CHAPTER 1
INTRODUCTION

Background and Rationale

Fragile X syndrome (FXS) is the most common inherited form of mental retardation (MR) with a prevalence of 1:4,000 in males and 1:8,000 in females. MR can vary from mild to severe, and long narrow face, prominent jaw, large ears along with macroorchidism in most post-pubertal men are typical features in FXS individuals. These patients also typically feature behavioral and social problems, including hyperactivity, attention deficit and autistic-like behaviors. The main cause of mutation is the expansion of the repeated CGG trinucleotide within the 5′ UTR of the first exon of the *FMR1* gene and subsequent hypermethylation at the CpG island in the promoter region of this gene leading to transcriptional silence of the FMR1 mRNA and the absence of FMRP translation. The polymorphic CGG repeat lengths of most normal individuals range from 6-59 repeats with two or more stabilizing AGG interspersions occurring once every 9 or 10 CGG blocks, while premutation carriers (60-200 repeats) and full mutation alleles (>200 repeats) have one or no AGG interspersions with long uninterrupted CGG tracts (>34 pure CGG triplets) at the 3′ end of the repeat segments, indicating that AGG substructures play a crucial role in maintaining repeat stability. Haplotype analysis of previous studies in Chinese, Japanese and Caucasian populations using the adjacent microsatellite markers around the *FMR1* gene have revealed linkage disequilibrium between the fragile X mutation and its flanking markers so that fragile X chromosomes can display several haplotypes, but only a few of these haplotypes account for most fragile X patients, with statistically significant differences from the normal groups. Some isolated ethnic groups, such as Finns, Swedes and Tunisian Jews, present a striking founder effect of a single dominant haplotype shared by the majority of FXS patients. In order to ascertain whether or not FXS in Thai patients has a specific haplotype association, Limprasert and colleagues (2001) analyzed the haplotype at the *FRAXA* locus in Thai subjects. They used both microsatellite and single nucleotide polymorphism (SNP) as the polymorphic markers to establish the haplotype. Although they found no significant specific haplotype differences between normal control and FXS groups, they found a specific haplotype association in the normal CGG repeat groups. Therefore, this study planned both a haplotype analysis using twofold samples of FXS patients and

more SNP markers that are dispersed through the *FMR1* gene, and an examination of the AGG interruption pattern to provide a clear insight into the molecular basis of FXS in the Thai population.

Review of the Literature

Fragile X syndrome (FXS), a single gene disorder, is the most common form of inherited mental retardation and the second most frequent cause of mental retardation after Down syndrome (Fu et al., 1991; Rousseau et al., 1995; Poon et al., 1999). The syndrome is an unusual X-linked dominant disorder with reduced penetrance of 80% in males and 30% in females (Jin and Warren, 2000). This disease is an example of a dynamic mutation that results from the instability of the CGG repeats and subsequent CGG repeat expansion (Jin and Warren, 2000; Ennis et al., 2001; Oostra and Chiurazzi, 2001). The estimated prevalence in males (1:4,000) is higher than in females (1:8,000), reflecting the locus of the disease causing gene on the X chromosome (Murray et al., 1996; Turner et al., 1996; de Varies et al., 1997; Morton et al., 1997; Berry-Kravis et al., 2002). The name of the disorder, moreover, comes from the presence of a visible fragile site in cytogenetics analysis, designated as *FRAXA* (Fragile site, X chromosome, A site), at Xq27.3 near the end of the long arm of the X chromosome (Jin and Warren, 2000; Schaeffer et al., 2003).

