UNIVERSITÀ DEGLI STUDI DI SASSARI

Ph.D SCHOOL IN BIOMOLECULAR AND BIOTECHNOLOGICAL SCIENCES

CURRICULUM BIOCHEMISTRY AND MOLECULAR BIOLOGY

HEMOGLOBINOPATHIES IN MOUNTAINOUS REGION OF THUA THIEN HUE, VIETNAM

Le Phan Minh Triet
Ph.D Thesis

Tutor:
Professor Bruno Masala
Professor Nguyen Viet Nhan

Director:
Professor Claudia Crosio

Cycle XXVI (2010 - 2013)

INDEX

Summary

Due to its high prevalence, β-thalassemia in Southeast Asia is a major public health problem. Therefore, development of genetic counseling and prenatal diagnosis programs by the local government should be a priority. A few limited works have been done to provide the groundwork for such programs in North, South and Central Vietnam. In this Doctoral Thesis, I determined the spectrum of β-thalassemia mutations in two Districts of the Central province of Thua Thien Hue (A Luoi and Nam Dong).

A community-based assessment of thalassemias and hemoglobinopathies was conducted to estimate the prevalence of hemoglobinopathies and to assess their molecular basis. 1100 participants were enrolled including 83.73% of the minorities (the Taoi, the Pako, the Cotu…) and 16.27% of the Kinh.

The blood samples were firstly screened by complete blood count and osmotic fragility test. Hemoglobinopathies were diagnosed by the combination of electrophoresis (Isoelectric focusing of native tetramers and Acid Urea Triton-polyacrylamide gel electrophoresis of the dissociated globin chains) and High Performance Liquid Chromatography (Cation Exchange HPLC and Reversed Phase HPLC). Mutations at the level of β and α globin genes were identified after DNA extraction, amplification by PCR of the affected gene, and DNA sequencing.

Four different mutations of the β^o-thalassemia type (i.e. characterized by the absence of β-globin synthesis *cis* to the mutation) were observed in five subjects:

- Two of these showed the *A*AG→*T*AG nonsense mutation at codon 17, which gives rise to a shortened, unstable, globin of 16 amino acid residues. One resulted a compound heterozygote in combination with the β^E gene (genotype β^0/β^E) and one was β^0/β.

- One showed the G→T substitution at the IVS-I nt 1, which completely prevents splicing of mRNA, combined with the β^E gene (genotype β^0/β^E). Due to the presence of Hb F synthesis ameliorating the clinical severity, the γ globin genes were also sequenced. The C→T substitution in the promoter region, at position -158 with respect to the Cap site of the $^G\gamma$ gene (also known as the *Xmn*I polymorphism) was found in both chromosomes. This is a further observation of an increased Hb F synthesis during erythropoietic stress, under the control of this mutation.

- One was the carrier of a four bp deletion (-TTCT) involving codons 41/42 (a frameshift mutation)(genotype β^0/β).

- One was the carrier of the G insertion at codons 14/15 (a frameshift mutation) (genotype β^0/β).

A total of six samples (three β^E/β, one β^E/β^E, two β^E/β^o) were examined. Sequencing confirmed that the G→A missense substitution at codon 26 (*G*AG→*A*AG) of the β globin gene is responsible for Hb E.

The prevalence of the hemoglobinopathies appears to be higher within the minorities than the main Kinh population.

1. INTRODUCTION

1.1 Structure and function of human hemoglobins

Hemoglobins (Hb) are globular proteins having the fundamental role to carry oxygen (O_2) molecules and carbon dioxide (CO_2) molecules throughout the body, and it serves also to destroy the physiologically important nitric oxide molecule (1). It has evolved to perform its transport functions in a highly efficient manner: a. the oxygen affinity of Hb allows nearly complete saturation with oxygen in the lungs, as well as efficient oxygen unloading in the tissues; b. its affinity increases with oxygenation, resulting in the sigmoid shape of the oxygen dissociation curve; and c. deoxy-Hb binds protons and oxy-Hb releases protons. This last property, which is known as the "alkaline Bohr effect", also facilitates oxygen loading in the lungs and unloading in the tissues (see Fig. 1 and 2). The Perutz models of oxygenated and deoxygenated Hb provide important insights into the structural basis of these three major features of the equilibria of oxygen with Hb (2).

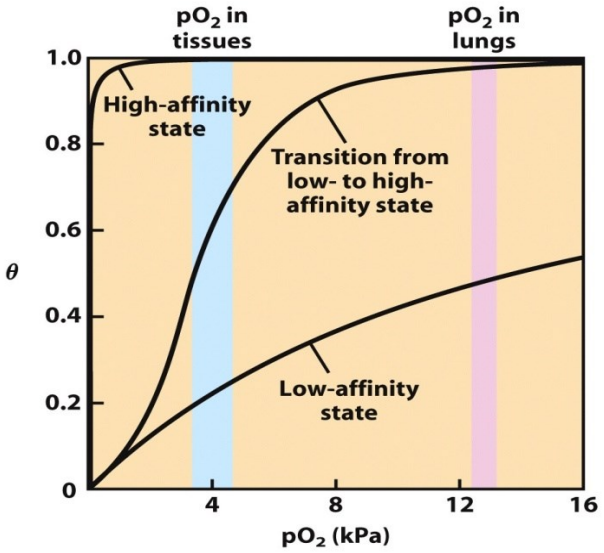

Figure 1. The sigmoid (cooperative) binding curve of human Hb. The sigmoid binding curve can be analyzed as a hybrid curve reflecting the transition from a low-affinity to a high-affinity state. Because of its cooperative binding, as manifested by a sigmoid binding curve, Hb is more sensitive to the small differences in O_2 concentration between the tissues and the lungs, allowing it to bind oxygen in the lungs (where pO_2 is high) and release it in the tissues (where pO_2 is low).

The roles of different parts of the Hb molecule in its equilibria have been deduced from its amino acid sequence, its helical conformation, models derived from x-ray crystallography

(3,4) studies of the kinetics of reactions of Hb with ligands (5) and observations utilizing nuclear magnetic resonance (6). The concentration of Hb within human red cells is extraordinarily high (14 g/dl), and its efficiency as an oxygen carrier is enhanced by its packaging in flexible cells of optimal shape for the diffusion of gases.

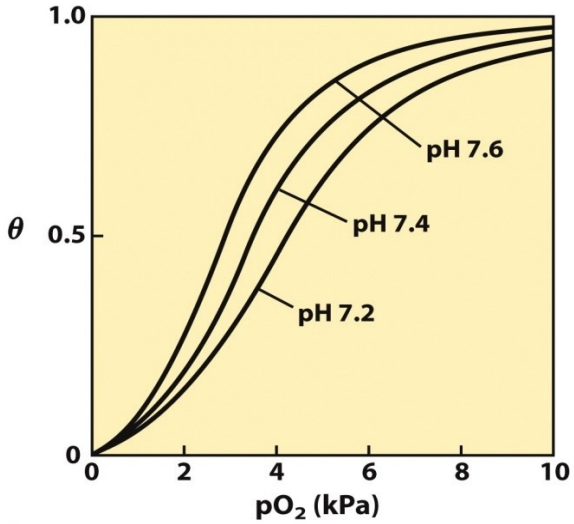

Figure 2. Effect of pH on O_2 binding to Hb (the Bohr effect). The shift of pH from 7.4 to 7.6 (as it is in the lungs) increases O2 affinity, whereas the shift to 7.2 (as in the tissues) decreases O2 affinity thus facilitating its release.

As shown in Fig. 3, Hb molecule consists of four polypeptide subunits, two identical α-globin chains and two identical β-globin chains, covalently bound with a heme prostetic pigments, one in each of the subunits, which are held together by ionic bonds, hydrogen bonds, hydrophobic interactions, and van der Waals forces (1). The heme group contains a positively-charged iron (Fe^{2+}) molecules (Fig. 4) which can reversibly bind to oxygen molecules to be transported to the various areas of the body. As the heme groups bind or release their oxygen loads, the overall Hb undergoes conformational changes which alters their affinity for oxygen (Fig. 1,2).

The heme group, which is typically hidden within the various subunits, is covalently bound, by means of the fifth coordination bond, to a different nitrogen atom belonging to a nearby histidine group, the “proximal His” at position 93 of the primary structure (8 of the F helix) (2). This histidine chain, together with other hydrophobic residues, stabilizes the heme group within each subunit. The O_2 molecule binds to the side of the iron ion that is opposite of the proximal His by the sixth coordination bond, in front of another His residue which is

located at position 64 (E7). His E7 is not directly bound to the heme group nor to the O_2 molecule (it is thus termed as the "distal His") even though it has a very important function: O_2 binds to heme with the axis at an angle because the perpendicular arrangement is sterically blocked by His E7. This effect weakens the binding of O_2 to Hb (2).

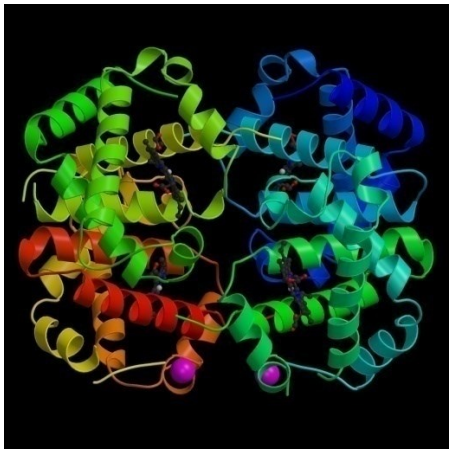

Figure 3. Quaternary structure of Hb. The four subunits are shown in different colors. The heme groups are shown in dark gray. http://ucsdnews.ucsd.edu/newsrel/supercomputer/04-08ProteinDataBank.asp.

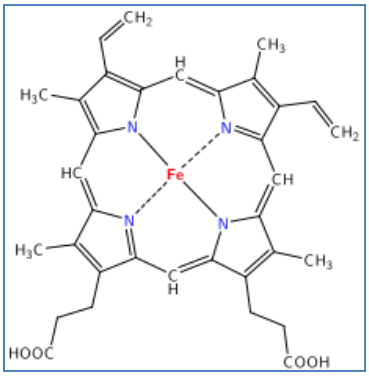

Figure 4. The heme group. The heme consists of a ring structure, protoporphyrin IX, which binds an iron atom in its Fe^{2+} oxydation state. The Fe atom has six coordination bonds, four bonded to the flat porphyrin molecule and two perpendicular to it. Fifth and sixth coordination bonds are perpendicular to the flat cyclic ring structures; one is bound to a N atom of a His residue (the proximal His residue) whereas the other serves as the binding site for an O_2 molecule. This group is found in myoglobin, Hb, cytochromes, and many other heme proteins.

Because His is positively charged, its close proximity to the negatively charged iron ion prevents the iron ion from becoming too oxidized, which would inhibit the binding of oxygen molecules. This is critical to the Hb's function of oxygen transport, since oxygen can bind to $Fe2^{+}$, but not $Fe3^{+}$. Also, the size, shape, and location of this distal histidine chain limits the

amount of CO2 molecules that will bind to the heme group (7). Because the heme group has a greater natural affinity for carbon monoxide than for oxygen, the lack of this distal histidine chain would allow heme groups to bind significantly more to carbon waste than to oxygen, preventing Hb proteins from providing cells with the necessary oxygen molecules for metabolic activities.

After an oxygen molecule binds to one of the heme groups of any subunit, other subunits undergo conformation changes exposing their own heme groups, thus giving the entire Hb structure greater oxygen affinity (8). The bond between oxygen the oxygen atom and the iron ion pulls the iron molecule closer to its heme group, which then pulls the proximal distal histidine chain backwards into the Hb molecule (8). This pull creates a strain on the other subunits, breaking ionic bonds in such a way that reveals their obscured individual heme groups. This positive cooperation allows binding at one subunit to increase the binding affinity at other subunits (9). As the result, the whole deoxyHb tetramer (T state) changes its conformation collapsing into the oxyHb (R state). A representation of this pulling is showed in Fig. 5.

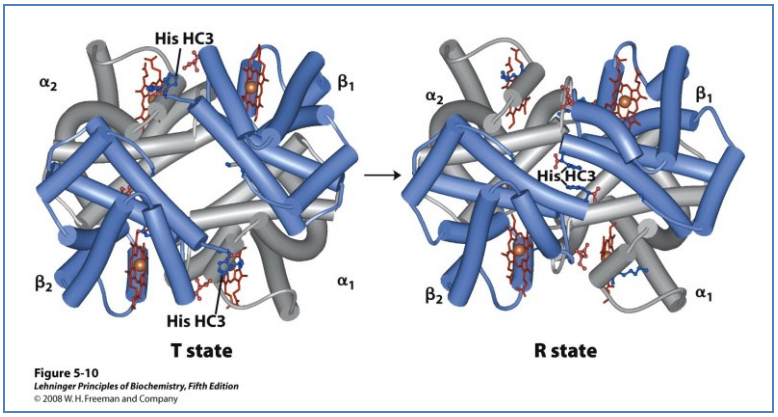

Figure 5. The T → R transition. In these depictions of deoxyHb the *α* subunits are blue and the *β* subunits are gray. Positively charged side chains and chain termini involved in ion pairs are shown in blue, their negatively charged partners in red. The transition from the T state to the R state shifts the subunit pairs substantially, affecting certain ion pairs. Most noticeably, the His HC3 residues at the carboxyl termini of the *β* subunits, which are involved in ion pairs in the T state, rotate in the R state toward the center of the molecule, where they are no longer in ion pairs. Another dramatic result of the T → R transition is a narrowing of the pocket between the *β* subunits.

The behaviour of Hb after binding of protons or CO_2, which causes a conformational change in the protein facilitating the release of oxygen indicates it is an allosteric protein. The sigmoidal (cooperative) curve which makes it efficient in binding (taking up O_2 in lungs), and

efficient in unloading (unloading O_2 in tissues) is a demonstration of an allosteric response to the binding of molecules other than O_2 in the heme pocket. In people acclimated to high altitudes, the concentration of 2,3-Bisphosphoglycerate (2,3-BPG) in the blood is increased, which allows these individuals to deliver a larger amount of oxygen to tissues under conditions of lower oxygen tension. This phenomenon, where molecule Y affects the binding of molecule X to a transport molecule Z, is called a heterotropic allosteric effect.

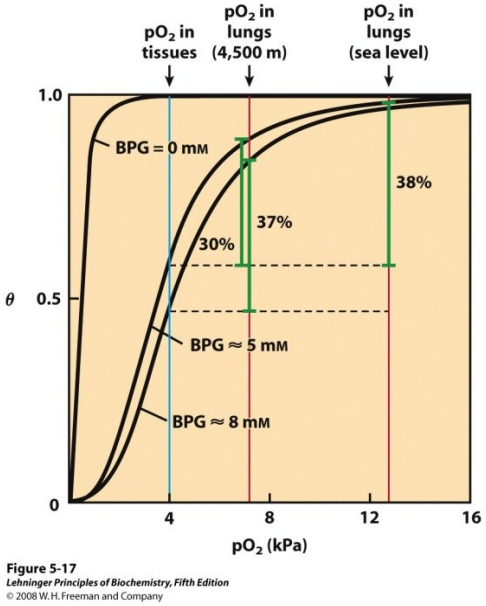

Figure 6. Effect of BPG on oxygen binding to Hb. The 2,3-BPG concentration in normal human blood is about 5 mM at sea level and about 8 mM at high altitudes. Hemoglobin binds to oxygen quite tightly when BPG is entirely absent, and the binding curve seems to be hyperbolic. At sea level, Hb is nearly saturated with O_2 in the lungs, but just over 60% saturated in the tissues, so the amount of O_2 released in the tissues is about 38% of the maximum that can be carried in the blood. At high altitudes, O_2 delivery declines by about one-fourth, to 30% of maximum. An increase in BPG concentration, however, decreases the affinity of Hb for O_2, so approximately 37% of what can be carried is again delivered to the tissues.

Fetal Hb (HbF, α2γ2), which is found in the developing fetus, binds oxygen with greater affinity than adult Hb. This means that the oxygen binding curve for HbF is left-shifted (i.e., a higher percentage of Hb has oxygen bound to it at lower oxygen tension), in comparison to that of adult Hb (HbA). As a result, fetal blood in the placenta is able to take oxygen from maternal blood.

Hemoglobin also carries nitric oxide (NO) in the globin part of the molecule. This improves O_2 delivery in the periphery and contributes to the control of respiration. NO binds reversibly to a specific Cys residue in globin; the binding depends on the state (R or T) of the Hb. The resulting S-nitrosylated Hb influences various NO-related activities such as the control of vascular resistance, blood pressure and respiration. NO is not released in the cytoplasm of erythrocytes but transported by an anion exchanger called AE1 out of them (10).

1.2. Localization and organization of globin genes

Human Hbs are tetrameric molecules encoded at two separate *loci*, the β-like globin gene cluster located on chromosome 11p15.5 (close to the olfactory receptor genes) and the α-like globin gene cluster on the terminus of chromosome 16p13.3 (close to heterochromatic gene encoding a putative RNA-binding protein) (11,12). As shown in Fig. 7 and 8, in each cluster the active genes are arrayed on the chromosome in the same order they are expressed developmentally: 5'-ε(embryonic)-^{G}γ(fetal)-^{A}γ(fetal)-δ(minor adult)-β(major adult)-3', and 5'-ζ(embryonic)-α_2(fetal and adult)-α_1(fetal and adult)-3', respectively. Hemoglobin production is characterized by two switches: the production of embryonic Gower1 ($\zeta_2\varepsilon_2$), Gower2 ($\alpha_2\varepsilon_2$) and Portland ($\zeta_2\gamma_2$) Hbs switches around the first two months of gestation to the production of two different fetal Hbs (HbF) ($\alpha_2{}^{G}\gamma_2$ and $\alpha_2{}^{A}\gamma_2$) followed, just before birth, to the major adult HbA ($\alpha_2\beta_2$) and minor adult HbA_2 ($\alpha_2\delta_2$) tetramers. As the result of the second switch, at birth the circulating Hb contains from 70 to 80% of HbF whereas 6 months later HbF covers less than 4-5 % of the total. The final adult Hb pattern is reached at 1 year of life (Fig. 8). At that time HbA comprises ~97%, HbA_2 ~2% and HbF ~1%. It is at this stage that mutations affecting the β gene become clinically apparent. At birth, HbF contains ^{G}γ and ^{A}γ chains in the 70:30 ratio. The switch from fetal to adult Hb production is not complete, so that in the small amounts of HbF which persist in adult life the proportion of the two chains reverses to 40:60. The switch from fetal to adult Hb is not due to changes in stem cell populations but rather to changes in programs of gene expression occurring in the progeny of a single stem cell population. All adults have residual amounts of HbF, present in a subset of erythrocytes called F-cells which also contains HbA. The levels of HbF and F-cells in adults vary considerably, and are largely genetically controlled (13).

1.2.1 Detailed chromosomal organization of the human globin genes.

Two other α-like globin genes have been identified and characterized in the α-globin gene cluster, but their roles, if any, in encoding globin polypeptides are still uncertain. The θ-globin gene located to the 3' or C-terminal side of the duplicated α-globin genes (16). It is more closely related to the α-globin genes than to the ζ-globin genes and is expressed at low levels in erythroid cells (17,18). Clear homologs to the θ-globin gene are found in the homologous position in other mammalian α-like globin gene clusters.

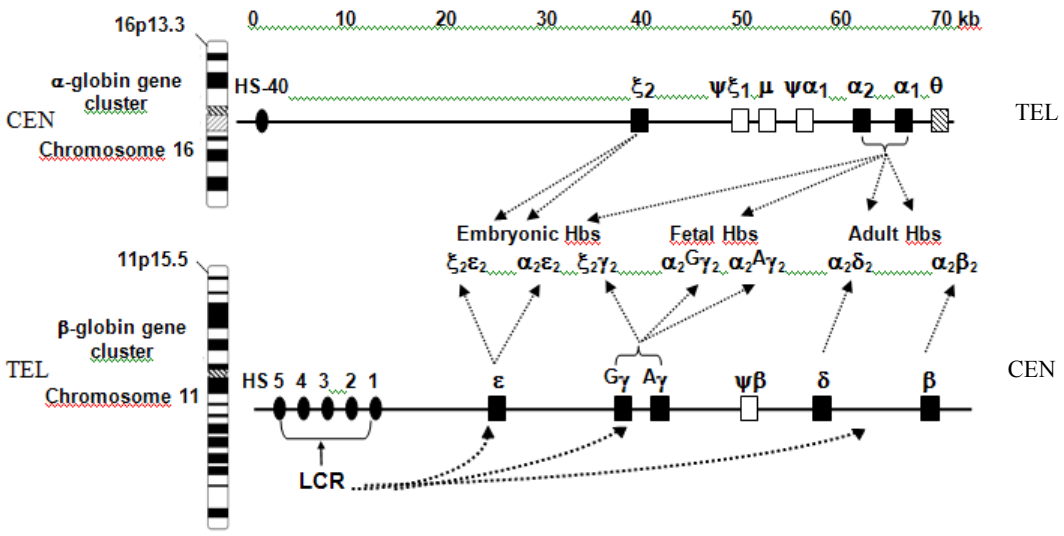

Figure 7. Basic organization of human α- and β-globin gene complexes and expression of the globin genes during ontogenesis. The locations of the α-globin gene complex very close to the telomere of the short arm of chromosome 16 and the β-globin gene complex on the short arm of chromosome 11 are shown. The genes are shown as boxes, named according to the globin polypeptide that is encoded. In both diagrams, the 5'→3' transcriptional orientation is from left to right. The orientations with respect to the centromere (CEN) and telomere (TEL) are opposite; the α-like globin genes are transcribed toward CEN whereas the β-like globin genes are transcribed toward TEL. The composition of Hb produced at progressive developmental stages is given between the clusters. The stage specific interaction between the β-LCR and the cluster is also schematized.

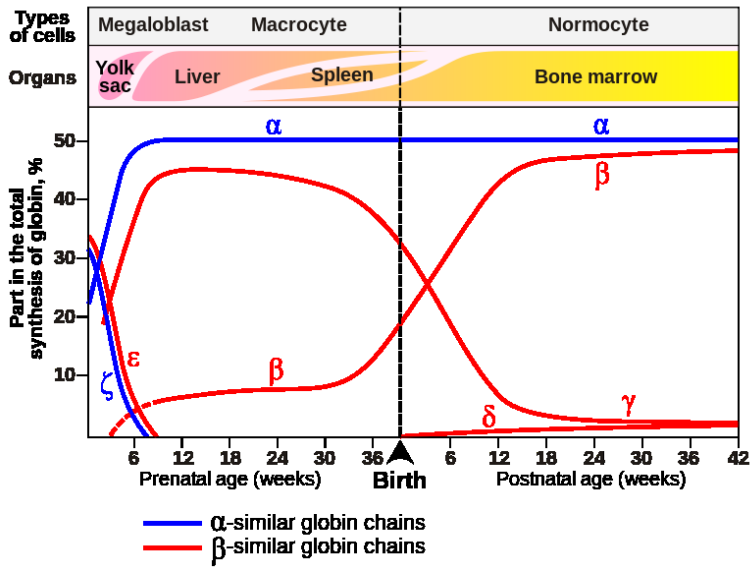

Figure 8. Globin chains production during ontigenesis. The production of embryonic Gower1 ($\zeta_2\varepsilon_2$), Hb switches around the first two months of gestation to the production of HbF ($\alpha_2\gamma_2$) followed, just before birth, to the major adult HbA ($\alpha_2\beta_2$) and minor adult HbA_2 ($\alpha_2\delta_2$) tetramers.

The μ-globin gene is located just 3' of the ψζ1-globin pseudogene (19, 20); it was initially called ψα2 (21) but with more accurate sequencing it is clear that this gene does not contain mutations that would render it inactive. It is a distant relative, being equally divergent from both α-globin and ζ-globin genes. Both the θ-globin gene and the μ-globin gene are transcribed and spliced in erythroid cells, albeit at much lower levels than the α-globin gene. Curiously, no Hb containing the θ-globin chain or the μ-globin chain has been identified, even by the sensitive mass spectrometry (20). Furthermore, the predicted structure of the θ-globin chain suggests that it would be unlikely to function normally as a Hb subunit (22). Thus these genes remain a puzzle. They tend to be retained over mammalian evolution, hence indicating constraint for some function. They are expressed at the RNA level but do not appear to be translated into a polypeptide. Perhaps they or their RNA transcripts play some role that has yet to be discovered.

1.2.2 Structure of globin genes

The coding region of each globin gene in humans (as well as in other vertebrates) is interrupted at two positions by tracts of noncoding DNA which were called intervening sequences (IVSs) or introns (23). In the β-like globin genes, the introns interrupt the sequence between codons 30 and 31 and between codons 104 and 105; in the α-globin gene family, the

IVSs interrupt the coding sequence between codons 31 and 32 and between codons 99 and 100 (Fig. 9). Interruptions due to introns occur at precisely the same position in the aligned primary sequence of the α- and β-globin chains. It has been hypothesized that the presence of the introns at these positions predates the separation of α-globin and β-globin genes, occurred in an ancestral vertebrate about 500 million years ago (24). The first IVS (IVS-1) is smaller than the IVS-2 in both α- and β-globin genes, but IVS-2 of the human β-globin gene is much larger than that of the α-globin gene.

The intron sizes of the ζ-like globin genes differs from that of the α-like globin genes since the introns in the α and ψα genes are small, fewer than 150 base pairs (bp), than those of the ζ and ψζ genes (25). The first introns of the ζ and ψζ genes are much larger than their second introns. The IVSs sequences interrupting the coding sequences of structural genes are removed during maturation of mRNA. Figure 9B shows that IVSs are transcribed into globin precursor mRNA (26), and next excised in such a way that the proper ends of the coding sequences joins to give rise the mature mRNA (27). This posttranscriptional processing of mRNA precursors has been termed "splicing". A crucial prerequisite for the proper splicing of globin precursor mRNAs is the presence of specific nucleotide sequences at the exons-introns junctions.

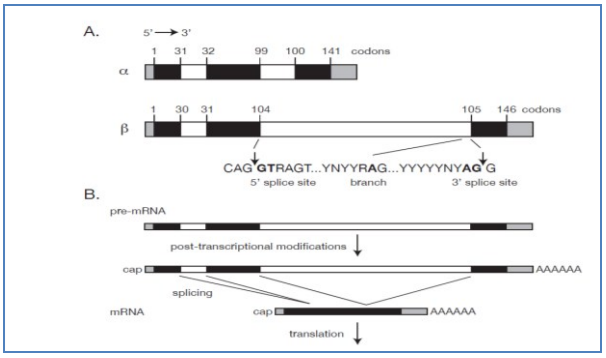

Figure 9. Structure and expression pathway of globin genes. A. General structure of globin genes. The coding sequences are separated by two IVSs (white boxes) into three exons. The first exon has a short 5' untranslated region (UTR, gray box) followed by a coding region (black box). The central exon codes for protein, while the third ends with a 3' UTR. The sizes of the boxes indicate the size of the tracts of the genes. Codon numbers are also given. Under the second intron of the β-globin gene are indicated the "consensus sequence". These are critical for the correct splicing process. Similar sequences are present in all introns. The vertical arrows indicate the splice site junctions within the consensus sequences where cleavage occurs during the process of cutting the introns and joining exons.
B. The pathway for expression of globin genes. The RNA transcript is shown with short boxes corresponding to the UTRs (gray), coding regions (black) and introns (white), with processing and splicing steps occurring in the nucleus to form mRNA (15).

Two different "consensus sequences" have been almost universally found at the 5' (donor) and 3' (acceptor) splice sites of introns (28). These are shown in Figure 9A, with the consensus surrounding the "branch point" A involved in the initiation of splicing. The doublets GT and AG (shown in boldface) at the 5' and 3' ends, respectively, of the intron, are invariant and thought to be absolutely required for proper splicing. This is referred to as the GT-AG rule. Mutations that alter those dinucleotide sequences or create similar consensus sequences at new sites in a globin gene are known determine abnormal processing of mRNA precursors thus representing the molecular basis for some types of thalassemia.

The globin genes nomenclature has been recently standardized following the Guideline for Human Gene Nomenclature. It is showed in Table 1.

