CHAPTER 1

INTRODUCTION

1.1 Statement and significance of the problem.

Sanguansermsri (1988) reported a haematological study of a β-thalassemia-major child from northern Thailand who also carried the abnormal haemoglobin, first called haemoglobin Chiang Mai (Hb Chiang Mai). For the further study of the hereditary of β-thalassemia phenotype and the abnormal globin synthesis of this child, his father and mother were used as subjects for clinical experiments. The elevated level of Hb A_2 synthesis in both father and mother, 7.6 and 5.8% respectively, indicated β-thalassemia-trait phenotype. Sitthipreechacharn (1994) reported a further haematologic study of the β-thalassemia-major child's father whose also inherited Hb Chiang Mai synthesis. His pathology was normal clinically and hematologically with the exception of having 24.6% of this abnormal haemoglobin according to a study using High Liquid Performance Chromatography (HPLC). Hb Chiang Mai was clearly identified as a separate band that moved to the anode slower than normal Hb A in cellulose acetate gel electrophoresis at pH 8.5 indicating the change of total negative charge of its globin chains. Finally, the electrophoresis in 6M urea buffer of α-globin chain isolated from Hb Chiang Mai moved to the anode faster than the normal α-

globin chain. This finding showed that the mutation caused a change of total negative charge of Hb Chiang Mai's α-globin chain.

Hb Chiang Mai had not been studied for protein and DNA sequences before, so the type of mutation was not known. Specific amplification by using Polymerase Chain Reaction of α_1- and α_2-globin genes prior to sequencing was the preliminary method applied for the detection of mutation in Hb Chiang Mai. The point mutation, when occurs, may generate a new palindromic sequence that can be recognized by the specific restriction enzyme. Restriction enzyme digestion of the α-globin gene PCR product containing the point mutation is reliable and easy to perform for screening of this haemoglobin variant in a large population.

Human α_1- and α_2-globin genes are non-allelic gene loci located adjacent to each other on chromosome 16. Their coding region was shown to be 95% identical in base sequence composition from 5'-end and the difference was found at the 3'-end of each gene. Amplification of both α-globin genes, however, has been complicated by their unusually high GC-content (average 60% with peaks as high as 70-80%). Such segments are known for their thermal stability and for a tendency to form secondary structures which interfere with the annealing of the primers, and then with the extension reaction during the amplification process.

Molchanova and co-workers (1994) have reported the successful amplification of α_1- and α_2-globin genes in an asymmetric manner, by using the ratio of 1:25 of the forward primer to the reverse primer. The majority of PCR products generated was a single strand DNA fragment which appropriated for sequencing with Sequenase Version 2.0 (United States Biochemical, Cleveland, OH, USA). However, their PCR system was not repeatable in this laboratory even with the optimization of some important parameters in PCR, such as formamide, $MgCl_2$, concentration and the annealing temperature.

In this study, Hb Chiang Mai's α_1- and α_2- globin genes were studied for DNA sequences. Mutations in those genes were detected by using Chain-termination cycle sequencing which could use either single- or double-stranded DNA templates. The preliminary part of the present study was the optimization of the symmetric PCR of α_1- and α_2-globin genes prior to sequencing and the mutations observed from the sequencing data were confirmed by using specific restriction enzyme digestion.

1.2 Literature reviews.

1.2.1. Structure and function of haemoglobin.

Haemoglobin is an intensely colored pigment which imparts the red color to the blood. It is the most abundant blood protein in human and presents more than 95% of the soluble protein content of the erythrocytes. The primary functional role of the haemoglobin is to transport oxygen from alveolar capillaries of the lungs to the body tissues. The associated function is binding of carbondioxide and protons by deoxy-haemoglobin, thereby serving to buffer the blood on the venous side of the circulation (Honig and Adams III, 1986).

Normal human haemoglobin molecules are composed of four subunits, each containing a protein globin chain and an iron-porphyrin heme moiety. The heme moiety of haemoglobin is composed of a porphyrin (tetrapyrrole) ring structure which is combined with iron to form a stable complex (Figure 1). The iron atoms in the heme groups constitue the oxygen-binding sites, and accordingly each molecule of haemoglobin is capable of combining with four molecules of oxygen.

The normal haemoglobin molecule contains two different types of protein globin chains, designated α and β, forming a tetramer molecule (Hb A), $\alpha_2\beta_2$.

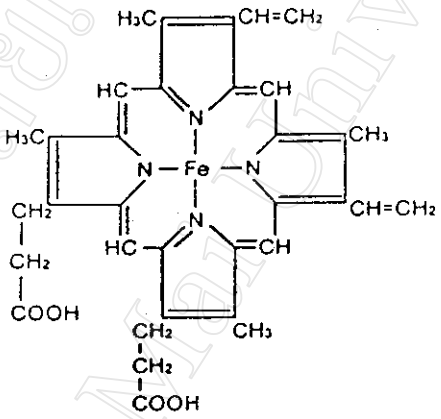

Figure 1. Structure of the heme group (Honig and Adams III, 1986).

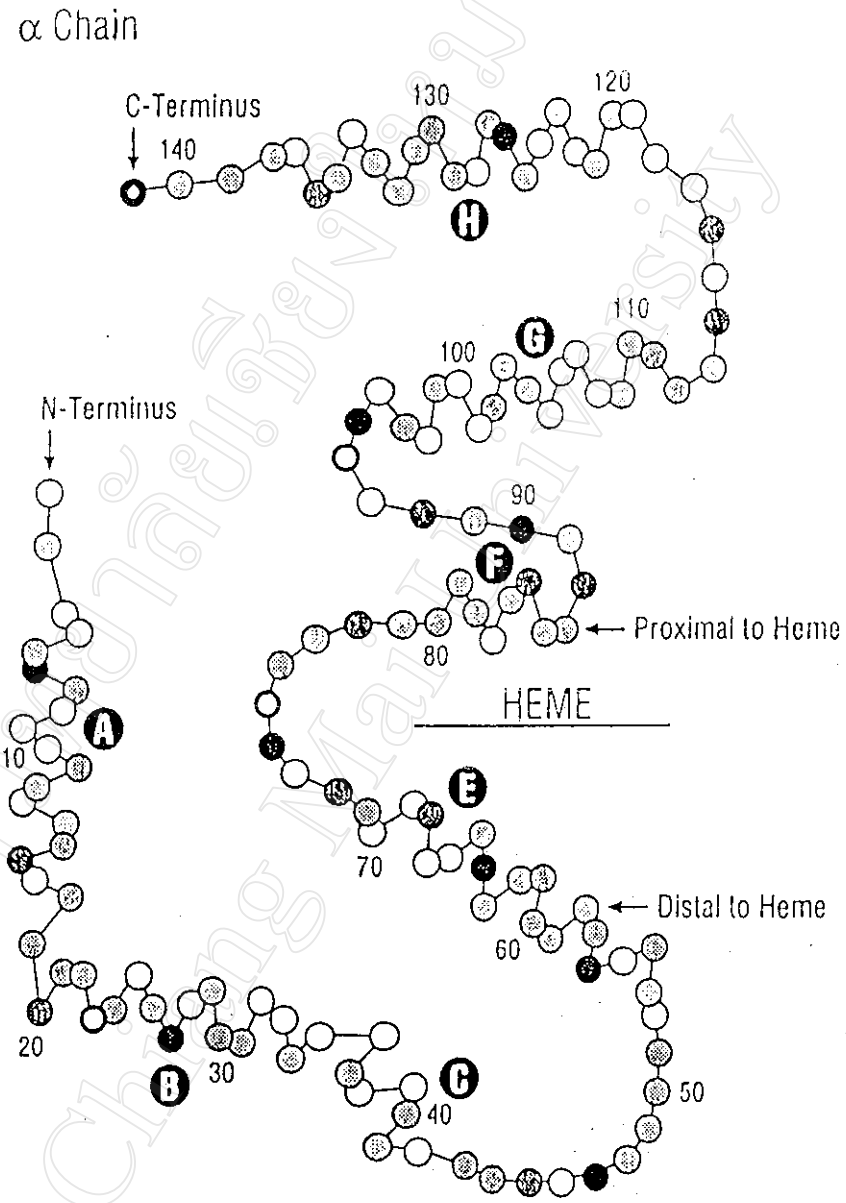

Figure 2. Representation of the secondary structure of the human α-subunit, showing the helical and non-helical regions of the globin chain. The α-chains have a generally similar distribution of helical and non-helical regions as those of the β-chain, but they lack a D-helix (Imai, 1997)

The naturally occurring proteins only portions of the polypeptide structure exist in the α-helical configuration, in spite of the increased stability that the α-helix confers on the protein. The non-helical regions in proteins responsible in the interactions between amino acid groups that result in the stabilization of the secondary structure. Nearly 80% of the polypeptide chains of haemoglobin have been shown to exist in the α-helical configuration, representing a higher percentage than of most other proteins. The more highly oderred structure is believed to contribute significantly to the highly stable molecular configuration of the globin polypeptides. The β-globin chains contain 8 helical segments, designated A-H (Honig and Adams III, 1986). The α-chains have a generally similar distribution of helical and non-helical regions as those of the β-chain, but lack of a D-helix (Figure 2).

1.2.2 Normal human hemoglobins.

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 (adult haemoglobin ; $\alpha_2\beta_2$) normally makes up 94-97% of the haemoglobin in the circulating erythrocytes of adults and is therefore the primary determinant of haemoglobin-related functional properties of the blood. Approximately 5% of Hb A in normal adults is

Table 1. Normal human hemoglobins and their globin subunits (Honig and Adams III, 1986)

Hemoglobins	α-like Subunits	β-like Subunits	Tetramer Composition
Hb A	α	β	$\alpha_2\beta_2$
Hb A_2	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	$\zeta_2\gamma_2$

present as a glycosylated derivative, Hb A1c, in which the amino-terminal valyl residues of the β-chains are in keto-amine linkage with glucose (Bunn *et al.*, 1975).

Hemoglobin A_2 (Hb A_2 ; $\alpha_2\delta_2$) is structurally very similar to Hb A but represents a minor haemoglobin fraction, accounting for only 2 to 3% of the haemoglobin 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.