The predominant manifestation of FXS patients is mental retardation (MR) ranging from mild to severe with the average adult IQ being 40 in males and 80 in females (Bontekoe et al.,2001; Oostra and Chiurazzi, 2001; Berry-Kravis et al., 2002; Schaeffer et al., 2003). Women and girls with high intellectual functioning have a higher percentage of cells with the normal active X chromosome, suggesting that intellectual functioning in FXS females can be correlated to the X inactivation ratios (Berry-Kravis et al., 2002). Besides the varying degrees of severity of MR, the common physical appearance of affected adults includes the typical facial features of long narrow face and large protruding ears, and macroorchidism (large testis) in most post-pubertal males (Jin and Warren, 2000; Bontekoe et al., 2001; Oostra and Chiurazzi, 2001; Berry-Kravis et al., 2002; Schaeffer et al., 2003). Other observed features are a prominent forehead and jaw, midfacial hypoplasia, high arched palate, macrocephaly, hyperextensible joints, loose and wrinkled skin over the hand, single palmar crease and other abnormal dermatoglyphic patterns such as a vertical hallucal crease and flat feet. These phenotypes are most prominent in adult

patients, but not every FXS patient shows all the abnormal physical characteristics. Behavioral and social dysfunctions such as attention deficit, hyperactivity, anxiety, perseveration, tactile defensiveness and autistic-like behaviors are often the most problematic issues faced by families of affected individuals. For example, about 90% of pre-pubertal boys with FXS show hyperactivity and virtually all have attentive difficulties. FXS girls frequently have attention problems and may occasionally manifest hyperactivity (Berry-Kravis et al., 2002). The majority of carriers, another type of FXS mutation, manifests usual gene expression and result in normal phenotypes. Recent evidence has also shown that approximately 20% of carrier females exhibit premature ovarian failure (POF) before the age of 40, leading to the risk of early estrogen deprivation and subsequent low bone density (Hundscheid et al., 2003; Mandel et al., 2004). Some carrier males develop a progressive neurodegenerative disorder, named the fragile X tremor/ataxia syndrome, at the beginning of their 50s or 60s. The characteristics of this syndrome are intention tremor, cerebellar ataxia, short term memory loss, Parkinsonism, peripheral neuropathy and generalized brain atrophy (Jacquemont et al., 2003; Mandel et al., 2004).

The polymorphic CGG repeat can expand during transmission. The most important factor for the expansion of a premutation allele is the sex of the transmitting parent (Murray et al., 1997). The expansion from premutation to full mutation only occurs through female transmission (Fu et al., 1991; Murray et al., 1997; Jin and Warren, 2000; Bontekoe et al., 2001; Schaeffer et al., 2003). The risk of expansion from a premutation to a full mutation increase in a following generation depends on the size of the premutation and the position of the individual in the pedigree. In the size of the premutation, the risk of expansion increases with the number of CGG repeats. For example, the larger premutation alleles of over 100 repeats expand to a full mutation in 100% of cases within one generation (Murray et al., 1997). In the position of the individual in the pedigree, brothers of normal transmitting males (NTMs) are at low risk (~9%), while grandsons and great grandsons have much higher risk (40% and 50% respectively). This particular form of anticipation has become known as the Sherman paradox (Fu et al., 1991; Jin and Warren, 2000; Bontekoe et al., 2001). Male premutation carriers (NTMs) never transmit a full mutation to their daughters (Bontekoe et al., 2001). Transmission of a premutation from father to daughter is usually associated with small (<10 repeats) increases or decreases in size, hence daughters of NTMs are never at risk for the fragile X phenotype (Fu et al., 1991; Macpherson et al., 1994). The reason for this comes from some kind of selection mechanism. Full mutation males have only premutation alleles in their sperm (Reyniers et

al., 1993). This mechanism appears to protect the male germline against transmission of full mutation (Jin and Warren, 2000; Bontekoe et al., 2001). Transition from a normal sized allele to a premutation allele has never been observed in fragile X pedigrees (Eichler et al., 1996).

