# CHAPTER III

# RESULTS

## 3.1 Cloning and characterization of AcMRJPs cDNA

The full length of AcMRJP4 and AcMRJP5 cDNA family was prepared by PCR of the first stranded cDNA with specific primers which designed from AmMRJPs cDNA sequence in GenBank Database. A pair of primers for AcMRJP4 cDNA amplification were 5'-ATA TCC TAG AAA AAA AAT GAC AAA ATG GTT GC-3' (MRJP4) and 5'-GGG GTA CCC TTT TTT TTT TTT TTT TTT TTT TTT TTT TA-3' (RMJ2). A pair of primers for AcMRJP5 cDNA amplification were 5'-CTG TCG TTT GCA AAA TAT TTG CAG C-3' (MRJP5_2) and 5'-GGG GTA CCC TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TA-3' (RMJ2) (Table 2.1 and Figure 2.1). The expected full length of AcMRJP4 and AcMRJP5 cDNA were about 1,600 bp and 1,900 bp as estimated from their corresponding size of AmMRJPs cDNA.

PCR product of 4 different sizes were detected when amplified by primers for AcMRJP4 cDNA (Figure 3.1). The sizes of PCR products were 400, 1,500, 1,600 and 2,000 bp in length. The PCR product size of 1,500, 1,600 and 2,000 bp were recovered from agarose gel by QIAquick gel extraction kit and used for cloning. PCR products were ligated with pGEM®-T easy vector and electro-transformed to *E. coli* JM 109 host.

Amplification of first stranded cDNAs with specific primers (MRJP5_2 and RMJ2) of AcMRJP5 cDNA, only one PCR product of 1,900 bp was obtained (Figure 3.2). The PCR product size of 1,900 bp was the expected size as calculated from AmMRJP5 cDNA. This PCR product was then clone as described before.

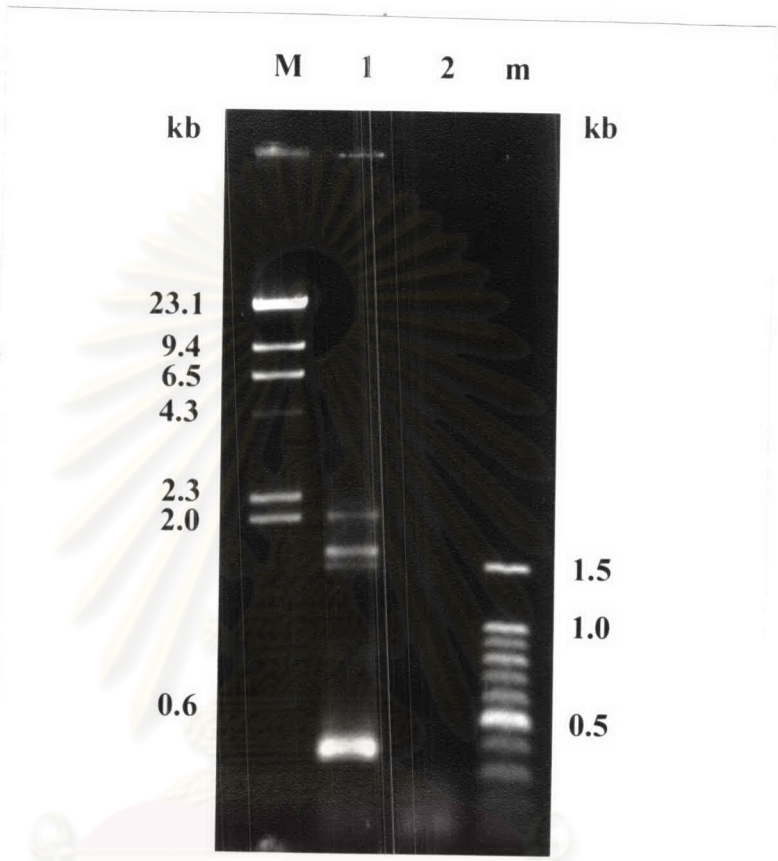


Figure 3.1 PCR amplification for full length cDNA of AcMRJP4

Lane M = λ / *Hin*d III standard molecular weight marker

Lane 1 = The amplification products of first strand cDNA

Lane 2 = Negative control

Lane m = A 100 bp DNA ladder

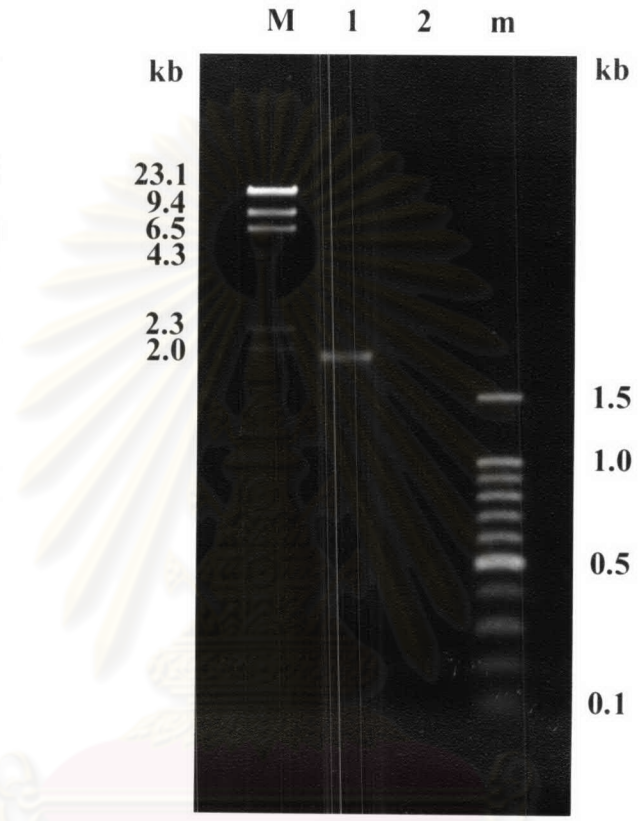


Figure 3.2 PCR amplification for full length cDNA of AcMRJP5

Lane M = λ / *Hin*d III standard molecular weight marker

Lane 1 = The amplification products of first strand cDNA

Lane 2 = Negative control

Lane m = A 100 bp DNA ladder

The recombinant clones were screened by blue-white screening on selective plate. The white colonies were randomly picked for culture. The recombinant plasmids were extracted from the cell culture. After plasmid extraction, recombinant plasmids were mapped by digestion with restriction endonucleases. If they contained MRJP cDNA sequence, they should have the restriction map related to restriction map predicted from *A. mellifera* MRJPs cDNA. The restriction enzymes *Eco*RI and *Ssp*I were selected for characterized the MRJP cDNA sequence. The pGEM®-T easy vector multiple cloning region was flanked by recognition sites for the restriction enzyme *Eco*RI. For AmMRJPs sequence, the restriction enzyme *Eco*RI could not cut any AmMRJP cDNA sequence except in AmMRJP6. There is 1 recognition site of *Eco*RI in AmMRJP6 cDNA, after digestion 2 fragments of 324 and 1125 bp will be obtained. Thus, the recombinant plasmids containing other AmMRJP cDNA, restriction enzyme *Eco*RI can be used to determined size of the insert. The pGEM®-T easy vector has 2 *Ssp*I recognition sites and all AmMRJP cDNA sequences also contain *Ssp*I recognition sites. The position of recognition sites for the restriction enzyme *Ssp*I are differed in each AmMRJP cDNA sequence. So, it can be used to characterized the family of MRJP cDNA. Restriction fragments size of various recombinant plasmids of pGEM®-T AmMRJP3-6 cDNA after digestion with *Ssp*I are shown in Table 3.1.

Recombinant plasmids containing 1,500, 1,600, 1,900 and 2,000 bp insert DNA were characterized by restriction endonucleases digestion with *Eco*RI and *Ssp*I (Figure 3.3, 3.4 and 3.5). The recombinant plasmid containing 1,500 bp insert DNA, when digested with *Eco*RI two fragments of 380 and 950 bp was obtained. In addition, the digested product of restriction enzymes *Ssp*I was 790, 1200 and 2300 bp (Figure 3.3).

Comparison of size of digested product to those of AmMRJP cDNA (Table 3.1). It showed that 1500 bp cDNA insert was most likely be AcMRJP6 cDNA.

Digestion of the recombinant plasmids containing 1,600 bp insert DNA with restriction enzymes *Eco*RI, two DNA fragments size of 1,600 and 3,000 bp was obtained. It showed that *Eco*RI can not cut within the cDNA insert. In addition, the digested product of restriction enzymes *Ssp*I was 410, 800, 1150 and 2200 bp (Figure 3.4). These digested products sizes were similar to those obtained from AmMRJP4 cDNA (Table 3.1). Therefore, it was mostly that the 1600 bp cDNA insert was AcMRJP4.

The recombinant plasmids containing 1,900 bp insert DNA were digested with restriction enzymes *Eco*RI. The digested products size of 1,900 and 3,000 bp were detected. It showed that restriction enzymes *Eco*RI can not cut within the cDNA insert. In addition, the digested products size of 790, 1850 and 2150 bp were obtained after digested with *Ssp*I (Figure 3.5). The sizes of digested product were compared with those of AmMRJP cDNA (Table 3.1). The result showed that 1,900 bp cDNA insert might be AcMRJP5.

The recombinant plasmids containing 2,000 bp insert DNA were digested with restriction enzymes *Eco*RI. The digested products size of 2,000 and 3,000 bp were detected. It showed that restriction enzymes *Eco*RI can not cut within the cDNA insert. In addition, the digested products size of 1300, 1400 and 2150 bp were obtained after digested with *Ssp*I. The sizes of digested product were compared with those of AmMRJP cDNA (Table 3.1). The result showed that 2,000 bp cDNA insert might be AcMRJP3.

These four types of recombinant plasmids were further identified by DNA sequencing using M13 forward and M13 reverse primers. The nucleotide sequences were

compared with the DNA sequences deposited in the GenBank database. The results of nucleotide sequence and restriction pattern of recombinant plasmids showed that 1,500, 1,600, 1,900 and 2,000 bp cDNA insert were AcMRJP6, AcMRJP4, AcMRJP5 and AcMRJP3 respectively.

**AcMRJP3**

Recombinant plasmid containing 2,000 bp cDNA insert from one clone was sequenced using M13 forward and M13 reverse primer. Nucleotide sequences obtained (Figure 3.6) was almost 100% identical with AcMRJP3 cDNA previously reported by Srisuparbh (2002). The single nucleotide substitutions also called Single Nucleotide Polymorphism (SNP) were found in this nucleotide sequence. These changes leaded to both silent or non-silent substitution. From deduced amino acid sequence, amino acid residue at position 66, 529, 533 and 536 were change when compared with those of Srisuparbh's AcMRJP3 cDNA (Figure 3.7).

The consensus polyadenylation signal sequences were observed. The sequences AATAAATAAAATAAA contained two separated or three partially overlapping consensus polyadenylation signal sequences (AATAAA) is located 14 bp upstream from the poly(A) tail.

Table 3.1 Restriction map of AmMRJP3-AmMRJP6 cDNA in pGEM®-T easy vector digested with restriction enzyme *Ssp*I

| Family | Direction of insert DNA | Predicted size of digested DNA fragment (bp) | Reference |
|---|---|---|---|
| MRJP3 | + | 182, 1159, 1250, 2256 | Klaudiny *et al.* (1994) |
| MRJP3 | - | 182, 783, 1159, 2723 | Klaudiny *et al.* (1994) |
| MRJP4 | + | 182, 391, 732, 1072, 2228 | Klaudiny *et al.* (1994) |
| MRJP4 | - | 182, 391, 755, 1072, 2205 | Klaudiny *et al.* (1994) |
| MRJP5 | + | 42, 182, 717, 1875, 2167 | Albert *et al.* (1999a) |
| MRJP5 | - | 42, 182, 694, 1875, 2190 | Albert *et al.* (1999a) |
| MRJP6 | + | 84, 182, 842, 1128, 2210 | Albert *et al.* (2004) |
| MRJP6 | - | 84, 182, 737, 842, 2601 | Albert *et al.* (2004) |

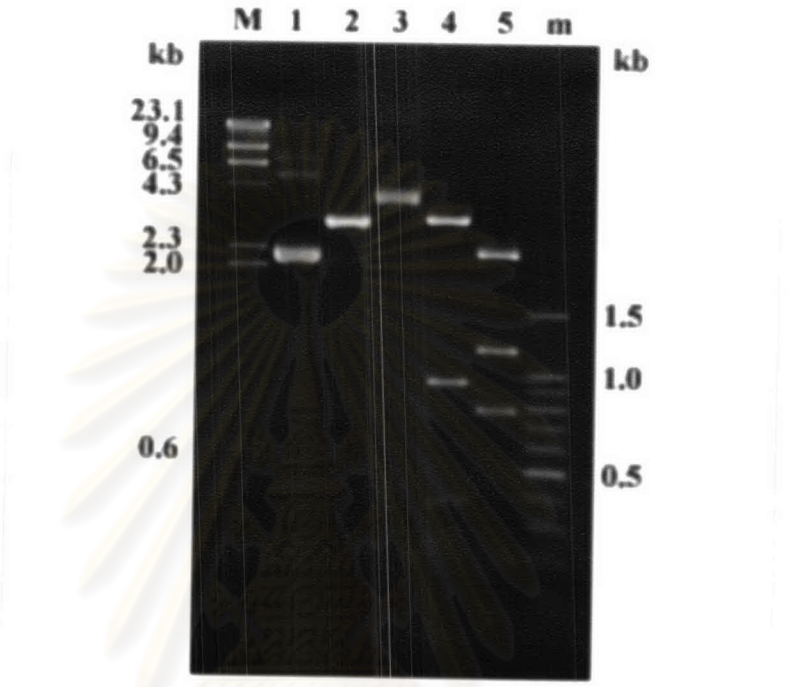


Figure 3.3 Restriction analysis of recombinant plasmid containing 1,500 bp cDNA insert.

Lane M = A λ/*Hin*d III standard DNA marker

Lane 1 = Undigested pGEM

Lane 2 = pGEM digested with *Eco*R I

Lane 3 = Undigested recombinant plasmid

Lane 4 = Recombinant plasmid digested with *Eco*R I

Lane 5 = Recombinant plasmid digested with *Ssp* I

Lane m = A 100 bp DNA ladder

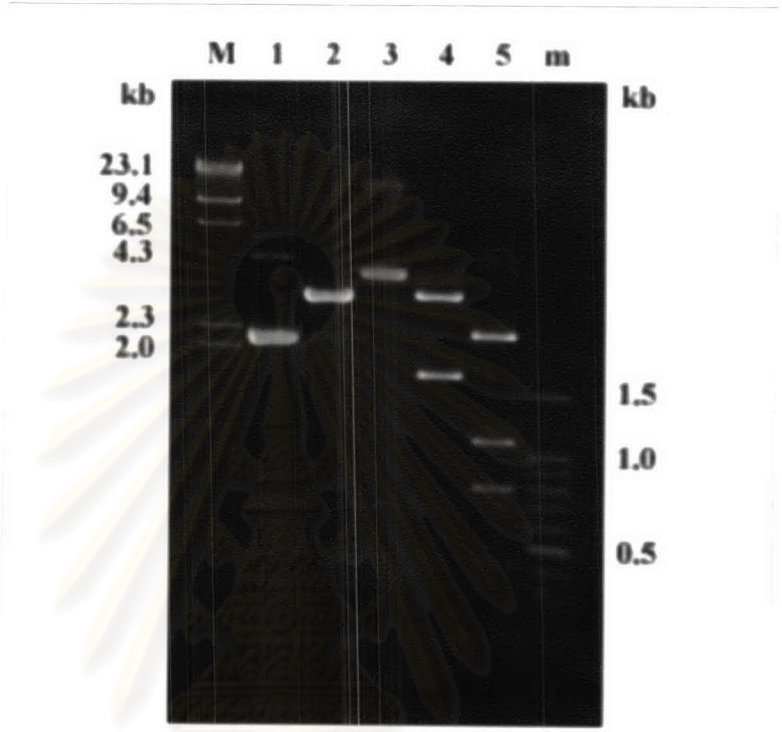


Figure 3.4 Restriction analysis of recombinant plasmid containing 1,600 bp cDNA insert.

Lane M = A λ/*Hin*d III standard DNA marker

Lane 1 = Undigested pGEM

Lane 2 = pGEM digested with *Eco*R I

Lane 3 = Undigested recombinant plasmid

Lane 4 = Recombinant plasmid digested with *Eco*R I

Lane 5 = Recombinant plasmid digested with *Ssp* I

Lane m = A 100 bp DNA ladder

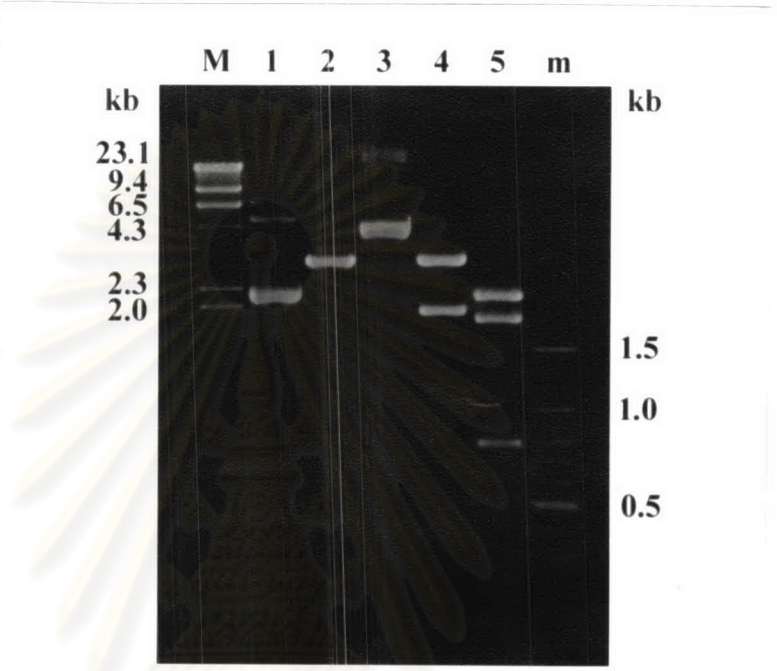


Figure 3.5 Restriction analysis of recombinant plasmid containing 1,900 bp cDNA insert.

