice site but which are not normally used. During RNA processing the newly created site is utilized preferentially, leading to aberrant splicing. The mutations affecting for mRNA processing include the mutations of the splice junction, consensus-sequence mutations, cryptic splice-site mutations in introns or exons and poly(A) addition-site mutations. For examples, the substitution of C for T in the adjacent nucleotide, IVS 1 position 6, only mildly affects RNA splicing even though it activates the same three cryptic donor sites the IVS-5 mutations (Treisman, *et al* 1983), a T-G substitution in position 116 of IVS 1 leading to a new 3' acceptor site (Dobkin, *et al* 1983) and the codon 26 GAG-AAG and codon 27 GCC-TCC are associated with only a minor activation of the alternative splicing pathway (Orkin, *et al* 1984). Substitution or small deletions result in ineffective clevage of the

mRNA transcript and cause mild β^+ thalassemia (Orkin, *et al* 1985). **Third**, mutations affecting mRNA translation involve either the initiation or elongation phases of globin synthesis. These mutations cause complete absence of β-globin production; hence resulting in β^0 thalassemia. The mutations affecting the initiation codon which produce β^0 thalassemia. They are single base substitutions (Hattori, *et al* 1991, Jankovic, *et al* 1990, Lam, *et al* 1990). Approximately half the β-thalassemia alleles are characterized by premature β-chain termination, reflecting frameshift or nonsense mutations. For examples, frameshift mutations are 4-bp (-TTCT) deletion at codons 41/42 and base addition (+A) at codon 71/72 (Kazazian, *et al* 1984). Nonsense mutations are the A-T substitution at codon 17 and G-T substitution at codon 43 generates stop codon that prematurely terminates β-globin mRNA (Table 1.2).

1.6.2 Deletions

In contrast to the α thalassemia, the β thalassemia are rarely caused by major gene deletion. There are two notable exceptions; a family of upstream deletion, which downregulate the β-globin LCR; and another which involve only the β-globin genes themselves (Figure 1.6).

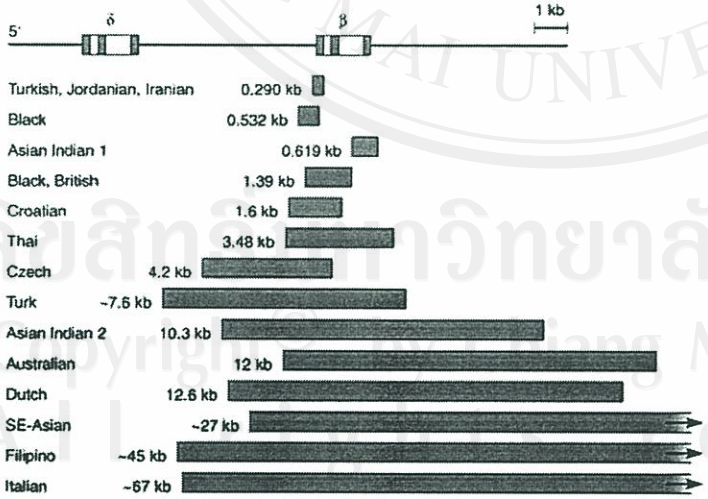

Figure 1.6 Deletions of the β-globin gene and flanking DNA associated with β^0-thalassemia (Forget 2001).

Table 1.2 Summary of some molecular defects causing β-thalassemia (Weatherall and Clegg 2001).

Mutations	Type	Distribution
1.Transcription mutation		
Promoter mutations		
ATA box		
-28 (A → G)	β^+	Chinese
-28 (A → C)	β^+	Kurdish
-29 (A → G)	β^+	American black, Chinese
-30 (T → C)	β^+	Chinese
-30 (T → A)	β^+	Eastern Mediterranean
-31 (A → G)	β^+	Japanese
-32 (C → A)	β^+	Taiwanese
CACCC box		
-86 (C → A)	β^+	Lebanese, Thai
-86 (C → A)	β^+	Italian
-87 (C → G)	β^+	Mediterranean
-87 (C → T)	β^+	German, Italian
-87 (C → A)	β^+	Yugoslavian, American black
-88 (C → T)	β^+	American black, Asian Indian
-88 (C → A)	β^+	American black, Asian Indian
-90 (C → T)	β^+	Portuguese
-92 (C → T)	β^+	Mediterranean
-101 (C → T)	β^+	Italian, Eastern Mediterranean

