II. LITERATURE REVIEW

1. Structure and expression of human hemoglobins

Human hemoglobins have a tetrameric structure made up of two α- or α-like globin chains and β- or β-like globin chains, each containing a heme-Fe++ prosthetic group which can reversibly bind oxygen. The α-related globin chains and β-related globin chains are encoded in two separate gene clusters located on chromosome 16 and 11, respectively (Deisseroth et al.,1977; Deisseroth et al.,1978). During the embryologic development there is an ordered switching of globin chain synthesis (Bunn et al., 1986). The β-globin gene cluster encodes globin subunits specific to the embryonic, fetal, and adult developmental stages; ε, γ, and β-globins, respectively. In contrast, the α-globin gene cluster encodes only two globin proteins; ζ–globin gene is expressed during the embryonic period and α - globin gene is expressed during the fetal-adult period. In the very early embryonic stage, Hb synthesis is restricted to the yolksac and the production of Hb Gower I ($\zeta_2 \varepsilon_2$), Hb Gower II($\alpha_2 \varepsilon_2$), and Hb Portland ($\zeta_2 \gamma_2$). Subsequently, at about 8 weeks of gestation, the fetal liver takes over , synthesizing predominantly Hb F($\alpha_2 \gamma_2$) and a small amount(<10 percent) of Hb A ($\alpha_2 \beta_2$). Between about 18 weeks and birth , liver is progressively replaced by bone marrow as the major site of red cell production, and this is accompanied in the later stages of gestation with a reciprocal switch in production of Hb F and Hb A, which continues until the end

of the first year when Hb F production has dropped to less than 2 percent. Throughout development a balance between α-like and β-like globin gene expression must be maintained so that functional hemoglobin tetramer can be assembled. The defect in the thalassemias is the loss of this balance (Heywood et al., 1964). In α-thalassemia, a relative deficiency in α-chains results in the accumulation of excess of γ or β globin(Clegg et al.,1967). Excess γ and β chains present in the newborn and adult can self assemble into γ_4 (Hb Bart's) and β4 (Hb H) homotetramer,respectively resulting in red cells damage and dysfunction.

2. The α - globin gene cluster

The 29 kb α-globin gene cluster on chromosome 16 comprises the duplicated α genes(α2 and α1), an embryonic α-like gene (ζ2), three pseudogenes (ψζ1, ψα2, ψα1) and θ1 gene arranged in the order 5'-ζ2-ψζ1-ψα2-ψα1-α2-α1-θ1-3' (Fig.1) (Lauer et al., 1980). Each of these genes has been fully sequenced and the functional status of all but one (θ1) is now clearly defined. This gene family is thought to have evolved by a series of duplications followed by sequence divergence punctated by insertions and deletions of DNA surrounding the coding parts of the genes (Lauer et al.,1980;Maniatis et al.,1981). Comparison of many globin genes within and between species demonstrates that some structural features have been conserved (Bunn et al., 1986) ; Stamatoyannopoulos et al., 1987). In general, the globin genes are compact (1 to 2 kb) and are divided into three exons by

