I. INTRODUCTION

I.1 Statement and Significance of the Problem

The clinical disorders associated with severe α–thalassemia range from the moderately severe to transfusion-dependent anemia characteristic of hemoglobin H (Hb H) disease to the lethal Hb Bart's hydrops fetalis syndrome (Bunn and Forget, 1986 ; Weatherall and Clegg, 1981 ; Kattamis *et al.*, 1988). The risk of inheriting one of these conditions is substantial in certain parts of Southeast Asia where the prevalence of α–thalassemia may be more than 40 % (Fucharoen and Winichagoon, 1987 ; Na-Nakorn and Wasi, 1970 ; Sicard *et al.*, 1979). In these groups, reliable diagnosis, genetic counseling and prenatal diagnosis, are significant health care issues. Methods that quantify the number of α–globin genes diagnose accurately the common deletional form of α–thalassemia (Bunn and Forget, 1986 ; Weatherall and Clegg, 1981). However, these methods predictably fail to identify nondeletional α–thalassemia genes which cause syndromes of greater clinical severity than those caused by deletions alone (Kattamis *et al.*, 1988 ; Chan *et al.*, 1985 ; Trent *et al.*, 1986).

The most prevalent variety of nondeletional α–thalassemia is Hb Constant Spring (Fucharoen and Winichagoon, 1987; Sicard *et al.*, 1979). It is caused by a single base mutation at the termination codon of the α_2-globin gene (TAA--->CAA). Therefore, the translation of α^{CS}- mRNA proceeds through this codon until the next stop codon is reached, resulting

in the 172 amino acids-globin. The α^{CS}-mRNA has found to be unstable, which is mediated by translation into the 3' noncoding region (Hunt *et al.*, 1982 ; Weis and Liebhaber,1990). Thus, the expression of the α^{CS} gene is low, resulting to a thalassemia phenotype (Steinberg and Adams, 1983 ; Clegg and Weatherall , 1974).

Hb Constant Spring was first observed on conventional hemoglobin electrophoresis (Weatherall and Clegg, 1981). Because it is unstable, its amount is small in heterozygote (< 1%) and it moves very close to the position of Hb A_2 during alkali electrophoresis, it is frequently not possible to identify this abnormal band (Pootrakul *et al.*, 1981). This has led to the development of several DNA-based diagnostic approaches. However, the methods of allele specific hybridization either with restriction fragment of genomic DNA (Kosasih *et al.*, 1989; Laig *et al.*, 1990) or with α_2-specific PCR product (Kropp *et al.*, 1989; Hsie *et al.*, 1989) needed a technical skill and involved in radioisotope labeling system, which is hazard and difficult to handle. Although nonradioisotope labeling system was also modified to use (Fucharoen *et al.*, 1989), it cannot overcome the trouble involved in hybridization method and may even get more trouble from the less sensitive labeling system. Another method presented by Kropp *et al.* (1991), the one step polymerase chain reaction which accomplished dual effects of the α_2-specific amplification and the discrimination of normal and Hb CS gene by allele specific fluorescent primers, was dubious. Moreover, it needed fluorescent labeling and

fluorescent detection system, which are not normally found in common laboratory.

In this study, the simple method of polymerase chain reaction (PCR) and restriction fragment length polymorphism (RFLP) techniques were applied to detect the abnormal termination codon of α_2-globin gene and Hb CS gene in northern Thailand populations and in nondeletional Hb H disease patients. The method involved two steps of detection. The first step was to detect the normal termination codon from the α_2-specific PCR product with restriction enzyme Mse I. The second step involved a semi-nested PCR with a special primer that could create restriction site for detection of Hb CS gene. This newly modified method is simple, reliable and suitable for population screening and routine diagnosis.

I.2 Literature Review

I.2.1 The Hemoglobin Tetramer

The primary function of the erythrocyte is transport of oxygen from the lungs to the peripheral tissues. This function is dependent upon hemoglobin which normally constitutes over 90 % of the soluble protein in the red cell. The hemoglobin molecule consists of two α– or α–like globin chains and two β– or β-like globin chains , each containing a heme - Fe^{++} prosthetic group which can reversibly bind oxygen. Allosteric interactions among the four globin chains result in the physiologically appropriate sigmoid oxygen binding curve and in adjustments in the oxygen binding affinity to local tissue pH (Bohr effect), temperature, pCO_2 and levels of erythrocyte 2,3 diphosphoglycerate (2,3 DPG) (Bunn and Forget, 1986). The function of the erythrocyte is therefore dependent upon balanced synthesis of α–and β- globin chains and their subsequent assembly into the functional hemoglobin tetramer (Liebhaber, 1989).

I.2.2 Balanced Expression of the Human Globin Genes

The α–related globin chains and the β-related globin chains are encoded in two separate gene clusters located on chromosome 16 (Deiseroth *et al.*, 1977) and 11 (Deiseroth *et al.*, 1978) , respectively. During embryological development there is an ordered switching of globin chain synthesis (Bunn and Forget, 1986). The β-globin gene cluster encodes

globin subunits specific to the embryonic ,fetal, and adult development stages ; ε- , γ- , and β- globins, respectively. In contrast, the α–globin gene cluster encodes only two globin proteins; ζ gene expresses during embryonic period and α gene expresses during fetal-adult period. Throughout development a balance between α–like and β-like globin gene expression must be maintained so that functional hemoglobin tetramers can be assembled. The defect in the thalassemias is the loss of this balance (Weatherall *et al.*, 1965). In β-thalassemia, a relative deficiency of β-chains results in the accumulation of unpaired α-globin chains (Clegg and Weatherall, 1967). The α–globin chains form insoluble aggregates which can disrupt the erythrocyte membrane (Nathan and Gunn, 1966) and may have direct toxic effects upon the developing erythroblasts (Rachmilewitz *et al.*, 1976). In α-thalassemia, a relative deficiency in α–chains results in the accumulation of excess of γ- or β-globin (Clegg and Weatherall, 1967). Excess γ– and β– chains present in the newborn and adult can self assemble into Hb Bart's (γ_4) and Hb H (β_4) homotetramers, respectively. Hb Bart's and Hb H are significantly more stable than the α–globin aggregates, and do not cause the intramedullary hemolysis seen in β–thalassemia. Instead, they precipitate as the red cell ages resulting in red cell damage and premature clearance by the reticuloendothelial system, thus shortening the red cell life span (Nathan *et al.*, 1966; 1969). The phenotype of thalassemia therefore reflects a deficiency of normal functional hemoglobin tetramers and the accumulation of excess γ– and β-globin resulting in erythrocyte damage and dysfunction (Liebhaber, 1989).

I.2.3 Structure and Expression of Genes on α–Globin Cluster

The α–globin gene cluster is located adjacent to the 16p telomere (16 p 13.3 - pter) (Nicholls *et al.*, 1987). The 29 Kb cluster contains a total of seven genes and pseudogenes, as shown in Fig. 1.