Hemoglobin F (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 (Singer *et al.*, 1951). From 10 to 15% of Hb F in the blood exist as a derivative (Hb F-1) in which the γ-chain amino terminal amino groups are acetylated (Schroeder *et al.*, 1962).

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 (Maurer *et al.*, 1970) 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 in the deoxy state interacts strongly with 2,3-DPG to produce a significant lowering of the oxygen affinity of the haemoglobin. Hb F, however, exhibits considerably less interaction with 2,3-DPG (Tyuma and Shimizu, 1970), 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. Haemoglobin gestation and development.

The normal pattern of human haemoglobin production during early development involves a success of "switches" in which the synthesis of a set of globin subunits is activated for a specific period during development, and then is "switched off" and replaced by a different set of globin polypeptides. This process begins in the early embryo and is largely completed by about six months of age postnatally (Figure 3).

The earliest recognizable haemoglobin-containing cells in the human embryo appear during the fourth week of gestation, arising from mesenchyme of yolk sac origin. In embryos of less than 6 weeks of gestation, hemoglobins Gower-1 ($\zeta_2\varepsilon_2$) and Gower-2 ($\alpha_2\varepsilon_2$) predominate, and may account for as much as 66% of the total haemoglobin

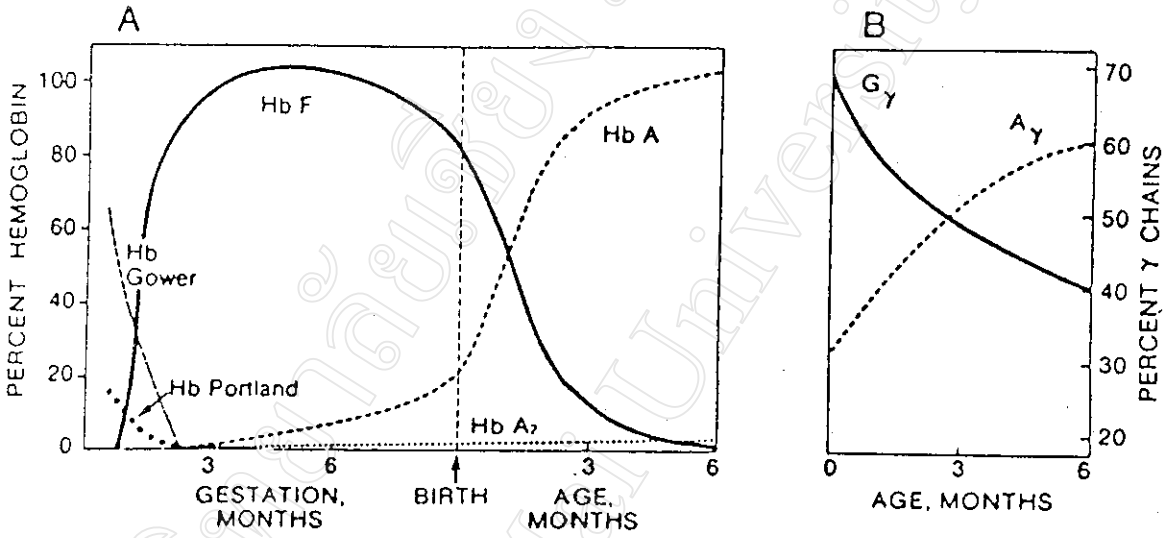

Figure 3. The changes in haemoglobin composition pattern during gestation and development. A) The changes of haemoglobins synthesis of the human embryo, fetus,and infant. And B), the changes in the ratio of the $^{G}\gamma$ and $^{A}\gamma$ subunits of Hb F in the early months of life. (Honig and Adams III, 1986)

(Hecht *et al.*, 1966) with Hb Portland ($\zeta_2\gamma_2$) representing as much as 20% (Pataryas and Stamatoyannopoulos, 1972). Hb F has also been identified in the earliest embryos that have been examined ; the synthesis of Hb F increases rapidly, and by 8 weeks of gestation Hb F accounts for at least 90% of the haemoglobin in the erythrocytes (Huehns *et al.*, 1964) and remains the predominant haemoglobin form through fetal life and in the neonatal period. Hb A is detectable before 8-10 weeks of gestation (Pataryas and Stamatoyannopoulos, 1972 ; Kazazian and Woodhead, 1973) and thereafter accounts for 4-13% of total haemoglobin synthesis. The major switch from Hb F to Hb A takes place in the first few weeks of life, with Hb F falling to less than 3% of the total by 6 months of age and reaching levels of less than 2% by one year (Colombo *et al.*, 1976). Hb A_2 ($\alpha_2\delta_2$) is present in the blood only in trace amount at birth (Horton *et al.*, 1962) and increases to 2-3% of the total haemoglobin by one year of age.

The drop off in the level of Hb F that takes place in the early months of life is also accompanied by a change in the ratio of $^G\gamma$ and $^A\gamma$ subunits which make up the total Hb F. The usual ratio of $^G\gamma$: $^A\gamma$ at birth is about 7:3, and this ratio characteristically reverses to about 2:3 by 6 months of age. where it remains throughout adult life (Schroeder *et al.*, 1971). In appoximately one-third of individuals, however, the

typical newborn $^{G}\gamma$: $^{A}\gamma$ ratio appears to persist in the adult (Huisman *et al.*, 1977).

Haemoglobin pattern changed during human development carries a number of practical implications with regard to the expression of abnormalities of the different globin chains : Genetic abnormalities of the γ-chains will be expressed throughout fetal development when Hb F is produced, but will have little or no pathologic potential beyond the first months of life when the switch to Hb A production occurs. Because α-chain synthesis also begins early in gestation, genetic abnormalities involving the α-chain will be fully expressed during fetal life and at birth, as well as throughout adult life. Abnormalities of the β-chain, on the other hand, are not fully expressed until the Hb F to Hb A switch has been completed 3 to 4 months after birth. With the application of sensitive analytical techniques, however, abnormal β-chains can often be detected at birth and even in the first trimester fetus.

1.2.4 The localization and organization of globin genes.

Hb A ($\alpha_2\beta_2$) was normally makes up 94-97% of the haemoglobin in the circulating erythrocytes of adults. The α-and β-globin genes were inherited independently and thereby occupied the sites on separate chromosomes (Smith and Torbert, 1958). The further experiments of the chromosomal localization of globin genes have shown that the

human β-globin gene is on chromosome 11 and α-globin gene is on chromosome 16. These findings, taken together with the previously established linkage data, demonstrate convincingly that the γ-, δ-, and β-globin gene loci are localized to chromosome 11 and the duplicated α-globin gene loci to chromosome 16 (Deisserooth *et al.*, 1976, 1977 and 1978).

1.2.4.1. The β-globin gene cluster.

Physical linkage maps of the γ-, δ-, and β-genes have been reported by several groups from data developed from blotting hybridization analyses of restriction enzyme fragments of genomic DNA (Flavell *et al.*, 1978 ; Mears *et al.*, 1978 ; Little *et al.*, 1979 ; Bernards *et al.*, 1979 ; Fritsch *et al.*, 1979). These gene maps shows the globin genes to be arranged in the order 5'...ε-^{G}γ-^{A}γ-δ-β...3', with the genes being separated by the approximately distance shown in Figure 4 (Fritsch *et al.*, 1980).

1.2.4.2. The α-globin gene cluster.

The two non-allelic α-globin genes have been mapped by genomic blotting, and were shown to occupy position *ca.* 3.7 kb apart (Orkin, 1978 ; Surrey *et al.*, 1978). This finding was confirmed by blotting hybridization analyses of cloned human DNA fragments (Lauer *et al.*,

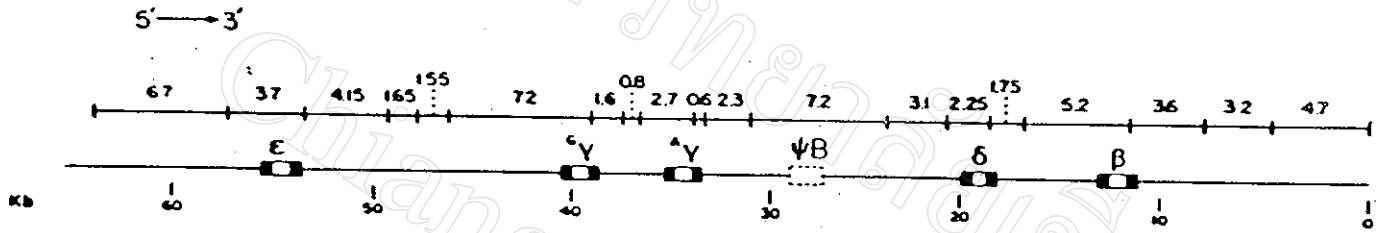

Figure 4. Linkage map of the human β-globin gene cluster. The location of the ε-, $G\gamma$-, $A\gamma$-, δ-, and β-globin genes and a β-pseudogene (ψβ) are indicated (Honig and Adams III, 1986).