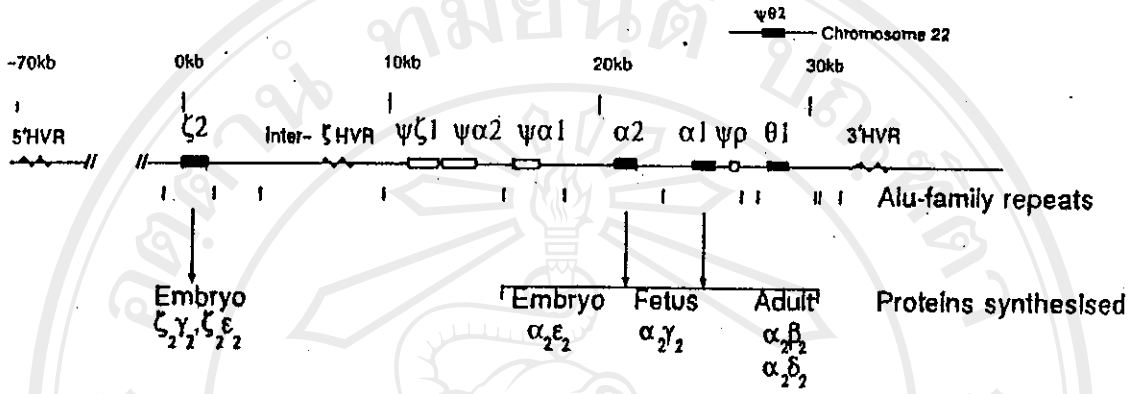

Figure 1. The organization of the α-globin complex.Filled boxes indicate functional genes and open boxes pseudogenes. Position O represents the ζ-globin mRNA CAP site. The nonlinked θ2 gene is shown on chromosome 22. Hypervariable regions (HVR) are denoted by zig-zag lines. The positions of Alu family repeats are shown below the complex. The hemoglobins synthesized at each stage of development are indicated below the α complex. Each molecule of hemoglobin comprises a tetramer of two α- like (α- or ζ-) and two β-like (β-, γ-, δ- or ε-) globin chains. (according to Higgs et al., 1989)

two noncoding intervening sequences (IVS). The positions of the introns are conserved in all globin genes and may reflect the early (>500 million years ago) evolutionary events that brought together the functional domains of the protein. For example, exon 2 encodes the entire heme binding region and $\alpha_1\beta_2$ contacts, whereas exon 3 encodes $\alpha_1\beta_1$ contacts;residues that contribute to Bohr effect and 2,3-diphosphoglycerate binding are more randomly distributed. The sizes and sequences of the introns are quite different and may vary from one individual to another within the tandemly repeated segments of the ζHVRs. Certain sequence motifs are conserved in most eukaryotic genes. Evidence from a variety of experimental systems and natural occurring varients has demonstrated that these sequences are essential to the process involved in gene expression and are often mutated in the nonfunctional pseudogenes. Sequence and heteroduplex analyses have shown that the α- globin genes are imbedded within two homologous segments, each approximately 4 kb long,which can be further divided into three homologous subsegments (X, Y, and Z). The Zα1- and Zα2-segments contain the α1- and α2-globin genes, respectively. The first intron in each gene is identical, but the second intron of α1 is seven bases longer and differs by three bases from the same region of the α2-gene. In the third exon, there are 18 base substitutions and a single base deletion in α2 relative to α1, all located within the 3' nontranslated region. The two mature α–globin mRNA only diverge in strcture in this 3'-nontranslated region and therefore encode identical α - globin proteins, but the α2 gene is functionally

the major globin gene in humans, encoding two to three times as much α-globin protein as the adjacent α1 gene (Stephen et al., 1986).

The two ζ-genes are also highly homologous(Proudfoot et al., 1982). The ψζ1 differs from the functional ζ2 gene by only three base substitutions in the protein coding region, one of which creates the nonsense mutation GAG-->TAG at codon 6. Homozygous deletion of the ψζ1 genes is compatible with normal development, thus supporting its lack of function (or dispensability) in vivo. The two other pseudogenes are ψα1 and ψα2. The ψα1 gene contains six separate inactivating mutations (Proudfoot et al., 1980). The ψα2 gene, probably the oldest of the three pseudogenes, has diverged so far from the normal α-globin gene structure that its identity was only first recognized by computer analysis (Hardison et al., 1986). The θ1 gene is highly similar in structure to the α-globin genes (Hsu et al.,1988). There is, to date, no evidence that the θ1 globin protein is expressed *in vi vo* and its function does not appear to be essential to normal fetal development (Fischel-Ghodsian et al.,1987). For this reason, the θ1 globin gene is as a provisional pseudogene.

Moreover, there are five separate regions within or adjacent to the α-globin cluster which are structurally hypervariable (hypervariable regions, HVR). Each of these regions is composed of multiple tandem repeats of a short GC rich sequence ranging in size from 17(3' HVR) to 57(5' HVR) bp. The repeats in all five of these regions share an 11 base core sequence 5'-GNGGGG(N)ACAG- 3'. The α globin 3'HVR is located approximately 4 kb 3' of the θ1 globin gene(Jarman et al., 1986). The inter- ζ HVR is located between the ζ2 and ψζ1 gene(Goodbourn et al.,1982) and the two ζHVRs lie within intron 1 of the ζ2 and ψζ1 globin gene (Proudfoot et al., 1982). More recently, a fifth HVR (call the 5' HVR) has been located 100 kb 5' to the ζ-globin gene (Jarman et al., 1988).