Lane M = A λ/*Hin*d III standard DNA marker

Lane 1 = Undigested pGEM

Lane 2 = pGEM digested with *Eco*R I

Lane 3 = Undigested recombinant plasmid

Lane 4 = Recombinant plasmid digested with *Eco*R I

Lane 5 = Recombinant plasmid digested with *Ssp* I

Lane m = A 100 bp DNA ladder

```
CLUSTAL X (1.81) multiple sequence alignment

AcMRJP3         ATGACAAAGTGGTTGTTGCTGGTGGTGTGTCTTGGTATAGCTTGTCAAGATGTGACAAGC
2000bpRC        ATGACAAAATGGTTGCTGCTGGTGGTGTGTCTTGGTATAGCTTGTCAAGATGTGACAAGC
                ******** ****** **********************************************

AcMRJP3         GCAGCTGTGAACCATCAAAGAAAATCTTCAAAAAATTTGGCACATTCGATGAAGGTGATC
2000bpRC        GCAGCTGTGAACCATCAAAGAAAATCTTCAAAAAATTTGGCACATTCGATGAAGGTGATC
                ************************************************************

AcMRJP3         TACGAATGGAAACATATTGATTATGATTTTGGTAGCGTTGAAAGAAGAGATGCTGCGATT
2000bpRC        TACGAATGGAAACATATTGATTATGATTTCGGTAGCGTTGAAAGAAGAGATGCTGCGATT
                ***************************** ******************************

AcMRJP3         AAATCTGGCGAATTTGATCACACAAAAAATTACCCTTTCGATGTGGATAGATGGCGTGAT
2000bpRC        AAATCTGGCGAATTTAATCACACAAAAAATTACCCTTTCGATGTGGATAGATGGCGTGAT
                ************** ********************************************

AcMRJP3         AAGACATTTGTCACCGTAGAAAGGTTCGATGGTGTACCTTCTTCTTTGAACGTGGTAACT
2000bpRC        AAGACATTTGTCACCGTAGAAAGGTTCGATGGTGTACCTTCTTCTTTGAACGTGGTAACT
                ************************************************************

AcMRJP3         AATAAAAAAGGCAAAGGTGGACCTCTTCTACATCCATATCCTGATTGGTCGTGGGCGAAC
2000bpRC        AATAAAAAAGGCAAAGGTGGACCTCTTCTACATNNNNNNNNNNNNNNNNNNNNNNNNNN
                *********************************

AcMRJP3         TATAAAGATTGCTCTGGAATTGTGAGCGCTTTCAAAATTGCGGTCGACAAATTCGACAGA
2000bpRC        NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN

AcMRJP3         TTATGGGTTCTGGACTCAAGTCTTGTCAATAATAATCAACCCATGTGCTCTCCAAAATTG
2000bpRC        NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN

AcMRJP3         GTAACCTTCGATTTGAATACCTCAAAATTGCTTAAGCAAGTCGAGATACCACATAATATT
2000bpRC        NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN

AcMRJP3         GCCGTAAATGCCACCACCGAATGGGGAGAATTAGTATCACTAGCTGTTCAAGCTGTAGAT
2000bpRC        NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN

AcMRJP3         CCTACGAATACTATGGTGTACATAGCAGACGAAAGAGGTGAAGCTTCAATCATCTATCAA
2000bpRC        NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN

AcMRJP3         AATTCCGACGATTCCTTCCATCGATTGACTTCCAATACTTTCGATTACGATCCCAGATAT
2000bpRC        NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN

AcMRJP3         ACCAAACTGACAGTCGCTGGAGAAAGTTTCACAGTGAAAAATGGAATTTGTGGAATTGCA
2000bpRC        NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
```

(continued)

Figure 3.6 Alignment of the partial nucleotide sequence of recombinant plasmid containing 2,000 bp cDNA insert (2000bpRC) with AcMRJP3 cDNA (Srisuparbh, D., 2002). Conserve residues are indicated by asterisks.

```
AcMRJP3         CTTAGTCCCGTGACGAACAATCTTTATTACAGTCCTCTCGCTTCTCACAGTTTGTATTAT
2000bpRC        NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN

AcMRJP3         GTTAACACAGAACAATTCAGGAATCCACAATATGAAGAAAATAACGTCCAATATGAAGGA
2000bpRC        NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN

AcMRJP3         TCCCAAGATATTTTGAACACTCAATCATTCGCTAAAGCAGTATCGAAAAATGGCGTCGTT
2000bpRC        NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN

AcMRJP3         TTCTTGGGACTCGTGAGTAATTCAACTGTTGGCTGTGTGAATGAACATCAAGTACTTCAG
2000bpRC        NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN

AcMRJP3         AAAGAAAATTTTGATGTTGTCGCTCAGAATGAAGAGACACTTCAAATGATCGTTAGTATG
2000bpRC        NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN

AcMRJP3         AAAATCATGCAAGATCTTCCACAATCCGGCAGAATTAATGATCCAGGAAATGAATATATG
2000bpRC        NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN

AcMRJP3         TTGGCTTTAAGTAACAAAATGCAAAAAATAATAAACAATGATTTTAATTTCAACGACGTA
2000bpRC        NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN

AcMRJP3         AATTTCCGAATTTTGGGTGCGAATGTAAATCACTTAACAAGAAACACTCGTTGCGCAAAA
2000bpRC        NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN

AcMRJP3         TCTAATAATCAGAATGCTAACAATCAGAATGCTAACAATCAAAATGCTACCAATCAGAAT
2000bpRC        NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN

AcMRJP3         GATACCAACCAGAATGATAATGGTACCAACAGGAGGAATGGTAACAACCAAAATGGTAAC
2000bpRC        NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN

AcMRJP3         AGACAAAATGATAATAAACAGAATGATAACAAGCAGAATGCTAACAAGCAGAATGCTAAC
2000bpRC        NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN

AcMRJP3         AAGCAGAATGCTAACAAGCAAAATGATAACAAGCAAAATGATAACAAGCAAAATGGTAAC
2000bpRC        NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN

AcMRJP3         AGACAAAATGATAATAGGCAGAATGATAACAAGCAAAATGATAATAGGCAGAATGATAAC
2000bpRC        NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNCAAGCAAAATGATAATAGGCAGAATGATAAC
                                              ********************************

AcMRJP3         AAGCAAAATGGTAACAGACAAAATGGTAATAGACAGAATGATAACAAGCGGAATGGTAAC
2000bpRC        AAGCAAAATGGTAACAGACAAAATGATAATAGACAGAAAGATAACCAGCGGAATGGTAAC
                ************************* ************ ****** **************

AcMRJP3         AGGCAAAATGATAATAGACAGAATGATAACAAGCGGAATAGTAACAGGCAAAATGATAAT
2000bpRC        AGGCAAAATGATAATAGACAGAATGATAACAAGCGGAATAGTAACAGGCAAAATGATAAT
                ************************************************************

AcMRJP3         AGACAGAATGATAACAAGCGGAATGGAAACAGGCAAAATGATAACAAGCAAAATGATAAC
2000bpRC        AGACAGAATGATAACAAGCGGAATGGTAACAGGCAAAATGATAACAAGCAAAATGATAAC
                ************************* ********************************
```

Figure 3.6 (continued)

```
AcMRJP3         AAGCAAAATGATAACAGGCAGAATGATAACAATCAGAATGATAATCAGAATGATAATAAT
2000bpRC        AAGCAAAATGATAACAGGCAGAATGATAACAATCAGAATGATAATCAGAATGATAATAAT
                ************************************************************

AcMRJP3         CGAAATAATCAAGCTCATCATTCTTAA--------------------------------
2000bpRC        CGAAATAATCAAGCTCATCATTCTTAAAAATCACATTAAATCAATTAATTATCAATTAAA
                ***************************

AcMRJP3         ------------------------------------------------------------
2000bpRC        ATCAATTAATTAGGATGTAAACCAAATTATTTTTTAAAATATTTTTTCGATGTAAACAAA

AcMRJP3         ------------------------------------------------------------
2000bpRC        ATTTTTTTAAATCTTTCATTATATTATAAATAAATAAAATAAATATCGTTTTCGCATAAA

AcMRJP3         -------------------------
2000bpRC        AAAAAAAAAAAAAAAAAAAAAAAAA
```

Figure 3.6 (continued)

```
CLUSTAL X (1.81) multiple sequence alignment

AcMRJP3          MTKWLLLVVCLGIACQDVTSAAVNHQRKSSKNLAHSMKVIYEWKHIDYDFGSVERRDAAI
2000bpRC         MTKWLLLVVCLGIACQDVTSAAVNHQRKSSKNLAHSMKVIYEWKHIDYDFGSVERRDAAI
                 ************************************************************

AcMRJP3          KSGEFDHTKNYPFDVDRWRDKTFVTVERFDGVPSSLNVVTNKKGKGGPLLHPYPDWSWAN
2000bpRC         KSGEFNHTKNYPFDVDRWRDKTFVTVERFDGVPSSLNVVTNKKGKGGPLLH---------
                 *****:******************************************

AcMRJP3          YKDCSGIVSAFKIAVDKFDRLWVLDSSLVNNNQPMCSPKLVTFDLNTSKLLKQVEIPHNI
2000bpRC         ------------------------------------------------------------


AcMRJP3          AVNATTEWGELVSLAVQAVDPTNTMVYIADERGEASIIYQNSDDSFHRLTSNTFDYDPRY
2000bpRC         ------------------------------------------------------------


AcMRJP3          TKLTVAGESFTVKNGICGIALSPVTNNLYYSPLASHSLYYVNTEQFRNPQYEENNVQYEG
2000bpRC         ------------------------------------------------------------


AcMRJP3          SQDILNTQSFAKAVSKNGVVFLGLVSNSTVGCVNEHQVLQKENFDVVAQNEETLQMIVSM
2000bpRC         ------------------------------------------------------------


AcMRJP3          KIMQDLPQSGRINDPGNEYMLALSNKMQKIINNDFNFNDVNFRILGANVNHLTRNTRCAK
2000bpRC         ------------------------------------------------------------


AcMRJP3          SNNQNANNQNANNQNATNQNDTNQNDNGTNRRNGNNQNGNRQNDNKQNDNKQNANKQNAN
2000bpRC         ------------------------------------------------------------


AcMRJP3          KQNANKQNDNKQNDNKQNGNRQNDNRQNDNKQNDNRQNDNKQNGNRQNGNRQNDNKRNGN
2000bpRC         -------------------------------QNDNRQNDNKQNGNRQNDNRQKDNQRNGN
                                                ****************.***:**:****

AcMRJP3          RQNDNRQNDNKRNSNRQNDNRQNDNKRNGNRQNDNKQNDNKQNDNRQNDNNQNDNQNDNN
2000bpRC         RQNDNRQNDNKRNSNRQNDNRQNDNKRNGNRQNDNKQNDNKQNDNRQNDNNQNDNQNDNN
                 ************************************************************

AcMRJP3          RNNQAHHS
2000bpRC         RNNQAHHS
                 ********
```

Figure 3.7 Alignment of the partial deduced amino acid sequence of recombinant plasmid containing 2,000 bp cDNA insert (2000bpRC) with AcMRJP3 (Srisuparbh, D., 2002). Conserve residues are indicated by asterisks.

**AcMRJP4**

Recombinant plasmids containing 1,600 bp cDNA insert which was expected to be AcMRJP4 cDNA as study by digestion with restriction enzyme *Eco*RI and *Ssp*I. Further identification of this cDNA insert, the recombinant plasmids containing 1,600 bp cDNA insert were sequenced by four primers (M13 forward and M13 reverse, MRJP4_2 and MRJP4_4) that shown in Table 2.1 and Figure 2.1. Two internal sequencing primers (namely MRJP4_2 and MRJP4_4) were designed from nucleotide sequences obtained using M13 forward and M13 reverse primers. Recombinant plasmids from two clones were sequenced along the entire length. The nucleotide sequences of the insert from these two recombinant plasmids, partial sequence of the insert DNA from another clone and AcMRJP4 cDNA sequence retrived from EST library of *A. cerana* hypopharyngeal glands (GenBank Acc. CB350335) were assembled (Appendix C). The complete nucleotide sequence was compared with the DNA sequence deposited in GenBank database using nucleotide Blast (BlastN) and translated protein Blast (BlastX). The result showed that similar to AmMRJP4 cDNA. This sequence was most likely to be AcMRJP4 cDNA.

AcMRJP4 cDNA indicated a length of 1608 bp (including poly(A) tail) The sequence contained an open reading frame (nucleotides 1-1458) which encoded 485 amino acid residues (Figure 3.8). The sequence AATAAAATAAA containing two partially overlapping consensus polyadenylation signal sequences (AATAAA) was located 15 bp upstream from the poly(A) tail. The computer sequence analysis predicted that the signal peptidase cleavage site was located between Gly 20 and Ala 21. A comparison of AcMRJP4 and AmMRJP4 nucleotide sequences analysis by blast N

program revealed an identity of 89%. For blast X program protein sequence of AcMRJP4 was showed 79% sequence identity and 85% sequence similarity to AmMRJP4. But program was calculated by disregard partial amino acid sequence of the C-terminal. The deduced amino acid (without putative signal peptide) composition of AcMRJP4 comprised of 36.77% hydrophobic, 35.70% neutral and 27.53% hydrophilic amino acid residues. The essential amino acid content was 42.4%.

The pI value was estimated to be 5.84. The estimated molecular weight was 52.8 kDa. Seven putative *N*-glycosylation sites were found.

The nucleotide sequence and deduced amino acid sequence of AcMRJP4 cDNA were compared with the nucleotide sequence and deduced amino acid sequence of AmMRJP4 cDNA. The result of alignment was shown in Figure 3.9 and Figure 3.10.

```
ATGACAAAATGGTTGCTGTTGATGGCATGCCTTGGCATAGCTTGTCAAAATATTAGAGGT   60
 M  T  K  W  L  L  L  M  A  C  L  G  I  A  C  Q  N  I  R  G ▲

GCCGTTGTTCGAGAAAATTCCTCGAGAAAAAAATTAACAAATACGTTGAACGTGATTCAC  120
 A  V  V  R  E  N  S  S  R  K  K  L  T  N  T  L  N  V  I  H

GAATGGAAGTATGTCGATTATGATTTCGGTAGCGACGAAAAAAGGCAAGCTGCGATTCAA  180
 E  W  K  Y  V  D  Y  D  F  G  S  D  E  K  R  Q  A  A  I  Q

TCTGGCGAATATGATCGTACGAAAAATTATCCTCTTGACGTCGATCAATGGCATGATAAG  240
 S  G  E  Y  D  R  T  K  N  Y  P  L  D  V  D  Q  W  H  D  K

ACTTTTGTCACTATGTTAAGATACGATGGTGTGCCTTCCTCTTTGAACGTGGTATCTGAC  300
 T  F  V  T  M  L  R  Y  D  G  V  P  S  S  L  N  V  V  S  D

AAAACTGGCAACGGTGGACCGCTTCTACAACCTTATCCCGATTGGTCATTTGCTAAGTAT  360
 K  T  G  N  G  G  P  L  L  Q  P  Y  P  D  W  S  F  A  K  Y

GAAGATTGCTCTGGAATCGTGAGCGCCAACAAAATTGCTATCGACGAATATGAGAGATTG  420
 E  D  C  S  G  I  V  S  A  N  K  I  A  I  D  E  Y  E  R  L

TGGGTTCTGGACTCGGGCCTTGTCAATAATATTCAACCTATGTGTTCTCCAAAATTGCTT  480
 W  V  L  D  S  G  L  V  N  N  I  Q  P  M  C  S  P  K  L  L

GCCTTTGATTTGACTACTTCGAAATTGCTCAAGCAAGTCGAGATACCGCACGATGTTGCC  540
 A  F  D  L  T  T  S  K  L  L  K  Q  V  E  I  P  H  D  V  A

GTAAATGCCACCACAGGAAAGGGCGGATTAGCATCTTTAGCTGTTCAAGCTATGGATTCT  600
 V  N  A  T  T  G  K  G  G  L  A  S  L  A  V  Q  A  M  D  S

GTAAATACTATGGTGTACATGGCAGATAACAAAGATGATGCTTTAATTGTCTACCAAAAT  660
 V  N  T  M  V  Y  M  A  D  N  K  D  D  A  L  I  V  Y  Q  N

GCCGATGATTCTTTCCATCGATTGTCTTCCCACATTTCCAATCACAACTTTAGATCTGAC  720
 A  D  D  S  F  H  R  L  S  S  H  I  S  N  H  N  F  R  S  D

AAAATGTCGCAAGAAAATCTCACCTTGAAAGAAGTAGACAACAGAGTTTTTGGAATGGCA  780
 K  M  S  Q  E  N  L  T  L  K  E  V  D  N  R  V  F  G  M  A

CTTAGTTCCGTGACGCATAATCTTTATTATAGTCCTCTCTCTTCTCAGAATTTATATTAC  840
 L  S  S  V  T  H  N  L  Y  Y  S  P  L  S  S  Q  N  L  Y  Y

GTTAACACAACATCGTTAATGAACTCGCAAAATCAAGGAAATGACGTGCAGTATGAAAGT  900
 V  N  T  T  S  L  M  N  S  Q  N  Q  G  N  D  V  Q  Y  E  S

GTCCAAGACGTTTTCAGCAGTCAATTATCCGCTAAAGCAGTATCGAAAAATGGCGTACTC  960
 V  Q  D  V  F  S  S  Q  L  S  A  K  A  V  S  K  N  G  V  L

TTTTTCGGATTCACGAATAATACTCTTGGTTGCTGGAATGAGCATCAGTCACTTGACAGA 1020
 F  F  G  F  T  N  N  T  L  G  C  W  N  E  H  Q  S  L  D  R

CAAAATATCGATATTGTAGCTCGAAATGAGACGCTTCAAATGGTCGTTGGTATGAAGATT 1080
 Q  N  I  D  I  V  A  R  N  E  T  L  Q  M  V  V  G  M  K  I

AAGCAAAACCTTCCACAATCTGGCAAAGTTAATAATACACAAAGAAATGAACATTTGTTG 1140
 K  Q  N  L  P  Q  S  G  K  V  N  N  T  Q  R  N  E  H  L  L
```

