# CHAPTER 4

## Discussion

Alpha-thalassemia, a common genetic defect, is a major health problem throughout Southeast Asia including Thailand [72]. Recent information of the prevalence of α-thalassemias is essential to the prevention and control of this genetic disorder. In this study, recent information of the prevalence of α-thalassemia genotypes were investigated in 638 primigravidarum (first time of pregnancy) volunteers who came to the antenatal clinic at Maharaj Nakorn Chiang Mai Hospital, northern Thailand. Identification of the common genotypes of α-thalassemia such as $\alpha^0$-thalassemia $--^{SEA}$ type, $\alpha^+$-thalassemia $-\alpha^{3.7}$and $-\alpha^{4.2}$ types, and HbCS were determined by standard methods of gap-PCR and PCR-RFLP, respectively.

In this study, all volunteers were primigravidarum because one of the major objective in this study was to detect α-thalassemia carrier genes in the first time of pregnancy subjects and then provided the genetic counseling for the high and at risk couples to prevent or avoid having babies with severe α-thalassemias. Early and immediate prevention and control of α-thalassemias in primigravidarum should be an effective strategy to prevent and control of this genetic defect in public health. Sample sizes in this study of 638 cases were statistically large enough for analysis (calculated by the one sample proportion method). All primigravidarum volunteers were lifetime residents in 15 out of 17 provinces of northern Thailand.



When the prevalence of α-thalassemias of this study was compared with other studies in northern Thailand by statistical analysis using the Pearson chi-squire of SPSS for Windows software version 13.0, the prevalence of $\alpha^+$-thalassemia $-\alpha^{3.7}$ (15.52%) and $-\alpha^{4.2}$ (0.31%) types, and HbCS (4.23%) of my data was not statistically different from the data of Hundrieser et al.[73] and Lemmens-Zygulska et al. [74] ($p$ value > 0.05). The prevalence of $\alpha^0$-thalassemia $--^{SEA}$ type (12.23%) in my study was not statistically

different from the data of Lemmens-Zygulska et al. [74] and Wanapirak et al. [75] with $p$ value > 0.05. It was noticeable that the study subjects of Lemmens-Zygulska et al. [74] were all males collected from the annual conscription examination of the Royal Thai Army in 4 districts of Chiang Mai. While my subjects were all females from 15 provinces of northern Thailand, nevertheless, both groups of subjects were carefully included only volunteers who were born in and residing in northern Thailand. The study subjects of Wanapirak et al. [75] were pregnant women volunteers at Maharaj Nakorn Chiang Mai Hospital which mostly living in Chiang Mai. However, the prevalence of $\alpha^0$-thalassemia $--^{SEA}$ type of my result was significantly higher than the result of Hundrieser et al. [73] with $p$ value < 0.05. It might be due to the migration of people from central, a lower of α-thalassemia frequencies, to the areas of northern Thailand [74].

To my knowledge, this study is the first report able to detect the deletional HbH disease in northern Thai primigravidarum (18/638, 2.82%). These pregnant subjects should have received special medical attention because they might be developed an acute hemolysis when exposed to oxidant drug or infection [3] and their fetuses could be affected from the anemic pregnancy. The existence of the deletional HbH disease indicated that there must be some interaction of $\alpha^0$- and $\alpha^+$-thalassemia consistently existed in this area. However, this study could not encounter the nondeletional HbH-CS disease in pregnant subjects. It might be due to the low frequency of HbCS (4.23%) compared to the high frequency of $\alpha^+$-thalassemia (16.61%) in northern Thailand leading to much more lower probability of HbCS and $\alpha^0$-thalassemias interaction than $\alpha^+$- and $\alpha^0$-thalassemias interaction.



Although PCR is a standard tool for gene analysis with reliable and accurate enough for identification of α-thalassemia genotypes. However, it is not applicable for screening of large population due to the sophisticated equipments, high cost of reagents, requiring well trained technician and tedious laboratory intervention. Therefore, many groups of researcher implemented the immunoassay as tools of immunodiagnosis for α-thalassemias. Various methods in immunoassay had been used to detect α-thalassemias such as radioimmunoassay [12, 13], ELISA using monoclonal antibody [16, 18], an ELISA strip [17] and sandwich-type immunochromatographic strip [19].

