Chapter 4

Discussion

1. Characterization of purified vitellin

The N-terminal amino acid sequence of 78 and 87 kDa subunits of purified Vt are highly similar to subunits that have average size 80-90 kDa of Vt/Vg in other shrimps, whereas those of 104 kDa subunit is highly similar to approximately 100 or 200 kDa subunit of Vt/Vg from other shrimps as shown in the Table 12. Although only 9 or 10 amino acid residues from the N-terminus of purified Vt 78, 87 and 104 kDa subunits were determined, the conservation of this portion appears to be very high among several shrimps. The N-terminal amino acid sequence of the 78 and 87 kDa subunits were exactly the same as the deduced amino acid sequence of Vg from *P. merguiensis* at the N-terminal region, just after the signal peptide; which correspond to the 78 kDa subunit predicted from the Vg deduced amino acid sequence. This implies that the 78 kDa subunit of Vt is post-translationally modified to produce the 87 kDa subunit. Results from Auttarat *et al*. (2006) indicated that Vt is a lipoglycoprotein. The N-terminal amino acid sequence of the 104 kDa subunit confirms that the RTRR cleavage site is used in Vg processing in banana shrimps. Since it corresponds to the predicted 203 kDa subunit, this subunit must experience further protease processing by the subtilisin endoprotease family or other endoprotease enzymes. Further, by comparing the deduced amino acid sequences from decapod crustaceans, N-terminal amino acid sequence and cleavage site are confirmed to be conserved. Alignment of the deduced amino acid sequences of *P. merguiensis* Vg with other species at N-terminal region including signal peptide and consensus cleavage sequences (R-X-[K/R]-R or R-X-X-R) for processing by subtilisin-like endoproteases are shown in Fig. 35A and 35B, respectively.

Table 12 The N-terminal amino acid sequence of the purified Vt/Vg subunits in shrimps and crab.

Species	Subunits (kDa)	N-terminal amino acid sequence	Reference
P. merguiensis	78, 87 (Vt)	APWGADVPR	This study
	104 (Vt)	SIDSSVIADF	This study
L. vannamei	78 (Vt)	APWGADVPRXSTEXPVT	Garcia-Orozco *et al*., 2002
M. japonicus	91 (Vt)	APWGADLPRXSTEXPISGSP	Kawazoe *et al*., 2000; Tsutsui *et al*., 2000
P. semisulcatus	72 (Vt)	APWGA	Avarre *et al*., 2003
	79 (Vt)	VQYTVAEK	Avarre *et al*., 2003
	79 (Vt)	MNHPVLPK	Avarre *et al*., 2003
	100 (Vt)	SIDSSV	Avarre *et al*., 2003
	207 (Vt)	SIDSSV	Avarre *et al*., 2003
	74 (Vg)	APWGADLP	Avarre *et al*., 2003
	199 (Vg)	SIDSSV	Avarre *et al*., 2003
M. rosenbergii	90 (Vt)	APWPSGTNLX	Okuno *et al*., 2002
	90 (Vt)	RREEQKVTGT	Okuno *et al*., 2002
	102 (Vt)	SIDLRQISHL	Okuno *et al*., 2002
	90 (Vg)	APWPSGTNLXSKEXP	Okuno *et al*., 2002
	90 (Vg)	RREEQKVTGTVELDI	Okuno *et al*., 2002
	199 (Vg)	SIDLSQISHIFDKLY	Okuno *et al*., 2002
P. hypsinotus	90 (Vt)	APWPSNLPXXSTEEPIAG	Tsutsui *et al*., 2002
	100 (Vt)	SIDFSXLXHLFDKLY	Tsutsui *et al*., 2002
	85 (Vg)	APWPSNLP	Tsutsui *et al*., 2004
	190 (Vg)	SIDFSSLS	Tsutsui *et al*., 2004
C. feriatus	29 (Vt)	APYGSTIQL	Mak *et al*., 2005
	76 (Vt)	APYGSTIQLCSTEEP	Mak *et al*., 2005
	105 (Vt)	VEYTRSGSHQQGFELDVFP	Mak *et al*., 2005

The amino acid composition of purified *P. merguiensis* Vt was determined and shows similarities to those of purified Vt from other shrimps; *Penaeus chinensis* (Chang *et al*., 1996), *P. monodon* (Chang *et al.*, 1993; Quinitio *et al*., 1990), *M. japonicus* (Vazquez-Boucard *et al*., 1986), *P. semisulcatus* and *L. vannamei* (Tom *et al*., 1992), *M. ensis* (Qui *et al*., 1997) and *M. rosenbergii* (Lee *et al*., 1997b) as shown in Table 8. Purified Vt protein from *P. merguiensis* was acidic with a pI value of 5.3. Purified *P. monodon* Vt was also reported to have more acidic than basic amino acids (Chang *et al*., 1993). The amino acid composition of *P. merguiensis* Vt is largely the same as deduced from the Vg cDNA except for cysteine content. This was higher in the deduced amino acid of Vg than in the purified Vt, where cysteine was not detected. Low or undetectable levels of cysteine were also found in other shrimps. The results from this study and Auttarat *et al*. (2006) show that the number of polypeptide subunits of purified Vt is the same when observed by SDS-PAGE with or without β–mercaptoethanol. This indicates that Vt subunits are not held together by disulfide bonds or have very weak disulfide bond interactions between subunits.