Table 1. Nomenclature of the globin genes*

Gene	Nomenclature
ζ	HBZ
α1	HBA1
α2	HBA2
θ	HBQ
ε	HBE
G_γ	HBG2
A_γ	HBG1
β	HBB
δ	HBD

*Aguileta G., Bielawski J.P., Ziheng Y. Proposed nomenclature for the α and β globin gene family. Genes Genet. Syst. 81, 367-371, 2006.

1.2.3. Regulation of globin genes expression

Studies conducted during the past three decades have revealed much about the regulation of the human globin genes. These include information about DNA sequences needed *in cis* such as: promoters, upstream regulatory sequences, proximal enhancers and distal enhancers. The knowledge on these regulatory elements not only facilitated a better comprehension on the molecular basis of hemoglobinopathies but also was the most useful for attempts in improving the conditions of patients with hemoglobinopathies.

Promoters

A promoter is the DNA sequence needed for accurate initiation of transcription. It is a DNA segment that interacts with RNA polymerase II and its accessory factors (such as TFIID and TFIIB) to determine the start site of transcription; this is the basal promoter (29). Five motifs have been associated with basal promoters, and these are found in the promoters of human globin genes (Figure 10.A). They include the familiar TATA box to which TBP binds, along with the BRE to which TFIIB binds and the Inr and DPE motifs to which components of TFIID binds (29).

Efstratiadis et al. (30) revealed the presence of the ATAAA sequence (motif) about 25-30 bp 5' to the start site of transcription of the globin genes, by far considered the most convincing evidence in its consensus role. More recent studies on other promoters revealed the presence and roles of additional motifs close to the start site of transcription. Matches to these motifs can be found readily at the appropriate positions in the human globin genes (Fig. 10.A). Except BRE, each of he motifs in Fig. 10A has been implicated in function by the discovery of a mutation in cases of β-thalassemia (from now on: β-thal). Every base in the TATA box has been altered in one or another thalassemia, and mutations in Inr, MTE and DPE also are associated with this syndrome.

Upstream regulatory sequences

Adjacent to the basal promoter is the upstream regulatory region (29), which runs from about positions -40 to -250 (Figure 10). The CCAAT box appears to be the only one motif highly expressed in this region. Different proteins bind to this motif (30), and it has been implicated in promoter function because of its presence in many promoters and the results of mutagenesis and binding studies (31).

Two other motifs which are found to bind to proteins are the CACC box, which is bound by transcription factors in the Krüppel-like zinc finger class (KLF) and the WGATAR which is the binding site for GATA-1 and related proteins (32,33). Mutations in almost every position in the proximal CACC box have been associated with β-thal.

Proximal Enhancers

The enhancers are sequences that increase the activity of promoters; they can be located on either side of a gene or internal to it, and they can act at considerable distances from genes (34). Two enhancers have been found close to genes in the β-globin gene cluster, one that is 3' to HBB and one that is 3' to HBG1 (Figure 10.A). In both cases the enhancers are less than 1 kb downstream of the polyA additional signal for the respective genes. The HBG1 enhancer

was discovered as the only DNA segment in a 22kb region surrounding the γ-globin genes for DNA segments that boosted expression of a reporter gene driven by a γ-globin gene promoter in transfected erythroid cells (35).

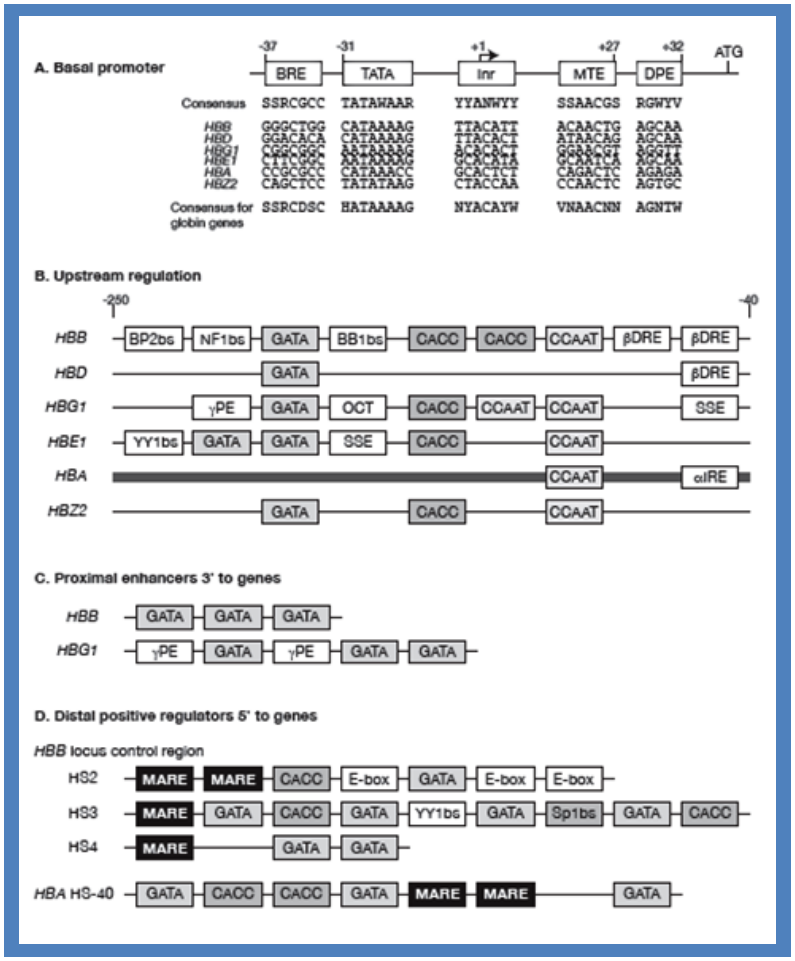

Figure 10. Motifs and binding sites in cis-regulatory modules of globin genes. (A) Motifs in the basal promoter, as review by Maston et al. (29). Numbers along the top are relative to the transcription start site as +1, and ATG denotes the translation start site. Corresponding positions in the globin genes are given for each motif, followed by the consensus derived for the globin genes. Some ambiguous nucleotides are indicated. (B) Motifs in the regulatory regions immediately upstream of the basal promoters. Motifs are indicated by sequence (CCAAT, CACC, and GATA), by the name (βDRE, αIRE, γPE, OCT) or by the protein name followed by "bs" for "binding site" (BP2bs, NF1bs, BB1bs). Boxes found in several upstream regions are shaded. The boxes were placed in the correct order but spacing is not indicated. The thick line for the HBA upstream regions (both HBA1 and HBA2) denotes that it is a CpG island. (C) Motifs in the proximal enhancers. (D) Motifs in distal positive regulators, including three hypersensitive sites of the β-globin LCR and HS-40 for the α-globin gene cluster (15).

Distal Enhancers

Both the α-like and β-like globin gene clusters are regulated by distal control regions. The β-like globin cluster is regulated by the distal LCR (36), and the α-like globin gene cluster is regulated by HS-40 (37). In both cases, deletion of the distal control region is associated with thalassemia. Without the LCR, erythroid expression of a β-globin transgene is not seen in all mouse lines (38), whereas with the LCR, the β-globin transgene is expressed at a high level in erythroid cells in almost all mouse lines, indicating strong enhancement and a reduction in position effects (39). A schematization of how an enhancer might affect gene transcription by bringing distal DNA is showed in Figures 11 and 12.

A clear picture on the mechanism of gene regulation was described in a Melanesian variant of the α-thal due to a single-nucleotide polymorphism located between the adult α-globin genes and their enhancers. The finding that this mutation creates a novel promoter provides support for a mechanism of gene regulation by facilitated chromatin looping (40). The mechanism is showed in Fig. 12.

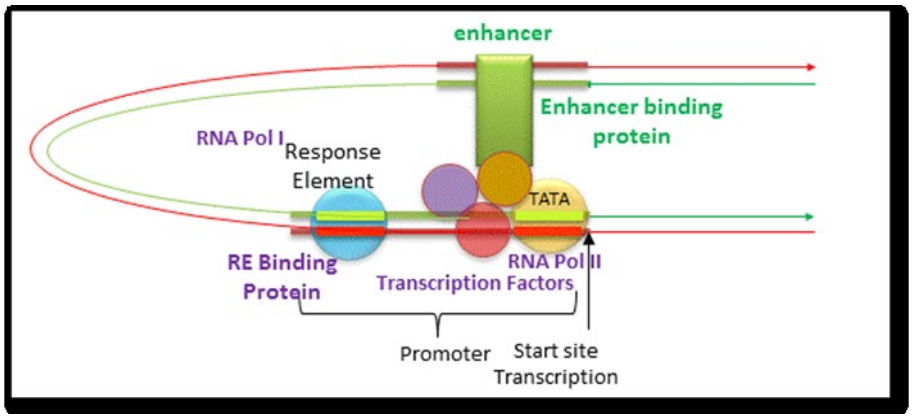

Figure 11. An enhancer might affect gene transcription by bringing distal DNA (perhaps thousands of base pairs away from the start site of transcription) close to the promoter for a gene. The figure also explains how a silencer would work in an analogous fashion.

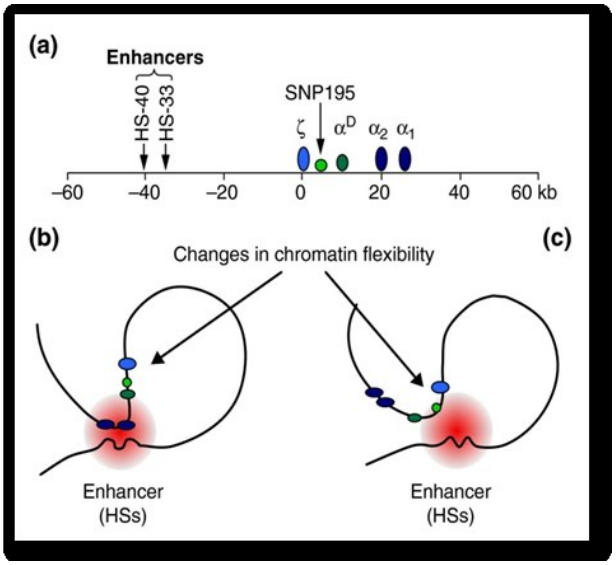

Figure 12. Possible mechanism for the downregulation of α-globin gene expression in hemoglobin H disease. (a) Schematic diagram of the human α-globin locus. The ζ-globin gene (light-blue oval) is expressed in the embryonic stage of development and is silenced at around 6 to 8 weeks of gestation. The α-globin genes (dark-blue ovals) are activated in fetal liver, and then in bone marrow in the adult. The physiological levels of α-like globin gene expression depend on the actions of upstream enhancers (HS-33 and HS-40), mainly HS-40, which is located 40 kb 5' of the ζ-gene. The single nucleotide polymorphism (SNP) 195 is shown as a green circle. (b,c) An interpretation for the SNP195 promoter-induced downregulation of the α-globin genes. The effective interaction between proteins bound by the enhancer (depicted as a red circle) and the α-globin genes is essential for their high-level expression, and is accomplished by chromatin looping. (b) In the normal locus the SNP195 region is lightly acetylated and chromatin flexibility favors interaction between the enhancer and the α1- and α2-genes. (c) When the SNP195 promoter site (green circle) is activated in HbH disease, histone acetylation is increased and the chromatin becomes more flexible as a consequence, resulting in a change in loop size. This change means that the enhancer now preferentially interacts with the new promoter, and no longer influences expression of the globin genes.

1.3. Disorders of the synthesis of hemoglobin

The study of the disorders of the synthesis of Hb is considered a paradigm for understanding insights into the cellular, molecular biology and pathophysiology of inherited genetic disorders. As many as 1570 disorders of Hb synthesis and of the structure have been so far identified and characterized. These are collected, and continuously updated, in the Globin Gene Server (42). Study of these disorders has established the principle of how a mutant genotype can alter the function of the encoded protein, which in turn can lead to a distinct clinical phenotype. Genotype/phenotype correlations have, in turn, provided important understanding of pathophysiological mechanisms of disease.

1.3.1. Classification of hemoglobinopathies

Disorders affecting the synthesis of the globin chains are classified into two categories: qualitative Hb-pathies (also known as hemoglobinopathies, Hb-pathies), and quantitative Hb-pathies (also known as thalassemias). Rarely, a single mutation affects both the structure and expression level of the affected gene, and these have been called quali-quantitative Hb-pathies or thalassemic hemoglobinopathies. Thus the latter were better classified as "dominant" β or α-thal (41).

The hemoglobinopathies (or qualitative Hb-pathies) thus refer conditions characterized by the inheritance of mutated Hbs which contain α or β globin chains with changes at the level of the primary structure of globin which is synthesized at a normal rate. The affected Hb is also called "variant Hb" or "variant globin". In most cases these Hb-pathies are due to a point mutation at the level of the structural gene which changes the meaning of the codon thus replacing the amino acid with a different one (missense mutations). Some sequence changes have little or no effects on Hb functions but are useful polymorphisms for genetic studies. A large number, however, are characterized by an altered Hb function (43). Variant Hbs are designated by letters of the alphabet or by the place names where the condition was first discovered. Even though researchers have identified more than 1100 structural hemoglobin variants, only three (Hb S, Hb C, and Hb E) are widespread. The homozygous state for the Hb S (sickle cell) gene results in sickle cell anemia, whereas the compound heterozygous state for the sickle cell and Hb C genes results in Hb SC disease. Hb SC disease, although milder, also has important public health implications. Hb E, the commonest variant globally, is innocuous in its heterozygous and homozygous states, but since it is synthesized less effectively than Hb A, it interacts with β-thal to produce an extremely common condition called Hb E β-thal, which is an increasingly important health burden in many parts of Asia.

As shown in Table 2, which is the updated summary of the 1578 Hb variants so far identified, all the α and non-α globin genes are subjected to mutations: 325 + 393 concern the α globin genes, 828 the β gene, 105 the δ, 58 the $^{A}\gamma$ and 69 the $^{G}\gamma$ genes, whereas 51 could be considered due to mutations disturbing both the primary structure of the globin chain and the expression level of the affected gene (41).

Table 2. Human Hemoglobin Variants
Summary of mutations categories (Updated August, 2013)
*http://globin.cse.psu.edu/hbvar/menu.html

Query	Count of results
Total entries in database	1578
Total hemoglobin variant entries	1170
Total thalassemia entries	460
Total entries in both variant and thalassemia categories	51
Entries involving the alpha1 gene	325
Entries involving the alpha2 gene	393
Entries involving the beta gene	828
Entries involving the delta gene	105
Entries involving the Agamma gene	58
Entries involving the Ggamma gene	69
Entries with an insertion mutation	65
Entries with a fusion gene mutation	9
Entries with a deletion mutation	201
Entries with a substitution mutation	1279
Hemoglobins with high oxygen affinity	94
Unstable hemoglobins	145
Methemoglobins	9

Table 3 shows the relatively few variants at the $^{G}\gamma$ and $^{A}\gamma$ globin genes having altered functional properties (44).

Table 3.

Variant Hb F showing altered properties

Variant Hb F*	Properties
F-Texas [$^{A}\gamma$5(A2)Glu→Lys]	Three times increased acetylation
F-Xinjiang [$^{A}\gamma^{T}$25(B7)Gly→Arg]	Unstable
F-Clamart [$^{G}\gamma$17(A14)Lys→Asn]	Mild microcytemia
F-Bron [$^{G}\gamma$20(B2)Val→Ala]	Hypochromia and microcytosis (probably due to an interaction with an α-thal)
F-Bonheiden [$^{G}\gamma$38(C4)Thr→Pro]	Severe haemolytic anaemia
F-Cincinnati, [$^{G}\gamma$41(C7) Phe→Ser]	Mild cyanosis; presumed decreased oxygen affinity
F-Lodz [$^{G}\gamma$44(CD3) Ser→Arg]	Mild hemolysis
F-Shanghai [$^{G}\gamma$66(E10)Lys→Arg]	Presumed altered functional properties
F-M-Fort Ripley [$^{G}\gamma$92(F8)His→Tyr]	Cyanosis; presumed low oxygen affinity
F-La Grange [$^{G}\gamma$101(G3) Glu→Lys]	Mildly unstable; increased oxygen affinity with lower heme–heme interaction
F-Poole [$^{G}\gamma$130(G3)Glu→Lys]	Unstable; hemolytic anaemia

*References are found at http://globin.cse.psu.edu.

The thalassemias constitute a heterogeneous group of naturally occurring, inherited mutations characterized by abnormal globin gene expression resulting in total absence or quantitative reduction of α- or β-globin chain synthesis in human erythroid cells. α-Thal is associated with absent or decreased production of β-chains, whereas in the β-thal, there is absent or decreased production of α-chains. In those cases in which some of the affected

globin chain is synthesized, early studies demonstrated absence of an amino acid substitution. The thalassemia gene appears to be allelic with the structural gene encoding α- or β-globin (41).

Increase the knowledge of the molecular basis of β-thal has followed and depended on progress and technical advances in the fields of biochemistry and molecular biology. In particular, recombinant DNA and polymerase chain reaction-based technologies (PCR) have contributed to a virtual explosion of new informations on the precise molecular basis of most

Table 4. Classification of hemoglobin disorders

I. QUANTITATIVE DISORDERS OF GLOBIN CHAIN SYNTHESIS/ACCUMULATION	**C. De novo and acquired α-thalassemia**
The thalassemia syndromes	*α-Thalassemia with mental retardation syndrome (ATR):*
A. β-Thalassemia	Due to large deletions on chromosome 16 involving the α-globin genes
Clinical classification: β-Thalassemia minor or trait	**II. QUALITATIVE DISORDERS OF GLOBIN STRUCTURE: STRUCTURAL VARIANTS OF HEMOGLOBIN**
β-Thalassemia major	**A. Sickle cell disorders**
β-Thalassemia intermedia	SA, sickle cell trait
Biochemical/genetic classification:	SS, sickle cell anemia/disease
β^0-Thalassemia	SC, HbSC disease
β^+-Thalassemia	S/β thal, sickle β-thalassemia disease
δ-Thalassemia	S with other Hb variants: D, O-Arab, other
γ-Thalassemia	SF, Hb S/HPFH
Lepore fusion gene	**B. Hemoglobins with decreased stability (unstable hemoglobin variants)**
δβ-Thalassemia	Mutants causing congenital Heinz body hemolytic anemia
εγδβ-Thalassemia	Acquired instability-oxidant hemolysis: Drug-induced, G6PD deficiency
HPFH	**C. Hemoglobins with altered oxygen affinity**
"Dominant" β-thalassemia (structural variants with β-thalassemia phenotype)	*High/increased oxygen affinity states:*
β-Thalassemia with other variants:	Fetal red cells
HbS/β-thalassemia	Decreased RBC 2,3-BPG
HbE/β-thalassemia	Carboxyhemoglobinemia, HbCO
Other	Structural variants
B. α-Thalassemia	*Low/decreased oxygen affinity states:*
Deletions of α-globin genes:	Increased RBC 2,3-BPG
One gene: α^+-thalassemia	Structural variants
Two genes in *cis*: α^0-thalassemia	**D. Methemoglobinemia**
Two genes in *trans*: homozygous α^+-thalassemia (phenotype of α^0-thalassemia)	*Congenital methemoglobinemia:*
Three genes: HbH disease	Structural variants
Four genes: Hydrops fetalis with Hb Bart's	Cytochrome b5 reductase deficiency
Nondeletion mutants:	*Acquired (toxic) methemoglobinemia*
Hb Constant Spring	**E. Posttranslational modifications**
Other	Nonenzymatic glycosylation
Due to mutations of the ATRX transcription factor gene on chromosome X	Amino-terminal acetylation
α-Thalassemia associated with myelodysplastic syndromes (ATMDS):	Amino-terminal carbamylation
Due to mutations of the ATRX gene	Deamidation

Ref. Forget and Bunn (41)

forms of thalassemia. Table 4 shows a comprehensive classification of the several, different, Hb disorders. The growth of this knowledge has matched the acquisition of detailed

information on the structure, organization, and function of the normal human globin genes. As a result, there have been a progressively clearer and increasingly complex picture of the molecular pathology of this group of genetic disorders. A major conclusion that was drawn was that a relatively limited number of phenotypes can result from a surprisingly large number of varied genotypes.

1.3.2. Pathophysiology of hemoglobinopathies

The causes for the anemia observed in the different forms of thalassemia are primary and secondary. In β-thal, a reduced synthesis of β-globin chains of Hb A ($\alpha 2\beta 2$) will result in an overall deficit of Hb accumulation in erythrocytes and cause a hypochromic, microcytic anemia with a low mean corpuscular hemoglobin concentration in affected erythrocytes. This is the case in both the heterozygous and homozygous states. In the homozygous state, however, another pathophysiological process worsens the anemia and is responsible for the major clinical manifestations in the syndrome referred to as β-thal major or Cooley's anemia.

The continued synthesis in normal amounts of normal α-globin chains results in the accumulation, within the erythroid cells, of excessive amounts of α chains. These chains cannot find complementary globin chains with which to bind, these chains form insoluble aggregates and precipitate within the cell, causing membrane damage and premature destruction of the red cells. The α-chain aggregates are called inclusion bodies or Heinz bodies. The process of inclusion body formation occurs not only in mature erythrocytes, but in particular in the erythroid precursor cells of the bone marrow. As a result, there is extensive intramedullary destruction of erythroid precursor cells, a process that is called "ineffective erythropoiesis". The severity of the clinical manifestations in β-thal generally correlates well with the amount of the free α-chain pool and the degree of α- to non-α-globin chain imbalance. Therefore, the fortuitous coinheritance of α-thal together with homozygous α-thal reduces the degree of α- to non α-globin chain imbalance and leads to a milder clinical course (1). Similarly, coinheritance of β-thal with conditions that are associated with increased levels of synthesis of γ-chains of HbF ($\alpha_2\gamma_2$) leads to less imbalance between α- and non α-globin chain synthesis, resulting in decreased formation of α-chain inclusion bodies, increased effective production of red cells, and their prolonged survival in the circulation. The clinical course in most cases of homozygous β-thal is severe. Although anemia is not evident at birth, severe hypochromic, microcytic, hemolytic anemia develops during the first year of

life and a regular transfusion program must be undertaken to maintain an adequate circulating hemoglobin level.

Owing to changes in the philosophy and practice of transfusion therapy, the clinical manifestations of homozygous β-thal in childhood have changed considerably over the past decades. With modern transfusion therapy, most children will develop normally, with few or no skeletal abnormalities, and will have a reasonably good quality of life. To avoid iron overload from transfusional hemosiderosis, transfusion therapy is usually coupled with a vigorous program of iron chelation, typically using parenterally administered desferrioxamine and more recently, orally effective iron-chelating agents. Although it is possible to maintain iron balance with such a management program, compliance is often difficult to realize. Iron overload eventually develops in most patients and is the major cause of morbidity and mortality in young adults. The one therapy that is curative is bone marrow or stem cell transplantation, which is being increasingly practiced when feasible. The hope for the future is the development of even more effective oral iron-chelating agents and improved approaches to gene therapy.

Studies of the molecular basis of β-thal have demonstrated that the gene defects responsible for the disorder are quite heterogeneous (41). In contrast to α-thal, in which deletions in the α-globin gene cluster account for most of the mutations, the molecular defects associated with β-thal are usually point mutations involving only one (or a limited number of) nucleotide(s), but resulting in a major defect of β-globin gene expression either at the transcriptional or posttranscriptional level, including translation.

Practically every conceivable type of defect in gene expression has been identified in one form or another of β-thal. Over 200 point mutations have been identified. Some deletion types of β-thal have also been described. In cases of β-thal in which β-globin gene expression is not totally absent (so-called β^+-thal), the β-chain that is synthesized is usually structurally normal. There is a syndrome called dominant β-thal in which a highly unstable, structurally abnormal β-globin chain is synthesized, resulting in inclusion body formation in the heterozygous state. The coinheritance of HbE ($\alpha_2\beta_2$ Glu→Lys) with β-thal is very prevalent in Southeast Asia and results in markedly variable and heterogeneous clinical manifestations, the basis for which is poorly understood. Finally, there are a number of β-thal-like disorders, called δβ-thal and hereditary persistence of fetal hemoglobin (HPFH), that are distinguished from the more typical forms of β-thal by the presence of a substantial elevation of HbF in

heterozygotes, as well as in homozygotes and compound heterozygotes. These disorders are usually due to deletions of different sizes involving the β-globin gene cluster, although nondeletion types of these disorders have also been identified.

The great heterogeneity of molecular lesions causing β-thal may appear at first glance to create insurmountable problems in putting this knowledge to practical use in the form of prenatal diagnosis and genetic counseling.

The availability of rapid and accurate PCR reaction-based assays for the detection of specific mutations in small samples of DNA has resulted in a number of surveys for the detection of the prevalence of different β-thal mutations in various population groups. The results of these surveys indicate that a given mutation is usually found only within one racial group and not another. Furthermore, a small number of different mutations, usually five or six, frequently accounts for 90% or more of the cases of β-thal in a given population group. Thus, it is possible to devise efficient and precise prenatal diagnosis programs using DNA-based approaches. Such programs have led to a striking decrease in the number of births of infants with homozygous β-thal in many countries where the disease is prevalent.

1.4. Molecular basis of thalassemia

As already mentioned, the β-thal are characterized by a quantitative reduction in the production of β-globin chains of HbA. More than 200 β-thal alleles have now been characterized (42) involving mutations affecting any of the steps in the transcription of the β-globin gene, posttranscriptional processing of its pre-mRNA, or the translation of its mRNA into protein. The vast majority of simple β-thal are caused by point mutations within the gene or its immediate flanking sequences, although small deletions involving the β gene may also occur. In the case of β-chain production is totally abolished by the mutation, it is referred to as β^{o}-thal, whereas reduced output of β-chains (of normal structure) produces β^{+}-thal, with the mildest forms sometimes referred to as β^{++}- or "silent" β-thal. These common forms of β-thal are inherited as mendelian recessives. Some structurally abnormal β-chain variants are also associated with quantitative deficiencies of β-globin chain production and have a phenotype of β-thal, in which case they are referred to as "thalassemic hemoglobinopathies,". This is the case, for example, of HbE (β26 Glu→Lys). In others, the β-globin variants are so unstable that they undergo very rapid postsynthetic degradation giving rise to a functional deficiency. These hyperunstable β-chain variants act in a dominant negative fashion, causing

a disease phenotype even when present in the heterozygous state, and hence have been referred to as “dominantly inherited β-thal.” (46).

The δβ-thal and hereditary persistence of fetal hemoglobin (HPFH) are disorders related to the β-thal that also involve down-regulation of β-globin gene expression. In δβ-thal the γ gene is also affected and there is variable compensation from increased HbF production. These disorders result from more extensive deletions within the β-globin gene cluster, as do the deletion forms of HPFH in which HbF compensation is sufficient to minimize hematological abnormalities. Other forms of HPFH result from mutations in the promoters of the β-globin genes that not only increase HbF production in adult life but are accompanied by reduced β-chain production.

In this section, the molecular mechanisms underlying the different types of β-thal are described.

The common denominator is absent or decreased synthesis of β-globin chains, resulting in the accumulation of excess α-globin chains that are responsible for the pathophysiology of the disorder. The severity of the phenotype is usually related to the degree of imbalance between α- and non-α-globin chain synthesis, and the size of the free α-chain pool. Hence severity is related to the type of β allele (β^o, β^+, β^{++}), ameliorated by an interacting α-thal (by reducing the α-chain excess) and any increased production of γ-chains (that decrease the excess of free a α-chains by binding to them to form HbF).

1.4.1. The β-thalassemia

1.4.1.1. Nondeletion types of β-thalassemia

These defects account for the majority of the β-thal alleles. They involve single base substitutions, small insertions or deletions within the gene or its immediate flanking sequences, and affect almost every known stage of gene expression. Allele frequency varies widely from one population to another though within any population there are usually a small number of common alleles together with many alleles that are rare for that population. These defects are listed in Table 5 according to the mechanism by which they affect gene function: transcription, RNA processing, or RNA translation. The mutations are continuously updated and the list is accessible electronically through the Globin Gene Server Website (42). Heterozygotes have minimal anemia but hypochromic (mean corpuscular haemoglobin, MCH, 18–24 pg), microcytic (mean corpuscular volume, MCV, 65–80 fL) red blood cells.

They are characterized by an increased proportion of HbA2 (normal <3.2%, β-thal trait 3.5%–6.0%), and HbF levels that vary from normal (<1.0%) to slightly raised (1.0%–3.0%).