(continued)

Figure 3.8 Nucleotide and deduced amino acid sequences of AcMRJP4. Initiation and termination of translational codons and putative polyadenylation signal are boldfaced. The signal peptidase cleavage site was indicated by arrow. *N*-linked glycosylation sites are underlined.

```
GCTTTAACCAACAAAAAGCAGGACGTGCTAAACAACGATCTTAATCTCGAACATGTGAAC 1200
 A  L  T  N  K  K  Q  D  V  L  N  N  D  L  N  L  E  H  V  N

TTCCAAATTTTGGATGCTAATGTAAACGACTTGATACGGAATAGTCGTTGCGCAAATTCT 1260
 F  Q  I  L  D  A  N  V  N  D  L  I  R  N  S  R  C  A  N  S

GACAATCAGGATAATAATCAACATAATTATAATCATAATCAAGTTCGTCATTCTTCAAAA 1320
 D  N  Q  D  N  N  Q  H  N  Y  N  H  N  Q  V  R  H  S  S  K

TCTGACAATCAGAATAACAATCAACATAACAATCAAGCTTATCATTCTTCAAAGTCTGAC 1380
 S  D  N  Q  N  N  N  Q  H  N  N  Q  A  Y  H  S  S  K  S  D

AATTGGGATAACAATAACAATCAAGCTCATCATTCCTCAAAATTTGATAATCAGAATAAC 1440
 N  W  D  N  N  N  N  Q  A  H  H  S  S  K  F  D  N  Q  N  N

AATCAATATAACAATTAGGTTCATCATTCTTCATCAAATCATGTTAAATCTGATAATTAA 1500
 N  Q  Y  N  N  *

TCTTTTCTCGATGTAAGTCAAATATTTTAAAAAATTTCATTACATTATAAAACGAATAAA 1560

ATAAATATCGTTTTTCGCATAAAAAAAAAAAAAAAAAAAAAAAAAAA 1608
```

Figure 3.8 (continued)

```
CLUSTAL X (1.81) multiple sequence alignment

AcMRJP4         -----------------------------------ATGACAAAATGGTTGCTGTTGAT
AmMRJP4         GTCACTTGTAAAATATTTGTAATATCCTAGAAAAAAAAATGACAAAATGGTTGCTGTTGAT
                                                   ************************

AcMRJP4         GGCATGCCTTGGCATAGCTTGTCAAAATATTAGAGGTGCCGTTGTTCGAGAAAATTCCTC
AmMRJP4         GGTATGCCTTGGCATAGCTTGTCAAAATATTAGAGGTGGCGTTGTTCGAGAAAATTCCTC
                ** ********************************** **********************

AcMRJP4         GAGAAAAAAATTAACAAATACGTTGAACGTGATTCACGAATGGAAGTATGTCGATTATGA
AmMRJP4         GGGAAAAAACTTGACAAATACGTTGAACGTGATTCACAAATGGAAGTATCTCGATTATGA
                * ******* ** ************************ *********** **********

AcMRJP4         TTTCGGTAGCGACGAAAAAAGGCAAGCTGCGATTCAATCTGGCGAATATGATCGTACGAA
AmMRJP4         TTTCGATAACGACGAAAGGAGGCAAGCTGCGATTCAATCTGGCGAATATGATCGTACAAA
                ***** ** ********  ************************************* **

AcMRJP4         AAATTATCCTCTTGACGTCGATCAATGGCATGATAAGACTTTTGTCACTATGTTAAGATA
AmMRJP4         AAATTATCCTCTTGACGTCGATCAATGGCACAACAAGACTTTTCTCGCTGTAATAAGATA
                ******************************  * ********* ** ** *  *******

AcMRJP4         CGATGGTGTGCCTTCCTCTTTGAACGTGGTATCTGACAAAACTGGCAACGGTGGACCGCT
AmMRJP4         CAATGGTGTGCCTTCCTCTTTGAACGTGGTATCTGACAAAACTGGCAACGGTGGACGACT
                * ********************************************************  **

AcMRJP4         TCTACAACCTTATCCCGATTGGTCATTTGCTAAGTATGAAGATTGCTCTGGAATCGTGAG
AmMRJP4         TCTACAACCGTATCCTGATTGGTCATTTGCCAAGTACGAAGATTGCTCTGGAATCGTGAG
                ********* ***** ************** ***** ************************

AcMRJP4         CGCCAACAAAATTGCTATCGACGAATATGAGAGATTGTGGGTTCTGGACTCGGGCCTTGT
AmMRJP4         CGCTCATAAAATTGCTATCGACGAATATGAGAGATTGTGGGTTCTGGATTCGGGTCTCGT
                ***  * ************************************** ***** ** **

AcMRJP4         CAATAATATTCAACCTATGTGTTCTCCAAAATTGCTTGCCTTTGATTTGACTACTTCGAA
AmMRJP4         CAATAATACGCAACCCATGTGTTCTCCAAAACTGTTCGCTTTTGATCTTAATACCTCGCA
                ********  ***** *************** ** * ** ****** * * *** *** *

AcMRJP4         ATTGCTCAAGCAAGTCGAGATACCGCACGATGTTGCCGTAAATGCCACCACAGGAAAGGG
AmMRJP4         ATTGCTCAAGCAAGTCGAGATACCGCACGATGTTGCC---------ACCACAGGAAAGGG
                ************************************         **************

AcMRJP4         CGGATTAGCATCTTTAGCTGTTCAAGCTATGGATTCTGTAAATACTATGGTGTACATGGC
AmMRJP4         CGAATTAGTATCTTTAACTGTTCAAGCTATGGATTCGACAAATACTATGGTGTACATGGT
                ** ***** ******* ********************   ********************

AcMRJP4         AGATAACAAAGATGATGCTTTAATTGTCTACCAAAATGCCGATGATTCTTTCCATCGATT
AmMRJP4         AGACAACAAAAATA---CTTTGATCATCTACCAAAATGCCGATGATTCTTTTCATCGATT
                *** ****** **    **** **  ************************** ********

AcMRJP4         GTCTTCCCACATTTCCAATCACAACTTTAGATCTGACAAAATGTCG-CAAGAAAATCTCA
AmMRJP4         GTCTTCCCACACTTTGAATCACAACTCT-GACAAAATGTCAGATCAACAAGAAAATCTCA
                *********** **  ********** * **    *    *  **  *************
```

(continued)

Figure 3.9 Alignment of the nucleotide sequence of AcMRJP4 cDNA with AmMRJP4 cDNA published sequence (GenBank Acc. Z26319). Conserve residues are indicated by asterisks.

```
AcMRJP4         CCTTGAAAGAAGTAGACAACAGAGTTTTTGGAATGGCACTTAGTTCCGTGACGCATAATC
AmMRJP4         CCTTGAAAGAAGTAGACAACAAAGTTTATGGAATGGCACTTAGTCCCGTGACGCATAATC
                ********************* ***** *************** ***************

AcMRJP4         TTTATTATAGTCCTCTCTCTTCTCAGAATTTATATTACGTTAACACAACATCGTTAATGA
AmMRJP4         TTTATTACAATTCTCCGTCTTCTGAGAATTTGTATTATGTTAACACAGAATCGTTAATGA
                ******* * * ***  ****** ******* ***** *********  ***********

AcMRJP4         ACTCGCAAAATCAAGGAAATGACGTGCAGTATGAAAGTGTCCAAGACGTTTTCAGCAGTC
AmMRJP4         AATCGGAAAATCAAGGAAATGACGTGCAATATGAAAGAGTCCAAGACGTTTTCGACAGTC
                * *** ********************** ******** ***************  *****

AcMRJP4         AATTATCCGCTAAAGCAGTATCGAAAAATGGCGTACTCTTTTTCGGATTCACGAATAATA
AmMRJP4         AATTAACCGTTAAAGCAGTATCGAAAAATGGCGTACTCCTTTTCGGACTCGCGAATAATA
                ***** *** ***************************** ******** ** *********

AcMRJP4         CTCTTGGTTGCTGGAATGAGCATCAGTCACTTGACAGACAAAATATCGATATTGTAGCTC
AmMRJP4         CTCTTAGTTGCTGGAACGAGCATCAGTCACTTGACAGACAAAATATCGATGTCGTAGCTC
                ***** ********** ********************************** * *******

AcMRJP4         GAAATGAG---ACGCTTCAAATGGTCGTTGGTATGAAGATTAAGCAAAACCTTCCACAAT
AmMRJP4         GAAATGAGGACACGCTTCAAATGGTCGTTAGTATGAAGATTAAGCAAAACGTTCCACAAT
                ********   ****************** ******************** *********

AcMRJP4         CTGGCAAAGTTAATAATACACAAAGAAATGAACATTTGTTGGCTTTAACCAACAAAAAGC
AmMRJP4         CTGGCAGAGTTAATAATACGCAAAGAAATGAATATTTGTTGGCTTTAAGCGACAGAAACC
                ****** ************ ************ *************** * *** *** *

AcMRJP4         AGGACGTGCTAAACAACGATCTTAATCTCGAACATGTGAACTTCCAAATTTTGGATGCTA
AmMRJP4         AGAACGTGCTAAACAACGATCTTAATCTCGAACACGTGAACTTCCAAATTTTGGGCGCTA
                ** ******************************** ******************  ****

AcMRJP4         ATGTAAACGACTTGATACGGAATAGTCGTTGCGCAAATTCTGACAATCAGGATAATAATC
AmMRJP4         ACGTAAACGACTTGATACGGAATAGTCGTTGCGCAAATTTTGACAATCAGGATAATAATC
                * ************************************* ********************

AcMRJP4         AACATAATTATAATCATAATCAAGTTCGTCATTCTTCAAAATCTGACAATCAGAATAACA
AmMRJP4         ACTATAATCATAATCATAATCAAGCTCGTCATTCTTCAAAATCTGACAATCAGAATAACA
                *  ***** *************** ***********************************

AcMRJP4         ATCAACATAACAATCAAGCTTATCATTCTTCAAAGTCTGACAATTGGGATAACAATAACA
AmMRJP4         ATCAACATAACGATCAAGCTCATCATTCTTCAAAGTCTAACAATCGGCATAACAATAACG
                *********** ******** ***************** ***** ** ***********

AcMRJP4         ATCAAGCTCATCATTCCTCAAAATTTGATAATCAGAATAACAATCAATATAACAATTAGG
AmMRJP4         ATTAAGCTCATCATTTTTCAAAATTTGATAATCAGAATAACAATCAGAATAACGATTAAT
                ** ***********  ***************************  ***** ****

AcMRJP4         TTCATCATTCTTCATCAAATCATGTTAAATCTGATAATTAATCTTTTCTCGATGTAAGTC
AmMRJP4         ATAATAATCAATTTTATCATTCTTTAAAATCTGTTAATTAATCTTTTCTCGATGTAAGTC
                 * ** **   *  *   **  * * ******* ************************

AcMRJP4         AAATATTTTAAAAAA-TTTCATTACATTATAAAACGAATAAAATAAATATCGTTTTTC-G
AmMRJP4         AAATATTTTAAAAAAATTTCATTACATTATAAAACGA-TAAAATAAATATCGTTTTTTTG
                *************** ********************** ******************  *

AcMRJP4         CATAAAAAAAAAAAAAAAAAAAAAAAAAAA
AmMRJP4         CATAAT------------------------
                *****
```

Figure 3.9 (continued)

```
CLUSTAL X (1.81) multiple sequence alignment

AcMRJP4          MTKWLLLMACLGIACQNIRGAVVRENSSRKKLTNTLNVIHEWKYVDYDFGSDEKRQAAIQ
AmMRJP4          MTKWLLLMVCLGIACQNIRGGVVRENSSGKNLTNTLNVIHKWKYLDYDFDNDERRQAAIQ
                 ********.***********.******* *:*********:***:****..**:******

AcMRJP4          SGEYDRTKNYPLDVDQWHDKTFVTMLRYDGVPSSLNVVSDKTGNGGPLLQPYPDWSFAKY
AmMRJP4          SGEYDRTKNYPLDVDQWHNKTFLAVIRYNGVPSSLNVVSDKTGNGGRLLQPYPDWSFAKY
                 ******************:***::::**:**************** *************

AcMRJP4          EDCSGIVSANKIAIDEYERLWVLDSGLVNNIQPMCSPKLLAFDLTTSKLLKQVEIPHDVA
AmMRJP4          EDCSGIVSAHKIAIDEYERLWVLDSGLVNNTQPMCSPKLFAFDLNTSQLLKQVEIPHDVA
                 *********:****************** ********:****.**:*************

AcMRJP4          VNATTGKGGLASLAVQAMDSVNTMVYMADNKDDALIVYQNADDSFHRLSSHISNHNFRSD
AmMRJP4          T---TGKGELVSLTVQAMDSTNTMVYMVDNKN-TLIIYQNADDSFHRLSSHTLNHN--SD
                 .   **** *.**:******.******.***: :**:**************  ***  **

AcMRJP4          KMS--QENLTLKEVDNRVFGMALSSVTHNLYYSPLSSQNLYYVNTTSLMNSQNQGNDVQY
AmMRJP4          KMSDQQENLTLKEVDNKVYGMALSPVTHNLYYNSPSSENLYYVNTESLMKSENQGNDVQY
                 ***  **********:*:*****.*******.. **:******* ***:*:********

AcMRJP4          ESVQDVFSSQLSAKAVSKNGVLFFGFTNNTLGCWNEHQSLDRQNIDIVARN-ETLQMVVG
AmMRJP4          ERVQDVFDSQLTVKAVSKNGVLLFGLANNTLSCWNEHQSLDRQNIDVVARNEDTLQMVVS
                 * *****.***:.*********:**::****.**************:**** :******.

AcMRJP4          MKIKQNLPQSGKVNNTQRNEHLLALTNKKQDVLNNDLNLEHVNFQILDANVNDLIRNSRC
AmMRJP4          MKIKQNVPQSGRVNNTQRNEYLLALSDRNQNVLNNDLNLEHVNFQILGANVNDLIRNSRC
                 ******:****:********:****:::::*:****************.************

AcMRJP4          ANSDNQDNNQHNYNHNQVRHSSKSDNQNNNQHNNQAYHSSKSDNWDNNNNQAHHSSKFDN
AmMRJP4          ANFDNQDNNHYNHNHNQARHSSKSDNQNNNQHNDQAHHSSKSNNRHNNND----------
                 ** ******::*:****.***************:**:*****:* .***:

AcMRJP4          QNNNQYNN
AmMRJP4          --------
```

Figure 3.10 Alignment of the deduced amino acid sequence of AcMRJP4 cDNA with AmMRJP4 cDNA published sequence (GenBank Acc. Z26319). Conserve residues are indicated in asterisks. **:** means amino acid which have the same group of side chains and similar size while **.** means amino acid which have the same group of side chains but different size.

**AcMRJP5**

Recombinant plasmids containing 1,900 bp cDNA which might be AcMRJP5 cDNA as analyzed by *Eco*RI and *Ssp*I digestion. Further identification was performed by sequencing the insert cDNA using five primers as showed in Table 2.1 and Figure 2.1.

The recombinant plasmids were sequenced using M13 forward and M13 reverse primer. Three internal sequencing primers (primer name MRJP5_A, MRJP5_B and MRJP5_C) were designed from nucleotide sequences obtained. The inserts of two recombinant clones were sequenced along the entire length and the insert of one recombinant clone was partially sequence. The nucleotide sequence of two recombinant clone and partial sequence of the another clone were assembled (Appendix D). The nucleotide sequence was compared with the DNA sequence deposited in GenBank database using nucleotide Blast (BlastN) and translated protein Blast (BlastX). The result shown that similar to AmMRJP5 cDNA. This sequence was most likely to be AcMRJP5 cDNA.

AcMRJP5 cDNA indicated a length of 1881 bp (including poly (A) tail) the sequence contained an open reading frame (nucleotides 1-1740) which encoded 579 amino acid residues (Figure 3.11). The sequence AATAAAATAAA containing two partially overlapping consensus polyadenylation signal sequences (AATAAA) was located 14 bp upstream from the poly (A) tail. The 3'-terminal sequence was observed nonanucleotide repeat sequence, GATAGAATG that encoded to tripeptide (DRM).

The computer sequence analysis predicted that the signal peptidase cleavage site was located between Gly 20 and Ala 21. A comparison of AcMRJP5 and AmMRJP5 nucleotide sequences analyed by blast N program revealed an identity of 91%. For blast

X program, protein sequence of AcMRJP5 showed 90% sequence identity and 96% sequence similarity to AmMRJP5.