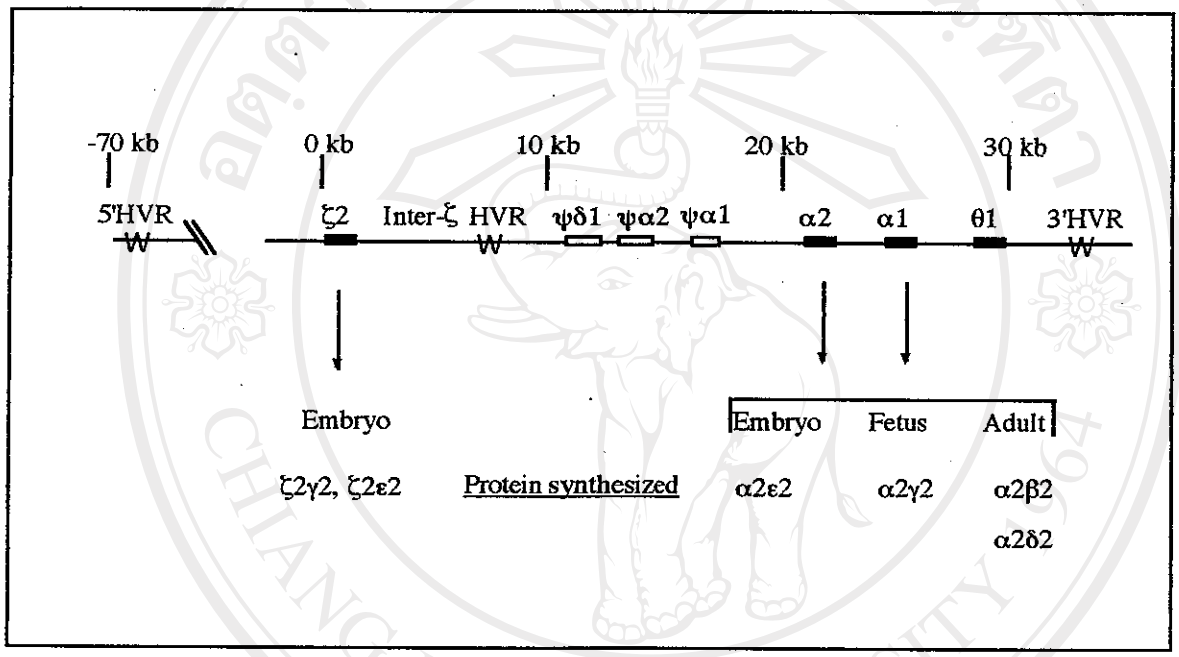


Figure. 1 The Organization of the α–Globin Complex

Filled box indicate functional genes and open boxes pseudogenes. Position 0 represents the α–globin mRNA CAP site. Hypervariable regions are denoted by zig-zag lines. The hemoglobins synthesized at each stage of development are indicated below the complex. Each molecule of hemoglobin comprises a tetramer of two α–like (α or ζ) and two β-like (β, γ, δ or ε). (According to Higgs *et al.*, 1989)

There are four functional genes in the α–globin gene cluster. The ζ-globin gene expresses during the first five weeks of in utero development by the primitive erythroblasts of the yolk sac. Hb Gower I (ζ_2-ε_2) is the first functional hemoglobin tetramer detected in the embryo, closely followed by the appearance of Hb Portland (ζ_2-γ_2) (Huehns *et al.*, 1961 ; Kaltsoya *et al.*, 1966). During the sixth week of development erythropoiesis shifts to the liver. The transcription of the ζ-globin gene decreases at this point and the transcription of the two α–globin genes, α_1 and α_2, is activated. The two α–globin genes function during both fetal and adult life (Liebhaber and Kan, 1981).

Until recently, θ-globin gene have been clearly characterized. It is shown that θ-globin gene has similar patterns of developments to that of α–globin gene. The θ-globin mRNA shows a parallel increase at 5 to 8 weeks postconception and is expressed in cord blood and adult reticulocytes. Its encoded protein has not yet been detected in vivo, therefore, its functioned has yet to be determined (Albitar *et al.*, 1992).

I.2.4 Structure and Expression of α_1 and α_2 Globin Genes

The α_1- and α_2-globin genes are composed of three exons interrupted by two short introns. These genes are highly similar with structural divergence limited to intron 2 and exon 3. In intron 2, there are two single base substitutions and a 7 bp deletion in α_2 relative to α_1. In exon 3 there are 18 base substitutions and a single base deletion in α_2

relative to α_1, all located within the 3' nontranslated region (Liebhaber *et al.*, 1980). The comparative nucleotide sequences of α_1 and α_2 are shown in Table 1. The two mature α–globin mRNAs only diverge in structure in this 3' untranslated region and therefore encode identical α–globin proteins.

Subsequent to the structural characterization of the two α–globin mRNAs, it was possible to devise a number of assays which could distinguish the two mRNA species bases upon the structural divergence in the 3'-nontranslated regions. These studies demonstrated that α_2- globin mRNA was present at approximately 2.6 folds higher level than α_1-mRNA in the reticulocytes of normal individuals (Liebhaber *et al.*, 1981 ; 1985 ; 1986). These and related studies further demonstrated the same excess of α_2-globin mRNA in bone marrow (Liebhaber and Kan, 1981) and early fetal erythroid tissue (Orkin and Goff, 1981) suggesting that the excess in α_2-globin gene expression is transcriptional in origin and that this ratio is not subject to developmental switching (Liebhaber, 1989).

The fact that α_2-globin gene has a dominant role in α-globin synthesis was confirmed in an analysis of the α-globin mRNA and protein expression in eight individuals with distinct α-globin structural mutations at either the α_1- or α_2-globin locus (Liebhaber *et al.*, 1986). The results demonstrated that the α_2-gene encodes approximately 2.6 fold more protein than the α_1-gene, paralleling the 2.6 to 1 ratio of α_2 : α_1-mRNA levels. The parallel between the α_2 : α_1 ratio of mRNA levels and protein