2. Characterization of cDNA encoding Vg from the ovary

Vt from the ovary and Vg from the hepatopancreas of *P. semisulcatus* are the gene products of one gene (Khayat *et al*., 1994b). In this study, cDNA encoding Vg was cloned based on the N-terminal amino acid sequence of 78 and 87 kDa subunits and conserved regions of Vg/Vt from other decapod crustaceans. Firstly, Vg cDNA sequence was isolated by one step RT-PCR and RACE approach, then the Vg cDNA sequence re-construction was confirmed by amplifying and sequencing a large full-length Vg PCR product. The ORF of Vg in *P. merguiensis* is 7.7 kb, approximately similar to the transcript size of the Vg ORF in other shrimps such as *L. vannamei* (7.7 kb) (Raviv *et al*., 2006), *P. semisulcatus* (7.7 kb) (Avarre *et al*., 2003), *M. rosenbergii* (7.6 kb) (Okuno *et al*., 2002), *P. hypsinotus* (7.6 kb) (Tsutsui *et al*., 2004) and *P. monodon* (7.7 kb) (Tiu *et al*., 2006). The Vg cDNA of *P. merguiensis* consists of 7,961 nucleotides. Sequence analysis revealed the presence of a single large ORF of 7,758 bp that is translated to 2,586 amino acids including a signal peptide. The deduced Vg sequence was predicted to be cleaved between amino acid residues 18 and 19 and exported from the cell. From Fig. 35A and 35B, the multiple alignment of deduced Vg sequences from *P. merguiensis*, *P.*

monodon, *M. japonicus*, *L. vannamei*, *P. semisulcatus*, *M. ensis*, *P. hypsinotus*, *M. rosenbergii*, *C. quadricarinatus* and *P. trituberculatus* shows that all have a signal peptide, and also have a conserved cleavage site (R-X-[K/R]-R or R-X-X-R) that is recognized by the subtilisin endoprotease family (Barr, 1991; Chen *et al*., 1997). This cleavage site is functional, as determined by N-terminal amino acid sequence in *P. merguiensis* (banana shrimp) (this study), *P. semisulcatus* (green tiger shrimp) (Avarre *et al*., 2003), *M. rosenbergii* (giant freshwater prawn) (Okuno *et al*., 2002) and *P. hypsinotus* (coonstriped shrimp) (Tsutsui *et al*., 2004) as shown in Table 12. Based on the Vg primary structure, cleavage at this site will produce two subunits of molecular masses of 78 and 203 kDa. Since the N-terminal amino acid sequence and SDS-PAGE analysis of purified Vt isolated from the ovary shows identical sequence, but a minor 104 kDa subunit instead of the expected 203 kDa subunit, there must be additional processing. For example, subtilisin endoprotease cutting between amino acid residues 1,731 and 1,734 as shown in Fig. 52 will produce 110 and 93 kDa subunits instead of 203 kDa subunit, another endoprotease enzyme may cut at a nearby site. In *M. rosenbergii*, the second cleavage occurs five amino acid residues downstream of an RDRR site (Arg1742 to Arg1745), yielding a protein with N-terminal sequence RREEQKVTGTVELDI (Okuno *et al*., 2002). In *C. feriatus*, the Vg precursor is cleaved five amino acid residues downstream of the KLSR site (Lys1781 and Arg1784) for the second cleavage site (Mak *et al*., 2005). In *P. semisulcatus*, the N-terminal amino acid sequences of 79 kDa subunits are VQYTVAEK and MNHPVLPK, may indicate that the enzyme responsible for this second cleavage recognizes a broader motif. Thus, it is difficult to identify the second cleavage site in *P. merguiensis* Vg. The alignment of the possible secondary cleavage site region by subtilisin endoprotease and N-terminal sequence at second cleaved site from reported species is shown in Fig. 53.

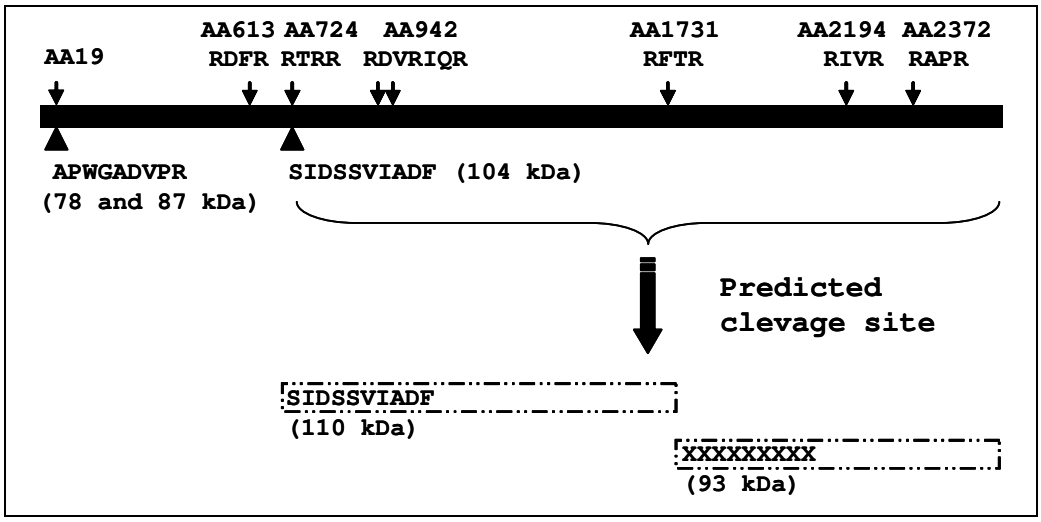

Fig. 52 Schematic view of *Peneaus merguiensis* Vg potential cleavage sites.

N-terminal amino acid sequence positions of Vt subunits and predicted cleavage site at residue 1,735 and predicted subunit production based on use of the cleavage site at 1,735 are shown. The downward vertical arrows indicate the location of predicted cleavage sites that have the RXXR consensus motif for subtilisin endoprotease. The upward vertical arrows represent the N-terminal amino acid sequence of Vt subunits. The X shows unknown amino acid sequence. The dash line boxes show the predicted sizes of peptides after cleavage at position 1,735 by endoprotease.