Transcriptional mutations

As shown in Table 5, as many as 26 mutations involve the conserved 100 bp DNA sequences that precede the site of the initiation of transcription and represent the β-globin promoter (including the functionally important CACCC, CCAAT, and ATAA boxes), whereas 8 mutations are described in the stretch of 50 nucleotides in the 5' untranslated region (UTR). These mutations affect transcription of the gene giving rise to mild β^{+}-thal. Mutations in all of the three conserved sequence motifs in the β promoter, the two CACCC, CCAAT, and ATAA boxes, have been identified in different patients with β-thal (47). These defects have been described in diverse ethnic groups, with ethnic variation in phenotype. Black individuals homozygous for the -29 A→G mutation have an extremely mild disease (48), whereas a Chinese individual homozygous for the same mutation had thalassemia major (49).The cause of this remarkable difference is not known but may be related to the different chromosomal backgrounds on which the apparently identical mutations have arisen or to the C-T polymorphism at position -158 upstream of the $^{G}\gamma$-globin gene (Xmn1-$^{G}\gamma$ site), which is associated with increased HbF production under erythropoietic stress. The Xmn1-$^{G}\gamma$ site is present in the β chromosome carrying the −29 A→G mutation in Blacks but absent in that of the Chinese (50). The C-T mutation at position −101 to the β-globin gene appears to cause an extremely mild deficit of β-globin. In fact, heterozygotes are "silent" with borderline reduced/normal red cell indices (51,52).

Since the CAP +1 A-C mutation was described (53) 7 additional mutations in the 5' UTR stretch have been characterized. The heterozygotes for this class of mutations are silent, with the hematological indices of α thal carrier. It is possible that the CAP mutation decreases β-globin gene transcription or decreases the efficiency of posttranscriptional addition of $m^{7}G$ (capping) and mRNA translation.

Mutations of the splice site junction affecting RNA processing

The process of the IVSs excision from the precursor mRNA to produce functional mRNA (splicing) must be very accurate. Sequences critical in the splicing process are the invariant dinucleotides GT at the 5' (Donor) and AG at the 3' (Acceptor) splice junctions in the introns. Table 5 shows that 27 mutations affect either of the invariant dinucleotides in the splice junction completely abolish normal splicing and produce the phenotype of β^{o}-thal.

These mutations can be base substitutions that change one or the other of invariant dinucleotides or short deletions that remove them. Genes bearing these mutations appear to transcribe normally and, although some alternative splicing occurs using "cryptic" donor or acceptor sites, the misspliced mRNA do not translate into functional β-globin.

Mutations of splice site consensus sequence affecting RNA processing

The invariant "donor" and "acceptor" dinucleotides essential for the splicing of introns are flanked by sequences that are rather well conserved and a consensus sequence can be recognized at the exon-intron boundaries. These encompass the last three nucleotides of the exon and the first six nucleotides of the intron for the 5' donor site; and the last 10 nucleotides of the intron and the first nucleotide of the exon for the 3' acceptor site. Thirteen different mutations (Table 5) within the consensus sequences at the splice site consensus sequence have been demonstrated to reduce the efficiency of normal splicing to varying degrees and produce a β-thal phenotype ranging from mild to severe. Mutations at IVS1 position 5, G→C, T, or A considerably reduce splicing at the mutated donor site compared with normal. As shown in Fig. 13, the mutations appear to activate the use of three otherwise "cryptic" donor sites (i.e. sequences that mimic the consensus sequence for a splice site but are not normally used). Two of these are in exon 1, and one in IVS1 (54). On the other hand, the substitution T→C in the adjacent nucleotide, IVS1 position 6, only mildly affects normal RNA splicing even though it activates the same three cryptic donor sites as seen in the IVS1–5 mutants (54). Although the IVS1-6 T→C mutation is generally associated with milder β-thal, it has been shown that in some cases apparently identical mutations can be severe (55) perhaps because of the chromosomal background on which the mutations have arisen.

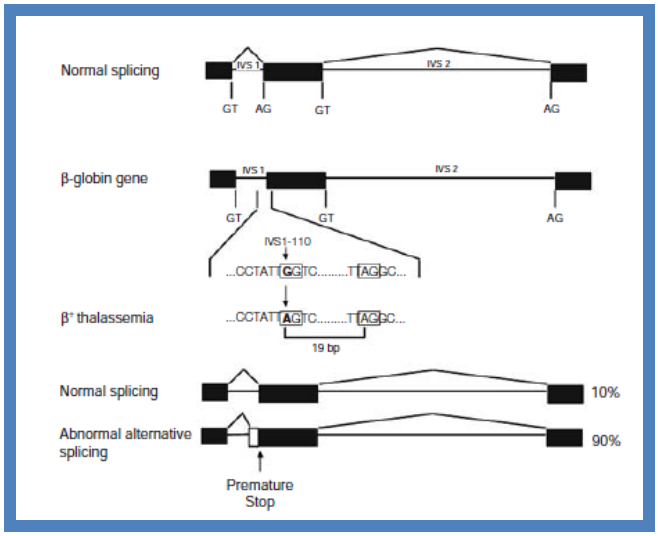

Figure 13. Alternative splicing of precursor β-globin mRNA resulting in β^+-thal associated with the IVS1–110 G→A mutation. The G→A substitution at position 110 creates an AG dinucleotide within a potential 3' acceptor splice site resulting in approximately 90% mRNA aberrantly spliced and 10% normally spliced (45).

Mutations creating new alternative splice sites in introns

Splicing mutation may be due also to base substitutions in IVS that generate new splicing signals, which are alternative to the normal splice sites. Six mutations have been identified in the β-globin gene: two in IVS1 and four in IVS2 (Table 5)(45). Depending on the site and nature of the mutation, the phenotype may be either β^+ or β^o-thal. The splicing mutation at position 110 of IVS1 was the first base substitution identified in a β-thal gene (56,57). This mutation is one of the most common type of β-thal in the Mediterranean population. The mutation is a G→A substitution that creates an acceptor AG in a positive consensus sequence environment, 19 bp 5' to the normal acceptor AG of IVS1 (Fig. 13). This newly created alternative splice site is preferentially used in 80%–90% of the transcripts, whereas the normal splice site is used in only 10%-20% of the transcripts (58,59), thus giving the phenotype of β^+ thal.

Another β-thal gene with a T→G substitution in position 116 of IVS1, leads to a newly created 3' acceptor site. In this case, the normal acceptor sequence, although intact, is not used, and little or no normal β mRNA is produced resulting in a β^o-thal phenotype (60). Three other β-thal genes are due to substitutions within IVS2 that generate new donor sites (Fig. 14). They include the IVS2 position 654 C→T, 705 T→G, and 745 C→G. In each case, an upstream acceptor site at position 579 is activated such that the normal 5' donor site at

exon 2/IVS2 is spliced to the activated site at position 579 and the newly created donor site is spliced to the normal 3' acceptor site at IVS1/exon 3. This two-stage splicing results in the retention of 73 bp of IVS2 in the misspliced β mRNA for the IVS2 654 mutation. Variable amounts of splicing from the normal donor to the normal acceptor also occurs, resulting in phenotypes that range from β^+ to β^o-thal. Several individuals heterozygous for the IVS2 654 C-T mutation with a severe phenotype have been described (50). The unusually severe disease in these individuals is thought to result from the accumulation of the misspliced mRNA, which was predicted to translate into a highly unstable β chain variant with a dominant negative effect. Some normal splicing of IVS2 occurs; therefore, this mutation results in β^+ rather than β^o-thal.

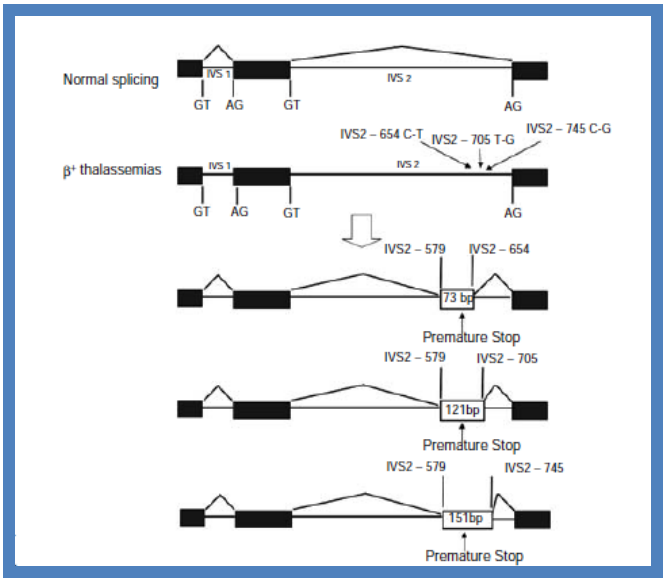

Figure 14. Alternative splicing of precursor β-globin mRNA due to three base substitutions; C→T at position 654, T→G at position 705, and C→G at position 745 in IVS2 of the β-globin gene. Each of these mutations creates a 5' donor site that is preferentially spliced to the normal 3' acceptor site, whereas the same acceptor site is activated upstream at position 579 in IVS2 and spliced to the normal donor site at the exon 2–IVS2 junction. This results in the incorporation of 73, 121, and 151 bp of IVS2 into the aberrantly spliced mRNA associated with each of the different IVS2 mutations, respectively (45).

Mutations creating alternative splice sites in exons

Four mutations have been identified in exon 1 that are associated with activation of cryptic or alternative splice sites (50). Three modify the cryptic splice site spanning codons

24-27 in exon 1 so that it closely resembles the physiological consensus splice sequence AAGGTGAGT and activates it. The reduction in normal splicing is the molecular cause for the mild β^+-thal phenotype of these variants, including the β^E allele which is particularly common in Southeast Asia where it can reach a frequency of 75%. Interaction of the β^E allele with β-thal genes is responsible for a large proportion of the thalassemia major observed in Southeast Asia. Similarly, the A→G mutation in codon 19 activates another cryptic donor site spanning codons 17–19 in exon 1 resulting in a reduced level of normally spliced β mRNA that contains the codon 19 mutation encoding Hb Malay (61).

Mutations modifications causing abnormal posttranscriptional modifications

The nascent globin mRNA molecule is submitted to posttranslational modifications at both of its ends: a methylated 'm^7G' cap structure is added at the 5' end, and a poly adenilic (A) tail is added at the 3' end of the mRNA. Ten different β-thal mutations (Table 5) have been associated with defective polyadenylation due to mutations involving the consensus sequence AATAAA required for the cleavage-polyadenylation reaction. Mutations involving the polyadenylation signal are: seven base substitutions at different positions of the consensus sequence, two short deletions of 2 and 5 bp each, and one deletion of the total AATAAA sequence. These mutations markedly decrease the efficiency of the cleavage-polyadenylation process but do not abolish it completely so that the associated phenotype is that of β^+-thal of moderate severity.

In the 3' UTR region downstream of the termination codon, a C→G substitution at nucleotide 6, and a 13-bp deletion at nucleotides 90, also result in β^+-thal (50).

Mutations resulting in premature termination of translation

Approximately half the β-thal alleles result from the introduction of premature termination codons, due to mutations creating a stop codon or to changes in the reading frame by insertion or deletion of a single or a few nucleotides. These frameshift defects cause premature termination further downstream when the next nonsense codon is reached (45,50). The AAG→TAG (Lys→Stop) at codon 17, and CAG→TAG (Gln→Stop) at codon 39 were the first nonsense mutations characterized and extensively studied. The former is was found in Chinese patients and the latter is the second most common cause of β-thal in the Mediterranean population accounting for 95% of the cases of β-thal in Sardinia.

The frameshift and nonsense mutations, inherited as a typical recessive β^o-thal character, result in premature termination within exon 1 and 2 with a couple of exceptions

terminating prematurely in exon 3 (45). In heterozygotes for such cases, no β-chain is produced from the mutant allele, resulting in a typical asymptomatic phenotype. In contrast, in mutations that produce in-phase termination later in the β-sequence, in exon 3, the mRNA does not undergo nonsense-mediated decay and produce normal amounts of mutant β mRNA, which presumably gets translated into variant β-chains (62).

Mutations affecting the initiation codon

Nine mutations affecting the initiation codon (ATG) have been described and all produce β^o thal (Table 5). One mutation involves an insertion of 45 bp between positions –22 and+23, thus affecting the initiation codon (63). The rest are single base substitutions, two affecting the first (A), three the second (T), and three the third (G) nucleotide of ATG. It is theoretically possible for mutant β mRNAs to be initiated at the next downstream initiation codons, which are located at codons 21 and 22 or codon 55.

Deletions of the β-glogin gene

These deletions involve only the β-globin gene and its flanking DNA without affecting any of the other β-like globin of the cluster and range from 105 bp to approximately 67 kb in size (Fig. 15). The phenotype associated with these deletions is that of β^o-thal. Two deletions remove the 3' end but preserve the integrity of the 5' end of the β-globin gene. A 0.6-kb deletion involving the 3' end of the β-globin gene is a relatively common cause of β^o-thal in Asian Indians (50).

Table 5. Nondeletion mutants that cause β-thalassemia

Mutation		Type	Distribution	References
I. Transcriptional mutations				
Promoter regulatory elements				
1)	−101 (C→T)	β^{++} (silent)	Mediterranean	4, 13
2)	−101 (C→G)	β^{++} (silent)	Ashkenazi Jew	68
3)	−92 (C→T)	β^{++} (silent)	Mediterranean	4, 13
4)	−90 (C→T)	β^{+}	Portuguese	4, 13
5)	−88 (C→T)	β^{++}	US Blacks, Asian Indians	4, 13
6)	−88 (C→A)	β^{+}	Kurds	4, 13
7)	−87 (C→G)	β^{++}	Mediterranean	4, 13
8)	−87 (C→T)	β^{++}	German, Italian	4, 13
9)	−87 (C→A)	β^{++}	US Blacks	4, 13
10)	−86 (C→G)	β^{+}	Thai, Lebanese	4, 13
11)	−86 (C→A)	β^{++}	Italian	4, 13
12)	−73 (A→T)	β^{++}	Chinese	10
13)	−32 (C→A)	β^{+}	Taiwanese	4, 13
14)	−32 (C→T)	β^{+}	Hispanic	198
15)	−31 (A→G)	β^{+}	Japanese	4, 13
16)	−31 (A→C)	β^{+}	Italian	8
17)	−30 (T→A)	β^{+}	Mediterranean, Bulgarian	4, 13
18)	−30 (T→C)	β^{+}	Chinese	4, 13
19)	−29 (A→G)	β^{+}	US Blacks, Chinese	4, 13
20)	−29 (A→C)	β^{+}	Jordanian	199
21)	−29 (G→A)	β^{+}	Turkish	200
22)	−28 (A→C)	β^{+}	Kurds	4, 13
23)	−28 (A→G)	β^{+}	Blacks, SE Asians	4, 13
24)	−27 (A→T)	β^{+}	Corsican	13
25)	−27 to −26 (−AA)	β^{+}	African American	201
26)	−25 (G→C)	β^{+}	African American	198
5′ UTR				
27)	CAP +1 (A→C)	β^{++} (silent)	Asian Indian	4, 13
28)	CAP +8 (C→T)	β^{++} (silent)	Chinese	4, 13
29)	CAP +10 (−T)	β^{++} (silent)	Greeks	4, 13
30)	CAP +20 (C→T)***	?	Bulgarian	4, 13
31)	CAP +22 (G→A)	β^{++}	Mediterranean, Bulgarian	4, 13
32)	CAP +33 (C→G)	β^{++} (silent)	Greek Cypriot	4, 13
33)	CAP +40 to +43 (−AAAC)	β^{+}	Chinese	4, 13
34)	CAP +45 (G→C)	β^{+}	Italian	202
II. RNA processing				
Splice junction				
1)	NS1−(−2) CD30 (AGG→GGG)	β^{0}	Sephardic Jews	4
2)	NS1−(−2) CD30 (AGG→CGG)	β^{0}	Italian Canadian	203
3)	NS1−(−1) CD30 (AGG→ACG) (Arg→Thr)	β^{0}	Mediterranean, US Blacks, N. African, Kurds, UAE	4, 13
4)	NS1−(−1) CD30 (AGG→AAG)	β^{0}	Bulgaria, UAE	4, 13
5)	NS1−1 (G→A)	β^{0}	Mediterranean	4, 13
6)	NS1−1 (G→T)	β^{0}	Asian Indian, SE Asian, Chinese	4, 13
7)	NS1−1 (G→C)	β^{0}	Italian Canadian, Japanese	203, 204
8)	NS1−2 (T→G)	β^{0}	Tunisian	4, 13
9)	NS1−2 (T→C)	β^{0}	US Blacks	4, 13
10)	NS1−2 (T→A)	β^{0}	Algerian, Italian	4, 13
11)	NS2−1 (G→A)	β^{0}	Mediterranean, US Blacks	4, 13
12)	NS2−1 (G→C)	β^{0}	Iranian	4, 13
13)	NS2−2 (T→A)	? β^{0}	Turkish	205
14)	NS2−2 (−T)	β^{0}	Chinese	4

Table 5. Continued

Mutation		Type	Distribution	References
15)	IVS1 –3' end del 17 bp	β^0	Kuwaiti	4, 13
16)	IVS1 –3' end del 25 bp	β^0	Asian Indian, UAE	4, 13
17)	IVS1 –3' end del 44 bp	β^0	Mediterranean	4, 13
18)	IVS1 –3' end duplication 22 bp	β^0	Thai	206
19)	IVS1 –130 (G→C)	β^0	Italian, Japanese, UAE	4, 13
20)	IVS1 –130 G→A	β^0	Egyptian	4, 13
21)	IVS1 –130 (+1) CD30 (AGG→AGC) (Arg→Ser)	β^0	Middle East	4, 13
22)	IVS2 –849 (A→G)	β^0	US Blacks	4, 13
23)	IVS2 –849 (A→C)	β^0	US Blacks	4, 13
24)	IVS2 –850 (G→C)	β^0	Yugoslavian	4, 13
25)	IVS2 –850 (G→A)	β^0	N European	4, 13
26)	IVS2 –850 (G→T)	β^0	Japanese	4, 13
27)	IVS2 –850 (–G)	β^0	Italian	4, 13
Consensus splice sites				
28)	IVS1 –5 (G→C)	β^0	Asian Indian, SE Asian, Melanesian	4, 13
29)	IVS1 –5 (G→T)	β^+	Mediterranean, N European	4, 13
30)	IVS1 –5 (G→A)	β^+	Mediterranean, Algerian	4, 13
31)	IVS1 –6 (T→C)	β^{++}	Mediterranean	4, 13
32)	IVS1 –(–3) CD29 (GGC→GGT)	β^+	Lebanese	4, 13
33)	IVS1 –129 (T→G)	β^+	Saudi Arabian	4, 13
34)	IVS1 –129 (A→G)		German	4, 13
35)	IVS2 –5 (G→C)	β^+	Chinese	4, 13
36)	IVS2 –843 (T→G)	β^+	Algerian	4, 13
37)	IVS2 –844 (C→G)	β^{++} (silent)	Italian	4, 13
38)	IVS2 –844 (C→A)	β^{++} (silent)	Ghanaian	207
39)	IVS2 –848 (C→A)	β^+	UB Blacks, Egyptian, Iranian	4, 13
40)	IVS2 –848 (C→G)	β^+	Japanese	4, 13
Cryptic splice sites in introns				
41)	IVS1 –110 (G→A)	β^+	Mediterranean	4, 13
42)	IVS1 –116 (T→G)	β^0	Mediterranean	4, 13
43)	IVS2 –654 (C→T)	β^0/β^+	Chinese, SE Asians, Japanese	4, 13
44)	IVS2 –705 (T→G)	β^+	Mediterranean	4, 13
45)	IVS2 –745 (C→G)	β^+	Mediterranean	4, 13
46)	IVS2 –837 (T→G)	?	Asian Indian	4, 13
Cryptic splice sites in exons				
47)	CD10 (GCC→GCA)		Asian Indian	4, 13
48)	CD19 (AAC→AGC) Hb Malay (Asn→Ser)	β^{++}	SE Asian	4, 13
49)	CD24 (GGT→GGA)	β^{++}	US Black, Japanese	4, 13
50)	CD26 (GAG→AAG) (Glu→Lys, Hb E)	β^+	SE Asian, European	4, 13
51)	CD26 (GAG→GCG) (Glu→Ala, Hb Tripoli)	β^+	Libyan	208
52)	CD27 (GCC→TCC) (Ala→Ser, Knossos)$^{++}$	β^+	Mediterranean	4, 13
3' UTR				
RNA Cleavage – Poly A signal				
53)	AATAAA→AACAAA	β^{++}	US Blacks	4, 13
54)	AATAAA→AATGAA	β^{++}	Mediterranean	4, 13
55)	AATAAA→AATAGA	β^{++}	Malay	4, 13
56)	AATAAA→AATAAG	β^{++}	Kurd	4, 13
57)	AATAAA→AA–AA	β^+	French, US Blacks	4, 13
58)	AATAAA→A----	β^+	Kurd, UAE	4, 13
59)	AATAAA→AAAAAA	β^+	Tunisian	209
60)	AATAAA→CATAAA	β^{++} (silent)	Chinese	210
61)	AATAAA→GATAAA	β^+	Czechoslovakian, Mediterranean, Yugoslavian, Canadian	211
62)	AATAAA→------	β^+	Nigerian	212

(continued)

Table 5. Continued

Mutation		Type	Distribution	References
Others				
63)	Term CD +6, C→G	β^{++} (silent)	Greek	4
64)	Term CD +90, del 13 bp	β^{+}	Turkish	4
65)	Term CD +47 (C→G	β^{++}	Armenian	4
III. RNA translation				
Initiation codon				
1)	ATG→GTG	β^{0}	Japanese	4
2)	ATG→CTG	β^{0}	Northern Irish	213
3)	ATG→AOG	β^{0}	Yugoslavian	4
4)	ATG→AGG	β^{0}	Chinese	4
5)	ATG→AAG^{+++}	β^{0}	N European	4
6)	ATG→ATC	β^{0}	Japanese	4
7)	ATG→ATA	β^{0}	Italian, Swedish	4
8)	ATG→ATT	β^{0}	Iranian	4
9)	45 bp insertion (−22 to +23)	?	Maori, Polynesian	35
Nonsense codons				
1)	CD6 GAG→TAG,	β^{0}	Brazilian	214
2)	CD7 GAG→TAG	β^{0}	English	4, 13
3)	CD15 TGG→TAG	β^{0}	Asian Indian, Japanese	4, 13
4)	CD15 TGG→TGA	β^{0}	Portuguese, Japanese	4, 13
5)	CD17 AAG→TAG	β^{0}	Chinese, Japanese	4, 13
6)	CD22 GAA→TAA	β^{0}	Reunion Island	4, 13
7)	CD26 GAG→TAG	β^{0}	Thai	4, 13
8)	CD35 TAC→TAA	β^{0}	Thai	4, 13
9)	CD37 TGG→TGA	β^{0}	Saudi Arabian	4, 13
10)	CD39 CAG→TAG	β^{0}	Mediterranean	4, 13
11)	CD43 GAG→TAG	β^{0}	Chinese, Thai	4, 13
12)	CD59 AAG→TAG	β^{0}	Italian American	215
13)	CD61 AAG→TAG	β^{0}	Blacks	4, 13
14)	CD90 GAG→TAG	β^{0}	Japanese	4, 13
15)	CD112 TGT→TGA	β^{0}	Slovenian	4, 13
16)	CD121 GAA→TAA*	β^{0}	Czechoslovakian	4, 13
Frameshift				
1)	CD1 −G	β^{0}	Mediterranean	4, 13
2)	CD2/3/4 (−9 bp, +31 bp)	β^{0}	Algerian	4, 13
3)	CD2−4, 5−9, 7, 10	β^{0}	Algerian	4, 13
4)	CD5−CT	β^{0}	Mediterranean	4, 13
5)	CD6 −A	β^{0}	Mediterranean, US Blacks	4, 13
6)	CD8 −AA	β^{0}	Mediterranean	4, 13
7)	CD8/9 +G	β^{0}	Asian Indian, Japanese	4, 13
8)	CD9 +TA	? β^{0}	Tunisian	209
9)	CD9/10 +T	β^{0}	Greek, Arab	4, 13
10)	CD11 −T	β^{0}	Mexican	4, 13
11)	CD14/15 +G	β^{0}	Chinese	4, 13
12)	CD15 −T	β^{0}	Malay	4, 13
13)	CD15/16 −G	β^{0}	German	4, 13
14)	CD15/16 +G	β^{0}	Chinese	216
15)	CD16 −C	β^{0}	Asian Indian	4, 13
16)	CD22/23/24 −7 bp (−AAGTTGG)	β^{0}	Turkish	4, 13
17)	CD24 −G; +CAC	β^{0}	Egyptian	4, 13
18)	CD24/25 −GGT	?	No additional information	4, 13
19)	CD25/26 +T	β^{0}	Tunisian	4, 13
20)	CD26 +T	β^{0}	Japanese	4, 13
21)	CD27/28 +C	β^{0}	Chinese, Thai	4, 13

Table 5. Continued

Mutation		Type	Distribution	References
22)	CD28 –C	β^0	Egyptian	4, 13
23)	CD28/29 –G	β^0	Japanese, Egyptian	4, 13
24)	CD31 –C	β^0	Chinese	4, 13
25)	CD35 –C	β^0	Malay	4, 13
26)	CD36/37 –T	β^0	Kurd, Iranian	4, 13
27)	CD37/38/39 del 7 bp (–GACCCAG)	β^0	Turkish	4, 13
28)	CD38/39 –C	β^0	Czechoslovakian	4, 13
29)	CD38/39 –CC	β^0	Belgian	4, 13
30)	CD40 –G	β^0	Japanese	4, 13
31)	CD40/41 +T	β^0	Chinese	4, 13
32)	CD41 –C	β^0	Thai	4, 13
33)	CD41/42 –TTCT	β^0	Chinese, SE Asian, Indian	4, 13
34)	CD42/43 +T	β^0	Japanese	4, 13
35)	CD42/43 +G	β^0	Japanese	4, 13
36)	CD44 –C	β^0	Kurdish	4, 13
37)	CD45 –T	β^0	Pakistani	4, 13
38)	CD45 +T	β^0	Turkish	217
39)	CD47 +A	β^0	Surinamese	4, 13
40)	CD47/48 +ATCT	β^0	Asian Indian	4, 13
41)	CD49 –C	β^0	Jordanian	199
42)	CD51 –C	β^0	Hungarian	4, 13
43)	CD53/54 +G	β^0	Japanese	4, 13
44)	CD54 –T	β^0	Swedish	4, 13
45)	CD55 –A	β^0	Asian Indian	218
46)	CD54/55 +A	β^0	Asian Indian	4, 13
47)	CD56–60 +14 bp	β^0	Iranian	4, 13
48)	CD57/58 +C	β^0	Asian Indian	4, 13
49)	CD59 –A	β^0	Italian	4, 13
50)	CD62/63/64 del 7 bp (–TCATGGC)	β^0	Asian Indian	219
51)	CD64 –G[+++]	β^0	Swiss	4, 13
52)	CD67 –TG	β^0	Filipino	4, 13
53)	CD71/72 +T	β^0	Chinese	4, 13
54)	CD71/72 +A	β^0	Chinese	4, 13
55)	CD72/73 –AGTGA, +T	β^0	British	4, 13
56)	CD74/75 –C	β^0	Turkish	4, 13
57)	CD76 GCT → –T	β^0	North African	220
58)	CD76 –C	β^0	Italian	4, 13
59)	CD82/83 –G	β^0	Czech, Azerbaijani	4, 13
60)	CD91–97 (– 22 bp)	β^0	Asian Indian	221
61)	CD93–98 del 6 bp (– CACCTTTG)[+++]	β^0	Japanese	4
62)	CD94/95 +C	β^0	Japanese	4, 13
63)	CD94/95/96 +T	β^0	Japanese	4, 13
64)	CD98 +T	β^0	Asian Indian	4, 13
65)	CD98 –TG	β^0	Japanese	4, 13
66)	CD99/90 –GT	β^0	Japanese	4, 13
67)	CD95 +A	β^0	SE Asian	4, 13
68)	CD106/107 +G	β^0	US Blacks, Egyptian	4, 13
69)	CD109 (GTG → GT–)	?	Irish	213
70)	CD120/121 +A[**]	β^0	Filipino	4, 13
71)	CD130/131 +GCCT	? β^0	German	222
72)	CD142/143 (–CC)	?	French Caucasian	217

* Unlike the majority, some heterozygotes for the CD121 G→T mutation do not have an unusually severe phenotype.

