CHAPTER I

INTRODUCTION

1.1 The biology of honey bee

The honeybee is one of the most important social insects in the world of agriculture and food industry. Many species of plant crops require pollination for increasing yield production. Commercial honey bee products such as honey, bee wax, royal jelly and pollen also have a high economic value in food industry.

The honeybee is an eusocial insect with three social castes: the queen, the infertile female workers and the males bees or drones. Each caste has different tasks to perform within the colony. A colony of honey bees has only one queen, a large number of workers and zero to a few hundred of drones depending on different annual period. The queen and worker bees are female, heterozygotes (diploid $2n = 32$) grown from fertilized eggs whereas drones are male, hemizygotes (haploid individuals) arising from unfertilized eggs (Wongsiri, 1989).

The queen, a mother of all members in the colony, is responsible for laying all the eggs for the colony and can exert marked influence on the behavior of the workers and the drones through the release of chemical signal called queen substances or pheromones. The queen mates with several drones and stores the sperm in her spermatheca for the remaining fertilization. After the queen is mated, the queen begins to lay eggs. The queen lays two types of eggs, fertilized and unfertilized. The unfertilized eggs develop into drones, while the fertilized eggs can develop into workers or queens, depending on what they are fed during development.

The vast majority of the individuals in a bee colony are worker bees. Distribution of task among workers depends on their ages. For the first two or three weeks, workers are called 'nurse' bees. They stay and work inside of the hive involved in synthesizing, secreting and feeding the royal jelly to young larvae and the queen. After this indoor period, they work in the field bring back nectar, pollen, water and propolis, and are called 'foragers'. The drones's only role in the hive is to mate with the queen (Robinson, 1991; Page and Peng, 2001).

Worker and queen bees hatch from fertilized eggs. There is no genetic difference between the egg, which produces a queen, and the egg, which develops into a worker. The difference between the castes is caused by the food given to the larvae by the nurse bees. Two different diets are fed to queen or worker larvae. The queen larvae are always mass-fed a diet of royal jelly throughout her entire life after emergence. Worker larvae are also mass-fed royal jelly for the first three days. After the fourth day the woker lavae are fed a mixture of woker jelly and pollen (Johansson, 1995).

Due to different dietary feeding which is particular rich in queen, mechanisms between queen and workers in the process of female caste determination and differentiation is observed. In addition, queen attains a larger size than workers and the reproductive organ is well developed to a mature stage and is able to lay several thousand eggs a day. In contrast, workers are smaller in size. The reproductive organ is not well developed but organs that are related with their tasks such as pollen baskets, mandibular, hypopharyngeal and wax glands are fully developed. Basically, the time required for development of the queen larvae to the adult stage is about 15.5 days. The life span of the adult queen is several years, while workers require 21 days for growing up with only a few months of life span (Krell, 1996).

In Thailand, there are five species of honey bee, *Apis dorsata*; the giant honey bees, *Apis florea*; the dwarf honey bees, *Apis andreniformis*; the small dwarf honey bees, *Apis cerana*; the eastern cavity-nesting honey bees, and *A. mellifera*; the cavity-nesting honey bee. For *A. mellifera*, this species is not native to Thailand, they are imported from Europe for a beekeeping purpose. *A. mellifera* and *A. cerana* can be kept and managed in hive for commercial beekeeping. Commercial beekeeping with *A. mellifera* is more successful than those of *A. cerana* due to non-aggressive behavior. However, *A. cerana* shows more disease resistance to bee mite especially *varroa jacobsoni* mite and exhibits better climatic adaptability than *A. mellifera* (Kavinseksan, 1994).

The taxonomic classification of *A. cerana* is as follows (Borror *et al.*, 1976; Gojmerac, 1980).

Phylum Arthropoda

Class Insecta

Order Hymenoptera

Family Apidae

Genus Apis

Species *Apis cerana*

1.2 The biology of royal jelly (RJ)

Royal jelly (RJ) or bee-milk looks like a yellowish white creamy with acidic material, constitutes the principal food of the queen bee and young larval of honey bees (Lercker, 1981). RJ is derived from secretions of both the hypopharyngeal and mandibular glands of nurse honey bee (Knecht and Kaatz, 1990; Lensky and Rakover, 1983). These glands are located in the head of nurse bees (Figure 1.1). RJ can be sold in various forms such as fresh RJ, freeze-dried RJ and mixed with other product. It is used as supplement food and used in cosmetics industry and has become a commercial product in Thailand. China is the world's largest producer and exporter of RJ while Japan has the highest domestic consumption of royal jelly, a large part of which is imported from other Asian countries.

Royal jelly has been demonstrated to possess several pharmacological activities in experimental animal. For example, RJ can lower the cholesterol in blood (Vittek, 1995), possesses the anti-inflammatory and wound healing properties (Fujii *et al.*, 1990) and can act as a potential immunomodulator for stimulating antibody production and immunocompetent cell proliferation in mice. However, RJ depresses humoral immune functions in rat (Sver *et al.*, 1996). In human, RJ consumption in the amounts of 50-100 mg per day can reduce total cholesterol levels by about 14% in people with moderately high cholesterol levels (Vittek, 1995).

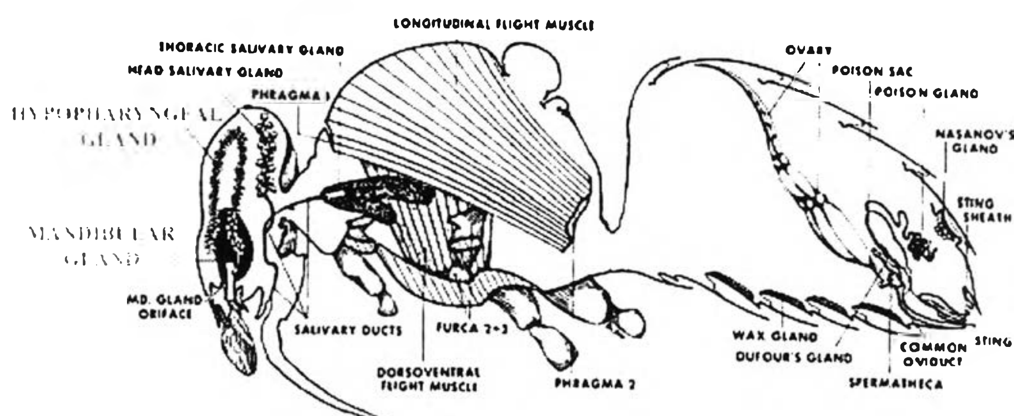

Figure 1.1 Diagram showing the organ systems of an adult female honeybee (Michener, 1974).

However, allergic reactions are the most common side effect when consume RJ. Allergic reactions from oral intake of royal jelly can range from very mild (e.g., mild gastrointestinal upset) to more severe reactions, including asthma, anaphylaxis (shock), intestinal bleeding, and even death in people who are extremely allergic to bee products (Thien *et al.*, 1996; Leung *et al.*, 1997 and Yonei *et al.*, 1997).

Several chemical analyses of royal jelly have been published. Royal jelly is acidic substance with pH between 3.6 to 4.2. The chemical composition of fresh RJ of *A. mellifera* contains moisture (66-70%), protein (12-17%), sugars (11-12.5%), lipids (3-5.5%) and mineral salts. In addition, the amino acids composition of RJ protein is also investigated. The RJ proteins are rich in essential amino acids (39.3% - 51.4%), the most abundance is aspartic acid (16.1%) and glutamic acid (10.19%). For lipids fraction, the major fatty acid is 10-hydroxy-2-decenoic acid (10-HDA) at an average concentration of 50.3% of the total fatty acid content. The sugars in royal jelly are dominant in fructose and glucose; fructose is prevalent in all RJ samples. In many cases, fructose and glucose together account for 90% of the total sugars (Howe, *et al.*, 1985, Palma, 1992 and Krell, 1996).