Mutations	Type	Distribution
5'UTR		
Cap site		
+ 1 (A → C)	β^+	Asian Indian
+ 40 to +43 (-AAAC)	β^+	Chinese
2. RNA processing		
Splice site junction mutations		
IVS-1 position 1 (G → A)	β^0	Mediterranean
IVS-1 position 1 (G → T)	β^0	Asian Indian, Chinese
IVS-1 position 2 (T → G)	β^0	Tunisian
IVS-1 position 2 (T → C)	β^0	American black
IVS-1 position 2 (T → A)	β^0	Algerian
IVS-1 position 130 (G → A)	β^0	Egyptian
IVS-2 position 1 (G → A)	β^0	Mediterranean
IVS-2 position 1 (G → C)	β^0	Iranian
IVS-2 position 849 (A → G)	β^0	American black
IVS-2 position 849 (A → C)	β^0	American black
IVS-2 position 850 (G → C)	β^0	Yugoslavian
IVS-2 position 850 (-G)	β^0	Italian
Consensus sequence mutations		
IVS-1 position 5 (G → C)	β^+	Asian Indian
IVS-1 position 5 (G → T)	β^+	Mediterranean
IVS-1 position 5 (G → A)	β^+	Greek
IVS-1 position 6 (T → C)	β^+	Mediterranean
IVS-1 position 128 (T → G)	β^+	Saudi Arabian
IVS-2 positions 4-5 (-AG)	β^0	Portuguese
IVS-2 position 848 (C → G)	β^+	Japanese

Mutations	Type	Distribution
IVS-2 position 844 (C → G)	β^+	Italian
IVS-2 position 837 (T → G)	β^+	Asian Indian
Cryptic splice sites in introns		
IVS-1 position 110 (G → A)	β^+	Mediterranean
IVS-1 position 116 (T → G)	β^0	Mediterranean
IVS-2 position 654 (C → T)	β^0	Chinese
IVS-2 position 705 (T → G)	β^0	Mediterranean
IVS-2 position 745 (C → G)	β^+	Mediterranean
Cryptic splice sites in exons		
Codon 19 (A → G) Hb Malay	β^+	Malaysian
Codon 24 (T → A)	β^+	American black
Codon 26 (G → A) Hb E	β^+	Japanese, Southeast Asian
Codon 27 (G → T) Hb Knossos	β^+	Mediterranean
Cleavage and polyadenylation mutations		
AATAAA → AACAAA	β^+	American black
AATAAA → AATAAG	β^+	Kurdish
AATAAA → AATAGA	β^+	Malaysian
AATAAA → AATGAA	β^+	Mediterranean
AATAAA → A(-AATAA)	β^+	Arab
AATAAA → AAAA(-AT)	β^+	French

Mutations	Type	Distribution
3. RNA translation		
Initiation codon mutations		
ATG → AGG	β^0	Korean, Chinese, Japanese
ATG → ACG	β^0	Yugoslavian, Swiss
ATG → GTG	β^0	Japanese
ATG → ATT	β^0	White
Nonsense mutations		
Codon 15 (TGG → TAG)	β^0	Asian Indian
Codon 15 (TGG → TGA)	β^0	Portuguese
Codon 17 (A → T)	β^0	Chinese
Codon 22 (G → T)	β^0	Reunion Islander
Codon 26 (G → T)	β^0	Thai
Codon 35 (C → A)	β^0	Thai
Codon 37 (G → A)	β^0	Saudi Arabian
Codon 39 (C → T)	β^0	Mediterranean, European
Codon 43 (G → T)	β^0	Chinese
Codon 61 (A → T)	β^0	American black
Codon 90 (G → T)	β^0	Japanese
Codon 112 (T → A)	β^0	Czechoslovakian
Codon 121 (G → T)	β^+	Polish, Swiss, British, Japanese
Codon 127 (C → T)	β^0	English
Frameshift mutations		
Codon 1 (-G)	β^0	Mediterranean
Codon 5 (-CT)	β^0	Mediterranean
Codon 6 (-A)	β^0	Mediterranean, American black