Most of the recent immunoassay utilized monoclonal antibody to capture the specific antigen for immunodiagnosis. Since monoclonal antibody is synthesized and secreted from hybridoma, therefore, generation, storage and cultivation of hybridoma are vital for the production and secretion of monoclonal antibody. The common problem of hybridoma was gradually lost of the synthesis and secretion of monoclonal antibody during long-term cultivation [20]. A more recent and advance recombinant DNA technology, therefore, was implemented in this study.

Recombinant DNA technology was used for the construction of scFv antibody. In the general scientific work, the starting materials for the generation of recombinant antibody were hybridoma and combinatorial libraries generating from B-lymphocytes [52]. In this study, mouse hybridoma producing monoclonal antibody specific to Hb Bart's was the starting material. The advantage of using hybridoma was the reliable cloning of the functional immunoglobulin genes. Furthermore, the procedure including RNA isolation, cDNA synthesis, antibody gene amplification and gene cloning were practical for the selection of the desired $V_H$ and $V_L$ genes. Importantly, mouse hybridoma used for the construction of the scFv antibody was confirmed of the ability to produce monoclonal antibody highly specific to Hb Bart's without cross reactivity to other hemoglobins [15]. Therefore, this mouse hybridoma was a good starting material for the generation of the scFv antibody.

Cloning strategy of $V_H$ and $V_L$ genes was based on the cloning of amplicon that amplified by PCR using the antibody-specific primers. These degenerate set of primers were previously used for the amplification of $V_H$ and $V_L$ genes. The success of amplification of $V_H$ and $V_L$ genes were proved by analysis of nucleotide and amino acid sequences. Nucleotide sequences of both $V_H$ and $V_L$ genes at 5′ and 3′ end showed the identical sequence to at least one pair of the degenerate set of primers. The forward and reverse primers of the degenerate set of primers used for amplification of $V_H$ gene were HB5 and HB6, and HF4, respectively. While the forward and reverse primers of the degenerate set of primers used for the amplification of $V_L$ gene were LB5, LB11 and LB13, and LF2, respectively. Previous studies [76] examined the amino acid sequences of the region proposed to form the variable domain of antibody. They proposed the variable domain of an antibody for antigen binding site in termed of complementarity

determining regions (CRDs). Kabat and Chothia numbering scheme is the most widely used standard for numbering the amino acid residues in variable domain of an antibody. Using this numbering scheme, our deduced amino acid sequence of the selected $V_H$ and $V_L$ genes that retrieved from translated nucleotide could be identified and located all of 3 CDRs. This result confirmed that our selected $V_H$ and $V_L$ genes were variable domain of an antibody gene.

Analysis of the nucleotide and amino acid sequences revealed that all of ten of $V_H$ gene were identical and unique whereas two populations of $V_L$ gene were observed. One of the $V_L$ population (5/10), the deduced amino acid sequences represented the premature termination and causing of the shortened polypeptide products. The deduced amino acid sequences of the other populations (5/10) were aligned with the available sequences in GenBank public database using BLAST program (http://blast.ncbi.nlm.nih.gov/Blast.cgi). We found that amino acid sequences of this populations of $V_L$ were identical to mouse anti-chloramphenicol immunoglobulin kappa light chain variable region (accession no. ACV40677) and anti-hTERT single-chain variable fragment (99%) (accession no. ABW90123). All of results indicated the success of $V_H$ and $V_L$ genes cloning.

