CHAPTER IV

DISCUSSION

Molecular cloning and characterization of AcMRJPs cDNA

MRJPs cDNA have been first characterized in *Apis cerana* by constructed the expressed sequence tag (EST) library of nurse honeybee hypopharyngeal glands. The nucleotide sequence from sixty-six recombinant clones show that the AcMRJP1, AcMRJP2, AcMRJP3 and AcMRJP4 were expressed with different expression level in hypopharyngeal glands of *A. cerana* nurse bee. According to the number of sequenced recombinant clones found, the AcMRJP1-3 showed higher expression level than AcMRJP4. The AcMRJP5 was not found in 66 recombinant clones which has been sequenced (Srisuparbh, *et al.*, 2003). At present, the complete nucleotide sequence of AcMRJP1 (Srisuparbh, *et al.*, 2003), AcMRJP2 (Imjongjailuk, *et al.*, 2004) and AcMRJP3 (Srisuparbh, 2002) and partial nucleotide sequence of AcMRJP4 (Srisuparbh, *et al.*, 2003) were reported. For AcMRJPs, RJ of *A. cerana* was directly purified and characterized. Three types of AcMRJP (AcMRJP1, AcMRJP2 and AcMRJP3) could be purified from *A. cerana* RJ, whereas AcMRJP4 and AcMRJP5 were not found in the purification process (Srisuparbh, *et al.*, 2003).

In this study, the reverse transcriptase polymerase chain reaction (RT-PCR) was selected tool for the *in vitro* synthesis of AcMRJP4 and AcMRJP5 cDNA using RNA extracted from hypopharyngeal gland of *A. cerana* nurse bee. The MRJP4 and MRJP5 specific primers were designed from AmMRJPs cDNA sequence in Genbank Database. The PCR product sizes of 400, 1,500, 1,600, and 2,000 bp were amplified by primers for

AcMRJP4 cDNA. In addition, one PCR product (1,900 bp) was amplified by primers for AcMRJP5 cDNA. The PCR product size of 1,500, 1,600, 1,900, and 2,000 bp from both amplification reactions were cloned by ligated with pGem® - T easy vector. The ligation products were electro-transformed to *E. coli* JM 109 host. The recombinant plasmids were extracted and used for identified the cDNA insert. The cDNA inserts in recombinant plasmids were identified by restriction mapping and nucleotide sequencing. Recombinant plasmids containing 1,500, 1,600, 1,900 and 2,000 bp insert DNA were characterized by restriction enzyme *Eco*RI and *Ssp*I. Sizes of each fragments from digestion were compared with the size of fragments expected from AmMRJPs cDNA sequence. The result showed restriction map of recombinant plasmid containing 1,500, 1,600, 1,900 and 2,000 bp insert DNA related to restriction map of AmMRJP6, AmMRJP4, AmMRJP5 and AmMRJP3 respectively. The recombinant plasmid containing 2,000 bp cDNA insert was sequenced using M13 forward and M13 reverse primer. The recombinant plasmid containing other sizes of insert DNA were sequenced along the entire length of cDNA insert. The nucleotide sequence of cDNA insert were blast against GenBank database. The result of nucleotide sequences and restriction pattern of recombinant plasmids showed that recombinant plasmid containing 1,500, 1,600, 1,900 and 2,000 bp insert DNA were related to AmMRJP6, AmMRJP4, AmMRJP5 and AmMRJP3, respectively. These cDNA inserts were designated to AcMRJP6, AcMRJP4, AcMRJP5 and AcMRJP3, respectively.

The result showed AcMRJP3, AcMRJP4, AcMRJP5 and AcMRJP6 were expressed in hypopharyngeal gland of *A. cerana* nurse bee. This result corresponded with the expression of MRJPs in *A. mellifera*. In *A. mellifera*, the expression of AmMRJP3

was localized in hypopharyngeal gland of nurse honeybee (Ohashi, *et al.*, 1997). And the AmMRJP4, AmMRJP5 and AmMRJP6 cDNA was identified from the head cDNA library of nurse bee (Klaudiny, *et al.*, 1994, Albert, *et al.*, 1999a, Albert and Klaudiny, 2004).

The complete nucleotide sequence of AcMRJP4, AcMRJP5 and AcMRJP6 cDNA were derived from three nucleotide sequences of recombinant clones obtained from RT-PCR cloning while partial nucleotide sequence of AcMRJP3 cDNA was obtained from one clone.

Partial nucleotide sequence of AcMRJP3 cDNA were almost 100% identity in overlap sequence with the sequence of AcMRJP3 cDNA ORF previously report from the same group (Srisuparbh, *et al.*, 2002).

In overlap nucleotide sequence of both AcMRJP3 cDNA, six positions of single nucleotide substitution or so call single nucleotide polymorphism (SNP) were observed, excepted of two position nucleotide different which under control of the primer. Four of six SNPs altered the amino acid in the encoded protein. This result correlated with the observed in *A. mellifera*, the SNPs were observed in AmMRJP3, AmMRJP4 and AmMRJP5 (Albert and Klaudiny, 2004). Moreover, nucleotide sequence of the 3' untranslated region (3' UTR) of AcMRJP3 was obtained in this study. The unique characteristic of polyadenylation signal sequence that contained two separated or three partially overlapping consensus sequences was found only in AcMRJP3 cDNA sequence. The same characteristic was found in AmMRJP3 at the same position (Klaudiny, *et al.*, 1994).