The cause of fragile X syndrome is the expansion of a CGG trinucleotide repeat. The highly polymorphic CGG trinucleotide repeat is located at the 5′ untranslated region (5′ UTR) in the first exon of the *Fragile X Mental Retardation 1 (FMR1)* gene (Fu et al., 1991; Kremer et al., 1991; Verkerk et al., 1991; Ashley et al., 1993; Eichler et al., 1993; Murray et al., 1996; Jin and Warren, 2000; Bontekoe et al., 2001; Oostra and Chiurazzi, 2001; Schaeffer et al., 2003). The *FMR1* gene, a highly conserved gene among different species, consists of 17 exons spanning ~38 kb of Xq27.3 (Ashley et al., 1993; Eichler et al., 1993; Jin and Warren, 2000; Oostra and Chiurazzi, 2001; Schaeffer et al., 2003). The CGG repeat is 250 bp distal to the fragile X-associated CpG island (Fu et al., 1991; Verkerk et al., 1991). Expansion and subsequent methylation can occur in both the CpG island promoter region and the CGG repeat itself (Murray et al., 1997; Bontekoe et al., 2001). This methylation blocks transcription of the *FMR1* gene, thus preventing protein production and leading to the fragile X phenotype (Murray et al., 1996, 1997; Bontekoe et al., 2001; Oostra and Chiurazzi, 2001; Schaeffer et al., 2003).

The *FMR1* gene can transcript into the 4.4 kb full length mRNA that encodes for the fragile X mental retardation protein (FMRP) with a maximum length of 632 amino acids and molecular mass of 70-80 KDa (Oostra and Chiurazzi, 2001). Analysis of the amino acid sequence of FMRP revealed the presence of two types of RNA-binging motif, two ribonucleoprotein K homology domains (KH domains), a nuclear localization signal (NLS), a nuclear export signal (NES) and clusters of arginine and glycine residues (RGG boxes), suggesting that the FMRP is an RNA-binding protein which shuttles between the nucleus and cytoplasm (Siomi et al., 1993, 1994; Jin and Warren, 2000; Schaeffer et al., 2003; Mandel et al., 2004). FMRPs, present mainly in the cytoplasm of the cell, bind the mRNA and other protein to form large messenger ribonucleoprotein complexes (mRNP) together with the polyribosomes, indicating that the FMRP is involved in posttranscriptional control of gene expression through such mechanisms as the export of mRNA from the nucleus to the cytoplasm, the regulation of mRNA translation including its own mRNA, and the control of mRNA stability and decay (Oostra and Chiurazzi, 2001; Schaeffer et al., 2003; Mandel et al., 2004). Multiple FMRP isoforms are produced by alternative splicing. FMRPs are found in a variety of

tissues and organs, but they are predominantly found in the central nervous system (CNS) and testis. The absence of FMRPs during synaptic development may lead to MR because anomalies in the number and morphology of dendritic spines appear in the brain of fragile X patients.

As described previously, the main cause of mutation in FXS patients (>95%) is the expansion of the CGG repeat. Such repeat expansions occur during meiosis or early embryonic development (Bontekoe et al., 2001). Both premutations and full mutations have been found to be meiotically and mitotically unstable, indicating mosaicism (Fu et al., 1991). The other mutations within the *FMR1* coding sequence are deletions and point mutation that are found in less than 1% of FXS mutations (de Boulle et al., 1993; Hammond et al., 1997; Oostra and Chiurazzi, 2001; Schaeffer et al., 2003). Due to the absence of FMRPs, individuals with deletions and point mutations have a usual FXS phenotype but they do not express the fragile site. This shows that the *FMR1* gene is the only gene involved in the pathogenesis of fragile X syndrome (Jin and Warren, 2000).

The CGG repeat alleles can be categorized into 5 groups, normal, intermediate or borderline, premutation, full mutation and mosaicism. Normal alleles have a range of 6-40 with 29, 30 and 36 CGG repeats as the most common alleles in the Thai population (Fu et al., 1991; Brown et al., 1996; Kunst et al., 1996; Murray et al., 1996; Rerkamnuaychoke et al., 1998; Limprasert et al., 1999). These alleles are stable upon transmission between generations (Bontekoe et al., 2001; Maddalena et al., 2001). Intermediate or borderline alleles have a range of 41-59 repeats. Most intermediate alleles are stable on transmission like normal alleles; however 10-30% of these alleles may be unstable, similar to premutation alleles. When these alleles are passed to the next generation, minor increases and decreases in the repeat number can occasionally occur (Maddalena et al., 2001; Nolin et al., 2003). Women with intermediate alleles can be considered normal or premutation, depending on the phenotype of their children. Premutation alleles, range from 60-200 repeats, are unstable during transmission to the next generation in both expansion to a different size of premutation or a full mutation and contraction. Because they are not hypermethylated with normal protein level expression, they are not associated with mental retardation, resulting in no phenotypic effect. Women with alleles of 60 repeats or higher are considered to be at risk for having affected children (Heitz et al., 1992; Turner et al., 1994; Murray et al., 1997; Jin and Warren, 2000; Bontekoe et al., 2001; Maddalena et al., 2001; Schaeffer et al., 2003). Full mutations typically have several hundred to several thousand repeats (over 200 repeats). They