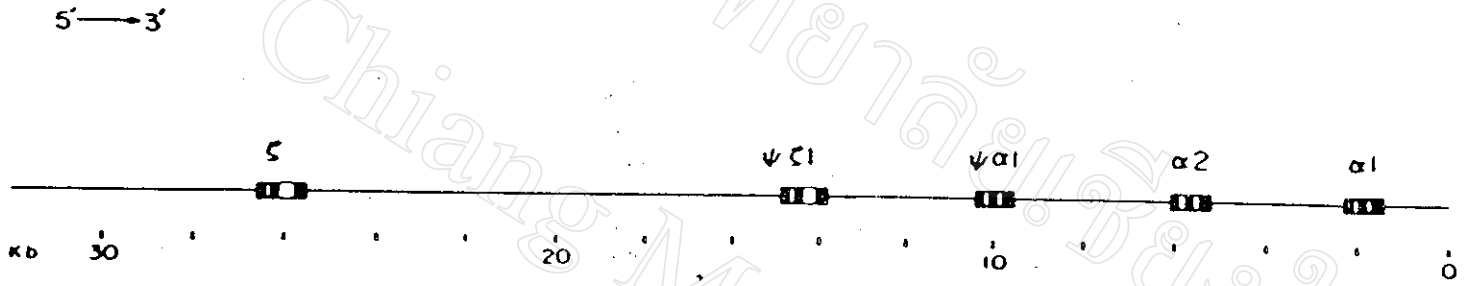

Figure 5. Linkage map of the human α-globin gene cluster. The location of the two adult (α_1 and α_2) and the embryonic (ζ) genes are shown, as well as two pseudogenes ($\psi\alpha_1$ and $\psi\zeta_1$) (Honig and Adams III, 1986)

1980) ; the latter study also showed that three other α-like sequences are closely linked to the α-globin gene pair. One of these, which has been shown to be a non-functional α-pseudogene, is located approximately 4 kb, 5' to the α-globin genes. The other two were determined by partial sequence analysis to correspond to the embryonic ζ-globin (Lauer *et al.*, 1980) (Figure 5). When the structures of the ζ-globin genes were examined in greater detail, however, one of these genes (the 3' gene) was found to have a translation termination codon at the position corresponding to amino acid residue 6 in the ζ-globin polypeptide (Proudfoot *et al.*, 1982). This gene could not, therefore, have normal expression, and is believed to be a pseudogene of recent evolutionary origin. The conclusion that the 5'-gene is the functional ζ-gene is also supported by other evidence : Pressley and co-workers (Pressley *et al*, 1980) described a Greek infant with severe α-thalassemia in whom they were able to demonstrate by gene mapping that the α-globin genes as well as the 3' ζ-gene were detectable. Cord blood from the infant contained ζ-chains in the form of Hb Portland ($\zeta_2\gamma_2$), which must have represented activity of the remaining ζ-globin gene. Although the α-gene cluster is less complex than that of the β-gene group, its organization follows the same general pattern. All of the genes in the α-gene family are transcribed from the same DNA strand (Lauer *et al.*,

1980) and the 5' to 3' orientation of the genes follows the order of their expression during development (i.e. ζ-->α).

1.2.5. The α_2- and α_1-globin genes family.

α_2- and α_1-globin genes sequences have been reported (Liebhaber *et al.*, 1980 for α_2-globin gene and Michelson and Orkin, 1980 and 1983 ; Liebhaber *et al.*, 1981 for α_1-globin gene sequences). These 2 linked genes have identical polypeptide products (Földi *et al.*, 1980) and the nucleotide sequences of their coding regions have also been shown to be identical. The sequences of their intron 1 has also been shown to be highly conserved, differing by only a single base, and it appears likely that even this difference may represent a polymorphic site rather than a point of actual non-homology. Intron 2, however, has shown a significant degree of dissimilarity between the α_2- and α_1-genes, with the latter having a longer and somewhat different sequence. The greatest divergence between the two α-globin genes was found to be in the 3' untranslated portion (Michelson and Orkin, 1980 and 1983) where at least 19 sequence differences could be identified.

The flanking regions of α_2- and α_1-globin genes have also been examined in considerable detail (Proudfoot and Maniatis, 1980 ; Michelson and Orkin, 1983) and revealed the presence of two extensive

regions of sequence homology, both of which are located 5' to the coding portions of the genes.

The existence of paired α-globin genes in human and other species is believed to be resulted from a duplication of an ancestral α-globin gene during an early stage of vertibrate evolution, probably more than 300 million years ago (Zimmer *et al.*, 1980). Independent evolution of the two linked genes over so long a period of time would be anticipated to produce considerable divergence of their sequences, but the very high degree of similarity between the two α-globin genes, as well as between their flanking regions, is at variance with this expectation. To explain this apparent inconsistency, Zimmer and co-workers (1980) have proposed a process of "concerted evolution" of this pair of linked genes. According to this mechanism, the α-globin genes may be shunted from one chromosome to another, by a process of gene conversion or unequal crossing over.

By repetitive cycles of gene duplication and loss in this manner, the 3' and 5' α-globin genes could effectively exchange places, and in doing so would "erase" any structural differences between the two that might have arisen from divergent evolution. This mechanism would be expected to require a significant degree of sequence homology between the α_2- and α_1-globin genes, to serve as sites complementary for the misalignment during synapsis. The extensive similarity of the sequences

in the coding regions as well as in the introns and the 5' flanking areas of the two genes would amply appear to meet this requirement.

Considerable sequence homology has also been shown to exist between the two embryonic ζ-globin genes (Proudfoot *et al.* 1982). Chromosomes bearing single or triplicated forms of this gene have been identified (Winichagoon *et al.*, 1982) suggesting that this gene pair may also have undergone concerted evolution by a process of interchromosomal exchanges of genetic information.

The complete sequence of the α-globin pseudogene has been determined, and has shown approximately 70% homology with the α-globin genes (Proudfoot and Maniatis, 1980). A replacement of its translation initiation codon as well as frameshift deletions in this gene account for its failure to be expressed with a polypeptide product.

1.2.6. Different rates of mRNA translation balance the expression of the two human α-globin gene loci.

The region of the two α-globin genes which encoded mature α-globin mRNA (exon 1, 2, and 3) are identical from the 5' capping site through the UAA termination codon (Liebhaber *et al.*, 1981 ; Michelson and Orkin, 1980) but considerable differences have been shown to exist between the 3' untranslated portions of the two genes. By taking advantage of these sequences differences, through the application of

specific probes or by other methods capable of distinguishing between the products of the two genes, it has become possible to determine the relative expression of the α_2- and α_1-globin genes in human erythroid cells (Orkin and Goff, 1981 ; Liebhaber and Kan, 1981).

In the study of Orkin and Goff (1981), reticulocyte mRNA was allowed to hybridize with a ^{32}P-labeled DNA probe that was fully complementary to the α_1-gene, followed by treatment with S1 nuclease to digest the unhybridized portions of the probe. Under these conditions the labeled probe fragments which were recovered from the α_1-mRNA hybrids were larger than those from the α_2-mRNA, because of the great similarity between the probe and the α_1-mRNA sequence. The two different sized fragments could then be separated and their radioactivity were determined. In normal human reticulocytes studied in this manner the α_2-mRNA was consistently found to be present in higher concentration than the α_1-mRNA, with a ratio of approximately 60:40.

Liebhaber and Kan (1981) examined the expression of the two α-globin genes by a different experimental approach : For these determinations they used mRNA isolated from reticulocytes or from bone marrow cells as a template for reverse transcriptase, from which complementary DNA (cDNA) copies were synthesized. The cDNA products were digested with the restriction endonuclease Hae III, which

recognizes unique sequences in the 3'-untranslated portion of the α_2-gene that are not present in the α_1-gene. The resulting α_2- and α_1-cDNA digestion products were therefore unequal in size, and could be readily separated and their relative quantities determined. This study, in aggreement with the findings of Orkin and Goff, also showed the α_2-mRNA to be present in relative excess, with the α_1-mRNA accounting only for 26% of the total of the two.

Additional insights about the relative expression of the α-globin genes were gained from a subsequent study by Liebhaber and Kan (1982) which involved a unique patient who had both α-thalassemia and a point-mutation α-globin gene abnormality. The patient had two α-globin genes, a normal α_1-gene, and an α_2-gene which contained the point mutation. When reticulocyte mRNA from the patient was analyzed by the cDNA method described above, an α_2/α_1 mRNA ratio of 74/26 was again obtained. However, when the patient's mRNA was translated in a cell-free globin-synthesizing system, the two different globin chains were surprisingly found to be synthesized at equal rates. It was therefore concluded that the α_1-mRNA must be translated with a three fold higher efficiency than the α_2-mRNA, to compensate for its correspondingly lower concentration in the cell. The mechanism for the apparently dissimilar rates of translation of the α_2- and α_1-mRNA's is

entirely unknown ; it was suggested that the nucleotide sequences of the 3' untranslated portions of the two α-globin genes may somehow be responsible for the observed differences in their transcription efficiency. Liebhaber and Kan have further speculated that the divergent 3' segments of the α_2- and α_1-genes may have evolved so as to maintain equal expression of the two genes, by counterbalancing the different translational efficiencies of their respective mRNA's.

1.2.7. Classification of the globin gene mutations.

Abnormalities resulting from the globin gene mutations are traditionally classified into three major categories, including the haemoglobin structural variants, the thalassemias, and the syndromes of hereditary persistence of fetal haemoglobin synthesis (HPFH). Most mutations can be readily assigned to one of these individual groups, although some are associated with phenotypic features which apply to more than one category (Honig and Adams III, 1986).

1.2.8. Mutation associated with globin structural abnormalities.

The haemoglobin structural variants have been further classified into 8 categories (Honig and Adams III, 1986) such as single point mutation, double point mutation which occured within the single globin

gene, frameshift mutation causing premature translation termination, etc.

1.2.8.1. Single point mutations which produce amino acid substitutions.

Boyer and co-workers (1978) examined the distribution of amino acid replacements in the human β-chain, including data from the known β-chain variants as well as from substitutions that were determined to occur in mammalian evolution ; they concluded that these mutations are generally uniformly distributed over the length of the gene. These findings (Figure 6) suggested that this group of mutations arose by a random process, and they further implied that there were no hypermutable "hot spots" that favor unusual mutation rates, at least within the coding regions of the gene.

On the other hand, it is also apparent from an inspection of the lists of known mutations that the amino acid substitutions are by no mean uniformly distributed among the different globin chains. Approximately 30% of the mutations reported are in the α-chains, with *ca.* 60% involving the β-chains, and only 10% affecting the δ- and γ-chains. This uneven distribution is presumably not related to variations in the mutation rates of the various globin chains, but rather appears to be explainable by a group of largely unrelated factors.