Table 1 : The Comparison of Nucleotide Sequences Between the Two α-Globin Genes

```
6551  gggactcccc tgcggtccag gccgcgcccc gggctccgcg ccagccaatg
10362 .......... .......... .......... .......... ..........

6601  agcgccgccc ggccgggcgt gcccccgcgc cccaagcata aaccctggcg
10412 .......... .......... .......... .......... ..........

6651  cgctcgcggc ccggcACTCT TCTGGTCCCC ACAGACTCAG AGAGAACCCA
10462 .......... .......... .......... .......... ..........

6701  CCATGGTGCT GTCTCCTGCC GACAAGACCA ACGTCAAGGC CGCCTGGGGT
10512 .......... .......... .......... .......... ..........

6751  AAGGTCGGCG CGCACGCTGG CGAGTATGGT GCGGAGGCCC TGGAGAGgtg
10562 .......... .......... .......... .......... ..........

6801  aggctccctc ccctgctccg acccgggctc ctcgcccgcc cggacccaca
10612 .......... .......... .......... .......... ..........

6851  ggccaccctc aactctcctg gccccggacc caaaccccac ccctcactct
10662 .......... .......... .......... .......... ..........

6901  gcttctcccc gcagGATGTT CCTGTCCTTC CCCACCACCA AGACCTACTT
10712 .......... .......... .......... .......... ..........

6951  CCCGCACTTC GACCTGAGCC ACGGCTCTGC CCAGGTTAAG GGCCACGGCA
10762 .......... .......... .......... .......... ..........

7001  AGAAGGTGGC CGACGCCCTG ACCAACGCCG TGGCGCACGT GGACGACATG
10812 .......... .......... .......... .......... ..........
```

```
7051   CCCAACGCGC TGTCCGCCCT GAGCGACCTG CACGCGCACA AGCTTCGGGT
10862  .......... .......... .......... .......... ..........

7101   CCCAACGCGC AACTTCAAGg tgagcggcgg gccgggagcg atctgggtcg
10912  .......... .......... .......... .......... ..........

7151   aggggcgaga tggcgccttc ctctcagggc agaggatcac gcgggttgcg
10961  .......... .......... ...g...... .......... ..........

7201   ggaggtgtag cgcaggcgga ggctgcgggc ctgggccgca ctgaccctct
11012  .......... .......... .......... .......^.. ..........
                                                 ctcggccc
7251   tctctgcaca gCTCCTAAGC CACTGCCTGC TGGTGACCCT GGCCGCCCAC
11069  .......... .......... .......... .......... ..........

7301   CTCCCCGCCG AGTTCACCCC TGCGGTGCAC GCCTCCCTGG ACAAGTTCCT
11119  .......... .......... .......... .......... ..........

7351   GGCTTCTGTG AGCACCGTGC TGACCTCCAA ATACCGTTAA GCTGGAGCCT
11169  .......... .......... .......... .......... ..........

7401   CGGTAGCCGT TCCTCCTGCC CGATGGGCCT CCCAACGGGC CCTGGTCCCC
11219  ....G...A. G.T..T.... .CT........ ...CC.A.C. ..........

7451   TCCTTGCACC GGCCCTTCCT GGTCTTTGAA TAAAGTCTGA GTGGGCGGCa
11269  .T.C...... C.TA.CC..GT.......... .......... ..........
```

Note : - The sequences obtained from Gene Bank locus HUMHBA4. The upper rows are the sequences of α_2 globin gene and the lower rows are the sequences of α_1 globin genes. Dot represents the same sequences as above. Capital letter represents exon.

expression was further substantiated by demonstrating that the two α-globin mRNAs are translated with equal efficiencies (Shakin and Liebhaber, 1986).

I.2.5 The α–Thalassemia and Its Molecular Defects

The α–thalassemia is a hemolytic anemia resulting from deficient synthesis of α–globin. The deficiency of α–globin results in insufficient production of functional hemoglobin and in the accumulation of unstable Hb Bart's or Hb H tetramers with consequent accelerated red cell destruction. α–thalassemia can be inherited or acquired and can result from defect(s) in, or deletion(s) of, one or more of the four α–globin genes (Liebhaber, 1989).

The great majority of α–thalassemias result from deletion of one (-α/; α–thalassemia 2) or both α-globin genes (--/; α–thalassemia 1) (Higgs and Weatherall, 1983). However, a number of nondeletion defects have clearly been etiologic in α–thalassemia.

There are two major subtypes of the single α–globin gene deletion. The most common results in the loss of 3.7 Kb of DNA and is commonly noted as the $\alpha^{3.7}$ deletion. A less common deletion results from the loss of 4.2 Kb of DNA and is noted as the $\alpha^{4.2}$ deletion. Both of these deletions results from unequal homologous recombination within the α–globin cluster. The two α–globin genes are imbedded in a large region of

homology which is divided by short divergent regions into three homology subsegments ; x, y and z (Lauer *et al.*, 1980). These regions of homology within and surrounding the two α–globin genes can mediate homologous unequal recombination. The crossover can occur between the two x-regions or the two z-regions resulting in the 4.2 Kb and 3.7 Kb deletions (as shown in Fig. 2), respectively. Because of the relative position of these crossovers, $\alpha^{3.7}$ is often referred to as the rightward deletion and the $\alpha^{4.2}$ as the leftward deletion.

The $\alpha^{3.7}$ deletion can be further subdivided based on the exact position of crossover within the z box (Higgs *et al.*, 1984). Type I occurs within the large region of homology extending from 863 nucleotide 5' of the two α–globin genes to the 7 base insertion/deletion divergence within the second intron. This is by far the most common form. Type II and III crossovers occur within the homology blocks extending between the second intron and the divergent segment in the 3'-nontranslated region of the third exon and in the small segment of homology surrounding the poly A addition site, respectively.

The $\alpha^{3.7}$ Type I deletion is extremely widespread. The gene frequency of the $\alpha^{3.7}$ mutation can reach very high levels. In the Mediterranean Basin the prevalence of the heterozygous state of $\alpha^{3.7}$ is approximately 5-10 % (Kanavakis *et al.*, 1988), in Southeast Asia approximately 10-20 % (Nicholls *et al.*, 1987), in certain region of West Africa the frequency approximates 20-30 % (Dozy *et al.*, 1979), and in

specific areas of India and Papua New Guinea the incidence is as high as 90% (Oppenheimer *et al.*, 1984 ; Fodde *et al.*, 1988). The $\alpha^{4.2}$ is most frequently found in Asian populations.

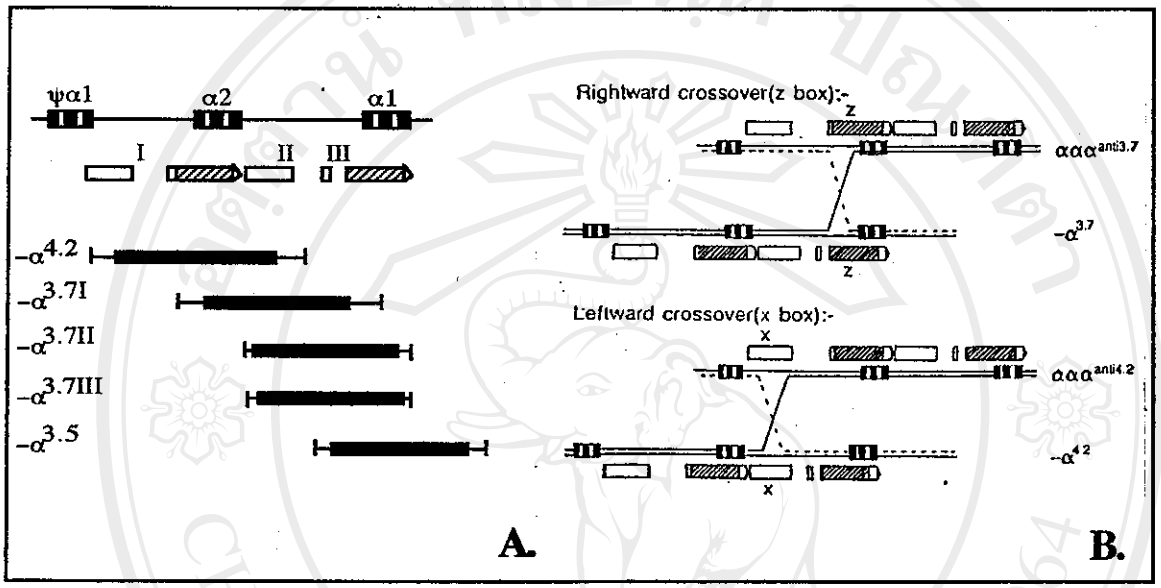


Figure 2 : The Single α-Globin Gene Deletion and Its Proposed Mechanisms (According to Higgs *et al.*, 1989)