undergo abnormally CpG island hypermethylation as well as resulting in the suppression of *FMR1* transcription and the absence of FMRP translation, leading then to the fragile X phenotypes (Bell et al., 1991; Heitz et al., 1991; Oberle et al., 1991; Pieretti et al., 1991; Jin and Warren, 2000; Bontekoe et al., 2001; Maddalena et al., 2001; Schaeffer et al., 2003). Additionally, 15-20% of affected individuals feature mosaicism, which is observed in both repeat size mosaicism and methylation mosaicism. Repeat size mosaics are referred to as a subpopulation of both full mutation and premutation or normal alleles. Methylation mosaics referred to full mutation with remaining unmethylation in subpopulations may have a higher level of intelligence than affected individuals with complete methylation due to probably the production of some FMRPs from the unmethylated allele (McCoukie-Rosell et al., 1993; Hagerman et al., 1994; Nolin et al., 1994; Sweets et al., 1995; Orrico et al., 1998; Schmucker et al., 1999; Maddalena et al., 2001).

Mutation analysis of fragile X syndrome requires the use of more than one methodology because no one technical approach can detect all mutation types in the *FMR1* gene equally well or with equal precision. The polymerase chain reaction (PCR) is a specific method for determination of the numbers of CGG trinucleotide repeats within the *FMR1* gene. Although this method has high sensitivity for estimation of many allele sizes in the normal and lower premutation range, it may fail to detect allele sizes in the upper premutation range and full mutation alleles. Rare patients who have cellular mosaicism for the CGG repeat may give the PCR result in the normal, intermediate, or premutation range, leading to a potential false negative misdiagnosis. If such a false negative is suspected, Southern blot analysis is used, which can detect the presence of all repeat expansion mutations including both large premutations and full mutatons, but it offers only a low resolution for estimation of CGG repeat number. Although Southern blot can also be used to determine the methylation status, it is time-consuming, expensive, labour intensive and requires a substantial sample of DNA, and a superior method, the methylation specific PCR (MS-PCR), has been developed for this latter function. Cytogenetic analysis using a folate and/or thymidine depletion culture medium to induce fragile sites is no longer an acceptable diagnostic method because both specificity and sensitivity are insufficient. Immunohistochemical staining for FMRP is a valid diagnostic method for protein analysis. Because severity of the fragile X syndrome phenotype is associated with FMRP expression, assessment of FMRP production in some patients has been proposed as a potential

prognostic indicator of disease severity (Maddalena et al., 2001; Saul and Tarleton, 2004).