Recently, compositions of fresh RJ from northern and southern *A. cerana indica* and *A. cerana japonica* were also examined and compared with that of *A. mellifera* (Table 1.1). Northern and southern *A. cerana* RJ contains 48.8 – 49.6%

หอสมุดกลาง สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

moisture, 20.1 – 22.6% crude proteins, 4.6 – 6.9% lipid and 1.8 – 1.7% ash. Interestingly, moisture content of *A. cerana indica* RJ from Thailand is lower than that of *A.cerana japonica* from Japan and *A mellifera*, whereas crude proteins content, acidity of *A. cerana indica* RJ from Thailand are higher than those of *A. cerana japonica* from Japan and *A mellifera* (Trongnipatt, 2002; Takenaka and Takenaka, 1996).

Table 1.1 Composition of fresh RJ of *A. cerana indica*, *A. cerana japonica* and *A. mellifera*.

Composition	*A. cerana* in Thailand *(North)*	*A. cerana* in Thailand *(South)*	*A. cerana japonica*	*A. mellifera*
Moisture (%)	48.8±4.4	49.6±0.1	65.3	68.3
Crude protein (%)	20.1±1.4	22.6±1.0	16.4	12.7
Carbohydrate (%)	ND	ND	9.4	11.9
Lipid (%)	4.6±0.2	6.9±0.1	7.4	6.1
Ash (%)	1.8±0.2	1.7±0.2	1.5	1.0
Acidity *	47.1±4.6	45.0±1.6	39.3	42.2
Reference	Trongnipatt (2002)	Trongnipatt (2002)	Takenaka and Takenaka (1996)	Takenaka and Takenaka (1996)

- Acidity: Volume of 1N NaOH (ml)/100g of fresh RJ)
- ND = Not determined

1.3 Molecular biology of royal jelly protein

Royal jelly consists of many proteins with different molecular masses. However, there is little information regarding its physical and chemical properties. The activities of some enzymes in RJ were determined: invertase species (Kratky, 1931; Vittec and Janci, 1966; Halberstadt, 1970), amylase (Brouwers, 1982; Vittec and Janci, 1966), ascorbinoxidase (Gontarski, 1949), catalase (Vittec and Janci, 1966), acid phosphatase (Halberstadt, 1980) and insulin-like peptide (Kramer *et al.*, 1982; Kramer *et al.*, 1980). Apart from proteins with enzymatic activity, RJ contains mainly proteins with unknown physiological functions which makes their isolation and charcterization more difficult.

Royallizin was the first royal jelly protein which the complete amino acid sequence was characterized. Royalizin of *A. mellifera* bee is composed of 51 amino acid residues, with the calculated molecular weight of 5.5 kDa. It was found to have potent antibacterial activity against Gram-positive bacteria (Fujiwara, *et al.*, 1990).

MRJP (from major royal jelly protein) is the important protein components of RJ which belonging to a family named Major Royal Jelly Proteins (MRJPs). In this family five species of proteins (MRJP 1–5) with molecular masses ranging from 49–87 kDa have been identified.

MRJPs of *A. mellifera* RJ have been extensively studied, which focuses on characterization of both cDNAs and proteins. The cDNA library prepared from the head of 8-day-old nurse honeybees (*A. mellifera*) was constructed and screened with a polyclonal antiserum (Schmitzova *et al.*, 1998). It was done in parallel with electrophoretic analysed and *N*-terminal sequencing of RJ proteins. It was found that RJ contained major proteins and that all the proteins belong to one protein family designated MRJP (from major royal jelly protein). The family consists of five main members (MRJP1, MRJP2, MRJP3, MRJP4 and MRJP5) which comprise 82-90% of total larval jelly protein. MRJPs contain high amount of essential amino acids (39.3-51.4%), presumably that MRJPs have nutritional function in honeybee larval food. Amino acid compositions of *A. mellifera* MRJPs are illustrated in Table 1.2. All members of MRJP are glycoprotein. In addition, The MRJP gene family encodes a

group of closely related proteins that share a common evolutionary origin with the yellow protein of *Drosophila melanogaster*.

Recently, three new cDNA members of MRJPs were identified as AmMRJP6, AmMRJP7, and AmMRJP8. These nucleotide sequences were assembled by using the honeybee brain expressed sequence tags (EST) sequence database, honeybee genomic sequence data and nucleotide sequence from amplification product of the nursehoneybee head cDNA library (Albert and Klaudiny, 2004). A summary for molecular characterization of cDNA and deduced amino acid sequences of the AmMRJPs is illustrated in Table 1.3

MRJP1

AmMRJP1 protein has the N-terminal amino acid sequence NILRGESLNKS. From deduced amino acid sequence, this protein shows high amount of the 10 essential amino acids (48%). This protein is the most abundant protein in RJ of *A. mellifera*, that comprises 31% of the relative content as determined by SDS-PAGE. This protein exhibits the apparent molecular weights of 55 kDa that could be resolved into at least eight protein bands with similar isoelectric points of pH 4.5-5.0 (Hanes and Simuth, 1992; Schmitzova *et al.*, 1998).

Treatment of native 56-kDa protein (AmMRJP1) with N-glycosidase F showed the molecular weight of digestion products of 47 kDa in SDS-PAGE. Which nearly the same to that of the putaive protein lacking the signal sequence (47.1 kDa) predicted from nucleotide sequences. This clearly showed that AmMRJP1 is a glycoprotein (Ohashi *et al.*, 1997, and Schmitzova *et al.*, 1998).

AmMRJP1 was reported to have three different forms; a monomer (55 kDa), oligomer (approximately 420 kDa) and water insoluble aggregates resulted from interaction with fatty acids (Simuth, 2001). In RJ, MRJP1 associates with a small peptide named apisimin (Bilikova *et al.*, 2002) and possibly with other compounds in a large complex of 420 kDa. The oligomeric form of AmMRJP1 was water-soluble (Kimura *et al.*, 1996; Simuth, 2001).

MRJP1 mRNA was found to be differentially expressed in the heads of early emerged honeybees (Kucharski *et al.*, 1998), nurse and also forager honeybees

(Ohashi *et al.*, 1997). Its expression was localized to hypopharyngeal glands (Ohashi *et al*, 1997), and also to a subset of Kenyon cells (intrinsic neurons) of mushroom bodies-presumed centers of learning and memory in the honeybee brain (Kucharski *et al.*, 1998). Therefore, it would seem that MRJP1 is not only functions as a component of larval food but also plays a role in the honeybee brain.

The cDNA encoding AmMRJP1 without signal peptide sequence (409 amino acid residues) was cloned into pQE32 vector for expression in *E. coli* system. The expressed recombinant protein was purified and characterized by SDS-PAGE. The molecular weight of recombinant protein was 47 kDa whereas those of native AmMRJP1 was 55 kDa (Judova *et al.*, 1998).

MRJP2

AmMRJP2, the second most abundant of RJ, contained the N-terminal amino acid sequence of AIVRENSPRNLEK. The relative content of this protein is 16% in total RJ protein with 47% of the essential amino acid composition (Schmitzova *et al*, 1998). The apparent molecular weight of native AmMRJP2 protein was 49 kDa that could be resolved into at least eight variants with different isoelectric points of pH 7.5-8.5.

The cDNA encoding AmMRJP2 without signal peptide sequence (422 amino acid residues) was cloned and expressed in *E. coli* expression system. The recombinant AmMRJP2 protein was purified and characterized by SDS-PAGE. The result showed the molecular weight of recombinant MRJP2 protein was 49 kDa which was the same as those of native AmMRJP2 protein (Schmitzova *et al.*, 1998; Bilikova *et al.*, 1999).