3. Molecular defects of α-thalassemia

The advent of molecular cloning, DNA sequencing, and restriction endonuclease mapping technique has led to an increased understanding of the molecular basis for the DNA rearrangement that causes α-thalassemia. The great majority of α-thalassemia results from deletions involving one or both of the α-globin genes and some of the mechanism underlying such deletions are beginning to be understood.

3.1 α-Thalassemia 2 or α^+thalassemia : one deleted α globin gene per chromosome - designated (-α/).

The mechanism by which the α^+thalassemia deletions occur has now been established, and it is clearly related to the underlying molecular structure of the α-globin complex (Lauer et al., 1980;Embury et al; 1980). Each α gene is locatd within a region of homology approximately 4 kb long, interrupted by two small nonhomologous regions. It is thought that the homologus regions result from an ancient duplication event. Subsequently, during evolution, these homologous segments were subdivided, presumably by insertions and deletions, to give three homologous subsegments referred to as X, Y, and Z (Fig. 2,A). The duplicated Z boxes are 3.7 kb apart, and the X boxes are 4.2 kb apart. Misalignment and reciprocal crossover between these segments at meiosis can give rise to chromosome with either single (-α) or triplicated (ααα) α-globin genes (Fig. 2B,2C). Such an occurrence between homologous Z boxes deleted 3.7 kb of DNA(referred to as a rightward deletion, $-\alpha^{3.7}$), whereas a similar crossover between the two X blocks deletes 4.2 kb of DNA (referred to as a leftward deletion, $-\alpha^{4.2}$). The corresponding triplicated α-genes arrangements are referred to as $\alpha\alpha\alpha^{anti\ 3.7}$ and $\alpha\alpha\alpha^{anti\ 4.2}$. The $-\alpha^{3.7}$ deletion is extremely widespread, being found in all populations studied. The $-\alpha^{4.2}$ is most frequently found in Asian (and related) populations although it has also been reported in numerous other populations (Pacific Islanders, Black, Mediterraneans, etc.). The linkage of these mutations to specific α-globin gene cluster haplotypes suggests that each has occurred by a number of independent mutational events with subsequent positive selection in certain populations (Yenchitsomanus et al., 1985). Its