Mutations	Type	Distribution
Codon 8 (-AA)	β^0	Turkish
Codons 8/9 (+G)	β^0	Asian Indian, Mediterranean
Codon 11 (-T)	β^0	Mexican
Codons 14/15 (+G)	β^0	Chinese
Codon 15 (-T)	β^0	Asian Indian, Thai
Codon 16 (-C)	β^0	Asian Indian
Codons 20/21 (+G)	β^0	Ashkenazi Jewish
Codon 24 (-G, +CAC)	β^0	Egyptian
Codons 25/26 (+T)	β^0	Tunisian
Codons 27/28 (+C)	β^0	Chinese
Codon 35 (-C)	β^0	Malaysian
Codons 36/37 (-T)	β^0	Kurdish (Iranian)
Codons 37-39 (-GACCCAG)	β^0	Turkish
Codons 38/39 (-C)	β^0	Czechoslovakian
Codon 41 (-C)	β^0	Thai
Codons 41/42 (-TTCT)	β^0	Asian Indian, Chinese, Southeast Asia
Codon 44 (-C)	β^0	Kurdish
Codon 47 (+A)	β^0	Surinamese black
Codon 54 (+G)	β^0	Japanese
Codon 64 (-G)	β^0	Swiss
Codon 71 (+T)	β^0	Chinese
Codons 71/72 (+A)	β^0	Chinese
Codons 74/75 (-C)	β^0	Turkish
Codon 76 (-C)	β^0	Italian
Codons 82/83 (-G)	β^0	Azerbaijani

Mutations	Type	Distribution
Codon 88 (+T)	β^0	Asian Indian
Codon 94 (+TG) β Agnana	β^+	Italian
Codon 95 (+A)	β^0	Thai
Codons 106/107 (+G)	β^0	American black
Codons 109/110 (-G) β Manhattan	β^+	Lithuanian
+1 codon 114 (-CT, +G) β Geneva	β^+	French
Codon 123 (-A) β Makabe	β^+	Japanese
Codons 123-125 (-ACCCCACC) β Khon Kaen	β^0	Thai
Codon 126 (-T) β Vercelli	β^+	Italian
Codons 127/128 (-AGG) β Gunma	β^+	Japanese
Codons 128/129 (-4, +5) (-GCTG, +CCACA) and codons 132-135 (-11) (-AAAGTGGTGGC)	β^+	Irish
Codons 134-137 (-10, +4) (-TGGCTGGTGT, +GCAC)	β^0	Portuguese

1.7 Clinical classification of β-thalassemia

The β-thalassemia can be classified on the clinical ground. The principle for this classification involves several clinical and hematological aspects of the patients. The β-thalassemia can be clinically divided into 3 group including β-thalassemia minor, β-thalassemia intermedia and β- thalassemia major (Weatherall and Clegg 2001).

1.7.1 β-thalassemia minor or thalassemia trait or heterozygous β-thalassemia

These terms are used interchangeably for people who have small red cells and mild or no anemia due to thalassemia. These patients are clinically well, and are usually only detected through routine blood testing. Hemoglobin value of patients with β-thalassemia trait are usually in the range of 8-10 g/dl (Weatherall and Clegg 1981). Physicians often mistakenly diagnose iron deficiency in people with thalassemia trait. Iron replacement dose not correct the condition. Most of the common forms of β-thalassemia trait are presented with increased levels of Hb A_2. The primary caution for people with β-thalassemia trait involves the possible problems that their children could be affected with β-thalassemia major if their partner also has β-thalassemia trait.

1.7.2 β-thalassemia intermedia

β-thalassemia intermedia always present with a broad spectrum of clinical picture ranging from a condition little milder than thalassemia major to symptomless disorder that is discovered only by chance examination of the blood (Modell and Berdoukas 1984). The majority of β-thalassemia intermedia is the part of compound heterozygote between β-thalassemia heterozygote and β-thalassemia variant; Hb E. The clinical features are characterized by significant anemia, spenomegaly and in the more severe forms, bone changes similar to the major form of the illness. The patient has β-thalassemia intermedia that is more severe than β-thalassemia trait, but not so severe as to require chronic transfusion as do the patients with β-thalassemia major. The hemoglobin level settles down to 7-10 g/dl (Ho, *et al* 1988).

