Molecular characterization of cockroach vitellogenins and vitellogenin receptor mechanisms

Tufail, Muhammad

(Degree)
博士（学術）

(Date of Degree)
2003-03-31

(Date of Publication)
2007-10-16

(Resource Type)
doctoral thesis

(Report Number)
甲2760

(URL)
https://hdl.handle.net/20.500.14094/D1002760

Doctoral Dissertation

Molecular characterization of cockroach vitellogenins and vitellogenin receptor mechanisms

ゴキブリのビテロジェニンおよびその受容体機構に関する分子生物学的研究

Muhammad Tufail

The Graduate School of Science and Technology, Kobe University

March, 2003

DEDICATIONS

To the souls of my parents, and brother Ch. Muhammad Anwar

CONTENTS

CHAPTER 1

Vitellogenins/vitellogenin receptors in general, reasons to study and present status of knowledge

1.1 Vitellogenins/vitellogenin receptors in general

Vitellogenins (Vgs) are precursors of the major egg yolk protein, vitellins (Vns), in insects and other oviparous animals. Vgs of most insect species are synthesized in the fat body cells, in tissue-, sex- and stage-specific manners, secreted into the hemolymph and then ultimately taken up by the developing oocytes via recepter-mediated endocytosis (reviewed by Byrne et al., 1989; Raikhel and Dhadialla, 1992). During these processes Vgs and Vns are modified through cleavage, glycosylation, lipidation, and phosphorylation (reviewed by Raikhel and Dhadialla, 1992; Hagedorn et al., 1998; Giorgi et al., 1999). After incorporation into the eggs, Vns represent a major component of egg yolk proteins and serve as storage proteins to provide a source of nutrients for the developing embryo (reviewed by Kunkel and Nordin, 1985; Bownes, 1986; Byrne et al., 1989). In insects, a sex-linked "female-protein" was first found in the hemolymph of the silkmoth *Hyalophora cecropia* and was shown to participate in yolk formation by Telfer (1954). The major source of this protein was identified as the fat body and it was named "vitellogenin" as the precursor of Vn or yolk protein by Pan et al. (1969).

The synthesis of Vg in most, if not all, insect species is regulated at the transcriptional level. The sex- and tissue-associated, and hormone-mediated developmental specificities of Vg transcription have been reported for many insect species (Bownes, 1986). Most insect Vgs are synthesized as a primary Vg gene product of about 200-240 kD and are processed into large (140-190 kD) and small (40-60 kD) subunits [(as has been observed in locust, silkworm and mosquito (Dadhiala and Raikhel, 1990, cicada (Lee et al., 2001)] before being secreted into the hemolymph. In

cockroaches, however, there exist also the processed products of ~90-110 kD (medium polypeptides) in addition to the large and small polypeptides (Tufail et al., 2001; Tufail and Takeda, 2002). In bees and wasps (higher Hymenoptera), the primary Vg gene product, with a molecular mass of about 180 kD is secreted without processing (Kageyama et al., 1994). In contrast, the vitellogenins of higher Diptera (appropriately called yolk proteins) are quite different from those of other insects. In *Drosophila melanogaster,* for example, three major yolk proteins of 46, 45 and 44 kD exist, each encoded by single-copy genes located on the X-chromosome (Bownes et at., 1993), and are not processed. Similar examples are found in other higher Diptera such as the fruit fly, *Ceratitis capitata* (Rina and Savakis, 1991) and the blow fly, *Calliphora erythrocephala* (Martinez and Bownes, 1994).

In insects, the yolk proteins may be divided into four categories: 1) Vitellin of most insect species, which is derived from vitellogenin, the precursor which is synthesized in the fat body cells in a female specific manner, secreted into the hemolymph and then ultimately taken up by the developing oocytes; 2) Yolk proteins (YPs) of higher Diptera, such as *D. melanogaster,* which are the products of the genes expressed both in the female fat body cells and ovarian follicle cells and are incorporated into developing oocytes; 3) Proteins such as the egg specific protein (ESP) of *Bombyx mori* produced in ovarian follicle cells and incorporated into the developing oocytes; and 4) Proteins such as the 30 kDa protein of *Bombyx mori* produced in the fat body cells, but not in a sex specific manner, which is secreted into the hemolymph and sequestered into the developing oocytes.

The accumulation of Vg by the developing oocytes during oogenesis is a characteristic phenomenon of all oviparous animals (Stifani et al., 1990; Raikhel nd Dhadialla, 1992). Internalization of Vg by the oocyte is achieved through receptor-mediated endocytosis, a ubiquitous mechnism for the selective intake of macromolecules by animal cells (Goldstein et al., 1979; Anderson and Kaplan, 1983). Although several types of yolk protein precursors are accumulated by insect oocytes (see reviews: Raikhel and Dhadialla, 1992, Izumi et al., 1994; Rajaratnam, 1996,

Sappington and Raikhel, 1998), Vg, with few exceptions, is by far the most abundant in all insect species.

Receptors where Vg anchors are localized in coated pits on the surface of growth competent oocytes which are engulfed with the yolk protein precursor into cells by receptor mediated endocytosis (Telfer et al., 1982; Byrne et al., 1989; Raikhel and Dhadialla, 1992). Yolk accumulation by growing oocytes via receptor-mediated endocytosis has been demonstrated in many species (Schneider, 1996; Grant and Hirsh, 1999; David et al., 2001). Extensive ligand and receptor tracer studies have led to a model for general endocytic trafficking routes in the cell (Mukherjee et al., 1997). Ligand-receptor complexes are thought to cluster in clathrin coated pits, which pinch off from the surface as clathrin-coated vesicles. The coated vesicles, after removing their coats, fuse with each other and with early endosomes. The acidification of early endosomes results in dissociation of ligand-receptor complexes. The receptors are then sent back to the cell surface for reuse, whereas the ligands are transported to late endosomes and ultimately lysosomes. Yolk storage vesicles are thought to be the oocyte equivalent of late endosomes or lysosomes but with low proteolytic activity (Schneider, 1996).

Molecular characterization of VgRs from different species, including both vertebrates [such as chickens (Bujo et al., 1994; Okabayashi et al., 1996), frog (Part et al., 1998) and rainbow trout (Davail et al., 1988)] and invertebrates [such as a nematode (Grant and Hirsh, 1999), the mosquito (Sappington et al., 1996) and a yolk protein receptor (YPR) of *D. melanogaster* (Schonbaum et al, 1995)] demonstrated that they all are the members of the low density lipoprotein receptor (LDLR) gene superfamily. The deduced amino acid sequences of the known VgRs revealed that these receptors, like other LDLR family members, contain five distict domains: clusters of cysteine-rich repeats constituting the ligand-binding domain, epidermal growth factor (EGF)-like repeats, repeats containing a YWTD motif that are proposed to form a B-propeller domain (Springer, 1998), a tansmembrane domain anchoring the receptor in the plasma membrane, and a cytoplasmic domain containing at least one copy of the

NPXY sequence (di-leucine or leucine-isoleucine motif in insect VgR/YPRs) which is involved in receptor internalization via coated pits (Goldstein et al., 1985; Chen et al., 1990). The insect VgR/YPRs are, however, large (180-214 kD) membrane-bound proteins, roughly twice the size of vertebrate VgRs (95-115 kD) (Ferenz, 1993; Sappington and Raikhel, 1995).

1.2 Reasons to study vitellogenins and present status of knowledge

Vitellogenin genes and cDNAs of different insect species so far cloned have shown that the primary structure of the Vg polypeptides were highly conserved (Chen et al., 1997; Lee et al., 2000). It was also reported that the Vg genes of oviparous animals, both invertebrates and vertebrates, are well conserved and form a gene superfamily. Evolutionary conservation of Vg genes has raised a possibility that Vg gene and the product (Vg) might be used as a molecular marker to indicate the phylogenetic relationships. Previousely Nose et al. (1997) has constructed a phylogenetic tree of insects, using the amino acid sequences of the Vgs of 6 species. The tree was in agreement with the accepted ones based on the morphological characteristics and ribosomal DNAs, thus suggesting Vg is one of the candidate molecular markers. Since Vg genes and cDNA cloning has been done in very few species in hemimetabolous insects, we have, therefore, chosen the cockroaches as a representative of polyneopteran insects for the purpose.

The Vg genes and/or cDNAs also provide excellent tools for studying the molecular basis of gene regulation. Recently, for example, a partial Vg cDNA from the German cockroach, *Blattella germanica* was obtained and utilized for monitoring the Vg gene expression (Martin et al., 1998; Comas et al., 1999). The sensitivity was quite high as compared to the conventional methods like immunological methods and pulse-chase techniques. The analyses at molecular level provided a reliable means to analyse the hormonal regulation on gene expression (Martin et al., 1998; Comas et al., 1999).