The deduced amino acid (without putative signal peptide) composition of AcMRJP5 comprised of 42.04% hydrophobic, 24.51% neutral and 33.45% hydrophilic amino acid residues. The essential amino acid content was 51.9%.

The pI value was estimated to be 8.75. The estimated molecular weight was 66.2 kDa. Five putative *N*-glycosylation sites were found.

The nucleotide sequence and deduced amino acid sequence of AcMRJP5 cDNA were compared with the nucleotide sequence and deduced amino acid sequence of AmMRJP5 cDNA. The result of alignment was shown in Figure 3.12 and Figure 3.13.

The deduced amino acid of AcMRJP5 inferred from AcMRJP5 cDNA show the extensive repeat region located between amino acid residue 367 and 520. This repeat region located at the C-terminal of this protein. The repetitive region consists of a 51-fold repeated tripeptide motif with dominance of DRM sequence motifs (Figure 3.13). The result from nucleotide sequencing of 3 recombinant clone contained AcMRJP5 cDNA insert show the repeat region of the AcMRJP5 was polymorphism that invariant in repeated tripeptide motif (Appendix D).

The repetitive region of *A. cerana* was located at the same position as found *in A. mellifera* but smaller in size, that occurred 51 times compared with 58 times in *A. mellifera* (Albert *et al.*, 1999a)

```
ATGACAAGTTGGTTGTTGCTGGTGGTGTGCCTTGGCATAGCTTGTCAAGGTATCACAGGC   60
 M  T  S  W  L  L  L  V  V  C  L  G  I  A  C  Q  G  I  T  G ▲

GCCACTGTTCGAGAAAATTCTTCGAGAAATTTGGCAAATTCGATGAACGTGATTCACGAA  120
 A  T  V  R  E  N  S  S  R  N  L  A  N  S  M  N  V  I  H  E

TGGAAGTATCTTGATTATGACTTCGGTAGCGACGAAAAAAGACAAGCTGCGATTCAATCT  180
 W  K  Y  L  D  Y  D  F  G  S  D  E  K  R  Q  A  A  I  Q  S

GGCGAATATGACCATACGAAAAATTATCCCTTCGATGTCGATCGATGGCATGATATGACT  240
 G  E  Y  D  H  T  K  N  Y  P  F  D  V  D  R  W  H  D  M  T

TTTGTCACCGTACTAAGATACAAAGGTGTACCTTCCTCTTTAAACGTGATATCTAAGAAA  300
 F  V  T  V  L  R  Y  K  G  V  P  S  S  L  N  V  I  S  K  K

ATTGGCAACGGTGGACCTCTTCTGCAGCCATATCCTGATTGGTCGTGGGCGAACTATAAA  360
 I  G  N  G  G  P  L  L  Q  P  Y  P  D  W  S  W  A  N  Y  K

GATTGCTCTGGAATCGTGAGCGCTTACAAAATTGCGATCGACAAGTTCGACAGATTGTGG  420
 D  C  S  G  I  V  S  A  Y  K  I  A  I  D  K  F  D  R  L  W

GTTCTGGACTCAGGTATTATCAATAATACTCAACCCATGTGTTCACCAAAATTGCATGTC  480
 V  L  D  S  G  I  I  N  N  T  Q  P  M  C  S  P  K  L  H  V

TTTGATCTCAATACCTCACAGCAGATTAAGCAAGTTATGATGCCGCATGATATTGCCATA  540
 F  D  L  N  T  S  Q  Q  I  K  Q  V  M  M  P  H  D  I  A  I

AATGCCACTACAGGAAAAGGAGGACTAGAAAATCTAGTTGTTCAAGCTATGGATCCTATG  600
 N  A  T  T  G  K  G  G  L  E  N  L  V  V  Q  A  M  D  P  M

AATACTCTGGTGTATATGGCAGATAACAAGGGTGATGCTTTAATTGTTTATCAAAATTCC  660
 N  T  L  V  Y  M  A  D  N  K  G  D  A  L  I  V  Y  Q  N  S

GATGATTCCTTCCATCGATTGACTTCCAACACTTTCGATTACGATCCCAAATATATCAAA  720
 D  D  S  F  H  R  L  T  S  N  T  F  D  Y  D  P  K  Y  I  K

ATGATGGCCGCAGGAGAAAGTTTCACATTGCAAGATGGAATTTTTGGAATGGCACTCAGT  780
 M  M  A  A  G  E  S  F  T  L  Q  D  G  I  F  G  M  A  L  S

CCCATGACAAACAATCTTTATTACAGTCCTCTCGCTTCTCGCAGTTTGTATTATGTTAAT  840
 P  M  T  N  N  L  Y  Y  S  P  L  A  S  R  S  L  Y  Y  V  N

ACGAAACCCTTCATGAAATCACAATATGGAACAAATAACGTACAACATGAAGGTGTCCAA  900
 T  K  P  F  M  K  S  Q  Y  G  T  N  N  V  Q  H  E  G  V  Q

GATATTTTCAATACTCAATCAATTGCTAAAATAATGTCGAAAAATGGCGTTCTCTTTTTC  960
 D  I  F  N  T  Q  S  I  A  K  I  M  S  K  N  G  V  L  F  F

GGTCTCATGAATAATTCAGCTATTGGTTGTTGGAATGAGCACCAACCACTTCAGAGACAA 1020
 G  L  M  N  N  S  A  I  G  C  W  N  E  H  Q  P  L  Q  R  Q

AATATGGATATGGTCGCTCAGAATGAAGAGACACTTCAAACGGTCGTTGCTATGAAAATG 1080
 N  M  D  M  V  A  Q  N  E  E  T  L  Q  T  V  V  A  M  K  M

ATGCATCTCCCACAATCCAACAGGATGAATAGGATGCATAGGATGAATAGTATGAATAGA 1140
 M  H  L  P  Q  S  N  R  M  N  R  M  H  R  M  N  S  M  N  R
```

(continued)

Figure 3.11 Nucleotide and deduced amino acid sequences of AcMRJP5. Initiation and termination of translational codons and putative polyadenylation signal are boldfaced. The signal peptidase cleavage site was indicated by arrow. *N*-linked glycosylation sites are underlined.

```
ATGGATAGGATGGATAGAATGGATAGGATGGATAGGATGGATAGGATGGATAGGATGGAT 1200
 M  D  R  M  D  R  M  D  R  M  D  R  M  D  R  M  D  R  M  D

AGGATGGATAGGATGGATAGGATGGATAGGATGGATAGAATGGATAGAATGGATAGGATG 1260
 R  M  D  R  M  D  R  M  D  R  M  D  R  M  D  R  M  D  R  M

GATAGGATGGATATAATGGATAGGACGAATAAAATGGATAGGATGGATAGGATGGATATA 1320
 D  R  M  D  I  M  D  R  T  N  K  M  D  R  M  D  R  M  D  I

ATGGATAAGATGAATAAAATGGATAGGATGGATAGTATGATTAGAATAGATAAAATGGAT 1380
 M  D  K  M  N  K  M  D  R  M  D  S  M  I  R  I  D  K  M  D

AGAATGGATAGAATGCATAGAATAGATATAATGAATAGAATGGATAGAATGGATAGAATG 1440
 R  M  D  R  M  H  R  I  D  I  M  N  R  M  D  R  M  D  R  M

GACACAAGAATAGATACAAGAATGGACAGAATGGATAGAATGGATAAAATGGATAAGATA 1500
 D  T  R  I  D  T  R  M  D  R  M  D  R  M  D  K  M  D  K  I

AATAAGATGCATAGGATGGGTAGGATGGATAGGATGGATAGAATGAATAGAATGAATAGA 1560
 N  K  M  H  R  M  G  R  M  D  R  M  D  R  M  N  R  M  N  R

CAAATGAATGAATATATGATGGCTTTAAGTATGAAATTACAGAAATTTATAAACAATGAT 1620
 Q  M  N  E  Y  M  M  A  L  S  M  K  L  Q  K  F  I  N  N  D

TATAATTTCAACGAAGTAAATTTCCGAATTTTGGCTGCAAATGTAAACGATTTAATAATG 1680
 Y  N  F  N  E  V  N  F  R  I  L  A  A  N  V  N  D  L  I  M

AACACTCGTTGTGCAAATTCTAACAATCAGAATGATAATCAAAATAAGCATAATAATTAA 1740
 N  T  R  C  A  N  S  N  N  Q  N  D  N  Q  N  K  H  N  N  *

GGTAGTCGTTCTTTATATTAAAATCTGTTAATTAGTCTTTTCTCGACTATAAACCAAATA 1800

TTGTTTCAAATTTCTTTATATTATAAATGAATAAAATAAATATCGTTTTTGCTTAAAAAA 1860

AAAAAAAAAAAAAAAAAAAAA 1881
```

Figure 3.11 (continued)

```
CLUSTAL X (1.81) multiple sequence alignment

AcMRJP5         ------------------------------------------------------ATGA
AmMRJP5         TACACTGCGTTCTCTTGAAACTGTCGTTTGCAAAATATTTGCAGCATCCAAGAACAATGA
                                                                      ****

AcMRJP5         CAAGTTGGTTGTTGCTGGTGGTGTGCCTTGGCATAGCTTGTCAAGGTATCACAGGCGCCA
AmMRJP5         CAACTTGGTTGTTGCTGGTCGTGTGCCTTGGCATAGCTTGTCAAGGTATCACAAGCGTCA
                *** *************** ********************************** *** **

AcMRJP5         CTGTTCGAGAAAATTCTTCGAGAAATTTGGCAAATTCGATGAACGTGATTCACGAATGGA
AmMRJP5         CTGTTCGAGAAAATTCTCCGAGAAAGTTGGCAAATTCGATGAACGTGATTCACGAATGGA
                ***************** ******* **********************************

AcMRJP5         AGTATCTTGATTATGACTTCGGTAGCGACGAAAAAAGACAAGCTGCGATTCAATCTGGCG
AmMRJP5         AGTATCTCGATTATGATTTTGGTAGCGACGAAAGGAGGCAAGCTGCGATGCAATCTGGCG
                ******* ******** ** *************  ** *********** **********

AcMRJP5         AATATGACCATACGAAAAATTATCCCTTCGATGTCGATCGATGGCATGATATGACTTTTG
AmMRJP5         AGTATGACCATACGAAAAATTATCCCTTCGATGTCGATCAATGGCGTGGTATGACTTTTG
                * ************************************ ***** ** ***********

AcMRJP5         TCACCGTACTAAGATACAAAGGTGTACCTTCCTCTTTAAACGTGATATCTAAGAAAATTG
AmMRJP5         TAACCGTACCAAGATACAAAGGTGTACCTTCTTCTTTGAACGTGATATCTGAGAAAATTG
                * ******* ********************** ***** ************ *********

AcMRJP5         GCAACGGTGGACCTCTTCTGCAGCCATATCCTGATTGGTCGTGGGCGAACTATAAAGATT
AmMRJP5         GCAACGGTGGACGACTTCTACAACCGTATCCTGATTGGTCGTGGGCGAACTATAAAGATT
                ************  ***** ** ** **********************************

AcMRJP5         GCTCTGGAATCGTGAGCGCTTACAAAATTGCGATCGACAAGTTCGACAGATTGTGGGTTC
AmMRJP5         GCTCTGGAATAGTGAGCGCTTACAAAATTGCGATCGACAAGTTCGACAGATTGTGGATTC
                ********** ******************************************** ***

AcMRJP5         TGGACTCAGGTATTATCAATAATACTCAACCCATGTGTTCACCAAAATTGCATGTCTTTG
AmMRJP5         TGGACTCAGGTATTATCAATAATACTCAACCCATGTGTTCACCAAAATTGCATGTCTTTG
                ************************************************************

AcMRJP5         ATCTCAATACCTCACAGCAGATTAAGCAAGTTATGATGCCGCATGATATTGCCATAAATG
AmMRJP5         ATCTCAATACCTCACATCAGCTTAAGCAAGTTGTGATGCCGCACGATATTGCCGTAAATG
                **************** *** *********** ********** ********* ******

AcMRJP5         CCACTACAGGAAAAGGAGGACTAGAAAATCTAGTTGTTCAAGCTATGGATCCTATGAATA
AmMRJP5         CCAGCACAGGGAATGGGGGACTCGTATCACTAGTTGTTCAAGCTATGGATCCTGTGAATA
                *** ***** ** ** ***** * *  ************************ ******

AcMRJP5         CTCTGGTGTATATGGCAGATAACAAGGGTGATGCTTTAATTGTTTATCAAAATTCCGATG
AmMRJP5         CTATCGTGTATATGGCAGATGACAAAGGTGATGCTTTAATCGTCTACCAAAATTCTGACG
                ** * *************** **** ************** ** ** ******** ** *

AcMRJP5         ATTCCTTCCATCGATTGACTTCCAACACTTTCGATTACGATCCCAAATATATCAAAATGA
AmMRJP5         AATCTTTCCATCGATTGACTTCCAACACTTTCGATTACGATCCCAAATATATCAAAATGA
                * ** ************************************************************
```

(continued)

Figure 3.12 Alignment of the nucleotide sequence of AcMRJP5 cDNA with AmMRJP5 cDNA published sequence (GenBank Acc. AF004842). Conserve residues are indicated by asterisks.

```
AcMRJP5     TGGCCGCAGGAGAAAGTTTCACATTGCAAGATGGAATTTTTGGAATGGCACTCAGTCCCA
AmMRJP5     TGGACGCGGGAGAAAGTTTCACAGCGCAAGATGGAATTTTTGGAATGGCACTCAGTCCCA
            *** *** ***************  ***********************************

AcMRJP5     TGACAAACAATCTTTATTACAGTCCTCTCGCTTCTCGCAGTTTGTATTATGTTAATACGA
AmMRJP5     TGACAAACAATCTTTATTACAGCCCTCTTTCTTCTCGCAGTTTGTATTATGTTAATACAA
            ********************** *****  **************************** *

AcMRJP5     AACCCTTCATGAAATCACAATATGGAACAAATAACGTACAACATGAAGGTGTCCAAGATA
AmMRJP5     AACCATTCATGAAATCAGAATATGGAGCAAATAACGTACAATATCAAGGTGTCCAAGATA
            **** ************ ******** ************** ** ***************

AcMRJP5     TTTTCAATACTCAATCAATTGCTAAAATAATGTCGAAAAATGGCGTTCTCTTTTTCGGTC
AmMRJP5     TTTTCAACACTGAATCGATTGCTAAAATAATGTCGAAAAATGGCGTTCTCTTTTTCGGCC
            ******* *** **** ****************************************** *

AcMRJP5     TCATGAATAATTCAGCTATTGGTTGTTGGAATGAGCACCAACCACTTCAGAGACAAAATA
AmMRJP5     TCATGAATAATTCAGCTATTGGTTGTTGGAACGAGCATCAACCACTTCAGAGAGAAAATA
            ****************************** ***** *************** ******

AcMRJP5     TGGATATGGTCGCTCAGAATGAAGAGACACTTCAAACGGTCGTTGCTATGAAAATGATGC
AmMRJP5     TGGATATGGTCGCTCAGAATGAAGAGACTCTTCAAACGGTCGTTGCTATGAAAATGATGC
            **************************** ********************************

AcMRJP5     ATCTCCCACAATCCAACAGGATGAATAGGATGCATAGGATGAATAGTATGAATAGAATGG
AmMRJP5     ATCTCCCACAATCCAACAAGATGAATAGGATGCATAGGATGAATAGAGTGAATAGAGTGA
            ****************** **************************  ******** **

AcMRJP5     ATAGGATGGATAGAATGGATAGGATGGATAGGATGGATAGGATGGATAGGATGGATAGGA
AmMRJP5     ATAGAATGGATAGAATGGATAGAATAGATAGGATGGATAGGATGGATAGGATGGATACAA
            **** ***************** ** ******************************** *

AcMRJP5     TGGATAGGATGGATAGGATGGATAGGATGGATAGAATGGATAGAATGGATAGGATGGATA
AmMRJP5     TGGATACAATGGATAGAATAGATAGGATGGATAGGATGGATAGAATAGATAGGATAGATA
            ******  ******** ** ************** ************ ******** ****

AcMRJP5     GGATGGATATAATGGATA----------------------------------------
AmMRJP5     GGATGCATACAATGGATACAATGGATACAATGGATAGAACAGATAAGATGAGTAGCATGG
            ***** *** ********

AcMRJP5     ---GGACGAATAAAATGGATAGGATGGATAGGATGGATATAATGGATAAGATGAATAAAA
AmMRJP5     ATAGGATGGATAGAATGGATAGGGTGGATAGGATGGATACAATGGATAGAACAGATAAGA
               *** * *** ********** *************** ********  *   **** *

AcMRJP5     TGGATAGGATGGATAGTATGATTAGAATAGATAAAATGGATAGAATGGATAGAATGCATA
AmMRJP5     TGAGTAGCATGGATAGGATGGATAGAATGGATAGGGTGGATACAATGGATACAATGGATA
            **  *** ******** ***  ****** ****   ****** ******** **** ***

AcMRJP5     GAATAGATATAATGAATAGAATGGATAGAATGGATAGAATGGACACAA------GAATAG
AmMRJP5     CAATGGATAGAATGGATAGGATGGATAGGATGGATAGAATGGATAGAATGGATAGGATGG
            *** **** **** **** ******** ************** * **      * ** *

AcMRJP5     ATACAA------GAATGGACAGAATGGATAGAATGGATAAAATGGATAAGATAAATAAGA
AmMRJP5     ATACAATGGATAGAACAGATAAGATGAGTAGGATAGATAGAATGGATAAAATAGATAGAA
            ******      ***  ** *  ***  *** ** **** ********* *** ***  *

AcMRJP5     TGCATAGGATGGGTAGGATGGATAGGATGGATAGAATGAATAGAATGAATAGACAAATGA
AmMRJP5     TGGATAGGATGGATAGGACAAATAGAATGGATAGAATGAATAGGATGAATAGACAAATGA
            ** ********* *****   **** ***************** ***************
```