The strategy for the construction of the full length gene ($V_H$-linker-$V_L$ format) was performed by SOE-PCR and the full length gene was inserted in the correct orientation via the multiple cloning sites of pET28a(+) expression vector. Therefore, the amplicon of the full length gene must contained the recognition nucleotide sequences of two different restriction enzymes at both ends for the production of different protruding end that were compatible with both ends of pET28a(+) vector. For this purpose, we designed a new set of primers based on the creation of the nucleotide sequences including the linker for assembly the $V_H$ and $V_L$ gene into the full length gene, recognition sites of two different restriction enzymes at each end and a stop codon. These designed primer set were very useful for the amplification of the full length gene that consist of *Nde* I restriction site at 5′ end, *Bam*H I restriction site, a stop codon at 3′ end and a linker sequences between $V_H$ and $V_L$ gene. The amplicon of the full length gene was suitable for subsequently cloning into pET28a(+) expression vector via *Nde* I and *Bam*H I restriction sites. Comparison with the previous known sequence, the results showed the success of the construction of the full length scFv gene.

The pET expression vector and *E. coli* BL21(DE3) host strain was the most widely used vector-host system for the expression of many recombinant scFv antibodies (htpp://refold.med.monas.edu.au). In this study, the expression strategy of recombinant scFv antibody was based on the expression of pET28a(+)-scFv anti-Hb Bart's vector in *E. coli* BL21(DE3) host strain. Expressed protein in form of scFv antibody fusion with N-terminal histidine was used for the detection and purification. When the expressed scFv antibody was evaluated by SDS-PAGE, it showed the predicted molecular weight of 28 kDa in form of inclusion bodies. When confirmed by Western blot analysis using HRP conjugated mouse anti-His tag, it could be detected the 6 residues of N-terminal histidine that fused with scFv antibody.

Basically, the expression of recombinant proteins in *E. coli* can be in soluble or insoluble form (inclusion bodies) depending on the nature of the expressed proteins and expression system including vector-host used and bacterial culture conditions [77, 78]. However, most of the scFv antibody tends to form inclusion bodies when expressed in *E. coli* [79-81]. The high rate of protein synthesis in reducing environment of bacterial cytoplasm results in the insufficient of the correct protein refolding. These effect leads to the formation of insoluble aggregates form called inclusion bodies [82]. There are advantages of the expressed protein in form of inclusion bodies. First, recombinant protein was highly expressed in level around 10-40% of the total protein. Second, inclusion bodies were easily isolated from bacterial protein due to differences in their size and density as compared with cellular protein. Third, expressed protein in form of inclusion bodies was resisted to proteolytic attack by cellular proteases due to the denature form and compact of inclusion bodies. Last, there is homogeneity of the protein of interest in inclusion bodies which helps in reducing the number of purification steps to recover pure protein [78, 83, 84].

Alternatively, the expression of recombinant proteins in *E. coli* might be soluble using the appropriate expression vector, matching to the bacterial host strain and optimized culture conditions (www.novagen.com). First, the expression vector that enhanced the solubility and/or folding of the recombinant proteins was previously described. For example, pET41 and pET42 vector provided the glutathione-S-transferase (GST) tag that itself is highly soluble. The pET32 vector provided the thioredoxin (Trx) tag that

catalyzed disulfide bond formation. The pET22, pET25, pET27 and pET36 provided a signal sequence for translocation a signal peptide tag recombinant proteins into the periplasmic space where is more favorable environment for folding and disulfide bond formation [85]. Second, *E. coli* strain Origami(DE3) that contained *trx*B and *gor* mutation were used for the enhancement of the disulfide bond formation of the recombinant proteins in the cytoplasm of bacteria. Mutation of *trx*B (encoding thioredoxin reductase) and *gor* (encoding glutathione reductase) in this host strain means the absent of both reductase enzymes leads to more oxidizing environment in bacterial cytoplasm for folding and disulfide bond formation [77]. Third, the optimized culture conditions could increase the solubility of recombinant proteins for example, the induction of protein at lower temperature (20-25ºC) leads to the reduction of the growth rate of bacteria, therefore, expressed protein having more time to refold into the native structures. The addition of 0.4 M sucrose to the culture medium for inducing the osmotic shock response increases the internal level of glutamate and proline thus providing the conditions more favorable for protein folding [78]. However, all of strategies for increasing the solubility of the recombinant proteins that mentioned above did not guarantee the desire results. I tried to optimize the induction of protein at lower temperature (20ºC), however, the expressed protein remained in form of inclusion bodies.