The rate of expansion of the CGG repeat depends on its initial size and the purity of the repeat. The CGG repeat at FRAXA is not pure but it is frequently interrupted by the AGG triplets (Eichler et al., 1994, 1995; Hirst et al., 1994; Kunst and Warren, 1994; Macpherson et al., 1994; Murray et al., 1997; Snow et al., 1994; Zhong et al., 1995). In stable normal alleles, the CGG region is interspersed with two AGG interruptions after every 9 or 10 CGG repeats and the longest tract of uninterrupted CGG repeats is usually found at the 3′ end of the repeat (Eichler et al., 1994; Murray et al., 1997; Bontekoe et al., 2001; Maddalena et al., 2001). The $(CGG)_6AGG$ pattern is found particularly in Asian populations (Chen et al., 1997; Larsen et al., 1999). The lack of the most distal 3′-AGG interruption is an important mutational event in the generation of unstable alleles predisposing to premutation and full mutation, especially when tracts contain 34 perfect CGG repeats (Eichler et al., 1994; Jin and Warren, 2000; Oostra and Chiurazzi, 2001). Repeat instability is also known to have an effect on the absence of 5′ AGG interspersions that is a novel factor in several clearly known factors including CGG repeat length, AGG interspersion patterns and the DXS548-FRAXAC1 markers associated haplotype (Penagarikano et al., 2004). Most premutation alleles have either no AGG or one AGG interruption at the 5′ end of the repeat and the 3′ CGG tract is greater than 34 CGG triplets (Snow et al., 1993, 1994; Eichler et al., 1994; Zhong et al., 1995; Jin and Warren, 2000; Bontekoe et al., 2001). Analysis of premutation alleles in FXS carriers reveals that 70% of these alleles contain one AGG interruption (Eichler et al., 1994). It shows that the premutation is unstable and may expand to a full mutation (Jin and Warren, 2000). Therefore, the AGG substructure of the *FMR1* plays a crucial role in the maintenance of repeat stability because the AGG triplets are thought to anchor the region during replication, preventing replication slippage (Eichler et al., 1996; Jin and Warren, 2000; Maddalena et al., 2001).

Results from previous investigations of haplotype association studies using microsatellite markers, DXS548 and FRAXAC1, which are located ~150 Kb and ~7 Kb proximal to the CGG repeat region, respectively, have demonstrated that linkage disequilibrium between the fragile X mutations and the specific alleles of these adjacent polymorphic markers has been found in various Caucasian populations (Richards et al., 1992; Buyle et al., 1993; Hirst et al., 1993; Oudet et al., 1993; Kunst and Warren, 1994; Macpherson et al., 1994; Richards et al., 1994; Chiurazzi et al., 1996a, 1996b;

Syrrou et al., 1996; Eichler et al., 1996; Zhong et al., 1996; Bonaventure et al., 1998; Larsen et al., 1999; Mingroni-Netto et al., 1999; Pekarik et al., 1999; Tzeng et al., 2005). FXS patients are associated with several haplotypes, however only a few founder haplotypes account for most of the fragile X chromosomes with statistically significant differences from normal chromosomes suggesting a founder effect of the fragile X mutation (Buyle et al., 1993; Chiurazzi et al., 1996; Limprasert et al., 2001). Due to the discrepancy in nomenclature of microsatellite alleles from many independent research groups, collation and comparison of the increasing data among various ethnic groups is quite difficult. The comparisons of nomenclature of microsatellite alleles were shown in Appendix A, Tables 20 and 21 (Richards et al., 1992; Jacobs et al., 1993; Oudet et al., 1993; Macpherson et al., 1994; Rousseau et al., 1995; Chiurazzi et al., 1996; Eichler et al., 1996; Chiurazzi et al., 1999; Mingroni-Netto et al., 1999; Limprasert et al., 2001; Tzeng et al., 2005). Limprasert et al. (2001) studied haplotype analysis at the *FRAXA* locus in Thai subjects, based on 3 microsatellites, DXS548, FRAXAC1 and FRAXE, and 2 single nucleotide polymorphisms, ATL1 and IVS10+14C/T. They observed that most diverse haplotypes came from different FRAXE alleles; hence they analyzed haplotypes featuring the 4 remaining markers, DXS548-FRAXAC1-ATL1-IVS10+14C/T. They found no specific haplotype association when comparing normal control and FXS groups. However, they found a specific SNPs haplotype association in the 30 CGG repeat group.