Figure 3.12 (continued)

```
AcMRJP5         ATGAATATATGATGGCTTTAAGTATGAAATTACAGAAATTTATAAACAATGATTATAATT
AmMRJP5         ATGAATATATGATGGCTTTAAGTATGAAATTACAGAAATTTATAAATAATGATTATAATT
                ********************************************** ************

AcMRJP5         TCAACGAAGTAAATTTCCGAATTTTGGCTGCAAATGTAAACGATTTAATAATGAACACTC
AmMRJP5         TCAACGAAGTAAACTTCCGAATTTTGGGTGCAAATGTAAACGATTTAATAATGAATACTC
                ************ ************ ************************** ****

AcMRJP5         GTTGTGCAAATTCTAACAATCAGAATGATAATCAAAATAAGCATAATAATTAAGGTAGTC
AmMRJP5         GTTGTGCAAATTCTGACAATCAGAATAACAATCAAAATAAGCATAATAATTAAGATGATC
                ************** *********** * ************************ *  **

AcMRJP5         GTTCTTTATATTAAAATCTGTTAATTAGTCTTTTCTCGACTATAAACCAAATATTGTTTC
AmMRJP5         GTTCTTTATATTAAAATCTGTTAATCAGTCTTTTCTCGA-TATAAACCAAATATTCTTTA
                ************************ ************ *************** ***

AcMRJP5         AAATTTCTTTATATTATAAATGAATAAAATAAATATCGTTTTTGCTTAAAAAAAAAAAAA
AmMRJP5         AAATTTCTTTATATTATAAATGAATAAAATAAATAT---TTTTGCATGAT----------
                ************************************   ****** * *

AcMRJP5         AAAAAAAAAAAAA
AmMRJP5         -------------
```

Figure 3.12 (continued)

```
CLUSTAL X (1.81) multiple sequence alignment

AcMRJP5         MTSWLLLVVCLGIACQGITGATVRENSSRNLANSMNVIHEWKYLDYDFGSDEKRQAAIQS
AmMRJP5         MTTWLLLVVCLGIACQGITSVTVRENSPRKLANSMNVIHEWKYLDYDFGSDERRQAAMQS
                **:****************..******.*:**********************:****:**

AcMRJP5         GEYDHTKNYPFDVDRWHDMTFVTVLRYKGVPSSLNVISKKIGNGGPLLQPYPDWSWANYK
AmMRJP5         GEYDHTKNYPFDVDQWRGMTFVTVPRYKGVPSSLNVISEKIGNGGRLLQPYPDWSWANYK
                **************:*:.****** *************:****** **************

AcMRJP5         DCSGIVSAYKIAIDKFDRLWVLDSGIINNTQPMCSPKLHVFDLNTSQQIKQVMMPHDIAI
AmMRJP5         DCSGIVSAYKIAIDKFDRLWILDSGIINNTQPMCSPKLHVFDLNTSHQLKQVVMPHDIAV
                ********************:*************************:*:***:******:

AcMRJP5         NATTGKGGLENLVVQAMDPMNTLVYMADNKGDALIVYQNSDDSFHRLTSNTFDYDPKYIK
AmMRJP5         NASTGNGGLVSLVVQAMDPVNTIVYMADDKGDALIVYQNSDESFHRLTSNTFDYDPKYIK
                **:**:*** .********:**:*****:************:******************

AcMRJP5         MMAAGESFTLQDGIFGMALSPMTNNLYYSPLASRSLYYVNTKPFMKSQYGTNNVQHEGVQ
AmMRJP5         MMDAGESFTAQDGIFGMALSPMTNNLYYSPLSSRSLYYVNTKPFMKSEYGANNVQYQGVQ
                ** ****** ********************:***************:**:****::***

AcMRJP5         DIFNTQSIAKIMSKNGVLFFGLMNNSAIGCWNEHQPLQRQNMDMVAQNEETLQTVVAMKM
AmMRJP5         DIFNTESIAKIMSKNGVLFFGLMNNSAIGCWNEHQPLQRENMDMVAQNEETLQTVVAMKM
                *****:**********************************:********************

AcMRJP5         MHLPQSNRMNRMHRMNSMNRMDRMDRMDRMDRMDRMDRMDRMDRMDRMDRMDRMDRMDRM
AmMRJP5         MHLPQSNKMNRMHRMNRVNRVNRMDRMDRIDRMDRMDRMDTMDTMDRIDRMDRMDRIDRI
                *******:******** :**::*******:********** ** ***:********:**:

AcMRJP5         DR------MDIMDRTNKMDRMDRMDIMD-----------KMNKMDRMDSMIRIDKMDRM
AmMRJP5         DRMHTMDTMDTMDRTDKMSSMDRMDRMDRVDRMDTMDRTDKMSSMDRMDRMDRVDTMDTM
                **      ** ****:**. ***** **             **..***** * *:*.** *

AcMRJP5         DRMHRIDIMNRMDRMDRMDTRIDT--RMDRMDRMDKMDKINKMHRMGRMDRMDRMNRMNR
AmMRJP5         DTMDRMDRMDRMDRMDRMD-RMDTMDRTDKMSRIDRMDKIDRMDRMDRTNRMDRMNRMNR
                * *.*:* *:********* *:**  * *:*.*:*:****::*.**.* :**********

AcMRJP5         QMNEYMMALSMKLQKFINNDYNFNEVNFRILAANVNDLIMNTRCANSNNQNDNQNKHNN
AmMRJP5         QMNEYMMALSMKLQKFINNDYNFNEVNFRILGANVNDLIMNTRCANSDNQNNNQNKHNN
                ******************************.***************:***:*******
```

Figure 3.13 Alignment of the deduced amino acid sequence of AcMRJP5 cDNA with AmMRJP5 cDNA published sequence (GenBank Acc. AF004842). Conserve residues are indicated in asterisks. : means amino acid which have the same group of side chains and similar size while . means amino acid which have the same group of side chains but different size.

**AcMRJP6**

Further identification of 1500 bp cDNA insert of recombinant plasmid was performed by sequencing of the insert DNA using four primers that showed in Table 2.1 and Figure 2.1. The insert DNA was sequenced using M13 forward and M13 reverse primers. One internal sequencing primer (MRJP6_A) was designed from nucleotide sequences obtained. Another primer (MRJP4_2) that had been used for sequence analysis of AcMRJP4 was used for sequence analysis of this DNA insert, eventhough one base mismatch at 5' termini was found in this primer.

The sequence of AcMRJP6 was obtained from sequencing of the insert DNA from three recombinant clones (Appendix E). One of this sequences show 100% identity with assembled sequence. The nucleotide sequence obtained was compared with the DNA sequence deposited in GenBank database using nucleotide Blast (BlastN) and translated protein Blast (BlastX). The result showed that the sequence is nearly the same to AmMRJP6 cDNA. This sequence was designated AcMRJP6.

AcMRJP6 cDNA had a length of 1450 bp (including poly(A) tail). The sequence contains an open reading frame (nucleotides 1-1308) which encoded 435 amino acid residues (Figure 3.14). The sequence AATAAAATAAA containing two partially overlapping consensus polyadenylation signal sequences (AATAAA) was located 14 bp upstream from the poly(A) tail. The computer sequence analysis predicted that the signal peptidase cleavage site was located between Ser 20 and Ala 21. A comparison of AcMRJP6 and AmMRJP6 nucleotide sequences analysis by blast N program revealed an identity of 92%. For blast X program protein sequence of AcMRJP6 was showed 88% sequence identity and 93% sequence similarity to AmMRJP6. The deduced amino acid

(without putative signal peptide) composition of AcMRJP6 comprised of 42.89% hydrophobic, 29.40% neutral and 27.71% hydrophilic amino acid residues. The essential amino acid content was 46.8%.

The pI value was estimated to be 6.44. The estimated molecular weight was 47.4 kDa. Two putative *N*-glycosylation sites were found.

The nucleotide sequence and deduced amino acid sequence of AcMRJP6 cDNA were compared with the nucleotide sequence and deduced amino acid sequence of AmMRJP6 cDNA. The result of alignment was shown in Figure 3.15 and Figure 3.16.

**AcMRJPs characterization**

A summary for molecular characterization of cDNA and deduced amino acid sequences of AcMRJPs are illustrated in Table 3.2

The amino acid composition of AcMRJPs are illustrated in Table 3.3. Four AcMRJPs contained high amounts of the 10 essential amino acid: MRJP5 (51.9%), MRJP1 (47.4%), MRJP6 (46.8%) and MRJP2 (45%). MRJP5 is rich in Arg and Met (8.9% and 12.5%). MRJP6 is rich in Ile and Phe (7.0% and 4.1%). MRJP3 and MRJP4 have a lower overall content of essential amino acids, but they possess relatively higher amount of some of them; MRJP3 had Arg (5.6%), Lys (6.6%) and MRJP4 had Leu (8.6%), Val (7.7%).

Alignment of deduced amino acid sequence of AcMRJP1-6 was show in Figure 3.17. Four cysteine residues are conserved in these proteins. The regions of high sequence similarity was found in N-terminal region of the protein. Protein sequence of PYPDWS, DCSGIVS, RLWVLDS, NLYYSP and LYYVNT are conserved among AcMRJP.

```
ATGACAAAATGGTTGCTGCTGATAGTGTGTCTTAGCATAGCTTGTCAAGATGTCACAAGC    60
 M  T  K  W  L  L  L  I  V  C  L  S  I  A  C  Q  D  V  T  S ▲

GCGATTCATCGAAGAAAATCTTCAAAAAATTTGGAACATTCGATGAACGTGATTCACGAA   120
 A  I  H  R  R  K  S  S  K  N  L  E  H  S  M  N  V  I  H  E

TGGAAATATCTTGATTATGATTTCGATACCAATGAAAAAAAACAAGCTGCGATTCAATTT   180
 W  K  Y  L  D  Y  D  F  D  T  N  E  K  K  Q  A  A  I  Q  F

GGTGAATACGACTATACGAAAAATTATCCCTTTGACGTCGATCAATGGCATGATAAGACT   240
 G  E  Y  D  Y  T  K  N  Y  P  F  D  V  D  Q  W  H  D  K  T

TTTGTCGCTGTAATAAGATACGATGGTGTACCTTCCTCTTTGAACGTGATATCTGACAAA   300
 F  V  A  V  I  R  Y  D  G  V  P  S  S  L  N  V  I  S  D  K

ACTGGCAACGGTGGACGCCTTCTCCAACCGTATCCTGATTGGTCGTGGACGAACTATAAA   360
 T  G  N  G  G  R  L  L  Q  P  Y  P  D  W  S  W  T  N  Y  K

GATTGTTCTGGAATCGTGAGCGTTTACAAAATTGCGATTGACAAATTCGACAGATTGTGG   420
 D  C  S  G  I  V  S  V  Y  K  I  A  I  D  K  F  D  R  L  W

GTTCTGGACTCAGGTCTTATTAATAATATTCAACTTATGTGTTCTCCAAAATTGCTTGCC   480
 V  L  D  S  G  L  I  N  N  I  Q  L  M  C  S  P  K  L  L  A

TTTGATCTGACAACTTCGAAATTGCTCAAGCAAGTCGAGATACCGTACGATATTGCTGTA   540
 F  D  L  T  T  S  K  L  L  K  Q  V  E  I  P  Y  D  I  A  V

AATGCCAGCACAGGAATGGGAGGACTCGTCTCATTAGTTGTTCAAGCTATGGATCCTATG   600
 N  A  S  T  G  M  G  G  L  V  S  L  V  V  Q  A  M  D  P  M

AATACTATGGTATATATAGCAGATGACAGAGGTGACGCTTTAATCATCTATCAAAATTCC   660
 N  T  M  V  Y  I  A  D  D  R  G  D  A  L  I  I  Y  Q  N  S

GATGATTCTTTCCATCGATTGAGTTCCAATACTTTTGATAACGATCCCAGATATTCTGAA   720
 D  D  S  F  H  R  L  S  S  N  T  F  D  N  D  P  R  Y  S  E

TTGACGGTCGCGGGAGAAAGTTTCACAGTGCATGATGGAATTTTTGGAATGGCACTTAGT   780
 L  T  V  A  G  E  S  F  T  V  H  D  G  I  F  G  M  A  L  S

CCTGTGACGAACAATCTTTATTATAGCCCTCTCACTTCTCACAGTTTGTATTACGTTAAC   840
 P  V  T  N  N  L  Y  Y  S  P  L  T  S  H  S  L  Y  Y  V  N

ACGGAACCATTTATGAAATCACAATATGGAGAAAATAATATACAATATGAAGGAATTCAA   900
 T  E  P  F  M  K  S  Q  Y  G  E  N  N  I  Q  Y  E  G  I  Q

GATATTTTCAACACTCAATCATCCGCTAAAGTAATGTCGAAAAATGGCGTCCTTTTCTTC   960
 D  I  F  N  T  Q  S  S  A  K  V  M  S  K  N  G  V  L  F  F

GGACTTGTGAATAATTCAGCTATTGGTTGTTGGAACGAGCATCAACCACTTCAGAAACAA  1020
 G  L  V  N  N  S  A  I  G  C  W  N  E  H  Q  P  L  Q  K  Q

AATATGGATATGGTCGCTCAGAATGAAGAGACACTTCAAATAATCACTAGTGTGAAAATT  1080
 N  M  D  M  V  A  Q  N  E  E  T  L  Q  I  I  T  S  V  K  I

ATACAAAATCTTCCATATTCCGGAAAGATGAATAGAATTCACAAGAATGAATATATGTTG  1140
 I  Q  N  L  P  Y  S  G  K  M  N  R  I  H  K  N  E  Y  M  L
```

(continued)

Figure 3.14 Nucleotide and deduced amino acid sequences of AcMRJP6. Initiation and termination of translational codons and putative polyadenylation signal are boldfaced. The signal peptidase cleavage site was indicated by arrow. *N*-linked glycosylation sites are underlined.

GCTTTAAGTAACAGAATGCAGAAAATAGTAAACAATGATTTTAATTTCAACGACATAAAT 1200
A L S N R M Q K I V N N D F N F N D I N

TTCCGAATATTGGGTGCGAATGTAAAGAACTTAATAAAAAACACTCGTTGTGCAAATTCT 1260
F R I L G A N V K N L I K N T R C A N S

AAAAATCAGAATAACAATCAAAAGAAACATAAGAATCAAGCTCAT**TAG**ATCTTTTCCAAG 1320
K N Q N N N Q K K H K N Q A H *

ATCATATTAAATTCTATAGATTAATTTTTTCTCGTGGTAAATCAAATATTTTTAAAAATT 1380

TATTTGCATTATAAATT**AATAAAATAAA**TATCATTTTCGCATAAAAAAAAAAAAAAAAAA 1440

AAAAAAAAAA 1450

Figure 3.14 (continued)

```
CLUSTAL X (1.81) multiple sequence alignment

AcMRJP6         ------------------------------------------------------------
AmMRJP6         ATTAAATATTTGCAGCTTCCCTCCATAAGTGTTTCCATATATCTTAATTGTATTTATTTG


AcMRJP6         ----------------------------------------ATGACAAAATGGTTGCTGCT
AmMRJP6         CAATCTTTCATTTATCTAACACGAAATATTTTGTAGAAAAATGACAAATTGGTTACTGCT
                                                        ******** ***** *****

AcMRJP6         GATAGTGTGTCTTAGCATAGCTTGTCAAGATGTCACAAGCGCGATTCATCGAAGAAAATC
AmMRJP6         GATAGTGTGTCTTAGCATAGCTTGTCAAGATGTCACAAGTGCGATTCATCAAAGAAAATC
                ************************************* ********** *********

AcMRJP6         TTCAAAAAATTTGGAACATTCGATGAACGTGATTCACGAATGGAAATATCTTGATTATGA
AmMRJP6         TTCAAAAAATTTGGAACATTCGATGAACGTGATTCACGAATGGAAATATATCGATTATGA
                ************************************************** * ********

AcMRJP6         TTTCGATACCAATGAAAAAAAACAAGCTGCGATTCAATTTGGTGAATACGACTATACGAA
AmMRJP6         TTTTGGTAGTGATGAAAAAAGACAAGCTGCGATTCAATCTGGCGAATACGATTATACGAA
                *** * **   ********* ***************** *** ******** ********

AcMRJP6         AAATTATCCCTTTGACGTCGATCAATGGCATGATAAGACTTTTGTCGCTGTAATAAGATA
AmMRJP6         AAATTATCCTTTCGACGTCGATCAATGGCATAATAAGACTTTTCTCGCTGTAATAAGATA
                ********* ** ****************** *********** ***************

AcMRJP6         CGATGGTGTACCTTCCTCTTTGAACGTGATATCTGACAAAACTGGCAACGGTGGACGCCT
AmMRJP6         CGATGGTGTACCTTCCTCTTTGAACGTGATATCTGAGAAAATTGGCAACGGTGGATGCCT
                ************************************ **** ************* ****

AcMRJP6         TCTCCAACCGTATCCTGATTGGTCGTGGACGAACTATAAAGATTGTTCTGGAATCGTGAG
AmMRJP6         TCTACAACCGTATCCTGATTGGTCGTGGGCGAACTATAAAGATTGTTCTGGAATAGTGAG
                *** ************************ ************************** *****

AcMRJP6         CGTTTACAAAATTGCGATTGACAAATTCGACAGATTGTGGGTTCTGGACTCAGGTCTTAT
AmMRJP6         CGCTTACAAAATTGCGATCGACAAATTCGACAGATTGTGGGTTCTGGACTCGGGTCTTAT
                ** *************** ********************************** ********

AcMRJP6         TAATAATATTCAACTTATGTGTTCTCCAAAATTGCTTGCCTTTGATCTGACAACTTCGAA
AmMRJP6         TAATAATATTCAACTTATGTGTTCTCCAAAATTACTTGCCTTTGATCTCAATACCTCAAA
                ********************************** ************** *  ** ** **

AcMRJP6         ATTGCTCAAGCAAGTCGAGATACCGTACGATATTGCTGTAAATGCCAGCACAGGAATGGG
AmMRJP6         GTTGCTTAAACAAATCGAGATACCACATAATATTGCCGTAAATGCCAGCACAGGAATGGG
                 ***** ** *** **********  *  ******* ************************

AcMRJP6         AGGACTCGTCTCATTAGTTGTTCAAGCTATGGATCCTATGAATACTATGGTATATATAGC
AmMRJP6         AGGACCCGTATCGCTAGTTGTTCAAGCTATGGATCCTATGAATACTACGGTGTATATAGC
                ***** *** **  ********************************** *** ********

AcMRJP6         AGATGACAGAGGTGACGCTTTAATCATCTATCAAAATTCCGATGATTCTTTCCATCGATT
AmMRJP6         AGACGACAGAGGTGACGCTTTAATCATCTATCAAAATTCTGATGATTCTTTCCATCGATT
                *** ************************************ ********************
```