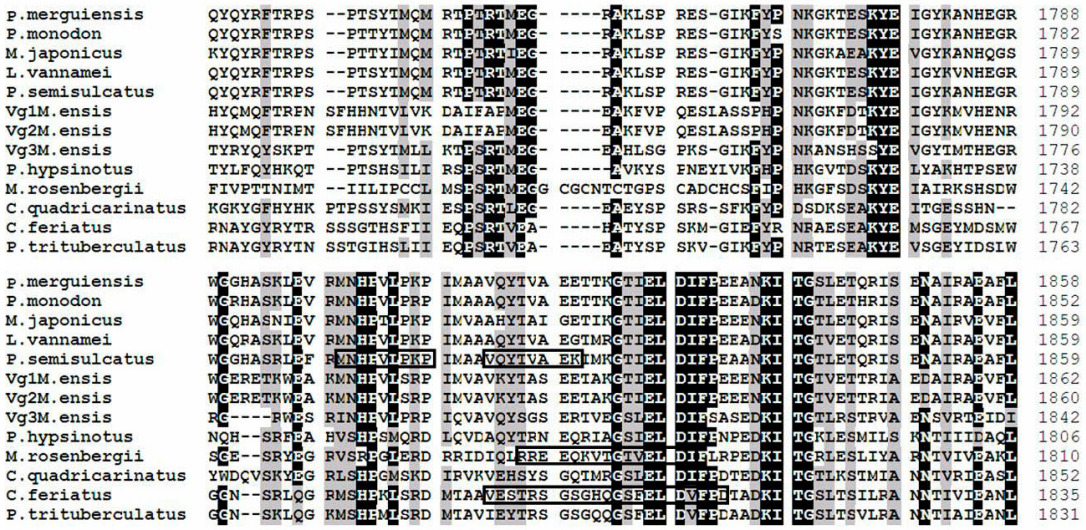

Fig. 53 Alignment of the partially amino acid sequence of decapod crustacean Vgs at region that possible to be second cleavage site.

N-terminal amino acid sequence products form the second cleaved sites are surrounded by a black bordered boxes. Numbers in right hand side indicate the amino acid positions counted from signal peptide of each Vg. The (-) indicates a gap introduced into the amino acid sequence to allow for the maximum degree of identity in the alignment. Identical amino acid residues are in white letters and highlighted in black. The similar amino acid residues are indicated by grey background.

3. Comparison of the deduced amino acid sequence of Vg from *P. merguiensis* by alignment BLAST and phylogenetic trees

The deduced amino acid sequence of *P. merguiensis* Vg was compared using identity ratio and relationship throughout its whole length, including the signal peptide, to other decapod crustacean species by alignment. Results from identity ratio and phylogenetic tree correspond well. The deduced amino acid sequence of Vg from *P. merguiensis* has the most identity (91.4%) and closest relationship to the green tiger shrimp *P. semisulcatus* Vg (Avarre *et al*., 2003). It is also very similar (86.9-50.3%) to the Vg from other members of the penaeus group: the white shrimp *L. vannamei* (Raviv *et al*., 2006), the black tiger shrimp *P. monodon* (Tiu *et al*., 2006), the kuruma shrimp *M. japonicus* (Tsutsui *et al*., 2000) and the sand shrimp *M. ensis* (Tsang *et al*., 2003; Kung *et al*., 2004). The similarity is lower (42.6-36.5%) when compared to the Vg of the crayfish *C. quadricarinatus* (Adbu *et al*., 2002), the coonstriped shrimp *P. hypsinotus* (Tsutsui *et al*., 2004) and the giant freshwater prawn *M. rosenbergii* (Okuno *et al*., 2002). The lowest identity (less than 35.0%) was found in crabs; the red crab, *C. feriatus* (Mak *et al*., 2005) and the Japanese blue crab, *P. triuberculatus* (GenBank accession number: DQ000638), not surprising because these are already believed to be more distantly related (Porter *et al*., 2005).

At present, the complete primary structure of Vg is available for eleven crustacean species, limited to decapods: six Penaeidea, two Caridea, one Astacidea and two Portunoidea species. The phylogenetic tree constructed in this study is based on the amino acid sequences of the Vgs from these eleven decapod species. The unrooted tree shows four distinct lineage groups and this is in agreement with their taxonomic classification (Porter *et al*., 2005). It also shows that the *P. merguiensis* Vg is most closely related in decreasing order to the Vg from *P. semisulcatus*, *L. vannamei*, *P. monodon*, *M. japonicu*s and *M. ensis*, all in Penaeidea. The unrooted tree is presented because there are two possible root positions. Either, our preliminary results with *Daphnia magna*, accession number AB114859 (Kato *et al*., 2004) as an outgroup, the root is located between Penaeids (Dendrobranchiata) and Caridea/Astacidea/Portunoidea (Pleocymata) as shown in Fig. 54, this result matches the previously-published taxonomy of these species (Porter *et al*., 2005). Alternatively, midpoint rooting and the published Vg phylogenetic tree analysis in another shrimps suggest that the root is on the branch to *C. feriatus* (Portunoidea) (Tiu *et al*., 2006). It is noted that one of the Vg gene in *M. ensis* (Vg3) has only 54% sequence

identity to the other two (Vg1 and Vg2) (Kung *et al*., 2004). The degree of divergence suggests an early gene duplication event followed by gene differentiation between Vg3 and Vg1/Vg2. Given the placement of Vg3 on the tree, predating Penaeid speciation, there is even the possibility that corresponding, extra Vg genes exist in other shrimp, but currently no specific data is available. Since the corresponding gene has not been identified, despite work on Vg in a number of species, I predict that if it is functional it produces very little protein.