From immunoblotting analysis, the 50 kDa protein that had the same molecular weight with MRJP2 was detected in hypopharyngeal gland of nurse bee but not in that of forager bee (Kubo *et al.*, 1996). Interesting; MRJP2 mRNA was found to be expressed in heads of experienced foragers when using microarrays and northern blot analysis (Kucharski and Malezka, 2002).

MRJP3

AmMRJP3 exhibited a size polymorphism as detected by SDS-PAGE (Albert *et al*, 1999b).. The apparent molecular masses of MRJP3 are between 60 and 70 kDa. They have almost identical N-terminal amino acid sequence: AAVNHQ(R/K) KSANNLAHS, with the amino acid sequence deduced from RJP57-1 cDNA. A relative content of AmMRJP3 is approximately 26% of total RJ protein. The essential amino acid content is 39.3% (Schmitzova *et al.*, 1998). The deduced amino acid of MRJP3 contains a repetitive region at the C-terminal part, repetitive motifs of XQNXX, typically with 20 repeated units (Klaudiny *et al.*, 1994). Polymorphism at DNA level was also found. PCR analysis of MRJP3 gene using genomic DNA as template show that repetitive region of the MRJP3 gene was highly polymorphic with as many as five alleles was found in 10 individuals from the same colony (Albert *et al*, 1999b). In other species, the study of repetitive sequence motifs in Giant bee, *Apis dorsata* showed that repettitive sequence also existed in MRJP3 gene liked those in *A. mellifera* (Albert *et al.*, 2002).

The AmMRJP3 protein was reported to have two different forms; a monomer (70 kDa) and trimer (210 kDa) (Okamoto *et al.*, 2003). MRJP3 mRNA and protein are expressed specifically in hypopharyngeal gland of nurse honeybees (Kubo *et al.*, 1996, Ohashi *et al.*, 1997). The amount of AmMRJP3 mRNA is 8% of total mRNA (Klaudiny *et al.*, 1994).

MRJP4

Only one clone (RJP57-2) from cDNA library of nurese bee head containing cDNA encoding MRJP4. It showed the lowest expression level (2% of total mRNA). The deduced amino acid of AmMRJP4 contains 44.5% essential amino acid content that was lower in overall essential amino acid content, but possesses high amount of some amino acids Leu (9.7%) and Val (8%). The calculated isoelectric focusing point of AmMRJP4 was 6.2 (Klaudiny *et al.*, 1994, Schmitzova *et al*, 1998).

MRJP4 could not be obtained from purification experiment and SDS-PAGE analysis of RJ. It can be characterized by two-dimensional gel electrophoresis of RJ following by N-terminal amino acid analysis. From two-dimensional gel electrophoresis of European honeybee (*A. mellifera*) RJ, two spots; HBRJ E27 and

E28, were identified as MRJP4. They possessed the N-terminal amino acid sequence (GVVRENSSRK) identical to those of MRJP4 (RJP57-2) cDNA previously reported by Klaudiny *et al.* (1994). For Africanized honeybee RJ, five spots; HBRJ A24 to A28, were identified as MRJP4, and possessed the N-terminal amino acid sequence of AVVRENSSRK. The N-terminal amino acid sequence of the 2 species were different by only one amino acid residue. The results from the study of Africanized and European honeybee RJ showed that the average molecular weight of MRJP4 was 60 kDa and the isoelectric focusing point was 5-6 (Sano *et al.*, 2004).

MRJP5

AmMRJP5 exhibited two different molecular weights (77 kDa and 87 kDa) on SDS-PAGE. They possess an identical N-terminal amino acid sequence of "VTV (R/N) E (N/Q) SPR". The relative content of AmMRJP5 was 9% of total RJ proteins and contains 51.4% essential amino acid, dominant in Arg (9%) and Met (11.4%) (Schmitzova *et al.*, 1998).

The deduced amino acid of AmMRJP5 inferred from MRJP5 cDNA showed the extensive repeat region located between 367 th and 540 th amino acid residues. This repeat unit was composed of a consensus sequence (GATAGAATG) which encodes for tripeptide (DRM): aspartic acid (D), arginine (R) and methionine (M) occurred 58 times and interrupts a conserved region of the MRJP consensus sequence. This repeat region was located at the C-terminal of this protein and invariant in repetitive unit size (Albert *et al.*, 1999a).

The MRJP5 repetitive region was characterized in Giant honeybee, *A. dorsata*. The repetitive region was located at the same position as found in *A. mellifera* but smaller in size, and occurred 23 times compared with 58 times in *A. mellifera* (Albert *et al.*, 2002). From two-dimensional gel electrophoresis, the MRJP5 proteins were found in both the Africanized and the European honey bee RJ. The MRJP5 protein from this 2 species posses the identical N-terminal amino acid sequence (VTVRENSPRK), however, molecular weight and pI value were different (Sano *et al.*, 2004).

MRJP6-8

Recenty, three new members of the MRJP family (MRJP6, MRJP7 and MRJP8 were identified in *A. mellifera* (Albert and Klaudiny, 2004). Novel cDNA of MRJP family were identified by using the honeybee brain expressed sequence tags (EST) sequence database, honeybee genomic sequence data and nucleotide sequence from amplification product of the nurse honeybee head cDNA library. Excluding MRJPs, cDNA coding for orthologues of *Drosophila* yellow protein was reported. From its homology with the yellow-f gene product of *Drosophila*, the cDNA was designated am-yellow-f (Albert and Klaudiny, 2004).

MRJP6 cDNA sequence contained an ORF of 1,314 nucleotides encoding for 437 amino acid residues. Sequence analysis revealed highly homologous to MRJP5, but does not have repeat sequence encoding the tripeptide motif (DRM). The 5' non-coding region of MRJP6 cDNA sequence contains a 3' part of intron 0 with the conserved AG motif, intron 0 is found in MRJP 1 genomic sequence (Malecova *et al.*, 2003).

MRJP7 cDNA sequence was assembled from only the honeybee brain EST sequence database. The MRJP7 cDNA contained an ORF of 1,332 nucleotides encoding for 443 amino acid residues. The deduced amino acid sequence of MRJP7 shows high homology to MRJP2 protein.

MRJP8 cDNA sequence was found only one clone in the honeybee brain expressed sequence tags (EST) sequence database. The complete MRJP8 cDNA sequence was identified by using genomic sequence database and nucleotide sequence from amplification product of the nurse honeybee head cDNA library. The MRJP8 cDNA contained an ORF of 1,251 nucleotides encoding for 416 amino acid residues.

Table 1.2 Amino acid composition of *Apis mellifera* MRJPs

	MRJP1	MRJP2	MRJP3	MRJP4	MRJP5	MRJP6*
Ala	3.9	6.2	4.9	4.3	3.8	5.8
Arg	**3.4**	**3.8**	**4.9**	**4.1**	**9.0**	**3.1**
Asn	6.9	11.3	15.9	13.8	8.7	11.0
Asp	8.6	7.1	7.5	7.5	12.0	6.5
Cys	2.5	1.5	1.1	1.3	1.0	1.2
Gln	3.9	5.1	7.1	6.3	3.8	5.3
Glu	3.9	3.8	3.8	3.9	2.5	4.1
Gly	5.6	6.0	6.4	4.1	4.0	5.0
His	**2.3**	**2.4**	**2.2**	**3.9**	**1.8**	**2.6**
Ile	**6.0**	**5.1**	**4.0**	**3.2**	**4.8**	**7.4**
Leu	**9.5**	**8.2**	**6.8**	**9.7**	**5.2**	**7.9**
Lys	**5.1**	**6.9**	**5.8**	**5.0**	**4.3**	**6.0**
Met	**3.5**	**2.4**	**2.2**	**2.4**	**11.4**	**3.6**
Phe	**4.2**	**4.4**	**1.7**	**2.2**	**2.6**	**3.8**
Pro	3.7	3.1	2.5	2.2	2.6	2.9
Ser	8.1	5.8	5.9	8.4	6.2	8.2
Thr	**6.3**	**4.6**	**4.0**	**4.7**	**5.6**	**3.4**
Trp	**1.2**	**1.3**	**0.9**	**1.3**	**1.1**	**1.4**
Tyr	4.4	3.5	3.1	3.9	3.3	5.0
Val	**6.5**	**7.5**	**6.8**	**8.0**	**5.6**	**5.8**
Ess. aa.	48 %	47 %	39.3 %	44.5 %	51.4 %	45 %

Percent content of amino acid in native protein was obtained by computer analysis of its sequence (Schmitzova *et al.*, 1998). Essential amino acids are marked in boldface.

