CHAPTER I
INTRODUCTION

1.1 STATEMENT AND SIGNIFICANCE OF THE PROBLEM

The α-thalassemias are the most common single-gene disease in the world (1). They are characterized by a reduction or complete absence of α-globin gene expression. Normal individuals have two α-genes on each chromosome 16 (αα/αα). The loss of one (-α) or both (--) of these *cis*-linked genes are the most common causes for α-thalassemias. Patients with Hb H disease (--/-α) develop chronic hemolytic anemia of variable severity, whereas fetuses with Hb Bart's hydrops fetalis (--/--) die either *in utero* or shortly after birth as a result of severe intrauterine anemia. Although individuals with three functional α-genes (-α/αα) are clinically and hematologically silent and carriers with α-thalassemia trait (-α/-α or --/αα) only result in very mild hypochromic microcytic anemia, couples with these genotypes are at risk of having a hydrop baby or offspring with Hb H disease (2). In Thailand between 15% and 30% of the population are carriers (3). It can be estimated that about 1,250 fetuses will have Hb Bart's hydrops fetalis and about 7,000 fetuses with Hb H disease per year, which based on a total birth rate of about one million per year (4). These reports indicate that the thalassemias are the major health problem in Thailand.

Homozygosity of the α-thalassemia-1 gene resulting in Hb Bart's hydrops fetalis is a disorder almost always incompatible with postnatal life. Pregnant women with an affected fetus may present in the late second or third trimester of pregnancy with complications of pregnancy and/or fetal distress. The fetus typically is hydropic and severely anemic, and frequently in these cases, prenatal diagnosis is required (5). Because of the method for treatment of the Hb Bart's hydrops fetalis in present is not

appropriate. Thus, it seems reasonable to develop the preventive strategies to avoid the serious genetic diseases.

In Thailand, the --determinant most commonly results from a deletion of approximately 20-kb that removes both α-genes but leaves the ζ_2- and $\psi\zeta_1$-genes intact ($--^{SEA}$) (6). Thus, the detection of α-thalassemia-1 gene carriers in a routine clinical laboratory was the only $--^{SEA}$ determinant. Nevertheless there have been several reports (7,8) of Thai individuals with α-thalassemia-1 in whom there is a complete deletion of the ζ-α complex on one chromosome. They are called the $--^{THAI}$ and $--^{FIL}$ determinants (6). Since these deletions extend beyond the ζ-genes at the 5' end and beyond the α-genes at the 3' end they cannot be positively identified in heterozygotes ($--/\alpha\alpha$) by routine PCR of the $--^{SEA}$ determinant (9,10) and may therefore be missed during genetic counseling and prenatal diagnosis.

Therefore, we have developed a PCR technique that allows rapid detection of these 3 determinants in the same time. This technique takes advantage of the sensitivity of multiplex PCR (11,12) to detect the α-thalassemia-1 ($--^{SEA}$, $--^{THAI}$, and $--^{FIL}$) determinants which there have been report in Thai population.

In this study, multiplex PCR was applied to characterize the molecular mutation and incidence of the non-southeast Asian deletion type of α-thalassemia-1 in patients with Hb H disease. This simple multiplex method should greatly facilitate the genetic screening and molecular diagnosis of these determinants in populations where α-thalassemia are prevalent.

1.2 LITERATURE REVIEW

1.2.1 Structure and function of hemoglobin (13-15)

Hemoglobin is the oxygen-carrying protein molecule found in red blood cells (RBC) of vertebrates. Hemoglobin is responsible for the transport of oxygen from the lungs to other body tissues and the transport of carbon dioxide from tissues to the lungs. Hemoglobin represents >95% of the solid constituents of RBCs, and more is known about the chemical structure and function of this oxygen-transporting protein than any other protein molecule. The hemoglobin molecule is composed of four polypeptide subunits (the protein or globin components) and four heme groups (one on each globin chain) (Figure 1). At the center of each heme group is an iron atom (Fe^{2+}) (Figure 2). The hemoglobin molecule is about 6% heme and 94% globin. The heme molecule contains the iron atom that is responsible for the reversible binding of the oxygen molecule in its normal physiological role. Only when the iron is the ferrous (Fe^{2+}) oxidation state can oxygen bind. If the iron is in the ferric (Fe^{3+}) state, heme does not bind oxygen, and the hemoglobin molecule is referred to as methemoglobin or ferrihemoglobin.

The hemoglobin molecule is ellipsoid in shape and roughly 5.5 x 5.5 x 7.0 nm in size. The molecular weight of normal adult human hemoglobin Hb A is 66,000 daltons. The four polypeptide units of normal adult hemoglobin, Hb A, are two α-chains and two β-chains having 141 and 146 amino acid residues each, respectively, arranged in a truncated tetrahedron (Figure 3).

The four polypeptide chains comprise the functional globin unit of hemoglobin. However, embryonic hemoglobin, present in the first 2 months of intrauterine life, is formed with the pairing of one α-chain and one γ-chain. This initial pairing produces a non-functional hemoglobin dimer (i.e., it cannot efficiently deliver oxygen). The functional hemoglobin tetramer is formed with the combination of two dimers (two α-chains and two γ-chains). The α-chains are identical to those in normal adult hemoglobin (Hb A), and the γ-chains are similar, but not identical, to the β-chain.

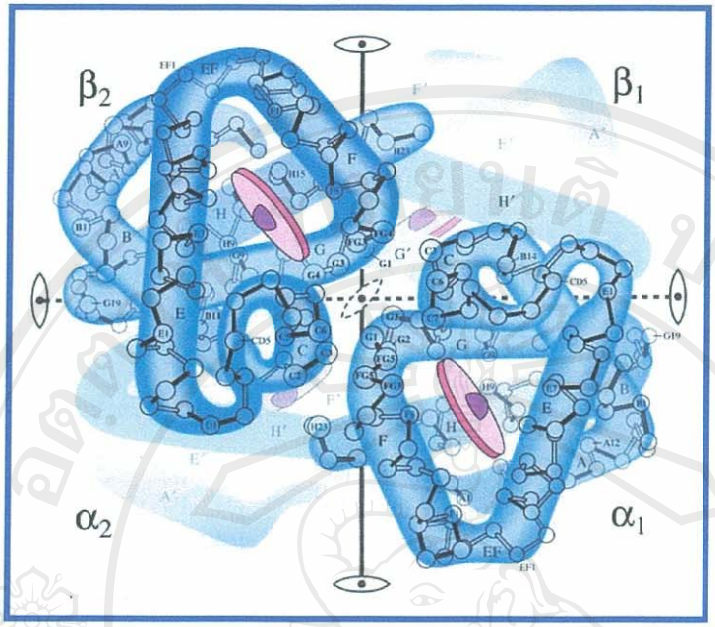

Figure 1. Hemoglobin structure (Available from: http://chemlearn.chem.indiana.edu/S125/Biochem/Background%20fo.../Spectro.html Accessed August 17, 2003)

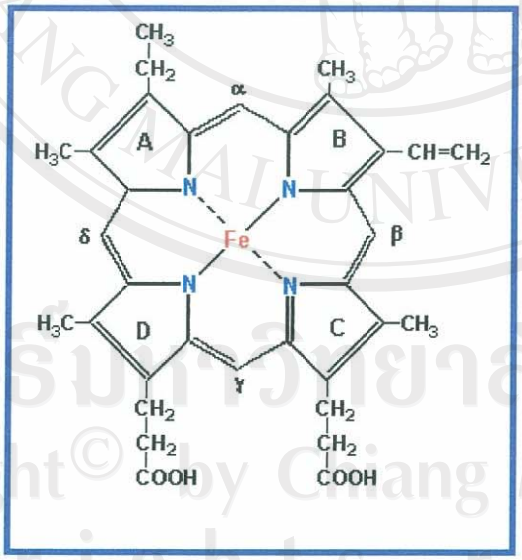

Figure 2. Structure of heme group (Available from: http://ntri.tamuk.edu/homepage-ntri/lectures/protein/hemoglobin/hempage.html. Accessed September 9, 2003)

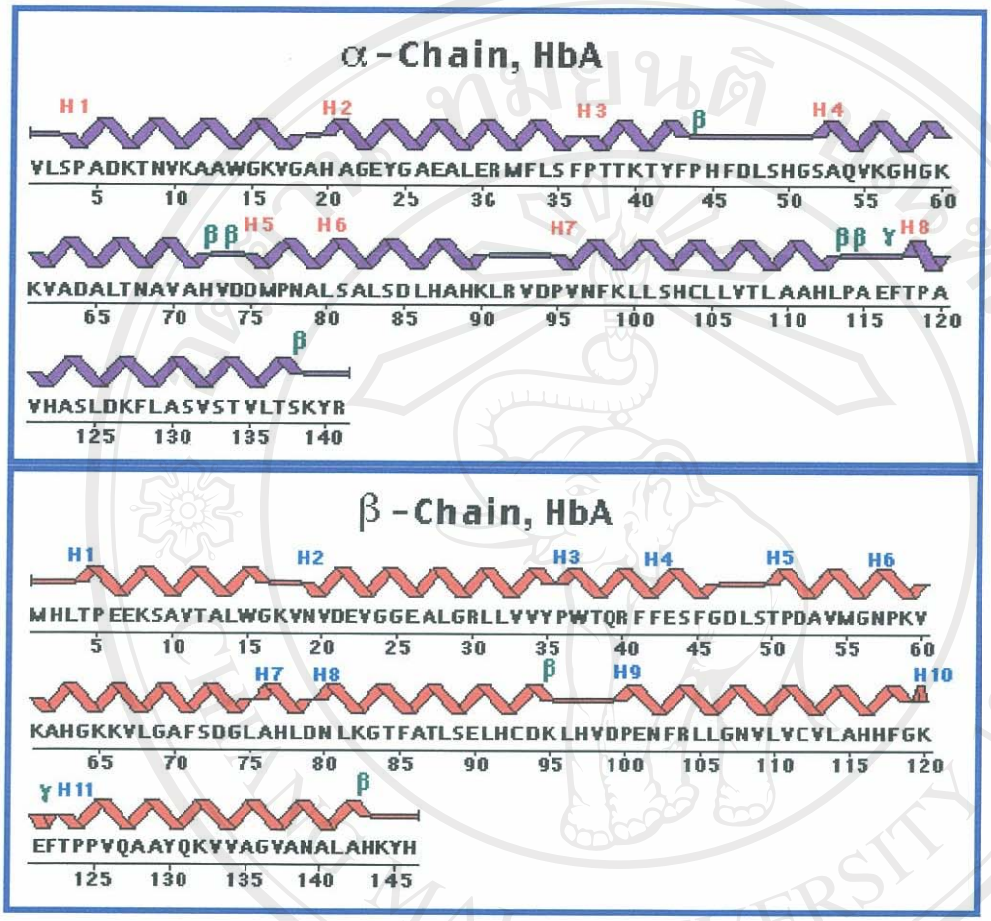

Figure 3. Representation of the primary structure of the human α- and β-globin chains. The four polypeptide units of normal adult hemoglobin, Hb A, are two α-chains and two β-chains having 141 and 146 amino acid residues each, respectively, arranged in a truncated tetrahedron.