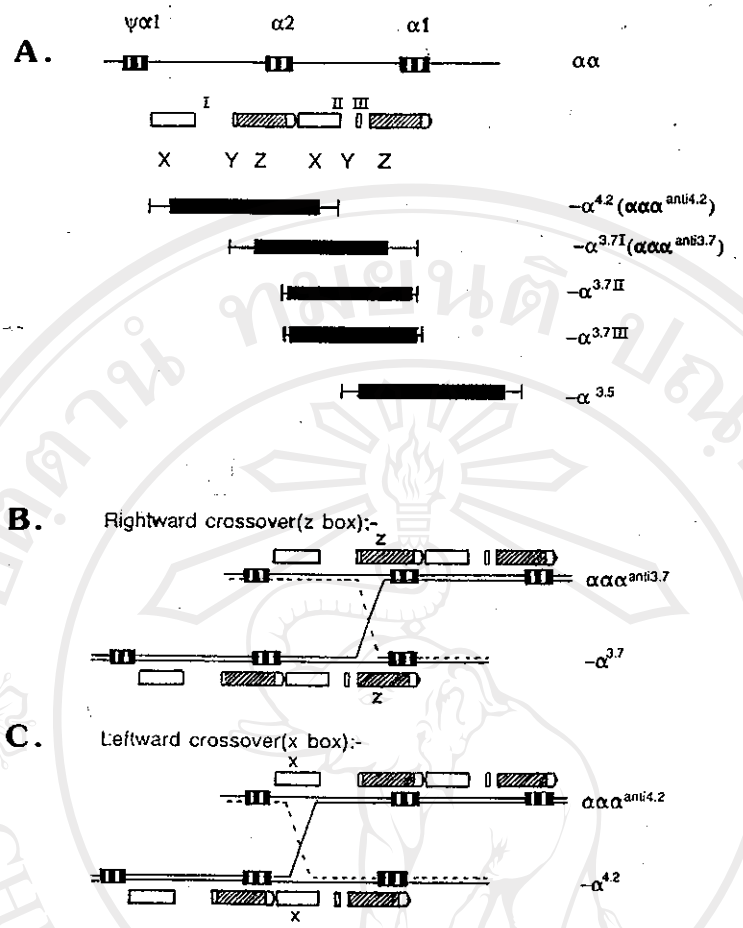

Figure 2. The mechanism of unequal crossover that give rise to the -α 3.7and -α4.2 deletions. A.The fine structural detail of the duplicated α-globin gene is shown as in the figure 1. B. The rightward crossover occurs when genetic exchange takes place between the misaligned homologous Z boxes giving rise to chromosomes with either one (-α 3.7) or three (αααanti3.7) α-globin genes. C. The leftward crossover occurs when genetic exchange takes place between the misaligned homologous X boxes giving rise to chromosomes with either one (-α 4.2) or three (ααα anti4.2) α-globin genes. (according to Higgs et al., 1989)

prevalence correlates geographically with that of malaria (Flint et al., 1986). In the Mediterranean Basin the prevalence of heterozygous state of $-\alpha^{3.7}$ is approximately 5-10% (Kanavakis et al., 1988) although in certain areas such as Cyprus it can reach higher than 30%, in Southeast Asia approximately 10-20%, in certain regions 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).

In both the $-\alpha^{4.2}$ and the $-\alpha^{3.7}$ deletions, the remaining α-globin gene is identical to the α1-globin gene. However, it can not be assumed that this remaining gene functions at the same low level as the native α1 gene since each of these deletions has resulted in a major change in the structure of the cluster as a whole. Studies which have assessed the expression of the $-\alpha^{3.7}$ gene have in fact demonstrated that it is expressed at 1.8 fold higher levels than the α1-globin gene in the native α-globin cluster (Liebhaber et al., 1985). In addition a number of studies suggest that the $-\alpha^{3.7}$ deletion may result in a less significant loss of α- globin synthesis than the $-\alpha^{4.2}$ deletion (Fodde et al., 1988).

3.2 α Thalassemia 1 or α^{o}thalassemia due to total deletion of two α globin genes per chromosome - designated (--/)

With the exception of the $(\alpha\alpha)^{RA}$ mutant (a deletion from the α-cluster in which both α-genes remain intact) , all of the deletions

described in this section either completely or partially [$-(\alpha)^{5.2}$ and $-(\alpha)^{20.5}$] delete both α globin genes, and therefore no α chain synthesis is directed by these chromosomes *in vivo* (Fig. 3). Those α thalassemia defects that have been fully characterized result from illegitimate or nonhomologous recombination events (Nicholls et al., 1987). The deletions range in size from rather small (5.2 kb) to those which remove the entire cluster [--Fil, -- Thai (Winichagoon et al., 1984) and a new deletion of > 47 kb in a Northern European family (Fischel - Ghodsian et al., 1988 ; Fortina et al., 1988)].

The two most common deletions, (--SEA) and (--Med), occur in Southeast Asia and the Mediterranean Basin, respectively. These two along with two additional deletions(--SA, and $--(\alpha)^{20.5}$) are all approximately the same size (20 - 30 kb) and remove both α-globin genes (--) but leave the functional ζ2 gene (Nicholls et al., 1987).

3.3 Nondeletional α-thalassemia

These comprise a distinct and apparently heterogenous group of mutations that lead to absent or greatly reduced expression of α genes that are structurally intact, as determined by restriction endonuclease mapping (Kan et al., 1979 ; Pressley et al.,1981; Higgs et al.,1981). Several lesions of the nondeletional α-thalassemia have been described (Higgs et al., 1989) all but one occur in the α2 globin gene(Table 1). The two mutations modifying the initiation codon (ATG --> ACG in the α2 gene; ATG --> GTG in the α1 gene) prevent the synthesis of the α