- Amino acid composition of AmMRJP6 was obtained by computer analysis employing the program ProtParam (Albert and Klaudiny, 2004).

Table 1.3 Molecular characterization of cDNAs and deduced amino acid sequences of AmMRJP.

Family	DNA insert size* (bp)	Deduced amino acid (residues)	No. of N-glycosylation site**	Amino acid residues without signal peptide**	Molecular weight (kDa)**	Reference
MRJP1	1444	432	3	416	46.8	Schmitzova *et al.* (1998)
MRJP2	1579	452	2	435	48.9	Schmitzova *et al.* (1998)
MRJP3	1719	467	1	528	59.5	Klaudiny *et al.* (1994)
MRJP4	1625	464	8	449	50.9	Klaudiny *et al.* (1994)
MRJP5	1966	598	4	581	68.0	Albert *et al.* (1999a)
MRJP6	1529	437	5	417	47.6	Albert *et al.* (2004)
MRJP7	1427	443	4	426	48.7	Albert *et al.* (2004)
MRJP8	1329	416	8	400	45.1	Albert *et al.* (2004)

* including polyA tail/ **Partial of data obtained from Schmitzova *et al.* (1998)

1.4 Genomic structure of MRJP

Recently the genomic structure of the gene coding for AmMRJP1 and the putative promoter regions of five known AmMRJP genes were analyzed (Malecova *et al.*, 2003). The AmMRJP1 gene sequence spans over 3,038 bp and contains six exons separated by five introns. Putative promoters were predicted upstream of all MRJP genes. The predicted promoters contain the TATA motif (TATATATT), highly conserved both in sequence and position across AmMRJP gene families. Ultraspiracle (USP) transcription factor (TF) binding sites in putative promoter regions and clusters of dead ringer TF binding sites upstream of these promoters were predicted computationally. The juvenile hormone (JH) was proposed to be a physiological regulator for the binding of USP and acted as a regulator of MRJPs expression. Moreover, MRJPs gene family (MRJP1-5) were a single-copy gene per haploid honeybee genome when examined by Southern blot analysis (Malecova *et al.*, 2003).

1.5 Characterization of MRJPs in *Apis cerana*

Although molecular characterization of AmMRJPs is well studies, there is virtually limited information available on molecular studies of major protein genes of Asian honeybee (*A. cerana indica*). Previously, Takenaka and Takenaka (1996) reported the chemical composition of *A. cerana* royal jelly compared with *A. mellifera* royal jelly. They showed that the levels of proteins, 10-HDA and glucose/fructose ratio in RJ of this 2 species were different. Analysis of water soluble proteins in RJ by SDS-PAGE revealed 21 protein bands in each species where 14 protein bands were shared between the royal jellys of these bees. Four (bands 6, 7, 12 and 16) of six major bands (bands 4, 6, 7, 12, 16 and 21) in the *A. cerana* RJ were more heavily stained than those of *A. mellifera*. In addition, two protein bands (no.10 and 11 with the range of 42.7-66.2 kDa in size) were major and specific to *A. mellifera* RJ.

Recently, the proteins in *A. cerana* RJ was purified and characterized using Q-sepharose and Sephadex G-200 column chromatography (Srisuparbh *et al.*, 2003). The *N*-terminal and internal peptide sequencing were used to clarify the proteins that were purified from *A.cerana* RJ. The molecular weights of denatured proteins were determined by SDS-PAGE. Three types of MRJPs (MRJP1, MRJP2 and MRJP3) that homologue to AmMRJP were found in *A. cerana* RJ. AcMRJP1 was reported to have

two different forms, a monomer (50 kDa) and oligomer (300 kDa) with isoelectric points of 5.2-5.7 and 5.7, respectively. The molecular weight of AcMRJP2 was 55 kDa which had isoelectric points of 7.0-8.0. The native form of AcMRJP3 had the molecular weight of 115 kDa, whereas denatured form was 80 kDa. The isoelectric point of AcMRJP3 was 8.3. In addition, 66 clones from EST library of hypopharyngeal glands were sequenced in unidirection. Partial cDNA sequences of AcMRJP2, AcMRJP3 and AcMRJP4 were obtained. However, AcMRJP5 cDNA was not found in this library.

Cenphakdee (2003) isolated partial nucleotide sequence of AcMRJP3 and complete nucleotide sequence AcMRJP4, AcMRJP5 and AcMRJP6 cDNAs from hypopharyngeal gland of *Apis cerana* nurse bee by Reverse transcription-PCR (RT-PCR) using specific primers designed from *A. mellifera* cDNAs.

Recently, Su *et al.* (2004) constructed a cDNA library from eight-day-old worker heads of *Apis cerana cerana*. A probe derived from part of *Apis cerana* genomic AcMRJP3 segment was used to screen this library. A total of 120 positive clones of MRJPs were screened out from the cDNA library. The positive clones were amplified and sequenced with T3/T7 primer.

1.6 Advantages of MRJPs in other biological system

Several advantages of MRJPs in other biological system have been studies. The MRJP1 enhanced the cell proliferation of rat hepatocytes (Kamakura *et al.*, 2001b), stimulated the growth of human lymphocytes in a serum-free medium (Watanabe *et al.*, 1996) and showed an antifatigue effect in mice (Kamakura *et al.*, 2001a). The MRJP3 had immunosuppressive functions in a mammalian immune system. Recently, Okamoto *et al.* (2003) reported that AmMRJP3 has potent antiallergic activity by inhibition of interleukin–4 (IL–4), IL–2 and IFN–γ production. Intraperitoneal administration of AmMRJP3 inhibited anti–OVA IgE and IgG1 levels in the serum of immunized mice indicating clinical significance on potent immunoregulatory effects of AmMRJP3.

1.7 Recombinant protein expression

Recombinant DNA techniques are used for expression of genes from a variety of sources in different host cells. A number of factors should be taken into consideration when producing a protein in a heterologous host. All different available expression systems have their own characteristics which may have a large impact on the expressed protein or the use thereof.

1.7.1 Prokaryotic expression systems

Prokaryotic expression systems are most widely used. This is due to the large knowledge about these systems, the availability of commercial vectors and strains, and the use of inexpensive growth media. The most widely and routinely used host for heterologous protein production is the bacterium *E. coli*.

1.7.1.1 *Escherichia coli*

Bacterial expression systems are commonly used for production of heterologous gene products of both eukayotic and prokayotic origin. The *E. coli* system has several advantages compared to other systems: (i) it is well characterized both genetically and physiologically, (ii) commercially available vectors and host strains are available, (iii) *E. coli* can be grow in inexpensive and simple media, (iv) *E. coli* has a short generation time (20 min) and (v) a number of cell manipulation techniques using simple laboratory equipment are available.