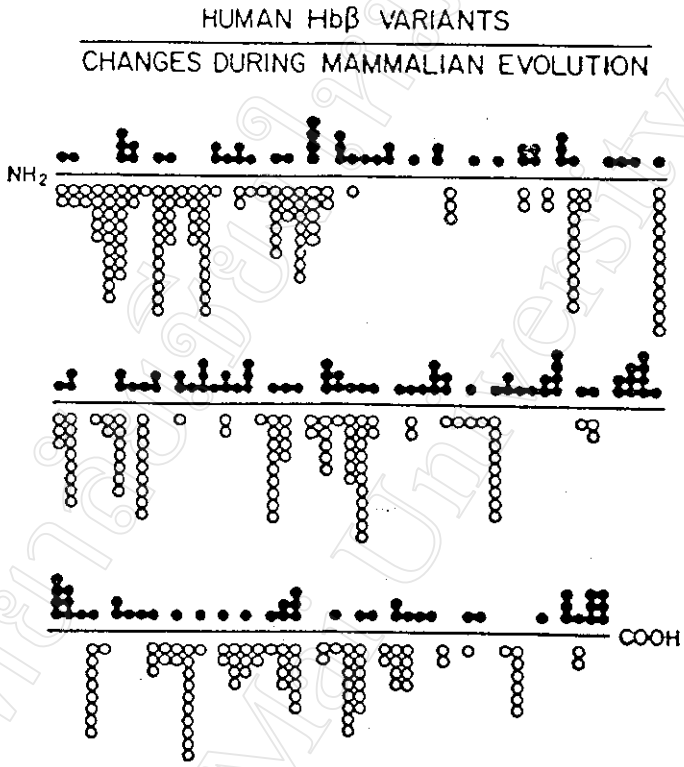

Figure 6. The distribution of nucleotide base-change mutations of the human β-chain. The β-chain sequence is represented by the solid line segments, with its amino acid carboxyl termini as shown. Each of the solid circles represents the site of a known β-chain mutation, with the open circles indicating the locations of substitutions which occured during mammalian evolution. (Honig and Adams III, 1986)

Firstly, the globin mutations which produce abnormalities of haemoglobin function and those which give rise to haematologic disease are disproportionately represented among the β-chain variants, and this difference alone accounts for a large part of the relative excess of known β-chain variants. One explanation for the apparent concentration of these mutants in the β-chain is that only a single pair of β-globin gene alleles are normally expressed, as contrast to the α- and γ-globin genes which normally have two pairs of alleles ; because of this difference dysfunction β-chains will generally be produced in relatively higher concentration, and will consequently have a greater likelihood of causing significant symptoms or other clinical findings that will facilitate their identification. Comparable abnormalities in the α- or γ-chains, on the other hand, will usually be diluted by the 3-fold excess of the normal gene products, so that their associated haematologic disorders may be sufficiently mild to escape detections. An additional factor is that the β-globin gene is not expressed to any significant extent in fetal life or in the perinatal period, with the fetus and newborn thus being protected from deleterious effects of mutations affecting the β-chains. It is concievable that the more severe examples of these mutants might be incompatible with survival of the fetus in utero if they were fully expressed, and in accordance with this notation relatively few of

the known α- and γ-chain structural mutants produce significant clinical disease (Honig and Adams III, 1986).

1.2.8.2. Double point mutations which produce the amino acid substitutions.

Several of the haemoglobin structural variants are characterized by the presence of two different amino acid substitutions within a single globin chain. In most of these mutants the amino acid replacements are located at widely separated sites in the globin chain, for example Hb J-Singapore ; $\alpha^{78Asn-->Asp}$ and $\alpha^{79Ala-->Gly}$ (Honig and Adams III, 1986). An additional intriguing feature for these variants is that virtually all of their individual point mutations have also been observed in single-substitution globin variants, which, moreover, often co-exist in the same population.

A unique example has also been reported of an identical point mutation that was present in both the α_2- and α_1-globin genes of a single chromosome (Liebhaber *et al.*, 1984). This abnormality was identified in an individual with Hb I ($\alpha^{16Lys-->Glu}$) whose percentage of the variant haemoglobin was found to be unexpectedly high. This type of abnormality might well have arisen by a process of gene conversion, involving the α^A gene of α^A, α^I linked pair.

1.2.9. Structural haemoglobin variants with α-thalassemia-like expression.

Mutation involving single nucleotide substitutions in the globin genes have been estimated to occur at a rate of 10^{-9}-10^{-8} nucleotide substitutions per nucleotide site per year (Kimura, 1978 ; Boyer *et al.*, 1978) A notable feature of the base-change mutations is that certain base substitutions occur with greater frequency than others, both in the globin genes and in those of other protein systems. Bunn and co-workers (1972) tabulated the base substitutions of 115 haemoglobin mutants and showed that the number of purine-purine transitions (involving adenine-guanine substitutions) was substantially larger than would be expected from random replacement. Similarly, a comparison of base substitution frequencies among a large group of vertibrate globins showed that G-->A and A-->G transitions were represented at a considerably greater frequency than could be predicted by chance (Zuckerkandl *et al.*, 1971). The latter study also demonstrated that transversions (purine-purine transitions) occured at close to expected frequencies, while pyrimidine-pyrimidine transitions were far less frequent than would be anticipated from random substitution.

A number of examples have been described of α-chain structural variant alleles which are localized to chromosomes that also contain an α-thalassemia determinant. In each of these that have been examined by

gene mapping an α-thalassemia-2 deletion has been identified indicating that the mutant allele represents the sole α-globin gene in these abnormal chromosomes. As a result of this linkage arrangement, individuals expressing these abnormal haemglobins also exhibit the α-thalassemia-2 phenotype. The mutant gene for one of the haemoglobin variants in this group, Hb Mahidol (Q), is localized to a chromosome having a leftward α-globin gene deletion (Pagnier *et al.*, 1982). Others of these variants, including Hb Evanston (Honig *et al.*, 1984), Hb G Philadelphia (Surrey *et al.*, 1980 ; Sancar *et al.*, 1980 ; Felice *et al.*, 1982), Hb J Tongariki (Bowden *et al.*, 1982) and Hb Hasharon (Giglioni *et al.*, 1980) have the rightward form of deletion. Two types of mechanisms could presumably have led to the formation of these doubly affected chromosome, and from what is presently known about these haemoglobin variants either appears to be equally plausible. On the other hand, these linked abnormalities could have arisen by point mutations in already existing α-thalassemia-2 chromosomes. The alternative mechanism would involve genetic recombination between a chromosome carrying a mutant allele, with either a normal or a gene-deletion chromosome.

A second group of variant hemoglobins with α-thalassemia-like expression includes Hb Constant Spring and the other extended-chain hemoglobins which result from mutations that involve the normal

translation-termination codon (Honig and Adams III, 1986). These α-chain variants are synthesized at a very low rate (Weatherall *et al.*, 1965) which is the apparent cause for the thalassemia-like changes that they produce. The reason for their low rate of synthesis, however, is not entirely clear, and it appears likely that more than one underlying abnormality may be contributing factors. Some evidences indicate that the synthesis of the Constant Spring α-chain may be prematually terminated during the process of erythroid cell maturation (Derry *et al.*, 1984).

A third group of globin structural variants that produce the α-thalassemia phenotype includes several highly unstable α-chain variants, all of which have been shown to undergo very rapid post-synthetic proteolytic degradation, for example Hb Suandok (Sanguansermsri *et al.*, 1979). The thalassemia-like expression of these variants contrasts with that of the unstable haemoglobin groups described earlier, of which the more severe forms produce Heinz-body hemolytic anemia.

In individuals who are heterozygous for the unstable variants as indicated, the abnormal haemoglobin is usually either undetectable or is present in very small amounts, usually no more than 2% of the total haemoglobin. Individuals with Hb Suandok or Hb Petah Tikva who also inherit a two α-gene deletion chromosome, exhibit the Hb H disease phenotype. These compound heterozygotes also characteristically have a

considerably larger percentage of the variant haemoglobin in their erythrocytes than is observed in the usual heterozygote. It appears likely that the presence of a substantial excess of uncombined β-subunits in the erythrocytes of these individuals may promote the incorporation of the unstable α-chains into more stable subunit tetramer molecules, thereby accounting for their larger quantity of the haemoglobin variant.

1.2.10. Hb Q-Thailand (Hb Mahidol) and Hb Westmead, the abnormal hemoglobins with α-chain variant.

Hb Westmead ($\alpha_1{}^{122\ His\text{-->}Gln}\ \beta_2$), as firstly reported by Fleming and co-workers (1980), is one of the most common haemoglobin variants in Guangxi, a province in Southern China. Hb Westmead is slightly unstable and the child who carries this variant is sufferred from mild anemia (Jiang *et al.*, 1991). As reported by Jiang and co-workers (1991), Hb Westmead cannot be separated from Hb A by electrophoresis or isoelectrofocusing (IEF). However, by using the selective amplification of α_2-globin gene of a carrier and then analyzing with the restriction enzyme *Stu* I, the results showed that the CAC-->CAG mutation was located at codon 122 of the α_2-globin gene.

Hemoglobin Q (Hb Q) was first described by Vella and co-workers (1958) in a Chinese man and further cases, which have occured in people of Chinese or South East Asian origin, have been reviewed by

Sagnet and co-workers (1968). It is an α-chain abnormal haemoglobin, and its association with α-thalassemia is one of the few examples of interaction between an abnormal haemoglobin and α-thalassemia.

On paper electrophoresis at pH 9.0 (Cradock-Watson *et al.*, 1959), Hb Q migrates between Hb A and Hb G and on chromatography on ion-exhange resin at pH 6.0. According to Huisman and Prins (1957), is widely separated from Hb A. These characteristics of Hb Q suggested that it might have a mutation of an acidic residue to histidine by analogy with other such haemoglobin, *e.g.* Hb Hasharon $\alpha^{47Asp\text{-->}His}$ (Halbrecht *et al.*, 1967), and Hb J Iran $\beta^{77His\text{-->}Asp}$ (Rahbar *et al.*, 1967). At alkaline pH values the histidyl residue is neutral, but at acidic pH values it is positively charged so that the charge difference, and consequently the separation, is enhanced.