(Available from: http://ntri.tamuk.edu/homepage-ntri/lectures/protein/hemoglobin/hempage.html. Accessed September 9, 2003)

Two types of γ-chains are synthesized, differing only in a glycine or alanine in position 136 of the γ-chain.

Each hemoglobin molecule has a life span the same as the RBC it is contained in, roughly 120 days. When old RBC are destroyed, the heme is metabolized into iron which is reused, and to porphyrin, which is degraded to bile pigments and excreted by the liver. The four polypeptide chains of the hemoglobin protein are broken down to amino acids which may be reused in other metabolic processes including the synthesis of new protein.

Normal adult human males and females have about 16 and 14-g, respectively, of hemoglobin per 100 ml of blood. Each RBC contains about 29×10^{-12} g hemoglobin. RBC comprise 40-50% of the volume of whole blood. Determination of these quantities is useful in diagnosing hematological disorders. Significant deviation from the values occur under altered conditions of RBC production and destruction.

1.2.2 The human hemoglobin types (16)

All of the hemoglobins listed in Table 1 shared similar chemical and functional properties, but each also has a number of unique characteristics.

Hemoglobin A (Hb A, $\alpha_2\beta_2$) normally makes up 94-97% of the hemoglobin in the circulating erythrocytes of adults and is therefore the primary determinant of hemoglobin-related functional properties of the blood. Approximately 5% of Hb A in normal adults is present as a glycosylated derivative, Hb A_{1c}, in which the amino-terminal valyl residues of the β-chains are in keto-amine linkage with glucose.

Hb A_2 ($\alpha_2\delta_2$) is structurally very similar to Hb A but represents a minor hemoglobin fraction, accounting for only 2 to 3% of the hemoglobin in the erythrocytes of normal adults. The functional properties of Hb A_2 are nearly identical to those of Hb A. It is present in the low concentration thereby play no significant role in the transportation of oxygen in the blood. However, in several forms of β-thalassemia, the level of Hb A_2 is significantly elevated.

Table 1. Normal human hemoglobins and their globin subunits (16)

Hemoglobins	α-like subunits	β-like subunits	Tetramer composition
Hb A	α	β	$\alpha_2\beta_2$
Hb A2	α	δ	$\alpha_2\delta_2$
Hb F	α	$^{G}\gamma, ^{A}\gamma$	$\alpha_2\gamma_2$
Hb Gower-2	α	ε	$\alpha_2\varepsilon_2$
Hb Gower-1	ζ	ε	$\zeta_2\varepsilon_2$
Hb Portland	ζ	γ	$\zeta_2\gamma_2$

Hb F ($\alpha_2\gamma_2$) is distinguished from other normal human hemoglobins by its resistance to denaturation by alkali. This property form the basis for the most widely used method for the quantitative measurement of Hb F. From 10 to 15% of Hb F in the blood exists as a derivative (Hb F-1) in which the γ-chain amino terminal amino groups are acetylated.

The oxygen affinity of pure Hb F is quite similar to Hb A. however, blood that contains mainly Hb F exhibits a characteristically higher oxygen affinity than that of blood containing Hb A (17). These observations are explained by differences between the interaction of Hb A and Hb F with 2,3-diphosphoglycerate (2,3-DPG), which is present in high concentration in all normal human erythrocytes. Hb A is the deoxy state interacts strongly with 2,3-DPG to produce a significant lowering of the oxygen affinity of the hemoglobin. Hb F, however, exhibits considerably less interaction with 2,3-DPG (18), and the presence of 2,3-DPG in the erythrocytes has a relatively minor effect on the oxygen binding property of Hb F.

1.2.3 Genetic control and synthesis of hemoglobin (19,20)

The structure and genetic control of human hemoglobin is summarized in Figure 4. Human adult hemoglobin is a heterogeneous mixture of proteins consisting of a major component, Hb A, and a minor component, Hb A_2, which constitutes about 2.5% of the total. In intrauterine life, the main hemoglobin is Hb F. The structure of these hemoglobins is similar. Each consists of two separate pairs of identical globin chains. Except for some of the embryonic hemoglobins, all the normal human hemoglobins have one pair of α chains: in Hb A these are combined with β-chains ($\alpha_2\beta_2$), in Hb A_2 with δ-chains ($\alpha_2\delta_2$), and in Hb F with γ-chains ($\alpha_2\gamma_2$).

Human hemoglobin shows further heterogeneity, particularly in fetal life, and this has important implications for understanding the thalassemias and for approaches to their prenatal diagnosis. Hb F is a mixture of molecular forms with the formulas $\alpha_2\gamma_2^{136Gly}$ and $\alpha_2\gamma_2^{136Ala}$. The γ chains containing glycine at position 136 are designated $^{G}\gamma$-chains, and those containing alanine at this position are called $^{A}\gamma$-chains. At birth the

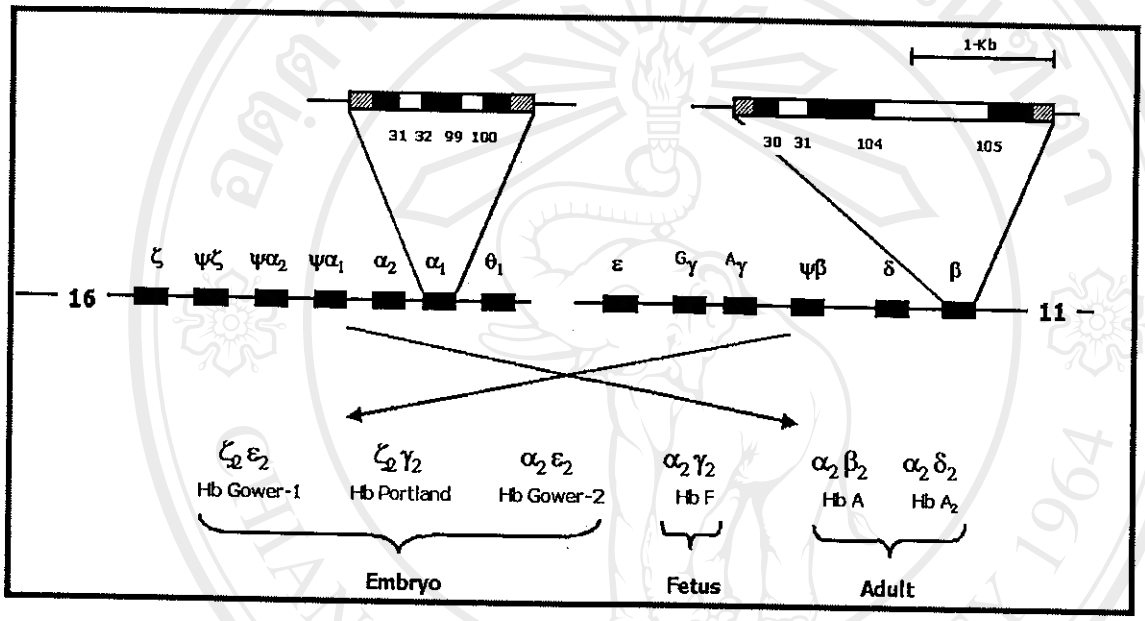

Figure 4. The genetic control of human hemoglobin. The main globin gene clusters are on chromosome 11 and 16. At each stage of development different genes in these clusters are activated or repressed. The different globin chains directed by individual genes are synthesized independently and combine with each other in a random fashion as indicated by the arrows (21).

ratio of molecules containing $^{G}\gamma$ chains to those containing $^{A}\gamma$-chains is about 3:1; this ratio varies widely in the trace amounts of hemoglobin F present in normal adults.

Before the eight-week of intrauterine life, there are three embryonic hemoglobins, Hb Gower-1 ($\zeta_2\varepsilon_2$), Hb Gower-2 ($\alpha_2\varepsilon_2$), and Hb Portland ($\zeta_2\gamma_2$). The ζ and ε chains are the embryonic counterparts of the adult α and β, γ, and δ chains, respectively. ζ-Chain synthesis persists beyond the embryonic stage of development in some of the α thalassemias; so far, persistent ε-chain production has not been found in any of the thalassemia syndromes. During fetal development there is an orderly switch from ζ to α and from ε- to γ-chain production, followed by β- and δ-chain production after birth.

1.2.4 Globin gene clusters (21)

Although there is some individual variability, the α-gene cluster usually contains one functional ζ gene and two α-genes, designated α_2 and α_1. It also contains four pseudogenes; $\psi\zeta_1$, $\psi\alpha_1$, $\psi\alpha_2$, and θ_1. The latter is remarkably conserved among different species. Although it appears to be expressed in early fetal life, its function is unknown; it seems unlikely that it can produce a viable globin chain. Each α-gene is located in a region of homology approximately 4-kb long, interrupted by two small nonhomologous regions(22). It is thought that the homologous regions have resulted from gene duplication and that the nonhomologous segments may have arisen subsequently by insertion of DNA into the noncoding regions around one of the two genes. The exons of the two α-globin genes have identical sequences. The first intron in each gene is identical, but the second intron of α_1 is nine bases longer and differs by three bases from that in the α_2 gene(23). Despite their high degree of homology, the sequences of the two α-globin genes diverge in their 3' untranslated regions 13 bases beyond the TAA stop codon. These differences provide an opportunity to assess the relative output of the genes, an important part of the analysis of the α-thalassemias(24). It appears that the production of α_2-mRNA exceeds that of α_1 by a factor of 1.5 to 3. ζ

$_1$- and ζ_2-genes are also highly homologous. The introns are much larger than those of α-globin genes, and, in contrast to the latter, IVS-1 is larger than IVS-2. In each ζ-gene, IVS-1 contains several copies of a simple repeated 14-bp sequence that is similar to sequences located between the two ζ-genes and near the human insulin gene. There are three bases changes in the coding sequence of the first exon of ζ_1, one of which gives rise to a premature stop codon, thus making it an inactive pseudogene.

The regions separating and surrounding the α-like structural genes have been analyzed in detail. Of particular relevance to thalassemia is the fact that this gene cluster is highly polymorphic (25). There are five hypervariable regions in the cluster: one downstream from the α_1-gene, one between the ζ- and $\psi\zeta$-genes, one in the first intron of both the ζ-genes, and one 5' to the cluster. These regions have been found to consist of varying numbers of tandem repeats of nucleotide sequences. Taken together with the single-base restriction fragment length polymorphisms (RFLPs), the variability of the α-globin gene cluster reaches a heterozygosity level of approximately 0.95. Thus, it is possible to identify each parental α-globin gene cluster in the majority of persons. This heterogeneity has important implications for tracing the history of the thalassemia mutations.