The most important limitation is the inability of *E. coli* to carry out complex post-translational modification typical of eukayotes, such as glycosylation, myristylation, phosphorylation, etc., and its limit ability to carry out extensive disulfide bond formation and assembly of heterologous proteins into multisubunit assemblies. It is sometimes difficult to achieve a sufficiently high level of expression because of problems such as protein degradation, or inefficient translation due to mRNA structural features or eukaryotic codon usage that are not optimal for *E. coli*.

Several different vectors base on the lac operon are used for high-level expression of foreign proteins in *E. coli*, including:

The trp-lac (tac) promoter. Tac is a hybride trp-lac promoter containing the –35 region of the trp promoter fused to the –10 region of the lacUV5 promoter; it is regulated by the lac repressor and is independent of cAMP regulation mediated by the crp gene product (Amannn *et al.*, 1983, de Boer *et al.*, 1983).

The trp-lac (trc) promoter. Trc is another version of the lac repressor-regulated hybrid trp-lac promoter containing the –35 region of the trp promoter fused to the –10 region of the LacUV5 promoter (Amann and Brosius, 1985). The only difference between the trc and tac promoters is the distance separating the –35 and –10 regions of the promoter. In the trc promoter, these two elements are separated by a consensus distance (17 bp), whereas in the tac promoter, they are separated by 16 bp. This differnce has little or no effect on the expression levels of foreign proteins (Amann and Brosius, 1985).

The lac promoter. Any general-purpose vector (pUC, pTZ, pSK, pBluescript, pGEM, etc.) designed for blue/white screening for clones containing inserts of foreign DNA can be used to express a foreign protein, usually as a fusion protein with amino acids encoded by the amino terminus of the lacZ gene and/or the polylinker sequence. Although the lac promoter is not as strong as the tac or trc promoters, the high copy number of most general-purpose vectors allows expression of foreign proteins at respectable levels.

Bacteriophage T7 promoter. The pET series of vectors, originally developed by Studier *et al.*, 1990) and since expanded, allow regulated expression of foreign genes by bacterophage T7 RNA polymerase. These vectors typically carry the colicin E1 (colE1) replicon of pBR322 and confer resistance to ampicillin or kanamycin. For protein production, a recombinant plasmid is transferred to an *E. coli* strain containing a chromosomal copy of the gene for T7 RNA polymerase.

1.7.2 Eukaryotic expression systems

1.7.2.1 Mammalian cells

Protein production in mammalian cell lines or mammals has the advantage over other expression systems that the native conformation of proteins from mammalian origin is retained. The protein can be secreted by the mammary glands

and it can be harvested by milking (Houbedine, 1994, Maga and Murray, 1995). Cell lines of mammals are commonly used and are under less ethical debate. A great disadvantage of the use of mammalian cells is their requirement for complex and expensive medium. The medium is enriched with serum or serum components which is costly and increases the risk for contamination with blood borne pathogens and toxins.

1.7.2.2 Yeast

As a eukaryote, it is more likely to provide a suitable environment for the folding of foreign eukaryotic proteins than *E. coli*, and can be more readily used for secretion of proteins. Yeast can be grown rapidly on simple media, and to high cell density. There are several expression plasmids available for production of heterologous proteins in yeast. Plamids can also be stabily integrated into the chromosome by homologues recombination, when regions on the plasmid are homologue to regions in the yeast chromosome. Advantages with the yeast system are optimized fermentation conditions, less complex culture medium and ability to perform post-translational modifications. However, there are some draw-backs, including comparatively low expression levels and different in post-translational modifications compared to animal cells (Pichuantes *et al.*, 1996).

1.7.2.3 Insect cells

Insect cell lines are mainly used for intracellular production of proteins. To introduce the foreign gene recombinant baculovirus is often used. The baculovirus expression system has provided an efficient and effective way to synthesize foreign proteins in eukaryotic cells. The use of insect cells for heterologous protein production is a more cost-effective alternative to mammalian cell cultures (Altmann *et al.*, 1999; McCarroll and King, 1997). The maintenance of cell lines is relatively easy and the yield of correctly folded and processed target proteins is generally high, whereas all post-translational modifications of proteins from higher mammals are not always correct. The advantages of using insect cells over mammalian cells are that insect cells use less expensive growth medium, low expression background and lower growth temperature (Pfeifer 1998).

1.7.2.4 Plant cells

Plant cells also provide an inexpensive and convenient system for large-scale production of recombinant proteins. Several efficient plant-based expression systems have been developed. Plant-derived biopharmaceutical proteins, such as antibodies, vaccines, human blood products, hormones and growth regulators, are reaching the late stages of commercial development (Fischer and Emans, 2000). Industrial enzymes, such as phytase and glucanase, are other examples of proteins that are produced in plants (Twyman *et al.*, 2003). In vegetative plants the recombinant proteins are secreted into their exudates (e.g. leaves or roots), enabling proteins to be collected continuously (Borisjuk *et al.*, 1999; Komarnytsky *et al.*, 2000). Some plant-based transient expression systems allow rapid screening of recombinant protein production (Fischer *et al.*, 1999). Plants are considered to be much safer than both microbes and animals as they generally lack human pathogens, oncogenic DNA sequences and endotoxins (Twyman *et al.*, 2003). The ability of plants to post-translationally modify recombinant proteins was demonstrated by production of functional antibodies (Twyman *et al.*, 2003). However, there are some differences in post-translational modifications compared to mammalian cells, particularly with respect to glycan-chain structures (Twyman *et al.*, 2003). Disadvantages include allergic reactions to plant protein glycosylations and other plant antigens, contamination with plant secondary metabolites, mycotoxins or pesticides. Issues regarding separation of genetically modified plants to avoid spreading in the environment should not be ignored (Doran 2000, Hood *et al.* 2002).

The strong and constitutive cauliflower mosaic virus 35S (CaMV 35S) promoter is often chosen to drive transgene expression in dicotyledonous species. Since the CaMV 35S promoter has a lower activity in cereals, the maize ubiquitin-1 (*ubi*-1) promoter is preferred (Twyman *et al.*, 2003). The compartment in which the recombinant protein accumulates strongly influences the interrelated processes of protein folding, assembly and posttranslational modification (Schillberg *et al.*, 2003). As a result, sub-cellular targeting can also be used as a general method to increase the yield of recombinant proteins. Targeting recombinant proteins to oil bodies or to the plasma membrane can also facilitate protein purification (Twyman *et al.*, 2003). Maize has been used to commercially produce recombinant avidin and glucuronidase

(Hood *et al.*, 2002; Lamphear *et al.*, 2002), recombinant antibodies, laccase, trypsin and aprotinin (Hood *et al.*, 2002). Potato is the major host for vaccine production and transgenic potato tubers have been administered to humans in at least three different clinical trials (Twyman *et al.*, 2003).

1.8 Biology of Potato (*Solanum tuberosum* L.)

The taxonomic definition of potato is as follows:

Division Magnoliophyta

Class Magnoliopsida

Order Solanales

Family Solanaceae

Genus Solanum

Species *Solanum tuberosum* L.

Potato *Solanum tuberosum* is the most important noncereal food crop and is the fourth major world food crop after wheat, rice and maize in terms of total global food production (CIP and FAO, 1995; Jones, 1994). Potato was used as animal feed and as raw material for manufacture of starch, alcohol, and other food products (CIP, 1984; SalaZar, 1996; Michael, 1996). The crop is eaten by over one billion people world-wide; it is a part of the diet of half a billion consumers in developing countries. It is grown in 79% contries of the world. It is the first major food crop where biotechnology has been successfully applied (Bajaj, 1987).