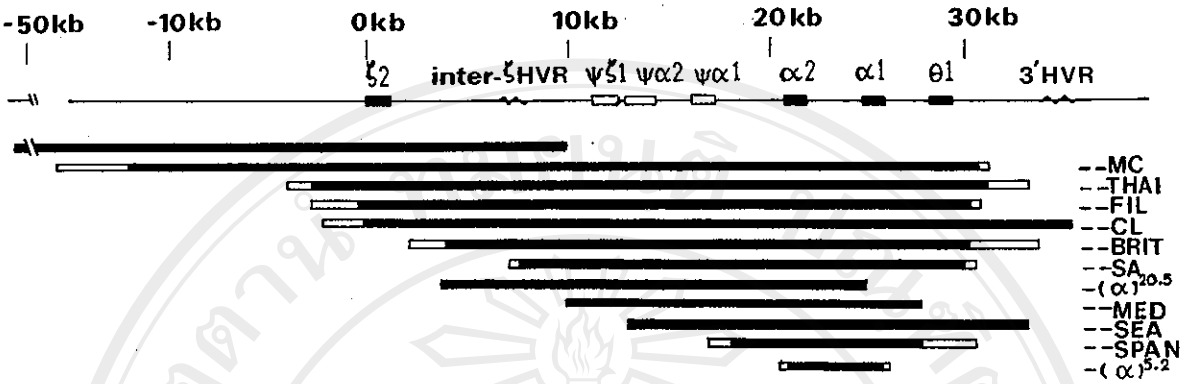

Figure 3. summerizes the deletions that give rise to α^{o}-thalassemia. The extent of each deletion is represented as a solid box and the uncertainty of the brakepoints is indicated by the open boxes. (according to Higgs et al., 1989)

chains because of abnormal initiation of translation(Pirastu et al., 1984; Morle et al., 1985). The 5 nuclotides deletion at the 5' end of the first intron makes splicing at that site impossible leading to abnormal processing of the α globin mRNA precursor, while the mutation in the hexanucleotide (AATAAA --> AATAAG) affects the polyadenylation of the 3' end of the mRNA(Orkin et al.,1981;Felber et al.,1982;Higgs et al.,1983). Some mutations in the third exon result in the formation of unstable α chain or a termination codon and subsequently in an α chain deficiency. Four elongated α chains, each with a total of 172 amino acid residues, are the result of different mutation within the normal termination codon (Hb Constant Spring $\alpha^{CS}\alpha$, Hb Icaria $\alpha^{IC}\alpha$, Hb Koya Dora $\alpha^{KD}\alpha$, and Hb Seal Rock $\alpha^{SR}\alpha$).

The most common nondeletional α-thalassemia is Hb Constant Spring(α^{CS}). It is particularly prevalent in Southeast Asia with frequencies as high as 4% in Thailand (Wasi et al., 1980). As reported by Laig et al.(1990), it is up to 6-10% in Northern and Northeastern Thailand. This abnormal globin arises from a single base substitution at the termination codon of the α2 globin gene (TAA -->CAA), allowing the ribosome to read into the 3' untranslated region until an in-phase termination codon is reached within the polyadenylation signal AAUAAA. The encoded α^{CS} protein contains an additional 31 amino acids at its carboxyl-terminus (Clegg et al., 1971).

Table 1. Non-deletion mutants that caused α-thalassemia.
(according to Higgs et al., 1989)

	Affected Gene	Affected Sequence	Mutation	Geographical Distribution	Comments
RNA processing	α 2	IVS I donor site	GAGGTGAGG->GAGG	Mediterranean	Aberrant splicing
	α 2	poly(A) signal	AATAAA->AATAAG	Middle East,Mediterranean	↓efficiency of 3' end processing
	α 2	Initiation codon	CCACCATGG->CCACCACGG	Mediterranean	↓ mRNA translation
	α 1	Initiation codon	CCACCATGG->CCACCGTGG	Mediterranean	↓ mRNA translation
	-α	Initiation codon	CCACCATGG->CCACCGTGG	Black	↓ mRNA translation
	$-\alpha^{3.7II}$	Initiation codon	CCACCATGG->CC-CATGG	North African,Mediterranean	↓ mRNA translation
	α 2	Exon III	α116 GAC -> UAG	Black	In phase termination
	α 2	Termination codon	α142 TAA -> CAA	Southeast Asian	Hb Constant Spring
	α 2	Termination codon	α142 TAA -> AAA	Mediterranean	Hb Icaria
	α 2	Termination codon	α142 TAA ->TCA	Indian	Hb Koya Dora
	α 2	Termination codon	α142 TAA -> GAA	Black	Hb Seal Rock
	-α	Exon I	α30/31 GAGAGG->GAG-G	Black	Reading frameshift
Post translatioal	α 2	Exon III	α125 Leu -> Pro	Southeast Asian	Hb Quong Sze
instability	α 2	Exon III	α109 Leu -> Arg	Southeast Asian	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

4. Interaction between α-thalassemia determinants

There are potentially several hundred different interactions that could take place between the large number of α-thalassemia determinants that have been described. Phenotypically these interactions result in one of four broad categories.