Nucleotide sequence of AcMRJP4, AcMRJP5 and AcMRJP6 cDNAs showed high homology with those of AmMRJP4 (89%), AmMRJP5 (91%) and AmMRJP6 (92%), respectively. The identical consensus polyadenylation signal sequence was observed almost at the same position in these cDNA sequences.

The sixteen nucleotides sequence from start codon of AcMRJP3, AcMRJP4 and AcMRJP6 cDNA sequence were in the primer sequence. So, the five to six amino acid residues deduced from the sequence were the same in AcMRJP3, AcMRJP4 and AcMRJP6. Some of them may differ from native protein. It was excepted in AcMRJP5 because the forward primer was designed before start codon.

The deduced amino acid of AcMRJP5 that inferred from their cDNA show the extensive repeat region located between amino acid residue position 367 and 540. The repeat unit was dominant in tripeptide DRM that occurred 51 times and interrupted a conserved region of the MRJP consensus sequence. It located deep inside the protein molecule (Figure 3.17). The AcMRJP5 repetitive region was compared with the MRJP5 repetitive region previously reported in *A. dorsata* and *A. mellifera*. The result showed the repetitive region of these 3 species was located at the same position but differ in repeat unit sizes. The repeat unit in *A. cerana*, *A. dorsata* and *A. mellifera* occurred 51 times, 23 times and 58 times, respectively (Albert, *et al.*, 1999a, Albert, *et al.*, 2002).

The polymorphism of AcMRJP5 was found in this study. It variant in repeat unit sizes (Appendix D) like that observed in AmMRJP3 repetitive region (Albert, *et al.*, 1999b).

The duduced amino acid sequence of AcMRJP4-6 inferred from their cDNAs were used to predict for signal peptidase clevage site by using SignalP V2.0.b2 programe.

The predicted N-terminal amino acid sequence of AcMRJP4, AcMRJP5 and AcMRJP6 was AVVRENSSRK, ATVRENSSRN and AIHRRKSSKN, respectively. Comparison of the N-terminal amino acid sequence of MRJP4 and MRJP5 obtained from this study with those previously reported in *A. mellifera* and *Apis mellifera scutellate* species, the predicted N-terminal amino acid sequence of AcMRJP4 was 100% identical with the N-terminal amino acid sequence of Africanized honeybee (*Apis mellifera scutellate*) MRJP4 and differ only one amino acid residue where compared with N-terminal amino acid sequence of *A. mellifera* MRJP4 (GVVRENSSRK). The predicted N-terminal amino acid sequence of AcMRJP5 were corresponded with those in 2 honeybee species (*A. mellifera* and *A. mellifera scutellate*) that possed VTVRENSPRK (Sano, *et al.*, 2004), they differ in only three amino acid residues. Since MRJP6 could still not purified from all species of honeybee, therefore there is no report for the N-terminal amino acid sequence. Using the same criterial as those for AcMRJP4 and AcMRJP5, the predicted N-terminal amino acid sequence of AcMRJP6 should be AIHRRKSSKN. From this comparison can be proposed that the signal peptidase clevage site of AcMRJP6 may be correct position. It seems likely that this difference in N-terminal amino acid residue may originate from the species differences.

Deduced amino acid sequence after eliminated the signal peptide were used to predicted the putative N-glycosylation site, molecular weight (MW), theoretical pI and amino acid composition.

The number of putative N-glycosylation site of MRJP4 and MRJP5 in *A.cerana* were 7 and 5 sites compared with 8 and 4 sites in *A. mellifera*, respectively. Moreover, two putative N-linked glycosylation sites were found in AcMRJP6 that lower than 5 sites

observed in AmMRJP6. The *O*-linked glycosylation site was not found in any AcMRJPs like in *A. mellifera* (Schmitzova, *et al.*, 1998).

The predicted pI values of AcMRJP4, AcMRJP5 and AcMRJP6 were 5.84, 8.75 and 6.44, respectively. These were roughly comparable to values determined by IEF analysis of purified AcMRJPs like that reported in AcMRJP1-3. The pI values of purified AcMRJP1-3 were close to the pI values predicted from deduced amino acid of AcMRJP1-3 (Srisuparbh., 2002). In native protein, many factors effected the pI values of MRJP such as polymorphism, the SNPs and phosphorylation sites in the protein molecule.

The predicted MW of AcMRJP4, AcMRJP5 and AcMRJP6 were 52.8, 66.2 and 47.4 kDa, respectively. The MW of native proteins were expected to be larger than that of predicted MW, because they were reported to be glycoproteins. The attachment of the sugar chain to the protein can alter the size of this proteins (Schmitzova, *et al.*, 1998, Srisuparbh, *et al.*, 2003).