The result revealed high expression of the scFv antibody in bacterial lysate as 47% of total protein (53 mg). Recovery yield of the scFv antibody was decreased in each step of purification (100 to 37%) while the purity of scFv antibody was increased (47 to 90%). The purification procedure of the scFv antibody against Hb Bart's was based on the specific binding of N-terminal 6 histidine residues with Ni-NTA affinity chromatography under denaturing condition, then, directly refolded in redox refolding buffer by dilution method. Conventional methods for protein refolding have 3 types including dilution, dialysis and on-column refolding. Dilution refolding is the process to dilute out the denaturant by diluting the solubilized protein into a suitable refolding buffer. Dialysis is the process to dilute out the denaturant by exchange the buffer. On-column refolding is the basic principle that denatured protein bound to a column resin or residing within the pores of a gel filtration resin are tethered or sequestered and if protein experienced a gradual decrease in the concentration of denaturant, denatured

protein will refold. The most effective and widely used was the refolding by dilution [78, 84]. After the refolding step by dilution, the recovery yield of the scFv antibody was 2% (1 mg/53 mg of scFv protein starting from bacterial lysate). Loss of the scFv antibody resulted from the aggregation of protein in the refolding step. Previous studies reported the variation of the refolding and recovery efficiency in different antibodies as 0.63-20 mg per liter of bacterial culture [85-89]. For this study, the amount of scFv antibody was approximately 1 mg starting from 1 L of bacterial culture. Higher scale of recovery yield might be improved by changing the expression system of scFv antibody to be in soluble form or optimized the formula of the refolding buffer that enhanced the solubility of the scFv antibody.

Although the low recovery yield of the scFv antibody was observed, however, the specificity of the scFv antibody to Hb Bart's was retained when compared with the parent monoclonal antibody. The specificity of the scFv antibody to Hb Bart's was demonstrated by Western blot analysis and indirect ELISA. Both investigative methods showed the same specific binding of the scFv antibody to Hb Bart's without cross reactivity to HbA ($\alpha_2\beta_2$), HbF ($\alpha_2\gamma_2$), HbS ($\alpha_2\beta^S{}_2$), HbE ($\alpha_2\beta^E{}_2$), $HbA_2$ ($\alpha_2\delta_2$) and HbH ($\beta_4$). It was clearly shown that the scFv antibody only react to the specific folding region of the $\gamma_4$ globin chain but not any folding regions or side chains of the $\gamma_2$ globin chain.

The sensitivity of the scFv antibody and the parent monoclonal antibody to Hb Bart's were determined by dot blot ELISA. The detection sensitivity of Hb Bart's by both antibodies was 5 μg/μL of Hb Bart's hydrops fetalis hemolysate. Dot blot ELISA indicated the similar detection sensitivity of the scFv antibody and the parent monoclonal antibody.



The mean affinity constant ($K_a$) of the scFv antibody against Hb Bart's and the parent monoclonal antibody were $8.31 \pm 1.28 \times 10^5$ $M^{-1}$ and $2.11 \pm 0.20 \times 10^6$ $M^{-1}$, respectively. The mean $K_a$ of the scFv antibody was less than the parent monoclonal antibody as 2.5 fold. Generally, the mean $K_a$ of antibodies was range between $10^6$-$10^{12}$ $M^{-1}$ [90]. The lower $K_a$ of the scFv antibody than the parent mAb has previously been reported. It might be affected by the incomplete folding [91] and/or the conformation change at the antigen binding site of the scFv antibody [77]. However,

the binding affinity of the scFv antibody might be improved by randomized, site-directed mutagenesis and chain shuffling [92]. Further study of protein engineering to improve the affinity might increase the advantage of the recombinant antibody usage in various scientific fields.