A. The duplicated x,y and z box arrangement containing the α-globin genes. Nonhomologous regions (I, II, and III) are indicated. The extend of each deletion is indicated by the solid boxes and the limits of the break points are represented by solid lines.

B. The proposed crossovers between the homologous regions that lead to the $\alpha^{-3.7}$ and $\alpha^{-4.2}$.

In addition to a single α–globin gene deletion, a variety of more extensive deletions occur within the α–globin gene cluster which adversely affect α–globin gene expression. These deletions range in size from rather small (5.2 Kb) to those which remove the entire cluster (as shown in Fig.3). The two most common deletions, (-- SEA) and (-- Med) , occur in Southeast Asia and the Mediterranean Basin, respectively. These two deletions remove both α–globin genes but spare the functional ζ gene (Nicholls *et al.*, 1987).

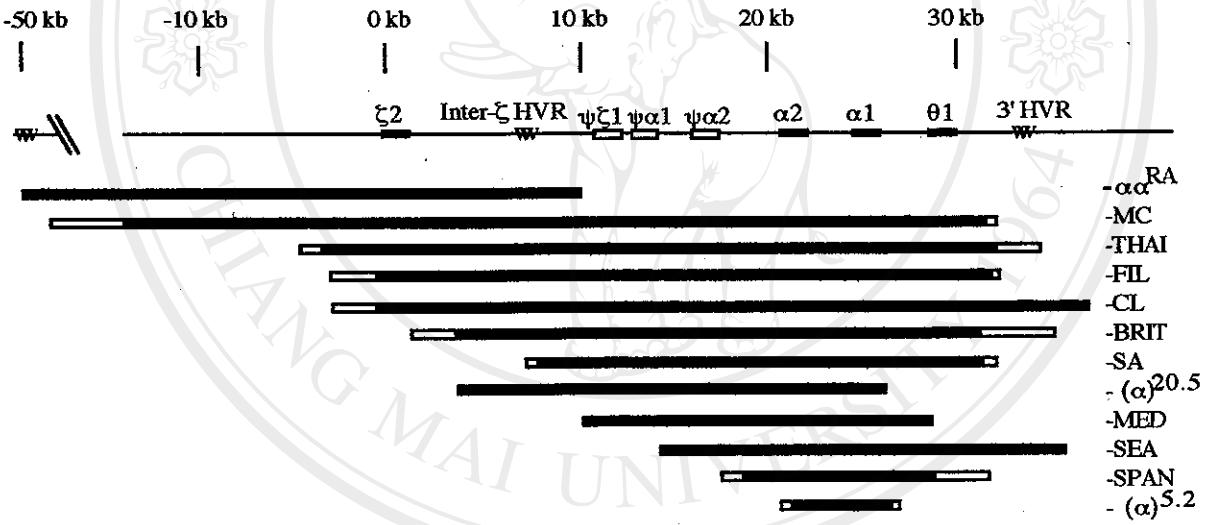


Figure 3 : The Deletion of Both α-Globin Genes

The extent of each deletion is represented as a solid box and the uncertainty of the breakpoints is indicated by the open boxes. (According to Higgs *et al.*, 1989)

A number of mutations within the α–globin have been clearly etiologic in α–thalassemia. This nondeletion category is included both single base substitutions as well as very small deletions. This is summarized in Table 2.

Table 2 : Nondeletional Mutants That Cause α-Thalassemia
(According to Higgs *et al.*, 1989)

	Affected Gene	Affected Sequence	Mutation	Geographical Distribution	comments
RNA	α2	IVS I donor site	GAGGTGAGG->GAGG	Mediterranean	aberrant splicing
Processing	α2	poly(A) signal	AATAA->AATAAG	Mediterranean, Middle East	↓ efficiency of 3'end processing
	α2	Initiation codon	ACCATGG->ACCACGG	Mediterranean	↓ mRNA translation
	α1	Initiation codon	ACCATGG->ACCGTGG	Mediterranean	↓ mRNA translation
	–α	Initiation codon	ACCATGG->ACCGTGG	Black	↓ mRNA translation
	$-\alpha^{3.7II}$	Initiation codon	ACCATGG->AC-CATGG	Mediterranean, N.Africa	↓ mRNA translation
	α2	Exon III	α116GAC->UAG	Black	in phase termination
	α2	Termination codon	α142 TAA->CAA	Southeast Asean	Hb CS
	α2	Termination codon	α142 TAA->AAA	Mediterranean	Hb Icaria
	α2	Termination codon	α142 TAA->TCA	Indian	Hb Koya Dora
	α2	Termination codon	α142 TAA->GAA	Black	Hb Seal Rock
	–α	Exon I	α30/31GAGAGG->GAGG	Black	reading frameshift
Posttranslational	α2	Exon III	α125 Leu->Pro	Southeast Asean	Hb Quong Sze
instability	α2	Exon III	α109 Leu->Arg	Southeast Asean	Hb Suan Dok
	α	Exon III	α110 Ala->Asp	Middle Eastern	Hb Petah Tikvah
	–α	Exon I	α14 Trp->Arg	Black	Hb Evanston
Uncharacterized	α	Unknown	Not determined	Black	
	α	Unknown	Not determined	Greek	'Karditsa' mutation

I.2.6 Determinants of α–Thalassemic Phenotypes

The severity of the α–thalassemic phenotype is directly related to the loss of α–globin synthesis. This in turn relates to three major factors:

1) the number of genes affected
2) the degree to which the specific mutation decreases the expression of the affected gene (partial or complete loss of function)
3) the degree to which the effected gene normally contributes to α–globin synthesis.