The conventional views as it that the potato is a good source of energy, but it containes substantial amounts of high quality proteins and essential vitamins, minerals and trace elements. Potato protein containes a higher proportion of the essential amino acid lysine than most cereal protein, but is deficient in the sulphur-containing amino acids, methionine and cysteine. The essential amino acid composition of plant foodstuffs is an important aspect of human diet in developing countries where nutritional reliance is often on plant protein from a single source. In this situation the improvement of plant products to yield protein with a balanced amino acid content is beneficial. Both conventional breeding and the techniques of genetic engineering have been applied to improve the quality of seed storage protein. Genetic manipulation is

an attractive approach for the modification of potato tuber protein quality. Specific changes can be effected to existing favoured cultivars by the addtion of a single gene without the reassortment of genes which occurs in breeding of cultivars exhibiting tetrasomic inheritance. A prime source of carbohydrate, vitamins and minerals, the potato also supplies valuable protein to the human diet.

1.9 Biology of rice (*Oryza sativa* L.)

The taxonomic definition of rice is as follows:

Division Magnoliophyta

Class Liliopsida

Order Poales

Family Poaceae

Genus *Oryza*

Species *Oryza sativa* L.

Rice (*Oryza sativa* L.) is one of the most important crops in the world and is the staple food for about three billion people (Toenniessen, 1996). Production and consumption are concentrated in Asia where more than 90% of all rice is produced and consumed (David, 1991). Rice is the significant commodity for Thai's economy. It is the country's staple food, by-products of rice is also importance for human and animal consumption (Tassongchant, 1987).

There are 3 subspecies of *O. sativa*, *indica* (long grain), *japonica* (round grain) and *javanica* (medium grain). The *indica* rice concentrates in the warm climate belt, from Indochina, Thailand, India, Pakistan, Brazil and Southern U.S.A.. The *japonica* is mostly grown in cold climate countries, Japan, Korea, northern China and California. The *javanica* is only grown in Indonesia (Oka, 1991).

Khao Dawk Mali 105 (KDML 105) is a famous aromatic rice variety of Thailand and worldwide because of its aromatic, soft and tender cooked rice. KDML 105 is popularly grown under rainfed lowland in the North and Northeast of Thailand.

However, KDML 105 itself is a tall variety and can not produce high enough grain yield. KDML 105 is also photoperiod sensitive which restricted its multiple cropping per year. Moreover, KDML 105 is susceptible to many insect pets and diseases, although this variety can resist to several adversed planting conditions such as moderate degree of drought salted soil or acid soil (Tassongchant, 1987).

1.10 Gene transfer methods in plants

1.10.1 Vector-free gene transfer

Vector-free or dircet gene transfer systems introduced foreign DNA as a naked molecule without a biological vector. There are broadly three approaches to direct gene transfer, which can be described as chemical, electrical and physical. Each is designed to overcome the barrier to DNA uptake that is presented by the cell wall and the plasma membrane. The plant cell wall is a densely structured organelle in which the cellulose microfibrils are intercalated with pectins, hemicelluloses and proteins, which together can prevent the diffusion of large molecules, such as nucleic acids. Furthermore, since DNA is a charged and hydrophilic molecule it does not move freely through the lipid bilayer of the plasma membrane. For these reasons it has been considered advantageous to develop direct gene transfer systems that utilize protoplasts: cells that have had the cell walls removed by enzymatic digestion.

1.10.1.1 Polyethylene glycol (PEG) treatment

The PEG method involves in plasma-membrane destabilizing agent. This type of chemical transformation uses mostly protoplasts that are simply incubated with DNA in buffers containing PEG (Weising *et al.*, 1988). The exact mechanism of action of these diverse agents is not known, but is regarded as destabilizing membrane structures (Kahl and Weising, 1993). However, the production of transgenic plants with PEG treatment was often restricted to certain plants in which the regeneration system from protoplasts is well established. Moreover, this method suffers from many limitations. The choice of explant for protoplast isolation is often restricted to embryogenic cell suspension. It is very difficult to initiate and maintain these cultures and the regeneration capacity of these cultures has been shown to decline gradually with increasing age of the cultures (Jähne *et al.*, 1995). Further, plant regeneration

from protoplasts is labor-intensive, inefficient, time consuming, and is strongly genotype dependent.

1.10.1.2 Electroporation

The use of electroporation for gene delivery has been preferred over PEG because it was found to be more efficient (Zhang *et al.*, 1988). Electroporation is the application of high-voltage electric pulses to cells to induce transient membrane pores, allowing entry of macromolecules including DNA (Rathus and Birch, 1991) However, this method also suffers from the same drawbacks as the PEG method because this also relies mainly on the use of protoplasts for the introduction of foreign genes.

1.10.1.3 Biolistics method or microprojectile bombardment

The microprojectile bombardment system as a method of gene delivery in intact cells and tissues has been enthusiastically employed in cereals soon after its development (Christou, 1997). This method is based on high velocity bombardment of plant cells with DNA-coated microprojectiles (tungsten or gold) accelerated by gun powder discharge or pressurized helium or electric current (Sanford, 1990). This method has been widely used in the production of transgenic plants. The reason for the popularity of this method is that transgenic plants can be obtained following bombardment of any regenerable tissue, thereby eliminating the requirement of isolation and regeneration of protoplasts (Cao *et al.*, 1991). This technique is widely used to deliver foreign DNA to regenerable cells without the burdens of *Agrobacterium* related host specificity limitations. Like other methods, the particle acceleration has its drawbacks too. In some cases, the copy number and rearrangement of the introduced DNA is high, thereby rendering transgene prone to gene silencing and causing genomic changes (Bover *et al.*, 1996). The other major drawback is the restricted availability of the equipment because of its high cost (Christou and Ford, 1995).

1.10.2 *Agrobacterium*-mediated transfer of foreign genes into target plants

Genes can be introduced into the plant cell by using the bacteria *Agrobacterium tumefaciens*. *Agrobacterium* introduces a plasmid, T-DNA, in the

plant cell which integrates randomly in the plant genome. By replacing part of the T-DNA for the gene of the desired protein, the gene can be introduced stably in the plant cell.

A. tumefaciens is exploited by many plant biologists in molecular and genetic studies to introduce DNA into plants. Although best known for this practical application, the transfer of DNA from bacterium to plant comprises fundamental biological processes, many of which are largely uncharacterized.

Agrobacterium carries three genetic components required for plant cell transformation. Two of these, the T-DNA and the virulence (*vir*) region, are located on the large (roughly 200 kb) Ti plasmid (Fig. 1.2) (Binns and Thomashow, 1988). The T-DNA is the DNA segment that can be transferred from *Agrobacterium* to the plant cell. The *vir* region is organized into six complementation groups that are either absolutely essential for (*virA*, *virB*, *virD* and *virG*) or that enhance the efficiency of (*virC* and *virE*) plant cell transformation. The third bacterial component of the T-DNA transfer process resides in the *Agrobacterium* chromosome. Three chromosomal virulence loci, *chvA* and *chvB* and *pscA*, encode products involved in the binding of *Agrobacterium* to plant cells during the infection process (Zambryski, 1988).

During infection by *Agrobacterium*, a piece of DNA is transferred from the bacterium to the plant cell (Fig. 1.2). The piece of DNA is a copy of a segment called the T-DNA (Transferred DNA). It is carried on a specific plasmid, the Ti-plasmid (Tumor-inducing). The T-DNA is delimited by 25-bp direct repeats that flank the T-DNA. Any DNA between these borders will be transferred to a plant cell (Walden, 1993). Wild-type T-DNA encode enzymes for the synthesis of the plant growth regulators; auxin and cytokinin, and the production of these compounds in transformed plant cells results in the tumorous phenotype. In addition, wild-type T-DNA also encodes enzymes for the synthesis of novel amino acid derivatives called opines. The Ti-plasmid encodes enzymes for their catabolism; hence, *Agrobacterium* has evolved to genetically commandeer plant cells and use them to produce compounds that they uniquely can utilize as a carbon / nitrogen source (Kahl and Weising, 1993).