In the present study, we report on cloning and characterization of the three

complete Vg cDNAs, the two (Vg1 and Vg2) being that of the American cockroach, *Periplaneta americana* (Tufail et al., 2000; 2001) and one being that of *Leucophaea maderae* (Tufail and Takeda, 2002). We also show their use as a probe in Northern blot analyses to assess the tissue- and sex-specificity of the Vg gene expression in both the cockroach species. The Vg cDNA probe was also used to investigate the stage specific expression of Vg gene in the female fat body cells of *P. americana*. The expression of Vg gene was first detected by Northern blot analysis in the fat body cells of 2-days-old adult females, and the hemolymph Vgs were first detected by Western blot analysis in 4-days-old adult females. These results demonstrate that the *P. americana* Vg cDNA clone will be a sensitive probe to study the hormonal regulation on Vg gene expression.

The deduced amino acid sequences from three Vg cDNAs described here in this report were confidently aligned among each other and among the other known insect Vgs and a molecular phylogenetic tree constructed was in agreement with those constucted previousely based on the morphological characteristics and the molecular markers. We also clarify, on molecular basis, the processing patterns of cloned Vg molecules from both the cockroach species.

Moreover, we have also cloned and sequenced another Vg cDNA of *L. maderae* which has stretches of amino acid sequences different from the one reported previously (Tufail and Takeda, 2002). The complete nucleotide sequence of this Vg (which we named as Vgb) consists of 5915 bp which encodes a deduced amino acid sequence of 1911 amino acid residues (including a putative signal peptide sequence) in a single open reading frame. The comparison of base sequences of both *L. maderae* Vgs revealed that the difference was due to mutations (addition/deletion of a base (s) or addition of a base (s) at one position followed by deletion of a base (s) at another position and vice versa) in the base sequence of Vgb (the other Vg) which made its amino acid sequence different from the one reported previously (Tufail and Takeda, 2002). The both *L. maderae* Vgs were, however, showing 96% similarity in the protein primary structure.

Next, we also report here for the similarity in Vn antigenicity among 10 selected

cockroach species, belonging to two superfamilies, using the anti-*P. americana* Vn-antisera. The results demonstrate that the Vn antigenicities are at least very similar within the members of the same superfamily.

1.3 Reasons to study vitellogenin receptor and present status of knowledge

The molecular analyses of VgRs, both from vertebrates and invertebrates, revealed that they all are the members of LDLR gene superfamily. This family of receptors mediate endocytosis and lysosomal targeting of a diverse array of macromolecules and impact directly or indirectly most, if not all, physiological processes including reproduction, development and nutrition and many pathophysiological processes. The VgRs are particularly involved in reproduction. The chicken VgR, for example, has been shown to import very low density lipoprotein (VLDL), riboflavin-binding protein, and alpha$_2$-macroglobulin into growing oocytes (Stifani et al., 1990; Mac Lachlan et al., 1994). Indeed, mutations that abrogate expression of the VgR result in non-egg laying hens. In *D. melanogaster*, the YPR is encoded by the gene yolkless (*yl*) (Schonbaum et al., 1995). Female sterility occurs in insects having genetic deficiency of *yl* (Schonbaum et al., 1995). Analysis of oocytes produced by yl^{-1} females shows drastic reduction in numbers of coated pits and coated vesicles and very little protenacious yolk (DiMario and Mahowald, 1987).

The insect oocyte provides an excellent model system for studying receptor-mediated endocytosis because of the high intensity of protein uptake. This system could also be a promising target for the pest control. For example, interruption of the receptor-ligand interaction would block egg production, and the knowledge can be exploited for the safer pest control strategies.

To manipulate this system/strategy the basic information required is a thorough understanding of structures, interactions, regulation and expression of the proteins involved (both the ligand and the receptor). We have recently characterized the molecular structures of two Vg molecules, the ligands, from the American cockroach, *P. americana* (Tufail et al., 2000 and 2001). We now report the cDNA cloning and

structural analysis of their counterpart, the receptor, from this cockroach species, and describe that the VgR we cloned is a new LDLR family member having only 6 repeats in the second ligand-binding domain. This novel insect VgR consists of 1709 amino acid residues and shares a significant homology with other LDLR family members and particularly with Vg/Yp receptors reported from the mosquito and *Drosophila*. The cytoplasmic tail of *P. americana* VgR contains a leucine-isoleucine internalization signal, the motif that seems to be common in insect VgRs/YPRs (di-leucine in *Drosophila* YPR), unlike the tight-turn-tyrisine motif (NPXY) of other members of LDLR gene family. The phylogenetic (neighbour-joining) analysis shows that the mosquito VgR and YPR of *Drosophila* are more closely related than with the cockroach VgR.

During cloning of VgR, we also found a clone encoding receptor tyrosine kinase (RTK) from the previtellogenic ovaries of *P. americana*. The complete cDNA for RTK was of 4128 residues which encoded a deduced amino acid sequenece of 1294 residues long including 22 residues for the putative signal peptide. The deduced amino acid sequence was aligned confidently with other members of the RTK family of receptors. We are, however, not clear about its role in the previtellogenic ovaries at the moment. One possibility is that RTK may be involved in the cell proliferation, especillay the follicle cells.

In the present study, we report on cloning and characterization of cockroach Vgs and VgR mechanisms which elusidates the structures, expressions/regulations and processing patterns (in Vgs) of these proteins. We also show the use of these proteins to indicate the phylogenetic relationships. Moreover, we also examine the cross reactivity among Vg molecules from ten cockroach species.

1.4 LITERATURE CITED

Anderson, R. G. W., Kaplan, J., 1983. Receptor-mediated endocytosis. Mod. Cell Bil. 1, 1-52.

Bownes, M., 1986. Expression of genes coding for vitellogenin (yolk protein). Annu. Rev. Entomol. 31, 507-531.

Bownes, M., Ronaldson, E., Mauchline, D., Martinez, A., 1993. Regulation of vitellogenesis in *Drosophila*. Int. J. Morph. and Embry. 22, 349-367.

Bujo, H., Hermann, M., Kaderli, M. O., Jacobsen, L., Sugawara, S., Nimpf, J., Yamamoto, T., Schneider, W. J., 1994. Chicken oocyte growth is mediated by an eight ligand repeat member of the LDL receptor family. EMBO J. 13, 5165-5175.

Byrne, B. M., Gruber, M., Ab, G., 1989. The evolution of egg yolk proteins. Progr. Biophys. Molec. Biol. 53, 33-69.

Chen, J. S., Sapington, T. W., Raikhel, A. S., 1997. Extensive sequence conservation among insect, nematode and vertebrate vitellogenins reveals ancient common ancestry. J. Mol. Evol. 44, 440-451.

Chen W. J., Goldstein J. L., Brown M. S., 1990. NPXY, a sequence often found in cytoplasmic tails, is required for coated pit-mediated internalization of the low density lipoprotein receptor. J. Biol. Chem. 265, 3116-3123.

Comas, D., Piulachs, M. D., Belles, X., 1999. Fast induction of vitellogenin gene expression by juvenile hormone III in the cockroach *Blattella germanica* (L.) (Dictyoptera, Blattellidae). Insect Biochem. Mol. Biol. 29, 821-827.

Davail, B., Pakdel, F., Bujo, H., Perazzolo, L. M., Waclawek, M., Schneider, W. J., Le Menn, F., 1988. Evolution of oogenesis: the receptor for vitellogenin from the rainbow trout. J. Lipid Res. 39, 1929-1937.

David, S. R., Matthew, G., Brandy, C., Debra, M. H., Suzanne, S., Danielle, S., 2001. Yolk protein endocytosis by oocytes in *Drosophila melanogaster*: immunofluorescent localization of clathrin, adaptin and the yolk protein receptor. 47, 715-723.

Dhadialla, T. S., Raikhel, A. S., 1990. Biosynthesis of mosquito vitellogenin. J. boil. Chem. 265, 9924-9933.

DiMario, P. J., Mahowald, A. P., 1987. Female sterile (1) yolkless: a recessive female sterile mutation in *Drosophila melanogester* with depressed numbers of coated pits and coated vesicles within the developing oocytes. J. Cell Biol. 105, 199-206.

Ferenz, H.-J., 1993. Yolk protein accumulation in *Locusta migratoria* (R. and F.) (Orthoptera: Acrididae) oocytes. Int. J. Insct Morphol. Emryol. 22, 295-314.

Giorgi, F., Bradley, J. T., Nordin, J. H., 1999. Differential vitellin polypeptide processing in insect embryos. Micron. 30, 579-596.

Goldstein, J. L., Anderson, R. G. W., Brown, M. S., 1979. Coated pits, coated vesicles, and receptor mediated endocytosis. Nature. 279, 679-685.