1.7.3 β-thalassemia major

β-thalassemia major is the most severe form of β-thalassemia still sometimes called Cooley's anemia. Genotypically, it usually results from the compund hetrozygous state for two different β-globin gene mutation, from the homozygous state for the same mutation or compound heterozygote for β-thalassemia mutation including Hb E. The onset of the clinical symptom is usually before 2 years old. The β-thalassemia major present with severe anemia, early growth retardation, icteric and splenomegaly when they become older. Progressive expansion of bone marrow in response to anemia leads to the classical thalassemia facies. A steady-state hemoglobin level is less than 6 g/dl (Weatherall and Clegg 1981). Blood transfucions are needed for the β-thalassemia major. Transfusions ultimately produce iron overload and sometime present brown pigmentation of the skin. Chelation therapy, usually with the iron-binding agent is needed to prevent death from iron-mediated organ injury.

1.8 Distribution and frequency of β-thalassemia

Thalassemia has a high incidence in a broad region extending from the Mediterranean basin and part of Africa, through the Middle East, the Indian subcontinent, Southeast Asia, Malanesia and out into the Pacific (Figure 1.7). For β-thalassemia, which is a unique single-base mutations. These mutations are, as proposed, a natural selective advantage for malarial infection in the past thraughout the endemic regions of the world. In Thailand, β-thalassemia has a high incidence and it is a serious health problem. The overall incidence of β-thalassemia is 4.8 %, with the highest incidence of ~11 % in northern Thailand (Flatz, *et al* 1985).

Figure 1.7 Distribution of β-thalassemia across the world. The pale color represents regions where the β-thalassemia is common (Weatherall and Clegg 1981).

1.9 Making diagnosis for β-thalassemia/ hemoglobinopathies

The diagnosis of thalassemia and hemoglobinopathies involves 3 crucial steps which are :

a) History retrieve : This is to obtain the information of disease transmission in the patient's family. This step is normally performed in the clinic by physicians.

b) Physical examination : This step is also carried out by the clinicians to evaluate the clinical symptom of the patients. Most of the thalassemia disease can be diagnosed in this step. However, specific type of the β-thalassemia and hemoglobinopathies, particularly in those with heterozygous state, cannot be identified just only by eary physical examination. Thus, the third step is required.

c) Laboratory investigations.

1.9.1 Laboratory Investigations of β-thalassemia

There are two steps for detection of β-thalassemia; Screening and Diagnotic tests.

1.9.1.1 Screening Tests

The screening test are the tests used to determine characteristics of the red blood cells. The screenig tests for β-thalassemia include One Tube Osmotic Fragility Test (OFT), red blood indicies (Mean Corpuscular Volumn; MCV, Mean Corpuscular Hemoglobin; MCH), Dichlorophenol Indolphenol (DCIP) Precipitation test and Hb E-screenig test (Gomber, *et al* 1997, Raghaven, *et al* 1991, Sanguansermsri, *et al* 1998). These methods aim to exclude normal red blood cells from hypochromic microcytic red blood cells. In β-thalassemia (homozygote, heterozygote and β-thalassemia variant, e.g Hb E) the results of all these tests are ultimately positive whereas those generated from normal erythrocytes are negative.

1.9.1.2 Diagnostic Tests

Diagnostic tests comprise of hemoglobin studies and DNA analysis. These methods aim to determine the specific type of the β-thalassemia encountered in the patients.

a) Hemoglobin (Hb) studies

The Hb studies for making the diagnosis of β-thalassemia and hemoglobinopathies include; 1) idenfification of hemoglobin types in blood samples by cellulose acetate electrophoresis at alkali medium (pH 8.2-8.6) or by weak cation exchange high performance liquid chromatography (HPLC), 2) quantification of Hb A_2 or E by anion exchange microcolumn chromatography or by cellulose acetate electrophoresis with elution and 3) quantification of Hb F by Betke alkali denaturation test (Hb F). For instance, Hb pattern of A_2F is observed in those homozygous for

β^0-thalassemia while EF is seen in those compound heterozygous for β^0-thalassemia and Hb F alleles, Hb A_2 level of more than 3.5 % but less than 10 % is the diagnostic criterion for β-thalassemia heterozygote. Hb F level is slightly to modestly increased in heterozygous β-thalassemia. However, in β-thalassemia disease, its level ranges from modestly to markedly high.