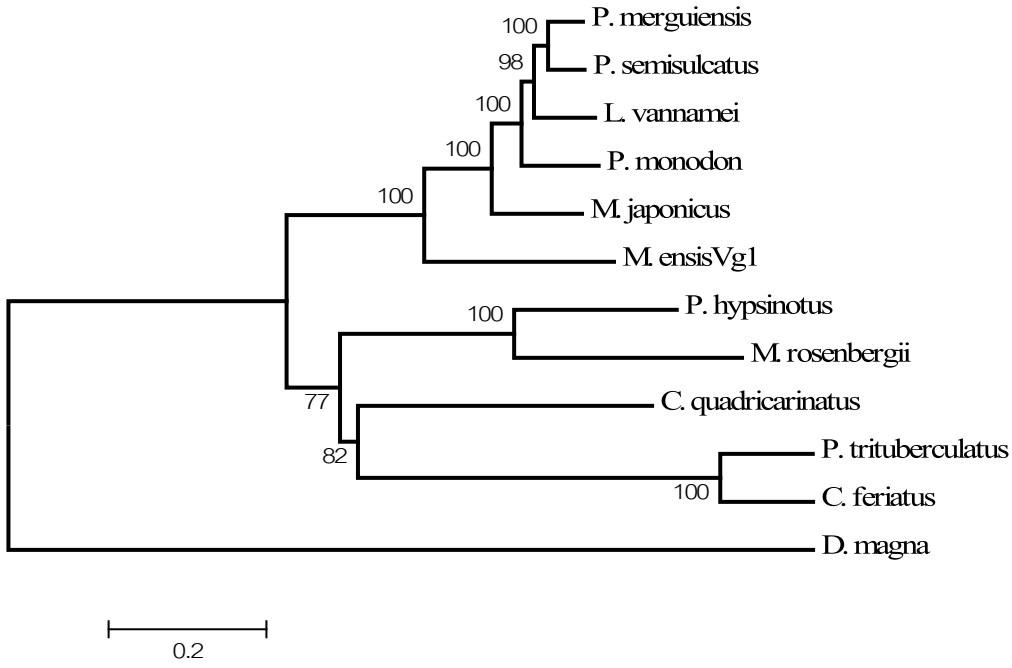

Fig. 54 Phylogenetic tree analysis of crustacean Vgs.

The rooted tree was constructed with MEGA (Kumar, *et al*., 2004) using the neighbor-joining method (Saito and Nei, 1987) based on the alignment of amino acid sequences of ten crustacean Vgs using Clustal W (Thomson, *et al*., 1994). The distance method used was poisson correction; otherwise default parameters were used. Percentage bootstrap values for 1000 replicate analyses are shown at branch points which indicated how reliable of this tree. The standard bar at the bottom shows the branch length corresponding to 0.2 changes per residue.

The results from BLAST, phylogenetic tree analysis and tertiary structure prediction of the N-terminal region of deduced Vg sequence from banana shrimp confirms that the lipoprotein N-terminal domain is conserved in a number of proteins that are involved in lipid transport. The results from this study agree with Babin *et al*. (1999) that Vg, insect Apo, ApoB, Retin and MTP are all members of the same multigene superfamily: large lipid transfer protein (LTP). Even if they have low overall conservation across taxonomic groups, the biological functioning of lipoproteins depends on this region, and these proteins most likely have a common ancestor (Chen *et al*., 1997).

Interestingly, the Clot from decapod crustacean (shrimp and crayfish) also has the lipoprotein N-terminal domain and the results from phylogenetic analysis showed that the Clot has a closer relationship with Vg from daphnia (crustacean, arthropod) and insects (hexapoda, arthropod) than Vg from shrimps or other decapod crustaceans. Surprisingly it appears that crustaceans Clot are evolutionarily closely related to Vgs but have a different function. The results from this study corresponded with the sequence analysis from Hall *et al*. (1999a) which revealed that the Clot in crayfish is homologous to Vgs, and the Clot and Vgs are all lipoproteins which share sequence similarity at N-terminal (lipoprotein terminal domain) and C-terminal (vWF type D domain) regions to other lipoproteins (Apo, Retin and MTP) as shown in Fig. 37 and Fig. 38. Additionally, the phylogenetic trees of N-terminal region also indicated that all Vgs from decapod crustaceans are closer to Apo, ApoB and Retin than to Vg from other species and does not have phosvitin or polyserine regions. This is an extremely unexpected result, as it is assumed Vg evolved once, and indeed Vg from all other species (vertebrates, nematodes, coral, oyster etc) are descended from a single ancestral Vg. This makes Vg in decapod crustaceans paralogous to all other Vgs. Paralogues often retain common biochemical activities and can acquire new functions in the course of evolution through gene rearrangements (Doolittle, 1995; Henikoff *et al*., 1992). Furthermore, investigators have established the relationship between crustacean Vg and mammalian ApoB at the level of immunological cross activity (Warrier and Subramoniam, 2003). Currently no clear evidence exists for why Clot lost its ancestral Vg function, and why another LTP family protein took over Vg function in decapod crustaceans. There may be specific features of vitellogenesis unique to decapod crustaceans, either as the cause or as a result of the functional changes of these LTP genes. Note that all decapod crustaceans with Vg sequence are paralogous

to all other Vgs, including the crustacean *D. magna*. However, Vg sequences are not available for any other crustacean, so to determine when the gene functional change occurred, it would be necessary to isolate and sequence the Vg gene from a number of other crustacean species.

The internal region of *P. merguiensis* Vg showed little or no similarity to non-crustacean proteins in common with other crustaceans whereas its C-terminal domain showed some homology to vWF in mammals. vWF originally described in domains are a characteristic feature of all reported Vgs. Tsutsui *et al*. (2004) suggested that the central segment has little or no similarity to non-crustacean proteins because this region forms a structure specific to the kind and amount of lipids required in crustacean species. Generally, if the central segment has no enzyme function or protein interaction, but is required only to bind lipid, different selection pressure would be expected, perhaps leading to loss of sequence similarity.

4. Tertiary structure modelling of N-terminal region Vg

Structural and functional relationships exist among Vg, ApoB and MTP and the structural differences between these proteins relate to their different lipid binding and lipid transfer properties (Mann *et al*., 1999).