Since the α_2-globin gene is expressed at 2.6 fold higher levels than the α_1-globin gene, the defect on the α_2-globin gene results in the more severity than the defect on the α_1-globin gene. The clinical phenotype of α–thalassemia resulting from nondeletion α_2-mutations (-- /$\alpha^T\alpha$), had a significantly more severe Hb H disease than those with the α_1-mutation (--/α α^T) (Paglietti *et al.*, 1986). However, the deletion of one α–globin gene in the various (-α/) chromosome behaves like neither α_1 nor α_2 (Bowden *et al.*, 1987). It is found that the levels of α_1-globin mRNA encoded by the α_1-gene on the $\alpha^{3.7}$ deletion chromosome is 1.8-fold higher than that of the α–globin gene in the intact α–globin cluster (Liebhaber *et al*, 1985). The impact of the $\alpha^{3.7}$ deletion should therefore be intermediate between the nondeletion α_1 and the nondeletion α_2.

In summary, studies to date strongly suggest that the loss of a single α–globin gene can result in a mild, moderate, or severe loss of α–globin

synthesis depending on the mode of loss (deletion versus nondeletion) and the locus involved (α_1 versus α_2). A quantitative interpretation of these data is summarized in Table 3 (Liebhaber, 1989).

Table 3 : Predicted Impact of α–Thalassemia Mutation on α–Globin Synthesis (According to Liebhaber, 1989)

Genotype	Units of α–Globin Synthesis
$\alpha\ \alpha$	2.6 + 1.0 = 3.6
$\alpha\ \alpha^T$	2.6 + 0.0 = 2.6
$\alpha^{-3.7}$ or $\alpha^{-4.2}$	3.6 / 2 = 1.8
$\alpha^T\alpha$	0.0 +1.0 = 1.0

I.2.7 The Clinical Syndromes of α–Thalassemia

It is possible to subdivide thalassemic clinical syndromes into four categories : silent carrier , α–thalassemic trait, Hb H disease, and α–thalassemic hydrops fetalis. However, there can be significantly overlap among these groups based upon the spectrum of clinical severity within each category which reflects the variety of genetic background (as in I.2.6). It may not always be possible to categorize an individual patient in one of the groups on the basis of clinical data alone.

1. Silent Carrier (three functional α–genes): the hematological parameters such as hemoglobin concentration, red cell indices and number of red cells are within normal limit.

2. α–Thalassemic Trait (two functional α– genes) : there are mild hematological changes but no major clinical abnormality.

3. Hb H disease (one functional α–gene) : the clinical severity is considerable variability. The details of this category are given below (I.2.8)

4. Hb Bart's hydrop fetalis (no functional α–gene) : the loss of all four α–globin genes is incompatible with life, resulting in mid- to late-gestational stillbirth of an hydropic fetus. The predominant physical finding are generalized and massive edema, ascites, gross enlargement of the liver with a spleen which may be normal or only slightly enlarged, and a large friable placenta.

I.2.8 Hb H Disease

Hb H disease is the most severe of the α–thalassemia phenotypes compatible with life. This syndrome occurs most commonly in Southeast Asia and in the Mediterranean Basin. The most common genotype associated with Hb H disease is (--/-α). Other less frequent genotypes can also result in Hb H disease including coinheritance of deletion and nondeletion defects (--/ααᵀ ; --/αᵀα) or homozygosity for nondeletion defects (αᵀα/αᵀα) (Pressley *et al.*, 1980; Pagliette *et al.*, 1986 ; Kattamis *et al.*, 1988).

The heterogeneity in the genetic basis of Hb H disease appears to correlate with its clinical severity. The common deletional form of Hb H (--/-α) is rarely symptomatic. The loss of globin gene expression by a nondeletion defect (--/$\alpha\alpha^T$) is also usually quite mild (Moi *et al.*, 1987). However, the loss of the α_2 gene by nondeletions (--/$\alpha^T\alpha$) results in a more severe phenotype than the loss by deletion or by nondeletional loss of α_1 (--/$\alpha\alpha^T$) (Moi *et al.*, 1987 ; Paglietti *et al.*, 1986 ; Katamis *et al.*, 1988). This relationship between genotype and phenotype in Hb H disease severity is consistent with the greater impact of α–thalassemic mutations on the α_2 locus. In addition to deletion defects, the homozygosity of nondeletion defects on the α_2-globin gene can result in mild Hb H disease (Paglietti *et al.*, 1986).

The usual Hb H phenotype includes microcytosis, mild to moderate anemia with hemoglobin levels between 9 to 12 g/dl and splenomegaly. The red cell morphology is usually abnormal with basophilic stippling, target cells and teardrop forms. Brilliant cresyl blue supervital stain shows the Hb H precipitates as multiple small punctate inclusions , which appear as large single inclusions after splenectomy. The anemia of Hb H disease results from a deficiency in Hb A ($\alpha_2\beta_2$) and the formation of unstable Hb H homotetramers (β_4). Premature removal of red cells containing Hb H precipitates occurs predominantly in the spleen.

Hb H disease is usually not severe enough to interfere with the activities of daily living, longevity, or reproductive function. However,

severe exacerbation of the anemia may occur coincident with pregnancy, infections , and exposure to oxidant drugs. Nevertheless, the loss of the α_2-gene function by nondeletion defects (--/$\alpha^T\alpha$) resulted in a more severe symptom with some patients being transfusion dependent (Moi *et al.*, 1987; Kattamis *et al.*, 1988; Paglietti *et al.*, 1986). Two reported cases of nondeletion Hb H disease (--/$\alpha^T\alpha$) were severe enough to result in neonatal death (Hb H Hydrops Fetalis) (Chan *et al.*, 1985; Trent *et al.*, 1986).

I.2.9 The Molecular Basis of α–Thalassemia in Thailand

Thalassemia is common throughout Southeast Asia. In Thailand, the incidence is in the range of 15-30 %. It could be estimate that about 1,250 fetuses will be born with Hb Bart's hydrop fetalis and about 7,000 fetuses with Hb H disease per year, based on a total birth rate of about 1 million per year (Winichagoon *et al.*, 1992). By the molecular method of Southern blot analysis, the gene frequency of α–thalassemia 1 has been shown to be around 0.02 in every region of the country. The gene frequencies of α–thalassemias in different regions of Thailand were summarized in Table 4.

Table 4 : Gene Frequencies of α–Thalassemias in Different Regions of Thailand (According to Panich *et al.*, 1991).

Region	α–Thal 1	α–Thal 2	Hb CS
Central (Bangkok)	0.0185 [1]	0.0837 [1]	0.008 [2]
North (Chiangmai)	0.0236 [2] 0.058 [5]	0.0991 [3]	0.0331 [4]
Northeast	0.0234 [6]	0.1719 [6]	0.055 [6]
South (Songkhla)	0.0217 [7]	0.0650 [7]	0.0292 [7]

Note : α-thalassemia 1 and α–thalassemia 2 were detected by DNA methods in all studies. Hb CS was diagnosed through direct DNA analysis except in Songkhla where it was assumed that newborns with 0.5-4 % Hb Bart's without α–globin gene deletion should have Hb CS gene.