** This frameshift leads to predicted truncated β variant of 138aa with an abnormal carboxy terminal. Heterozygotes do not appear to have an unusually severe phenotype.

*** Occurs in-cis to the IVS2-745 C→G mutation.

+ Also occurs in-cis to 7201-bp deletion involving δ gene.

++ Occurs in-cis to IVS1-5G-A.

+++ Probably de novo.

A recently described deletion, was present in compound heterozygosity with HbS in a patient from Cape Verde Islands affected by mild sickle cell disease (64). Such deletion removes 7.7 kb, starting in IVS2 of the β-globin gene and extending 7.1 kb downstream, within the *Kpn* I family of L1 receptor elements. The other, rather rare deletions, are of particular functional and phenotypic interest. These differ in size, though are all located in the same region (from positions -125 to +78 relative to the mRNA cap site) in the β promoter, which includes the CACCC, CCAAT, and TATA elements. They are associated with unusually high levels of HbA2 and variable increases of HbF in heterozygotes. Codrington et al. (65) proposed that deletion of the β promoter removes competition for the upstream β locus control region (LCR) and limiting transcription factors, thus allowing enhanced expression of the cis δ and γ genes. Carriers of these deletions have a variable and slight increases in HbF,though sufficient to partially compensate for the complete absence of β-globin in homozygotes as suggested by two individuals homozygous for different deletions in this group who are not transfusion dependent with a mild disease (45).

1.4.1.2. Dominantly inherited β-thalassemia (thalassemic structural variants)

In this category are included the hyperunstable β-globin chain structural variants that are synthesized at a normal rate but are so unstable that they are not able to form tetramers, thus resulting in a quantitative deficiency of β-globin chains (1). Unlike the typical β-thal, the hyperunstable β-globin chain variants cause a disease phenotype even in heterozygotes, thus the alternative term, "dominantly inherited β-thal." Apart from the usual features of heterozygous β-thal such as increased levels of HbA2 and decreased β-globin chain synthesis, this group of β-thal is characterized by severe dyserythropoiesis and intraerythroblastic inclusion bodies. Numerous dominantly inherited β-thal alleles associated with similar clinical features have now been identified in families of disparate ethnic origins (1,45). Some are due to missense mutations in which the amino acid substitution is directly responsible for an unstable globin chain or mutants in which deletion or insertion of intact codons disrupts the structure. Others are caused by premature termination codons (nonsense mutations) in exon 3, which fail to undergo nonsense-mediated decay of the RNA. Frameshifts may also result in aberrant splicing producing elongated or truncated β-globin chain variants with abnormal carboxy terminal ends (Table 6).

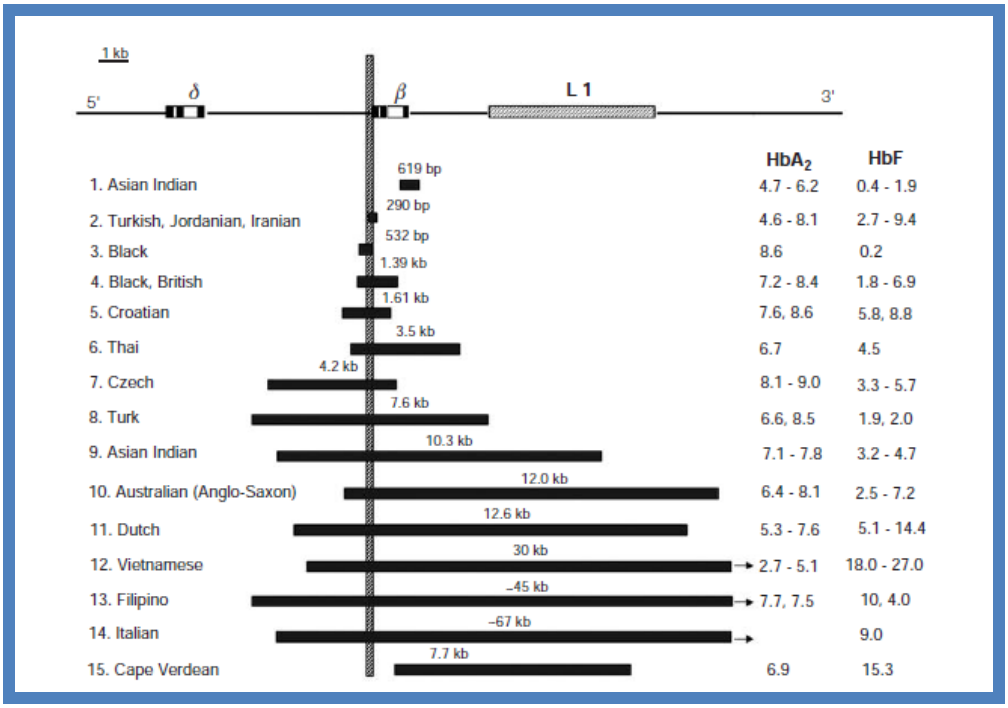

Figure 15. Deletions at the β-globin gene. The L1 hatched box represents the 6.4 *Kpn* repeat 3' of the β-globin gene. The horizontal black boxes represent deletions. The 3' ends of three deletions (Vietnamese, Filipino, and Italian) have not been characterized as indicated by an arrow. The vertical shaded box represents the common region (from positions −175 to +128 relative to the β mRNA CAP sites) lost in all the deletions except for the 619 bp (Asian Indian) and the 7.7 kb (Cape Verdean). The corresponding levels of HbA2 (%) and HbF (%) observed in heterozygotes for the different deletions are shown on the right (45).

Missense mutations

Table 6 shows 12 examples of variant Hb due to missense mutations. Hb Terre Haute (β106 Leu→Arg) is a variant causing β-thal intermedia is 41 This mutant, first described as Hb Indianapolis, was observed in members of two families in whom affected members had mild hemolytic anemia with 2%– 4% reticulocytosis, globin-chain biosynthesis with an α/non-α ratio of approximately 1.0 in bone marrow erythroblasts and of approximately 2.0 in peripheral blood reticulocytes. (67,68).

In the example of Hb Chesterfield, an abnormal peak in the position expected for the β-globin chain variant but without detectable corresponding protein was demonstrated by globin-chain biosynthesis studies (69).

Deletion or insertion of intact codons

Nine examples of deletions or insertions of entire codons allow the reading frame to remain in phase, and the remaining amino acids are normal Table 6). Both Hb Korea and Hb

Gunma have 145 amino acid residues each; in Hb Gunma, the β127-128 Gln-Ala dipeptide is replaced by a Pro residue due to the deletion of three bases (AGG) (70); in Hb Korea, the deletion of three bases (GGT) removes one of the Val residues from codons 32–34 (71).

Other β-globin chain variants have extra residues and include the insertion of Arg in codons 30-31 and insertion of a single proline in codons 124-126. In all cases, no trace of abnormal hemoglobin could be detected by the standard techniques of IEF, high performance liquid chromatography, or heat stability tests. The mechanism that could explain the lack of detection of these structural β-globin chain variants is that the affected amino acids are involved in $\alpha_1\beta_1$ contacts. In the normal β-globin chain, β30 Arg (B12), β33 Val (B15), β34 Val (B16), β108 Asn (G10), β112 Cys (G14), β124 Pro (H2), β125 Pro (H3), β127 Gln (H5), and β128 Ala (H6) are essential for $\alpha_1\beta_1$ dimer formation (43). Deletion or substitution of these critical amino acids would be likely to perturb, or even prevent, the formation of $\alpha_1\beta_1$ subunits and lead to the loss of about half of the β-globin chains.

Another example in this category involved a deletion of 12 nucleotides and an insertion of 6 nucleotides, leading to the substitution of the normal Val-Ala-Gly-Val by Gly- Arg in codons 134-137 and a β-globin subunit that was two amino acids shorter than normal. Affected individuals of this Portuguese family had moderately severe anemia, splenomegaly, and leg ulcers (72). Another example of a hyperunstable β chain variant in this category is Hb Stara Zagnora (codons 137-139, -6 bp) found in a Bulgarian by Efremov (73). A deletions of 6 bp spanning codons 137-139 replaces Val-Ala-Asn with Asp and results in a hyperunstable variant associated with hemolysis and dyserythropoietic anemia with globin chain imbalance (α/β synthesis ratio 1.40); HbA_2 levels were normal.

Premature termination (nonsense mutation)

The GAA→TAA termination codon at codon 121, which leads to the synthesis of a truncated β-globin chain is perhaps the most common dominantly inherited β-thal allele. Significant amounts of mutant β-globin mRNA were found in individuals with such mutations, showing the presence of the truncated β-globin variant has been difficult (62).

Elongated or truncated variants with abnormal carboxy terminal ends

The elongated or truncated β-globin variants in this group have arisen from frameshift mutations generating distal premature termination codons. Elongated β-globin subunits could also arise from aberrant splicing of precursor mRNA. As many as 24 different variants of this category have been described (Table 6).

Table 6. Structural variants associated with a β-thalassemia phenotype

Mutations		Exon	Hb variant	Phenotype	Ethnic group
I. Missense mutations					
1)	CD 28 (CTG→CGG) Leu to Arg	1	Hb Chesterfield*	Thalassemia intermedia, inclusion bodies	English
2)	CD 32 (CTG→CAG) Leu to Glu in cis with CD 98 (GTG→ATG) Val to Met, Hb Köln	2	Hb Medicine Lake*	Severe thalassemia	US Caucasian
3)	CD 60 (GTG→GAG) Val to Glu	2	Hb Cagliari*	Thalassemia intermedia, inclusion bodies	Italian
4)	CD 106 (CTG→CGG) Leu to Arg	3	Hb Terre Haute	Thalassemia intermedia	N European, French
5)	CD 110 (CTG→CCG) Leu to Pro	3	Hb Showa-Yakushiji	Thalassemia intermedia	Japanese
6)	CD 114 (CTG→CCG) Leu to Pro	3	Hb Durham NC/Hb Bressla*	Thalassemia trait/Thalassemia intermedia	US Irish, Italian
7)	CD 115 (GCC→GAC) Ala to Asp	3	Hb Hradec Kralove	Thalassemia intermedia, inclusion bodies	Czech
8)	CD 127 (CAG→CCG) Gln to Pro	3	Hb Houston	Thalassemia intermedia, inclusion bodies	US English
9)	CD 127 (CAG→CGG) Gln to Arg	3	Hb Dhappe	Thalassemia intermedia	French
10)	CD 128 Ala to Pro	3	Hb Mont Saint Aignan	Thalassemia trait, hemolytic anemia	French Caucasian
11)	CD 132 Lys to Gln	3	Hb K Woolwich	Thalassemia trait	West Indian
12)	CD 134 Val to Glu	3	Hb North Shore	Thalassemia trait	Finnish
II. Deletion or insertion of intact codons → destabilization					
1)	CD 3 (+T), CDS (—C) Leu-Thr-Pro to Ser-Asp-Ser	1	Hb Antalya	Thalassemia trait/ thalassemia intermedia	Turkish
2)	CD 30–31 (+CGG) +Arg	1/2		Thalassemia intermedia, inclusion bodies	Spanish
3)	CD 32–34 (—GGT) Val-Val to Val	2	Hb Korea*	Thalassemia intermedia	Korean
4)	CD 33–35 (—TGGTCT) Val-Val-Tyr to 0–0-Asp	2	Hb Dresden	Thalassemia intermedia, hemolysis	German
5)	CD 108–112 (—12bp) Asn-Val-Leu-Val-Cys to Ser	3		Thalassemia trait	Swedish
6)	CD 124–126 (+CCA) +Pro	3		Thalassemia intermedia	Armenian
7)	CD 127–128 (—AGG) Gln-Ala to Pro	3	Hb Gunma	Thalassemia trait/ thalassemia intermedia	Japanese
8)	CD 134–137 (—12, +6) Val-Ala-Gly-Val to Gly-Arg	3		Thalassemia intermedia	Portuguese
9)	CD 137–139 (—TGGCTA) Val-Ala-Asn to Asp	3	Hb Stara Zagnora*	Thalassemia intermedia, hemolysis	Bulgarian
III. Premature termination					
1)	CD 121 (GAA→TAA) Glu to Term (120aa)**	3		Thalassemia trait/ thalassemia intermedia Inclusion bodies	N. European Japanese
2)	CD 127 (CAG→TAG) Gln to Term (127aa)	3		Thalassemia trait/ thalassemia intermedia	English

Table 6. Continued

Mutations		Exon	Hb variant	Phenotype	Ethnic group
IV. Frameshifts or aberrant splicing → elongated or truncated variants with abnormal carboxy terminal					
1)	IVSII: 2,3 (+11, –2)	IVSII		Thalassemia trait/thalassemia intermedia	Iranian
2)	IVSII: 4, 5 (–AG)→aberrant splicing	IVSII		Thalassemia intermedia, inclusion bodies	Portuguese
3)	IVSII: 535 to CD108 (+23, –310, +28) to CD125–108 Leu-Leu-Glu-Asn to Val-Pro-Ser-Val-Thr-Leu-Phe-Phe-Asp	IVSII(Exon 3)	Hb Jambol*	Thalassemia intermedia, hemolysis	Bulgarian
4)	CD 91 (CTG → _CG) → 156aa	2	Hb Morgantown	Thalassemia intermedia	Irish
5)	CD 94 (+TG) → 156aa	2	Hb Agnana*	Thalassemia intermedia, inclusion bodies	S. Italian
6)	CD 100 (–CTT, +TCTGAGAACTT) → 158aa				S. African
7)	CD 104 (AGG→AG) → 156aa***	2		Thalassemia intermedia	German Caucasian
8)	CD 109 (GTG → TG) → 156aa	3	Hb Manhattan	Thalassemia intermedia	Lithuanian
9)	CD 113 (GTG → TG) → 156aa	3		Thalassemia intermedia, inclusion bodies	Canadian, N European
10)	CD 114 (–CT, +G) → 156aa	3	Hb Geneva	Thalassemia intermedia, inclusion bodies	Swiss-French
11)	CD 118 (–T)	3	Hb Saint Sava	Thalassemia intermedia	French Caucasian
12)	CD 120–121 (+a) → 138aa	3		Thalassemia trait	Philippine
13)	CD 123 (–A) → 156aa	3	Hb Makaba	Thalassemia intermedia, inclusion bodies	Japanese
14)	CD 123–125 (–ACCCACC) → 135aa	3	Hb Khon Kaen+	Severe thalassemia intermedia with HbE	Thai
15)	CD 124 (–A) → 156aa	3		Thalassemia intermedia, inclusion bodies	Russian
16)	CD 124–126 (+CCA) → Pro-Pro-Val to Pro-Pro-Val	3		Thalassemia intermedia	Armenian
17)	CD 125 (–A) → 156aa	3		Thalassemia intermedia	Japanese
18)	CD 126 (–T) → 156aa	3	Hb Vercelli*	Thalassemia intermedia, inclusion bodies	N Italian
19)	CD 126–131 (–17 bp) → 132aa	3	Hb Westdale+	Severe thalassemia intermedia with HbE, thalassemia major homozygote	Asian Indian, Pakistani
20)	CD 128–129 (–4, +5, –11) → 153aa	3		Thalassemia intermedia, inclusion bodies	Irish
21)	CD 131–132 (–GA) → 138aa	3		Thalassemia intermedia	Swiss
22)	CD 131–134 (–11 bp) → 134aa	3		Thalassemia intermedia, inclusion bodies	Spanish
23)	CD 140–141 (–C) → 156aa	3	Hb Florida	Thalassemia intermedia, inclusion bodies, hemolysis	Argentinian Spanish

Note: Some of these variants are not associated with elevated A_2 in heterozygous state, for example, Hb Dresden, Hb Jambol, Hb Morgantown, Hb Sfax, Zagora.
* Spontaneous mutations.
** Several families reported including one spontaneous mutation.
*** Coinheritance of extra α-globin gene (genotype $\alpha\alpha\alpha/\alpha\alpha$) contributed to unusually severe thal intermedia in proband.
+ Difficult to evaluate phenotypes of heterozygotes as only homozygote and compound heterozygotes reported.

1.4.2. The α-thalassemia

α-Thal is due to a decreased amount of α-globin chain synthesis. Unlike the β-globin chain, there are no developmental or adult substitutes for the α chain. Since two α globin genes are present in each haploid genome, mutations may occur in one to four alleles. Types and clinical forms of α-thal thus depend on the number of affected genes (see Table 7).

Table 7. Different types of α-Thalassemia (75,76)

Number of functional alleles of the α-globin chain	**Number of nonfunctional alleles of the α-globin chain**	**Name of disease**	**Symptoms**
4	0	Healthy individual	Normal hematological profile
3	1	α-thalassemia silent carrier	- Clinically asymptomatic - Blood test usually normal, hemoglobin normal - Slight changes in size of red blood cells (microcytic) - Slightly lighter color of red blood cell (hypochromic)
2	2	α-thalassemia trait	- Mild anemia - Small red blood cells (microcytic) - Light, pale color of red blood cells (hypochromic)
1	3	Hemoglobin H(HbH) disease	- Moderate to severe anemia - Small red blood cells (microcytic) - Light, pale color of red blood cells (hypochromic) - Mild jaundice - Fatigue - Hepato-splenomegaly - Skeletal deformities (in some cases)
0	4	α-thalassemia major or hemoglobin Barts hydrops fetalis (HbBarts) syndrome	- Intrauterine death of shortly after birth - Skeletal deformities - Cardiovascular problem - Improper brain growth - Enlarged placenta - Hepato-splenomegaly

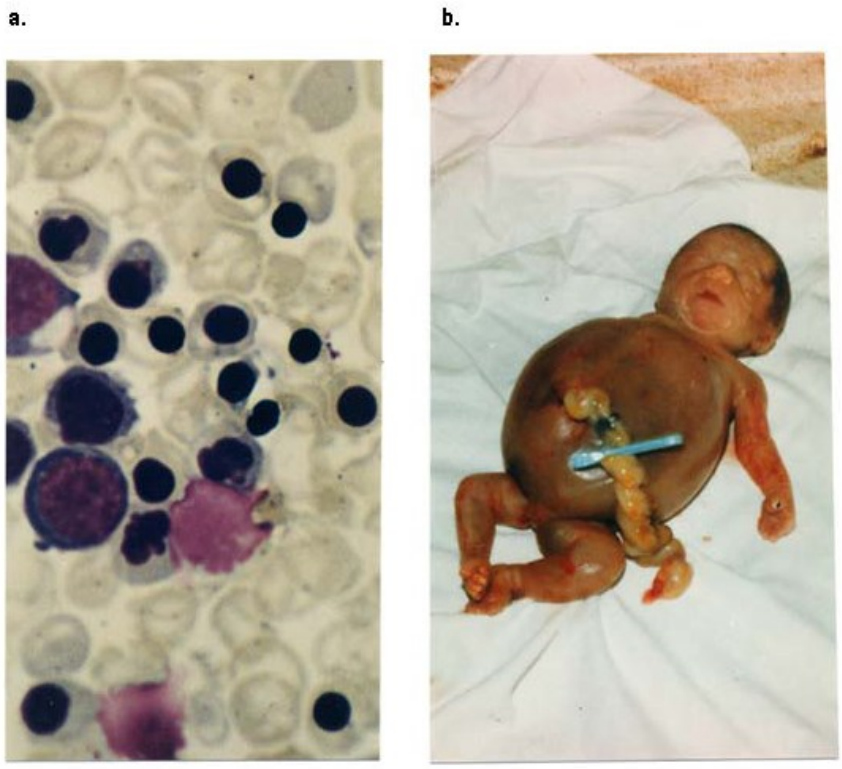

Figure 16. The Haemoglobin Bart's hydrops syndrome. a. peripheral blood film with immature red-cell precursors and hypochromic, microcytic, red cells showing anisocytosis and poikilocytosis; b. stillborn hydropic infant (75)

Mutations

The molecular basis of α-thal is now understood in great detail. Normally, the α-like genes are arranged along chromosome 16 in the order in which they are expressed in development (telomere-ζ2-α2-α1-centromere). Furthermore, it has been described that the cluster lies in a telomeric, gene-rich region of the genome, surrounded by widely expressed genes (see Fig. 7, page 9). Full expression of the α-like genes is critically dependent on the presence of a regulatory element (called HS-40), which lies 40 kb upstream of the cluster (toward the telomere). As many as 50 deletions removing one (-α/) or both (--) genes have been characterized and of these six ($-\alpha^{3.7}$/, $-\alpha^{4.2}$/, --SEA/, --MED/, $-\alpha^{20.5}$/ and --FIL/) represent by far the most common causes of α-thal worldwide. In addition, many different point mutations affecting the structural genes have been identified, these cause the less common nondeletional forms of α-thal ($\alpha^{T}\alpha$/). This information has allowed researchers and hematologists to establish logical and robust screening programs for identifying patients with α-thal. This in turn allows clinicians to provide accurate genetic counseling and prenatal

diagnosis for the severe syndromes of α-thal, including Hb Bart's Hydrops Fetalis and transfusion-dependent forms of HbH disease. When a mutation(s) completely abolishes expression from a chromosome this is called α^{o}-thal and when the mutation(s) only partially downregulate expression from the chromosome this is called α^{+}-thal (1,75)

α^{+}-thalassaemia due to deletions

The α-globin genes are embedded within two highly homologous 4 kb duplication units [61-65]. One very common α-thal deletion is the rightward deletion, a 3.7 kb deletion caused by reciprocal recombination between Z segments producing a chromosome with only one functional α-gene ($-\alpha^{3.7}$ or rightward deletion) causing α-thal and an α-triplication allele without a thalassemic effect. Likewise a reciprocal recombination between mispaired X-boxes results in a 4.2 kb deletion, called leftward deletion ($-\alpha^{4.2}$). An increasing number of deletions resulting in the loss of a single α-gene are reported due to non-homologous recombination events, most of which are rare, or highly region specific.

α^{+}-thalassemia due to non-deletion types of α-thalassemia

α-thal is more frequently caused by deletion than single point mutations or nucleotide insertions and deletions involving the canonical sequences controlling gene expression. In general, the non-deletion α^{+}-thal determinants may give rise to a more severe reduction in α-chain synthesis than the deletion type of chromosomes. Many mutations have been described affecting mRNA processing, mRNA translation, and α-globin stability. Of these the most common non-deletional variants are the αIVSI(-5nt)α (in Mediterraneans), polyadenylation site mutations α2AATAAG, α2AATGAA and α2AATA-- (in the Mediterranean and Middle East)[71-74], termination codon mutations leading to elongated Hb variants, such as Hb Constant Spring (HbCS), Hb Icaria, Hb Koya Dora, Hb Seal Rock and Hb Pakse (middle East, Mediterranean and South East Asia) and structural mutations causing highly unstable α-globin variants; for example, Hb Quong Sze, Hb Suan Dok, Hb Petah Tikvah, Hb Adana, Hb Aghia Sophia . A regularly updated overview is provided by the HbVar web-site

α^{o}-thalassemia due to deletions

The complete or partial deletion of both α-genes in cis results in no α-chain synthesis directed by these chromosomes in vivo. Homozygotes for such deletions have the Hb Bart's Hydrops Foetalis Syndrome. Many deletions were described which remove the ζ- and α-genes and although heterozygotes appear to develop normally, it is unlikely that homozygotes could survive even the early stages of gestation since neither embryonic (ζ2γ2) nor fetal (α2γ2) haemoglobins could be made. Rare deletions causing α^{o}-thal remove the regulatory region,

which lies 40-50 kb upstream of the α-globin gene cluster leaving the α-genes intact. This region composed of four multispecies conserved sequences (MCS), called MCS-R1 to R4, correspond to the previously identified erythroid-specific DNAse1 hypersensitive sites referred to as HS-48, HS-40, HS-33 and HS-10. Of these elements, only MCS-R2 (HS-40), 40 kb upstream from the ζ globin mRNA cap-site has been shown to be essential for α globin expression. Ethnic origin may guide molecular diagnosis. Knowledge of the mutations found in a specific population may allow strategic choice in laboratory diagnostics, especially in selection of the molecular techniques to be applied (75).

In addition to these common forms of α-thal there are many rare and unusual molecular defects that have been identified. These are important because they provide an explanation for patients with hitherto undiagnosed anemia, and they help us to understand how the α cluster is normally regulated in vivo. Rarely, α-thal is caused by deletions that remove the α-globin regulatory element (HS-40). In general these mutations have been observed outside of the "malaria belt," indicating that they are sporadic genetic events that have not been selected during evolution. These natural deletions first indicated the existence of this unexpected form of long-range control of α-globin expression. There are also two rare forms of α-thal that are found in association with a variety of developmental abnormalities, and in particular with mental retardation (so-called α-thal with mental retardation, ATR syndromes). The first group of patients has large (>1 Mb) deletions from the tip of chromosome 16 including the α-globin genes (ATR-16). These usually result from chromosome truncations or translocations and in fact this syndrome provided the first examples in human genetics of subcytogenetic chromosomal translocations, which are now known to underlie many cases of unexplained mental retardation. The second group of patients is now known to have mutations in a trans-acting factor (called ATRX) encoded on the X-chromosome (ATR-X syndrome). These patients have α-thal with profound mental retardation, facial abnormalities, and urogenital anomalies. In this case it is thought that the X-encoded factor regulates expression of many genes, the α genes being but one target. Finally, there is a rare and unexplained form of α-thal that is seen as an acquired mutation in patients with myelodysplasia, hence called the ATMDS syndrome.

These patients inherit a normal complement of α genes (αα/αα) but later in life develop myelodysplasia and presumably acquire a clonal genetic abnormality during the course of their disease. It is interesting that the majority of these patients are elderly males who at some

stage of their disease have abnormal erythropoiesis. It has recently been shown that these patients have acquired mutations in the ATRX gene (1).

1.5. Distribution and prevalence of thalassemic hemoglobinopathies

Just after World War II, independent studies in Italy and in Mediterranean immigrants in the United States highlighted a remarkably high frequency of thalassemia in these populations. This geographical association was in fact responsible for its naming (77,78). After these pioneering observations hundreds of publications demonstrated that thalassemia is a real global health problem (79).

As shown in Fig. 17 and 18, it is clear that Hb disorders are not confined to any particular region, occurring widely across the world. It is established as well that the unusually high frequency of the Hb disorders is the result of natural selection due to the relative resistance of carriers against *Plasmodium falciparum* malaria (77). The unequal distribution of the different thalassemic genes in different countries reflects complex epistatic interactions between the genes for carriers for different varieties of these diseases with respect to the relative resistance of carriers to malaria (78,79). Other factors also help maintain the gene frequency for these conditions at a high level. These include: the practice of consanguineous marriage, the increased maternal age, gene drift and founder effects.

Thalassemic genes have been identified across Southern Europe from Portugal to Spain, Italy and Greece, as well as in a number of central European countries and parts of the former Soviet Union. They also affect the Middle East through to Iran, Pakistan, India, Bangladesh, Thailand, Malaysia, Indonesia and southern China, as well as countries along the north coast of Africa and in South America.

Hemoglobin disorders have spread through the migration of populations from endemic areas to countries where their prevalence in indigenous populations had been extremely low. Such countries include the USA, Canada, Australia, the United Kindom and France.

Although reliable data are still lacking for many regions of the world, recent data indicate that about 7% of the world's population is a carrier of an Hb disorder, and that 300,000 to 500,000 children are born each year with the severe homozygous states of these diseases (76,80).

The Hb variants of the greatest clinical significance are Hb S, C, and E. The Hb S gene has the greatest frequency in West Africa, where approximately 25% of individuals are

heterozygous; the Mediterranean, Caribbean, South and Central American, Arab, and East Indian populations also exhibit high frequencies of the Hb S allele (Fig. 19 and Table 8).