All model predictions from several servers gave the same main domains including the N-, C-, A-sheets and helical domain. N-, C-sheet and helical domain have the same structure when compared to 1LSH_A template. Not surprisingly, loops and some small secondary structures are different in the different models, since the N-terminal region of Vg sequence from banana shrimp shared only about 19.4% identities with lamprey LV template. It is expected with such low sequence conservation that models from different servers differ in details.

Additionally, the lamprey LV crystal structure (Thompson and Banaszak, 2002) used as a template has several breaks. In total there are five fragments, four fragments from the first polypeptide (LV1) or chain A, and another from the second polypeptide (LV2) or chain B. This means our sequence, and our models have extra loops not found in the template. LV is a lipoprotein molecule which is difficult to crystallize and analyze by X-ray crystallography, because the lipid does not have a fixed structure. Additionally, variation in lipid quantity from the purification process or due to different stages of ovarian development contributes to the problem, so the structure data in PDB are still not complete.

The A-sheet could not be predicted completely by any model, since the middle region has a lot of amino acid divergence, and has no significant homology with lamprey LV chain B structure. Amino acid residues in this region may only be required mainly as a source of amino acids for growth, and to fold into a lipid pocket for supplying nutrients to the developing embryo. If it does not have enzymatic function or amino acid interaction with other molecules, this region can tolerate for more amino acid change. The structure of lamprey LV (Anderson *et al*., 1998) shows that the A-sheet forms most of the molecular surface for the lipid cavity, and when compared to other LV family members, this domain contains the smallest number of conserved residues (16%), again, probably because it has the fewest protein-protein contacts and the most side chain contacts with lipid whereas the N-sheet and helical domains, which have the most protein-protein contacts are the most highly conserved-31% and 30%, respectively.

The EasyPred_Vg model shows a different direction for the N-sheet and its connection to other domains than the other models. There is not enough information to accurately model the connections between domains and the correct positions.

5. Expression of Vg gene in different tissues.

The site of yolk protein synthesis in crustaceans has been the subject of controversy for several decades. Recently, sites of Vg synthesis in decapod crustaceans have been determined (Table 13) using molecular techniques (RT-PCR and/or Northern blot). By detecting the presence of mRNA which reflects the site of expression of the yolk precursor gene, this is the most reliable criterion for determining the site of Vg synthesis. In the present study, RT-PCR results showed that the mRNA encoding Vg was present in both ovary and hepatopancreas of vitellogenic *P. merguiensis* females. The cDNA fragment of approximately 600 bp was not detected in muscle, heart or intestine of the same females. The Vg gene was also not expressed in the hepatopancreas of mature males and this result agrees with in *C. quadricarinatus* (Serrano-Pinto *et al*., 2004), *M. ensis* (Tsang *et al*., 2003) and *P. semisulcatus* (Avarre *et al*., 2003). In addition, earlier investigations in *P. semisulcatus* reported that Vg was produced in both the hepatopancreas and ovarian tissues through *in vitro* incubation studies (Fainzilber *et al*., 1992) and cell-free synthesis (Khayat *et al*., 1994a). It is now clear that Vg is synthesed both in the ovary and hepatopancreas in many crustaceans.

Table 13 Sites of Vg synthesis in decapod crustaceans.

Species	Sites of Vg synthesis	References
Penaeus merguiensis	Ovary, hepatopancreas	This study
Penaeus monodon	Ovary, hepatopancreas	Tiu *et al*., 2006
Litopenaeus vannamei	Ovary, hepatopancreas	Raviv *et al*., 2006
Penaeus semisulcatus	Ovary, hepatopancreas	Avarre *et al*., 2003
Marsupenaeus japonicus	Ovary, hepatopancreas	Tsutsui *et al*., 2000
Cherax quadricarinatus	Ovary, hepatopancreas	Serrano-Pinto *et al*., 2004
*Metapenaeus ensis*Vg1	Ovary, hepatopancreas	Tsang *et al*., 2003
*Metapenaeus ensis*Vg2	Hepatopancreas	Tsang *et al*., 2003
Pandalus hypsinotus	Hepatopancreas	Tsutsui *et al*., 2004
Macrobrachium rosenbergii	Hepatopancreas	Jasmani *et al*., 2004
Charybdis feriatus	Hepatopancreas	Mak *et al*., 2005

6. Relative quantification of Vg mRNA expression during ovarian development by real-time-PCR

Our real-time PCR results confirm that Vg synthesis occur in both ovary and hepatopancreas of vitellogenic *P. merguiensis* females as determined by the multiplex RT-PCR approach. Changes in relative Vg mRNA values differed between these two tissues. In ovary, the relative value of Vg mRNA is highest at the early vitellogenic stage (stage 2), and thereafter those levels rapidly decreased. On the other hand, the highest relative value of Vg mRNA in the hepatopancreas was found at the vitellogenic stage (stage 3) of ovarian development, and dropped in the late vitellogenic stage. The dynamic patterns of Vg mRNA expression in the hepatopancreas and ovary of intact *P. merguiensis* are similar to those of intact *M. japonicus* (Tsutsui *et al*., 2000; 2005).

Ovary is the major site of Vg synthesis in *P. merguiensis* since the relative values in ovary are higher than in hepatopancreas at all stages of ovarian development, and this pattern is also found in *L. vannamei* (Raviv *et al*., 2006). In *M. rosenbergii*, hepatopancreas is the principal site of Vg synthesis with ovary being only a minor contributior since Vg mRNA expression was negligible in the ovary of both intact and eyestalk ablated animals (Jayasankar *et al*., 2002). In *P.*

monodon, ovary and hepatopancreas produce equal amounts of Vg1 transcripts when determined by RT-PCR approach (Tiu *et al*., 2006).