Residues α74 and α75 occupy positions EF3 and EF4 in the helical notation. This is at the surface of the globin protein molecule in a non-helical region which is not involved in any interchain contacts and the substitutions would not be expected to cause any serious disturbance of the molecular structure. The two variants are unusual rather than abnormal hemoglobins, and the fact that in spite of this they form considerably less than half of the adult haemoglobin in heterozygotes, would agree with the suggestion of Lehmann and Carrell (1968) that the gene for the human α-chain is duplicated.

Lorkin and co-workers (1970) have examined six more samples of Hb Q from three different sources. Three of the samples came from three unrelated individuals found in a survey in Iran, none of them had any known connections with South East Asia. Two samples came from Thailand, from a mother and her child, the latter having Hb Q-H disease, and the sixth from a Chinese person living in California. They reported the structural of Hb Q differs from Hb A by a substitution of aspartic acid by histidine either at position α74(EF3) or at position α75(EF4).

Lie-Injo and co-workers (1979) reported studies of Hb Q in the two patients from Kuala Lumpur and Hong Kong which carried Hb-$\alpha_2^{74\ Asp\text{-->}His}\ \beta_2$. They were identical to Hb Q Thai and Hb Mahidol found in Thais and Hb G Taichung found in Chinese from Taiwan. Although Hb Q in the trait carrier from Hong Kong has not been fully identified, it is most probable that it is Hb-$\alpha_2^{74Asp\text{-->}His}\ \beta_2$ because aspartic acid at position 74 or 75 is replaced by histidine in the a-chain tryptic peptide 9 and is from a Chinese (Hb-$\alpha_2^{75Asp\text{-->}His}\ \beta_2$ is found in Iran). Hb Q (or G) of the carrier from Kuala Lumpur is structurally different ; it has a mutation in α-chain where at position number 30, glutamic acid is replaced by glutamine.

Pootrakul and co-workers (1972) reported a study of the four unrelated families with Hb Mahidol ($\alpha_2^{74Asp-->His}$ β_2) which were identical to Hb G-Taichung and one type of Hb Q. The abnormal haemoglobin gene could be traced to Chinese ancestors in all. Eight were found to be heterozygous for Hb Mahidol ; they were normal clinically and hematologically with the exception of having 17.5-21.7% (average 20.2) of Hb Mahidol. Another three were found to have Hb Mahidol/α-thalassemia disease which showed the clinical and hematological behavior similar to Hb H disease. But in Hb Mahidol/α-thalassemic disease, in addition to Hb H disease, a large amount of Hb Mahidol (77-96%) was present with a striking absence of Hb A. The findings are perfectly compatible with the one α-chain locus model, one of the two allelic positions being occupied by the Hb Mahidol gene and another by the severe α-thalassemia gene resulting in the absence of Hb A. The two linked α-chain loci model requires an assumption that there is a linkage between the Hb Mahidol gene and an α-thalassemia-1. This has yet to be proved or disproved. The presence of 1.22% of Hb Bart's at birth in one of their Hb Mahidol heterozygotes argues equally well for both models. The failure to observe segregation of the relative amounts of Hb Mahidol in Hb Mahidol trait did not appear to be compatible with the two independent α-chain loci model. However, the data were limited.

Sitthiprechacharn (1994) reported a study of one sample from the Northern Thailand which carried the abnormal haemoglobin which was first called Hb Chiang Mai. According to the cellulose acetate gel electrophoresis at pH 8.5, in which the hemoglobins will be separated according to their different total globin's negative charge, Hb Chiang Mai was clearly indentified by the slower moving to the anode comparing with normal Hb A (Figure 7). By using DEAE-Sephadex A-50 column chromatography to separate and quantitate the normal and the abnormal haemoglobin fractions by applying a shallow salt gradient to the column, the elution rates are related to the isoelectric points of the different hemoglobins. Hb Chiang Mai was released from the column at pH 8.0 comparing with Hb A_2 at pH 8.2, Hb A at pH 7.9, and Hb F at pH 7.2. and was founded approximately 27.17% while Hb A was decreased to 52.08% (Figure 8). In addition to the use of High Liquid Performance Chromatography (HPLC), Hb Chiang Mai was also clearly isolated from Hb A and founded as 24.6% of the total haemoglobin (Figure 9). Finally, from the globin electrophoresis in 6M urea buffer, the procedure based on the separation of globin chains on a weak cation exchange, carboxymethyl-cellulose using phosphate buffers of low ionic strength under denaturing condition, α-globin chain isolated from Hb Chiang Mai moved faster to the anode comparing with α-globin chain from Hb A (Figure 10). This showed the mutation causing the changes

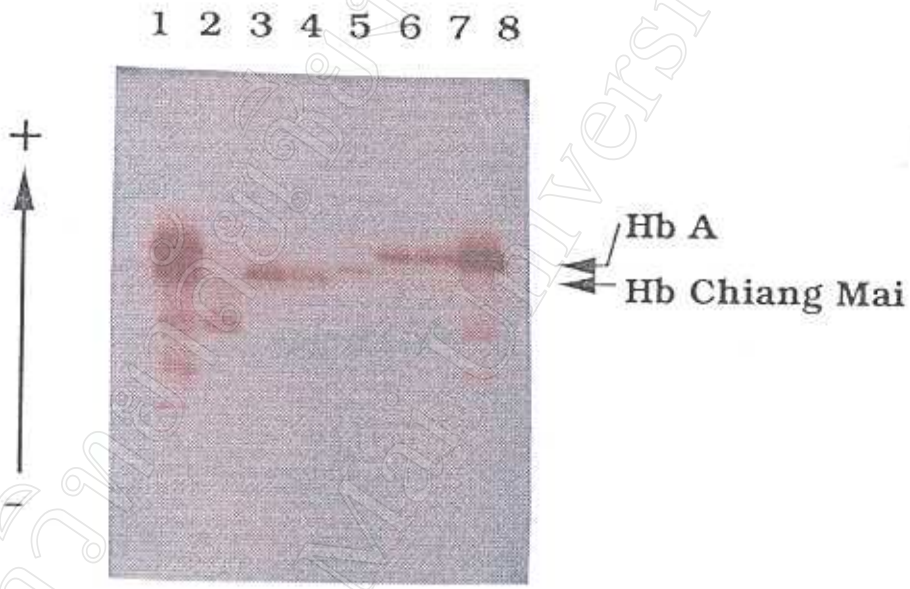

Figure 7. Cellulose acetate gel electrophoresis of Hb Chiang Mai. Lane 1 : crude hemolysate containing Hb Chiang Mai, lane 2 : Hb A_2/E, lane 3 and 4 : purified Hb Chiang Mai, lane 5 : Hb F, lane 6 and 7 : Hb A and lane 8 : crude-normal haemolysate (Sitthipreechacharn, 1994).

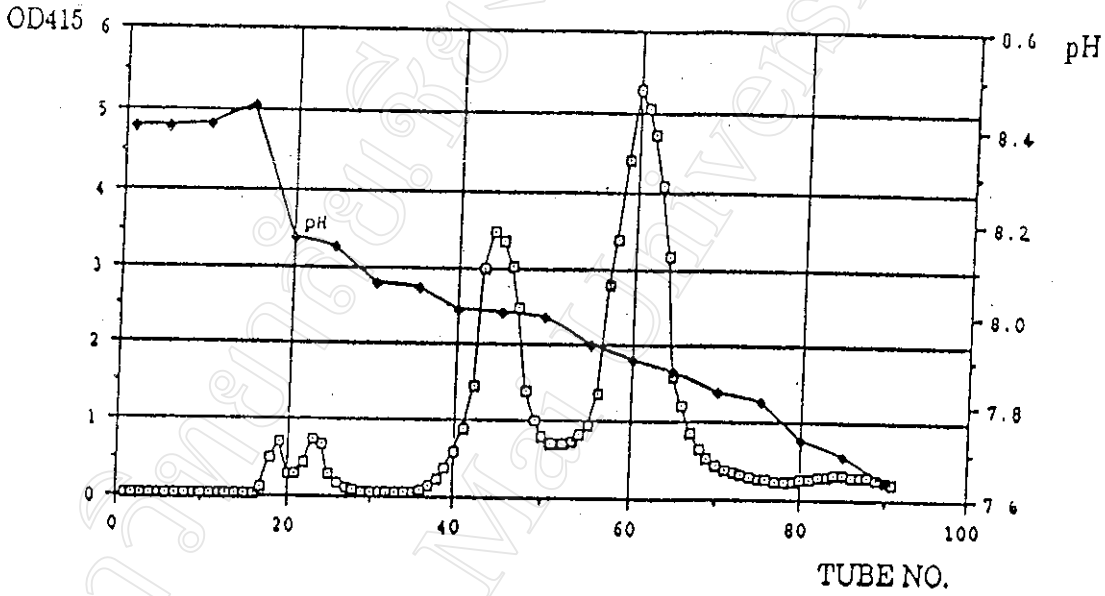

Figure 8. DEAE-Sephadex A-50 column chromatography of Hb Chiang Mai. It was released from the column at pH 8.0 comparing with Hb A_2 at pH 8.2, Hb A at pH 7.9, and Hb F at pH 7.2. and was founded approximately 27.17% while Hb A was decreased to 52.08%. (Sithipreechacharn, 1994).