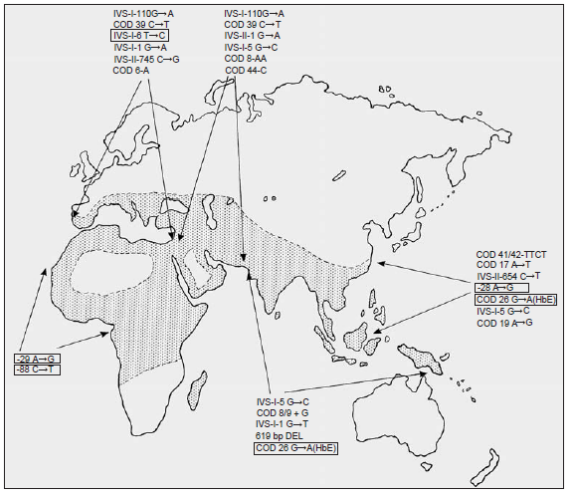

Figure 17. The world distribution and different mutations that cause β-thal. Those shown in boxes are milder mutations. Hb E is included in the milder mutation because of its β-thal phenotype (74).

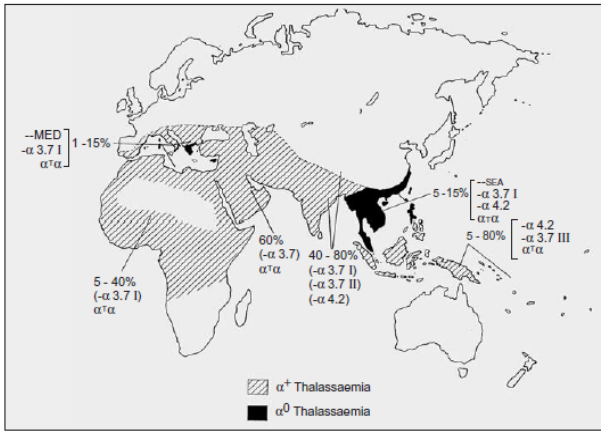

Figure 18. The world distribution of the α-thal. The α^{o}-thal is found at high frequencies in Asia and some Mediterranean countries. The α^{+}-thal have a much broader distribution. Some of the different subtypes of α^{o}- and α^{+}-thal are shown. --SEA indicates the Southeast Asian deletion form of α^{o}-thal $-\alpha^{4.2}$ or $-\alpha^{3.7}$ relate to different forms of α^{+}-thal with their particular deletion sizes, $-\alpha^{3.7}$ is further subdivided into types I, II, and III, again depending on the different deletion sizes. α^{T} denotes nondeletion α-thal (74.)

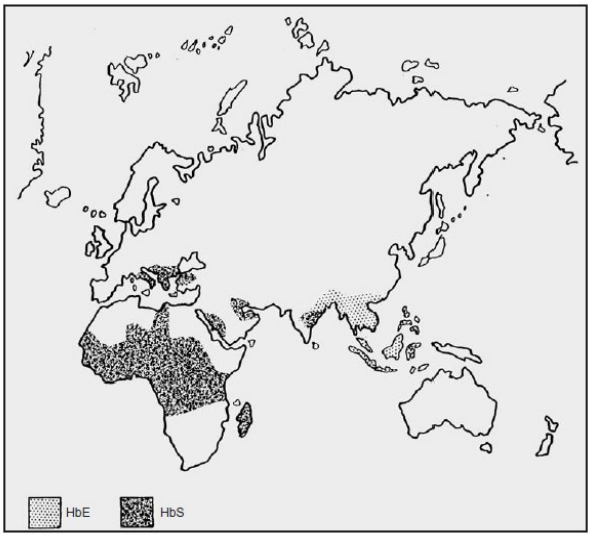

Figure 19. The world distribution of hemoglobins S and E (74).

Table 8. Percent carrier frequencies for common hemoglobin disorders, by WHO region (74,82)

Region	HbS	HbC	HbE	β Thalassemia	α^0 Thalassemia	α^+ Thalassemia
Americas	1–20	0–10	0–20	0–3	0–5	0–40
Eastern Mediterranean	0–60	0–3	0–2	2–18	0–2	1–80
Europe	0–30	0–5	0–20	0–19	1–2	0–12
Southeast Asia Africa	40	0	0–70	0–11	1–30	3–40
Western Pacific	0	0	0	0–13	0	2–60

Other common Hb variants include Hb E and Hb C (Fig. 19 and Table 8). Hb E is the most common structural variant in the world. The distribution of Hb E extends from Eastern India through Southeast Asia and has its highest incidence in Thailand, Laos, and Cambodia. The highest frequency of the Hb E allele was reported in Eastern Thailand near the Vietnamese border, an area termed the ‘‘Hemoglobin E triangle’’, where the carrier rate is as high as 25% to 30%. The Hb E variant also is seen sporadically in parts of China and the Indonesian islands. In areas of the United States that have large populations of immigrants from Southeast Asia, this variant is found more commonly than Hb S in newborn-screening programs. Hb C is restricted to parts of West Africa and is estimated to have a gene frequency in that population of approximately 25%.

1.5.1 Distribution of hemoglobinopathies in Southeast Asia and in Vietnam

Southeast Asia consists of 10 countries with a total population of about 400 million. The ethnic origins of people living in these countries are very heterogeneous. The Mon-Khmer and Tai language-speaking people occupy Thailand, Laos, Cambodia and some parts of Vietnam, Myanmar and Malaysia. The west includes the Burmese (Tibeto-Burman) and the Northeast is the Vietnamese (Austro-Asiatic). The Malayopolynesians (Austronesian) live in Malaysia, Indonesia, Brunei, the Philippines and a number of Pacific island nations. Chinese and Indians are relatively newcomers spread throughout the region.

Hemoglobinopathies are the most common genetic disorders among the people living in Southeast Asia. In Southeast Asia α-thal, β-thal, Hb E and Hb Constant Spring (CS) are prevalent. The gene frequencies of α-thal reach 30-40% in Northern Thailand and Laos, 4.5% in Malaysia and 5% in the remote island of the Philippines, whereas β-thal vary between 1 and 9%. Hb E is the hallmark of Southeast Asia attaining a frequency of 50-60% at the junction of Thailand, Laos, and Cambodia. Hb CS gene frequencies vary between 1 and 8%. These abnormal genes in different combinations lead to over 60 different thalassemia syndromes, making Southeast Asia the locality with the most complex thalassemia genotypes (83).

In Vietnam, both β-thal and Hb E are prevalent and represent one of the most common forms of hemolytic anemia. Three-quarters of the population of more than 80 million are of the Kinh ethnic group, and the remaining comprise over 50 different minority ethnic groups. β-thal and Hb E are distributed in all regions in Vietnam. The carrier rate for β-thal varies between 1.5% and 25% (84). The higher incidence was in the ethnic minority groups, especially Tay (11%) and Moung (25%) in the North and Pako (8.33%) and Catu (14%) in the central region of the country. The prevalence of Hb E in Kinh groups was not much different throughout the country, 1–9%. The highest frequencies of Hb E were in the minority groups, Ede (41.7%), Pako (6%), and Van Kieu (23%) who live in the central plateau. The frequency of β-thal studied in the Kinh people and the Vietnamese refugees in the United States was between 1% and 2%, and Hb E in Saigon was about 3.2% (84,85).

Table 9 estimates that more of 300,000 children are born each year with either thalassemia or sickle cell anemia in the world (81). One unexpected features of these data are the finding that the serious forms of β-thal are almost equally divided between β-thal major and HbE/β-thal as the result of the extremely high frequency of HbE disorder in India, Bangladesh, Myanmar, and many other Southeastern Asia countries. The symptomatic forms

of β-thal, HbH disease, and Hb Bart's hydrops fetalis are also mainly confined to Southeast Asia.

One of the most striking features of these data is the finding that the severe forms of β-thal are almost equally divided between β-thal major and Hb E\β-thal, which reflects the extremely high frequency of the latter disorder throughout Southeast Asia. The symptomatic forms of β-thal, Hb H disease, and Hb Bart's hydrops fetalis are also mainly confined to Southeast Asia.

Table 9. Estimation of the annual number of births with the different Hb disorders (82)

Annual births with major hemoglobin disorders	
β-thalassemia major	22,989
HbE β-thalassemia	19,128
HbH disease	9,568
Hb Bart's hydrops (α°/α°)	5,183
SS disease	217,331
S β-thalassemia	11,074
SC disease	54,736

However, as suggested by Weatherall (79), these results must be reviewed with caution. Since large scale population surveys of the Hb variants were carried out on relatively small samples from limited numbers of centres in individual countries, they have to be considered as a rough approximation of the true global load of the thalassemias. More recent micromapping, which have achieved adequate sample-size estimates from many different parts of an individual country, have shown that there is a remarkable heterogeneity in the distribution of these conditions within populations separated by only short geographical distances (86,87). This is particularly clear from the result showed in Fig. 20, which resumes frequencies and heterogeneity of β-thal mutations as the result of micromapping in Portugal, Spain, France, Italy and the island of Sardinia, Greece, Cyprus, United Kingdom. Similarly, Fig. 21 shows a summary of the α- and β-thal mutations heterogeneity in the Mediterranean area.

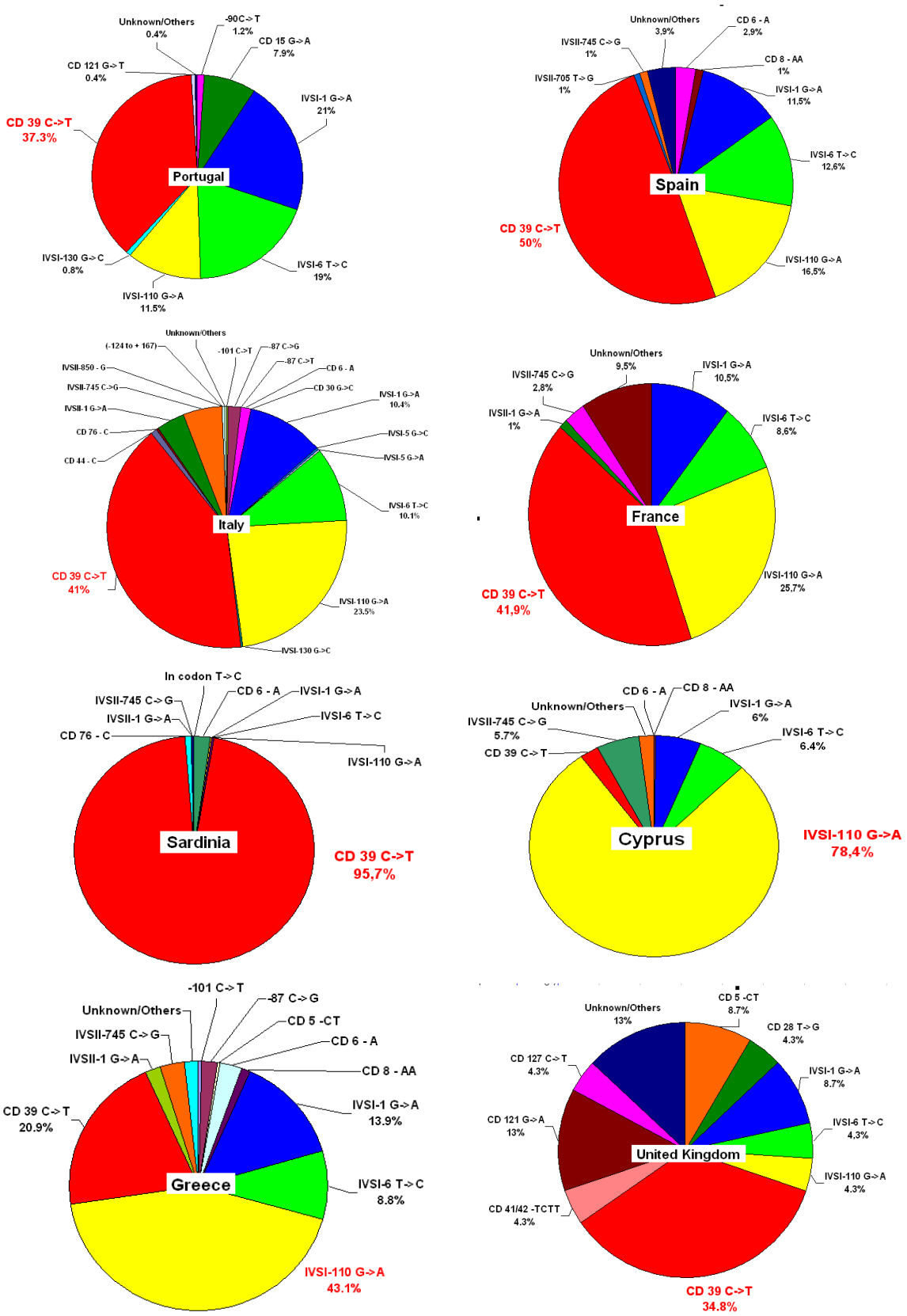

Figure 20. Heterogeneity and prevalence of the different β-thal mutations in eigth European countries. The prevalence of the most common mutation in the country is in red color (88)

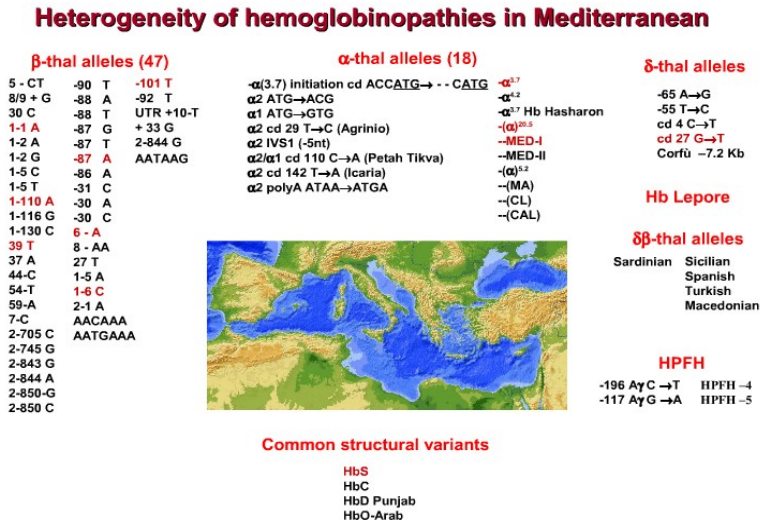

Figure 21. Heterogeneity and prevalence of the different β-thal mutations in Mediterranean.

To provide a better estimate of the global burden of the Hb disorders, in the future it will be therefore necessary to carry out micro-mapping in many of the high frequency regions.

2. RESEARCH OBJECTIVES

In Vietnam, most of researches on hemoglobinopathies were mainly conducted in the North and the South regions. While this research was in progress, Nguyen et al. (89) performed a preliminary study on hemoglobinopathies in Thua Thien Hue Province, the central area of Vietnam, which was restricted to a limited number of pregnant women. Epidemiologic researches of hemoglobinopathies are, in general, not systematic and molecular studies concerning Vietnam are incomplete.

Nam Dong and A Luoi, two mountainous districts of Thua Thien Hue Province have very poor socio-economic conditions and are inhabited mainly by ethnic minorities including Pako, Cotu, Taoi. People in these districts also face many risk factors for hemoglobinopathies such as malaria and consanguineous mating.

With the aim to contribute to a better knowledge of thalassemia defects in Vietnam, with the additional goal to promote future national screening and prenatal diagnosis programs of the severe forms, this research was carried out with following objectives:

1. Estimate the prevalence of hemoglobinopathies in the two mountainous Vietnamese districts of Thua Thien Hue: Nam Dong and A Luoi.

2. Assess the molecular basis of the different mutations and their prevalence in patients with hemoglobinopathies.

3. MATERIALS AND METHODS

3.1. Study design

This consisted with a cross-sectional study that was conducted in the period March 2012 to June 2013.

3.2. Study setting

Thua Thien Hue is a province in the North Central Coast region of Vietnam, approximately in the center of the country. Thua Thien Hue is divided into one city (Hue), two towns (Huong Tra and Huong Thuy) and six districts (two of which mountainous). The research study was conducted in the mountainous districts of Nam Dong and A Luoi.

3.2.1. Nam Dong district

Nam Dong is located in highlands and bounded by Quang Nam province in the south, Huong Thuy in the north, Phu Loc in the east and A Luoi in the west.

Figure 22. Map of the Nam Dong district.

- Area: 661,95 km^2
- Sea level: 500 m
- Population: 24.546 including Cotu, Kinh…
- Capital: Khe Tre

- 10 communes: Huong Giang, Huong Hoa, Huong Huu, Huong Loc, Huong Phu, Huong Son, Thuong Lo, Thuong Long, Thuong Nhat, Thuong Quang.

3.2.2. A Luoi district

A Luoi is located west in the highly moutainous area of A Shau valley bordering Laos

- Area: 1232,70 km^2
- Sea level: 700 m
- Population: 45.546 including Taoi, Pako, Kinh, Cotu…
- Capital: A Luoi
- 20 communes: A Dot, A Ngo, A Roang, Bac Son, Dong Son, Hong Bac, Hong Ha, Hong Kim, Hong Quang, Hong Thai, Hong Thuong, Hong Thuy, Hong Trung, Hong Van, Huong Lam, Huong Nguyen, Huong Phong, Nham, Phu Vinh, Son Thuy

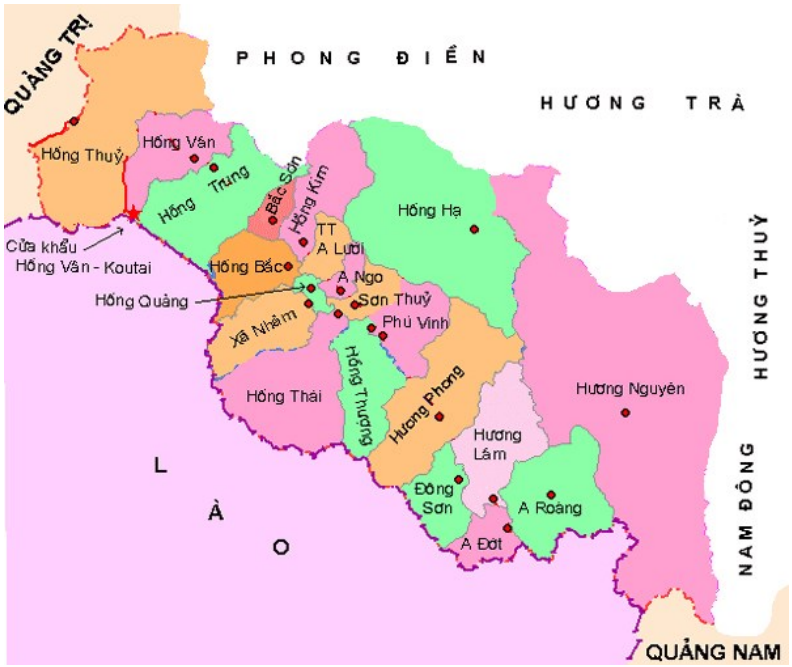

Figure 23. Map of the A Luoi district.

3.3. Population and sample

Sample size was calculated based on the following formula

$$n = Z^2_{\alpha/2} \frac{p(1-p)}{\Delta^2}$$

Where p: prevalence of disease, z: the upper α/2 point of the normal distribution, α: type I error (the error of rejecting a null hypothesis when it is actually true), Δ: one half the width of the desired sample confidence interval.

With **α**= 0.05, $Z_{\alpha/2}$ = 1.96; p = 12%, **Δ** = 0.02, **α** = 0.05, **n = 1015**

We decided to choose **n= 1100**

There are 21 communes in A Luoi and 11 communes in Nam Dong. Sample size in each commune was proportional to its population size and its ethnic distribution. Samples were randomly selected.

Table 10. Population and sample

	A LUOI			NAM DONG		
	Commune	**Population**	**Sample (n)**	**Commune**	**Population**	**Sample (n)**
1	Hong Thuy	2,957	34	Thuong Quang	1,871	38
2	Hong Van	2,723	35	Thuong Nhat	2,053	41
3	Hong Trung	1,911	51	Thuong Long	2,429	47
4	Bac Son	1,125	29	Thuong Lo	1,222	28
5	A Roang	2,613	45	Khe Tre	4,005	31
6	A Dot	2,247	41	Huong Son	1,382	31
7	Dong Son	1,333	20	Huong Phu	3,066	24
8	Huong Lam	1,999	36	Huong Loc	2,264	46
9	Huong Phong	419	4	Huong Huu	2,546	50
10	Phu Vinh	1,026	14	Huong Hoa	2,347	18
11	Hong Thuong	2,192	25	Huong Giang	1,361	31
12	Son Thuy	2,740	41			
13	A Ngo	3,080	48			
14	Nham	2,158	28			
15	Hong Thai	1,455	24			
16	Hong Quang	2,011	27			
17	Hong Bac	1,988	36			
18	Hong Kim	1,796	36			
19	A Luoi	6,965	100			
20	Hong Ha	1,629	22			
21	Huong Nguyen	1179	19			
Total		**45,546**	**725**		**24,546**	**385**

3.4. Methods

3.4.1. Socio-demographic data collection

The questionnaire for all participants were designed by study group. The final questionnaire collected data on demography: age, gender, ethnicity, geography.

3.4.2. Laboratory examination

Whole blood were collected using EDTA as anticoagulant.

3.4.2.1. Complete blood count (CBC)

Red blood cell indices and complete cell analysis were determined using an automated cell counter (Celldyn 3200, Abbott). We used following hematological parameters for the research: RBC ($x10^{12}$/L), Hb (g/dL), MCV (fl), MCH (pg). Anemia cases were diagnosed by Hb concentration.

Table 11. Hemoglobin levels to diagnose anemia (90)

Population	Non -Anaemia*	Anaemia*		
		Mild[3]	Moderate	Severe
Children 6 - 59 months of age	110 or higher	100-109	70-99	lower than 70
Children 5 - 11 years of age	115 or higher	110-114	80-109	lower than 80
Children 12 - 14 years of age	120 or higher	110-119	80-109	lower than 80
Non-pregnant women (15 years of age and above)	120 or higher	110-119	80-109	lower than 80
Pregnant women	110 or higher	100-109	70-99	lower than 70
Men (15 years of age and above)	130 or higher	110-129	80-109	lower than 80

The MCV and MCH levels are usually altered in Hb disorders. Samples with MCV < 80 fl and or MCH < 27 pg were analyzed in the next step.

3.4.2.2. Single tube osmotic fragility (OF)

After pipetting 20 µL of fresh homogeneous EDTA blood into a tube containing 5 mL of a 0.34% and 0,35% NaCl solution, the suspension is gently shaken. In normal conditions, the suspension should become clear in few minutes (OF test negative). The continued presence of turbidity after 5 min indicates reduced OF (or elevated osmotic resistance) and thalassemia (OF test positive)(91).

3.4.2.3. Hb electrophoresis

Samples with one of following abnormal haematological parameters were analyzed by electrophoretic techniques:

- Anemia
- MCV < 80 fl
- MCH < 27 pg
- OF test positive

Isoelectric focusing (IEF)

This technique separates proteins according to their isoelectric points (pI). A stable pH gradient is established in the gel by the addition of appropriate ampholytes. A protein mixture is placed in a well on the gel. With an applied electric field, proteins enter the gel and migrate until each reaches a pH equivalent to its pI. This is because of when pH = pI, the net charge of a protein is zero (Fig. 24).

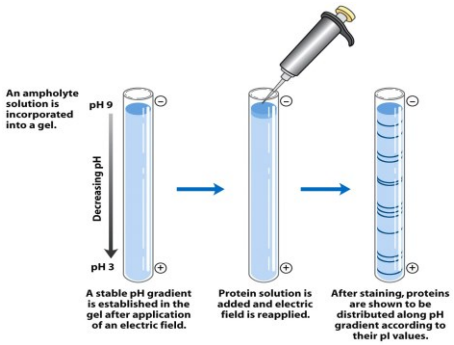

Figure 24. The principle of isoelectric focusing.

Immobilized pH gradient gels are acrylamide gel matrix co-polymerized with a pH gradient, which result in completely stable gradients except the most alkaline (>12) pH values. The immobilized pH gradient is obtained by the continuous change in the ratio of carrier ampholites.

Ampholytes are weak acid or base defined by its pK value. All of the amino acids found in proteins are ampholytes because they contain a carboxyl group that acts as an acid and an amino group that acts as a base. They are used in IEF after co-polymerization with acrylamide. They set up a stable pH gradient through the electrophoretic medium, and components being separated equilibrate to a point in the medium corresponding to their isoelectric point.

IEF is the most useful, and economical, technique for the separation of normal and abnormal hemoglobins (Fig. 25)(92).

In this work, IEF was as follows: peripheral blood samples were collected in EDTA as anticoagulant and washed three times in 0.15 mol/1 NaCI by centrifugation at 1500 X g. Washed red cells were lysed with 7.7 mmol/1 K.CN in water and adjusted to a final concentration of 6-8 g/1 of Hb. Higher concentrations caused poor separations of Hb fractions in the IEF system. 125 X 125 X 0.5 mm poly- acrylamide gels, 15 wells capacity, or 125 X 250 X 0.5 mm, 30-33 wells capacity, were cast onto silanized glass plates. 5% polyacrylamide gels were prepared by mixing 1.67 ml of 8.55 mol/1 acrylamide-0.06 mol/1 bisacrylamide stock solution, 622 µl Pharmalyte, pH 6.7-7.7 range (Pharmacia, Uppsala, Sweden), 50 /il Pharmalyte, pH 3.5-10, and 7.67 ml water. Polymerization was achieved by the addition of 16 µl TEMED and 14 JU.I of 1.75 mol/1 ammonium persulfate. Anodic and cathodic solutions were 40 mmol/1 glutamic acid and 1 mol/1 NaOH, respectively. Gels were prerun for 45 min at 400 V. Lysates were applied at the cathodal site in a volume of approximately 5 /il in slots precast in the gel, and electrophoresis was for 90 min at 1600 V constant voltage on the Consort E 3800 apparatus. Multiplier cell (Pharmacia, Uppsala, Sweden) thermostated at 4°C. Electrophoretic migration was performed at 4°C with the refrigerated bath circulator Wise Circu® WCR – P6 (Wisd – Daihan Scientific). Gels were than fixed and stained.

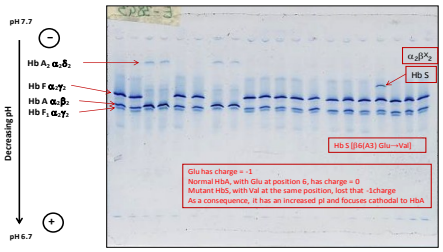

Figure 25. Isoelectric focusing of hemoglobins of newborns and adults.

Acid Urea Triton PolyAcrylamide Gel Electrophoresis (AUT-PAGE)

It is a variant of the SDS polyacrilamide gel electrophoresis, thus separating on the basis of molecular weight, charge and, additionally, of the properties of the –R group. Since Triton X-100 is a nonionic surfactant, it denatures native proteins to unfolded individual

polypeptides by deleting hydrophobic interactions whereas mass and the net charge remains unchanged. It is largely used in the detection of Hb variants though it is not useful for variants due to charged amino acid substitutions (93).

The method consists in a vertical polyacrylamide gel polymerized in the presence of 8M urea with migration in a acetic acid solution and the presence of Triton X-100, a non-ionic detergent which masks some charged residues.

In this work, globin chains were separated by electrophoresison 12% polyacrylamide gels, containing 6 M urea and 2% Triton X-100, in 5% acetic acid. The gel solution consisted of 5 ml of 60:0.4% acrylamide: bis-acrylamidc (Eastman Kodak Company); 1-25 ml of glacial acetic acid; 18-75 ml of 8 M urea; 0.125 ml of Temed (NNN'N'-tetra-methylcchylcncdiaminc); 0.5 ml of Triton X-100; and 15 mg of ammonium persulphate. The gel mixture was degassed by vacuum and then used for the gel.

Slab gels were 12 cm long, 16 cm wide and 1.5 mm thick. The well-former had 12 spaces. Gelling occurred within 30 min, but the gels were left for 2 h prior to use. The electrophoresis buffer was 5% acetic acid. The gels were pre-electrophoresed for 60 min at 200 V with the anode at the top. The current fell from 60 mA to 40 mA. The anode buffer was removed, the gels were overlayed with 1 M β-mercaptoethylamine.

Fresh electrophoresis buffer was added and a second pre-electrophorcsis performed for 45-60 min at 150 V. The current fell from 30 to 20 mA. The β-mercaptoethylamine was then removed entirely. Sample buffer consisted of 5 ml of 8 M urea, 0.5 ml of glacial acetic acid, 0.5 ml of 2-mercaptoethanol and 2 mg of pyronin Y. All procedures were performed at room temperature.