1 Tanphaichitr *et al.*, 1988. 2 Thonglairoam *et al.*, 1991. 3 Hundrieser *et al.*, 1988. 4 Laig *et al.*, 1990. 5 Makonkawkeyoon *et al.*, 1990. 6 Hundrieser *et al.*, 1990. 7 Sriroongrueng *et al.*, 1991.

I.2.10 Properties and Structure of Hb Constant Spring

Hemoglobin Constant Spring (Hb CS) is always found in relatively low quantities; it occurs in the following conditions: the homozygous state, compound heterozygous state with α-thalassemia 1 and the heterozygous state.

Hemoglobin Constant Spring is easily purified by chromatography on Amberlite IRC 50. When it is converted to globin and the constituent globin chains separated by urea chromatography, it is found that the variant consists of normal β-chains and abnormal α-chains. The parent component seems to be Hb CS_1 in which the chains is elongated by 31 amino acid residues at its C-terminal end. Hb CS_2 seems to be identical in structure except that it lacks the terminal tripeptide Val-Phe-Glu. Hb CS_3 has a shortened chain which ends at the tryptophan residue at position α 154. The most likely explanation for the production of multiple forms of hemoglobin Constant Spring is that they result from proteolytic degradation of the elongated chain by red-cell enzymes during storage and manipulation (Weatherall and Clegg, 1981). Pootrakul *et al.* (1976) have provided experimental evidence in favor of the degradation of hemoglobin Constant Spring being mediated through proteolysis.

Hemoglobin Constant Spring is usually observed in the form of two slowly migrating hemoglobin fractions on electrophoresis at alkaline pH. On starch gel electrophoresis in a Tris-EDTA-Borate buffer system, pH

8.5, these components migrate between the origin and hemoglobin A_2. They are called hemoglobin CS_1 and CS_2, respectively (see Fig. 4). This electrophoretic appearance is fairly reproducible provided fresh lysates are used, but when lysates are kept for more than a few days at 4 ^{o}C, there is an increase in the concentration of Hb CS_2 with the appearance of an extra band migrating just behind the origin (Hb CS_4) and a component which migrates slightly more anodally than hemoglobin A_2. This latter component, which is very variable in its appearance, is called Hb CS_3.

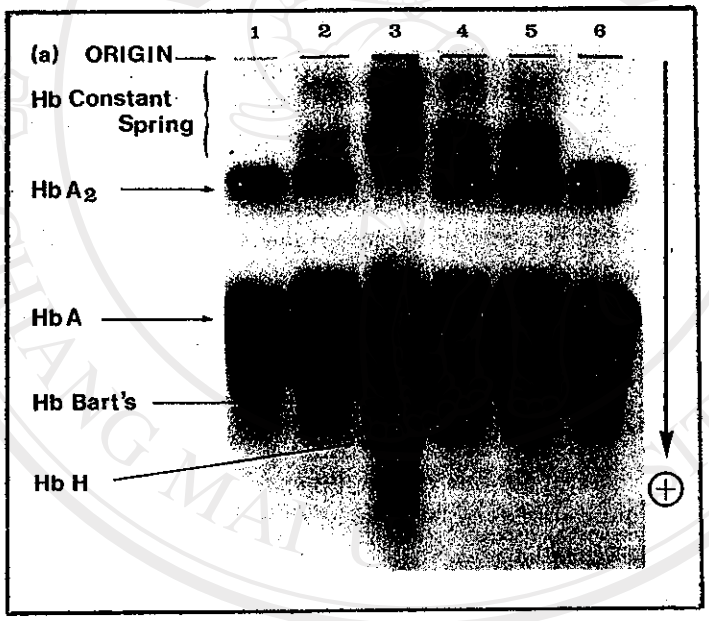


Figure 4 : Starch Gel Electrophoresis Analysis of Hb CS
(According to Weatherall and Clegg, 1981)

Lane 1 and 6 : Normal adults.
Lane 2, 4, and 5 : Hb CS Heterozygotes.
Lane 3 : Compound heterozygote of α-thalassemia 1 and Hb CS

I.2.11 Hb Constant Spring Gene and α-Thalassemia

The most common nondeletional α–thalassemia mutation is α_2 Constant Spring. It is particularly prevalent in Southeast Asia with frequencies as high as 4 % in Thailand (Wasi *et al.*, 1980). The molecular defect, a single base substitution at the termination codon of the α_2 globin gene (TAA--->CAA), allows the ribosome to read its mRNA into the 3' untranslated region until an in-phase termination codon is reached within the polyadenylation signal AAUAAA . The encoded protein contains an additional 31 amino acids at its carboxyl-terminus (Clegg *et al.*, 1971) and is incorporated into stable hemoglobin tetramer. This mutation markedly destabilizes its mRNA (Liebhaber and Kan, 1981; Hunt *et al.*, 1982). Thus, the expression of the α^{cs} gene is low, resulting in a thalassemic phenotype. The α^{cs} mRNA can be easily detected in bone marrow of individuals carrying this defect but is absent from their peripheral reticulocytes (Liebhaber and Kan, 1981). The red cells of heterozygotes contain approximately 1 % Hb Constant Spring.

The studies of Weiss and Liebhaber (1990) have revealed that the translation of ribosome into the 3' noncoding region of α^{cs} mRNA results in a significant decrease in mRNA stability. It is suggested that the present of specific structures or sequences within this region are critical to mRNA stability. This can also explain the similar effect of other α_2-termination codon mutations which cause α-thalassemia syndrome.

I.2.11 The Molecular Approaches in Detection of Hb Constant Spring

The difficulty and unreliable diagnosis of hemoglobin Constant Spring using standard electrophoresis has led to the development of more reliable DNA-based diagnostic methods, as described below.

Kosasih *et al.* (1988) used the techniques of allele specific hybridization with the α_2-globin gene fragment of genomic DNA. The genomic DNA was first digested with restriction enzyme Sst I and Hind III, this produced a 1.05-kb fragment from the 3'end segment of α_2-globin gene. Two nanodecamers were served as probes, one for normal and one for Hb CS gene, which differed by a replacement of T in the termination codon TAA with C. The probes ware labeled with P^{32} and then hybridized with the fragment, which had been separated by agarose gel electrophoresis and transferred to nylon membrane by Southern blotting.

The alike method, which differed only by the use of restriction enzyme, was used to find the distribution of Hb CS gene in Southeast Asian populations by Laig *et al.* (1990).

Hsie *et al.* (1989) combined the techniques of polymerase chain reaction and allele specific slot-blot hybridization. The 3'-end of the α_2-globin gene was amplified by PCR and then blotted onto nylon membrane. Subsequently, this membrane was hybridized with p^{32} allele specific

probes, one for the normal gene and the other for Hb CS gene. At almost the same time, Kropp *et al.* (1989) published the same techniques of detection Hb CS gene with a slightly different details. In 1990, Fucharoen *et al.* modified the method of Kropp *et al.* (1989) by using non-radioactive method in the labeling system.