Goldstein, J. L., Brown, M. S., Anderson, R. G. W., Russell, D. W., Schneider, W. J.,1985. Receptor-mediated endocytosis: concepts emerging from the LDL receptor system. Ann. Rev. cell Biol. 1, 1-39.

Grant, B., Hirsh, D., 1999. Receptor-mediated Endocytosis in the *Caenorhabditis elagans* oocyte. Mol. Biol. Cell. 10, 4311-4326.

Hagedorn, H. H., Maddison, D. R., Tu, Z., 1998. The evolution of vitellogenins, cyclorrhaphan yolk proteins and related molecules. Adv. Insect Physiol. 27, 335-384.

Izumi, S., Yano, K., Yamamoto, Y., Takahashi, S.Y., 1994. Yolk proteins from insect eggs: structure, biosynthesis and programmed degradation during embryogenesis. J. Insect Physiol. 40, 735-746.

Kageyama, Y., Kinoshita, T., Umesono, Y., Hatakeyama, M., Oishi, K., 1994. Cloning of cDNA for vitellogenin of *Athalia rosae* (Hymenoptera) and characterization of the vitellogenin gene expression. Insect Biochem. Mol. Biol. 24, 599-605.

Lee, J. M., Hatakeyama, M., Oishi, K., 2000a. A simple and rapid method for cloning insect vitellogenin cDNAs. Insect Biochem. Mol. Biol. 30, 189-194.

Lee, J. M., Nishimori, Y., Hatakeyama, M., Bae, T-W., Oishi, K., 2000b. Vitellogenin of the cicada *Graptopsaltria nigrofuscata* (Homoptera): analysis of its primary structure. Insect Biochem. Mol. Biol. 30, 1-7.

Mac Lachlan, I., Nimpf, J., Schneider, W. J., 1994. Avian riboflavin binding protein binds to lipoprotein receptors in association with vitellogenin. J Biol. Chem. 269, 24127-24132.

Martinez, A., Bownes, M., 1994. The sequence and expression pattern of the *Calliphoraerythrocephala* yolk protein A and B genes. J. molec. Evol. 38, 336-351.

Martin, D., Piulachs, M. D., Comas, D., Belles, X., 1998. Isolation and sequence of a

partial vitellogenin cDNA from the cockroach, *Blattella germanica* (L.) (Dictyoptera, Blattellidae), and characterization of the vitellogenin gene expression. Arch Insect Biochem Physiol. 38, 137-146.

Mukherjee, S., Ghosh, R. N., Maxfield, F. R., 1997. Endosysis. Physiol. Rev. 77, 759-803.

Nose, Y., Lee, J. M., Ueno, T., Hatakeyama, M., Oishi, K., 1997. Cloning of cDNA for vitellogenin of the parasitoid wasp, *Pimpla nipponica* (Hymenoptera: Apocrita: Ichneumonidae): vitellogenin primary structure and evolutionary considerations. Insect Biochem. Mol. Biol. 27, 1047-1056.

Okabayashi, K., Shoji, H., Nakamura, T., Hashimoto, O., Asashima, M., Sugino, H., 1996. cDNA cloning and expression of the *Xenopus laevis* vitellogenin receptor. Biochem. Biophys. Res. Commun. 224, 406-413.

Pan, M. L., Bell, W. J., Tefler, W. H., 1969. Vitellogenic blood protein synthesis by insect fat body. Science. 165, 393-394.

Part, F., Coward K, Sumpter, J. P., Tyler, C. R., 1998. Molecular characterization and expression of two ovarian lipoprotein receptors in the rainbow trout *Oncorhynchus mykiss*. Biol. Reprod. 58, 1146-1153.

Raikhel, A. S., Dhadialla, T. S., 1992. Accumulation of yolk proteins in insect oocytes. Annu. Rev. Entomol. 37, 217-251.

Rajaratnam, V. S., 1996. Isolation, characterization and complete nucleotide sequence of a *Galleria mellonella* cDNA coding for the follicle cell-specific yolk protein, YP4. Insect Bioch. Mol. Biol. 26, 545-555.

Rina, M., Savakis, C., 1991. A cluster of vitellogenin genes in the Mediterranean fruitfly, *Ceratitis capitata*: sequence and structural coservation in dipteran yolk proteins and their genes. Genetics. 127, 769-780.

Sappington, T. W., Hays, A.R., Raikhel, A. S., 1995. Mosquito vitellogenin receptor-purification, developmental and biochemical characterzation. Insect Biochem. Mol. Biol. 25, 807-817

Sappington, T. W., Kokoza, V. A., Raikhel, A. S., 1996. Molecular characterization of the mosquito vitellogenin receptor reveals unexpected high homology to the *Drosophila* yolk protein receptor. Proc. Natl. Acad. Sci. USA. 93, 8934-8939..

Sappington, T. W., Raikhel, A. S., 1998a. Molecular characteristics of insect vitellogenins and vitellogenin receptors. Insect Biochem. Mol. Biol. 28, 277-300.

Sappington, T. W., Raikhel, A. S., 1998b. Ligand-binding domains in vitellogenin receptors and other LDL-receptor family members share a common ancestral ordering of cysteine-rich repeats. J. Molec. Evol. 46, 476-487.

Schneider, W. J., 1996. Vitellogenin receptors: oocyte specific members of the low-density lipoprotein receptor supergene family. Int. Rev. Cytol. 166, 103-137.

Schonbaum, C. P., Lee, S., Mahowald, A. P., 1995. The *Drosophila* yolkless gene encodes a vitellogenin receptor belonging to the low density lipoprotein receptor superfamily. Proc Natl. Acad. Sci. USA. 92, 1485-1489.

Springer, T. A., 1998. An extracellular beta-propeller module predicted in lipoprotein

and scavenger receptors, tyrosine kinases, epidermal growth factor precursor and extracellular matrix components. J. Mol. Biol. 283, 837-862.

Stifani, S., Barber, D. L., Nimpf, J., Schneider, W. J., 1990. A single chicken oocyte plasma membrane protein mediates uptake of very low density lipoprotein and vitellogenin. Proc. Natl. Acad. Sci. USA. 87, 1955-1959.

Tefler, W. H., 1954. Immunological studies of insect metamorphosis. II. The role of a sex-limited blood protein in egg formation by the Cecropia silkworm. J. Gen. Physiol. 37, 538-558.

Telfer, W. H., huebner, E., Smith, S. D., 1982. The cell biology of vitellogenic follicles in *Hyalophora* and *Rhodnius*. In *Insect Ultrastructure*, eds. R. C. king, H. Akai, pp. 118-149. Plenum press, New York.

Tufail, M., Hatakeyama, M., Takeda, M., 2001. Molecular evidence for two vitellogenin genes and processing of vitellogenins in the American cockroach, *Periplaneta americana*. Arch. Insect Biochem. Physiol. 48, 72-80.

Tufail, M., Lee, J. M., Hatakeyama, M., Oishi, K., Takeda, M., 2000. Cloning of vitellogenin cDNA of the American cockroach, *Periplaneta americana* (Dictyoptera), and its structural and expression analyses. Arch. Insect Biochem. Physiol. 45, 37-46.

Tufail, M., Takeda, M., 2002. Vitellogenin of the cockroach, *Leucophaea maderae*: nucleotide sequence, structure and analysis of processing in the fat body and oocytes. Insect Biochem. Mol. Biol. 32, 1469-1476.

CHAPTER 2

Cloning of vitellogenin cDNA of the American cockroach, *Periplaneta americana* (Dictyoptera) and its structural and expression analyses

2.1 ABSTRACT

A cDNA expression library constructed from poly $(A)^+$ RNA prepared from vitellogenic female fat body cells of the American cockroach, *Periplaneta americana* (Dictyoptera) was screened using a polyclonal antiserum against the 100 kD polypeptide(s) from the egg extract. A partial Vg cDNA clone was obtained and sequenced. The 5' end portion of the cDNA was then obtained by the RACE method, cloned and sequenced. The combined complete Vg cDNA was 5854 bp-long and contained a single ORF encoding 1896 amino acids. The entire deduced amino acid sequence was aligned confidently with those of the known insect Vgs. A GL/ICG motif, a number of cysteines at conserved locations following this motif and a DGXR motif upstream of the GL/ICG motif were present near the C-terminal. The chemically determined N-terminal amino acid sequence of the 170 kD polypeptide from the egg extract completely matched the deduced sequence starting from just after one of the consensus (RXXR) cleavage sites, indicating the occurrence of post-translational cleavage in the fat body cells. The Vg gene begins to be expressed in the 2-day-old adult female fat body cells but is never expressed in ovaries or in male fat body cells. Hemolymph Vg was first detected by immunoblotting in 4-day-old adult females, two days after the beginning of gene expression. Western blot analysis of major yolk polypeptides in nine cockroach species belonging to the two superfamilies, Blattoidea and Blaberoidea, using the antisera against *P. americana* major yolk polypeptides showed that the similarities in Vn antigenicity are basically limited to within a

superfamily.