Single nucleotide polymorphisms (SNPs) are the most common type of genetic variation in the human genome, occurring about one in every 1,200 nucleotides on average. This genetic trait explains the basis of heritable variation in disease susceptibility. SNPs can be used in modern medicine as genetic markers for identifying disease genes by linkage studies in families, linkage disequilibrium in isolated populations, association analysis of patients and controls, loss of heterozygosity studies in tumors and studying genetic factors associated with complex disease (Wang et al., 1998; Brightwell et al., 2002). SNPs were used as alternative polymorphic markers in this study because they have several advantages over microsatellites. SNP alleles remain in the population without sustaining recurrent mutation much longer than microsatellite alleles do, and also they have low mutation rates (Mathews et al., 2001; Brightwell et al., 2002). In addition SNPs have two alternative bases (biallelic), so statistical computer programs can facilitate SNP data analysis. Individual SNPs are less informative, however, than microsatellite markers; the usefulness of SNP markers is primarily in situations where a large number of them can be used in an appropriate population size (Wang et al., 1998; Mathews et al., 2001;

Brightwell et al., 2002). SNPs are present in the genome at high density, and large collections of mapped SNPs provide powerful tools for a wide range of human genetic studies. Brightwell et al. (2002) generated an SNP map for the FRAX region of the X chromosome, covering the region of the X chromosome approximately 1 Mb proximal and 2 Mb distal to the FRAXA-CGG repeat. They identified 54 polymorphisms, including 51 SNPs and 3 insertion/deletions distributed over 2.6 Mb. This represented about 1 SNP every 1024 bp. Of 51 SNPs, only 12 SNP sites were actually in the *FMR1* gene (WEX8, WEX3, WEX1, WEX4, WEX5, WEX6, WEX20, WEX16, WEX17, WEX18, WEX51 and WEX52). Of the 12 SNPs in the *FMR1* gene, 7 novel SNPs and 5 SNPs are in the NCBI SNP database. In addition, SNP databases, which can be found at the NCBI web site (www.ncbi.nlm.nih.gov/SNP), show 112 SNPs distributed along the *FMR1* gene, most of them within intron. These reveal that there is a high-density SNP map for the FRAXA region, as summarized in Appendix A, Table 19.

There are many techniques for SNP genotyping. For example, PCR-RFLP (restriction fragment length polymorphism) involves appropriate restriction endonuclease digestion of PCR products and finally gel electrophoresis. Dot blot hybridization with an ASO (allele-specific oligonucleotide) probe can also be used for typing, but it is a somewhat complicated process, with several steps of hybridization, washing and detection after PCR amplification of the target SNP alleles (Hamajima, 2001; Ye et al., 2001). The biallelic-ARMs PCR technique, also named tetra-primer ARMS-PCR, SB-ASA (single-tube bi-directional amplification) or PCR-CTPP (PCR with confronting two-pair primers), is a rapid and reliable method for SNP haplotype analysis experiments (Hamajima, 2001; Jiang Z et al., 2001; Ye et al., 2001; Brightwell et al., 2002). This is a more economical method than the original ARMS (amplification refractory mutation system) because of its capability to detect 2 different SNP alleles simultaneously in a single PCR reaction. It is therefore a suitable technique of SNP scoring in a large population and can provide a wide range of SNPs for a research laboratory (Hamajima, 2001; Ye et al., 2001; Brightwell et al., 2002). The principal of the Biallelic-ARMS method, based on allele specific PCR, involves the lack of proofreading activity of DNA polymerase and the 3′-terminal base of allele specific primers (Newton CR et al., 1989; Jiang Z et al., 2001). Biallelic-ARMS employs 2 primer pairs comprising 2 control primers and 2 allele specific primers (ASPs) in the same PCR system. The primer pairs are designed for amplification of individual SNP alleles, especially introducing the last 3′ mismatched base of the ASPs (complementary to one of the two alleles). Such ASPs prime with target sequences at the site of the SNP in

opposite directions, so PCR amplification with ASPs and the opposing control primers can occur and generate two allele specific fragments different in size. The SNP genotype is identified on the length of amplified products of each allele after gel electrophoresis. Besides, both control primers also produce the largest fragments of PCR reaction called the control products, which can also serve as internal controls for accurate SNP typing (Hamajima, 2001; Jiang Z et al., 2001; Ye et al., 2001; Brightwell et al., 2002).

This study intend to examine not only the haplotype analysis using SNP markers that disperse through the *FMR1* gene but also AGG interruption patterns to provide insight into the molecular basis of FXS in Thai population.