The arrangement of the β-globin gene cluster on the short arm of chromosome 11 is shown in Figure 4. Each of the individual genes and their flanking regions has been sequenced (26-29). Like the α_1 and α_2 gene pairs, the $^{G}\gamma$ and $^{A}\gamma$ genes share a similar sequence. In fact, the $^{G}\gamma$ and $^{A}\gamma$ genes on one chromosome are identical in the region 5' to the center of the large intron yet show some divergence 3' to that position. At the boundary between the conserved and divergent regions, there is a block of simple sequence that may be a "hot spot" for the initiation of recombination events that have led to unidirectional gene conversion.

Like the α-globin genes, the β-globin gene cluster contains a series of single-point RFLPs, although in this case no hypervariable regions have been identified. The arrangement of RFLPs, or haplotypes, in the β-globin gene cluster falls into two

domains. On the 5' side of the β-gene, spanning about 32-kb from the ε-gene to the 3' end of the ψβ gene, there are three common patterns of RFLPs. In the region encompassing about 18-kb to the 3' side of the β-globin gene, there are also three common patterns in different populations. Between these regions there is a sequence of about 11-kb in which there is randomization of the 3' and 5' domains and hence where a relatively higher frequency of recombination may occur. The β-globin gene haplotypes are similar in most populations but differ markedly in individuals of African origin; these findings suggest that these haplotype arrangements were laid down very early during evolution, and they are consistent with data obtained from motochondrial DNA polymorphisms that point to the early emergence of a relatively small population from Africa with subsequent divergence into other racial groups. Again, they are extremely useful for analyzing the population genetics and history of the thalassemia mutations.

The regions flanking the coding regions of the globin genes contain a number of conversed sequences that are essential for their expression (30). The first is the TATA box, which serves accurately to locate the site of transcription initiation at the CAP site, usually about 30 bases downstream, and also appears to influence the rate of transcription. In addition, there are two so-called upstream promoter elements; 70 or 80 base pairs (bp) upstream is a second conserved sequence, the CCAAT box, and further 5', approximately 80 to 100-bp from the CAP site, is a CACCC homology box that can be either inverted or duplicated (30). These promoter sequences are also required for optimal transcription, and, as we shall see later, mutations in this region of the β-globin gene cause its defective expression. The globin genes also have conserved sequences in their 3' flanking regions, notably AATAAA, which is the polyadenylation signal site.

1.2.5 Developmental changes in globin gene expression (21)

One aspect of the human globin genes that is of particular importance is the regulation of the switch from fetal to adult hemoglobin. Since many of the thalassemias and related disorders of the β-globin gene cluster are associated with persistent

γ-chain synthesis, a full understanding of their pathophysiology must include an explanation for this important phenomenon, which plays a considerable role in modifying their phenotypic expression.

The complex topic of hemoglobin switching has been the subject of several extensive reviews (20,31). β-Globin synthesis commences early during fetal life, at approximately 8 to 10 weeks' gestation. Subsequently, it continues at a low level, approximately 10% of the total non-α-globin chain production, up to about 36 weeks' gestation, after which it is considerably augmented. At the same time, γ-globin chain synthesis starts to decline so that at birth there are approximately equal amounts of γ- and β-globin chains produced. Over the first year of life there is a gradual decline in γ-chain synthesis, and by the end of the first year this amounts to less than 1% of the total non-α-globin chain output. In adults the small amount of Hb F is confined to an erythrocyte population called F cells (Figure 5).

It is still not clear how this series of developmental switches is regulated. It is not organ specific but is synchronized throughout the developing hematopoietic tissues. Although environmental factors may be involved, the bulk of experimental evidence suggests that there is some form of "time clock" built into the hematopoietic stem cell. At the chromosomal level it appears that regulation occurs in a complex manner involving both developmental stage-specific *trans*-activating factors and the relative proximity of the different genes of the β-globin gene cluster to LAR. The elements involved in the stage-specific regulation of the human globin genes have not yet been identified.

Fetal hemoglobin synthesis may be reactivated at a low level in states of hematopoietic stress and occurs at higher levels in certain hematologic malignancies, notably juvenile myeloid leukemia. However, it is only in the hemoglobinopathies that high levels of Hb F production are seen with any consistency in adult life.

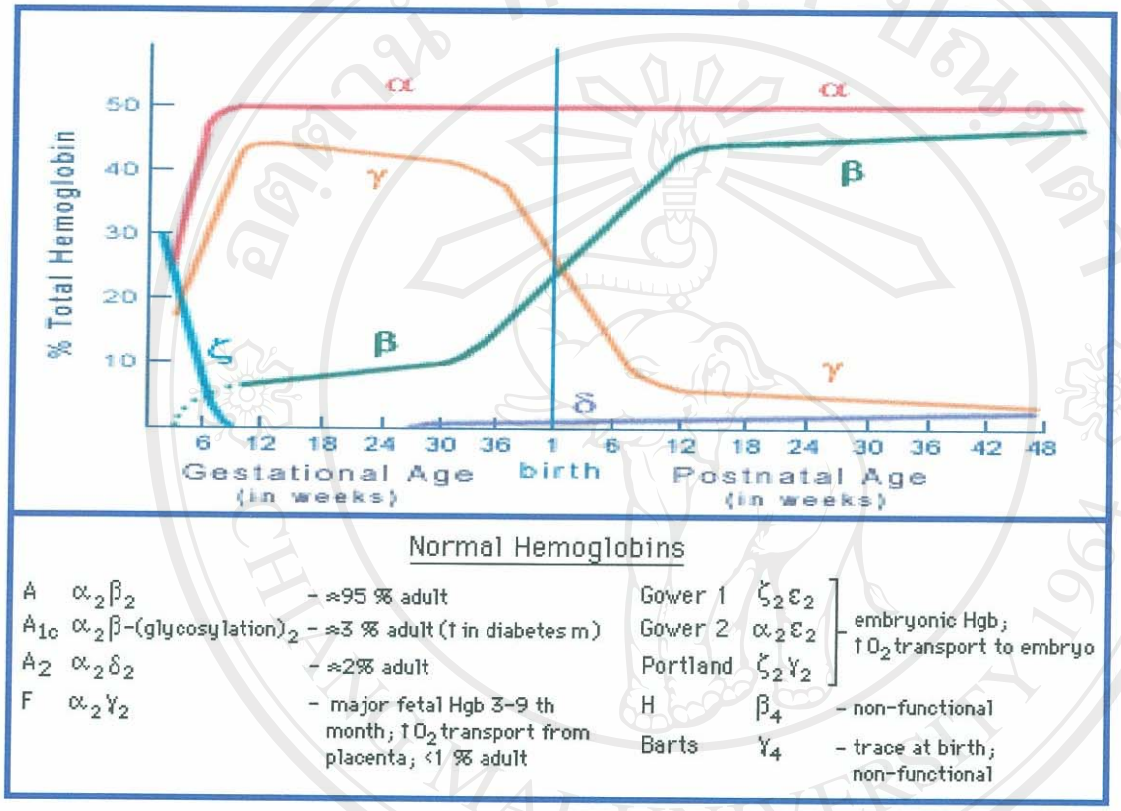

Figure 5. The changes in hemoglobin composition pattern during gestation and development

(Available from: http://www.thalassemia.com/medical/definition.shtml. Accessed July 14, 2002)

1.2.6 The thalassemias (32)

Thalassemia can be defined as a condition in which a reduced rate of synthesis of one or more of the globin chains leads to imbalanced globin-chain synthesis, defective hemoglobin production, and damage to the red cells or their precursors from the effects of the globin subunits that are produced in relative excess.

α-Thalassemia: Because there are two α-globin genes per haploid genome, four in all, the α-thalassemias are classified according to the relative output of both α-genes. When both α-globin genes on a chromosome are inactivated, the condition is called "α^0-thalassemia." The heterozygous genotype can be written --/αα. When one of the linked α-globin genes is inactivated, the condition is called "α^+-thalassemia", and the genotype can be written -α/αα in cases in which one of the α-globin genes in deleted, or α^Tα/αα if one of the linked α genes is inactivated by a mutation. In other words, α^0- and α^+-thalassemia describe an α-globin haplotype, that is the state of the *two* linked α-globin genes on a particular chromosome: in α^0-thalassemia, there is no output of α-globin from the particular chromosome; in α^+-thalassemia, there is some output, but usually only the product of a single α-globin locus. In some description of α-thalassemia, the less logical terms α-thalassemia-1 and -2 are used to describe α^0- and α^+-thalassemia, respectively.

β-Thalassemia: There are two main varieties of β-thalassemia, β^0-thalassemia, in which no β-globin chains are produced, and β^+-thalassemia, in which some β-chains are produced but at a reduced rate. Some forms of β-thalassemia are designated β^{++} to indicate that the defect in β-chain production is particularly mild.

The diagnostic feature of β-thalassemia is an elevated level of Hb A_2 in heterozygotes, which is found in most form of β^0- and β^+-thalassemia. There are, however, less common forms of β-thalassemia in which the Hb A_2 level is normal in heterozygotes. These so-called normal Hb A_2 β-thalassemia are also heterogeneous. They are classified into two varieties: type 1, in which there are no associated hematological changes; and type 2, in which the hematological findings are typical of

β-thalassemia trait with a rised Hb A_2 level. Type 1 normal Hb A_2 β-thalassemia is also called silent β-thalassemia. Both these forms of β-thalassemia with normal Hb A_2 levels are heterogeneous at the molecular level.

δβ-Thalassemia: The δβ-thalassemias are also heterogeneous. In some cases, no δ- or β-chains are synthesized. In the part, it was customary to classify these conditions according to the structure of the γ-chains of the Hb F that are produced, $^{G}\gamma^{A}\gamma(\delta\beta)^{0}$- and $^{G}\gamma(\delta\beta)^{0}$-thalassemia, for example. This was illogical and out of line with the classification of thalassemia according to the chain that is ineffectively synthesized. Thus, these conditions are best described as $(\delta\beta)^{0}$- and $(^{A}\gamma\delta\beta)^{0}$-thalassemias.

There are also $(\delta\beta)^{+}$ forms of δβ-thalassemia. In many of these conditions, an abnormal hemoglobin is produced that has normal α-chains combined with non-α-chains that are constituted by the N-terminal residues of the β-chain. These δβ fusion variants, collectively called the Lepore hemoglobins, are synthesized inefficiently and produce the clinical phenotype of δβ-thalassemia.

δ-Thalassemia: Several different mutations give rise to a reduced output of δ-chains and hence a reduced level of Hb A_2. These conditions are clinically silent and are of importance only insomuch that when they are inherited together with β-thalassemia, they may prevent an elevation of the level of Hb A_2.

εγδβ-Thalassemia: This rare form of thalassemia results from loss of either the whole or part of the β-like globin gene cluster. Homozygotes have not been encountered, presumably because the condition would not be compatible with life; heterozygotes have the clinical phenotype of β thalassemia with a normal Hb A_2 level.