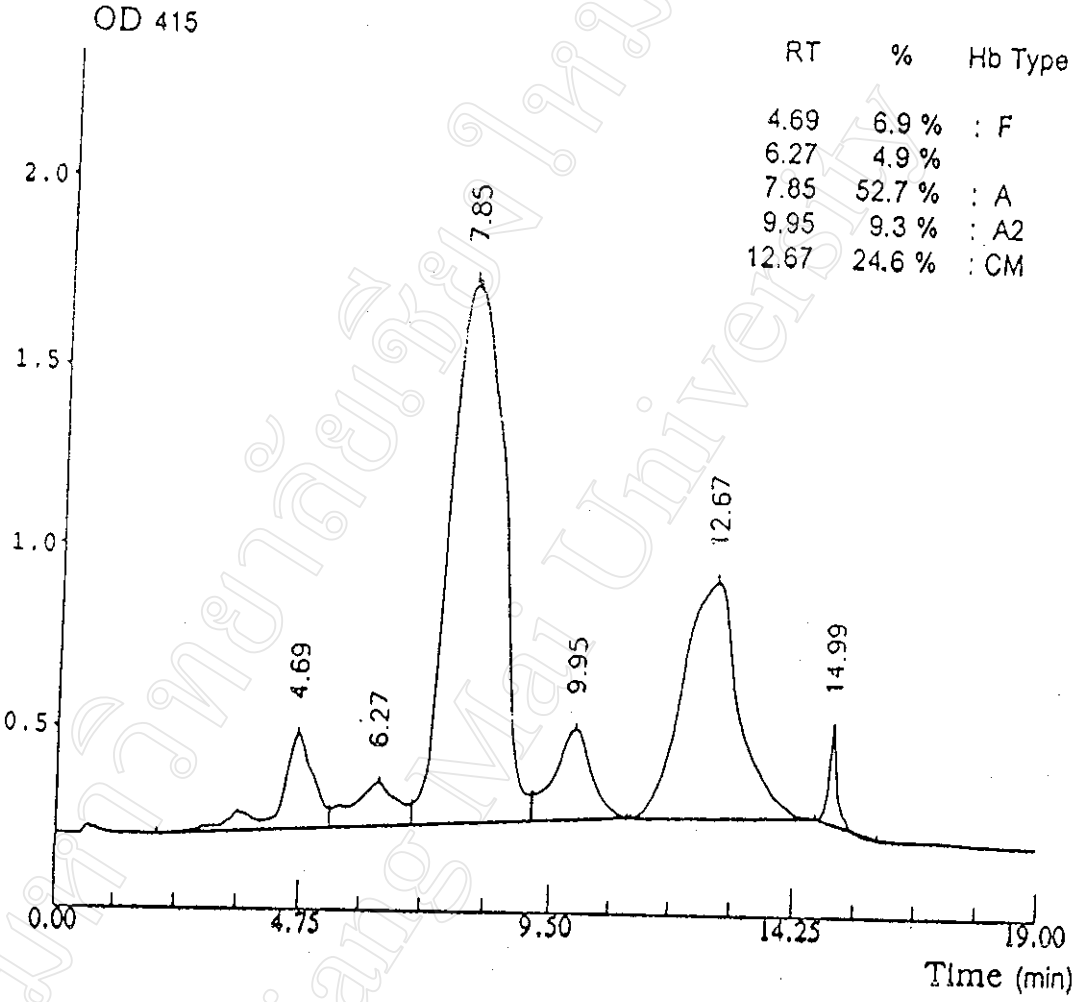

Figure 9. The separation and quantitation of Hb Chiang Mai by using High Performance Liquid Chromatography (HPLC) technique. Hb Chiang Mai was found as 24.6% while Hb A was 52.7%. (Sithipreechacharn, 1994).

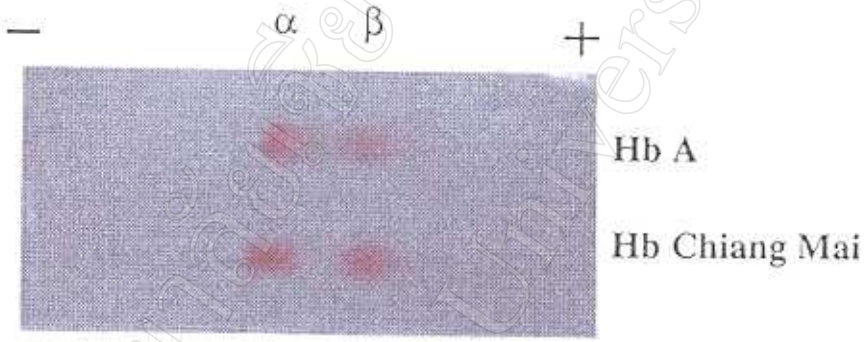

Figure 10. Hb Chiang Mai's α-globin electrophoresis in 6M urea buffer. (Sithipreechacharn, 1994).

the changes of total negative charge of Hb Chiang Mai occured in the α-globin chain. However, her studies did not indicated this mutation in the DNA level.

1.2.11. The Polymerase Chain Reaction.

Polymerase chain reaction (PCR) is a method inovated by Kary B. Mullis in 1983. It is the *in vitro* enzymatic amplification reaction of DNA or RNA segments normally less than 5 kb. The essential steps in each cycle are thermal denaturation of double-stranded target molecules, primer annealing to both DNA strands and enzymatic synthesis of DNA (Vosberg, 1989). A pair of primers complementary to both strands of a DNA molecule and flanking a target region of interest is used to direct DNA synthesis in repeated cycles in opposite and overlapping directions. In each cycle, both strands are templates for the generation of two new duplex molecules. The end results is an exponential increase in the total number of DNA fragments that include the DNA sequences between the PCR primers, which are finally represented at a theoretical abundance of 2^n, where n is the number of PCR cycle performed (Gibbs, 1990).

The PCR is efficient, specific and very sensitive. The exponential increase in the number of the product molecules is limited (Vosberg, 1989). Regarding efficiency, the theoretical upper limit of the number of product molecules is 2^n. That means every target sequence present at

the begining could, in 20 cycles, give rise to about a million progeny molecules. Under normal experimental conditions, this value is not obtained, however. A more realistic average efficiency of 85% per cycle (Saiki *et al*, 1985) reduces the overall yield from the theoretical upper limit of the reaction.

An important feature of this scheme is that the majority of the amplification products that are present following many PCR cycles are double-stranded DNA fragments of discrete length. The strands that are synthesized as copies of the original template are bounded at the 5' terminus by the oligonucleotide primer while the 3' terminus is determined by the position at which the DNA polymerase finishes its synthesis. In contrast, the products of polymerase extension resulted from priming of DNA strands that were produced during the PCR will have both their 5' and 3' termini defined by the position of the oligonucleotide primers as a consequences of "run off" synthesis. The overall consequence is that the fragments without discrete length are relatively rare compared to those with defined termini (Gibbs, 1990).

The general course of the reaction with DNA as the initial template is outlined in Figure 11. Each cycle is initiated by melting double-stranded DNA at 91-95°C (usually for 1 min) to obtain single-stranded templates. This step is followed first by annealing of the primer oligonucleotides, which are added in large molar excess over

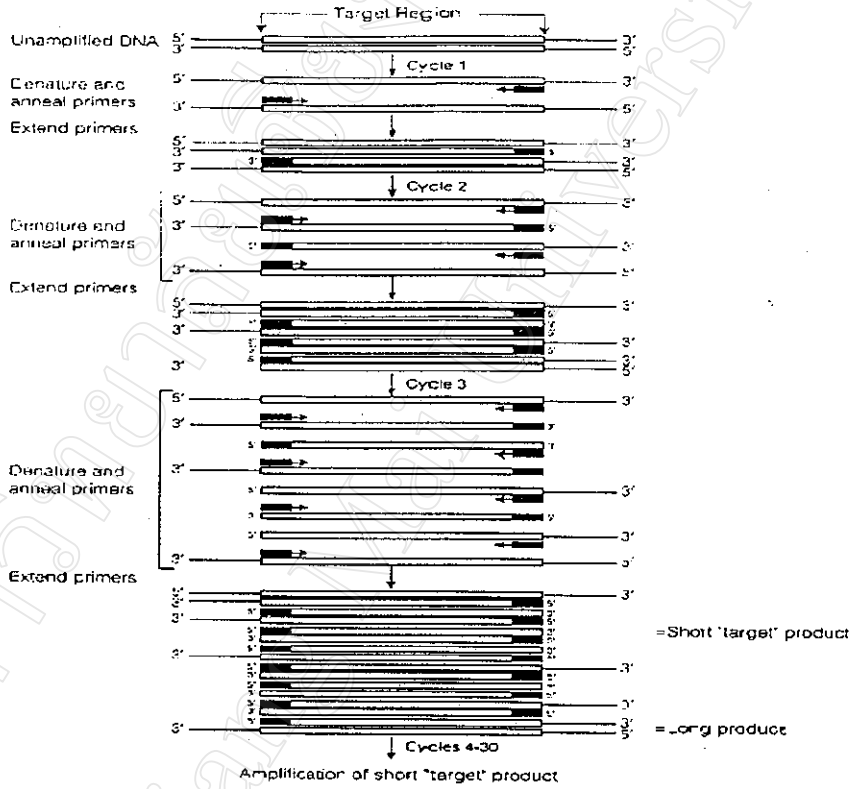

Figure 11. The diagrammatic representation of PCR strategies.

template strands. The temperature applied for primer annealing (occassionary between 50°C and 55°C), and DNA synthesis vary with the enzymes used and are dependent also to some degree on the base composition of the primers. The lower the G+C content, the lower the optimal temperature for the reaction (Kim and Smithies, 1988).

For the synthesis state in the cycle, a number of different DNA polymerases can be applied. The most frequently used enzyme is the heat-stable DNA polymerase of the archaeabacterium *Thermus aquaticus*, designated *Taq* polymerase (Chien *et al.*, 1976). It offers a number of advantages : first, it does not need to add in each cycle (essentially, with a good enzyme preparation, the addition of one unit is normally sufficient for the entire amplification running through 30 or more cycles). Secondly, by allowing synthesis at an elevated temperature, this reduces the chances of unintended oligonucleotide priming by destabilizing mismatch-pairing with unwanted target sequences, as may result from partial homology with random or related, but not identical sequences. This is particularly important if genes or transcripts, which originate from multigene families, are amplified. Mismatch-priming is less likely to occur at higher than lower annealing and/or polymerization temperatures (Saiki *et al.*, 1988). Thirdly, the availability of the heat-stable *Taq* polymerase was a critical prerequisite for the development of automatic equipment for PCRs (Vosberg, 1989).