Vg expression results allow the examination of the dynamics of tissue-specific contributions of Vg synthesis and indicate that the roles of these tissues in Vg synthesis are distinct. In a comparison of quantitative results from real-time PCR and ELISA in *P. merguiensis* (Auttarat *et al*., 2006), the concentrations of Vg in the hemolymph throughout the ovarian maturation stages are very much lower than those of Vt in the ovaries, parallel with the relative values of Vg mRNA in hepatopancreas and ovary. Vg concentration was not detected in hepatopancreas of banana shrimp revealing that after Vg synthesis in hepatopancreas, it is excreted into the hemolymph immediately. In ovary, the relative value of Vg mRNA is highest at stage 2 and then rapidly decreases at stage 3 and 4, meanwhile the increasing ovarian Vt concentrations correlates well with the GSI values during ovarian maturation. The explanation is that Vg mRNA is transcribed, translated and then processed to produce Vt within the ovary. So Vt accumulates inside the ovary, thus the concentration of Vt detected by ELISA correspond well with the GSI values (Auttarat *et al*., 2006).

7. Cloning of Vg cDNA from hepatopancreas and nucleotide-amino acid differences analysis

In this study, Vg cDNA at the 3' end from the hepatopancreas of *P. merguiensis* was cloned and sequenced. When the nucleotide sequence of this Vg cDNA was compared with the same region isolated from the ovary (GenBank Accession number AY499620), there were 18 nucleotide differences. These differences were distributed throughout the sequence but only 5 led to amino acid differences. Based on nucleotide-amino acid differences analysis, the probability that the real sequence differences are due to random laboratory error is less than 0.001, suggesting the existence of two different but very similar Vg genes, that may be tissue specific. Note that the patterns of Vg mRNA expression between the hepatopancreas and ovary are different and expression in the hepatopancreas is much lower than in the ovary at all stages of ovarian development (real-time PCR result). This is further evidence that there may be more than one Vg gene in *P. merguiensis*.

Several published papers have found that decapod crustaceans have multiple Vg genes expressed in ovary and hepatopancreas. There is evidence from tissue-specific expression

patterns, such as in *P. monodon* (Tiu *et al*., 2006), *M. ensis* (Tsang *et al*., 2003, Kung *et al*., 2004), *C. quadricarinatus* (Serrano-Pinto *et al*., 2004) and *C. feriatus* (Chan *et al*., 2005). Furthermore, in *C. feriatus* smaller Vg mRNA transcripts were detected in hepatopancreas, suggesting alternative splicing of the Vg gene to produce the smaller transcripts (Mak *et al*., 2005). However, for *M. japonicus* (Tsutsui *et al*., 2000; 2005) and *P. semisulcatus* (Avarre *et al*., 2003) the Vg gene sequences which are expressed in ovary and hepatopancreas are likely identical. Some publications have reported only one Vg cDNA sequence such as *L. vannamei* (Raviv *et al*., 2006), *M. rosenbergii* (Okuno *et al*., 2002; Jasmani *et al*., 2004) and *P. hypsinotus* (Tsutsui *et al*., 2004). Further work is required to fully determine the number of Vg genes and their tissue and timing of expression and effect of hormones or chemical reagents on Vg expression.

8. Proteomics analysis at differing GSI values of ovarian development

The proteomic analysis of ovary at differing GSI values of ovarian development in crustacean has not previously been performed; this work is the first preliminary study for proteomic analysis. Currently, there is relatively little shrimp sequence available, and no genome data. We analyze mass spectrums based on crustacean or all databases that are available from NCBI (http://www.ncbi.nlm.nih.gov), Swiss-Prot and TrEMBL (http://cn.expasy.org) for identifying proteins. Criteria for identifying proteins include percentage coverage, number of peptide matches and species that are more closely related to shrimp in terms of taxonomy. In *Drosophila* there are many studies of proteomics during development (Sakoyama and Okubo, 1981; Trumbly and Jarry, 1983; Alonso and Santaren, 2005) and also the complete genome sequence is available (Adams *et al*., 2000). This is very useful for the analysis in these studies. Functions of some identified proteins were determined using BLASTP by predicting the possible function from conserved domains, and from the interactive fly data base (http://flybase.bio.indiana.edu). Several important proteins were found to be involved in oogenesis and ovarian development.

A family of proteins containing the conserved motif DEAD (Asp-Glu-Ala-Asp), the DEAD box proteins are typified by the eukaryotic translation initiation factor eIF4A, and its members are believed to share the functional property of ATP-dependent RNA unwinding. One of the identified members of the family is vasa protein (spot 20) which was detected only at GSI

2.998. Vasa is the product of a maternally expressed gene in *Drosophila* and plays a role in the formation of the embryonic body plan. Vasa have been associated with two developmental processes. The first involves assembling the perinuclear region of the oocyte. Perinuclear cytoplasm is the precursor of the pole plasm which is required for localization of mRNA of the posterior determinant and serves as the cytoplasm of pole cells, the zygotic cells reserved in development for the establishment of gonads. In the middle phase of oogenesis, pole plasm moves from the perinuclear region to the pole, along with vasa protein. The integrity of the cell's cytoskeleton is crucial for the first stage of vasa localization in the periplasmic cytoplasm, as well as the second stage in the pole plasm. In the second process, vasa has helicase activity, unwinding RNA, which is a crucial process, since a three dimensional tangled or self annealed RNA would prevent transcription. The helicase function of vasa is not required for vasa localization, but it is required for the assembly of pole plasm. It is important for oocyte formation and in the specification of the posterior structures of the embryo (Lasko and Ashburner, 1988). Vasa interacts with the general translation initiation factor, and thus may regulate translation of specific mRNAs (Johnstone and Lasko, 2004).