γ-Thalassemia: There have been a few reports of deletions involving one or the other γ-globin genes. They have been identified by determining the level of $^{G}\gamma$- and $^{A}\gamma$-chains in Hb F and do not appear to be of clinical significance.

Hereditary Persistence of Fetal Hemoglobin (HPFH): As a form of β- or δβ-thalassemia. This is another heterogeneous group of disorders of hemoglobin synthesis that are characterized by persistent fetal hemoglobin synthesis in adult life in

the absence of major hematological abnormalities. By virtue of their interaction with the β-thalassemias, and from other evidence, it is apparent that many of these conditions are extremely well compensated forms of β- or δβ-thalassemia field because of the way in which it can modify the clinical phenotype of the β-thalassemias.

1.2.7 The molecular basis of α-thalassemia (21)

The most common mechanism that produces α-thalassemia is gene deletion that removes one or both α-globin genes from a single chromosome. However, the less common defects that lead to the impairment of α-globin production and do not involve gross deletion are associated as the nondeletion type of α-thalassemia.

The different classes of α-thalassemia are summarized in Table 2. The α-globin gene haplotype can be written αα, indicating the α_2 and α_1 genes, respectively. A normal individual has the genotype αα/αα. A deletion involving one (-α) or both (--) α genes can be further classified on the basis of its size, written as a superscript; thus, $-\alpha^{3.7}$ indicates a deletion of 3.7 kb including one α gene. When the sizes of the deletions have not yet been established, a superscript describing their geographic or family origin in useful; thus, $--^{SEA}$ describes a deletion of both a genes first identified in individuals of Southeast Asian origin. In those thalassemia haplotypes where both genes are intact, that is, nondeletion lesions, the nomenclature $\alpha^T\alpha$ is given, the superscript *T* indicating that this gene is thalassemic. However, when the precise molecular defect is known, as in Hb Constant Spring, for example, $\alpha^T\alpha$ can be replaces by the more informative $\alpha^{CS}\alpha$.

1.2.7.1 α-thalassemia-1

To date, 20 deletions (33) that involve both α-genes, and therefore abolish α-chain production from the affected chromosome, have been described (Figure 6).

Table 2. Classes of mutations that cause α-thalassemia (21)

α^0-thalassemia
Deletions involving α-globin gene cluster
Truncations of telomeric region of 16p
Deletions of HS40 region
α^+-thalassemia
Deletions involving α_2 or α_1 genes
Point mutations involving α_2 or α_1 genes
mRNA processing
IVS-I donor
IVS-I acceptor
Poly(A) signal
mRNA translation
Initiation codon
Exon I or II
Termination codon
Post-translational
Unstable α-globin
α-thalassemia mental retardation
ATR-16
Deletions or telomeric truncations of 16p
Translations
ATR-X
Mutations of XH2
Deletions
Missense
Nonsense
Splice site

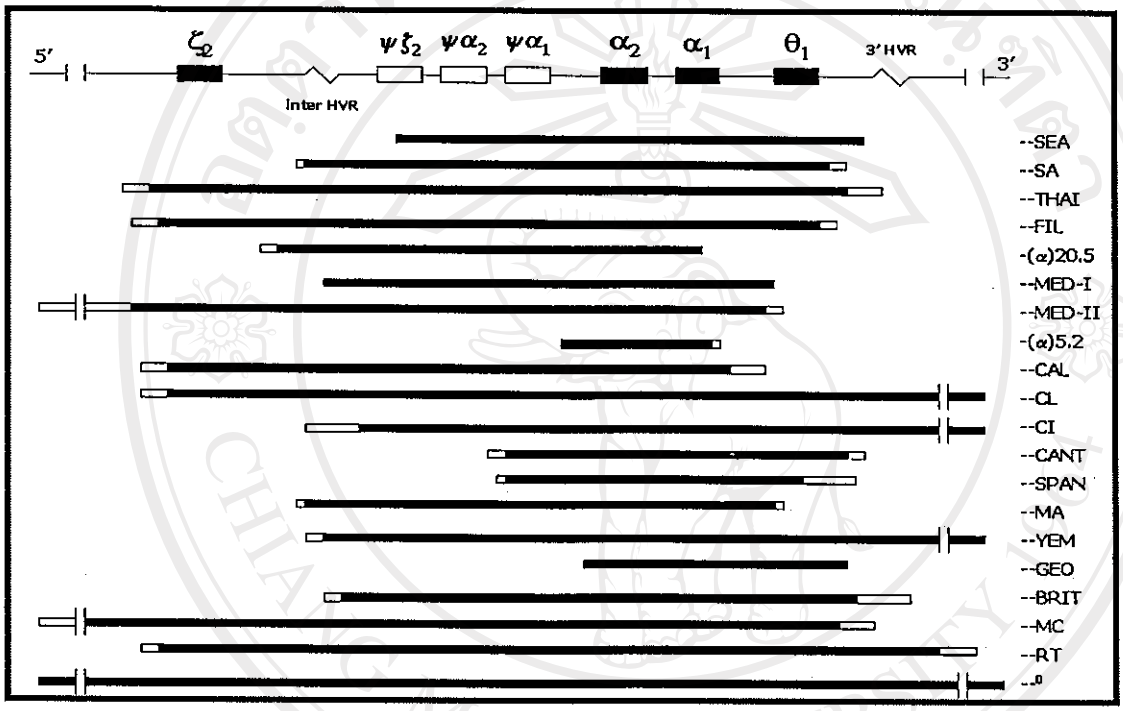

Figure 6. The deletions of the α-globin genes that result in α-thalassemia-1 (30)

Several of the 3' breakpoints fall within a 6- to 8-kb region at the 3' end of the α-globin complex, suggesting that this may represent a breakpoint cluster region with a high level of recombination. In at least five of the deletions, the 5' breakpoints also appear to cluster. This gives rise to a situation in which the 5' breakpoints are located approximately the same distance apart and in the same order a long a chromosome as their respective 3' breakpoints. It is possible that such staggered deletions may have arisen from illegitimate recombination events that delete an integral number of chromatin loops as they pass through their nuclear attachment points during replication, a mechanism that has also been suggested to underlie some of the deletion forms of HPFH. One of these deletions ($--^{MED}$) involves a more complex rearrangement that introduces a new piece of DNA bridging the two breakpoints in the α-gene cluster. This new sequence originates upstream from the α cluster and appears to have been replicated into the junction in a manner that suggests that the upstream segment of DNA also lies at the base of a replication loop. At least some of these deletions seem to have arisen by recombination events between *Alu* repeat sequences.

Several other mechanisms for the generation of α^0-thalassemia have been identified (20). In some cases this condition results from a terminal truncation of the short arm of chromosome 16 to a site 50-kb distal to the α-globin genes. It is interesting to note that the telomeric consensus sequence $(TTAGGGG)_n$ has been added directly to the site of the break. Since this mutation is stably inherited, it appears that telomeric DNA alone is sufficient to stabilize the broken chromosome end. This observation raises the possibility that other genetic diseases may result from chromosomal truncations.

Several deletions have now been identified that appear to down-regulate the α-globin genes by removal of the α-globin locus control region (HS40)(20). In each case the α-globin genes have been left intact, although in one the 3' breakpoint is found between the ζ_2 and $\psi\zeta_1$ genes, thus removing the ζ_2 gene. It appears that these deletions completely inactive the α-globin gene complex, just as deletions of the β-globin LAR inactivate the entire β-gene complex. So far, such deletions have not been observed in the homozygous state, presumably because they would be lethal.

1.2.7.2 α-Thalassemia-2

The most common forms of α^+-thalassemia ($-\alpha^{3.7}$ and $--\alpha^{4.2}$) involve the deletion of one or the other of the duplicated α-globin genes (Figure 7).

Each α-gene is located within a region of homology approximately 4-kb long, interrupted by two nonhomologous regions. It is thought that the homologous regions have resulted from an ancient duplication event and that subsequently they were subdivided, presumably by insertions and deletions, to give three homologous subsegments referred to as X, Y, and Z (Figure 8). 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 chromosomes with either single ($-\alpha$) or triplicated ($\alpha\alpha\alpha$) α-globin genes. Such an occurrence between homologous Z boxes deletes 3.7-kb of DNA (rightward deletion, $-\alpha^{3.7}$), while a similar crossover between the two X blocks deletes 4.2-kb of DNA (leftward deletion, $--\alpha^{4.2}$). The corresponding triplicated α-gene arrangements are referred to as $\alpha\alpha\alpha^{anti\text{-}3.7}$ and $\alpha^{anti\text{-}4.2}$(34;35). More detailed analysis of these crossover events indicates that they occur more commonly in the Z box, and at least three different $-\alpha^{3.7}$ deletions have been found, depending on exactly where the crossover has taken place. These are designated $-\alpha^{3.7I}$, $-\alpha^{3.7II}$, and $-\alpha^{3.7III}$, respectively. Other, rarer deletions of a single α-gene have been observed.

1.2.7.3 Nondeletion α-thalassemia

Since the expression of the α_2 gene is two to three times greater than that of the α_1 gene, it is not surprising that most of the nondeletion mutants discovered to date affect predominantly the expression of the α_2 gene; presumably this is ascertainment bias because of the greater phenotypic effect of these lesions. It is also possible that they have come under greater selective pressure. Like the β-thalassemia mutations, α-thalassemia mutations can be classified according to the level of gene expression they affect (Table 2). Several processing mutations have been identified. For example,

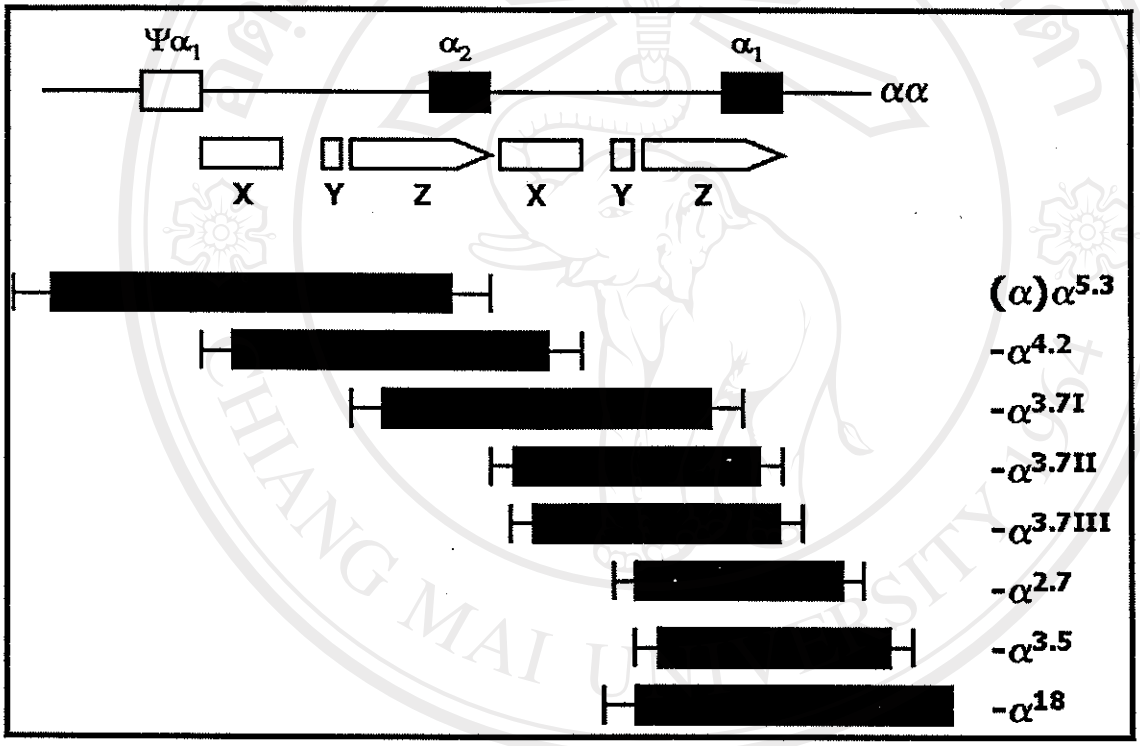

Figure 7. The deletions that underlie the α-thalassemia-2 (21)