Taq DNA polymerase is an 832 amino acids protein with an inferrred molecular weight of 93,920 dal, optimal polymerization activity is achieved at 75-80°C and half maximal activity is achieved at 60-70°C (Lawyer *et al.*, 1993). Its thermal stability as measured as a half-life of activity is between 45 and 96 min at 95°C and 9 min at 97.5°C (Saiki *et al.*, 1988 : Lawyer *et al.*, 1993). Under conditions of enzyme excess, the polymerase extends a primer at a maximum rate of 75 nucleotides per second at 70°C (Innis *et al.*, 1988 ; Abramson *et al.*, 1990). The enzyme is moderately processive, extending a primer at an average of 50-60 nucleotides before it dissociates (Abramson *et al.*, 1990), however, the fidelity of *Taq* DNA polymerase base substitution has been estimated to range between $3x10^{-4}$ and $3x10^{-6}$ errors per nucleotide polymerized, depending on reaction conditions (Eckert and Kunkel, 1992).

On the other hand, Vent$_R$® DNA polymerase, the enzyme purified from an *E. coli* strain that carries the Vent DNA polymerase gene from the archaeabacterium *Thermococcus litoralis*, is a higher fidelity thermophilic DNA polymerase. Its fidelity is 5-15 fold higher than that observed for *Taq* DNA polymerase (Eckert and Kunkel, 1991 ; Mattila *et al.*, 1991), derived in part from an integral 3' to 5' proof reading exonuclease activity (Mattila *et al.*, 1991 ; Kong *et al.*, 1993). The combination of *Taq* and Vent DNA polymerase (at reduced

concentration of Vent DNA polymerase) were facilitated the DNA amplification of longer than 5 kb (Barnes, 1994 ; Cheng *et al*, 1994). Presumably, when *Taq* polymerase misincorporates a dNTP, subsequent extension of the newly synthesized DNA either proceeds very slowly or completely stops. The additional proof reading polymerase of Vent$_R$® DNA polymerase serve to remove the misincorporated nucleotide, allowing *Taq* DNA poymerase to continue the extension of the new strand (Barnes, 1994 ; Cheng *et al.*, 1994). Moreover, more than 90% of Vent$_R$® DNA polymerase activity remains following 1 hour's incubation at 95°C. According to the more heat stable activity and higher fidelity caused by the proof reading activity of Vent$_R$® DNA polymerase, it was more attractive for DNA amplification by PCR prior to sequencing than *Taq* DNA polymerase.

1.2.12 High GC-content DNA template amplification by Polymerase Chain Reaction.

Since its conception, PCR has become a widely used molecular biology technique. One difficulty that has plagued the PCR technique is the amplification of GC-rich DNA sequences. Several methods have been developed to overcome this problem including the use of 7-deaza-2'-deoxy-guanosine (dcGTP) (McConlogue *et al.*, 1988) and nested PCR (Happi *et al.*, 1988). Some of the co-solvents such as formamide (Sarkar

et al., 1990), dimethyl sulfoxide (DMSO) (Pomp and Medrano, 1991 ; Filichkin and Gelvin, 1992; Chester and Marshak, 1993), and betaine (Henke *et al.*, 1997), were introduced to the PCR to reduced the Tm of the oligonucleotide primer and DNA template. Glycerine has been reported to stabilize the polymerase activity when longer incubation in denaturation step coupled with relatively high annealing and extension temperatures were used (Smith *et al.*, 1990). Complete denaturation of DNA template at relatively high temperatures and the higher annealing and extension temperatures, facilitated by using the more thermostable DNA polymerase such as Ampli*Taq* DNA polymerase or Vent$_R$® DNA polymerase, were reported to overcome such amplification (Varadaraj and Skinner, 1994).

Schuchard and co-workers (1993) reported the successful amplification of 179 and 111 bp PCR products contain avian *c-myc* proto-oncogene from hen oviduct genomic DNA extracted, with approximately 75% GC-content, by using the two step "Hot" PCR. Without formamide, the specific PCR products were obtained when the modified forward and reverse primers (Tms between 70-74°C) were used (70°C and 76°C annealing/extension temperature). However, the intensity of those specific products observing on the gel were obviously decreased, compared with the standard reaction mixture plus 5% formamide with both of the low or high Tm primer sets, or even the

combination of one low and the other high Tm primers when the native three-steps PCR protocol (50°C annealing and 72°C extension temperatures) was used. It seems to be that the most important parameter improving this amplification system was not exactly the two-step cycle condition, but the 5% formamide included in the reaction mixture and also only the short fragments (less than 300 bp) amplified.

"Touchdown" PCR, the method described by Don and co-workers (Don *et al.*, 1991) seems to improve the amplification of all type of the DNA fragments. The annealing temperature is decreased 1°C every 2 cycles until the exactly annealing temperature is reaching and so on for 20-30 cycles, generating either non-specific or specific products. The specific one will be generated when the last stage with the exact annealing temperature is reached. This can help to reduce the time used for annealing temperature optimization experiments. And it also seems to be appropriate for the primer set that cannot easily be modified, extended or completely "moved" to other sequences and therefore, not in the same Tm range.

1.2.13. α_2- and α_1-globin genes amplification by Polymerase Chain Reaction.

The amplification of the human α_2-and α_1-globin genes was very complicated caused by the strict parameters considered in primers

selection step. The coding region of α_2- and α_1-globin genes appear to be identical but considerable differences have been shown to exist between the 3' untranslated portions of the two genes. In the primer selection step, these facts were used for the consideration. 1) specific PCR products generated should be as short as possible (less than 1 kb was acceptable) while all 3 exons of each gene are still covered by the primers used, 2) the prevention of unappropriated primer sequences, such as the stretch of Gs or Cs and 3) the Tm of each primers should be relatively high (more than 60°C) which be facilitates the reduction of secondary structure of primer-template at the relatively high annealing and extension temperature is allowed in PCR. They were made the specific primers selection for each gene more difficulty. Shorter PCR products which are normally amplified easier than the other longer one was widely accepted and may cause the successful amplification of avian *c-myc* proto-oncogene either with 3 steps PCR or two step "Hot" PCR, when formamide included (Schuchard *et al.*, 1993). In primer selection step, by taking the advantage of the sequences that are shown to be identical at the 5' end and the differences that were existed in the 3' end of these 2 globin genes, the same sequences of forward primers that specific annealed to 5' end of both α-globin genes were used. The specific amplification of α_2- and α_1-globin genes were then indicated by the other different reverse primers. For reverse primers, in primer

selection step, however, the longer extension of each may show the Tm closer to the forward primer but always occupying the stretch of Cs or Gs (unusual sequences) which should be prevented. Then, the best 2 primer sets for these amplifications were selected according to Molchanova and co-worker (1994). According to their report, "asymmetric" PCR of α_2- and α_1-globin genes were individually performed by using a ratio of 1:25 of forward primer to reverse primer, 7.5% formamide and the native three-steps PCR protocol (68°C annealing temperature) were reported to enhance those specific globin genes amplification products. The majority of PCR products generated were the single-stranded DNA fragments which appropriated for sequencing with Sequenase Version 2.0 (United States Biochemical, Cleveland, OH, USA).

In attempting to amplify the two α-globin genes, the PCR reaction mixture and PCR protocol were firstly according to Molchanova and co-workers (1994). However, their PCR system was not repeatable in this laboratory. The further optimization of any parameters such as formamide and $MgCl_2$ concentration and the annealing temperatures did not improve the amplification system. The most probable failure of PCR may be due to their cycle condition.

The various formulas used to calculate the Tms of each primer-template combination considered different factors such as monovalent

salt concentration, base compositions and overall length of the oligonucleotide (Hecker and Roux, 1996). However, even the most sophisticated formulas do not take into account all of the factors influencing PCR (Rychlik, 1994). According to the two Tm-calculation equations that are widely used, the monovalent salt concentration and primer length were considered by equation 2, but not by equation 1.

$$Tm = 4(G+C) \times 2(A+T) \qquad \text{(equation 1)}$$

and

$$Tm = 81.5 + 16.6(\log 10\ [Na+] + 0.41(\%G+C) - (675/n) \qquad \text{(equation 2)}$$

; where [Na+] = the molar salt concentration (M)

and n = the oligonucleotide primer length (bp)

As shown in Table 2, Tms calculated by those two equations for each oligonucleotide primer used in α_2- and α_1-globin genes amplification (Molchanova *et al.*, 1994) show the differences in degrees celsious up to 15. According to Molchanova and co-workers (1994), the exact annealing temperature used in three-step PCR protocol was 68°C, which was nearly the value calculated from equation 2.

Table 2. The primer sequences and their melting temperatures calculated from the two most widely uesd equations. The values listed under 2+4[1] were calculated using the formula Tm(°C) = 2(A+T) x 4(G+C). The values listed under GC[2] were calculated using formula : Tm(°C) = 81.5 + 16.6 log [Na^+] + 0.41(%GC) - 675/N, where N is the length of the primer, and [Na^+] were 0.05 and 0.001 respectively. D[3] gives the difference between lowest and highest calculated melting temperature.

Primer	Primer sequence	Calculated Melting Temperature (°C) 2+4[1]	GC[2]	D[3]
945	5'-CGC GCT CGC GGC CCG GCA C-3'	72.00	61.06	10.94
944	5'-GGG GGG GAG GCC CAA GGG GCA AGA A-3'	82.00	62.42	19.58
2618	5'-GGG AGG CCC ATC GGG CAG GCG GAA C-3'	86.00	64.06	21.94

In this thesis, the sequencing methodology used for the study of α_2- and α_1-globin genes was Chain-termination cycle sequencing in which either single- or double-stranded DNA fragments can be used as templates for the reaction, therefore, the optimization of α_2- and α_1-globin genes PCR was concentrated to the symmetric PCR in which double-stranded PCR product was obtained. By using the PCR mixture that modified from McKeown and co-workers (1994) with the primer sets according to Molchanova and co-workers (1994), combining with 7.5% DMSO and the "Semi-touchdown" PCR protocol in which the PCR protocol was divided into 3 stages with different annealing temperatures optimized from the Tm value calculated from those two equations as indicated, specific 795 and 785 bp of α_2- and α_1-globin genes PCR products were enhanced.

1.2.14 Cycle sequencing.

Cycle sequencing, first described in 1989, is an integration of two technologies ; the dideoxy chain termination method of Sanger (Sanger *et al.*, 1977) and thermal cycling methodology as used in the PCR technique (Saiki *et al.*, 1985).