The processing and transfer of T-DNA are mediated by products encoded by the *vir* (virulence) region, which is also resident on the Ti-plasmid (Stachel and Nester, 1986). Those *vir* genes, whose products are dirctly involved in T-DNA processing and transfer, are tightly regulated so that expression occurs only in the presence of wounded plant cells, the targets of infection. Control of gene expression is mediated by the VirA and VirG proteins, a two-component regulatory system. VirA detects the small phenolic compounds released by wounded plants resulting in autophosphorylation (Fig 1.2, step 1). VirA phosphoryloation of VirG then leads to activation of *vir* gene transcription (Winans, 1992).

Following *vir* gene induction, the production of a transfer intermediate begins with the generation of the T-strand, an single strand copy of the T-DNA (Stachel *et al*, 1986). VirD1 and VirD2 are essential for this process (Filichkin and Gelvin, 1993). Together, VirD1 / VirD2 recognize the 25-bp border sequence and produce an single strand endonucleolytic cleavage in the bottom strand of each border (Fig 1.2, step 2). These nicks are used as the initiation and termination sites for T-strand production. After nicking, VirD2 remains tightly associated with the $5'$ end of the T-strand. The lone VirD2 at the $5'$ end gives the T-complex a polar character that may ensure that, in subsequent steps, the $5'$ end is the leading end. T-strand production is thought to result from the displacement of the bottom strand of the T-DNA between the nicks (Zupan and Zambryski, 1995).

The T-strand must travel through numerous membranes and cellular spaces before arrival in the plant nucleus. Thus, to preserve its integrity, it was hypothesized that the T-strand likely travels as an single strand DNA-protein complex. VirE2 is an inducible single strand nucleic acid-binding protein encoded by the *vir*E locus that binds without sequence specificity. VirE2 binds tightly and cooperatively, which means that a T-strand would be completely coated (Fig. 1.2, step 3). Consequently, degradation by nucleases would be prevented and, indeed, *in vitro* binding of VirE2 renders single strand DNA resistant to nucleolytic degradation. Finally, binding of VirE2 unfolds and extends single strand DNA to a narrow diameter of 2 nm, which may facilitate transfer through membrane channels. The T-strand along with VirD2 and VirE2 are termed the T-complex (Zupan and Zambryski, 1995).

Subsequently, the T-complex must exit the bacterial cell (Fig. 1, step 4) passing through the inner and outer membranes as well as the bacterial cell wall. It must then cross the plant cell wall and membrane (Fig. 1, step 5). Once inside the plant cell, The T-complex targets to the plant cell nucleus and crosses the nuclear membrane (Fig. 1, step 6), after which the T-strand becomes intergrated into a plant chromosome (Fig. 1, step 7). In the context of experimental chronology and relevant results, step 1 through 3 (Fig. 1) have been studied. Current and recent research has related to step 4 and 6. Entering the plant cell (step 5) and the mechanics of integration (step 7) are almost completely uncharacterized (Zupan and Zambryski, 1995).

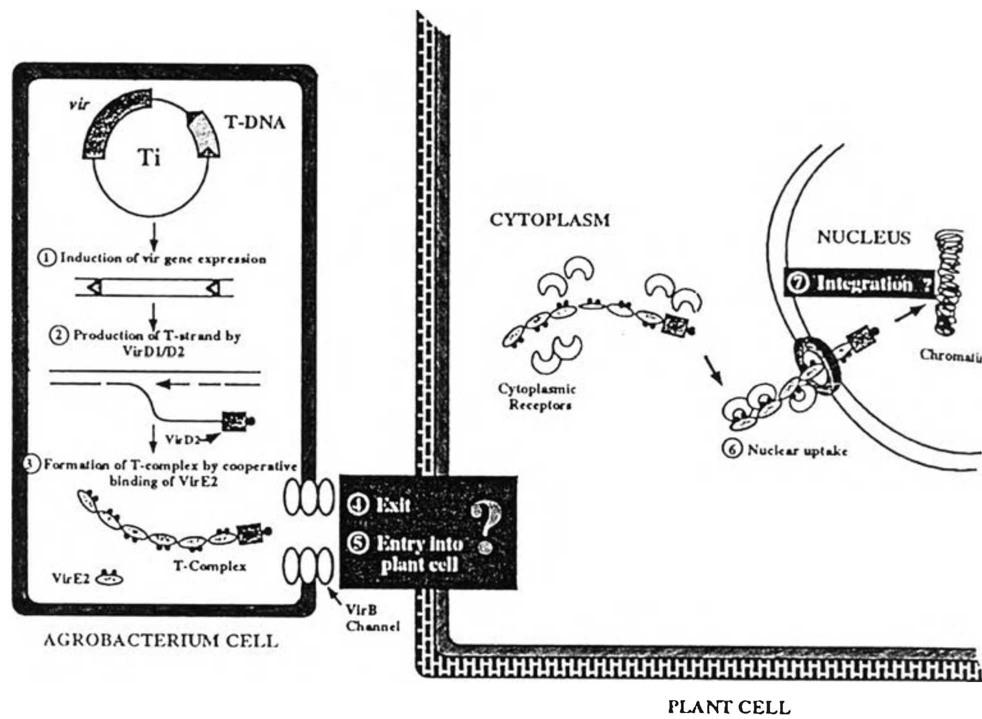

Figure 1.2 Basic steps in the transformation of plant cells by *A. tumefaciens* (Zupan and Zambryski, 1995).

1.11 Promoters

1.11.1 CaMV 35S promoter

The double-strand caulimovirus family has provided the single most important promoter for transgene expression in plants: the caulifower mosaic virus (CaMV) 35S promoter. The CaMV 35S is expressed in most cells of most plants, and in most tissued, although with diffeent efficiency according to plant species.

1.11.2 Tissue-specific promoters

Many plant genes that are higly specific for certain tissues have been characterized, and in most cased their promoters reproduce the gene expression pattern. Tissue-specific promoters are attracive for the production of proteins which must be concentrated in ceratin organs to limit toxicity or to improve harvesting efficiency.

1.11.2.1 Patatin

Patatin is a major soluble glycoprotein of potato tubers which, although it has lipid acyl hydrolase activity, is thought to serve a food storage role. The promoters of the tuber-specific patatin genes, called class I patatin genes (class II are expressed in root tips and vacuolar tissue), allow expression of heterologous proteins in tubers. In addition, they allow induction in sucrose-treated leaves (Grierson *et al.*, 1994, Kim *et al.*, 1994).

1.11.2.2 Granule-bound starch synthase (GBSS)

Starch is the major storage carbohydrate in higher plants and is found most often in special storage organs such as roots, seeds or tubers. Starch consists of a mixture of two glucose polymers: amylopectin (α-1,4-linked D-glucose residues with α-1,6-glucosidic branch points) and amylose (essentially linear chains of α-1,4-linked D-glucose (Banks and Greenwood, 1975). Granule-bound starch synthase (GBSS) is one of the key enzymes in the biosynthesis of starch and catalyses the formation of amylose. The promoter of GBSS gene is organ-specifically expressed in stolons and tuber. The expression was inducible in leaves by growth on relatively high concentraiotns of sucrose, fructose and glucose (Visser *et al.*, 1991).

1.12 Selectable marker genes

Selectable markers are used to select for the specific growth of transformed cells among a background of non-transformed individuals. This is important because transformation frequencies remain normally relatively low. Such markers allow growth, or at least viability, in the presence of the selective agent. Routinely, resistance to antibiotics or phytotoxins such as herbicides has been used. In the former case, antibiotic resistance genes derived from bacteria have been utilised (e.g. kanamycin and hygromycin), whereas in the latter case genes encoding products which are more tolerant to herbicides, for example glyphosate and phosphinothricin have been used (Walden, 1993).