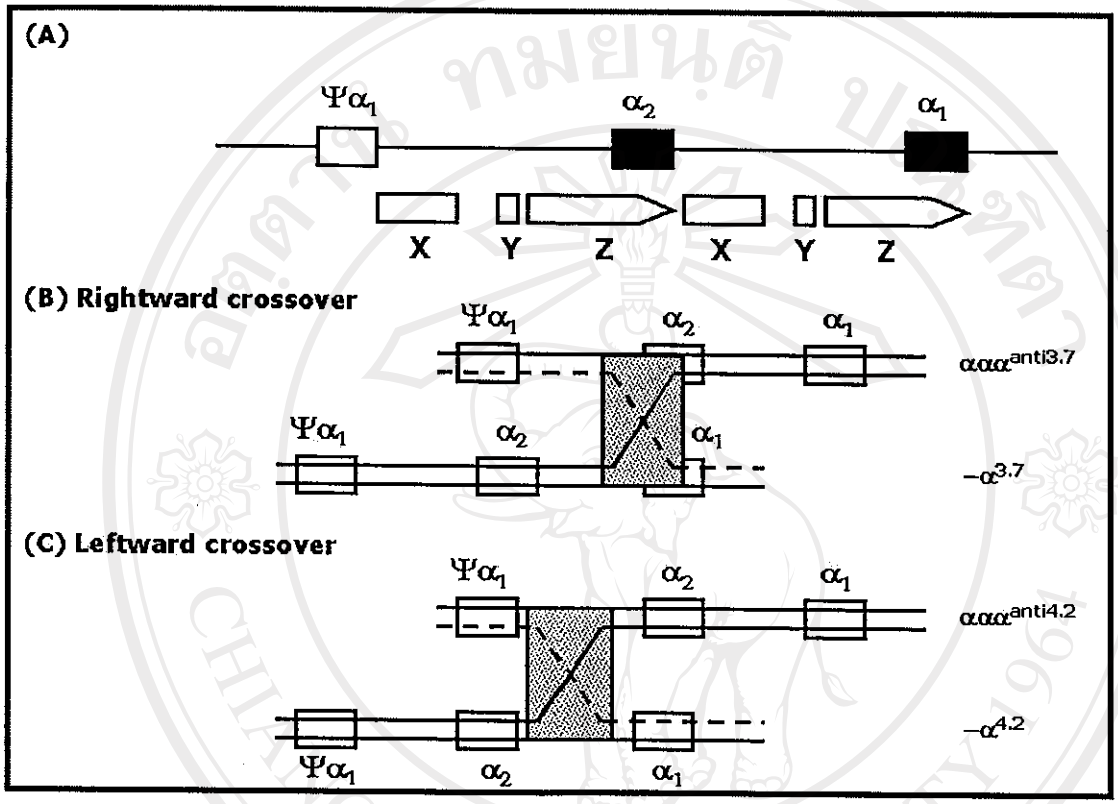

Figure 8. The mechanisms for the production of the common deletion forms of α-thalassemia-2 (21)

(A) the normal α-globin gene cluster showing the homology boxes X, Y, and Z

(B) the rightward crossover through the Z boxes, giving rise to the 3.7-kb deletion and a chromosome with 3 α-globin genes

(C) the leftward crossover through the Z boxes, giving rise to the 4.2-kb deletion and a chromosome containing 3 α-genes

a pentanucleotide deletion includes the 5' spice site of IVS-1 of the α_2-globin gene. This involves the invariant GT donor splicing sequence and thus completely inactivates the α_2 gene. A second mutant of this type, found commonly in the Middle East, involves the poly-A addition signal site (AATAAA $\rightarrow$ AATAAG) and downregulates the α_2 gene by interfering with 3' end processing.

A second group of nondeletion α-thalassemias result from mutations that interfere with the translation of mRNA. In one case, for example, the initiation codon is inactivated by a T $\rightarrow$ C transition, and, in another, efficiency of initiation is reduced by a dinucleotide deletion in the consensus sequence around the start signal. Five mutations that affect termination of translation and give rise to elongated α chains have been identified: Hb Constant Spring, Icaria, Koya Dora, Seal Rock, and Pakse. Each specially changes the termination codon TAA so that an amino acid is inserted instead of the chain terminating. This is followed by read-through of mRNA that is not normally translated until another "in-phase" stop codon is reached. Thus, each of these variants has an elongated α chain. It seem likely that the "read-through" of α-globin mRNA that is usually not utilized somehow reduces its stability. There are several nonsense mutations, one in exon 3 of the α_2-globin gene, for example (36). Finally, there are several mutations that cause α-thalassemia by producing highly unstable α-globin chains; they include Hb Quong Sze (37), Suan Dok (38), Petah Tikvah (39), and Evanston.

1.2.8 **Clinical manifestations** (15)

1.2.8.1 **Silent carrier (α-thalassemia-2 trait)**

α-thalassemia-2 trait has no consistent hematologic manifestations. The red cells are not microcytic, and Hb A_2 and Hb F are normal. During the newborn period, small amounts ($\leq$3%) of Hb Bart's (γ_4) may be seen by electrophoresis or column chromatography. This condition is most often recognized when an apparently normal individual becomes the parent of a child with Hb H disease after mating with a person

with α-thalassemia-1 trait. The mild excess of β-globin chains is probably removed in erythroblasts by proteolysis. α-Thalassemia-2 is particularly common in Melanesia, as well as in Southeast Asia and in American blacks, reaching a prevalence of >80% in north coastal Papua New Guinea. At the molecular level, α-thalassemia-2 has been found to be associated with two common gene deletions resulting from different nonhomologous crossing over events between the two linked α-globin genes: a 3.7-kb rightward deletion ($-\alpha^{3.7}$) resulting in a fused $\alpha_2\alpha_1$-globin gene, and a 4.2-kb leftward deletion ($-\alpha^{4.2}$) resulting in loss of the 5' (α_2) gene.

1.2.8.2 **α-Thalassemia trait (α-thalassemia-1 trait)**

α-Thalassemia-1 trait is characterized by levels of Hb A_2 in the low to low normal range (1.5-2.5%) and β/α synthetic ratios averaging 14:1. During the perinatal period, elevated amounts of Hb Bart's are noted (3-8%). Microcytosis is present in cord blood erythrocytes.

Studies of newborns from the archipelago of Vanuatu in the southwest Pacific and from Papua New Guinea, indicate that homozygotes for the rightward $-\alpha^{3.7III}$ deletion (where only a fused $\alpha_2\alpha_1$-globin gene, mostly of the α_2 type, remains) have lower Hb Bart's levels (3.50.±8%) than those of infants homozygous for the leftward $-\alpha^{4.2}$ deletion (where only the α_1-globin gene remains) (6.0 ±1.4%). These results suggest that the 5' α_2-globin gene has a higher output than the 3' α_1-globin gene, a conclusion supported by direct measurement of α_2 / α_1 mRNA ratios(24).

Hb H is not detected in hemolysates of peripheral red cells, probably because of rapid proteolysis of Hb H or free β-globin chains. However, about 1% of erythroblasts and marrow reticulocytes have inclusions. When an α-thalassemia gene occurs in persons who are also heterozygous for β-globin chain variant hemoglobins, such as Hb S, Hb C, or Hb E, the proportion of the abnormal hemoglobin is lower than seen in simple heterozygotes. The lower level of the abnormal hemoglobin is due to post-translational control because of higher affinity of β^A chains for a limited pool of α-globin chains, coupled with proteolysis of the uncombined $\beta^{variant}$ chains.

1.2.8.3 Hb H disease

Hb H disease is associated with a moderately severe but variable anemia, resembling thalassemia intermedia with osseous changes and splenomegaly, however, the clinical phenotype may be considerably milder. It occurs predominantly in Asians and occasionally in whites (Mediterraneans) but is rare in blacks.

Because Hb H is unstable and precipitates within the circulating red cell, hemolysis occurs. Hb H can be demonstrated by incubation of blood with supravital oxidizing stains such as 1% brilliant cresyl blue. Multiple small inclusions form in the red cells. Electrophoresis of a freshly prepared hemolysate at alkaline or neutral pH demonstrates a fast-moving component iron deficiency may reduce the amount of Hb H in the patient's red cells. A syndrome of Hb H disease associated with mental retardation, other congenital anomalies, and large deletions on chromosome 16 has been noted in several white families (Figure 9).

1.2.8.4 Hydrops fetalis with Hb Bart's

Hb Bart's hydrops fetalis occurs almost exclusively in southeastern Asians, especially Chinese, Cambodians, Thais, and Filipinos. Affected fetuses usually are born prematurely and either are stillborn or die shortly after birth. Marked anascara and enlargement of the liver and spleen are present. Severe anemia usually is present, with hemoglobin levels of 3-10 g / dl. The red cells are markedly microcytic and hypochromic and include target cells and large numbers of circulating nucleated red blood cells. These morphologic abnormalities and a negative Coombs test exclude hemolytic diseases due to blood group incompatibility. Hemoglobin electrophoresis reveals predominantly Hb Bart's, with a smaller amount of Hb H. a minor component identified as Hb Portland ($\zeta_2\gamma_2$) migrating in the position of Hb A is seen also. Normal Hb A and Hb F are totally absent.