The Sanger method of DNA sequencing is now a standard technique by which it is possible to obtain several hundred bases of DNA sequencing information per cycling reaction. The method requires

the use of specific terminators of DNA elongation, 2',3'-dideoxy-ribonucleoside triphosphates (ddNTPs). These analogues of the usual deoxyribonucleoside triphosphate (dNTPs) constituents of DNA lack a hydroxyl group at the 3'-position. They are therefore able to be incorporated into a growing DNA chain by DNA polymerase enzymes, but are unable to form phosphodiester bonds with the next incoming (d)dNTPs. The growing chain therefore terminates whenever a ddNTPs is incorporated. By performing four reactions, each with a specific ddNTPs analogue, it is possibed to terminate the growing oligo-nucleotide selectively at dG, dA, dC and dT position. A series of nested fragments is thus produced, each with a common origin but ending in a diffferent nucleotide. These can be resolved by high resolution denaturating gel electrophoresis to give an archetypal sequence ladder.

Visualization of the sequence information can be achieved using a number of different formats, depending on the labelling method used. Conventionally, radioisotope label (^{32}P, ^{35}S and more recently ^{33}P) have been used to produce autoradiagrams of DNA sequence ladders (Evan *et al.*, 1993). These methods are still widely used, but alternative non-radioactive methods are becoming increasingly common. These include the use of fluorescent or chemiluminescent labels, and direct in gel detection by silver staining. One major benefit of the fluorescent based labelling technologies is the ability to automate sequencers for the

analysis procedure (Ansorge *et al.*, 1987), which greatly enhances both the rate and accuracy of sequencing.

There are three different principal methods of incorporating labels into the DNA strand being sequenced ; using labelled primers, incorporating labelled nucleotides into the nascent DNA chain and using labelled dideoxynucleotide terminators (Prober *et al.*, 1987 ; Lee *et al.*, 1992). In contrast, the silver staining method does not require labelling of the DNA.

One of the limitations of the standard polymerization techniques developed for DNA sequencing is that only one copy of the labelled DNA is produced from each primed single stranded template. This means that the sensitivity of the entire procedure is limited by the amount of DNA which can be included in the reaction and the sensitivity of the detection method. However, the application of a thermostable DNA polymerase enzyme to this procedure means that these constraints are removed : one an initial labelled DNA is produced, it can be dissociated from the template by heating the reaction mixture to 94°C. T7 DNA polymerase, the usual enzyme for dideoxynucleotide sequencing, would degrade at this temperature, but the thermostable polymerase remains intact. The reaction mixture is then cooled to circa 55°C when another sequencing primer binds to the template DNA molecule. A second copy of the labelled DNA is then synthesised by

heating the reaction mixture to the optimum temperature for the thermostable polymerase, typically to 72°C. By repeating this process many folds, a linear amplification of the labelled, nested, dideoxy nucleotide-terminated DNA fragments is produced (Carothers *et al.*, 1989 ; Murray, 1989)

1.2.15. Cycle sequencing with Dye terminators and the Ampli*Taq* DNA polymerase, FS enzyme.

The high resolution of DNA sequencing has generated a considerable interest in DNA sequencing protocols for the analysis of PCR products (Gibbs, 1990). The sequencing strategies are broadly divided into two approaches : cloning and direct sequencing from PCR products. The cloning of the PCR products prior to sequencing has the advantage that standard sequencing methods can be used and when complex alleles are analyzed from a heterozygote the linkages of different base substitutions can be determined. The disadvantages of the cloning and sequencing mainly arise as a consequence of the frequency of errors that are introduced by the *Taq* polymerase during amplification. To overcome this difficulty it is necessary to analyse several independent isolates to derive a consensus sequence. In addition to the polymerase errors, the cloning method has the disadvantage of being

relatively time-consuming and labor intensive compared to direct DNA sequence analysis.

Direct sequencing can rarely be carried out by using standard sequencing protocols, presumably because of the presence of a low level of leftover primers and nonspecific PCR products present in the reactions, and also as the second strand of the PCR product can compete with the sequencing primer for the DNA template strand. Approaches have been devised to overcome these problems, by either partially purifying the double stranded PCR products or converting the double stranded fragments to single DNA strands. The first report of single-strand production by PCR was achieved by using an unbalanced ratio of the two oligonucleotide primers in "asymmetric PCR" (Gibbs, 1990). During asymmetric amplification, the supply of one of the PCR primers over the concentration of the other reverse primer are primed and extended in one direction only. The result is a linear increase in the amount of one strand relative to the other, and a template mixture that is much more suitable for dideoxynucleotide DNA sequencing than is a double-stranded PCR product.

The majority of direct DNA sequencing studies have been carried out using radioactive labels, but recently fluorescence DNA sequencing procedures have been adapted for the analysis of PCR products (Gibbs *et al.*, 1989). The fluorescent protocols take advantage of automated

devices for the direct on-line monitoring of the sequencing gel electrophoresis and computer-directed data entry.

1.2.16 ABI 310 "PRISM" DNA Sequencer.

The ABI 310 "PRISM"DNA Sequencer detects fluorescence from four different dyes that are used to identifiy the A, G, C, and T extension reactions. Each dye emits light at a different wavelength when excited by laser light ; thus, all four colors (and there are all four reactions) can be detected and distinguished in a single gel lane. This strategy improves sequencing accuracy because it eliminates problems caused by variations in electrophoretic mobility from lane to lane. It also increases the number of templates that can be analyzed on a single gel by a factor of four.

The fluorescent dye labels can be possibly incorporated into DNA extension products using either 5'-dye labelled primers (dye primers) or 3'-dye labelled dideoxynucleotide triphosphates (dye terminators). With dye terminator labelling, each of the four dideoxy terminators (ddNTPs) is tagged with a different fluorescent dye. Thus, the growing chain is simultanously terminated and labelled with the dye that corresponds to that base.

Cycle seuqencing with dye-terminators and Ampli*Taq*, FS enzyme is very useful when using internal unlabelled sequencing primers. The

reaction is carried out in a single tube, since each growing chain is simultaneously terminated and labelled with the dye that corresponds to the applicable base. False stops are not detected, and compression is rare, so the chemistry can be very successful in GC-rich regions. The single tube reaction and the removal of the excess dye terminator by ethanol precipitation make it fast and simple to perform (Parker *et al.*, 1996).

The polymerase used in cycle sequencing reaction was Ampli*Taq* DNA polymerase, FS which is a member of the *Taq* F667Y family, in which phenylalanine is substituted by a tyrosine at position 667; it can incorporate chain-terminating dideoxynucleotide triphosphates (ddNTPs) much more efficiently than the wild-type *Taq* DNA polymerase. It is an ultra-pure thermostable polymerase obtained by expression of a modified form of the *Thermus aquaticus* DNA polymerase gene in *E. coli*. The enzyme provides rapid nucleotide incorporation, and has no 3' to 5' exonuclease activity. Ampli*Taq* retains substantial activity after repeated exposure to high temperatures (94-96°C) thereby making it an ideal enzyme for cycle sequencing.

Dye terminators cycle sequencing had many advantages and disadvantages as follows :

Advantages :

1. Unlabelled sequencing primers can be used, making it possible to use one of the PCR primers as a sequencing primer. Alternatively, an unlabelled internal sequencing primer can be synthesized. This is a primer with a 3' end that begins internally to the 3' end of the PCR primer, so it does not prime other artifacts of the PCR, for example primer dimers. Using an internal primer often improves the signal-to-noise ratio, because fewer contaminating products are available for sequencing primers to anneal to an extend from.

2. The reaction is performed in a single tube, expediting set up.

3. False termination products do not appear. A dye-labelled dideoxy-nucleotide must be incorporated in order for a fragment to be detected.

4. Because this chemistry uses inosine rather than dGTP or deaza dGTP, compressions in GC-rich regions of DNA are rarely observed.

5. The Ampli*Taq*, FS enzyme requires much lower concentrations of dye terminators than the Ampli*Taq* enzyme. Excess dye can be eliminated by ethanol precipitation, instead of spin column chromatography or organic extraction.

6. This chemistry uses the lowest amount of starting DNA template of any of the fluorescent chemistries.

Disadvantages :

1. After amplification, if PCR primers contaminate the PCR products used as a template, both PCR primers can act as sequencing primers, resulting in generation of sequencing reaction products. This method is therefore not recommended for direct sequencing (that is, sequencing without purification of the PCR product).

2. The peak heights are less even than with dye primer sequencing using the Ampli*Taq* DNA polymerase, FS enzyme. More manual interpretation of the data must be performed to check for false negatives or false positives. It is also more important to minimize background noise in the data to avoid the case in which a noise peak combined with a low signal is misinterpreted as a heterozygote. Also, peaks that go off-scale can cause increased background noise making heterozygote determination more difficult. It might be necessary to perform gel purification of the PCR products if the other methods leave too many contaminating products in the preparation.

According to the non-repeatable asymetric amplification of α_2- and α_1-globin genes reported by Molchanova and co-workers (1994), the symmetric PCR products generated by the method described here were used instead as the templates for dye-terminator cycle sequencing. The PCR products were purified by using agarose gel electrophoresis and purification system as indicated before cycle sequencing.

1.2.17. Confirmation of single point mutation having occured in Hb Chiang Mai's α-globin gene by using specific restriction enzyme digestion.

The data derived from the sequencing of Hb Chiang Mai's α-globin gene was expected to indicate that some point mutations occured in its coding regions. This point mutation, may have generated the palindromic sequence recognized by some specific restriction enzymes. By using the specific restriction enzyme digestion experiment, this point mutation can be detected in a simple manner and possibly applied for a screening test of a large population. The identification by using the sequencing method was not appropriate because of the high cost and the complexity of the procedure.

1.3 Objectives.

1. To develop an amplification technique for Hb Chiang Mai's and normal's α_2- and α_1-globin genes by using Polymerase Chain Reaction.

2. To detect the point mutation which are expected to have occurred in Hb Chiang Mai's α-globin genes by using Chain-termination cycle sequencing technique and then confirm the point mutation by using the specific restriction enzyme digestion.