2.2 INTRODUCTION

The major yolk protein of most insects is Vn. The precursor, Vg, is synthesized extraovarially in the fat body cells in sex- and stage-specific manners, secreted into the hemolymph and incorporated into developing oocytes (reviewed in: Raikhel and Dhadialla, 1992; Hagedorn et al., 1998; Giorgi et al., 1999). In most, if not all, insect species the biosynthesis of Vg is under hormonal regulation. Juvenile hormone (JH) plays a major role in the induction of Vg synthesis in hemimetabolous insects including cockroaches. The action of JH or JH analogs upon Vg synthesis has been extensively studied, especially by using cockroaches (reviewed in: Engelmann, 1979; Wyatt, 1991; Wyatt and Davey, 1996). In most of those studies, Vg synthesis was measured by immunological methods and pulse-chase techniques, and was used as an index for hormonal activity. Recently, a Vg cDNA clone for the German cockroach, *Blattela germanica*, was obtained and utilized for monitoring Vg gene expression (Martin et al., 1998; Comas et al., 1999). The sensitivity was quite high compared to the conventional methods and the analyses at molecular level provided a reliable means to analyse the hormonal regulation on gene expression (Martin et al., 1998; Comas et al., 1999).

The American cockroach, *Periplaneta americana*, is another model species employed to examine hormonal regulation of Vg synthesis (Wyatt and Davey, 1996). Two Vns (Vn1 and Vn2) and corresponding Vgs (Vg1 and Vg2) have been identified in this species (Bell, 1970; Engelmann, 1979). Purification of these proteins revealed that both of these proteins consisted of multiple subunits. Vn1/Vg1 is composed of four major polypeptides and Vn2/Vg2 is composed of three major polypeptides (Storella et al., 1985; Kim et al., 1992; Kim and Lee, 1994). Most probably, thus, there are at least two Vg genes in *P. americana* and each of their products undergoes post-translational processing producing subunits with a smaller molecular mass.

The Vg genes and complete cDNAs have been cloned in 10 insect species [*Anthonomus grandis* (Coleoptera): Trewitt et al., 1992; *Aedes aegypti* (Diptera): Chen et al., 1994, Romans et al., 1995; *Bombyx mori* (Lepidoptera): Yano et al., 1994a, b; *Athalia rosae* (Hymenoptera): Kageyama et al., 1994, Nose et al., 1997; *Pimpla nipponica* (Hymenoptera): Nose et al., 1997; *Lymantria dispar* (Lepidoptera): Hiremath and Lehtoma, 1997a, b; *Riptortus clavatus* (Heteroptera): Hirai et al., 1998; *Graptopsaltria nigrofuscata* (Homoptera): Lee et al., 2000b; *Plautia stali* (Heteroptera): Lee et al., 2000a; *Blattella germanica* (Dictyoptera): Comas et al., 1999, 2000]. The deduced Vg amino acid sequences have been shown to align confidently along their entire lengths (Chen et al., 1997; Lee et al., 2000a). Nose et al. (1997) earlier proposed that Vg might be used as a supplemental molecular marker to construct insect phylogenetic trees. Nine species out of the 10 noted above for which the entire Vg amino acid sequences are available belong to the Paraneoptera (*R. clavatus*, *G. nigrofuscata* and *P. stali*) and the Oligoneoptera (*A. grandis*, *A aegypti*, *B. mori*, *A. rosae*, *P. nipponica* and *L. dispar*). Only one species (*B. germanica*) belongs to the Polyneoptera. Clearly cloning of Vg cDNAs from other species belonging to the Polyneoptera and the Paleoptera is needed.

In the present study, a complete Vg cDNA of *P. americana* was cloned and sequenced. The expression pattern of the Vg gene was examined by Northern blot analysis using portions of the cloned cDNA as probes. The similarity in Vn antigenicity was examined in nine cockroach species using the antisera raised against the *P. americana* Vns. The deduced amino acid sequence was aligned with those of the known insect Vgs and a molecular phylogenetic tree was constructed.

2.3 MATERIALS AND METHODS

2.3.1 Insects

All cockroach species used were from laboratory stocks which had been

maintained for many years under constant light conditions at 26°C. Newly eclosed adults were collected daily and kept separately under conditions of a 12:12 light-dark cycle at 26°C.

2.3.2 SDS-PAGE and Western Blot Analysis

SDS-PAGE and Western blot analyses were carried out basically as described previously (Hatakeyama et al., 1990; Takadera et al., 1996). Egg extracts (0.004 egg equivalent per lane) and hemolymph samples (0.05 µl equivalent per lane) from each species were electrophoresed on 7% polyacrylamide gels. Separated proteins were either stained with Coomassie blue or transferred to nitrocellulose membranes (Schleicher & Schuell, pore size 0.45 µm).

Antisera against each of the *P. americana* Vns were raised essentially as described previously (Hatakeyama et al., 1990). For each antigen injection, eggs (mature oocytes) dissected from 12-day-old adult females and stored at -20°C were homogenized with 100 µl of sample buffer (0.625 M Tris-HCl, pH 6.8: 10% SDS: 50% glycerol: distilled water=1:2:2:4.5:0.5, v/v). The homogenate was subjected to SDS-PAGE. After staining the gel, Vn bands were cut out, processed, and injected into New Zealand white rabbit females. Western blotting was carried out using the rabbit anti-*P. americana* Vn antisera diluted 1/1000 as the primary antiserum and the horseradish peroxidase-conjugated goat anti-rabbit Igs (Cappel) diluted 1/1000 as the secondary antiserum. ECL Western blotting detection reagents (Amersham-Pharmacia) were used for detection according to the supplier's protocol.

2.3.3 Construction of cDNA Expression Library, Immunological Screening and Nucleotide Sequencing.

Total RNA (669 µg) was prepared from fat body cells of 20 females (3- to 5-day-old adults) by the guanidine thiocyanate-CsCl method as described in Sambrook et al. (1989) and was used to obtain poly $(A)^+$ RNA (14.7 µg) using an mRNA

purification kit (Amersham-Pharmacia). A portion of the poly $(A)^+$ RNA (5 μg) was used to construct a cDNA expression library in the Uni-ZAP cloning vector using a lambda ZAPII XR library construction kit (Stratagene) according to the supplier's protocol.

The library was amplified and a portion (200-400 phages) of it was incubated with *Escherichia coli* XL1-Blue MRF' strain. The plaques were transferred to nitrocellulose membranes and screened immunologically with the anti-*P. americana* 100 kD polypeptide antiserum following the method described in Nose et al. (1997) based on Sambrook et al. (1989). After three successive screenings, the positive clones were excised into pBluescript II SK$^-$. The transformants were stored at -80°C as 50% glycerol stocks.

The selected clones were individually digested with appropriate restriction enzymes and subcloned into pBluescript II SK$^-$. Each clone was amplified by PCR using a Thermo sequenase cycle sequencing kit (Amersham-Pharmacia) with M13 forward and reverse primers and analyzed using a DNA sequencer (Li-Cor, 4200L) according to the supplier's protocol.

2.3.4 5`-end Amplification of cDNAs

The 5' end of cDNA was obtained by the rapid amplification of cDNA ends (RACE) method as described in Lee et al. (2000b) based on Ausubel et al. (1995). An adaptor-ligated double-stranded cDNA library was constructed from poly $(A)^+$ RNA using a Marathon cDNA amplification kit (Clontech), and was subjected to PCR with a gene-specific primer corresponding to 5`-end portion of the initial *P. americana* Vg cDNA clone and the adaptor primer according to the supplier's protocol. The PCR products (shown in Fig. 2) were separated on 1.2% agarose gels and DNA bands were cloned into the TOPO TA cloning vector (Invitrogen). Clones were cycle sequenced as described above.

2.3.5 N-terminal Amino Acid Sequencing of Vns

Egg extracts (0.04-0.1 egg equivalent per lane) were subjected to SDS-PAGE as described above. The proteins thus separated were transferred to a PVDF membrane (Millipore, Immobilon) and stained lightly with Ponseau S. Bands corresponding to *P. americana* Vns were each cut out and analyzed using a gas phase amino acid sequencer (Perkin Elmer, 492 Procise).