Hydropic infants have massive hepatosplenomegaly. Extreme extramedullary erythropoiesis occurs in response to the profound hypoxia and hemolytic anemia characteristic of this disease. The universal edema characteristic of the hydropsfetalis syndrome is a reflection of severe congestive heart failure and hypoalbuminemia in

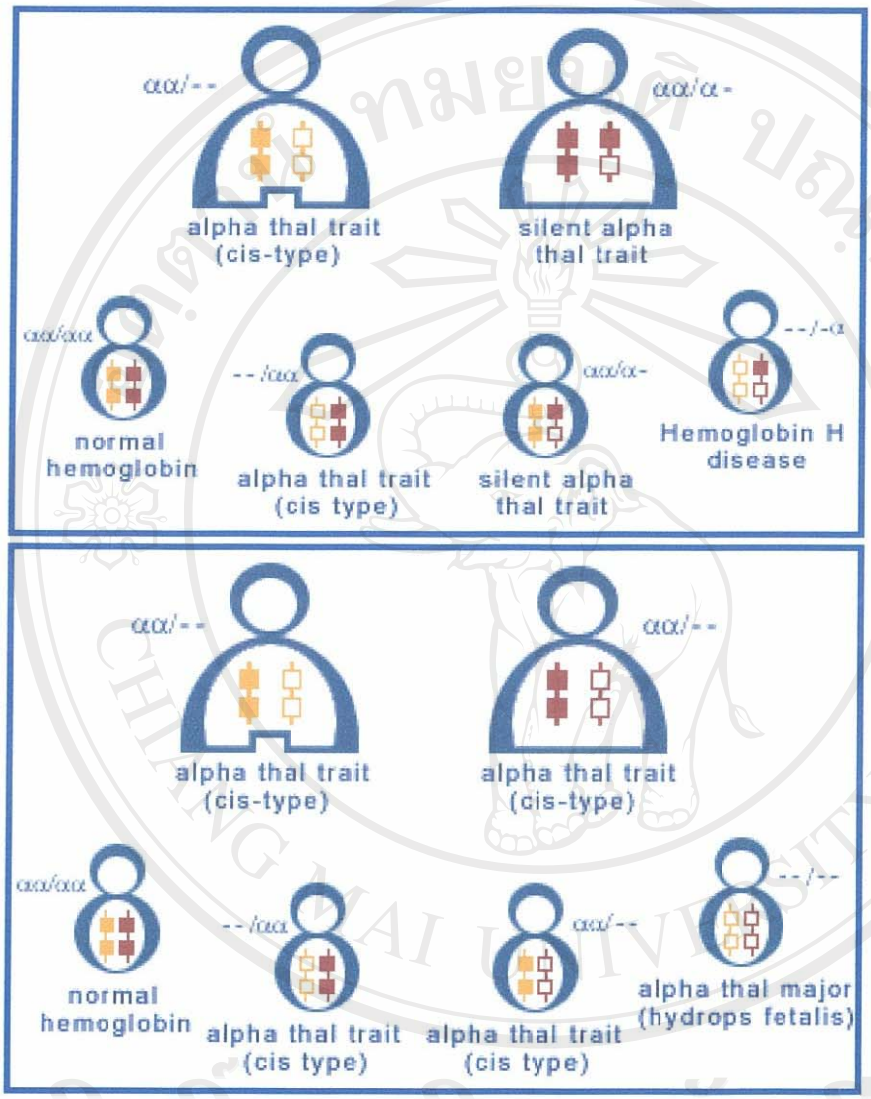

Figure 9. The genetics of α-thalassemia: The α-globin genes are represented as boxes. The open boxes represent gene deletions or otherwise inactivated α-genes. The black boxes represent normal α-genes (Available from: http://www.thalassemia.com/genetics/inherit3.shtml. Accessed July 14, 2002)

utero This is partly a consequence of anemia, but the strikingly abnormal oxygen affinity of the tetrameric Hb Bart's is probably the most important determinant of the severe tissue hypoxia. The oxygen dissociation curve of Hb Bart's lacks the normal sigmoid form due to noncooperativity during oxygen loading and unloading and is markedly shifted to the left. The shift is so great that little oxygen is released under conditions of low oxygen concentration in the tissues.

Infants with this syndrome do not die in an earlier trimester of pregnancy because of the presence of Hb Portland ($\zeta_2\gamma_2$). This hemoglobin does display cooperativity in a manner similar to that of Hb F and therefore has a much more favorable oxygen dissociation pattern than that of Hb Bart's. A high incidence of toxemia of pregnancy has been described in women carrying severely affected infants, providing an increased rationale for prenatal diagnosis of this condition.

1.2.9 The polymerase chain reaction (PCR) (40)

H. Ghobind Khorana and his colleagues first proposed the PCR in the early 1970s as a strategy to lessen the labor involved in chemical synthesis genes (41). Their ideas, however, did not seem practicable at a time when genes had not yet been sequenced, thermostable DNA polymerases had not been described, and synthesis of oligonucleotide primers was more of an art than a science. Not surprisingly, Khorana's ideas were quickly forgotten. The technique was independently conceived 15 years later, given its present name, and put into practice by Mullis and coworkers at Cetus Corporation, who described *in vitro* amplification of single-copy mammalian genes using the Klenow fragment of *Escherichai coli* DNA polymerase I (42-43). Even so, PCR would probably have remained a clumsy laboratory curiosity were it not for the discovery of thermostable DNA polymerases (44). The use of a thermostable polymerase from *Thermus aquaticus* (43) greatly increased the efficiency of PCR and opened the door to automation of the method.

1.2.9.1 General principles of PCR

The PCR technique was developed highly efficient amplification of DNA sequences of interest. In general, the procedure 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 strand. The DNA is denatured at high temperature and then reannealled in the presence of large molar excess of the oligonucleotides. The oligonucleotides, oriented with their 3' ends pointing toward each other, hybridize to opposite strand of the target sequence and prime enzymatic extension along the nucleic template in the presence of the 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 (Figure 10). This leads to the selective enrichment of specific DNA sequences so that they can be readily manipulated or detected.

1.2.9.2 Optimization of PCR

An ideal PCR would be the one with high specificity, efficiency, and fidelity. Studies indicate those essential components of PCR, the programming PCR, optional components of PCR, inhibitors of PCR and the type of DNA polymerase influence each of these five parameters. Unfortunately, adjusting conditions for maximum specificity may be not be compatible with high yield; likewise optimizing for the fidelity of PCR may result in reduced efficiency. Thus, when setting up a PCR, one should know which of the five parameters is the most important for its intended application and optimize PCR accordingly.

The programming PCR: PCR is an iterative process, consisting of 3 elements; denaturation of the template by heat, annealing of the oligonucleotide primers to the single-stranded target sequence, and extension of the annealed primers by a thermostable DNA polymerase. Typical denaturation conditions are 90-95^{o}C for 30-60

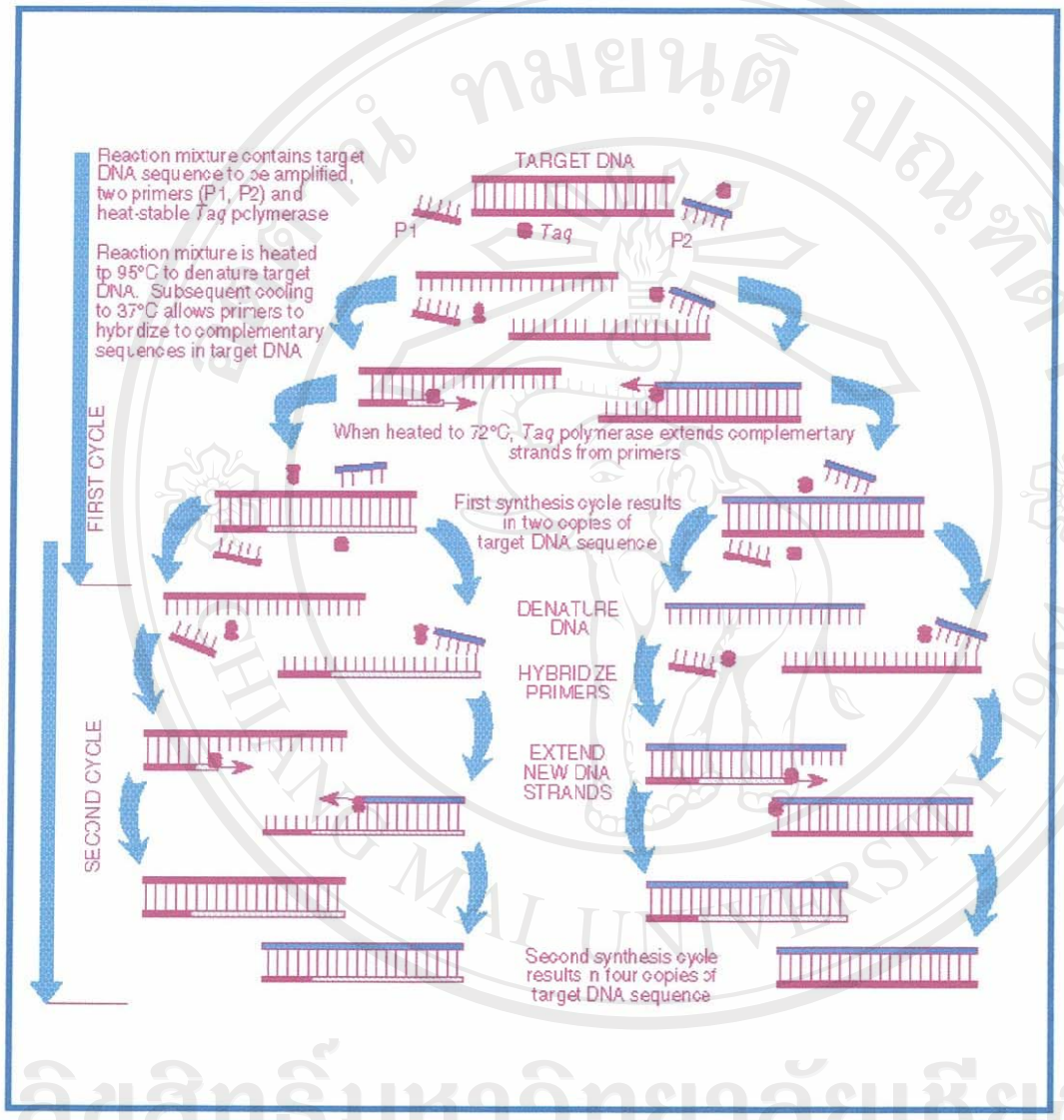

Figure 10. Sequence of amplification in the PCR

(Available from: http://aidshistory.nih.gov/imgarchive/pcr.html. Accessed September 9, 2003)

seconds. However, the higher temperature on the longer time may be appropriate, especially for G-C rich target. In contrast, too high or too long denaturation step led to unnecessary loss of enzyme activity. The half-life of *Taq* DNA polymerase is >2 hours at 92.5°C, 40 min at 97.5°C. Annealing is a very critical step for the specificity of the PCR products. The annealing temperature should be adjusted each reaction. Usually the temperatures in the range 55-72°C yield the best results. The estimate annealing temperature for each primer can be calculated by the following formula:

$2(A+T) + 4(G+C) - 5^{\circ}C$, or $5^{\circ}C$ below the true T_m of the primers.