1.12.1 Neomycin phosphotransferase

The most commonly used selectable gene is *npt*II. It was isolated from transposon Tn 5 of an *E. coli* strain. This gene codes for the enzyme neomycin phosphotransferase that detoxifies several aminoglycoside antibiotics such as kanamycin, geneticin (G418) and neomycin. However, kanamycin can be used as a selective agent during regeneration of protoplasts but it is not effective for selection of transformed calli. In addition, many calli recovered after kanamycin selection are unable to regenerate green plants (Ayres *et al.*, 1994; Toriyama *et al.*, 1988).

1.12.2 Hygromycin phosphotransferase

The gene *hpt* (or *hph*) was isolated from *E. coli*. It codes for the enzyme hygromycin phosphotransferase. This gene therefore causes resistance to the antibiotic compound hygromycin. The *hpt* / hygromycin B combination was successfully employed in the genetic transformations of tobacco, *Arabidopsis*, maize and rice (Schrott *et al.*, 1995). Hygromycin is a more potent phytotoxic compound than kanamycin; especially in cereal crops (Galun and Breiman, 1997). Hygromycin allows clear discrimination between transformed and non-transformed tissues and problems with albinos or the ferility of regenerants have not been reported (Ayres *et al.*, 1994).

1.12.3 Phosphinothricin acetyltransferase (*bar* or *pat*)

Two bacterial genes (*bar* from *Streptomyces hygroscopicus* and *pat* from *S. viridochromogenes*) have the ability to detoxify phosphinothricin (PPT) by acetylation. These genes code for phosphinothricin acetyltransferase which confers resistance to phosphinothricin (D'Halluin *et al.*, 1992). Phosphinothricin which is a structural analog of glutamine, inhibits the glutamine synthase (GS) of both plant and bacterial origin. Inactivation of the GS will lead to ammonia accumulation which is toxic to the cell. There are derivatives of this compound that are also used as herbicides (e.g. Bialaphos, Basta).

1.13 Reporter genes in plants

Reporter genes are coding sequences that, upon expression in the transgenic plant, provide a clear indication that genetic transformation did take place (Galun and Breiman, 1997). In general, reporter genes should have the following characteristics: 1) the genetic organization should be well described, 2) the gene products should not be present in the organism or tissue under study, 3) the gene products should be well characterized with regard to biochemical activity, 4) substrate dependence and stability, and 5) the product of the reaction catalyzed by the reporter gene product should be stable, easily detectable, and quantifiable (Crazzolara *et al.*, 1995). The most widely used systems of reporter genes is β-glucuronidase (GUS).

1.13.1 β-glucuronidase (GUS)

GUS is encoded by the *Escherichia coli uidA* gene (Jefferson *et al.*, 1986, 1987) and has became the most widely used reporter gene in plants. The protein has a molecular weight of 68.2 kDa. The best substrate currently available for histochemical localization of β-glucuronidase activity in tissues and cells is 5-bromo-4-chloro-3- indolyl glucuronide (X-Gluc). The product of glucuronidase activity on X-Gluc is colorless. Instead, the indoxyl derivative produced must undergo an oxidative dimerization to form the insoluble and highly colored indigo dye. This dimerization is stimulated by atmospheric oxygen, and can be enhanced by using an oxidation catalyst such as a potassium ferricyanide/ferrocyanide mixture. It is the efficient reporter that can be used to locate its expression in plant tissues and organs

without the need to extract the respective tissue. However, the substrate for dectection of GUS activity (X-Gluc) is expensive and the X-Gluc-stained plant material is killed by the respective GUS assay; it is therefore not a vital staining (Jefferson *et al.*, 1986, 1987).

1.14 Applications of biotechnology in plant breeding

Conventional breeding has been used regularly for crop improvement and production of a new variety. Traditional plant breeding involves making large numbers of crosses between a range of parents which have been selected by the breeder for their desirable attributes. Some of the progeny from these crosses will show a combination of the best traits from the parents. After continued genetic recombination and a number of cycles of selection, a new variety is obtained which has a combination of desirable characteristics and which is distinct from any other variety. This process is simple in outline but complex in practice. Plant breeding involves a wide range of skills. Given this complexity it can take up to 15 years between the initial cross and the commercial release of a new variety. Genetic engineering offers an advantage over conventional breeding in a way that only one or two characters will be introduced into crop species. The overall genotypic characters of that species remain unchange. As a further complication there are various limitations to plant improvement through conventional breeding. Not only are plant-breeding programmes long-term and therefore expensive, they also require the cultivation and analysis of large numbers of plants which takes both time and space. Additionally, in some cases, the breeder has limited genetic variation for incorporation into new varieties. Moreover, varieties derived from breeding program may be somewhat different from their parental lines. Horticultural characters like size and shape of fruit, color of flower, and processing quality of vegetable could be affected. Thus techiques which overcome any of these constraints are of interest and potential value to plant breeders (Connett and Barfoot, 1992).

Genetic engineering technologies have made possible the production of foreign proteins in plants, potentially for the manufacture of the proteins themselves, such as antibodies, vaccines or nutritional proteins with altered amino acid content; or for the production of proteins to change the characteristics of the plant.

One of the goals of plant genetic engineering has been to create crops that are tailored to provide better nutrition for humans and their domestic animals. A major target has been the improvement of the amino acid composition of seed proteins, because animals, including humans, are incapable of synthesizing 10 of the 20 amino acids needed for protein synthesis, and these "essential" amino acids must therefore be obtained from the diet. The essential amino acids that limit the nutritive value of potato protein are lysine, tyrosine and the sulfur-containing amino acids methionine and cysteine (Jaynes *et al.*, 1986) whereas methionine, lysine, threonine and tryptophan are deficient in rice (Kavakli *et al.*, 2000). Advances in plant tissue cultue techniques and gene transfer technology have opened up possibilities for modifying the amino acid contents of plants. One approach has been to manipulate the regulation of amino acid biosynthesis to increase the abundance of a particular amino acid. Mutant selection and engineering genes encoding key enzymes of amino acid biosynthetic pathways have been used to increase amino acids in crop plants (Matthews and Hughes, 1993). However, an increse in the free essential amino acids does not lead to an increase in the fixed content, and the amino acids could be leached out from the plant tissue and lost during boiling and other processing (Shaul and Galili, 1992; Falco *et al.*, 1995). An alternative approach is the insertion and the expression of genes encoding essential amino acid-rich proteins in transgenic plants.

There are many examples of plants that produce transgenic proteins serving a wide spectrum of purposes including their nutrient values enrichment (Schroeder *et al.*, 1991), Tabe and Higgings, 1998). In recent years, many cereals and legumes have been transformed, e.g. by increasing of free lysine content in rice seeds (Christou *et al.*, 1991, Lee *et al.* 2001). Transformed wheat (Becher *et al.*, 1994) and birdsfoot trefoil can accumulate the maize seed storage proteins rich in amino acids containing sulphur, gamma zein and beta zein (Bellucci *et al.*, 2002). In transgenic potato high-methionine or high-lysine proteins have been expressed (Tu *et al.*, 1998).

The proteins of honeybee royal jelly (RJ) represent a new potential for enrichment of crop plants in nutritive and physiologically active proteins. MRJP1 has relatively high content of essential amino acid. Therefore, it could be a potential ingredient of functional foods.

The objective of this study is to clone and characterize the cDNA and genomic DNA of AcMRJP1 and AcMRJP2 of *Apis cerana* and expression in *E. coli* and plant expression system. The knowledge obtained from this study will provide a fundamental basis for Major royal jelly protein 1 and 2 encoding genes of *A. cerana* and basic knowledge on heterologous expression of these genes.