(continued)

Figure 3.15 Alignment of the nucleotide sequence of AcMRJP6 cDNA with AmMRJP6 cDNA published sequence (GenBank Acc. AY313893). Conserve residues are indicated by asterisks.

```
AcMRJP6         GAGTTCCAATACTTTTGATAACGATCCCAGATATTCTGAATTGACGGTCGCGGGAGAAAG
AmMRJP6         GACTTCCAAAACTTTTGATAACGATCTCAGATATTCTGAACTGGCCGTCGCGGGAGAAAG
                ** ****** *************** ************* ** * **************

AcMRJP6         TTTCACAGTGCATGATGGAATTTTTGGAATGGCACTTAGTCCTGTGACGAACAATCTTTA
AmMRJP6         TTTCACAGTGCATGATGGAATTTTTGGAATGGCACTTAGTCCTGTGACGAACAATCTTTA
                ************************************************************

AcMRJP6         TTATAGCCCTCTCACTTCTCACAGTTTGTATTACGTTAACACGGAACCATTTATGAAATC
AmMRJP6         TTACAGCCCTCTCACTTCTCACAGTTTGTATTATGTTAACATGGAACCATTCATGAAATC
                *** ***************************** ******* ********* ********

AcMRJP6         ACAATATGGAGAAAATAATATACAATATGAAGGAATTCAAGATATTTTCAACACTCAATC
AmMRJP6         ACAATATGAAGAAAATAATATAGAATATGAAGGAATCCAAGATATTTTCAACACTCAATC
                ******** ************* ************* ***********************

AcMRJP6         ATCCGCTAAAGTAATGTCGAAAAATGGCGTCCTTTTCTTCGGACTTGTGAATAATTCAGC
AmMRJP6         GTCTGCTAAAGTAATGTCGAAAAATGGCGTCCTTTTCTTCGGACTTGTGAATAATTCAGC
                 ** ************************************************************

AcMRJP6         TATTGGTTGTTGGAACGAGCATCAACCACTTCAGAAACAAAATATGGATATGGTCGCTCA
AmMRJP6         TATTGGTTGTTGGAACGAGCATCAACCACTTCAGAGACAAAATATGGATATGGTCGCTCA
                ********************************** ************************

AcMRJP6         GAATGAAGAGACACTTCAAATAATCACTAGTGTGAAAATTATACAAAATCTTCCATATTC
AmMRJP6         GAATGAAAAGACACTTCAAATGATCATTAGCGTGAAAATTATACAAAATCTTGCATATTC
                ******* ************* **** *** ********************** *******

AcMRJP6         CGGAAAGATGAATAGAATTCACAAGAATGAATATATGTTGGCTTTAAGTAACAGAATGCA
AmMRJP6         CGGAAGGATGAATAGAATTCACAAGAATGAATATATGTTGGCTTTAAGTAACAGAATGCA
                ***** ******************************************************

AcMRJP6         GAAAATAGTAAACAATGATTTTAATTTCAACGACATAAATTTCCGAATATTGGGTGCGAA
AmMRJP6         GAAAATAGTAAACAATGATTTTAATTTCGACGAAGTAAACTTTCGAATTTTGGGTGCGAA
                **************************** ****  **** ** ***** ***********

AcMRJP6         TGTAAAGAACTTAATAAAAAACACTCGTTGTGCAAATTCTAAAAATCAGAATAACAATCA
AmMRJP6         TGTAAATAACTTAATAAAAAACACTCGTTGTGCAAAGTCTAACAATCAGAATAACAATCA
                ****** ***************************** ***** ****************

AcMRJP6         AAAGAAACATAAGAATCAAGCTCATT------AGATCTTTTCCAAGATCATATTAAATTC
AmMRJP6         AAATAAATATAAGAATCAAGCTCATTTAGATTAGATCTTTTTCAAGATCATATTAAAATC
                *** *** ******************      ********* *************** **

AcMRJP6         TATAGATTAATTTTTTCTCGTGGTAAATCAAATATTTTTAAAAATTTATTTGCATTATAA
AmMRJP6         TATAGATCAACTTTTTTTCGTGGTAAATCAAATATTTTTAAAAATTTATTTGCATTATAA
                ******* ** ***** ******************************************

AcMRJP6         ATTAATAAAATAAATATCATTTTCGCATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AmMRJP6         ATGAATAAAATAAATATCGTTTTCGCATG-----------------------------
                ** ************** *********

AcMRJP6         AAAAAAAAAAAAA
AmMRJP6         -------------
```

Figure 3.15 (continued)

```
CLUSTAL X (1.81) multiple sequence alignment

AcMRJP6          MTKWLLLIVCLSIACQDVTSAIHRRKSSKNLEHSMNVIHEWKYLDYDFDTNEKKQAAIQF
AmMRJP6          MTNWLLLIVCLSIACQDVTSAIHQRKSSKNLEHSMNVIHEWKYIDYDFGSDEKRQAAIQS
                 **:********************:********************:****.::**:*****

AcMRJP6          GEYDYTKNYPFDVDQWHDKTFVAVIRYDGVPSSLNVISDKTGNGGRLLQPYPDWSWTNYK
AmMRJP6          GEYDYTKNYPFDVDQWHNKTFLAVIRYDGVPSSLNVISEKIGNGGCLLQPYPDWSWANYK
                 *****************:***:****************:* **** **********:***

AcMRJP6          DCSGIVSVYKIAIDKFDRLWVLDSGLINNIQLMCSPKLLAFDLTTSKLLKQVEIPYDIAV
AmMRJP6          DCSGIVSAYKIAIDKFDRLWVLDSGLINNIQLMCSPKLLAFDLNTSKLLKQIEIPHNIAV
                 *******.************************************.*******:***::***

AcMRJP6          NASTGMGGLVSLVVQAMDPMNTMVYIADDRGDALIIYQNSDDSFHRLSSNTFDNDPRYSE
AmMRJP6          NASTGMGGPVSLVVQAMDPMNTTVYIADDRGDALIIYQNSDDSFHRLTSKTFDNDLRYSE
                 ******** ************* ************************:*:***** ****

AcMRJP6          LTVAGESFTVHDGIFGMALSPVTNNLYYSPLTSHSLYYVNTEPFMKSQYGENNIQYEGIQ
AmMRJP6          LAVAGESFTVHDGIFGMALSPVTNNLYYSPLTSHSLYYVNMEPFMKSQYEENNIEYEGIQ
                 *:************************************ ******** ****:*****

AcMRJP6          DIFNTQSSAKVMSKNGVLFFGLVNNSAIGCWNEHQPLQKQNMDMVAQNEETLQIITSVKI
AmMRJP6          DIFNTQSSAKVMSKNGVLFFGLVNNSAIGCWNEHQPLQRQNMDMVAQNEKTLQMIISVKI
                 ************************************:**********:***:* ****

AcMRJP6          IQNLPYSGKMNRIHKNEYMLALSNRMQKIVNNDFNFNDINFRILGANVKNLIKNTRCANS
AmMRJP6          IQNLAYSGRMNRIHKNEYMLALSNRMQKIVNNDFNFDEVNFRILGANVNNLIKNTRCAKS
                 ****.***:****************************:::*********:*********:*

AcMRJP6          KNQNNNQKKHKNQAH--
AmMRJP6          NNQNNNQNKYKNQAHLD
                 :******:*:*****
```

Figure 3.16 Alignment of the deduced amino acid sequence of AcMRJP6 cDNA with AmMRJP6 cDNA published sequence (GenBank Acc. AY313893). Conserve residues are indicated in asterisks. **:** means amino acid which have the same group of side chains and similar size while **.** means amino acid which have the same group of side chains but different size.

Table 3.2 Molecular characterization of cDNAs and deduced amino acid sequences of AcMRJP.

| Family | DNA insert size* (bp) | Deduced amino acid (residues) | No. of N-glycosylation site | Amino acid residues without signal peptide | Molecular weight (kDa) | pI | Reference |
|---|---|---|---|---|---|---|---|
| MRJP1 | 1421 | 433 | 3 | 413 | 46.7 | 5.40 | AF525776 |
| MRJP2 | 1565 | 463 | 2 | 446 | 50.6 | 7.78 | AF525777 |
| MRJP3 | 2005 | 608 | 6 | 588 | 67.3 | 8.79 | Srisuparbh (2002) |
| MRJP4 | 1608 | 485 | 7 | 465 | 52.8 | 5.84 | AY532368 |
| MRJP5 | 1881 | 579 | 5 | 559 | 66.2 | 8.75 | AY532369 |
| MRJP6 | 1450 | 435 | 2 | 415 | 47.4 | 6.44 | - |

The characteristics were predicted by computer analysis of AcMRJP cDNA sequence and deduced amino acid sequence without their signal peptides.

* including polyA tail

Table 3.3 Amino acid composition of AcMRJPs

| | MRJP1 | MRJP2 | MRJP3 | MRJP4 | MRJP5 | MRJP6 |
|---|---|---|---|---|---|---|
| Ala | 4.1 | 5.2 | 5.1 | 5.2 | 4.3 | 4.8 |
| **Arg** | **4.1** | **2.5** | **5.6** | **3.2** | **8.9** | **2.9** |
| Asn | 8.0 | 14.8 | 20.2 | 14.0 | 10.0 | 10.6 |
| Asp | 8.0 | 6.1 | 8.0 | 8.0 | 11.3 | 7.0 |
| Cys | 1.5 | 1.1 | 0.9 | 0.9 | 0.7 | 1.0 |
| Gln | 3.6 | 5.6 | 8.3 | 6.9 | 4.3 | 5.3 |
| Glu | 4.1 | 4.3 | 3.2 | 3.2 | 2.1 | 3.4 |
| Gly | 5.6 | 5.6 | 4.6 | 3.9 | 3.8 | 5.1 |
| **His** | **2.4** | **2.7** | **2.0** | **3.7** | **2.3** | **2.7** |
| **Ile** | **5.8** | **5.4** | **3.2** | **3.0** | **5.4** | **7.0** |
| **Leu** | **9.2** | **8.1** | **4.9** | **8.6** | **5.2** | **7.7** |
| **Lys** | **5.3** | **8.1** | **6.6** | **5.6** | **5.4** | **6.7** |
| **Met** | **2.7** | **2.7** | **1.4** | **2.2** | **12.5** | **3.4** |
| **Phe** | **4.1** | **3.8** | **3.1** | **2.8** | **2.9** | **4.1** |
| Pro | 3.9 | 3.4 | 2.6 | 2.2 | 2.7 | 3.1 |
| Ser | 9.2 | 5.6 | 6.0 | 9.7 | 5.5 | 8.0 |
| **Thr** | **5.8** | **4.5** | **4.3** | **4.3** | **3.9** | **4.6** |
| **Trp** | **1.2** | **1.1** | **1.0** | **1.3** | **1.1** | **1.4** |
| Tyr | 4.6 | 3.6 | 2.7 | 3.9 | 3.4 | 5.1 |
| **Val** | **6.8** | **6.1** | **6.3** | **7.7** | **4.3** | **6.3** |
| Ess. aa. | 47.4 % | 45 % | 38.4 % | 42.4 % | 51.9 % | 46.8 % |

Percent content of amino acid in native protein was obtained by computer analysis of its sequence employing the program ProtParam. Essential amino acids are marked in boldface.

```
CLUSTAL X (1.81) multiple sequence alignment


AcMRJP2         MTRWLFMVACLGIACQG----AIIRQ-NSAKNLENSLNVIHEWKYIDYDFGSEERRQAAI
AcMRJP3         MTKWLLLVVCLGIACQDVTSAAVNHQRKSSKNLAHSMKVIYEWKHIDYDFGSVERRDAAI
AcMRJP1         MTRWLFMVVCLGIVCQG-----TTSSILRGESLNKSLSVLHEWKFFDYDFDSDERRQDAI
AcMRJP5         MTSWLLLVVCLGIACQGITG--ATVRENSSRNLANSMNVIHEWKYLDYDFGSDEKRQAAI
AcMRJP6         MTKWLLLIVCLSIACQDVTS--AIHRRKSSKNLEHSMNVIHEWKYLDYDFDTNEKKQAAI
AcMRJP4         MTKWLLLMACLGIACQNIRG-AVVRENSSRKKLTNTLNVIHEWKYVDYDFGSDEKRQAAI
                ** **:::.**.*.**.             ..* :::.*::***..****.: *::: **

AcMRJP2         QSGEYDHTKNYPFDVDQWHDKTFVTILKYDGVPSTLNMISNKIGKGGRLLQPYPDWSWAE
AcMRJP3         KSGEFDHTKNYPFDVDRWRDKTFVTVERFDGVPSSLNVVTNKKGKGGPLLHPYPDWSWAN
AcMRJP1         LSGEYDYRKNYPSDVDQWHGKIFVTMLRYNGVPSSLNVISKKIGDGGPLLQPYPDWSFAK
AcMRJP5         QSGEYDHTKNYPFDVDRWHDMTFVTVLRYKGVPSSLNVISKKIGNGGPLLQPYPDWSWAN
AcMRJP6         QFGEYDYTKNYPFDVDQWHDKTFVAVIRYDGVPSSLNVISDKTGNGGRLLQPYPDWSWTN
AcMRJP4         QSGEYDRTKNYPLDVDQWHDKTFVTMLRYDGVPSSLNVVSDKTGNGGPLLQPYPDWSFAK
                  **:*  **** ***:*:.  **:: ::.****:**::::.* *.** **:******:::

AcMRJP2         NKDCSGIVSAFKIAIDKFDRLWVLDSGLINRTEPICAPKLHVFDLKNTKHLKQIEIPHDI
AcMRJP3         YKDCSGIVSAFKIAVDKFDRLWVLDSSLVNNNQPMCSPKLVTFDLNTSKLLKQVEIPHNI
AcMRJP1         YDDCSGIVSATKLAIDKCDRLWVLDSGLVNNTQPMCSPKLLTFDLTTSQLLKQVEIPHDV
AcMRJP5         YKDCSGIVSAYKIAIDKFDRLWVLDSGIINNTQPMCSPKLHVFDLNTSQQIKQVMMPHDI
AcMRJP6         YKDCSGIVSVYKIAIDKFDRLWVLDSGLINNIQLMCSPKLLAFDLTTSKLLKQVEIPYDI
AcMRJP4         YEDCSGIVSANKIAIDEYERLWVLDSGLVNNIQPMCSPKLLAFDLTTSKLLKQVEIPHDV
                 .*******. *:*:*: :*******.::*. : :*:*** .***..:: :**: :*:::

AcMRJP2         AVNATTGKGGLVSLVVQAMDP---MNTLVYIADHKGDALIVYQNSDDSFHRMTSNTFDYD
AcMRJP3         AVNATTEWGELVSLAVQAVDP---TNTMVYIADERGEASIIYQNSDDSFHRLTSNTFDYD
AcMRJP1         AVNATTGKGRLSSLAVQPLDCNINGDTMVYIADEKGEGLIVYHDSDNSFHRLTSKTFDYD
AcMRJP5         AINATTGKGGLENLVVQAMDP---MNTLVYMADNKGDALIVYQNSDDSFHRLTSNTFDYD
AcMRJP6         AVNASTGMGGLVSLVVQAMDP---MNTMVYIADDRGDALIIYQNSDDSFHRLSSNTFDND
AcMRJP4         AVNATTGKGGLASLAVQAMDS---VNTMVYMADNKDDALIVYQNADDSFHRLSSHISNHN
                *:**:*  * * .*.**.:*     :*:**:**.:.:. *:*:::*:****::*:  : :

AcMRJP2         PRYAKMTINGESFTLKNG-ICGMALSPVTNNLYYSPLASHGLYYVNTEPFMKSQFGDNNN
AcMRJP3         PRYTKLTVAGESFTVKNG-ICGIALSPVTNNLYYSPLASHSLYYVNTEQFRNPQYEENN-
AcMRJP1         PKFTKMTINGESFTTQSG-ISGMALSPMTNNLYYSPVASTSLYYVNTEQFRTSNYEQNA-
AcMRJP5         PKYIKMMAAGESFTLQDG-IFGMALSPMTNNLYYSPLASRSLYYVNTKPFMKSQYGTNN-
AcMRJP6         PRYSELTVAGESFTVHDG-IFGMALSPVTNNLYYSPLTSHSLYYVNTEPFMKSQYGENN-
AcMRJP4         FRSDKMSQENLTLKEVDNRVFGMALSSVTHNLYYSPLSSQNLYYVNTTSLMNSQNQGND-
                 :  ::   . ::.  .. : *:***.:*:******::* .******  : ..:   *

AcMRJP2         VQYEGSQDTLNTQSLAKAVSKDGVLFVGLVGNSALGCLNEHQPLQRENLELVAQNEKTLQ
AcMRJP3         VQYEGSQDILNTQSFAKAVSKNGVVFLGLVSNSTVGCVNEHQVLQKENFDVVAQNEETLQ
AcMRJP1         VHYEGVQNILDTQSSAKVVSKSGVLFFGLVGDSALGCWNEHRSLERHNIRTVAQSDETLQ
AcMRJP5         VQHEGVQDIFNTQSIAKIMSKNGVLFFGLMNNSAIGCWNEHQPLQRQNMDMVAQNEETLQ
AcMRJP6         IQYEGIQDIFNTQSSAKVMSKNGVLFFGLVNNSAIGCWNEHQPLQKQNMDMVAQNEETLQ
AcMRJP4         VQYESVQDVFSSQLSAKAVSKNGVLFFGFTNN-TLGCWNEHQSLDRQNIDIVARN-ETLQ
                :::*. *: :.:*  ** :**.**:*.*: .: ::** ***: *:::*:  **:. :***
```