The gels were stained for 30 min in 0.5% Coomassie Brilliant Blue, in 7% acetic acid, 30% methanol, and destained in the acetic acid-methanol by diffusion.

AUT-PAGE was performed with Protean® II xi Cell and Power Supply HV (Biorad).

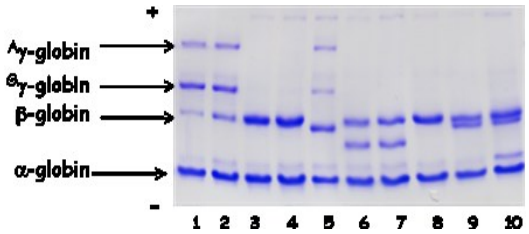

Figure 26. Separation of dissociated globin chains from newborn and adult lysates.

3.3.2.4. High Perfomance Liquid Chromatography (HPLC)

High-performance liquid chromatography (sometimes referred to as high-pressure liquid chromatography), HPLC, is a chromatographic technique used to separate a mixture of compounds with the purpose of identifying, quantifying or purifying the individual components of the mixture. HPLC has many uses including medical (e.g. detecting vitamin D levels in blood serum), legal (e.g. detecting performance enhancement drugs in urine), research (e.g. separating the components of a complex biological sample, or of similar synthetic chemicals from each other), and manufacturing (e.g. during the production process of pharmaceutical and biological products). Several methods are used to separate normal and abnormal Hbs.

As shown in Fig. 3.6, the stationary phase surface displays ionic functional groups (R-X) that interact with analyte ions of opposite charge.

This type of chromatography is further subdivided into cation exchange chromatography and anion exchange chromatography. The ionic compound consisting of the cationic species M^+ and the anionic species B^- can be retained by the stationary phase.

Cation exchange (CE) chromatography retains positively charged cations because the stationary phase displays a negatively charged functional group:

$$R\text{-}X^-C^+ + M^+B^- \leftrightarrow R\text{-}X\text{-}M^+ + C+ + B^-$$

Anion exchange (AE) chromatography retains anions using positively charged functional group:

$$R\text{-}X^+A^- + M^+B^- \leftrightarrow R\text{-}X^+B^- + M^+ + A^-$$

Note that the ion strength of either C^+ or A^- in the mobile phase can be adjusted to shift the equilibrium position and thus retention time.

The abnormal samples identified by means of electrophoresis were analyzed by two different HPLC techniques.

Cation Exchange – HPLC (CE-HPLC)

Ion-exchange chromatography (or ion chromatography) is a process that exploits differences in the sign and magnitude of the net electric charges of proteins at a given pH. It can be used for almost any kind of charged molecule including large proteins, small nucleotides and amino acids. It is often used in protein purification, water analysis, and quality control.

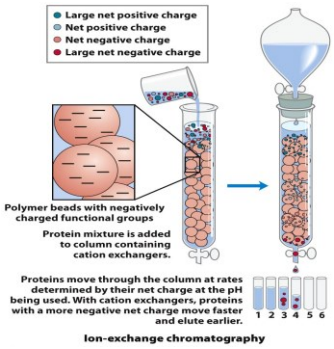

Figure 27. The principle of chromatography. It is shown a typical chromatogram obtained with an anion exchange column.

In this study, the Bio-Rad Variant®, a fully automated HPLC system largely used in the search for hemoglobinopathies, was used. It employs the β Thalassemia Short program, designed to separate and determine in 5–6 min the area percentages for HbA_2 and HbF and to provide qualitative determinations of abnormal Hbs. Windows for retention times have been established for presumptive identification of the most commonly occurring Hb variants. In Fig. 26 it is shown the chromatogram obtained separating Hb of a carrier of β-thal. Note that the quality of Hb is as in normal, the only difference being the higher level of the HbA_2.

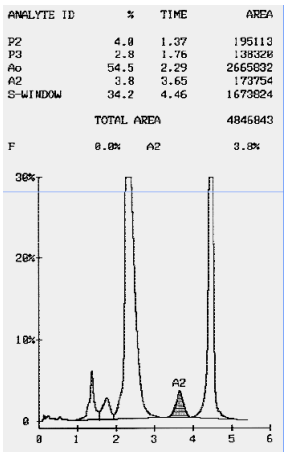

Figure 28. The Variant® CE-HPLC chromatogram. Lysate was from a heterozygous β-thal with increased HbA_2.

Reversed phase-HPLC (RP-HPLC)

In the reversed phase-HPLC a non-polar stationary phase and an aqueous, moderately polar mobile phase are used. One common stationary phase is a silica which has been surface-modified with R-Me2SiCl, where R is a straight chain alkyl group such as C18H37, C8H17, or C4H9. With such stationary phases, retention time is longer for molecules which are less polar, while polar molecules elute more readily (early in the analysis).

It is possible to increase retention times by adding more water to the mobile phase; thereby making the affinity of the hydrophobic analyte for the hydrophobic stationary phase stronger relative to the now more hydrophilic mobile phase. Similarly, it is possible to decrease retention time by adding more organic solvent to the eluent. RP-HPLC is so commonly used that it is often incorrectly referred to as "HPLC" without further specification.

In this research, RP-HPLC was performed on a Vydac C_4 column with a linear gradient of acetonitrile:water (from 20 to 60%) as described by Masala and Manca (94).

Blood samples collected in anticoagulant are diluted with 4-5 volumes of saline (0.9% NaCl) and washed three times by centrifugation. Lysates are prepared by mixing a volume of the washed red cells with an equal volume of distilled water and 0.4 volumes CCl_4, with occasional stirring. The mixture is centrifuged at 20,000 g for 10 min at 4°, and the clear Hb solution is collected using a pipette. An optimal separation occurs when a sample containing 80-120 μg globin is injected.

Developer A is consistent with 200 ml Acetonitrile and 1 ml TFA, and water to 1000 ml; Developer B contains 600 ml Acetonitrile and 1 ml TFA, and water to 1000 ml.

The equipment consists of two pumps programmed with a controller, the Vydac C4 column, an injector, a recorder, and a UV detector. The basic procedure is useful as a primary test on normal or abnormal adult and newborn lysates. The controller is programmed in order to create an 80 min linear gradient from 50 to 60% developer B (with a corresponding decrease in developer A from 50 to 40%) at the 1 ml/min flow rate. All procedures are carried out at room temperature.

As shown in Fig. 29, the normal elution pattern shows the appearance of the most hydrophilic heme group, followed by a pre-β peak, by the α chain of Hb A, by the $^{G}\gamma$, the $^{A}\gamma^{T}$ and the $^{A}\gamma$ chain of the Hb F. The $^{A}\gamma^{T}$ globin is the product of an allele of the $^{A}\gamma^{I}$ gene. Due to a point mutation, codon 75 may code for the Ile ($^{A}\gamma^{I}$) or for Thr ($^{A}\gamma^{T}$).

Variant globin chains having a substitution with a more hydrophobic amino acid elutes with a higher retention time. Globin having a substitution with a more hydrophilic amino acid elutes with a shorter retention time.

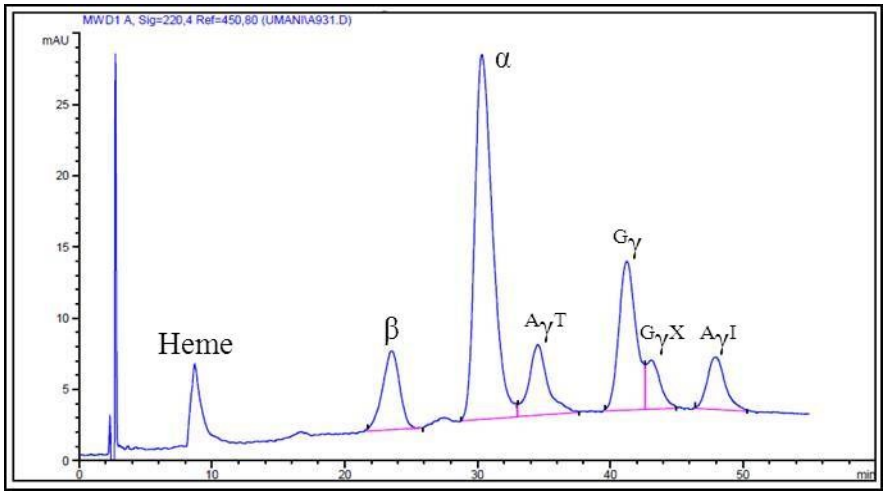

Figure 29. Separation of the globin chains of a newborn baby by RP-HPLC. Note the presence of a variant $^{G}\gamma$ gene.

3.4.3. Detection of mutations affecting globin genes

Selected samples indicating the presence of β-thal gene(s) or variant Hbs, or both, were analysed in order to identify the affected gene and the mutation.

DNA was extracted by the salting out method. This method involves salting out of the cellular proteins by dehydration and precipitation with a saturated NaCl solution.

Buffy coats of nucleated cells obtained from anticoagulated blood (EDTA) were resuspended in 15 ml polypropylene centrifugation tubes with 8 ml of nuclei lysis buffer. The cell lysates were digested overnight at 37°C with 0.1 ml of 10% SDS and 0.5 ml of a protease K solution (1 mg protease K in 1% SDS and 2 mM Na2EDTA). After digestion was complete, 1 ml of saturated NaCl (approximately 6M) was added to each tube and shaken vigorously for 15 seconds, followed by centrifugation at 3000 rpm for 15 minutes. The precipitated protein pellet was left at the bottom of the tube and the supernatant containing the DNA was transferred to another 15 ml polypropylene tube. Exactly 1 volume of room temperature absolute isopropanol was added and the tubes inverted several times until the DNA precipitated. The precipitated DNA strands were removed with a plastic spatula or pipette and transferred to a 1.5 ml microcentrifuge tube containing 100-200 µl TE buffer . The DNA was allowed to dissolve 2 hours at 37°C before quantitating. (95).

DNA was quantified with NanoDrop 8000 Spectrophotometer (Thermo Scientific) and DNA stored at -20^{0} C.

Selected samples were amplified for target globin gene or portion of it (α, β, γ genes) and sequenced to determine gene mutations.

3.4.3.1. *Amplification and isolation of α1 and α2 genes*

Selected amplifications of the entire α2 (2112 bp in length) and α1 loci (1753 bp) were performed using the α1α2 and α3α4 oligonucleotides. The PCR was carried out in a 50 µl reaction volume containing 200 ng genomic DNA, 0.3 µM of each primer, 375 µM dNTPs, 2.0 mM MgCl2, 5% of DMSO, 0.07% of β-mercaptoethanol, 1 M betaine, 2.5 units Taq polymerase, in a 1X buffer (10X stock: 166 mM (NH4)2SO4, 670 mM TrisCl pH 8.8, 670 µM Na2EDTA, 1.7 mg/ml BSA).

The amplification reaction included 35 cycles of denaturation at 95°C for 1 minute, annealing at 60.6°C or 57°C (α2 and α1, respectively) for 1 minute, and extension at 72°C for 3 minutes.

Primers were the follows:

PRIMER	Position	LOCUS	SEQUENCE (5'→ 3')
α1	32677- 32697	α2	AGCCTGGCCAAACCATCACTT
α2	34770- 34788	α2	TTCTGAGCCACTGCCTGCA
α3	36656- 36676	α1	GCTTTGTTTACCTGTTTAACC
α4	38388- 38407	α1	AAAAACTCAGGCACACACAG

3.4.3.2. *Amplification and isolation of Gγ and Aγ genes*

Complete Gγ and Aγ genes were amplified with the same chemical and thermal protocol, except for the primer pairs. For sequence reactions nested PCR from these fragments were needed.

Chemical protocol: 1X Buffer, 2 mM MgCl2, 375 µM dNTPs, 0.3 µM of each primer, 1M Betaine, 0.2 µg of genomic DNA, 2.5 U of polymerase. Thermal file: 3' at 95°C, 35 cycles of 1' at 95°C, 1' at 52°C and 3' at 72°C.

Primers were the follows:

PRIMER	Position	LOCUS	SEQUENCE (5'→ 3')
R159	38792- 38817	Aγ	TGAAACTGTGGTCTTTATGAAAATTG
R160	33855- 33880	Gγ	GCACTGAAACTGTTGCTTTATAGGAT
R161	34507- 34530		
	39443- 39466	Gγ Aγ	GGCGTCTGGACTAGGAGCTTATTG

3.4.3.3. *Amplification and isolation of β gene*

The complete β gene was directly amplified from genomic DNA with 3 different PCRs, each with its own chemical and thermal protocol.

-P1-P2 primers amplify a 706 bp region (-140 to +566) encompassing the proximal promoter to the beginning of IVS2.

Chemical Protocol: 1X Buffer, 3.5 mM MgCl2, 250 μM dNTPs, 0.3 μM of each primer, 0.3 μg of genomic DNA, 2.5 U of polymerase. Thermal file: 3' at 95°C; 35 cycles of 1' at 95°C, 45'' at 65°C and 1' at 72°C; additional final extension was added (4' at 74°C).

-β7-β8 primers amplify a 923 bp region (+333 to +1255) including part of exon 2 and IVS2. This fragment contains also the Ava II polymorphism for Orkin Haplotypes (96).

Chemical Protocol: 1X Buffer, 3 mM MgCl2, 250 μM dNTPs, 0.3 μM of each primer, 0.3 μg of genomic DNA, 2.5 U of polymerase. Thermal file: 3' at 95°C; 35 cycles of 1' at 95°C, 1' at 60°C and 1' at 72°C; additional final extension was added (4' at 74°C).

-β9-β10 primers amplify a 956 bp region (+799 to +1764) from IVS2 to 5'UTR of β-gene.

Chemical Protocol: 1X Buffer, 1.5 mM MgCl2, 250 μM dNTPs, 0.3 μM of each primer, 0.3 μg of genomic DNA, 2.5 U of polymerase. Thermal file: 3' at 95°C; 35 cycles of 1' at 95°C, 1' at 55°C and 1' at 72°C; additional final extension was added (4' at 74°C).

Primers were the follows:

PRIMER	Position	LOCUS	SEQUENCE (5'→ 3')
β7	62469- 62489	β	TCCTGATGCTGTTATGGGCAA
β8	63371-63391	β	AAAAGCAGAATGGTAGCTGGA
β9	62935- 62955	β	AAAAACTTTACACAGTCTGCC
β10	63881- 63900	β	ATTAGCTGTTTGCAGCCTCA

3.4.3.4. *Detection of Small Nucleotides Polymorphisms* (*SNPs) and microsatellites on the β-cluster*

To determin Orkin Haplotypes (96) configurations PCRs were set for each SNP or microsatellite.

All the PCR contained 1X Buffer, 250 μM dNTPs, 0.3 μM of each primer, 0.3 μg of genomic DNA, 2.5 U of polymerase. MgCl2 was used at different concentrations (from 1.5 to 4 mM). Thermal file was identical for all the PCR except for the annealing temperature (Ta): 5' at 95°C; 35 cycles of 1' at 95°C, 1' at Ta and 1' at 72°C; additional final extension was added (4' at 74°C). MgCl2 and annealing temperature can be found below.

-HS2_F-HS2_R [3.5 mM MgCl2, Ta=57°C]: these primers amplify a fragment that contains the microsatellite (AT)xNy(AT)z in β-LCR-HS2.

-5pEpsilon_F-5pEpsilon_R [2.5 mM MgCl2, Ta=51°C]: this pair isolate a region upstream to the ε gene in which a Hinc II restriction (GTCAAC) site is present.

-pseBeta_F-pseBeta_R [1.5 mM MgCl2, Ta=55°C] primers allowed to identify the Hinc II restriction site (GTTGAC) on the ψβ1 gene.

-3pseBeta_F-3pseBeta_R [3.5 − 4 mM MgCl2, Ta=55°C] isolate a region downstream the ψβ1 gene in which a Hinc II (GTTAAC) restriction site can be present.

-proBeta_F-proBeta_R [3.5 mM MgCl2, Ta=51°C] primer pair amplify a portion of the distal promoter region of β-globin gene where (AT)xTy motif is present.

-3pBetaBam_F-3pBetaBam_R [2.5 mM MgCl2, Ta=51°C] locate downstream the β-globin gene, in which a Bam HI restriction site (GGATCC) is present.

3.4.3.5. *PCR fragment purification and quantification*

Following amplification, all the reaction volume was electrophoresed through a 1-1.2% agarose, 1X TAE, etidium bromurated gel at 7.5 volts/cm for 45 hour in presence of a molecular weight marker (Marker III, VI and 100 bp, Roche). Bands were visualized on an UV transilluminator, quantified by comparison with the marker and excised with sterile scalpel.

DNA was recovered from agarose by means of the Montage Gel Extraction Kit (Millipore).

3.4.3.6. *Sequence Reaction*

The sequence reaction was performed by the BigDye Terminator v3.1 Cycle Sequencing Kit (Life Technologies). This kit is based on Sanger chain-terminating method and each terminator or dideoxyribonucleotide (ddNTP) is associated with a specific fluorochrome. Given that all 4 ddNTPs have different fluorochrome, the reaction can be performed in a single tube.

The 20 µl reaction volume mixture consisted of: 2 µl of 5X Buffer, 2 µl of Mix, 3.2 pmol of primer, 1-2 ng of PCR product per 100 bp.

Sequence primers were almost always the same ones used in PCR; sometimes, because of the length of the fragment or other necessities, specific sequence primers were designed.

Thermal file: 1’ at 96°C, 25 cycles of 10’’ at 96°C, 5’’ at 50°C and 4’ at 60°.

Excess of unincorporated dye terminators and other contaminants (unincorporated primers, salts, and so on) were completely removed prior to electrophoresis through Sigma Spin Post-Reaction Clean-Up Columns (Sigma).

Purified sample was mixed with 2 volumes of formamide and injected on a 3130 Genetic Analyzer (Life Technologies) capillary electrophoresis.

3.4.3.7. *Sequence analysis*

Sequence analysis was performed with the help of two software: Geospiza FinchTV (http://www.geospiza.com) and Another Plasmid Editor or APE (http://biologylabs.utah.edu/jorgensen/wayned/ape/).

FinchTV allowed to read and edit sequence files, while APE was used to align the obtained sequences with the reference β-globin cluster (AC #: U01317) and α-globin cluster (AC #: NG_000006).

3.4.3.8. *Multiplex Ligation-dependent Probe Amplification (MLPA)*

MLPA is a multiplex PCR method detecting abnormal copy numbers of up to 50 different genomic DNA or RNA sequences, which is able to distinguish sequences differing in only one nucleotide (97).

MLPA probes hybridise to the target, are legated together to form a ssDNA and are amplified by a single pair PCR primers. The resulting amplification produces marked fragments from 130 and 480 nt in length and can be analyzed by capillary electrophoresis. Comparing the peak pattern obtained to that of reference samples indicates which sequences show aberrant copy numbers (http://www.mrc-holland.com).

MLPA P140HBA probemix (MRC-Holland) is designed to detect copy number changes of 24 different sequences in the α globin genes region. In addition, the probemix contains one probe that detects the presence of the Constant Spring mutation.

In this study the MLPA reaction was performed with 100 ng of DNA and followed the provided protocol, with a 20 h hybridization.

Data were analyzed with the Coffalyser.Net MLPA data analysis software available at the manufacturer's website.

3.5. Ethical issue

The participants were explained about the research, its benefits and risks and asked to sign a consent to participate.

Study protocols were approved by Hue University of Medicine and Pharmacy Institutional Review Board. The participants did not have to pay for this research. The results of the research was confidential. Participants with hemoglobinopathies received counselling about the risk factors and where to get medical support.

3.6. Statistical method

Descriptive statistics including mean (±sd), proportions were calculated. Chi-square test and Fisher's exact test were used to compare proportions and independent t-test was used to compare means of the continuous variables. Gene frequencies were calculated by using Hardy-Weinberg law. All reported confidence interval were two-sided 95% confidence intervals and p-values <0.05 were regarded as statistically significant. All analyses were performed using SAS 9.3 and Stata 12.0

4. RESULTS

4.1. Demographic characteristics of the studied population

Table 12. General demographic characteristics

Ethnic	Male	Female	Total
Kinh	66 (6%)	113 (10.27%)	179 (16.27%)
Pako	164 (14.91%)	206 (18.73%)	370 (33.64%)
Taoi	86 (7.82%)	84 (7.64%)	170 (15.45%)
Cotu	178 (16.18%)	200 (18.18%)	378 (34.36%)
Pahy	0	1 (0.09%)	1 (0.09%)
Muong	0	1 (0.09%)	1 (0.09%)
Thanh	0	1 (0.09%)	1 (0.09%)
Total	**494 (44.91%)**	**606 (55.09%)**	**1100 (100%)**

Among 1100 participants, 494 (44.91%) were males and 606 (55.09%) were females; only 16,27% were Kinh and ethnic minorities, mainly the Cotu, the Pako, and the Taoi representing the remaining (83.73%) of the population (Table 12).

Table 13. Demographic characteristics of each district

Ethnic	A LUOI			NAM DONG		
	Male	Female	Total	Male	Female	Total
Kinh	39 (5.45%)	61 (8.53%)	100 (13.99%)	27 (7.01%)	52 (13.51%)	79 (20.52%)
Pako	164 (22.94%)	205 (28.67%)	369 (51.61%)		1 (0.26%)	1 (0.26%)
Taoi	86 (12.03%)	83 (11.61%)	169 (23.64%)		1 (0.26%)	1 (0.26%)
Cotu	36 (5.03%)	38 (5.31%)	74 (10.35%)	142 (36.38%)	162 (42.08%)	304 (78.96%)
Pahy		1 (0.14%)	1 (0.14%)			
Muong		1 (0.14%)	1 (0.14%)			
Thanh		1 (0.14%)	1 (0.14%)			
Total	**325 (45.45%)**	**390 (54.55%)**	**715 (100%)**	**169 (43.9%)**	**216 (56.1%)**	**385 (100%)**

Of 715 participants in A Luoi, there are 45.45% males, 54.55% females. The Kinh ethnic group accounts for 13.99%, three main minority ethnic groups - the Pako, the Taoi and the Cotu account for the remaining (Table 14).

In Nam Dong, there are 43.9% males, 56.1% females in a total of 385 participants. The Kinh ethnic group is 20.52%, the main minority ethnic group is the Cotu.

Table 14. Age distribution

Age (year)	Male	Female	Total
2-5	23	20	43
6-14	72	66	138
15-25	104	138	242
26-40	147	231	378
41-60	85	99	184
>60	63	52	115
Total	**494**	**606**	**1100**

Age distribution is similar between genders. Average age is 32.7 ± 18.84, and range is from 2 years old to 102 years old.

4.2. CBC and OF characteristics

Tables 15 to 17 show haematological parameter and anemic cases found.

Table 15. Mean value of hematological parameters

Parameters	Mean	SD
RBC	4.86	0.55
Hb	13.34	1.48
MCV	84.26	7.81
MCH	27.65	2.99

Table 16. Anemia rate of study group

	Male	Female	Total
Anemics	76 (6.91%)	104 (9.45%)	180 (16.36%)
Normals	418 (38%)	502 (45.64%)	920 (83.64%)
Total	**494 (44.91%)**	**606 (55.09%)**	**1100 (100%)**

Anemia prevalence resulted 16.36%.

Table 17. Anemia rate of the Kinh and the minority

	Kinh	**Minority**
Anemics	30 (16.76%)	150 (16.29%)
Normals	149 (83.24%)	771 (83.71%)
Total	**179 (100%)**	**921 (100%)**

Similarly, there is no difference in anemia prevalence between the Kinh (16.76%) and the minority (16.29%) ($p > 0.05$).

731 samples which had one of the following signs: anemia, MCV < 80, MCH < 27, OF positive were screened by electrophoresis techniques.

4.3. Electrophoresis and HPLC results

Of the 731 abnormal samples defined above, 136 cases were diagnosed as being the carriers of hemoglobinopathies by the combination of methods: IEF, AUT-PAGE, CE-HPLC, RP-HPLC, including: Hb E, β-thalassemia and Hb Constant Spring (α thalassemia).

Table 18. Hb E prevalence

HbE	**n**	**%**
Homozygous	3	0.27
Heterozygous	111	10.09
No	986	89.64
Total	**1100**	**100**

Two heterozygotes were found as being double heterozygotes Hb E/β thala. Hb E prevalence is 10.36% (10.09% heterozygous and 0.27% homozygous)

4.4. Distribution of hemoglobinopathies among the Kinh and the minorities populations

Results obtained indicate that the prevalence of the Hb E gene is 5 times higher in the minorities with respect to the Kinh population (Table 19).

Table 19. Hb E prevalence between the Kinh and the minority group

HbE	**Kinh**	**Minorities**
	5 out of 174 (2.79%)	109 out of 812 (11.83%)

Hb E prevalence in the ethnic minority group is statistically higher than that in the Kinh group with $p < 0.0001$: 11.83% versus 2.79%.

Similarly, as shown in Table 20, Hb CS seems to be present only in the minority group. However, as already underlined, it is possible that the separation methods adopted in this work did not allowed the identification of all the carriers of the elongated globin chain. Thus, the difference between 2 groups is not statistically significant with $p > 0.05$.

Table 20. Hb CS prevalence between the Kinh and the minority group

Hb CS	Kinh	Minorities
	0	5 out of 921 (0.54%)

4.5. The molecular basis of the identified hemoglobinopathies

Selected samples were analysed DNA to confirm genotype and describe type of mutation. β gene sequencing was carried out for Hb E and β-thalassemia cases. For Hb CS cases, the gene was α2.

4.5.1. Hb E

A total of six samples showing Hb E by the IEF, CE- and RP-HPLC were submitted to sequencing of the β-globin genes. Three resulted β^E/β, one β^E/β^E, and two β^E/β^o. Biochemicals aspects of a β^E/β^E subject are showed in Figg. 30 and 31, whereas that of a β^E/β^o are in Figg. 32 and 33.

Sequencing confirmed that the G→A missense mutation at codon 26 (*G*AG→*A*AG) of the β globin gene, corresponding to the Glu→Lys amino acid substitution, is responsible for Hb E (Fig. 34).

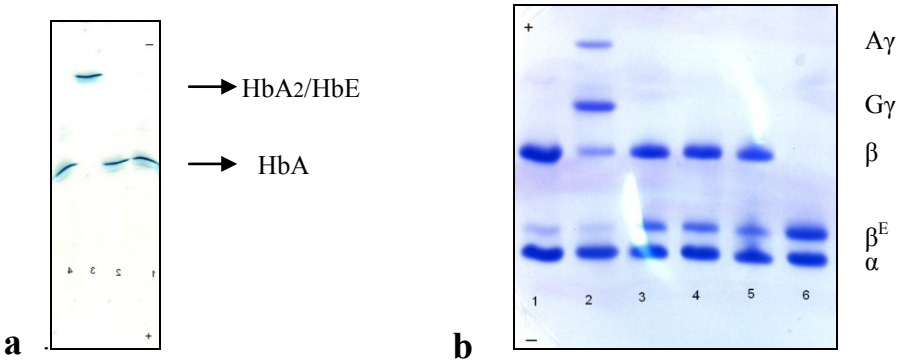

Figure 30. Electrophoresis of the lysate of a β^E homozygote. The patient is in lane 3 of panel a) and in lane 6 of panel b).

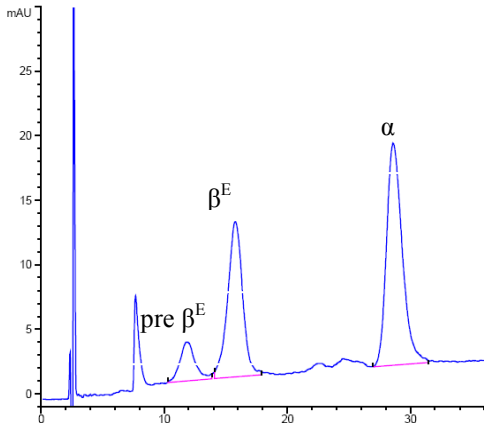

Figure 31. RP-HPLC. Separation of the globin chains of the homozygous β^E/β^E.