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

2.) α-Thalassemia trait (two functional α genes) : There are mild hematologic changes but no major clinical abnormality.

3.) Hb H disease(one functional α gene) : The pathophysiology resembles that of β-thalassemia intermedia.

4.) Hb Bart's hydrops fetalis (no functional α genes): The clinical picture is a pale edematous infant with signs of cardiac failure and prolonged intrauterine hypoxia.

5. Hb H disease

Hb H disease is the most severe form of α-thalassemia phenotype compatible with life. It most frequently results from the interaction of α-thalassemia 1 and α-thalassemia 2, and therefore it is predominantly found in Southeast Asia (commonly --SEA/-$\alpha^{3.7}$) and the Mediteranean basin (commonly --MED/-$\alpha^{3.7}$) where both α-thalassemia 1 and α-

thalassemia 2 are common. Hb H disease may also result from the interaction of nondeletion mutations affecting the predominant α2 globin gene ($\alpha^{Nco}\alpha$ / $\alpha^{Nco}\alpha$, $\alpha^{T\ SAUDI}$ / $\alpha\alpha^{T\ SAUDI}$, and $\alpha\alpha\alpha^{T\ SAUDI}$ / $\alpha\alpha$ $^{T\ SAUDI}$). In Algeria, homozygotes for the $-\alpha^{3.7III}$ defect ($-\alpha^{3.7III}/-\alpha^{3.7III}$) have typical Hb H disease (Whitelaw et al., 1986).

The genetic basis for Hb H disease is diverse and as more molecular defects are characterized, the underlying interactions will become even more complex. It is not yet clear to what extent this molecular diversity is reflected in the variable clinical and hematologic features of Hb H disease. The clinical picture of Hb H disease is usually thalassemia intermedia,although there is considerable variation in the severity of this condition. The predominant features are hypochromic, microcytic anemia , with jaundice and hepatospleeno-megaly. Since the main mechanism of the anemia is hemolysis rather than dyserythropoiesis, only one-third of patients have clinical evidence of an expanded erythron. The most common complication is the development of severe spleenomegaly with hyperspleenism. Others include infection, leg ulcers, gallstones and folic acid deficiency (Weatherall et al., 1988).

The hematologic features of Hb H disease are also quite variable. Hemoglobin levels ranging from 2.6 to 12.4 g/dl have been recorded, in association with reticulocytosis and typical thalassemic changes of the red cell indices. The hemoglobin consists of Hb A with a variable amount of Hb H and sometimes Hb Bart's. The proportion of Hb H varies from 2 to 40 percent. When peripheral blood is incubated with

redox dyes, this is reflected in the number of cells that contain typical Hb H inclusions.

As yet there have been no systemic attempts to correlate the genotype with the phenotype of Hb H disease. However, in general it appears that, as expected, patients with a nondeletion defect(affecting the predominant α2 gene) interacting with α- thalassemia 1 determinant (--/ $\alpha^T\alpha$) have higher levels of Hb H (β4), a greater degree of anemia and, anecdotally, a more severe clinical course than patients with the --/-α genotype. At the extreme of this spectrum, three patients have been described in whom severe Hb H disease was associated with hydrops fetalis (Chan et al., 1985; Trent et al., 1986).

In Thailand, where there is an abundance of well-documented cases of Hb H disease, it is known that despite the relatively homogenous nature of the molecular basis (--SEA/ $-\alpha^{3.7}$ in 80 percent) (Winichagoon et al., 1984), the clinical course is quite variable. This suggests that other genetic and environmental factors play an important role in the clinical and hematologic variation seen in this syndrome.

6. Purpose of the study

1) To optimize the conditions of non - radioactive Southern hybridization (PhotoGene Detection System^{TM})

2) To characterize the molecular basis of Hb H patients by non-radioactive Southern hybridization