Extension 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 extension is around 2 to 4 kb/min. Thus, 30 second usually enough for the extension of PCR products less than 1,000 nucleotides (45).

Inhibitors of PCR: Poor or nonexistent amplification may indicate the presence of inhibitors in the DNA sample. Numerous inhibitors of PCR have been described. These include ionic detergents (e.g., SDS and Sarkosyl) (46), phenol, heparin (47), polyanions such as spermidine (48), hemoglobin, and gel-loading dyes such as xylene cyanol, and bromophenol blue (49). Test for inhibitor in the template preparation by spiking original PCR mix with dilutions of known positive (demonstrably amplifiable) template. Re-extraction, ethanol precipitation, and/or centrifugal ultrafiltration may resolve the problem. Proteinase K carryover can serve to digest the *Taq* polymerase but is readily denatured by a 5-min incubation at 95°C.

Optional components of PCR: A number of cosolvents and additives have been reported to reduce unacceptably high levels of mispriming and to increase the efficiency of amplification of G+C rich templates. Cosolvents include formamide (1.25-10% v/v) (50), dimethylsulfoxide (up to 15% v/v) (51), and glycerol (1-10% v/v) (52). Additives include tetramethylammonium chloride (53), potassium glutamate (10-200 mM), ammonium sulfates (54), nonionic and cationic detergents (55), and certain as yet

unidentified "Specificity Enhancers" such as Perfect Match Polymerase Enhancer (Stratagene), Q-solution (QIAGEN), and GC-Melt (CLONTECH). Many of these additives and cosolvents inhibit PCR when used at high concentration, and the optimum concentration must be determined empirically for each combination of primers and template DNA.

Essential components of PCR:

1) A thermostable DNA polymerase to catalyze template-dependent synthesis of DNA. A wide choice of enzymes is now available that vary in their fidelity, efficiency, and ability to synthesize DNA products. For routine PCRs, *Taq* polymerase (0.5-2.5 units per standard 25-50-µl reaction) remains the enzyme of choice. The specific activity of most commercial preparations of *Taq* is ~80,000 units/mg of protein. Standard PCRs therefore contain 2×10^{12} to 10×10^{12} molecules of enzyme. Since the efficiency of primer extension with *Taq* polymerase is generally ~0.7, the enzyme becomes limiting when 1.4×10^{12} to 7×10^{12} molecules of amplified product have accumulated in the reaction.

2) A pair of synthetic oligonucleotides to prime DNA synthesis. Of the many factors that influence the efficiency and specificity of the amplification reaction, none are more crucial than the design of oligonucleotide primers. Careful design of primers is required to obtain the desired products in high yield, to suppress amplification of unwanted sequences, and to facilitate subsequent manipulation of the amplified product. Given that primers so heavily influence the success or failure of PCR protocols, it is ironic that the guidelines for their design are largely qualitative and are based more on common sense than on well-understood thermodynamic or structural principles.

3) Deoxynucleoside triphosphates (dNTPs). Standard PCRs contain equimolar amounts of dATP, dTTP, dCTP, and dGTP. Concentrations of 200-250 µM of each dNTP are recommended for *Taq* polymerase in reactions containing 1.5 mM $MgCl_2$. In a 50-µl reaction, these amounts should allow synthesis of ~6-6.5 µg of DNA, which should be sufficient even for multiplex reactions in which eight or more primer pairs are used at the same time. High concentrations of dNTPs (>4 mM) are inhibitory, perhaps because of

sequestering of Mg^{2+}. However, a satisfactory amount of amplified product can be produced with dNTP concentrations as low as 20 μM-0.5-1.0 pmole of an amplified fragment ~1-kb in length.

4) Divalent cations. It is beneficial to optimize the magnesium ion concentration. The magnesium concentration may effect all of the following: primer annealing, strand dissociation temperatures of both template and PCR product, product specificity, formation of primer-dimer artifacts, and enzyme activity and fidelity. *Taq* DNA polymerase requires free magnesium on top of that bound by template DNA, primers, and dNTPs. Although a concentration of 1.5-mM Mg^{2+} is routinely used, increasing the concentration of Mg^{2+} to 4.5-mM or 6-mM has been reported to decrease nonspecific priming in some cases. Note that the presence of EDTA or other chelators in the primer stocks or template DNA may disturb the apparent magnesium optimum.

5) Buffer to maintain pH. Tris-HCl, adjusted to a pH between 8.3 and 8.8 at room temperature, is included in standard PCRs at a concentration of 10 mM. When incubated at 72°C (the temperature commonly used for the extension phase of PCR), the pH of the reaction mixture drops by more than a full unit, producing a buffer whose pH is ~7.2.

6) Monovalent cations, standard PCR buffer contains 50 mM KCl and works well for amplification of segments of DNA >500 bp in length. Raising the KCl concentration to ~70-100 mM often improves the yield of shorter DNA segments.

7) Template DNA containing target sequences can be added to PCR in single- or double-stranded form. Closed circular DNA templates are amplified slightly less efficiently than linear DNAs. Although the size of the template DNA is not critical, amplification of sequences embedded in high molecular weight DNA (>10 kb) can be improved by digesting the template with a restriction enzyme that does not cleave within the target sequence.

When working at its best, PCR requires only a single copy of a target sequence as template. More typically, however, several thousand copies of the target DNA are seeded into the reaction. In the case of mammalian genomic DNA, up to 1.0 μg of DNA

is utilized per reaction, an amount that contains ~3 x 10^5 copies of a single-copy autosomal gene. The typical amounts of yeast, bacterial, and plasmid DNAs used per reaction are 10 ng, 1 ng, and 1 pg, respectively.

1.2.9.3 Selecting PCR primers

Listed below are several steps involved in the selection of oligonucleotide primers:

1) Analysis of the target gene for potential priming sites that are free of homopolymeric tracts, have no obvious tendency to form secondary structures, are not self-complementary, and have no significant homology with other sequences on either strand of the target genome.

2) Creation of lists of possible forward and reverse primers based on the criteria listed in the Table 3. The melting temperatures (T_m) of the nucleotides should obviously be calculated using the two T_m calculation equations that are far simpler to apply and are perfectly adequate for most purposes.

- The Wallace rule can be used to calculate the T_m for perfect duplexes 15-20 nucleotides in length in solvents of high ionic strength (e.g., 1 M NaCl):

$$T_m \text{ (in } ^\circ C) = 2(A+T) + 4(G+C)$$

; where (A+T) = the sum of the A and T residues in the oligonucleotide

(G+C) = the sum of the G and C residues in the oligonucleotide.

- The Baldino algorithm predicts reasonably well the T_m of oligonucleotides, 14-70 nucleotides in length, in cation concentrations of 0.4 M or less:

$$T_m \text{ (in } ^\circ C) = 81.5^\circ C + 16.6(\log_{10}[K^+]) + 0.41(\%[G+C]) - (675/n)$$

; where $[K^+]$ = the molar salt concentration (M)

and n = the number of bases in the oligonucleotide

3) Selection of well-matched pairs of forward and reverse primers that are similar in their content of G+C and will generate an amplified product of the appropriate size and base composition. The GC content of both primers and the amplified product should be similar and lie between 40% and 60%.

Table 3. Primer design: properties of oligonucleotides that influence the efficiency of amplification (40)

Property	Optimal design
Base composition	G+C content should be between 40% and 60%, with an even distribution of all 4 bases along the length of the primer.
Length	The region of the primer complementary to the template should be 18-25 nucleotides long. Members of a primer pair should not differ in length by >3 bp.
Repeated and self-complementary sequences	No inverted repeat sequences or self-complementary sequences > 3 bp in length should be present. Sequences of this type tend to form hairpin structures, which, if stable under PCR conditions, can effectively prevent the oligonucleotide from annealing to its target DNA.
Complementarity between members of a primer pair	The 3' terminal sequences of one primer should not be able to bind to any site on the other primer. Because primers are present at high concentration in PCR, even weal complementarity between them leads to hybrid formation and the consequent synthesis and amplification of primer dimers. If primer dimers form early in PCR, they can compete for DNA polymerase, primers, and nucleotides and so can suppress amplification of the target DNA. Formation of primer dimers can be reduced by careful primer design, by the use of hot start or touchdown PCR, and/or by the use of specially formulated DNA polymerases. When more than one primer pair is used in a single PCR, check that none of the 3' ends have detectable complementarity to any other primers in the reaction.
Melting temperatures (T_m)	Calculated T_m values of members of a primer pair should not differ by >5°C. The T_m of the amplified product should not differ from the T_m values of the primer pairs by >10°C. this property ensures that the amplified product will be efficiently denatured during each cycle of PCR.

Table 3.(continue) Primer design: properties of oligonucleotides that influence the efficiency of amplification

3' Termini	The nature of the 3' end of primers is crucial. If possible, the 3' base of each primer should be G or C. however, primers with a ...NNCG or ...NNGC sequence at their 3' termini are not recommended because the unusually high ΔG of the terminal GC bases promotes the formation of hairpin structures and may generate primer dimers.
Adding restriction sites, bacteriophage promoters, and other sequences to the 5' termini of primers	Useful sequences not complementary to the target DNA are commonly added to the 5' end of the primers. In general, the presence of such sequences does not significantly affect annealing of the oligonucleotide to its target DNA. These additional sequences include bacteriophage promoters and GC clamps. Restriction sites are a special case. Because the efficiency of cleavage of restriction sites located at the 5' termini of DNA molecules is poor, the primer should be extended by at least three additional nucleotides beyond the recognition sequence of the restriction enzyme.
Placement of priming sites	Depending on the purpose of the experiment, the placement of priming sites may be constrained by the location of mutations, restriction sites, coding sequences, microsatellites, or cis-acting elements. When designing primers for use on cDNA templates, it is best to use forward and reverse primers that bind to sequences in different exons. This allows amplification products derived from cDNA and contaminating genomic DNA to be easily distinguished.
Primers for degenerate PCR	When a short sequence of amino acids has been obtained by sequencing a purified protein, pools of degenerate oligonucleotides containing all possible coding combinations can be used to amplify the corresponding genomic or cDNA sequences.

4) Refining the length and/or placement of the oligonucleotides so that the 3'-terminal nucleotide is a G or a C. check that the two oligonucleotides do not display significant complementary. As a rule of thump, no more than three consecutive nucleotides on one primer should be complementary to the other primer.

1.2.9.4 Computer-assisted design of oligonucleotide primers

To avoid boredom, save time, and minimize problems, use computer programs to optimize the design, selection, and placement of oligonucleotide primers. Many stand-alone computer programs can be obtained to search sequences for priming sites that fit a set of user-defined parameters. Such programs generate a hierarchy of potentially specific primers whose melting temperatures have been calculated, generally using the nearest-neighbor method, in which the thermodynamic stability of the primer: template duplex is derived from the sum of the stacking interactions of neighboring bases.