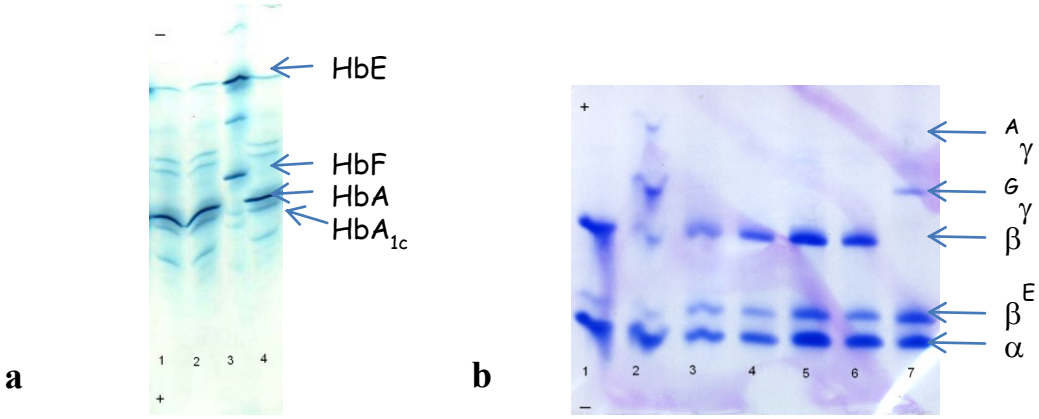

Figure 32. Electrophoresis of the lysate of a β^E/β^o compound heterozygote. Panel a. IEF. Lanes 1, 2, 4: normal adults. Panel b. AUT-PAGE. Lane 1: normal adult; lane 2: normal newborn; lane 3: β^E/β^o; lanes 3,4,5,6: β^E/β, lane 7: β^E/β^o.

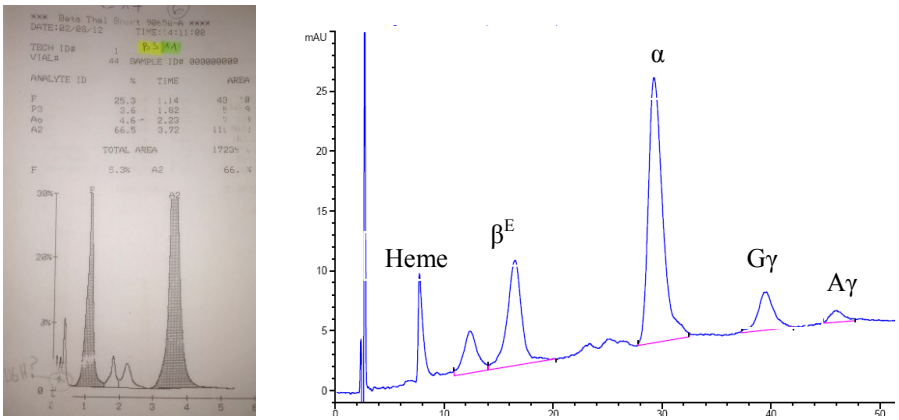

Figure 33. HPLC separation of Hb and globin chains of the β^E/β^o compound heterozygote. Panel a): Cation exchange HPLC showing the absence of Hb A and the presence of Hb F and

Hb E. Panel b): Reversed-phase HPLC showing the absence of β chains and the presence of the β^E globin and of the two different γ chains of Hb F.

The <u>G</u>AG→<u>A</u>AG mutation also affects the β-globin gene expression creating an alternate splicing site in the precursor mRNA at codons 25-27 of the β-globin gene. Thus, the formation of functional β^E-mRNA is decreased due to abnormal alternative splicing of precursor β-mRNA at a site 5' to the IVS-I. As the result, the Hb E heterozygote is mildly affected and the homozygosity is a benign disorder with a mild β-globin chain deficit which is comparable to that seen in a β^o-thal heterozygote. However, compound heterozygotes are often severely affected despite an increased production of Hb F scan be observed. Moreover, this variant has a weak interaction between α and β globin, causing instability when there is a high amount of oxidant (1,12).

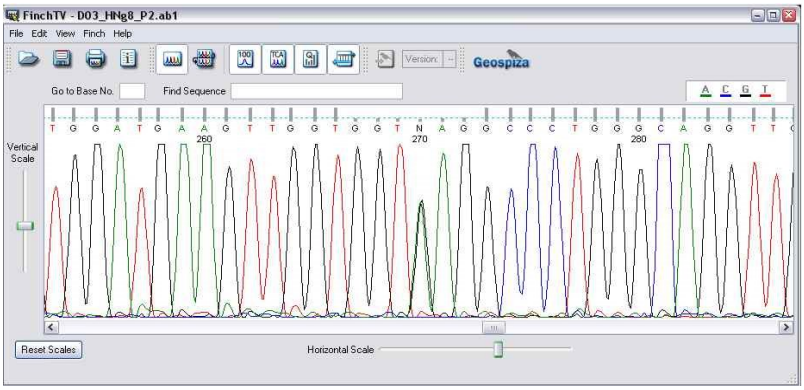

Figure 34. Sequencing of the β-genes of an homozigous for the Hb E. It is shown the: <u>G</u>AG→<u>A</u>AG at Codon 26 corresponding to the Glu→Lys substitution.

Patients having the β^E/β^E and β^E/β^o genotype were clinically similar to patients with β-thal of intermediate severity as described elsewhere (1).

4.5.2. β-thalassemia

5 cases were analyzed by the β-globin gene sequencing. Three were heterozygous and were identified thanks to the minor Hb A2 level determination by CE-HPLC (Fig. 35) whereas two were identified as β^o/β^E compound heterozygotes.

Four different mutations giving rise to the β^o-thal (i.e. absent β-globin chain synthesis under the control of the affected gene) were found and are listed in Table 21.

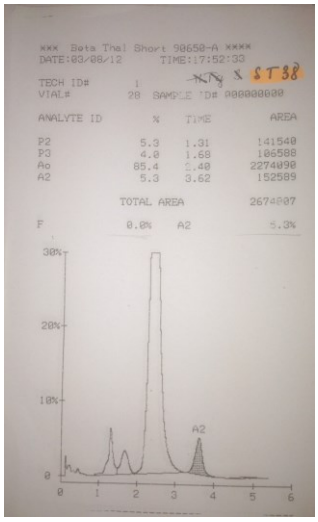

Figure 35. The typical CE chromatogram of the hemoglobins of a β-thal heterozygote. It is evident the clearly increased level of the Hb A_2.

Table 21. β^o-thal mutations

Sample	β-Genotypes	β^o-thal mutation
AL19	β^E/β^0	cod 17 AAG→TAG (Lys→Stop)
ST38	β^0/β	cod 17 AAG→TAG (Lys→Stop)
*BS11	β^E/β^0	IVS I-1 (G→T)
HV10	β^0/β	–TTCT cod. 41/42
HV2	β^0/β	+G cod 14/15

*The subject was homozygous for the C→T substitution in the promoter region of the $^G\gamma$ gene, at position -158 with respect to the Cap site (also known as the *Xmn*I polymorphism).

Two β-globin genes carried the AAG→TAG non-sense mutation at codon 17 (Lys→Stop) which is responsible for the premature interruption of the mRNA translation (Codon 17 (A→T); AAG(Lys)→TAG(stop codon) β^o, or β 17(A14) Lys→Stop (HGVS name: HBB:c.52A>T). This mutation was to first to be described at the molecular level in Chinese people in 1979 (98).

One patient with Hb E was also the carrier of the G→T substitution at the level of the invariant "donor" GT dinucleotides at the first nucleotide of the IVS I, which is essential for the splicing of introns. This patient was also the carrier of the Hb E mutation (IVS-I-1 (G→T); AG^GTTGGT→AGTTTGGT β^o, or β nt 143 G→T) (HGVS name: HBB:c.92+1G>T) (99). His genotype was thus β^0/β^E. Hematological parameters (RBC $5.35x10^{12}$/L, Hb 9.3 g/dL, MCV 59 fl, MCH 17.4 pg) and the presence of 25.3% of Hb F

prompted the sequence of the γ-globin genes to be analysed. The C→T substitution, at position -158 with respect to the Cap site, in the promoter region of both the $^{G}\gamma$ gene (also known as the *Xmn*I polymorphism) was found. This is a further evidence of an increased Hb F synthesis, during erythropoietic stress, under the control of this mutation (44).

One patient was the carrier of the 4 bp (-TTCT) deletion at the level of codons 41/42 which is responsible for a frameshift mutation resulting in a stop codon at the new codon 59 terminating translation (TTCTTT(Phe-Phe)→TT β^{o}, (Codons 41/42 (-TTCT); TTCTTT(Phe-Phe)→----TT β^{o}, or β 41-42 (-TTCT); modified C-terminal sequence) (HGVS name: HBB:c.126_129delCTTT). It is a typical β^{o}-thal heterozygote with modest anemia, microcytosis, and hypochromia, usually transfusion-dependent in homozygotes (100).

Finally, one subject was found to be heterozygote for the insertion of a nt (+G) at codons 14/15 (CTG TGG, Leu;Trp)→ CTG G TGG. The insertion modifies the C-terminal sequence: (15)Val-Gly-Gln-Gly-Glu-Arg-(21)Gly-COOH creating a stop codon at position 22 (HGVS name: HBB:c.45_46insG) (101)

4.5.3. Hb Constant Spring

By the two electrophoretic techniques and the two HPLC, 5 subjects were identified as having small amounts of an abnormal Hb migrating close to the minor Hb A_2 by IEF and CE-HPLC (Figg. 36 and 37).

Sequencing of the α-globin genes of four subjects showing the (possible) Hb Constant Spring showed the expected mutation at the level of α_2 gene at codon 142: *T*AA (stop codon) →*C*AA, (α_2 142, Stop→Gln; modified C-terminal sequence: (142)Gln-Ala-Gly-Ala-Ser-Val-Ala-Val-Pro-Pro-Ala-Arg-Trp-Ala-Ser-Gln-Arg-Ala-Leu-Leu-Pro-Ser-Leu-His-Arg-Pro-Phe-Leu-Val-Phe-(172)Glu-COOH) (HGVS name: HBA2:c.427T>C). This T→C mutation thus results in an extension of the α globin chain with 31 amino acid residues; a stop codon is found at the new codon 173. The globin chain is synthesized at a lower rate (102).

Hematologic parameters of a homozygote were: RBC $4.71x10^{12}$/L, Hb 12 g/dL, MCV 82 fl, MCH 24.9 pg, Hb A 94.6%, Hb F 0.4%, Hb A_2 1%, Hb CS 4%.

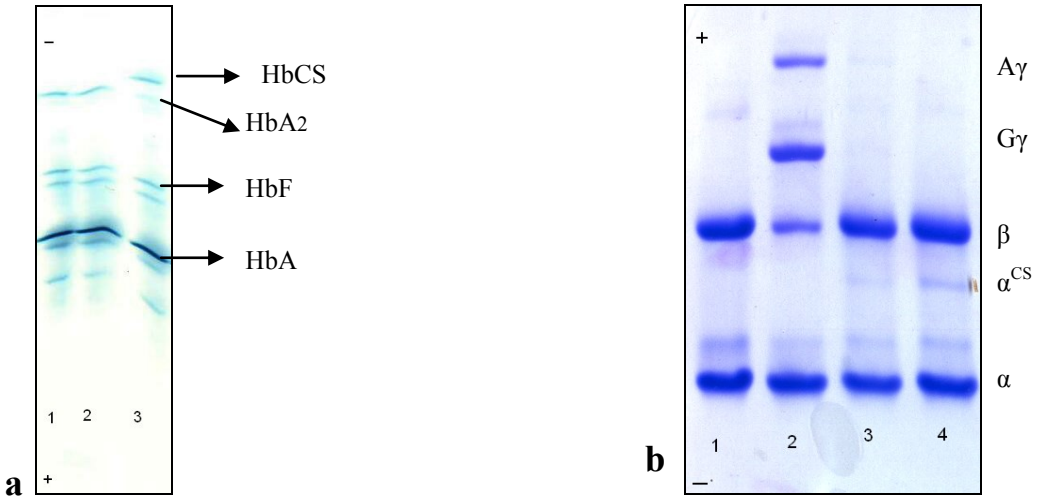

Figure 36. Electrophoresis of a minor Hb variant. Panel a): IEF. Lanes 1 and 2: normal adults; lane 3: $\alpha^{CS}\alpha/\alpha^{CS}\alpha$. Panel b): AUT-PAGE. Lane 1: normal adult; lane 2: normal newborn; lane 3 $\alpha^{CS}\alpha/\alpha\alpha$ and lane 4: $\alpha^{CS}\alpha/\alpha^{CS}$.

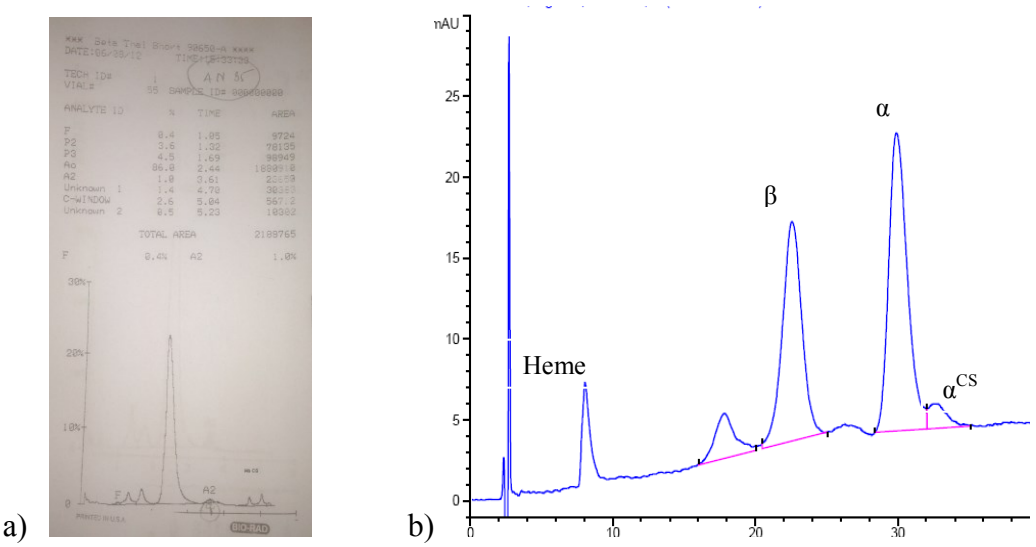

Figure 37. HPLC of the lysate of a subject with the $\alpha^{CS}\alpha/\alpha^{CS}\alpha$ genotype. Panel a): CE-HPLC of the Hb tetramers. Panel b): RP-HPLC of globin chains.

4.5.4. Haplotyping of the β-globin genes cluster

The assessment of the linkage of β-thal and Hb E mutations and β-globin gene polymorphisms with DNA polymorphisms in β-globin gene cluster after Orkin et al. (96) was possible in seven samples (Table 22).

Table 22. Haplotyping of the β-globin genes cluster

Sample and mutation	Hinc II 5' ε	Hind III Gγ	Hind III Aγ	Hinc II ψβ	Hinc II 3' ψβ	Ava II β	Bam HI 3' β	Orkin's haplotypes
1. β^E/β^E	-	+	-	+	-	+	+	recombinant
	+	-	-	-	-	+	+	I
2. β^E/β	-/-	+/+	-/-	+/+	-/-	+/+	+/+	recombinant
3. β^E/β	-	+	-	+	-	+	+	recombinant
	+	-	-	-	-	-	+	VII
4. β^E/β^0 IVS I-1	-/-	+/+	-/-	+/+	-/-	+/+	+/+	recombinant
5. β^0/β +G cod 14/15	+	-	-	-	-	+	+	I
	+	-	-	-	-	-	+	VII
6. β^0/β –TTCT cod. 41/42	+/+	-/-	-/-	-/-	-/-	+/+	+/+	I/I
7. β^0/β Stop at cod 17	+	-	-	-	-	+	+	I
	-	+	+	-	-	-	+	VI

Result of Table 22 suggests that β-globin gene cluster carrying the β^E mutation is rather polymorphic since at least two different haplotypes cohexist in this limited sample: the haplotype I as described by Orkin et al. (96) (found in one chromosome) and another which may be the result of recombinations at the level of polymorphic sites of the haplotype I (four chromosomes). The same conclusion was made by others (103) in six Vietnamese families living in South Carolina (USA) and in unselected peoples originating from Cambodia, Laos and Thailand (104) indicating a multiple origins of the β^E-globin gene mutation that contributed to its high frequency in South East Asia.

By analysis of restriction enzyme polymorphism this study obtained the expected evidence for the existence of at least four haplotypes: I, VI and VII and a recombinant, in cis to the four different mutations of Table 22.

5. DISCUSSION

From March, 2012 to July, 2012, 1100 people (715 from A Luoi and 385 from Nam Dong) were enrolled. This is the largest community study on hemoglobinopathy in Central Vietnam. Recently, Nguyen et al. (89) conducted a similar study among pregnant women in Thua Thien Hue with a smaller scale (n = 410). In 2010, another study was conducted in South Vietnam by O'Riordan et al. (108) with a sample size of approximately 9000. Other studies were mainly carried out in patients in clinics with a small sample size or in limited families living abroad.

In this study, 16.27% of participants were of the Kinh ethnicity and 83.73% belong to minority groups. This proportion corresponds with the demographic characteristics of the area under study although the Kinh make up 80-85% of the population in Vietnam. Three main ethnic minorities of the sample are Cotu (34.36%), Pako (33.64%) and Taoi (15.45%). The sample was representative for the community in Nam Dong and A Luoi.

The anemia prevalence resulted 16.36%, which is relatively high in comparison with the prevalence found in other regions. Beside hemoglobinopathy, common causes of anemia are iron deficiency and parasite (hookworm). No difference was found in anemia prevalence by gender or ethnicity.

MCV, MCH are 2 important parameters in hemoglobinopathy. MCV and MCH are low in patients with hypochromic microcytic anemia as is the case of hemoglobinopathies.

Osmotic fragility (OF) test was the first method used for screening of thalassemia and was introduced as a simple approach to detect carriers by Silvestroni and Bianco in the 1940s. This fast and simple method has been applied as a screening test in large populations (105).

Because of time and financial limits in this study Hb concentration, MCV, MCH and OF was helpful as the first step of screening. On the basis of those parameters, as many as 731 samples were continuously screened by electrophoresis techniques and then HPLC and sequencing of the globin genes.

The quite high 10.36% prevalence of the Hb E thalassemic variant was found which confirms it is the most common cause of hemoglobinopathies in South East Asia. The already discussed financial limitations did not allowed to performe CE-HPLC chromatography as a routine test so that many carriers of the low level Hb Constant Spring escaped from the observation in the electrophoretic methods. Thus, the 0.45% prevalence of the elongated

thalassemic variant is certainly largely underestimated as demonstrated by the observatyion of four homozygotes versus only one heterozygote.

For the same reasons it was no possible to calculate the prevalence of β-thal because the corrected diagnosis is possible only after the determination of the minor Hb A_2 which is the only parameter which allow the discriminanation of β-thal and α-thal. Few cases were identified because of in double heterozygosity with Hb E and other were selected because under treatment in the local hospital.

Hb E, the most common Hb variant among Southeast Asian populations results from a G to A substitution at codon 26 (*G*AG→*A*AG) in exon 1 of the β-globin gene. Hb E confers a survival advantage against *Plasmodium falciparum* and could be the logical explanation for its high global prevalence. It is the most prevalent abnormal hemoglobin in South East Asia with its frequency approaching 60% in Northeast regions of Thailand, Laos and Cambodia. Significant numbers were reported from other Asia countries such as Srilanka, North Eastern India, Bangladesh, Nepal, Vietnam and Malaysia. Additionally, population transmigration led to its emergence in United States and Canada (106, 107). The 10.36% prevalence obtained in this study is also relatively high and there are two main reasons to explain this prevalence: malaria is still in circulation in Nam Dong and A Luoi Districts and consanguineous marriage largely occurs in the minority groups.

Hb E occurs both in homozygous (EE) and heterozygous (EA or E trait) states and may co-inherits with α and β-thal, Hb S, Hb C and other hemoglobin variants. Hb EE and E trait are mild disorders and are associated with either mild or no anemia. In contrast, Hb E/β-thalassemia displayed a remarkable variability in its clinical severity from mildly asymptomatic state to a severe transfusion dependent anemia. This is because of the major mechanism underlying the pathophysiology of Hb E/β-thal is consistent with the inadequate or absent β-globin chain production.

The Hb E is a form of β^+ phenotype and the remainder of the 4 mutations account for β^0 phenotype in affected individuals. The Hb E frequency was highest. In our study, the β^E gene frequency was 0.0532 (0.0608 in the minority groups and 0.0014 in the Kinh group). The difference between the Kinh and the minority was significant.

Table 23 compares the β^E gene frequency described in South Vietnam, Central Vietnam (89, 108) and in this study. The frequency in the Kinh group obtained in this study is similar to that described in South Vietnam and to that described by Nguyen et al. (89) in pregnant women.

Table 23. Comparison of β^E gene frequency in South Vietnam and Central Vietnam

Population		**β^E gene frequency**
South Vietnam (O'Riordan S et al) (108)	Kinh	0.017
	Dao	0
	Tay	0.014
	Nung	0.010
	S'Tieng	0.356
	M'Nong	0.240
	Raclay	0.144
	Ede	0.314
Central Vietnam (Nguyen et al) (89)	Pregnant women (86.8% Kinh, 13.2% Cotu and Taoi)	0.017
Central Vietnam (this study)	Kinh	0.014
	Minority (Taoi, Cotu, Pako)	0.0608

Some discrepancies is seen in comparison with that found in Dao, Tay, Nung but lower than that of the S'Tieng, the M'Nong, the Raclay, the Ede (108).

As a whole, the β^E gene frequency found in this study is lower than that in Southeast Asian populations (Table 24).

Table 24. Comparison of β^E gene frequency in Southeast Asian populations

Population		**β^E gene frequency**
Eight Southeast Asian populations (Laig M et al) (109)	Karachi	0.507
	Ahom	0.342
	Northern Thailand	0.048
	Tak	0.102
	Khonkhaen	0.263
	Ubol Rachathani	0.287
	Cambodia	0.229
	So (adults)	0.497
Central Vietnam (this study)	Kinh	0.014
	Minority (Taoi, Cotu, Pako)	0.0608

Besides Hb E (which has the phonotype of a mild β^+-thal), four different mutations at the β-globin gene were found, all giving rise to the absence of β-globin production (β^o-thal): the <u>A</u>AG→<u>T</u>AG nonsense mutation at codon 17 (98), the G→T transversion at nt 1 of the

IVS-I (99), the four bp deletion (-TTCT) at codons 41/42 (100), and G insertion at codons 14/15 (101).

Table 25. Comparison of spectrum of β-thal mutations in Vietnam, South China and Thailand

Mutations	**Central Vietnam** (this study)	**North Vietnam** (110)	**South Vietnam** (84,111)	**Thailand** (84, 111)	**South China** (84, 111)
Codon 26 (GAG→AAG)	+	+	+	+	+
Codon 17 (AAG→TAG)	+	+	+	+	+
Codons 41/42(-TTCT)	+	+	+	+	+
IVS-I-1 (G→T)	+	-	+	+	-
Codon 14/15(+G)	+	-	-	+	-
Codon 95 (+A)	-	+	+	-	-
Codons 71/72 (+A)	-	+	+	+	+
IVS-II-654 (C→T)	-	-	+	+	+
-28 (A→G)	-	-	+	+	+
Codon 19 (A→G)	-	-	-	+	-
Codon 35 (C→A)	-	-	-	+	-
-86 (C→G)	-	-	-	+	-
IVS-I-5 (G→C)	-	-	-	-	+

There common mutations were found in all there regions of Vietnam, Thailand and China were Codon 26 (GAG→AAG), codon 17 (AAG→TAG) and codon 41/42(-TTCT). The common β-thalassemia in Southeast Asia have mutations of codon 17 (AAG→TAG), 41/42 (-TTCT), 71/72 (+A) and IVS-I-1(G→T). Moreover, one of the most common point mutations in the β-globin gene is known as Hb E (β codon 26, GAG→AAG, Glu→Lys) (112). Previous studies in different areas of Thailand have shown that 70-90% of β-thalassemia mutations are either frameshift codon 41/42(-TTCT) or nonsense codon 17 (AAG→TAG), both causing β^0-thalassemia (113).

The frameshift mutation at codon 41/42(-TTCT) is the most common mutation in South and Southeast China where its frequency reaches above 40%. The nonsense mutation at codon 17 (AAG→TAG) has a high prevalence in Southern Chinese provinces as well as in North Thailand (110). These three mutations also had highest frequencies in North and South Vietnam (84,110,111). Two mutations found in North Vietnam (codon 95, +A) and codons

71/72, +A) and two found in South Vietnam (IVS-II-654 (C→T), -28 (A→G) were not observed in this study. The mutation of IVS-I-1 (G→T) was found in Central and South Vietnam but not found in North Vietnam.

Mutation at codons 14/15 (insertion of G), was unobserved previously in Vietnam. In 1988, Chan et al. (101) detected a new frameshift due to an insertion of G between codons 14/15 of the β-globin gene in 2 unrelated Chinese patients with Cooley's anemia. As shown in Table 25, this mutation was found in this study and in Thailand and it is not common in β-thal. patients. This insertion of a nucleotide (+G) between codons 14 and 15 of the β-globin gene causes a frameshift mutation that results in premature termination 20 bp downstream. This accounts for the β^0 phenotype in affected individual. In our study, the patients who gets this mutation is a Kinh male, 45 years old at A Luoi district, genotype β^0/ β with Hb A2 4.6%.

In this research α-thal was not investigated. However, application of the electrophoretic analysis on all samples with anemia allowed the Hb Constant Spring to be identified in some subjects. This is a variant of the α2 gene having the stop codon 142 *T*AA mutated in *C*AA, coding Gln. Four cases had genotype $\alpha^{CS}\alpha$/ $\alpha^{CS}\alpha$ and one case with $\alpha^{CS}\alpha$/ αα. The production of an elongated globin is lower than the normal so that this disease expresses as a non-deletional α^+-thal.

The main area of the distribution of Hb CS is Southeast Asia. The Hb CS gene frequency was 0.033 in Northern Thailand and approximately 0.01 in Central Thailand and Cambodia. High frequencies between 0.05 and 0.06, were observed in Northeastern Thailand. A study in Laotians suggest that the Lao-speaking populations of the Mekong river basin in Northeastern Thailand and Laos have the highest frequencies of the Hb CS in Southeast Asia (109, 114).

The α^{CS} gene frequency of the minority group in this study was not possible to calculate. Previous studies indicated that differences exist in the prevalence of this mutation among the minorities population. These are shown in Table 26 and 27.