In 1991, Kropp *et al.*. developed a polymerase chain reaction-based method that used asymmetric priming and a temperature shift to accomplish dual effects, selective amplification of α_2 but not α_1 DNA and discrimination of normal and Hb Constant Spring gene by allele-specific fluorescence primers.

Chang *et al.* (1993) had developed a rapid method for detecting Hb Constant Spring. The method exploited PCR technique and restriction fragment length polymorphism. The procedures were divided into two separate parts, the detection of the base C and the detection of the bases AA at the termination codon of α_2-globin gene. The base C could be detected by cutting the PCR product, which had been created a Tth 111I restriction site by a designed mismatched primer, with restriction enzyme Tth 111I. The base AA was detected by the same way, the amplified created restriction site for restriction enzyme Hind III.

Makonkawkeyoon *et al.*, (1993) had developed a rapid method for detecting the chain termination mutations in the α_2-globin gene. The PCR product which was amplified from 7 bases deletion (compared to α_1-globin

gene) in intron II to the 3'end nontranslating region was specific to α_2-globin gene. The PCR product was then cut with restriction enzyme Mse I, which recognized the base TTAA within the termination codon. The PCR product that could not be cut with Mse I resulted from the mutation within this region, which mostly due to Hb CS.

I.3 Principle and Model of Study

I.3.1 PCR Technology

The polymerase chain reaction (PCR) is a method that can amplify the selected fragment of DNA in vitro. Since it was introduced by Mullis at the Cetus Corporation in 1983, the technique has been modified for many uses and has essentially revolutionized molecular biology (Guyer and Koshland, 1989). The general principles and the optimization of the PCR below are summarized from "PCR Technology : Principles and Applications for DNA Amplification" by Henry A. Erlich (Editor), 1992 and from "PCR Protocols : A Guide to Methods and Application", Innis et. al. (Editors), 1990.

I.3.1.1 General Principles

The PCR technique was developed to provide highly efficient amplification of DNA sequences of interest. In general, the procedures depends on the availability of sequences that flank region of interest. Two synthetic oligonucleotides are prepared using these flanking sequences, one complementary to each of the strands. The DNA is denatured at high temperature and then reannealled in the presence of large molar excess of the oligonucleotides. The oliginucleotides, oriented with their 3' ends pointing toward each other, hybridize to opposite strands of the target sequence and prime enzymatic extension along the nucleic template in the

present of the four deoxynucleotide triphosphates. The end product is then denatured again for another cycle. Since the product of one cycle can serve as templates for the next, the number of the product is increased exponentially as a function of cycle number. This leads to the selective enrichment of specific DNA sequences so that they can be readily manipulated or detected.

The Standard Reaction

The standard PCR is typically done in a 50-100-µl volume and, in addition to the sample DNA, contains 50 mM KCl, 10 mM Tris-HCL (pH8.4), 1.5 mM $MgCl_2$, 100 µg/ml gelatin, 0.25 µM of each primer, 200 µM of each deoxynucleotide triphosphate, and 2.5 units of Taq polymerase. The type of the DNA sample will be variable, but it usually has between 10^2 and 10^5 copies of template (e.g., 0.1 µg human genomic DNA). A few drop of mineral oil are often added to seal the reaction and prevent condensation.

The amplification can be conveniently perform in a DNA Thermal Cycler (Perkin-Elmer Cetus Instruments) using the "Step-Cycle" program set to denature at 94 °C for 20 seconds, anneal at the proper temperature for 20 seconds and polymerize at 72 °C for 30 seconds for a total of 30 cycles. These conditions can be used to amplify a wide range of target sequences with excellent specificity. However, some PCR application may need the optimizations , which are described below, for the best result.

I.3.1.2 Optimization Strategies for the PCR

Individuals reaction components and time/temperature parameters must be adjusted for efficient amplification of specific targets.

Primer Selection

There is no set of rules that will ensure the synthesis of an effective primer pair. However, the following guideline is useful for the designing of primers.

1. Where possible, select primers with a random base distribution and with a GC content similar to that of the fragment being amplified. Try to avoid primers with stretches of polypurines, polypyrimidines, or other unusual sequences.

2. Avoid sequences with significant secondary structure, particular at the 3' end of the primer.

3. Check the primers against each other for complementarily. In particular, avoiding primers with 3' overlaps will reduce the incidence of "primer dimer". Primer dimer occurs when the 3' ends of the two primers hybridize, they form a "primer template" complex, and primer extension results in a short duplex product

PCR primers are oligonucleotides, typically 15 to 30 bases long, and are complementary to sequences defining the 5' ends of the complementary template strands. Primer concentration between 0.1 and 1.0 μM are generally optimal. The use of higher concentrations of primer is not only expensive but also increases the chances of mispriming with the resultant formation of non-specific product. Using lower primer concentrations may also help reduce the primer dimer artifact.

Buffer and Magnesium Ion

Buffer usually provided along with the DNA polymerase enzyme in the form of 10x buffer. The company usually titrates for the optimal concentration of the components within the buffer; therefore , it is no need to change the composition of the buffer unless the special effect such as reduction of the secondary structure is needed.

The free magnesium ion concentration must be adjusted for specific PCR experiments. An optimal concentration usually between 1.5 and 4.5 mM. Too little free magnesium will result in no PCR product and too much free magnesium may produce a variety of unwanted products.

Deoxynucleotide Triphosphates

In the standard protocol, each dNTP concentration is usually presented at 50-200 μM. Higher concentrations may tend to promote

misincorporations by the polymerase and should be avoided. Lower concentrations may give higher fidelity and specificity with no reduction in product yield.

As deoxynucleotide triphosphates appear to quantitatively bind Mg^{2+}, the amount of dNTPs present in reaction will determine the amount of free magnesium available. Therefore, the significantly change of dNTPs may need the adjustment of the magnesium ion concentration.

DNA Polymerase

Until now there are many thermostable DNA polymerases available, but Taq DNA polymerase is still be the enzyme of choice, due to the historical profile and the widely distribution.

In the 100-µl reaction volume, 2.0 to 2.5 units of Taq DNA polymerase are recommended for most PCR applications. The enzyme can be added conveniently to a fresh master mix prepared for a number of reactions, therefore avoiding the tedious process and possible accuracy problems associated with adding individual 0.5 µl enzyme aliquots to each tube.

Target DNA

Target DNA should be kept in a low EDTA-containing buffer since the magnesium ion is essential for the reaction. It is sometimes recommended to dissolve the target DNA in purified water.

The target DNA concentration varies, depending on the type of sequences to be amplified. For plasmid DNA, nanogram amounts are good starting point. For genomic DNA, amounts ranging from 0.05-1.0 μg are typically used for amplifications of single loci.

Cycling Parameter

PCR is performed by incubating the samples at three temperatures corresponding to the three steps in a cycle of amplification: denaturation, annealing, and polymerization.