2.3.6 Northern Blot Analysis

Total RNA was extracted from fat body cells of *P. americana* according to the methods described in Chomczynski and Sacchi (1987), and was separated (5 μg per lane) on 1% agarose gels containing formaldehyde and then transferred to Hybond N+ membranes (Amersham-Pharmacia). Probes used were short fragments (nucleotide positions 696-1497 and 4132-5081, respectively, see Figs. 3 and 4) of *P. americana* Vg cDNA. Fifty ng of the DNA fragment was labeled using a Gene image random prime labeling kit (Amersham-Pharmacia). Hybridization and subsequent washes were done under stringent conditions as described previously (Kageyama et al., 1994; Takadera et al., 1996; Nose et al., 1997). Reactions were detected using a Gene image CDP-star detection module (Amersham-Pharmacia) following the supplier's protocol.

2.3.7 Comparison of Vg Amino Acid Sequences and construction of a phylogenetic tree

The entire amino acid sequences excluding the signal peptide sequences were multiple-aligned using the Clustal W computer program (Thompson et al., 1994). A molecular phylogenetic (neighbor-joining) tree was constructed based on the multiple-alignment with 1000 bootstraps. All gap sites were ignored in the alignment and Kimura's correction was used.

2.4 RESULTS

2.4.1 Identification of *P. americana* Vns and Vgs

The egg (mature oocyte dissected from the adult female) extract and adult female and male hemolymph samples from the American cockroach, *Periplaneta americana*, were subjected to SDS-PAGE. The polypeptides thus separated were stained with Coomasie blue (Fig. 1A). Three major polypeptide bands (170 kD, 100 kD and 50 kD) were detected in the egg extract and female hemolymph but not in male hemolymph. The results obtained basically agree with those reported previously, except for the 100 kD band which, seen here as a single band, was reported to consist of multiple bands (Storella et al., 1985). In addition to these major bands, there was a minor female-specific polypeptide band with molecular mass of 150 kD.

Polyclonal antisera were raised against each of the major polypeptides and used for Western blot analysis. The results showed a complicated reaction pattern (Fig. 1B-D). The anti-*P. americana* 170 kD polypeptide antiserum reacted strongly with the 170 kD polypeptide and 100 kD multisubunits. The anti-*P. americana* 100 kD multisubunits antiserum reacted strongly with the 100 kD, and weakly with the 170 kD polypeptide. The anti-*P. americana* 50 kD polypeptide antiserum reacted strongly with the 50 kD and the minor 150 kD polypeptide. The results may reflect organization of Vn multiple subunits and post-translational processing. The 100 kD multisubunits were first detected immunologically in 4-day-old adult female hemolymph (data not shown).

2.4.2 Cloning and Sequencing of the cDNA for *P. americana* Vg

A cDNA expression library was constructed from poly $(A)^+$ RNA prepared from adult female (3- to 5-day-old) fat body cells and was immunologically screened with the anti-*P. americana* 100 kD multisubunits antiserum. After three successive screenings, we obtained five short positive clones (2.5-3.7 kb) and constructed the restriction maps for them. The nucleotide sequences of these five clones were partially determined at both the 5' and 3` ends (each about 1 kb). The lengths determined varied

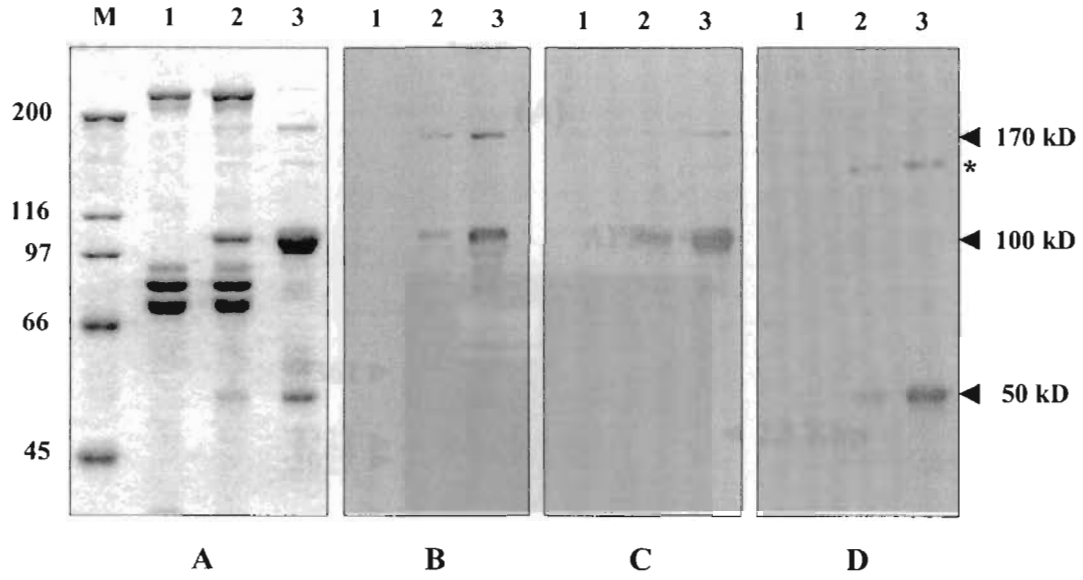


Fig. 1. Identification of *P. americana* Vns/Vgs by SDS-PAGE (**A**) and Western blotting (**B-D**). The SDS-PAGE gel was stained with Coomasie blue. **lanes 1**: hemolymph from 8-day-old adult males; **lanes 2**: hemolymph from 8-day-old adult females; **lanes 3**: egg extracts. Western blot analyses were conducted using antisera against *P. americana* 170 kD polypeptide (**B**), 100 kD multisubunits (**C**), and 50 kD polypeptide (**D**). The three major Vns/Vgs are indicated by arrowheads. The asterisk indicates a minor female-specific 150 kD polypeptide. Left-most lane (M) shows molecular mass markers (kD).

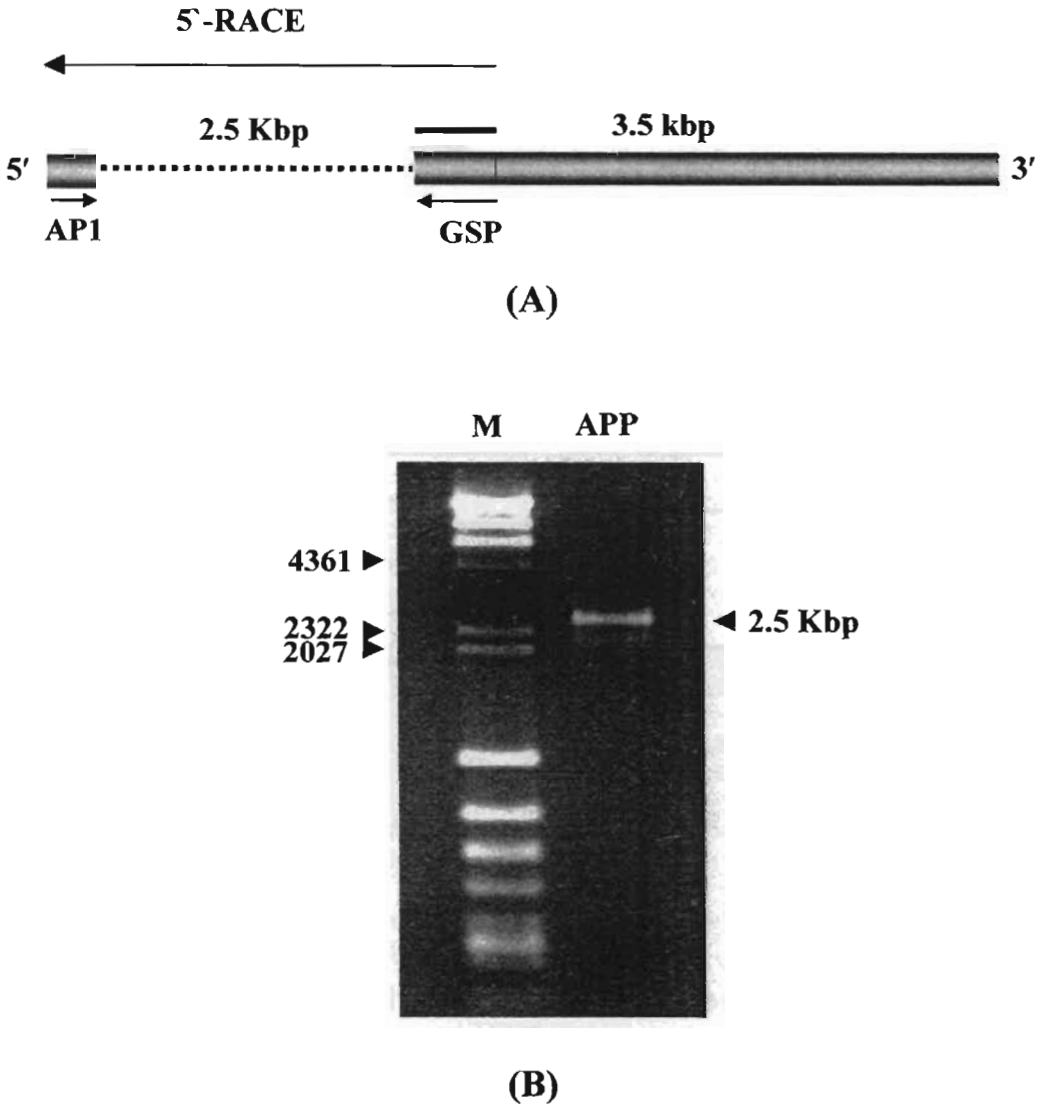


Fig. 2. 5′-end amplification of *P. americana* Vg cDNA. (A). A partial cDNA clone (3.5 kbp) has been shown on the 3` region; GSP, Gene specific primer (antisense); AP1, adapter primer; and a bold line is showing the overlapping region to identify the amplified portion (2.5 kbp). (B). M (lane-1) is molecular weight marker and APP (Lane-2) is the amplified PCR-prodct.