Another protein in the eIF4A family is known as ME31B (spot 50). ME31B also has maternal (ovarian germ-line) expression during oogenesis but not embryogenesis. ME31B has proved like vasa, to encode a member of the DEAD box protein family. Vasa is expressed in both male and female germ-line cells of *Drosophila*, but ME31B expressed only in female germ-line cells of the ovary. ME31B protein unwinds specific mRNA in an ATP-dependent fashion during oogenesis and early embryogenesis, thus ME31B is another translation control protein (Valoir *et al*., 1991). ME31B was expressed at all GSI values.

Rab 3, a ras related protein, is the identification of spot 49. This protein is involved in exocytosis by regulating a late step in synaptic vesicle fusion or regulating vesicle traffic in secretory and endocytic processes. The characterization of rab3A in squid synaptic vesicles extends their known distributions to invertebrates and points to a fundamental importance of these proteins in neurotransmitter release by regulating membrane flow in the nerve terminal (Chin and Goldman, 1992). Rab 3 was expressed only at GSI 0.995 and 2.998 of ovarian development, indicating that this stage of development Rab 3 is important for the developing ovary. Neurotransmitters like serotonin (5-hydroxytryptamine, 5-HT) and dopamine (DA) are

involved in the regulation of ovarian maturation and ovulation. Experimental results from Vaca and Alfaro (2000) showed that injection of serotonin caused the induction of ovarian maturation and spawning in *L. vannamei*, but at lower rates than unilateral eyestalk ablation. Serotonin may stimulate the release of the gonad-stimulating hormone (GSH) whereas DA inhibits gonadal maturation by inhibiting the release of GSH, and also triggers release of the gonad-inhibiting hormone (GIH) (Fingerman, 1997). The DA antagonist spiperone produced an increase of the GSI in *Procambarus clarkii*, when injected during early vitellogenesis (Rodriguez *et al*., 2002) and there are also evidence for a stimulatory effect of serotonin and spiperone on the ovarian maturation and spawning of the penaeoid shrimps: *Litopenaeus stylirostris* and *L. vannamei* (Alfaroa *et al*., 2004). Vitellogenesis-inhibiting hormone (VIH) is a member of neuropeptide hormones, and inhibits Vg synthesis at the Vg synthetic site (Khayat *et al*., 1998). Thus this protein was found in the ovary of banana shrimps possibly because of their sites of vitellogenesis are ovary and hepatopancreas.

Spot 17 is predicted to be similar to hypothetical protein 18 which is a Y-box protein and contains a cold shock domain (CSD). The high conservation of CSD proteins from bacteria to human suggests that Y-box proteins have essential roles in biological processes. There is evidence that Y-box proteins are involved in oogenesis. One of the *Xenopus* Y-box proteins, FRGY2, is a major component of ribonucleoprotein storage particles containing maternal mRNA within the *Xenopus* oocyte (Murray *et al*., 1992; Deschamps *et al*., 1992; Tafurl and Wolffe, 1993; Bouvet *et al*., 1995). This protein has been shown to stimulate the transcription of oocyte-specific genes, but to repress translation of the newly synthesized mRNA. Thus, the Y-box proteins are involved in the regulation of gene expression through both transcription and translation (Ranjan *et al*., 1993). The Y-box protein in *Drosophila*, Yps, is involved in the mRNA localization machinery, and necessary for the correct localization of oskar mRNA in the oocyte for germ cell formation (Wilhelm *et al*., 2000). The Y-box proteins in *Xenopus* seem to repress the translation of mRNA until an appropriate developmental time when the function of the stored mRNA is necessary (Yurkova and Murray, 1997). Y-box protein, BYB in *Bombyx mori* acts as a kind of nucleic acid chaperone and regulates reactions involving various protein-nucleic acid interactions, BYB also stimulated the reaction of some restriction endonucleases under cold conditions (Takiya *et al*., 2004).

The steps between gene transcription and mRNA translation, which include nuclear RNA processing, mRNA trafficking, and cytoplasmic mRNA degradation, are increasingly seen as important regulatory sites in diverse cellular processes. Many of these steps in mRNA metabolism appear to be regulated by RNA binding proteins containing the K-homology, or KH, domain with a conserved core sequence of VIGxxGxxI. This kind of protein was found at spot 33 (CG6203). One KH domain protein is vigilin, its is proposed roles include chromosome partitioning, facilitating translation and tRNA transport, and control of mRNA metabolism, including estrogen-mediated stabilization of Vg mRNA. Vigilin is an estrogen-inducible protein which binds specifically to a segment of the 3'-UTR of Vg mRNA, an area which has been implicated in the estrogen-mediated stabilization of Vg mRNA (Dodson and Shapiro, 1997). The estrogen-mediated induction of Vg mRNA is brought about both by an increase in the rate of Vg gene transcription and by stabilization of cytoplasmic Vg mRNA (Brock and Shapiro, 1982; Brock and Shapiro, 1983). Spot 33 (CG6203) was expressed only at GSI 2.998 and 8.273 of ovarian development and it plausibly has vigilin function.

In addition, CG7074 (spot 36) corresponds to the missing oocyte (mio) gene. The mio gene encodes a highly conserved protein that preferentially accumulates in pro-oocyte nuclei in the early prophase of meiosis I. The product of the mio gene acts in the oocyte nucleus to facilitate the execution of the unique cell cycle and developmental programs that produce the mature haploid gamete. Mio is not only required for the maintenance of the meiotic cycle but also for oocyte identity during oogenesis. There are at least two developmental stages; first, mio may be required for the specification and early differentiation of the oocyte, secondly, mio may act later in oogenesis to maintain oocyte identity (Lida and Lilly, 2004). In banana shrimp, the product of the mio gene was expressed at all differing GSI values of ovary development.