The homology deduced amino acid sequence from AcMRJPs cDNA showed high sequence homology among MRJPs sequence. The homologies were interrupted by the regions containing repetitive motif in AcMRJP3 and AcMRJP5 (Figure 3.17). From the sequence homologies it is possible to conclude that the AcMRJPs are members of one protein family.

All members of the MRJP family share an N-terminal hydrophobic sequence that functions as a cleavable signal peptide as well as putative N-linked glycosylation sites, suggesting that these proteins are also secreted from the cell.

The members of the AcMRJPs family posses some common structural feature: the conserved position of four cysteines; the presence of several blocks of conserved amino acid; and a highly hydrophilic character mainly in their C-termini. This indicates similarity in tertiary structures and the possibility of similar biological function. Structural difference could be expected for MRJP5 because of the long repetitive region located deeper inside the protein molecule. This structural features was found like in *A. mellifera* (Schmitzova, *et al.*, 1998).

It could presumed that AcMRJPs have nutritional function in honeybee larval food. The analysis of their amino acid composition showed that four of them (MRJP1, MRJP2, MRJP5 and MRJP6) contained a high amount of essential amino acids (45%-51.9%), comparable with other nutritional proteins [(casein (49.1%), chicken ovalbumin (51.6%) and quail ovalbumin (48.8%) (Schmitzova, *et al.*, 1998)], MRJP3 and MRJP4 possess a high amount of some essential amino acids; MRJP3 Arg (5.6%), Lys (6.6%) and MRJP4 Leu (8.6%) , Val (7.7%).

The amino acids composition of AcMRJPs, and their dominant content in RJ, indicated that they together represent a balanced mixture of the amino acid essential for nourishing both honeybee larvae and the queen. Some of MRJPs (MRJP3 and MRJP4) that contain lower amounts of essential amino acids and exist in RJ with low content may play other roles in the honeybee physiology. Kucharski, *et al.* (1998) reported that MRJPs may had an other physiological function since AmMRJP1 was also found to expressed in honeybee brain.

Three new members (MRJP4, MRJP5 and MRJP6) of the *A. cerana* MRJP cDNA were identified in this study. However, there proteins data had not been studied since

these proteins could not be detected and purified, because they existed in very low content. This corresponded to the result reported by Klaudiny, *et al.*, (1994) that AmMRJP4 mRNA was expressed in low level (2% of total mRNA). Two-dimensional gel electrophoresis of RJ follow by N-terminal amino acid analysis was a suitable method for studies these proteins that had the lower amounts in RJ (Sano, *et al.*, 2004). In addition, more extensive cDNA screening and sequencing could reveal new members of this protein family in *A. cerana*.

The origin of the MRJP family have been studied in *A. mellifera.* AmMRJP family showed similarity with yellow protein of *Drosophila melanogaster.* In structural features, four cysteines that conserved in MRJP also found in yellow protein (Albert, *et al.*, 1999a). It seem that MRJPs diverged from yellow protein to gain a novel nutritional function in honeybee. Recently, *A. mellifera* yellow protein was identified (Albert and Klaudiny., 2004). The nucleotide sequence and deduced amino acid of *A. mellifera* yellow cDNA was selected for phylogenetic analysis with AcMRJP cDNA (Figure 3.18-3.19)

The average sequence divergence between the same MRJP family from different bee species (e.g. MRJP1 of *A. cerana* and *A. mellifera*) was lower than that between divergence calculated from pairs of different protein family within the same species (e.g. comparisons between AmMRJP1-AmMRJP2, AcMRJP1-AcMRJP2 etc.). The bootstrapped phylogenetic tree revealed closer relationships between the same RJ protein families. The phylogenetic trees constructed from both nucleotide and amino acid sequence show the identical tree, it supported the occurrence of MRJPs through gene

duplication. The MRJP8 exhibited the earliest divergence within MRJPs gene families whereas MRJP3 and MRJP5 exhibited the recent divergence.

Overexpression of AcMRJP4 protein in *E. coli*

At the beginning of this study, MRJP4 cDNA have been identified in cDNA library of *A. mellifera* nurse honeybee head. But protein (MRJP4) was not identified in RJ. Therefore, the MRJP4 was selected for overexpression studied. However, the *E. coli* system was used in this study because it easily to culture, time saver and produced large amount of target protein. The pET system was used as cloning vector. The recombinant plasmids pET 19b contained AcMRJP4 cDNA were constructed and introduced to the expression host (*E. coli* Rosetta (DE3) pLysS). This host strain was designed to enhance the expression of eukaryotic proteins that contained codons rarely used in *E. coli* by supply the tRNA genes.

The production of the recombinant protein was observed at 1 hour after induction and showed highest expression level at 3 hours after induction with IPTG. The molecular weight of induced protein band was 53 kDa corresponded to the calculated molecular weight of recombinant AcMRJP4 protein deduced from DNA sequence (55.7 kDa). It can presumed that the induced protein was the recombinant AcMRJP4 protein. The His-tags in this recombinant protein were used for further purification and identification of this protein. Finally, the result showed that this system suitable for produced AcMRJP4 protein.