(continued)

Figure 3.17 Alignment of deduced amino acid sequence of AcMRJP cDNA. Accession number of AcMRJP as follows: AcMRJP1, AF525776; AcMRJP2, AF525777; AcMRJP4, AY532368; AcMRJP5, AY532369. AcMRJP3 amino acid sequence obtained from D. Srisuparbh (2002). Conserve residues are indicated by asterisks.

```
AcMRJP2         MIAGMKIKEELPHFVGSNKPVK-----------------------------------
AcMRJP3         MIVSMKIMQDLPQSGRINDPG-----------------------------------
AcMRJP1         MIVGMKIKEALPHVPIFDRYIN----------------------------------
AcMRJP5         TVVAMKMMH-LPQSNRMNRMHRMNSMNRMDRMDRMDRMDRMDRMDRMDRMDRMDRMD
AcMRJP6         IITSVKIIQNLPYSGKMNRIH-----------------------------------
AcMRJP4         MVVGMKIKQNLPQSGKVNNTQR----------------------------------
                 :..:*: . **     :

AcMRJP2         ------------------------------------------------------------
AcMRJP3         ------------------------------------------------------------
AcMRJP1         ------------------------------------------------------------
AcMRJP5         RMDRMDRMDRMDIMDRTNKMDRMDRMDIMDKMNKMDRMDSMIRIDKMDRMDRMHRIDIMN
AcMRJP6         ------------------------------------------------------------
AcMRJP4         ------------------------------------------------------------


AcMRJP2         ----------------------------------------------------DEYMLVLSNK
AcMRJP3         ----------------------------------------------------NEYMLALSNK
AcMRJP1         ----------------------------------------------------REYILVLSNR
AcMRJP5         RMDRMDRMDTRIDTRMDRMDRMDKMDKINKMHRMGRMDRMDRMNRMNRQMNEYMMALSMK
AcMRJP6         ----------------------------------------------------KNEYMLALSNR
AcMRJP4         ----------------------------------------------------NEHLLALTNK
                                                                     *::::*: :

AcMRJP2         MQKIVNNDFNFNDVNFRILGANVKELMRNTHCANFNNK--NNQKNNNQKNNNQN------
AcMRJP3         MQKIINNDFNFNDVNFRILGANVNHLTRNTRCAKSNNQNANNQNANNQNATNQNDTNQND
AcMRJP1         MQKMANNDYNFNDVNFRIMDANVNDLILNTRCENPNND-----NTPFKISIHL-------
AcMRJP5         LQKFINNDYNFNEVNFRILAANVNDLIMNTRCANSNNQ-----NDNQNKHNN--------
AcMRJP6         MQKIVNNDFNFNDINFRILGANVKNLIKNTRCANSKNQ-----NNNQKKHKNQAH-----
AcMRJP4         KQDVLNNDLNLEHVNFQILDANVNDLIRNSRCANSDNQDNNQHNYNHNQVRHSSKSDN--
                 *.. *** *:::**:*: ***:.*  *::* : .*.     :   :   :

AcMRJP2         ------------------------------------------------------------
AcMRJP3         NGTNRRNGNNQNGNRQNDNKQNDNKQNANKQNANKQNANKQNDNKQNDNKQNGNRQNDNR
AcMRJP1         ------------------------------------------------------------
AcMRJP5         ------------------------------------------------------------
AcMRJP6         ------------------------------------------------------------
AcMRJP4         ------------------------------------------------------------


AcMRJP2         ------------------------------------------------------------
AcMRJP3         QNDNKQNDNRQNDNKQNGNRQNGNRQNDNKRNGNRQNDNRQNDNKRNSNRQNDNRQNDNK
AcMRJP1         ------------------------------------------------------------
AcMRJP5         ------------------------------------------------------------
AcMRJP6         ------------------------------------------------------------
AcMRJP4         ----------------------------------------------------QNNNQ


AcMRJP2         -----NNNQKNNNQKNNNQKNNNQKNNNQNTNN--------
AcMRJP3         RNGNRQNDNKQNDNKQNDNRQNDNNQNDNQNDNNRNNQAHHS
AcMRJP1         ----------------------------------------
AcMRJP5         ----------------------------------------
AcMRJP6         ----------------------------------------
AcMRJP4         HNNQAYHSSKSDNWDNNNNQAHHSSKFDNQNNNQYNN-----
```

Figure 3.17 (continued)

**Phylogenetic relationships between AcMRJPs and AmMRJPs families**

Nucleotide and deduced amino acid sequences of AcMRJP4 (AY532368), AcMRJP5 (AY532369) and AcMRJP6 cDNA obtained from this study and sequence of AcMRJP1 (AF525776), AcMRJP2 (AF525777) and AcMRJP3 (Srisuparbh, 2002) cDNA previously reported were aligned with the sequence of AmMRJP1-AmMRJP8 cDNA and *A. mellifera* yellow-f protein cDNA retrieved from the GenBank database [ AmMRJP1 (AF000633), AmMRJP2 (AF000632), AmMRJP3 (Z26318), AmMRJP4 (Z26319), AmMRJP5 (AF004842), AmMRJP6 (AY313893), AmMRJP7 (BK001420), AmMRJP8 (AY398690) and AmYellowP (Albert and Klaudiny, 2004)].

Genetic distances of each MRJP were calculated at both nucleotide and protein levels. The lowest and highest divergence at the nucleotide level was 0.0660 (AcMRJP1-AmMRJP1) and 0.5066 (AcMRJP3-AmMRJP5), respectively. At the protein level, the lowest divergence was 0.0990 (AcMRJP1-AmMRJP1) whereas the highest divergence was 0.8556 (AcMRJP3-AcMRJP5) (Table 3.4).

The original data was then bootstrapped 1000, and 500 times for nucleotide and protein data, respectively. Bootstrapped neighbor-joining trees were then constructed (Figure 3.18 and 3.19). Relationships at both nucleotide and protein levels of MRJPs indicated phylogenetically closer relationships between the same MRJPs families from different species rather than different families of MRJPs within the same species.

The identical trees were obtained from two types data either nucleotide and amino acid sequences. The same families from different species were grouped together. The result showed *A. mellifera* Yellow-f protein was a monophyletic group distant from MRJPs. The MRJP8 exhibited the earliest divergence within MRJPs gene families.

Table 3.4 Estimated genetic distance among MRJPs families of *A. cerana* (Ac) , *A. mellifera* (Am) and *A. mellifera* yellow protein (YP) obtained from nucleotide (above diagonal) and deduced amino acid (below diagonal) sequences.

| | Ac1 | Ac2 | Ac3 | Ac4 | Ac5 | Ac6 | Am1 | Am2 | Am3 | Am4 | Am5 | Am6 | Am7 | Am8 | YP |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Ac1 | - | 0.2630 | 0.2872 | 0.3276 | 0.3530 | 0.2882 | 0.0660 | 0.2573 | 0.2846 | 0.3613 | 0.3731 | 0.2899 | 0.2736 | 0.3848 | 1.0910 |
| Ac2 | 0.4504 | - | 0.2131 | 0.3068 | 0.3332 | 0.2010 | 0.2910 | 0.0875 | 0.2237 | 0.3365 | 0.3573 | 0.2086 | 0.2018 | 0.3624 | 1.1438 |
| Ac3 | 0.4900 | 0.4032 | - | 0.3460 | 0.4817 | 0.1981 | 0.3103 | 0.2137 | 0.1075 | 0.3489 | 0.5066 | 0.1883 | 0.2422 | 0.3846 | 1.1233 |
| Ac4 | 0.5599 | 0.5790 | 0.6351 | - | 0.3995 | 0.2555 | 0.3419 | 0.3009 | 0.3590 | 0.0978 | 0.4393 | 0.2673 | 0.2825 | 0.4083 | 1.1892 |
| Ac5 | 0.6246 | 0.5709 | 0.8556 | 0.6969 | - | 0.2747 | 0.3706 | 0.3359 | 0.4326 | 0.4258 | 0.0966 | 0.2716 | 0.2609 | 0.4316 | 1.2190 |
| Ac6 | 0.4988 | 0.3627 | 0.3236 | 0.4668 | 0.4612 | - | 0.3163 | 0.2228 | 0.2201 | 0.2822 | 0.2785 | 0.0687 | 0.2106 | 0.3442 | 1.1964 |
| Am1 | 0.0990 | 0.4687 | 0.5183 | 0.5665 | 0.6319 | 0.5093 | - | 0.2817 | 0.3086 | 0.3797 | 0.3929 | 0.3136 | 0.2894 | 0.4028 | 1.0972 |
| Am2 | 0.4462 | 0.1683 | 0.4341 | 0.5214 | 0.5398 | 0.3930 | 0.4518 | - | 0.2256 | 0.3350 | 0.3462 | 0.2307 | 0.2020 | 0.3714 | 1.1227 |
| Am3 | 0.4670 | 0.3913 | 0.1760 | 0.6046 | 0.7936 | 0.3733 | 0.5032 | 0.4103 | - | 0.3472 | 0.4654 | 0.2185 | 0.2612 | 0.3888 | 1.1101 |
| Am4 | 0.6253 | 0.6560 | 0.6549 | 0.1969 | 0.7371 | 0.5060 | 0.6432 | 0.6135 | 0.5939 | - | 0.4383 | 0.2863 | 0.3149 | 0.4168 | 1.1986 |
| Am5 | 0.6348 | 0.6166 | 0.7548 | 0.7498 | 0.1898 | 0.4823 | 0.6580 | 0.5710 | 0.7121 | 0.7645 | - | 0.2687 | 0.2759 | 0.4461 | 1.2227 |
| Am6 | 0.5032 | 0.3680 | 0.3217 | 0.4936 | 0.4667 | 0.1036 | 0.5228 | 0.4061 | 0.3748 | 0.5250 | 0.4835 | - | 0.2253 | 0.3568 | 1.2031 |
| Am7 | 0.4586 | 0.4001 | 0.4447 | 0.5309 | 0.5100 | 0.4165 | 0.4988 | 0.3963 | 0.4625 | 0.5971 | 0.5275 | 0.4420 | - | 0.3426 | 1.1580 |
| Am8 | 0.6059 | 0.6192 | 0.6578 | 0.7297 | 0.7360 | 0.6390 | 0.6111 | 0.6299 | 0.6554 | 0.7887 | 0.7667 | 0.6390 | 0.6313 | - | 1.2479 |
| YP | 2.0515 | 2.0080 | 1.9782 | 2.2652 | 2.2173 | 2.0515 | 2.0773 | 2.1309 | 1.8883 | 2.4147 | 2.1877 | 2.0515 | 2.3122 | 1.8326 | - |

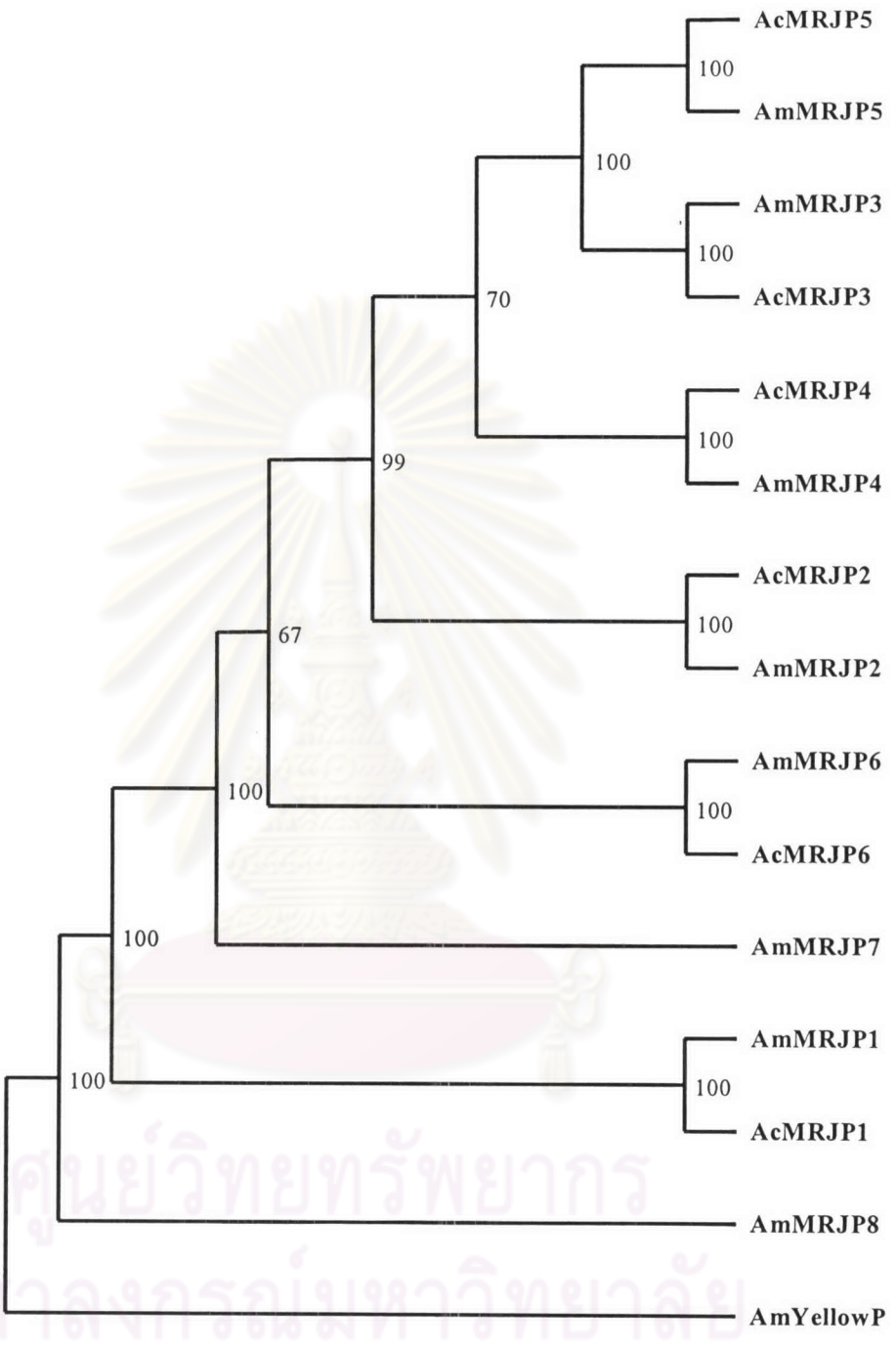


Figure 3.18 A bootstrapped tree illustrating relationship of MRJPs of *A. cerana*, *A. mellifera* and *A. mellifera* yellow protein. The original nucleotide sequence data was bootstrapped 1000 times. Values at the node indicate the percentage of times occurred out of 1000 trees.