Another maternal effect gene was found in the ovary of *P. merguiensis* as swallow (spot 57), one of the genes whose product is required for the localization of bicoid message during *Drosophila* oogenesis (Chao *et al*., 1991; Hegde and Stephenson, 1993) by a cytoskeletal organization mechanism (Pokrywka *et al*., 2000). The specification of anterior development is currently understood to be an aspect of maternal regulation of embryonic pattern. The bicoid mRNA is localized at the anterior tip of the egg (Berleth *et al*., 1988) and protein is translated from the localized message early in embryogenesis (Driever and Nusslein-Volhard,

1988). The bicoid protein is a trans-acting transcription factor which activates the zygotic expression of genes in a broad anterior domain (Driever and Nusslein-Volhard, 1989).

Calreticulin (spot 15) is a protein only found in the lumen of the endoplasmic reticulum (ER). The protein affects many cellular functions, both in ER lumen and outside the ER environment. In the ER lumen, calreticulin performs two major functions: chaperoning and regulation of Ca^{2+} homoeostasis. Molecular chaperones prevent the aggregation of partially folded proteins, increase the yield of correctly folded proteins and protein assemblies. In addition, calreticulin chaperones are involved in the quality-control process during the synthesis of a variety of molecules, including ion channels, surface receptors, integrins and transporters and are also involved in chaperoning of glycoproteins. Calreticulin modulates cell adhesion, integrin-dependent Ca^{2+} signalling and steroid-sensitive gene expression (Michalak *et al*., 1999). Calreticulin shows high expression at GSI values 2.998 and 8.273 of ovarian development.

Several spots were found to be cytoskeleton proteins. Some proteins may function to establish cell shape those found in all stage of ovarian development and some proteins may be involved in intracellular transport of molecules and organelles those over expressed in stage 2 and 3 of ovarian development. Tubulin is the major constituent of microtubules, involved in least three important developmental processes during *Drosophila* oogenesis: oocyte determination and growth, positioning of the anterior determinant bicoid mRNA and ooplasmic streaming, driven by vesicle translocation along microtubules. Maternal components synthesized in nurse cells are transported through cytoplasmic bridges to the oocyte, and microtubules are likely to serve as a polarized scaffold on which maternal RNAs and proteins are transported (Theurkauf *et al*., 1992). The unidirectional movements of the microtubule-associated motors, dyneins, as detected at spot 60, provide an important mechanism for the positioning of cellular organelles and molecules. Within the developing egg chambers of ovary, the dynein gene is predominantly transcribed in the nurse cell complex in *Drosophila*. The temporal and spatial pattern of dynein accumulation in the oocyte is remarkably similar to that of several maternal effect gene products that are essential for oocyte differentiation and axis specification. Microtubule motors participate in the transport of these morphogens from the nurse cell cytoplasm to the oocyte (Li *et al*., 1994). The driving force behind cell motility is the actin cytoskeleton. Spot 42 is a protein with homology to cofilin actin depolymerising factor, which regulates actin

filament dynamics. *Drosophila* cofilin is an important regulator of actin-based cell motility during *Drosophila* development. It also promotes cell movements during ovarian development and oogenesis, and is required for the migration of border cells during oogenesis (Chen *et al*., 2001).

The pleckstrin homology (PH) domain (spot 25) expressed only in stage 1, is involved in signal transduction and cytoskeletal function. It is possible that the PH domain of beta-spectrin plays a part in the association of spectrin with the plasma membrane of cells (Zhang *et al*., 1995). Spectrin is an elongated molecule that is a constituent of the submembrane cytoskeleton of epithelial cells, making up many tissues of the fly. Spectrin is illustrated by the follicular epithelium surrounding the egg. Spectrin influences cells in one of two ways. It can be thought of as part of an infrastructure that functions to stabilize cell shape and/or cell-cell contacts, or as a scaffold for the proper (stable) positioning of membrane bound proteins and other cytoskeletal elements or proteins involved in cell signaling and required in a development pathway that controls follicle cell monolayer and proliferation (Lee *et al*., 1997a). Furthermore, spectrin is involved in oogenesis, spectrin is a component of the fusome required for cyst formation by anchor mitotic spindles during asymmetric germ cell divisions, organizing a polarized microtubule-based transport system for RNA localization that leads to oocyte differentiation and oocyte determination (Cuevas *et al*., 1996; Deng and Lin, 1997).

Oligopeptide transporter, CPTB (spot 12) is a peptide transporters across the cell membrane. It recognizes and transports di-, tri- and tetra-peptides as part of two major functions in animals; absorption of small peptides in the epithelial cells of the intestinal tract and kidney tubules and re-uptake of neuropeptide degradation products in the nervous system (Yamashita *et al*., 1997) and might be related to the nutrition roles (Fei *et al*., 1998). The protein in spot 12 is expressed at all GSI values of ovarian development, and might be involved in reuptake of neuropeptides.

Serine proteases match the pattern of spot 40. These proteins have been implicated in many physiological processes in which specific proteins are activated by proteolytic cleavage. Some reported serine proteases are important in processes like hemolymph coagulation, melanization of pathogen surfaces, and antimicrobial peptide synthesis (Gorman and Paskewitz, 2001).

Ribosomal protein L4 (spot 59) specifically regulates the S10 operon, which codes for 11 r-proteins including L4 itself. L4 is unique among the regulatory ribosomal proteins because it regulates not only translation but also transcription of the S10 operon mRNA (Yates and Nomura, 1980). This protein might have a specific function for these stages of ovarian development.

Vg was found at spot 74, and was found in the ovary extracts which have GSI value 2.998 and 8.273. The size of the spot is surprisingly small as we expect an accumulation of Vt during ovarian development. The most likely reason is that Vt was lost during preparation process for 2-DE. Since Vt is a lipoglycocarotenoprotein, it might have been present in the orange pellet or lipid layer that were discarded after every centrifuge steps.

Overall, the proteomic maps of ovary extract at differing GSI value appeared most similar between GSI 2.998 and 8.273. In particular, there are fewer sports at GSI 0.995. GSI 0.995 precedes full ovarian development, so it is not surprising far more protein are seen in the ovary that have GSI 2.998 and 8.273, which is the period of most active ovarian development and oocytes synthesis.