Typical denaturation conditions are 90-95°c for 30-60 seconds. However, the higher temperature or the longer time may be appropriate, especially for GC-rich target. In contrast, too high or too long denaturation step lead to unnecessary loss of enzyme activity. The half-life of Taq DNA polymerase is >2 hours at 92.5°c, 40 minutes at 95°c, and 5 minutes at 97.5°c.

Annealing is a very critical step for the specificity of the PCR products. The annealing temperature should be adjusted for each reaction. Usually the temperatures in the range of 55-72°c yield the best results. The estimate annealing temperature for each primer can be calculate by the following formula: 2(no. of base AT) + 4(no. of base GC) - 5°c; or 5°c below the true Tm of the primers.

Polymerization usually set at 72°c, the time is usually depended on the length of the desired PCR products. At the optimum temperature of Taq DNA polymerase (75-80°c), the K_{cat} approaches 150 nucleotides/ second. At 72°c, the rate of polymerization is around 2-4 kb/minute. Thus, 30 seconds usually enough for the extension of PCR product less than 1000 nucleotides.

I.3.2 Restriction Fragment Length Polymorphism

Restriction fragment length polymorphism (RFLP) refers to inherited differences in sites for restriction enzymes that result in differences in the lengths of the fragments produced by cleavage with the relevant restriction enzyme. RFLPs are used for genetic mapping to link the genome directly to a conventional genetic marker and for the detection of some specific mutation related to genetic diseases.

I.3.3 Nested PCR

Nested PCR is the second round PCR using another pair of primers within the first PCR product. The objective of this procedure is to amplify a specific PCR product from the first round PCR. Semi-nested PCR is the nested PCR that uses only one primer different from the first PCR primers.

I.3.4 Amplified Created Restriction Site

The amplified created restriction site is the new technique that has been applied for the detection of many point mutations. The principle of this method is the introduction of a newly created restriction site by a mismatched primer next to the point that is wanted to detect. The PCR product produced by this primer will contain a site for restriction enzyme for either the wild type or the mutant, which can then be discriminated by RFLP techniques.

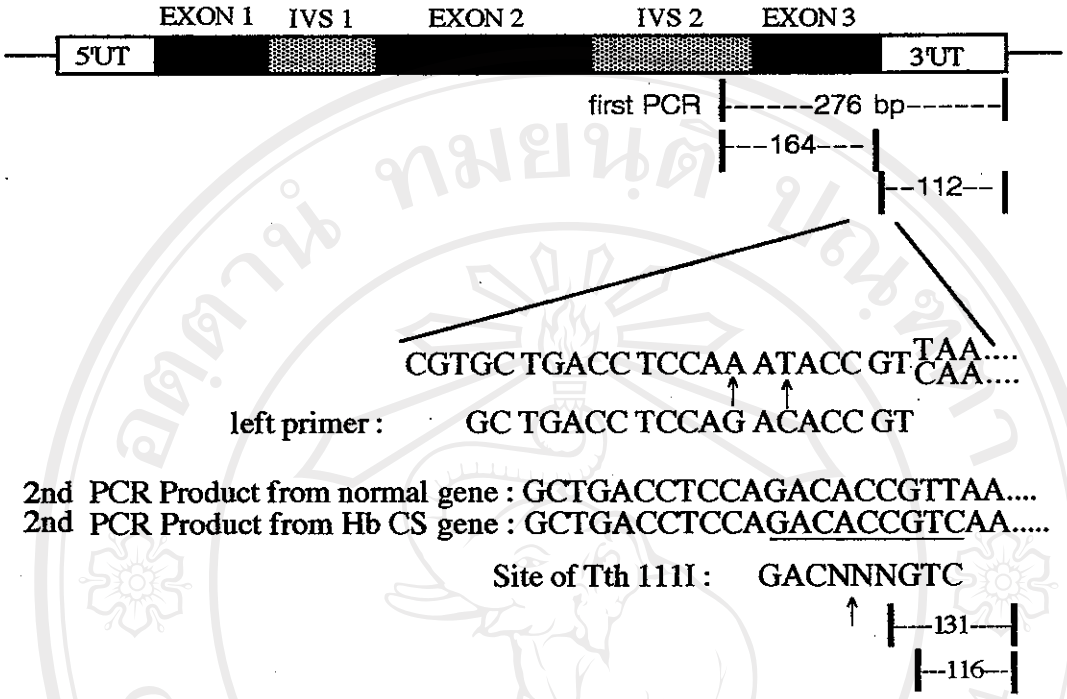
I.3.5 Model of Study

The site around the termination codon of α_1- and α_2- globin genes generates a restriction site for restriction enzyme Mse I (TTAA). Since there are 4 types of point mutations occurred in the termination codon of α_2-globin gene that cause nondeletional α-thalassemia (see text above); therefore, they can easily be detected by the property of not to be cut by

the mentioned restriction enzymes. The detection of abnormal termination codon on α_2 based on this principle was introduced by Makonkawkeyoon *et al.* (1993). The leftward primer is around the 7 base-deletion of α_2 compared to α_1 in intron 2 (7224-7244, according to Table 1), and the rightward primer is at the 3' nontranslated region (7479-7499, Table 1). The PCR product was 276 bp (7224-7499) long and specific to α_2-globin gene. The restriction fragments from normal subject were 164 bp- and 112 bp- long , while the only fragment of 276 bp was found in the subjects with abnormal stop codon.

To characterize the Hb Constant Spring gene, the PCR product from the sample that cannot be cut by Mse I from the step above are the subject to semi-nested PCR. With the leftward primer introduced by Chang *et al.*,(1993), the product with Hb CS gene will have the restriction site for restriction enzyme Tth 111I. The semi-nested PCR should have the product of 131 bp-long (7369-7499) and the restriction fragments of Hb CS gene should have 116 bp and 15 bp, corresponding to the restriction site of the enzyme.

The figure below shows the position of primers, the sequences of amplified α_2-globin DNA, the expected size of the PCR product after cutting with Mse I, the position of the Tth 111I restriction site created primer and its mismatches, and also the expected size of semi-nested PCR product and its fragments after cutting with Tth 111I.



7201 ggaggtgtag cgcaggcgga ggctgcgggc ctgggccgca ctgaccctct

7251 tctctgcaca gctcctaagc cactgcctgc tggtgaccct ggccgcccac

7301 ctccccgccg agttcacccc tgcggtgcac gcctccctgg acaagttcct

7351 ggcttctgtg agcaccgtgc tgacctccaa ataccgttaa gctggagcct

7401 cggtagccgt tcctcctgcc cgatgggcct cccaacgggc cctggtcccc

7451 tccttgcacc ggcccttcct ggtctttgaa taaagtctga gtgggcggca

Figure 5: Model of Study and Nucleotide Sequence of PCR Product

I.4 Objectives

1. To find the optimal conditions for the newly modified method, in order to detect the abnormal termination codon of α_2-globin gene and Hemoglobin Constant Spring gene.

2. To study the prevalence of Hb Constant Spring in the northern Thailand populations.

3. To detect the Hb Constant Spring gene in nondeletional Hb H patients.