Table 26. Comparison of α^{CS} gene frequency in South Vietnam, Central Vietnam

Population		**α^{CS} gene frequency**
South Vietnam (O'Riordan S et al) (108)	Kinh	0.001
	Tay	0.023
	Nung	0
	S'Tieng	0.030
Central Vietnam (Nguyen et al) (89)	Pregnant women (86.8% Kinh, 13.2% Cotu and Taoi)	0.0207
Central Vietnam (this study)	Kinh	absent
	Minority (Taoi, Cotu, Pako)	present

Table 27. Comparison of α^{CS} gene frequency in Southeast Asian populations

Population		**α^{CS} gene frequency**
Eight Southeast Asian populations (Laig M et al) (109)	Karachi	0
	Ahom	0
	Northern Thailand	0.0331
	Tak	0.0156
	Khonkhaen	0.0558
	Ubol Rachathani	0.0556
	Cambodia	0.0088
	So (adults)	0.1316
Central Vietnam (this study)	Kinh	absent
	Minority (Taoi, Cotu, Pako)	present

6. REFERENCES

1. Steinberg M.H., Forget B.G., Higgs D.R., Weatherall D.J. Disorders of haemoglobin. Cambridge University Press, Cambridge, 2009.
2. Perutz M.F. Mechanisms Regulating the reactions of human hemoglobin with oxygen and carbon monoxide. Annu. Rev. Physiol. 52:1-26, 1990.
3. Perutz M.F., Rossman M.G., Cullis A.F., Muirhead H., Will G., North A.G.T. Structure of haemoglobin: A three-dimensional Fourier synthesis at 5.5 Å resolution, obtained by x-ray analysis. Nature, 185:416-422, 1960.
4. Perutz M.F., Wilkinson A.J., Paoli M., Dodson G.G. The stereochemical mechanism of the cooperative effects in hemoglobin revisited. Annu. Rev. Biophys. Biomol. Struct. 27:1-34, 1998.
5. Fermi G., Perutz M.F., Shaanan B., Fourme R. The crystal structure of human deoxyhemoglobin at 1.74 Å resolution. J. Mol. Biol. 175:159-174, 1984.
6. Parkhurst L.J. Hemoglobin and myoglobin ligand kinetics. Annu. Rev. Phys. Chem. 30:503-546, 1979.
7. Boyer R. et al.2006. Myglobin & Hemoglobin http://www3.interscience.wiley.com:8100/legacy/college/boyer/0471661791/structure/HbMb/hbmb.htm.
8. Perutz M.F. Mechanisms regulating the reactions of human hemoglobin with oxygen and carbon monoxide. Annu. Rev. Physiol., 52: 1-26, 1990
9. Sadava D., Heller C.H., Orians G.H., Purves W.K., Hillis D.M. Life: The Science of Biology, 8th Edition. Massachusetts and Virginia. Sinauer Associates Inc. and W.H. Freeman and Company, 2008.
10. Rang H.P., Dale M.M., Ritter J.M., Moore P.K. Pharmacology, Fifth Edition. Elsevier. ISBN 0-443-07202-7.10, 2003
11. Steinberg M.H., Forget B.G., Higgs D., Nagel R.L. Disorders of Haemoglobin. Cambridge University Press, Cambridge, 2001.
12. Stamatoyannopoulos G., Majerus P.W., Perlmutter R.M., Varmus H. The Molecular Basis of Blood Diseases. Saunders Company, Philadelphia.
13. Manca L., Masala B. Disorders of the synthesis of human fetal hemoglobin. IUBMB Life, 60(2):94-111, 2008.
14. Zhang Z., Gerstein M. Large-scale analysis of pseudogenes in the human genome. Curr. Opin. Genet. Dev., 14:328-335, 2004.
15 Forget B.G., Hardison R.C. The normal structure and regulation of human globin gene clusters. In: Disorders of Hemoglobins: Genetics, Pathophysiology, and Clinical Management. 2nd Edition, Editors Steinberg M.H., Forget B.G., Higgs D.R., Weatherall D.J. Cambridge University Press, Cambridge, UK, pp. 46-61, 2009.
16. Hsu S., Marks J., Shaw J., et al. Structure and expression of the human θ1globin gene. Nature, 331:94-96, 1988.
17. Ley T.J., Maloney K.A., Gordon J.I., Schwartz A.L. Globin gene expression in erythroid human fetal liver cells. J. Clin. Invest., 83:1032-1038, 1989.
18. Albitar M., Peschle C., Liebhaber S.A. Thetaθ, ζ and ε globin messenger RNA are expressed in adults. Blood, 74:629-637, 1989.
19. Goh S.H., Lee Y.T., Bhanu N.V., et al. A newly discovered human α-globin gene. Blood, 106:1466-1472, 2005.
20. Cooper S.J., Wheeler D., De Leo A., et al. The mammalian αD-globin gene lineage and a new model for the molecular evolution of α-globin gene clusters at the stem of the mammalian radiation. Mol. Phylogenet. Evol., 38:439-448, 2006.

21. Hardison R.C., Sawada I., Cheng J-F., Shen C-K.J., Schmid C.W. A previously undetected pseudogene in the human α globin gene cluster. Nucleics Acids Res., 14:1903-11, 1986.
22. Clegg J.B. Can the product of the theta gene be a real globin? Nature, 329:465-6, 1987.
23. Tilghman S.M., Tiemeier D.C., Seidman J.G., et al. Intervening sequence of DNA identified in the structural portion of a mouse β-globin gene. Proc. Natl. Acad. Sci. USA., 75:725-729, 1978.
24. Goodman M., Czelusniak J., Koop B., Tagle D., Slightom J. Globins: A case study in molecular phylogeny. Cold Spring Harbor Symp. Quant. Bio., 52:875-90, 1987.
25. Proudfoot N.J., Gil A., Maniatis T. The structure of the human ζ-globin gene and a closely linked, nearly identical pseudogene. Cell, 31:553-563, 1982.
26. Tilghman S.M., Curtis P.J., Tiemeier D.C., Leder P., Weissmann C. The intervening sequence of a mouse β-globin gene is transcribed within the 15S β globin mRNA precursor. Proc. Natl. Acad. Sci. USA, 75:1309-1313, 1978.
27. Krainer A.R., Maniatis T., Ruskin B., Green M.R. Normal and mutant human β-globin pre-mRNAs are faithfully and efficiently spliced in vitro. Cell, 36:993-1005, 1984.
28. Padgett R.A., Grabowski P.J., Konarska M.M., Seiler S., Sharp P.A. Splicing of messenger RNA precursors. Annu. Rev. Biochem., 55:1119-1150, 1986.
29. Maston G.A., Evans S.K., Green M.R. Transcriptional regulatory elements in the human genome. Annu. Rev. Genomics Hum. Genet., 7:29-59, 2006.
30. deBoer E., Antoniou M., Mignotte V., Wall L., Grosveld F. The human β-globin promoter; nuclear protein factors and erythroid specific induction of transcription. EMBO. J., 7:4203-4212, 1987.
31. Efstratiadis A., Posakony J.W., Maniatis T., et al. The structure and evolution of the human β-globin gene family. Cell, 21:653-668, 1980.
32. Perkins A.C., Sharpe A.H., Orkin S.H. Lethal β-thalassaemia in mice lacking the erythroid CACCC-transcription factor EKLF. Nature, 375:318-322, 1995.
33. Asano H., Li X.S., Stamatoyannopoulos G. FKLF, a novel Kruppel-like factor that activates human embryonic and fetal β-like globin genes. Mol. Cell. Biol., 19:3571-3579, 1999.
34. Tjian R., Maniatis T. Transcriptional activation: A complex puzzle with few easy pieces. Cell, 77:5-8, 1994.
35. Bodine D., Ley T. An enhancer element lies 3' to the human Aγ globin gene. EMBO J., 6:2997-3004, 1987.
36. Hardison R., Slightom J.L., Gumucio D.L., Goodman M., Stojanovic N., Miller W. Locus control regions of mammalian β-globin gene clusters: Combining phylogenetic analyses and experimental results to gain functional insights. Gene, 205:73-94, 1997.
37. Higgs D., Wood W., Jarman A., et al. A major positive regulatory region located far upstream of the human α-globin gene locus. Genes Devel., 4:1588-1601, 1990.
38. Chada K., Magram J., Costantini F. Tissue- and stage-specific expression of a cloned adult β globin gene in transgenic mice. Prog. Clin. Biol. Res., 191:305-319, 1985.
39. Grosveld F., van Assendelft G.B., Greaves D., Kollias G. Position-independent, high-level expression of the human β-globin gene in transgenic mice. Cell, 51:975-985, 1987
40. Qiliang Li. A Melanesian α-thalassemia mutation suggests a novel mechanism for regulating gene expression. Genome Biol., 7:238-240, 2006.
41. Forget B.G., Bunn F.H. Classification of the Disorders of Hemoglobin. Cold Spring Harbor Perspect. Med., 2013
42. Patrinos G.P., Giardine B, Riemer C., Miller W., Chui D.H.K., Anagnou N.P., Wajcman H., Hardison R.C. Improvements in the HbVar database of human

hemoglobin variants and thalassemia mutations for population and sequence variation studies. Nucl. Acids Res. 32 Database issue: D537-541, 2004 http://globin.cse.psu.edu/hbvar/menu.html

43. Bunn H.F., Forget B.G. Hemoglobin: Molecular, Genetic and Clinical Aspects. Saunders, Philadelphia, 1986.
44. Manca L., Masala B. Disorders of the synthesis of human fetal haemoglobin. IUBMB Life, 60(2):94-111, 2008.
45. Thein S.L., Wood W.G. The molecular basis of β thalassemia, δβ thalassemia, and hereditary persistence of fetal hemoglobin. In: Disorders of Hemoglobins: Genetics, Pathophysiology, and Clinical Management - 2nd Edition, Editors Steinberg M.H., Forget B.G., Higgs D.R., Weatherall D.J. Cambridge University Press, Cambridge, UK, pp. 323-356, 2009.
46 Weatherall D.J., Clegg J.B. Inherited haemoglobin disorders: an increasing global health problem. Bull WHO, 79:704-712, 2001.
47. Chen X.W., Mo Q.H., Li Q., Zeng R., Xu X.M. A novel mutation of −73(A→T) in the CCAAT box of the β-globin gene identified in a patient with the mild β-thalassemia intermedia. Ann. Hematol., 86(9):653-657, 2007.
48. Safaya S., Rieder R.F., Dowling C.E., Kazazian H.H.J., Adams J.G. Homozygous β-thalassemia without anemia. Blood, 73:324-328, 1989.
49. Huang S-Z., Wong C., Antonarakis S.E., Ro-Lein T., Lo W.H.Y., Kazazian H.H.J. The same TATA box β-thalassemia mutation in Chinese and U.S. blacks: another example of independent origins of mutation. Hum. Genet., 74:162-164, 1986.
50. Thein S.L. Beta thalassaemia. In: Rodgers G.P., ed. Bailliere's Clinical Haematology, Sickle Cell Disease and Thalassaemia. London: Bailliere Tindall, pp:91-126, 1998.
51. Gonzalez-Redondo J.H., Stoming T.A., Kutlar A., et al. A C→T substitution at nt -101 in a conserved DNA sequence of the promoter region of the β-globin gene is associated with "silent" β-thalassemia. Blood, 73:1705-1711, 1989.
52. Maragoudaki E., Kanavakis E., Traeger-Synodinos J., et al. Molecular, haematological and clinical studies of the -101 C→T substitution of the β-globin gene promoter in 25 β-thalassaemia intermedia patients and 45 heterozygotes. Br. J. Haematol., 107:699-706, 1999.
53. Wong C., Dowling C.E., Saiki R.K., Higuchi R.G., Erlich H.A., Kazazian H.H.J. Characterization of β-thalassaemia mutations using direct genomic sequencing of amplified single copy DNA. Nature, 330:384-386, 1987.
54. Treisman R., Orkin S.H., Maniatis T.. Specific transcription and RNA splicing defects in five cloned β-thalassaemia genes. Nature, 302:591-596, 1983.
55. Camaschella C., Mazza U., Roetto A., et al. Genetic interactions in thalassemia intermedia: analysis of β-mutations, β-genotype, β-promoters, and β-LCR hypersensitive sites 2 and 4 in Italian patients. Am. J. Hematol., 48:82-87, 1995.
56. Spritz R.A., Jagadeeswaran P., Choudary P.V., et al. Base substitution in an intervening sequence of a β+ thalassemic human globin gene. Proc. Natl. Acad. Sci. USA, 78:2455-2459, 1981.
57. Westaway D., Williamson R. An intron nucleotide sequence variant in a cloned β+-thalassemia globin gene. Nucl. Acids Res., 9:1777-1782, 1981.
58. Busslinger M., Moschanas N., Flavell R.A. β+ thalassemia: aberrant splicing results from a single point mutation in an intron. Cell, 27:289-298, 1981.

59. Fukumaki Y., Ghosh P.K., Benz E.J, et al. Abnormally spliced messenger RNA in erythroid cells from patients with β+ thalassemia and monkey cells expressing a cloned β+-thalassemic gene. Cell, 28:585-593, 1982.
60. Metherall J.E., Collins F.S., Pan J., Weissman S.M., Forget B.G. β^o-thalassemia caused by a base substitution that creates an alternative splice acceptor site in an intron. EMBO J., 5:2551-2557, 1986.
61. Yang K.G., Kutlar F., George E., et al. Molecular characterization of β-globin gene mutations in Malay patients with Hb E-β-thalassaemia and thalassaemia major. Br. J. Haematol., 72:73-80, 1989.
62. Hall G.W., Thein S.L. Nonsense codon mutations in the terminal exon of the β-globin gene are not associated with a reduction in β-mRNA accumulation: a mechanism for the phenotype of dominant β-thalassemia. Blood, 83:2031-2037, 1994.
63. Blacklock H.A., Case J., Chan T., et al. Novel sequence insertion in a Maori patient with transfusion-dependent β-thalassaemia. Br. J. Haematol., 131(3):400–402, 2005.
64. Andersson B.A., Wering M.E., Luo H.Y., et al. Sickle cell disease due to compound heterozygosity for Hb S and a novel 7.7-kb beta-globin gene deletion. Eur. J. Haematol., 78(1):82-85, 2007.
65. Codrington J.F., Li H-W., Kutlar F., Gu L-H., Ramachandran M., Huisman T.H.J. Observations on the levels of Hb A2 in patients with different thalassemia mutations and a δ chain variant. Blood, 76:1246-1249, 1990.
66. Coleman M.B., Steinberg M.H., Adams J.G . Hemoglobin Terre Haute arginine β 106. A posthumous correction to the original structure of hemoglobin Indianapolis. J. Biol. Chem., 266(9):5798-5800, 1991.
67. Baiget M., Gomez Pereira C., Jue D.L., Johnson M.H., McGuffey J.E., Moo-Penn W.F. A case of hemoglobin Indianapolis [β112(G14)Cys→Arg] in an individual from Cordoba, Spain. Hemoglobin, 10(5):483-494, 1986.
68. De Biasi R., Spiteri D., Caldora M., et al. Identification by fast atom bombardment mass spectrometry of Hb Indianapolis [β112(G14)Cys→Arg] in a family from Naples, Italy. Hemoglobin, 12(4):323-336, 1988.
69. Thein S.L., Best S., Sharpe J., Paul B., Clark D.J., Brown M.J. Hemoglobin Chesterfield (β28 Leu→Arg) produces the phenotype of inclusion body β-thalassemia. Blood, 77:2791-2793, 1991.
70. Fucharoen S., Fucharoen G., Fukumaki Y., et al. Three-base deletion in exon 3 of the β-globin gene produced a novel variant (βGunma) with a thalassemia-like phenotype. Blood, 76:1894–1896, 1990.
71. Park S.S., Barnetson R., Kim S.W., Weatherall D.J., Thein S.L. A spontaneous deletion of β 33/34 Val in exon 2 of the β globin gene (Hb Korea) produces the phenotype of dominant β thalassaemia. Br. J. Haematol., 78:581-582, 1991.
72. Oner R., Oner C., Wilson J.B., Tamagnini G.P., Ribeiro L.M.L., Huisman T.H.J. Dominant β-thalassaemia trait in a Portuguese family is caused by a deletion of (G)TGGCTGGTGT(G) and an insertion of (G)GCAG(G) in codons 134, 135, 136 and 137 of the β-globin gene. Br. J. Haematol., 79:306-310, 1991.
73. Efremov G.D. Dominantly inherited β-thalassemia. Hemoglobin, 31(2):193-207, 2007.
74. Weatherall D.J., Williams T.N. Population genetics and global health burden. In: Disorders of Hemoglobins: Genetics, Pathophysiology, and Clinical Management - 2nd Edition, Steinberg M.H., Forget B.G., Higgs D.R., Weatherall D.J. (Eds). Cambridge University Press, Cambridge, UK, pp. 625-637, 2009.
75. Harteveld C.L., Higgs D.R. α-Thalassemia. Orphanet J. Rare Dis. 5(13):1-21, 2010.

76. Panja A., Ghosh T.K., Basu A. Genetics of thalassemia in Indian population. J. Commun. Nutr. Health, 1(1):39-46, 2012
77. Valentine W.N., Neel J.V. Hematologic and genetic study of transmission of thalassemia (Cooley's anemia: Mediterranean anemia). Arch. Intern. Med., 74:185-196, 1994.
78. Silvestroni E., Bianco I. Sulla frequenza dei portatori di malattia di morbo di Cooley e prime osservazioni sulla frequenza dei portatori di microcitemia nel Ferrarese e in alcune regioni limitrofe. Boll. Atti Acad. Med., 72:32, 1947.
79. Weatherall D.J. Thalassemia as a global health problem: recent progress toward its control in the developing countries. Ann. N.Y. Acad. Sci., 120:17-23, 2010.
80. Cao A., Kan Y.W. The prevention of thalassemia. Cold Spring Harb Perspect Med., 3: 01-15, 2012.
81. Christianson A., Howson C.P., Modell B. March of Dimes Global Report on Birth Defects. March of Dimes Birth Defects Foundation. New York, 2006.
82. Livingstone F.B. Frequencies of Hemoglobin Variants. Oxford University Press. New York, Oxford, 1985.
83. Fucharoen S., Winichagoon P. Haemoglobinopathies in Southeast Asia. Indian J. Med. Res., 134:498-506, 2011.
84. Saovaros Svasti M.L., Tran M.H., Munkongdee T., et al. Molecular analysis of β-thalassemia in South Vietnam. Am. J. Hematol., 71:85-88, 2002.
85. Nguyen C.K., Le T.T., Duong B.T., et al. β-Thalassemia/Haemoglobin E Disease in Vietnam. J. Tropical Pediat., 36:43-45, 1990.
86. Colah R., Gorakshakar A., Phanasgaonkar S., et al. Epidemiology of β-thalassaemia in Western India: mapping the frequencies and mutations in different geographic regions. Br. J. Haematol., 149(5):739-747, 2010.
87. de Silva S., Fisher C.A., Premawardhena A., et al. Thalassaemia in Sri Lanka: implications for the future health burden of Asian populations. Sri Lanka Thalassaemia Study Group. Lancet, 355:786–791, 2000.
88. Galanello R., Eleftheriou A., Trager-Synodinos J., Old J., Petrou M., Angastiniotis M. Prevention of thalassaemias and other haemoglobin disorders.Thalassemia International Federation Publications.Vol 1, 2003.
89. Nguyen H.V., Sanchaisuriya K., Nguyen D., et al. Thalassemia and hemoglobinopathies in Thua Thien Hue Province, Central Vietnam. Hemoglobin, 37(4):333-42, 2013.
90. WHO. Haemoglobin concentrations for the diagnosis of anemia and assessment of severity. VMNIS, 1-6, 2011.
91. Giordano P.C. Strategies for basic laboratory diagnostics of the hemoglobinopathies in multi-ethnic societies: interpretation of results and pitfalls. Int. J. Lab. Hematol., 5(5):465-479, 2013.
92. Masala B., Manca L. Detection of the common Hb F Sardinia [$^{A}\gamma$75(E19)Ile→Thr] variant by isoelectric focusing in normal newborn and in adult affected by elevated fetal hemoglobin syndromes. Clin. Chim. Acta., 198(3):195-202, 1991.
93. Manca L., Formato M., Demuro P., Pilo G., Gallisai D., Orzalesi M., Masala B. The γ globin chain heterogeneity of the Sardinian newborn baby. Hemoglobin, 10(5):519-528, 1986.
94. Masala B., Manca L. Separation of globin chains by the reversed-phase high-performance liquid chromatography. Methods Enzymol. 231:21-44, 1994.
95. Miller S.A., Dykes D.D., Polesky H.F. A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res., 16(3):1215, 1988.

96. Orkin S.H., Kazazian H.H. Jr, Antonarakis S.E., Goff S.C., Boehm C.D., Sexton J.P., Waber P.G., Giardina P.J. Linkage of β-thalassaemia mutations and β-globin gene polymorphisms with DNA polymorphisms in human β-globin gene cluster. Nature, 296(5858):627-631, 1982.
97. Schouten J.P., McElgunn C.J., Waaijer R., Zwijnenburg D., Diepvens F., Pals G. Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. Nucleic Acids Res., 30(12):e57, 2002.
98. Chang J.C., Kan Y.W. β^o thalassemia, a nonsense mutation in man. Proc. Natl. Acad. Sci. USA, 76(6):2886-2889, 1979.
99. Kazazian H.H. Jr., Orkin S.H., Antonarakis S.E., Sexton J.P., Boehm C.D., Goff S.C., Waber P.G. Molecular characterization of seven β-thalassemia mutations in Asian Indians. EMBO J., 3(3):593-596, 1984-
100. Kimura A., Matsunaga E., Takihara Y., Nakamura T., Takagi Y., Lin S., Lee H. Structural analysis of a β-thalassemia gene found in Taiwan. J. Biol. Chem., 258(5):2748-2749, 1983.
101. Chan V, Chan TK, Kan YW, Todd D. A novel β-thalassemia frameshift mutation (codon 14/15), detectable by direct visualization of abnormal restriction fragment in amplified genomic DNA. Blood, 72(4):1420-1423, 1988.
102. Clegg J.B., Weatherall D.J., Milner P.F. Haemoglobin Constant Spring: α chain termination mutant? Nature, 234(5328):337-340, 1971.
103. Nakatsuji T., Kutlar F., Kutlar F., Huisman T.H.J. Haplotypes among Vietnamese hemoglobin E homozygotes including one with a γ-globin gene triplication. Am. J. Hum. Genet., 38(6):981-983, 1986.
104. Antonarakis S.E., Orkin S.H., Kazazian H.H. Jr, Goff S.C., Boehm C.D., Waber P.G., Sexton J.P., Ostrer H., Fairbanks V.F., Chakravarti A. Evidence for multiple origins of the β^E-globin gene in Southeast Asia. Proc. Natl. Acad. Sci. U.S.A. 79(21):6608-6611. 1982.
105. Old J., Trager-Synodinos J., Petrou M., Galanello R., Angastiniotis M. Prevention of Thalassaemias and Other Haemoglobin Disorders.Thalassemia International Federation Publications. Vol.2, 2004.
106. Moiz B., Hashmi M.R., Nasir A., et al. Hemoglobin E syndromes in Pakistani population. BMC Blood Disord., 12(3):01-06, 2012.
107. Weatherall D.J. The inherited diseases of hemoglobin are an emerging global health burden. Blood, 115(22):4331-4336, 2010.
108. O' Riordan S., Hien T.T., Miles K., et al. Large scale screening for haemoglobin disorders in southernVietnam: implications for avoidance and management. Br. J. Haematol., 150(3):359–364, 2010.
109. Laig M., Pape M., Hundrieser J., et al. The distribution of the Hb Constant Spring gene in Southeast Asian populations. Hum. Genet., 84:188-190, 1990.
110. Filon D., Oppenheim A., Rachmilewitz E.A., et al. Molecular analysis of β-Thalassemia in Vietnam. Hemoglobin, 24(2):99-104, 2000
111. Le T.H., Pissard S., Pham H.V., et al. Molecular analysis of β-thalassemia in South Vietnam. Hemoglobin, 25(3):305-309, 2001.
112. Pornprasert S., Moriyama A., Kongthai K., et al. Detection of β thalassemia/hemoglobin E disease in samples which initially were diagnosed as homozygous haemoglobin. Europ. Clin. Lab., 5+6:1-5, 2013.
113. Steger H., Eigel A., Flatz G., et al. Hemoglobin E and codon 17 nonsense: two β-globin gene mutations common in Southeast Asia detected by the use of ARMS. Ann. Hematol., 67:119-120, 1993.

114. Facharoen S., Fucharoen G., Sae-ung N., et al. Thalassemia intermedia associated with the Hb Constant Spring EE Bart's disease in pregnancy: A molecular and hematological analysis. Blood Cells Mol. Dis., 39:195-198, 2007.

7. APPENDIX

Ethic approval

To:

The Chairman

The Scientific Committee of Hue University of Medicine and Pharmacy

Hue University of Medicine and Pharmacy

APPLICATION FOR ETHIC APPROVAL FOR IMPLEMENTATION OF SCIENTIFIC RESEARCH

Name of investigator: LE PHAN MINH TRIET

The title of study project:

Hemoglobinopathies in mountainous region of Thua Thien Hue, Vietnam

This study is a part of my study project fo my PhD thesis of the doctorate programme in Biomolecular and Biotechnological Sciences in Sassari University, Italy.

Supervisor: Prof. Bruno Masala, University of Sassari and Prof. Nguyen Viet Nhan, Hue University of Medicine and Pharmacy

Objectives of the project:

- Estimate the prevalence of hemoglobinopathies in 2 mountainous districts of Thua Thien Hue, Vietnam – Nam Dong and A Luoi
- Analyse gene mutations in patients with hemoglobinopathies.

The intended period of time for the study project:

From June 2011 to July 2013

The participants will be enrolled:

People from Nam Dong and A Luoi are randomly enrolled into the study

The procedures will be taken:

The study will be carry out in order to implement the following steps:

- Select the participants
- Perform clinical examinations and complete the recording form
- Collect the blood samples

Declaration of responsibility:

- Agreement (in writing or oral) from participants or through participants' guardians will be taken before entering them into study.
- I am aware that any medical procedure, which can be used on patients, may bring all the potential risks to patients' health and I also know with certainly that the procedure performed for collection of specimen will make no harm to patient.
- The patient data, record and results of examination and laboratory analysis will be kept only for scientific purpose and not for anything else.

I confirm that the information contained in this application is correct and true.

Hue, 05th June 2011

Investigator Signature

Le Phan Minh Triet

Approval

The Scientific Committee of Hue University of Medicine and Pharmacy

Hue University of Medicine and Pharmacy

Chairman and Rector

Prof. Dr. Cao Ngoc Thanh

Attestation of Authorship

I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person nor material which to a substantial extent has been accepted for the qualification of any other degree or diploma of a university or other institution of higher learning, except where due acknowledgement is made in the acknowledgements.

Name: Le Phan Minh Triet Sign:

ACKNOWLEGEMENT

Many people and organizations contributed to the completion of my PhD study and are too numerous to name although I acknowledge the contribution and assistance provided by everyone. Specifically I would like to express my sincere gratitude and appreciation to the following people and organizations that enabled me to study in Italy and Vietnam, and for the ideas and suggestions given to me during the research process.

First and foremost, I would like to express my sincere gratitude to my supervisor Professor Bruno Masala for the support of my PhD study and research, for his patience, motivation, enthusiasm, and immense knowledge. I will never forget. He has been my inspiration as I hurdle all the obstacles in the completion this research work.

I sincerely express my profound gratitude to Professor Nguyen Viet Nhan, my co-supervisor, who encouraged and helped me to complete this research.

I would like to thank Professor Piero Cappuccinelli. His constant inputs both in Italy and in Vietnam have helped to progress my studies.

I sincerely express my profound gratitude to Professor Cao Ngoc Thanh, Rector of Hue University of Medicine and Pharmacy, who introduced, encouraged and helped me to complete this research.

I would like to express my sincere thanks to Professor Claudia Crosio who facilitated my attendance to the PhD program at Sassari University.

I would like to thank Professor Nguyen Duy Thang, director of Hematology and Blood Transfusion Center – Hue Central Hospital for the support of my research.

I would like to acknowledge the generous financial contribution received from the Science and Technology Service of Thua Thien Hue province.

I sincerely say thanks to everyone in the laboratory in Sassari University was dedicated to helping me during the course of this study. In particular I will never forget the sincere help of Monica Pirastru, Paolo Mereu, Sandro Trova, and Luca Decandia who helped me make

all the techniques for my studies. Furthermore, I equally appreciate the support and wish happiness to Professor Laura Manca.

I owe Hematology department – Hue Central Hospital for their support across the studying process, making things happen to ensure the success of my work. Especially Dr.Phan Thi Thuy Hoa, who gave me valuable advice and shared the works with me during the research time.

I owe Nam Dong Health Center and A Luoi Health Center for enabling me to collect the samples.

I would like to thank my colleagues, my friends who helped, encouraged me and shared the works with me. I never forget the sincere help of Dr Le Van An, Dr Ngo Viet Quynh Tram, Dr Ha Thi Minh Thi, Dr Ton Nu Phuong Anh, Dr Duong Thi Hao, Dr Phu Thi Hoa, MSc Nguyen Hoang Bach, MSc Doan Huu Nhat Binh, Dr Le Bao Chi (in Vietnam), Antonio and Barbara, Dr Maura Fiammi, Phuoc Lai and Salvatore (in Italy).

With sincere gratitude and wish all good things, I want to send it to everyone in the Department of Hematology. I also would like to send best wishes success to the Dr Ha Nu Thuy Duong who worked closely with me, shared the works and encouraged me during the most difficult times.

With my profound gratitude I also sincerely thank the confidence and enthusiasm patients to participate, who helped me to complete this thesis.

Finally, and most importantly, I would like to thank my family for all their love and encouragement, my brother, my sister, especially my mother, who quietly care and sacrifice a lot to me.

Words cannot express the gratitude. Again sincerely thank you !