b) DNA analysis

The aim for characterization of thalassemia mutations at the DNA level has led to the application of various molecular techniques. In β-thalassemia, which point mutations and small deletions are the major causes, the PCR-based method is the best for mutations characterization. These include Allele-Specific Oligonucleotide (ASO) hybridization (Saiki, *et al* 1988), Reverse Dot Blot (RDB) Hybridization (Maggio, *et al* 1993, Sutcharitchan, *et al* 1995), Amaplification-Created Restriction Sites (ACRS) (Saiki, *et al* 1985) and Amplification Refractory Mutation System (ARMS) (Newton, *et al* 1989, Old, *et al* 1990). The most convenient source of DNA is blood obtained by venepuncture. However, other sources of DNA may be amniotic fluid cells obtained by amniocentesis at between 14 to 20 weeks of gestation, chrorionic villi obtained by ultrasound-guided transcervical or transabdominal sampling at between 8 to 10 weeks of gestation and fetal blood sampling obtained by cordocentesis at between 18 to 20 weeks of gestation. The hybridization of allele-specific oligonucleotide (ASO) probes to amplified genomic DNA bound to a nylon membrane in the form of dot was the first PCR-based method to be developed (Saiki 1988). The method is based on the use of two oligonucleotide probes for each mutation, one complementaty to the mutant DNA sequence and another to the wild-type DNA sequence at the same position. The probes are usually 5' end-labelled with either ^{32}P, biotin or horse radish peroxidase (HRP). The genotype of the DNA sample is determined by the presence or absence of the hybridization signal of both the mutation-specific and normal probes (Ristaldi, *et al* 1989). Like the ASO, RBD is also based on the use of two oligonucleotide probes and

the β-thalassemia is identified by presence or absence of hybridization signals generated from either the mutation-specific or normal probes or both. However, in RDB the oligonucleotide probes are bound to the nylon membrane instead, thus, making identification of more than one mutation can be undertaken simultaneously. ACRS uses primers that are designed to insert new bases into the amplified products in order to create a restriction-enzyme recognition site adjacent to the mutation sequence. This permits known mutations that normally do not alter a recognition site to be detected by restriction-enzyme digestion of the PCR products (Lindeman, *et al* 1991).

Amplification Refractory Mutation System (ARMS)

ARMS is also known as allele-specific PCR (AS-PCR) or PCR amplification of specific allele (PASA). It was first described by Newton and co-workers in 1989 to determine α-1-antitrypsin gene (Newton, *et al* 1989). The principle of this method involves amplification of target DNA using a common primer and either of two allele-specific primers, one complementary to the mutation (mutant primer) and another complementary to normal DNA at the same position (normal primer). A second pair of primers complementary to a different part of the genome is included in the PCR to amplify a fragment simultaneously in order to determine the success or failure of the amplification step. The method provides a quick screening means which dose not require any form of labelling, as the amplified products are visualized simply by agarose gel electrophoresis and ethidium bromide staining. In the β-thalassemia diagnostic field, Old, J. M. was the first to introduce this technique to identify the β-thalassemia mutation. Thereafter, the use of ARMS for β-thalassemia diagnosis has been increased rapidly owing to its simple technical steps (Old, *et al* 1990).

1.10 Prenatal diagnosis of β-thalassemia

The aim of prenatal diagnosis is to provide an accurate and rapid result as early in the pregnancy as possible. The first prenatal diagnosis for β-thalassemia was initially applied over 20 years ago through analysis at the protein level of fetal blood obtained by fetoscopy of placental aspiration (Clegg, *et al* 1966, Kan, *et al* 1975, Weatherall and Clegg 1981). As recombinant DNA technology evolved, DNA analysis become feasible (Kan and Dozy 1978, Kan, *et al* 1976, Orkin, *et al* 1978), but prenatal diagnosis was not widely applicable. Partly because available methods were cumbersome and time consuming, and partly because knowledge of the molecular basis of these disorders was limited. The discovery of the polymerase chain reaction (PCR) in the mid 1980s (Saiki 1988) led to the development of numerous methods for mutation analysis and subsequently to an explosion of information relative to the molecular genetics of the hemoglobinopathies. This information and many PCR-based analytical protocols form the basis of prenatal diagnostic strategies today.