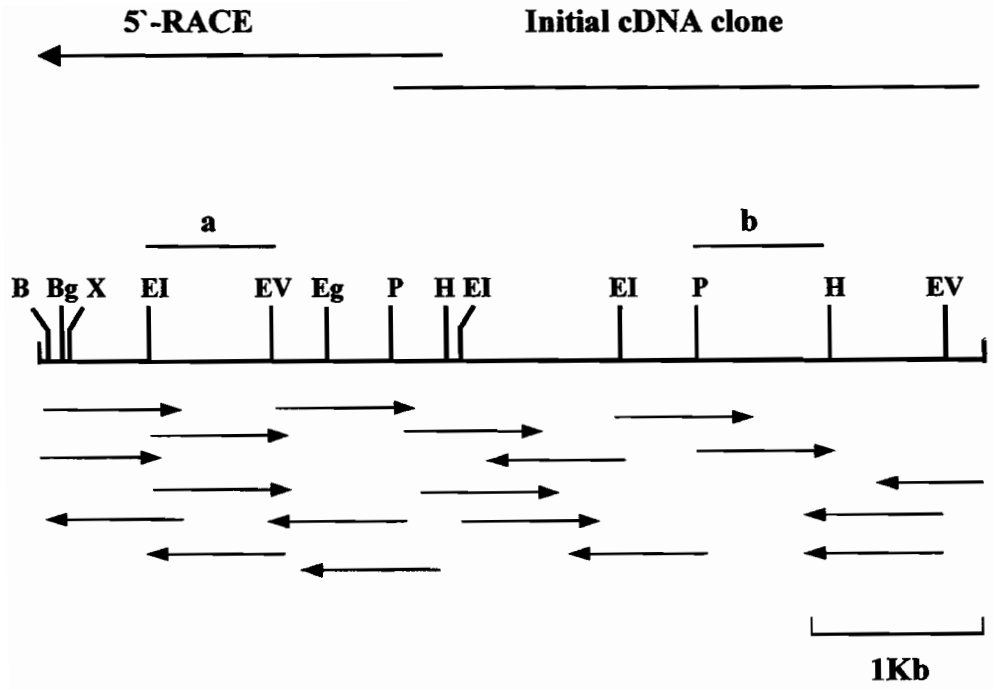
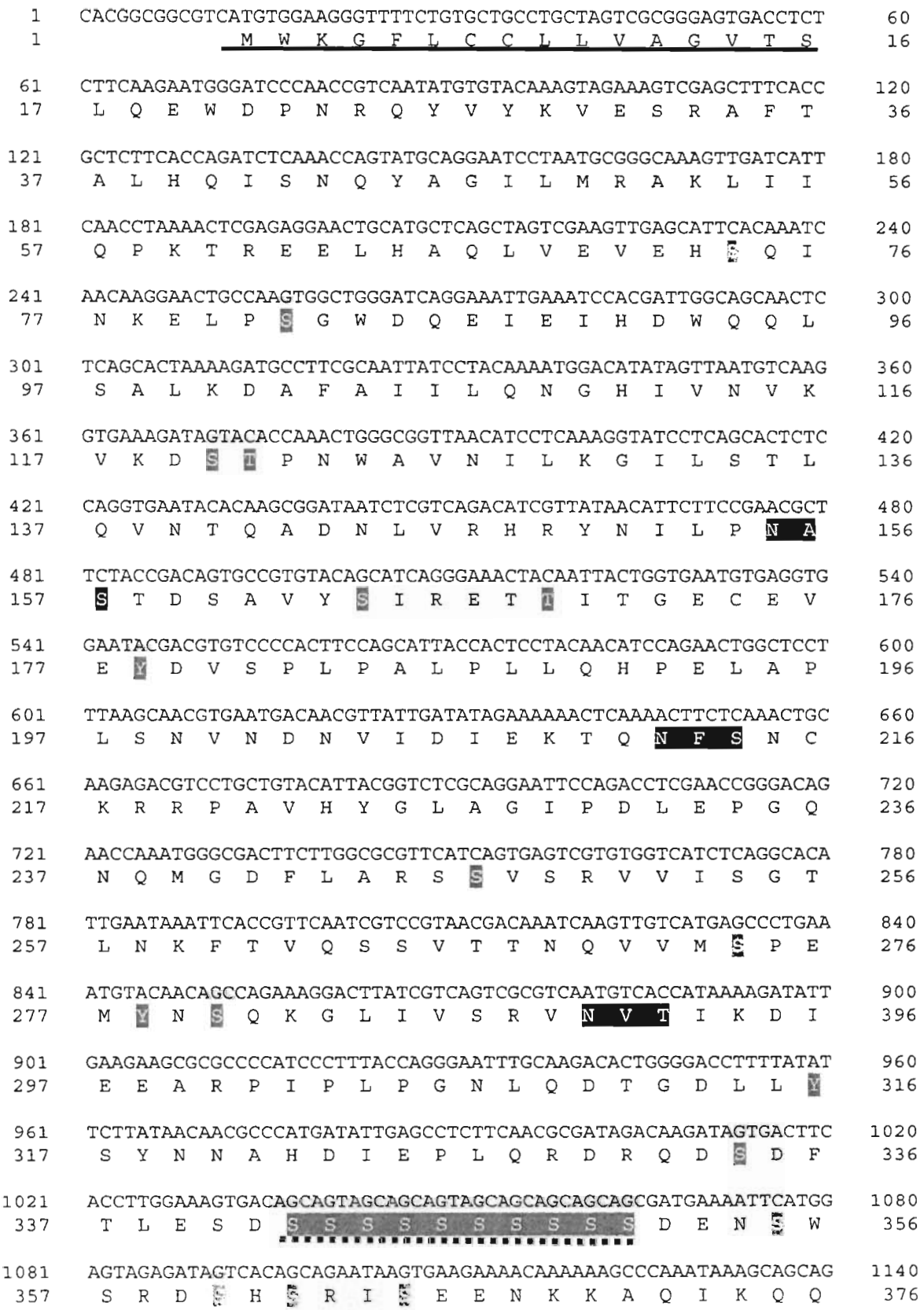


Fig. 3. Restriction map and sequence strategy of the cloned *P. americana* Vg cDNA and two probes (a and b) used for Northern blot analysis. B, *BamH*I; Bg, *Bgl*II; X, *Xho*I; EI, *EcoR*I; EV, *EcoR*V; Eg, *Eag*I; P, *Pst*I; H, *Hind*III.