Most of the programs use graphic tools, employ user-friendly interfaces, and rank potential primers and primer pairs according to the weight assigned to various parameters. Some of the program contain, for example, facile searching of databases for unintentional matches to the primer, optimization of conditions for the amplification reaction, translation of amino acid sequences into populations of degenerate oligonucleotides, and elimination of primers capable of forming stable secondary structures.

1.2.10 Multiplex PCR (40,56)

Multiplex PCR is the term used when more than one pair of primers is used in a PCR. The goal of multiplex PCR is to amplify several segments of target DNA simultaneously and thereby to conserve template DNA, save time, and minimize expense. Since its first description in 1988 (11), this method has been successfully applied in many areas of DNA testing, including analyses of deletions (57), mutations (58) and polymorphisms (59), or quantitative assays (60) and reverse-transcription PCR (61).

Unfortunately, multiplex PCR is very tricky to set up. Great care must be taken to ensure that all of the primers in the reaction have approximately the same melting temperature, that the primers are unlikely to interact with one another, and that the amplified products are of approximately the same size but can be unambiguously distinguished from one another by gel electrophoresis. The multiplex reaction must then be optimized as follows:

- Check that all target loci can be amplified efficiently in separate reactions using the same PCR program.
- Titrate the amount of each primer pair to achieve maximum amplification in separate reactions using the same PCR program and reaction conditions.
- Balance the amount of each primer pair in the multiplex reaction to achieve acceptable amplification of all target regions.

The last requirement usually causes the most problems. Frequently, one or two regions of the target DNA yield little or no amplified product, whereas all other primer pairs behave perfectly. The best option is to increase progressively the concentration of the recalcitrant primer primers while reducing the concentration of the well-behaved primer pairs. Consider adjusting the reaction conditions only if this strategy fails; for example, alter the concentrations of Mg^{2+} or KCl in the reaction mixture. However, if the yields of the multiplex reaction are systematically biased in favor of longer PCR products, the best option is to re-optimize the amount of each primer pair in a series of multiplex reactions containing increased concentrations of KCl (1.0-2.0-fold) and a constant concentration of Mg^{2+} (1.5 mM). Conversely, if the yields are biased in favor of smaller PCR products, increase the concentration of Mg^{2+} progressively (up to 4.5 mM) while maintaining a constant concentration of KCl.

1.2.11 DNA sequencing (62)

1.2.11.1 The dideoxy method of DNA sequencing

The Sanger technique uses controlled synthesis of DNA to generate fragments that terminate at specific points along the target sequence. A synthetic oligonucleotide

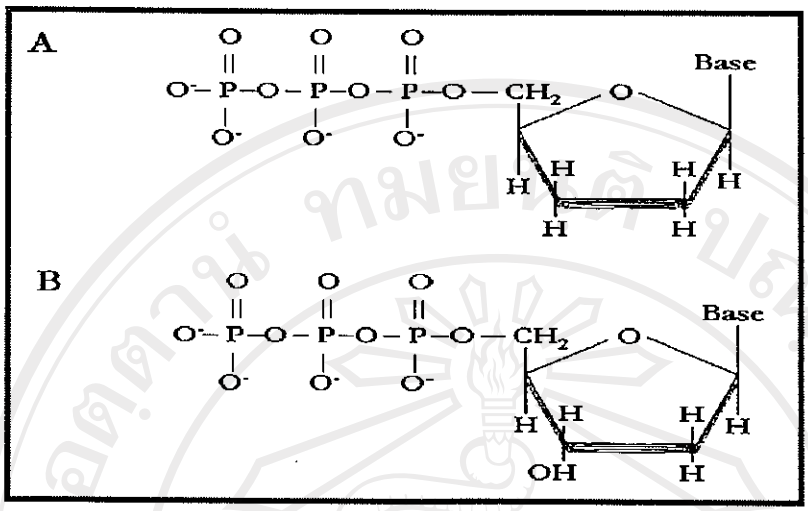

Figure 11. The structure of dideoxynucleoside triphosphates (ddNTPs) (62), which lacks a 3'-hydroxyl residue. (A), and the normal deoxynucleoside triphosphates (dNTPs) with the 3'-terminal nucleotide of DNA (B).

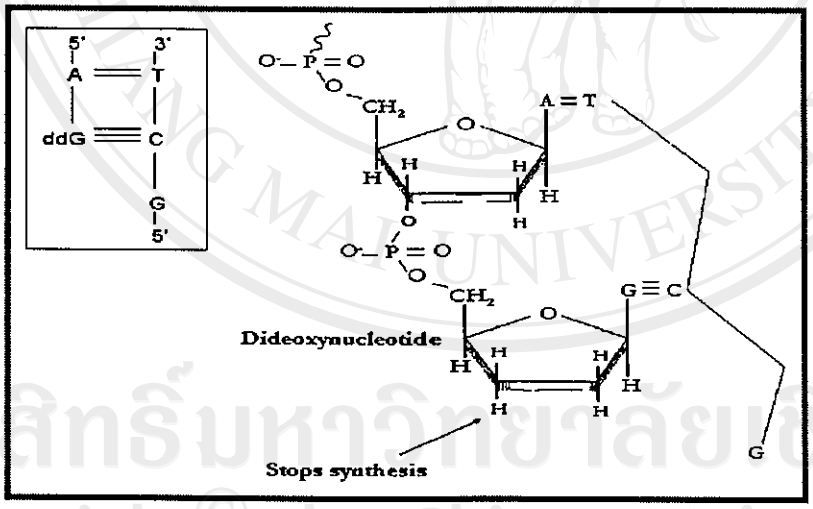

Figure 12. The ddNTP can be incorporated into the growing chain by DNA polymerase, where it acts as a terminator because it lacks the 3'-hydroxyl group required for formation of further 5'→3' phosphodiester bonds (62).

primer is annealed to a single-stranded DNA template. Four different sequencing reactions are set up each containing a DNA polymerase and the four normal dNTPs. One of the precursors or, in some cases, the primer is labeled radioactively with ^{32}P, ^{33}P, or ^{35}S or with a nonradioactive fluorescent tag. The four reactions also contain a small proportion of a 2',3'-ddNTP that carries a 3'-H atom on the deoxyribose moiety, rather than the conventional 3'-OH group (Figure 11). If a ddNTP molecule is incorporated into a growing DNA chain, the absence of a 3'-OH group prevents formation of a phosphodiester bond with the succeeding dNTP. Further extension of the growing chain is impossible. Thus, when a small amount of one of the ddNTPs is included with the four conventional dNTPs in a reaction mixture for DNA synthesis, there is competition between extension of the chain and infrequent, but base-specific, termination (Figure 12). The products of the reaction are a population of oligonucleotide chains whose lengths are determined by the distance between the 5' terminus of the primer used to initiate DNA synthesis and the sites of chain termination. In a sequencing reaction containing ddA, for example, the termination points correspond to all positions normally occupied by a deoxyadenosyl residue. By using the four different ddNTPs in four separate enzymatic reactions, populations of oligonucleotides are generated that terminate at positions occupied by every A, C, G, or T in the template strand. These populations of oligonucleotides can be separated by electrophoresis, and the locations of each band can be ascertained by autoradiography or the emission of fluorescence. When the four populations are loaded into adjacent lanes of a sequencing gel, the sequence of the newly synthesized strand can be read in a 5'-3' orientation by calling the order of bands from the bottom to the top of the gel.

1.2.11.2 Automated DNA sequencing with Dye-terminator systems

In sequencing with dye-terminators, the fluorescent dyes were attached to dideoxynucleotides that become incorporated at the 3' end of the products of sequencing reaction. Each of the four ddNTPs is labeled with a different dye linked to the nitrogenous base via a linker. Four chain-extension reactions can then be carried out with the same primer in a single tube, sparing considerable labor and cost. By

contrast to dye-primers, dye-terminator chemistry has the additional advantage that it eliminates noise arising from premature chain termination without attendant incorporation of dideoxynucleotides.

In its earlier phase of development, dye-terminator chemistry had major problems resulting from the attachment of bulky dye moieties to the dideoxy terminator molecules. Because substrate affinity was altered, it became necessary to tailor a specific set of dye-labeled terminators for use with each DNA polymerase. Unfortunately, the intensity of the signals generated by these polymerase-substrate pairs was frustratingly uneven, which reduced the accuracy of base calling and limited the range of readable sequence. During the late 1990s, this problem was largely solved by the use of modified enzymes as well as by the introduction of optimized sets of new dyes and linker arms. These improvements provided clearer peak patterns, minimal mobility shifts, and cleaner signals (Figure 13). As a consequence, the quality of the data from current dye-terminator sequencing became comparable to results obtained using dye-primers.

At present, two sets of second-generation dye terminators are available. One incorporates dichlororhodamine dyes (dRhodamine Terminator, PE Biosystems); the other incorporates the BigDyes (BigDye Terminator, PE Biosystems). The BigDyes are used in dye-primer sequencing. They contain a fluorescein isomer as the donor dye and four dichlororhodamine dyes as the acceptors. Both terminators work well with AmpliTaq polymerase FS. Convenient cycle-sequencing conditions are normally used, and the sequencing reactions can be carried out with any primer and with a wide variety of templates (single-stranded DNA, double-stranded DNA, or PCR-generated DNA). In addition, because reactions with AmpliTaq FS use far fewer dye terminators, tedious gel-filtration (spun column) steps to remove unincorporated dyes before gel loading can often be replaced by simple ethanol precipitation. Again, dGTP is replaced with dITP (deoxyinosine triphosphate) to reduce gel compressions.

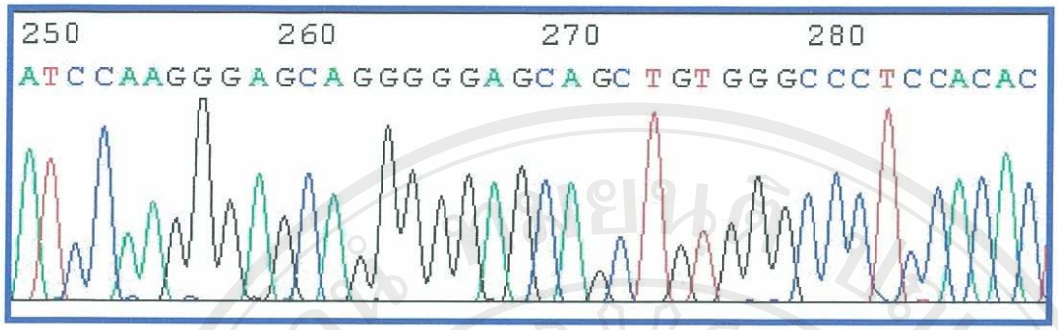

Figure 13. The output from the fluorescent sequencer is usually represented in a chromatogram format

1.3 OBJECTIVES

1. To characterize the molecular mutation of the non-southeast Asian deletion type of α-thalassemia-1 in-patients with Hb H disease by using multiplex polymerase chain reaction.
2. To study the incidence of the non-southeast Asian deletion type of α-thalassemia-1 in-patients with Hb H disease.