1.11 Background of the study

The β-thalassemia and β-hemoglobinophathies are the serious health problem in Thailand. There are ~6,250 patients with homozygote β-thalassemia, 97,500 patients with β-thalassemia/Hb E, 2,500 high risk couples/year for homozygote β-thalassemia and 13,000 high risk couples/year for β-thalassemia/Hb E. The prediction of birth with homozygote β-thalassemia is about 625 cases/year and that for β-thalassemia/Hb E is about 3,250 cases/year (Flatz, *et al* 1985). The average therapeutic cost for homozygote β-thalassemia is about 10,000 bath/person/month and about 4,950 bath/person/month for β-thalassemia/Hb E disease (Thai committee for hemotologic diseases 1990) . Thus, the government has launched the prevention and control scheme for these diseases. This policy comprises of four main programs; public education, carrier or heterozygote detection, genetic counselling and prenatal diagnosis (PND) (World Health Organization 1989).

The prevention and control of β-thalassemia disease at Maharaj Nakorn Chaing Mai Hospital involve two steps including selection of high risk couples and PND. The selection of high risk couples for homozygote β-thalassemia and β-thalassemia/Hb E disease is based on the results generated by the following tests which are Osmotic Fragility Test (OFT), Hb E-screening test, Microcolumn chromoatography for Hb A_2/E and /or High Performance Liquid Chromatography (HPLC). Identificatiion of high risk couples for β-thalassemia is based on the value of Hb A_2 as demonstrated in Table 1.3.

Teble 1.3 Identification criterion for high risk couples of β-thalassemia and hemoglobinopathies

% Hb A_2 of the couples	Diagnosis
<3.5 % in both	No risk
4.0 –9.0 % in both	High risk for homozygote β-thalassemia
>10% and 4.0-9.0 % in each or *vice versa*	High risk for β-thalasemia/Hb E disease

The PND in high risk couples is carried out by two methods. First, cordocentesis; this is to cellect fetal blood samples at 18 - 20 weeks gestation. These samples will be subsequenthy subjected for hemoglobin identification by HPLC to differentiate β-thalassemia disease (Hb A absent) from non-β-thalassemiadisease (Hb A present). Second, DNA studies; this category is performed entirely by directs nucleotide sequencing from PCR products. Sources of DNA could be amminocytes obtained by amniocentesis at between 14 to 20 weeks of gestation, chorionic villi obtained by ultrasound-guided transcervical or transabdominal sampling at between 8 to 10 weeks of gestation and fetal blood sample (Tammy 1994).

The hemoglobin analysis by HPLC is rapid, simple and cheap. However, it can only be carried out at late pregnancy. In addition, despite the direct nucleotide sequencing is able to detect known and unknown mutations of β-globin gene at early pregnancy, it has been realized to be very costly and sophisticate.

PCR-based protocols include those that are relatively simple and others that are technically challanging. Amplification Refractory Mutation System (ARMS) PCR is the PCR-based method that has been used to detect point mutations and small deletion of β-globin gene. However, the standard ARMS can only detect one mutation in one

reaction of PCR. The multiplex ARMS, in contrast, is able to identify more than one mutation in a PCR reaction and might be useful in prenatal diagnosis.

1.12 Opjectives

1. To optimize the single ARMS-PCR for detection of β-thalassemia mutations including codon 17 (A-T), codons 41/42 (-TTCT), codons 71/72 (+A), IVS I-nt 1 (G-T) and Hb E
2. To optimize the multiplex ARMS-PCR for detection of β-thalassemia mutations for simultaneous detection of codon 17 (A-T) + codons 41/42 (-TTCT), codon 17 (A-T) + Hb E, codons 41/42 (-TTCT) + Hb E and codons 41/42 (-TTCT) + codons 71/72 (+A) + codon 17.
3. To evaluate the optimal numbers of white blood cells for ARMS-PCR
4. To assess the application potential of ARMS-PCR technique in the β-thalassemia heterozygote screening
5. To assess the application potential of ARMS-PCR technique in HbE screening
6. To assess the application potential of ARMS-PCR for prenatal diagnosis (PND)
7. To access the applicability of ARMS-PCR and nucleotide sequencing for prenatal diagnosis of β-thalassemia major and β-thalassemia/Hb E disease