1 CACGGCGGCGTCATGTGGAAGGGTTTTCTGTGCTGCCTGCTAGTCGCGGGAGTGACCTCT 60
1 M W K G F L C C L L V A G V T S 16

61 CTTCAAGAATGGGATCCCAACCGTCAATATGTGTACAAAGTAGAAAGTCGAGCTTTCACC 120
17 L Q E W D P N R Q Y V Y K V E S R A F T 36

121 GCTCTTCACCAGATCTCAAACCAGTATGCAGGAATCCTAATGCGGGCAAAGTTGATCATT 180
37 A L H Q I S N Q Y A G I L M R A K L I I 56

181 CAACCTAAAACTCGAGAGGAACTGCATGCTCAGCTAGTCGAAGTTGAGCATTCACAAATC 240
57 Q P K T R E E L H A Q L V E V E H S Q I 76

241 AACAAGGAACTGCCAAGTGGCTGGGATCAGGAAATTGAAATCCACGATTGGCAGCAACTC 300
77 N K E L P S G W D Q E I E I H D W Q Q L 96

301 TCAGCACTAAAAGATGCCTTCGCAATTATCCTACAAAATGGACATATAGTTAATGTCAAG 360
97 S A L K D A F A I I L Q N G H I V N V K 116

361 GTGAAAGATAGTACACCAAACTGGGCGGTTAACATCCTCAAAGGTATCCTCAGCACTCTC 420
117 V K D S T P N W A V N I L K G I L S T L 136

421 CAGGTGAATACACAAGCGGATAATCTCGTCAGACATCGTTATAACATTCTTCCGAACGCT 480
137 Q V N T Q A D N L V R H R Y N I L P N A 156

481 TCTACCGACAGTGCCGTGTACAGCATCAGGGAAACTACAATTACTGGTGAATGTGAGGTG 540
157 S T D S A V Y S I R E T T I T G E C E V 176

541 GAATACGACGTGTCCCCACTTCCAGCATTACCACTCCTACAACATCCAGAACTGGCTCCT 600
177 E Y D V S P L P A L P L L Q H P E L A P 196

601 TTAAGCAACGTGAATGACAACGTTATTGATATAGAAAAAACTCAAAACTTCTCAAACTGC 660
197 L S N V N D N V I D I E K T Q N F S N C 216

661 AAGAGACGTCCTGCTGTACATTACGGTCTCGCAGGAATTCCAGACCTCGAACCGGGACAG 720
217 K R R P A V H Y G L A G I P D L E P G Q 236

721 AACCAAATGGGCGACTTCTTGGCGCGTTCATCAGTGAGTCGTGTGGTCATCTCAGGCACA 780
237 N Q M G D F L A R S S V S R V V I S G T 256

781 TTGAATAAATTCACCGTTCAATCGTCCGTAACGACAAATCAAGTTGTCATGAGCCCTGAA 840
257 L N K F T V Q S S V T T N Q V V M S P E 276

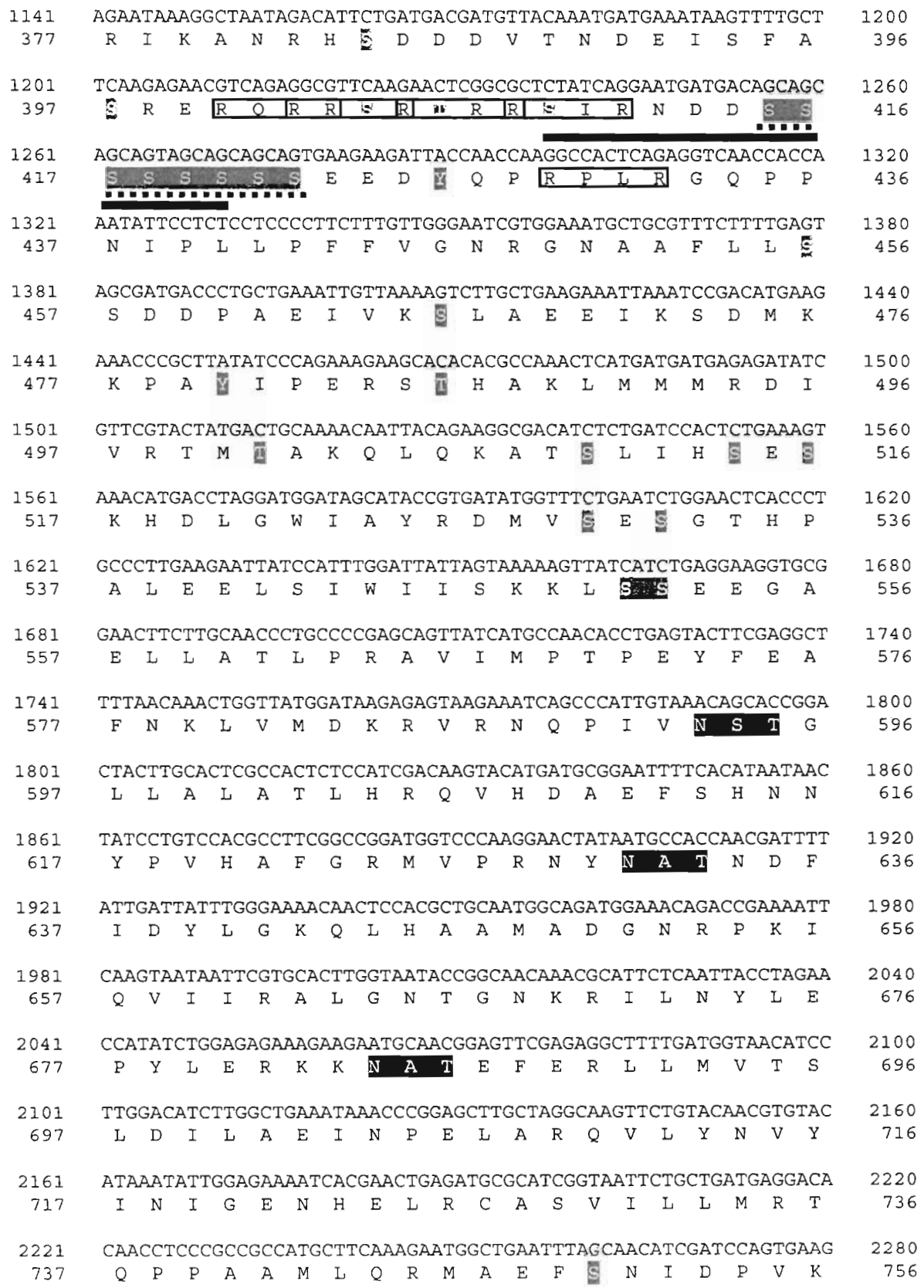
841 ATGTACAACAGCCAGAAAGGACTTATCGTCAGTCGCGTCAATGTCACCATAAAAGATATT 900
277 M Y N S Q K G L I V S R V N V T I K D I 396

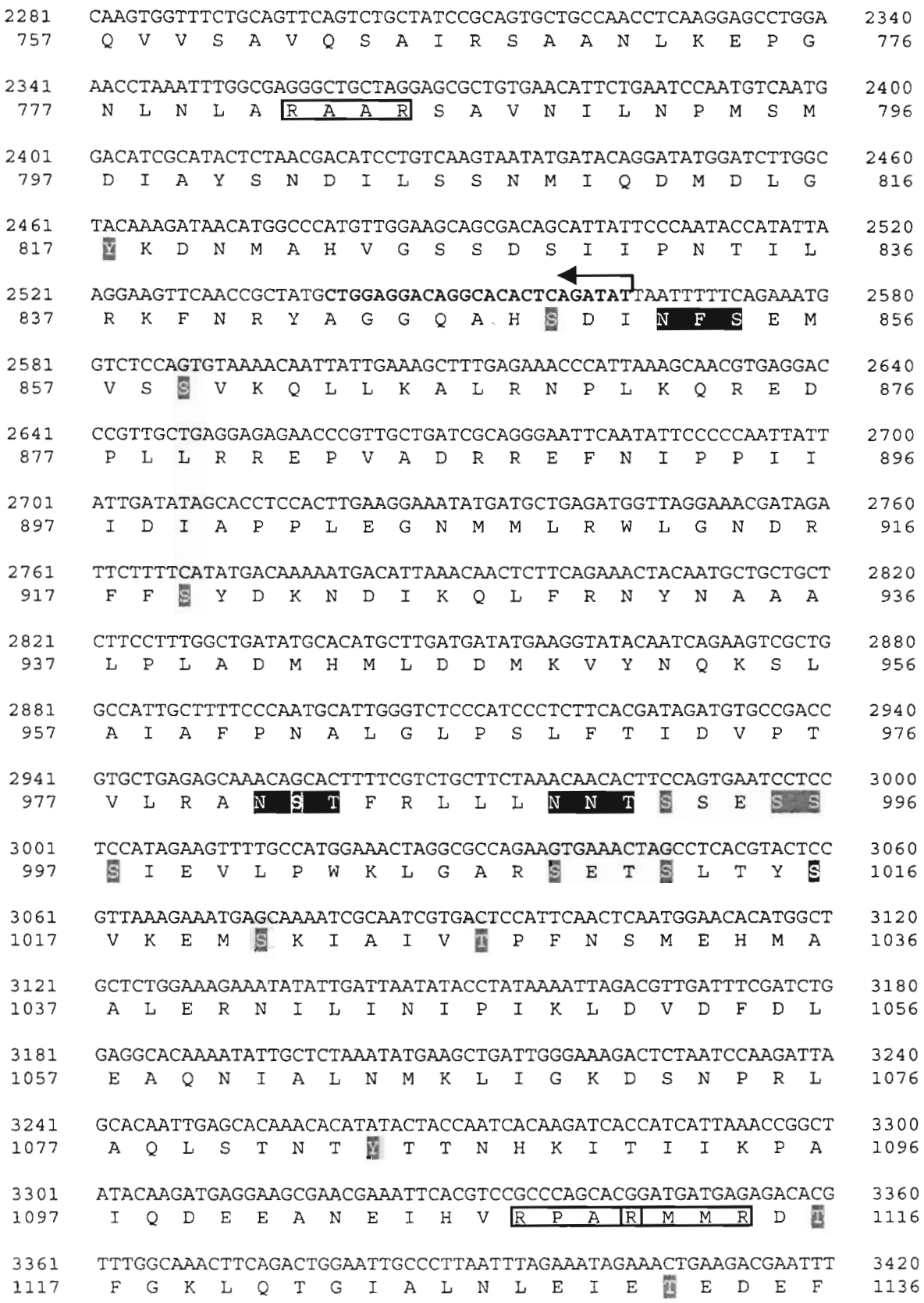
901 GAAGAAGCGCGCCCCATCCCTTTACCAGGGAATTTGCAAGACACTGGGGACCTTTTATAT 960
297 E E A R P I P L P G N L Q D T G D L L Y 316