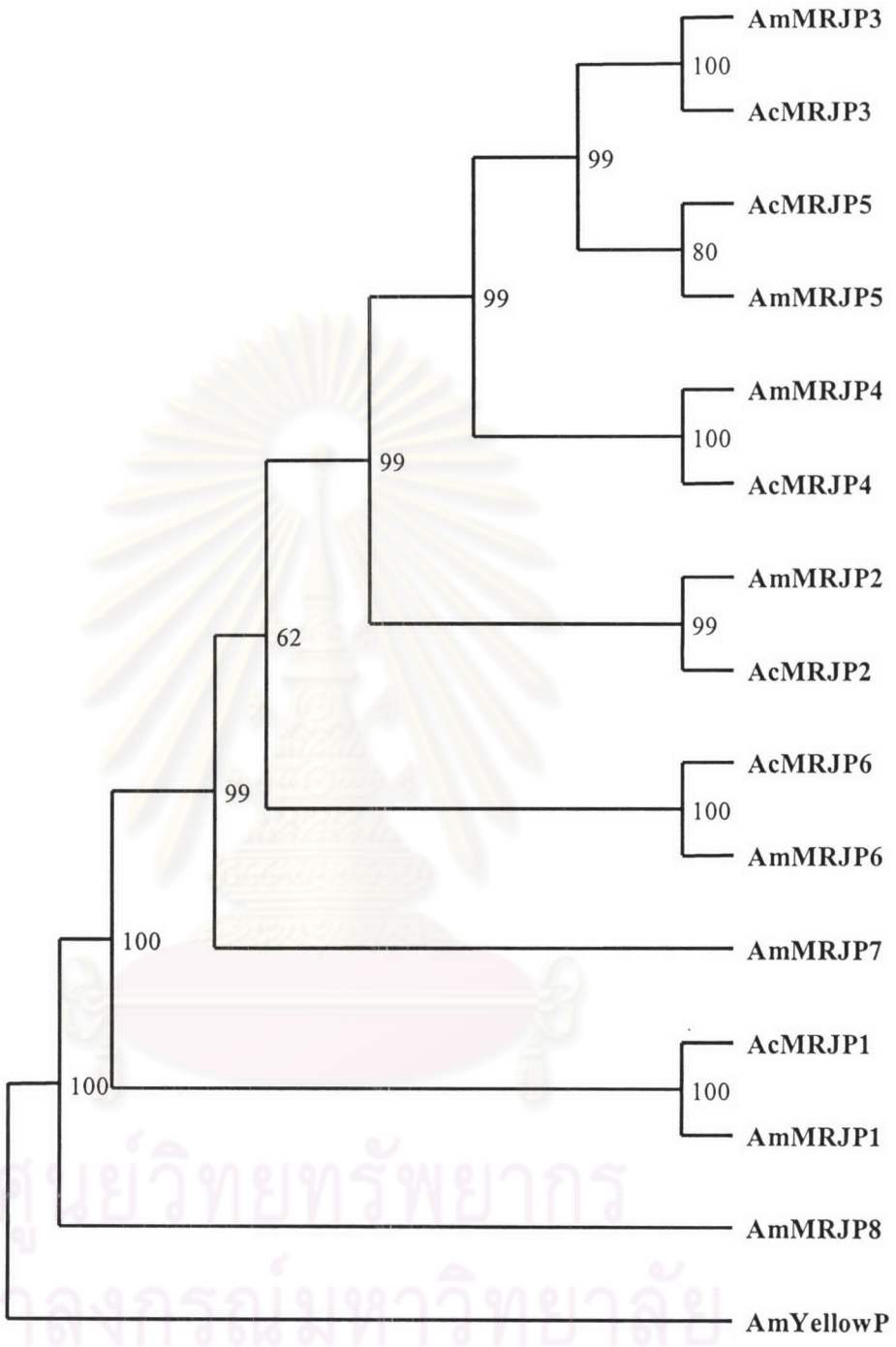


Figure 3.19 A bootstrapped tree illustrating relationship of MRJPs of *A. cerana*, *A. mellifera* and *A. mellifera* yellow protein. The original deduced amino acid sequence data was bootstrapped 500 times. Values at the node indicate the percentage of times occurred out of 500 trees.

### 3.2 Overexpression of AcMRJP4 protein in *E.coli*

In this study, The pET system was used for overexpression of AcMRJP4 protein in *E. coli* system. *E. coli* Rosetta (DE3) pLysS was selected for expression of the insert DNA, AcMRJP4 cDNA of pET 19b vector. The pET 19b contains T7 *lac* promotor, amplicillin resistance gene, *lac*I gene, sequence encoded a Histidine peptide (His-tags) at N-terminal of produced protein and sequence encoded protein sequences for enterokinase cleavage site. The His-tags was added for advantage in purification and identification of expressed protein while enterokinase cleavage site added for removed His-tags from the expressed protein.

*E. coli* Rosetta (DE3) pLysS contains T7 RNA polymerase gene and plasmid that harboured the rare tRNA genes, T7 lysozyme gene and chloramphenicol resistance gene.

For amplification of AcMRJP4 cDNA, primers were newly designed from AcMRJP4 cDNA sequence. The forward primers was designed over predicted N-terminal amino acid sequence without signal peptide and added *Nde*I restriction site to the 5'-end of the primer. The reverse primer was designed over stop codon and *Bam*HI restriction site was added to the 5'-end of the primer. Recombinant plasmid contained AcMRJP4 cDNA from transformant number 5 (MRJP405) was used as template DNA for PCR amplification process. The *Pfu* DNA polymerase that have 3' to 5' exonuclease activity was used in PCR reaction. After amplification reaction was completed, amplified product was eletrophoretically analyzed through 1 % agarose gel.

Only single PCR product of 1,400 bp was obtained (Figure 3.20). The PCR product was digested with proteinase K and purified by NucleoSpin® column. The pET 19b vector and purified PCR product were digested with *Nde*I and *Bam*HI restriction

endonuclease. The digested products were eletrophoretically analyzed and eluted from agarose gel by using QIAquick gel extraction kit. The *Nde*I-*Bam*HI digested PCR product was ligated with *Nde*I-*Bam*HI digested pET 19b vector, and then eletro-transformed into *E. coli* JM109. Twelve white colonies containing the recombinant plasmid were randomly picked for plasmid extraction and double digested with *Nde*I and *Bam*HI. The result of eletrophoretically analyzed show that all of twelve clones contained recombinant plasmid with AcMRJP4 cDNA fragment (Figure 3.21).

The transformed clone, transformant number 1 (clone name Exp401) was sent to Bioservice unit for plasmid extraction and sequencing. The T7 forward primer was used for sequencing. The amino acid sequence deduced from nucleotide sequence obtained showed that gene fragment had correct reading frame and 100% sequence identity with those of plasmid MRJP405 (Figure 3.22). The recombinant plasmid Exp401 and pET 19b vector were eletro-transformed to a competent *E. coli* Rosetta (DE3) pLysS. The transformants that grown on selective plate containing ampicillin and chloramphenicol were expected to be *E. coli* Rosetta containing the recombinant plasmid. The recombinant clones were identified by colony PCR method. The recombinant clone number one that contained the inserted gene and one clone that carried pET 19b vector (Figure 3.23) were induced by IPTG for AcMRJP4 protein production. For culture cell harbouring the vector DNA (contained T7 *lac* promoter), IPTG final concentration of 1 mM was used to induce the protein production. The cell pellet was collected before induction with IPTG as a reference, and then collected after induced at 1 hour interval for 5 hours. The cell pellets were analyzed by SDS-PAGE. The results showed that *E. coli* Rosetta carried pET 19b vector and *E. coli* Rosetta carried the recombinant plasmid at 0 hours of induction, the

expected protein band that overexpressed was not observed. The protein band of 53 kDa was observed at 1-5 hours after induction. The highest expression level was 3 hours after induction with IPTG. The molecular weight of induced protein band was 53 kDa, which corresponded well to 55.7 kDa, the calculated molecular weight of recombinant AcMRJP4 protein deduced from DNA sequence (Figure 3.24).

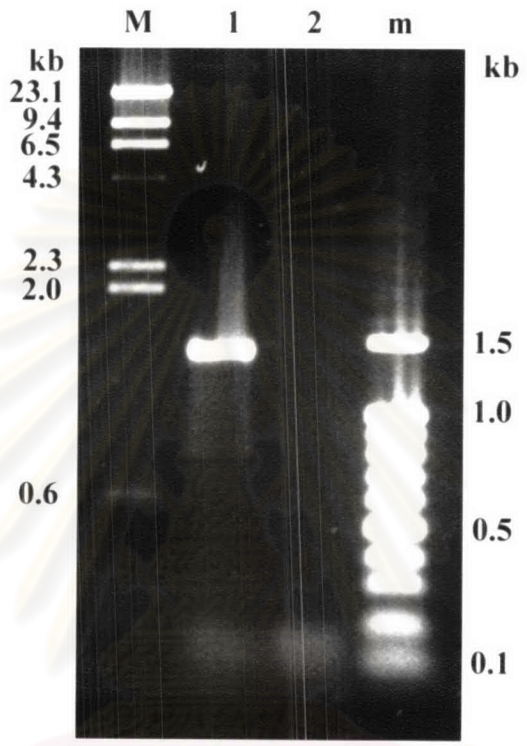


Figure 3.20 PCR amplification of AcMRJP4 cDNA without signal sq. for expression

Lane M = λ / *Hin*d III standard molecular weight marker

Lane 1 = The amplification products of AcMRJP4 cDNA

Lane 2 = Negative control

Lane m = A 100 bp DNA ladder

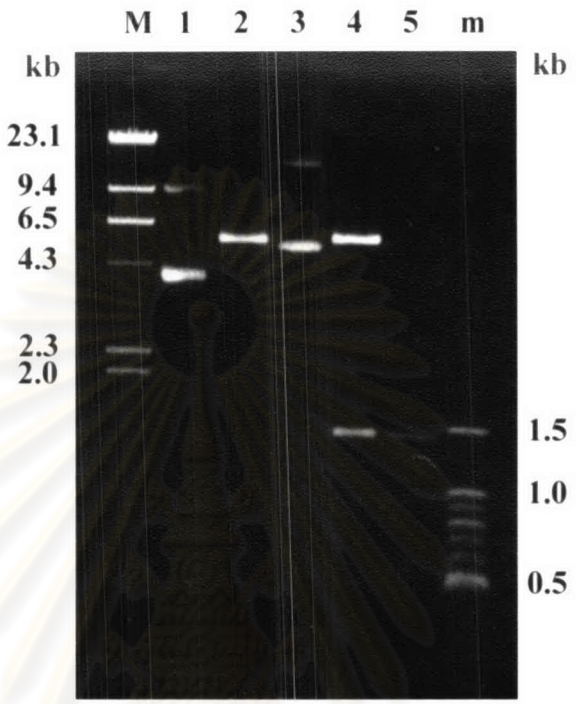


Figure 3.21 Cloning of pET19b expression vector containing AcMRJP4 cDNA in *E.coli* JM 109

Lane M = λ / *Hin*d III standard molecular weight marker

Lane 1 = undigested pET 19b

Lane 2 = pET 19b digested with *Nde* I and *Bam*HI

Lane 3 = undigested recombinant plasmid

Lane 4 = recombinant plasmid digested with *Nde* I and *Bam*HI

Lane 5 = AcMRJP4 insert digest with *Nde* I and *Bam*HI

Lane m = A 100 bp DNA ladder

```
CLUSTAL X (1.81) multiple sequence alignment

MRJP405         ----MTKWLLLMACPGIACQNIRGAVVRENSSRKKLTNTLNVIHEWKYVDYDFGSDEKRQ
Exp401          MGHHHHHHHHHHSSGHIDDDDKHMAVVRENSSRKKLTNTLNVIHEWKYVDYDFGSDEKRQ
                      :     :.  *  :: : **********************************

MRJP405         AAIQSGEYDRTKNYPLDVDQWHDKTFVTMLRYDGVPSSLNVVSDKTGNGGPLLQPYPDWS
Exp401          AAIQSGEYDRTKNYPLDVDQWHDKTFVTMLRYDGVPSSLNVVSDKTGNGGPLLQPYPDWS
                ************************************************************

MRJP405         FAKYEDCSGIVSANKIAIDEYERLWVLDSGLVNNIQPMCSPKLLAFDLTTSKLLKQVEIP
Exp401          FAKYEDCSGIVSANKIAIDEYERLWVLDSGL----------------------------
                ******************************

MRJP405         HDVAVNATTGKGGLASLAVQAMDSVNTMVYMADNKDDALIVYQNADDSFHRLSSHISNHN
Exp401          ------------------------------------------------------------

MRJP405         FRSDKMSQENLTLKEVDNRVFGMALSSVTHNLYYSPLSSQNLYYVNTKSLMNSQNQGNDV
Exp401          ------------------------------------------------------------

MRJP405         QYESVQDVFSSQLSAKAVSKNGVLFFGFTNNTLGCWNEHQSLDRQNIDIVARNETLQMVV
Exp401          ------------------------------------------------------------

MRJP405         GMKIKQNLPQSGKVNNTQRNEHLLALTNKKQDVLNNDLNLEHVNFQILDANVNDLIRNSR
Exp401          ------------------------------------------------------------

MRJP405         CANSDNQDNNQHNYNHNQVRHSSKSDNQNNNQHNNQAYHSSKSDNWDNNNNQAHHSSKFD
Exp401          ------------------------------------------------------------

MRJP405         NQNNNQYNN
Exp401          ---------
```

Figure 3.22 Alignment of the deduced amino acid sequence of AcMRJP4 cDNA transformant number 5 (MRJP405) with deduced amino acid sequence of AcMRJP4 cDNA in pET 19b vector transformant number 1 (Exp401). Conserve residues are indicated in asterisks. **:** means amino acid which have the same group of side chains and similar size while **.** means amino acid which have the same group of side chains but different size.

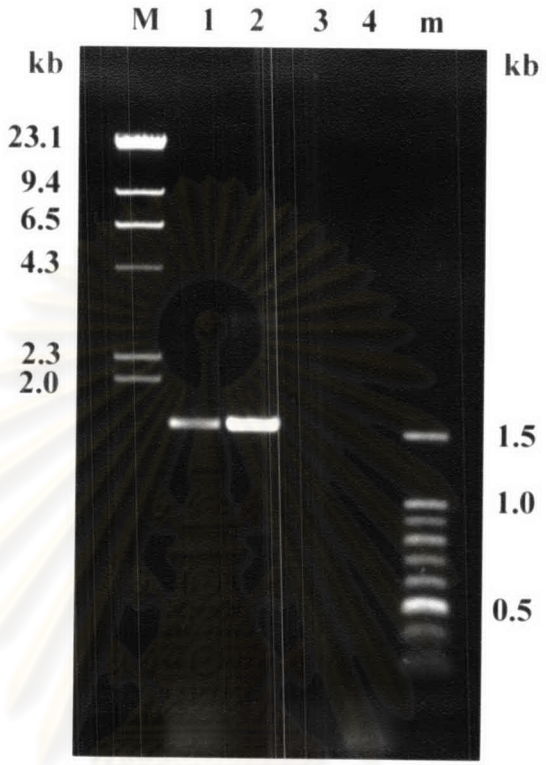


Figure 3.23 Colony PCR for identified cloning to expression host

Lane M = λ / *Hin*d III standard molecular weight marker

Lane 1 = Positive control 5 ng of vector containing AcMRJP4 cDNA

Lane 2 = The amplification products of vector containing AcMRJP4 cDNA in Rosetta host

Lane 3 = The amplification products of vector in Rosetta host

Lane 4 = Negative control

Lane m = A 100 bp DNA ladder

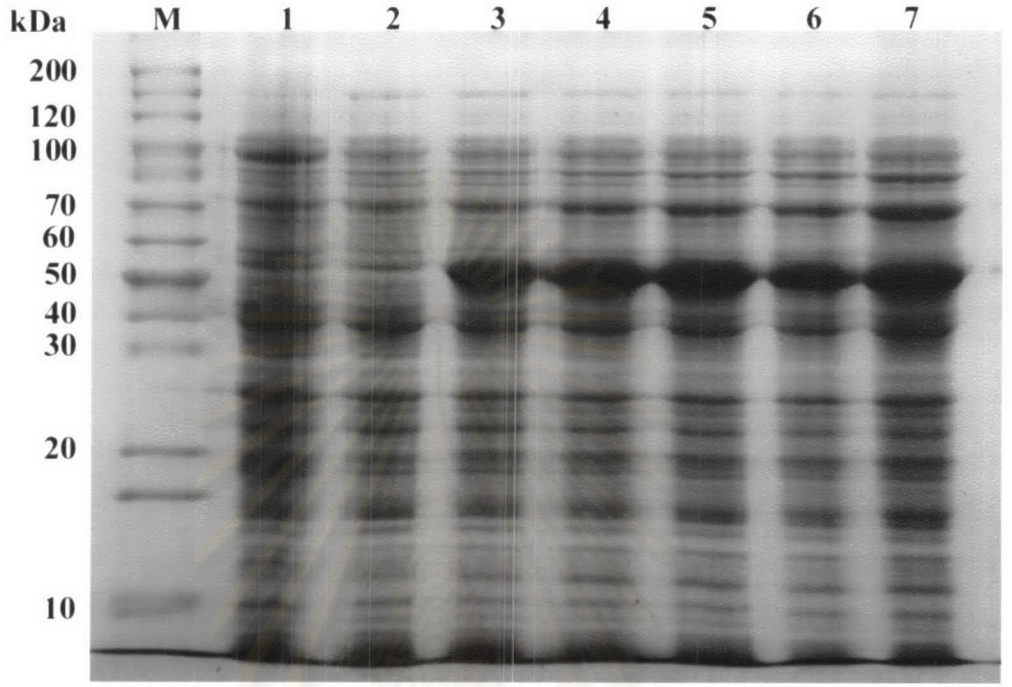


Figure 3.24 Protein pattern of crude extract of AcMRJP4 protein producing tranformant vary in induction time.

Lane M Protein molecular weight marker

Lane 1 *E. coli* Rosetta cells carried pET 19b vector at 0 hours induction.

Lane 2 *E. coli* Rosetta cells carried recombinant AcMRJP4 plasmid at 0 hours induction.

Lane 3 *E. coli* Rosetta cells carried recombinant AcMRJP4 plasmid at 1 hour induction.

Lane 4 *E. coli* Rosetta cells carried recombinant AcMRJP4 plasmid at 2 hours induction.

Lane 5 *E. coli* Rosetta cells carried recombinant AcMRJP4 plasmid at 3 hours induction.

Lane 6 *E. coli* Rosetta cells carried recombinant AcMRJP4 plasmid at 4 hours induction.

Lane 7 *E. coli* Rosetta cells carried recombinant AcMRJP4 plasmid at 5 hours induction.