961 TCTTATAACAACGCCCATGATATTGAGCCTCTTCAACGCGATAGACAAGATAGTGACTTC 1020
317 S Y N N A H D I E P L Q R D R Q D S D F 336

1021 ACCTTGGAAAGTGACAGCAGTAGCAGCAGTAGCAGCAGCAGCAGCGATGAAAATTCATGG 1080
337 T L E S D S S S S S S S S S S D E N S W 356

1081 AGTAGAGATAGTCACAGCAGAATAAGTGAAGAAAACAAAAAAGCCCAAATAAAGCAGCAG 1140
357 S R D S H S R I S E E N K K A Q I K Q Q 376





2281 CAAGTGGTTTCTGCAGTTCAGTCTGCTATCCGCAGTGCTGCCAACCTCAAGGAGCCTGGA 2340
757 Q V V S A V Q S A I R S A A N L K E P G 776

2341 AACCTAAATTTGGCGAGGGCTGCTAGGAGCGCTGTGAACATTCTGAATCCAATGTCAATG 2400
777 N L N L A R A A R S A V N I L N P M S M 796

2401 GACATCGCATACTCTAACGACATCCTGTCAAGTAATATGATACAGGATATGGATCTTGGC 2460
797 D I A Y S N D I L S S N M I Q D M D L G 816

2461 TACAAAGATAACATGGCCCATGTTGGAAGCAGCGACAGCATTATTCCCAATACCATATTA 2520
817 Y K D N M A H V G S S D S I I P N T I L 836

2521 AGGAAGTTCAACCGCTATGCTGGAGGACAGGCACACTCAGATATTAATTTTTCAGAAATG 2580
837 R K F N R Y A G G Q A H S D I N F S E M 856

2581 GTCTCCAGTGTAAAACAATTATTGAAAGCTTTGAGAAACCCATTAAAGCAACGTGAGGAC 2640
857 V S S V K Q L L K A L R N P L K Q R E D 876

2641 CCGTTGCTGAGGAGAGAACCCGTTGCTGATCGCAGGGAATTCAATATTCCCCCAATTATT 2700
877 P L L R R E P V A D R R E F N I P P I I 896

2701 ATTGATATAGCACCTCCACTTGAAGGAAATATGATGCTGAGATGGTTAGGAAACGATAGA 2760
897 I D I A P P L E G N M M L R W L G N D R 916

2761 TTCTTTTCATATGACAAAAATGACATTAAACAACTCTTCAGAAACTACAATGCTGCTGCT 2820
917 F F S Y D K N D I K Q L F R N Y N A A A 936

2821 CTTCCTTTGGCTGATATGCACATGCTTGATGATATGAAGGTATACAATCAGAAGTCGCTG 2880
937 L P L A D M H M L D D M K V Y N Q K S L 956

2881 GCCATTGCTTTTCCCAATGCATTGGGTCTCCCATCCCTCTTCACGATAGATGTGCCGACC 2940
957 A I A F P N A L G L P S L F T I D V P T 976

2941 GTGCTGAGAGCAAACAGCACTTTTCGTCTGCTTCTAAACAACACTTCCAGTGAATCCTCC 3000
977 V L R A N S T F R L L L N N T S S E S S 996

3001 TCCATAGAAGTTTTGCCATGGAAACTAGGCGCCAGAAGTGAAACTAGCCTCACGTACTCC 3060
997 S I E V L P W K L G A R S E T S L T Y S 1016

3061 GTTAAAGAAATGAGCAAAATCGCAATCGTGACTCCATTCAACTCAATGGAACACATGGCT 3120
1017 V K E M S K I A I V T P F N S M E H M A 1036

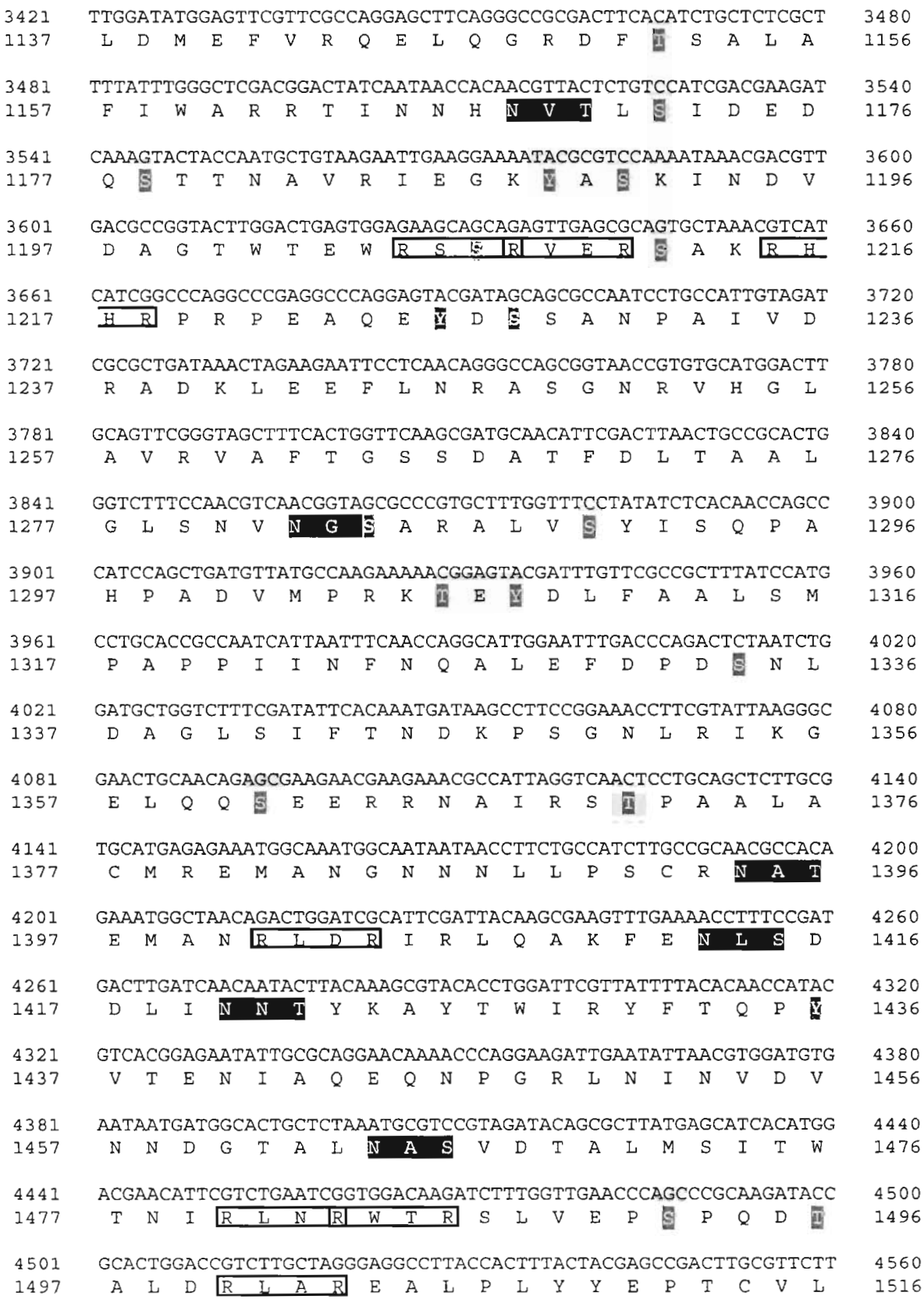
3121 GCTCTGGAAAGAAATATATTGATTAATATACCTATAAAATTAGACGTTGATTTCGATCTG 3180
1037 A L E R N I L I N I P I K L D V D F D L 1056

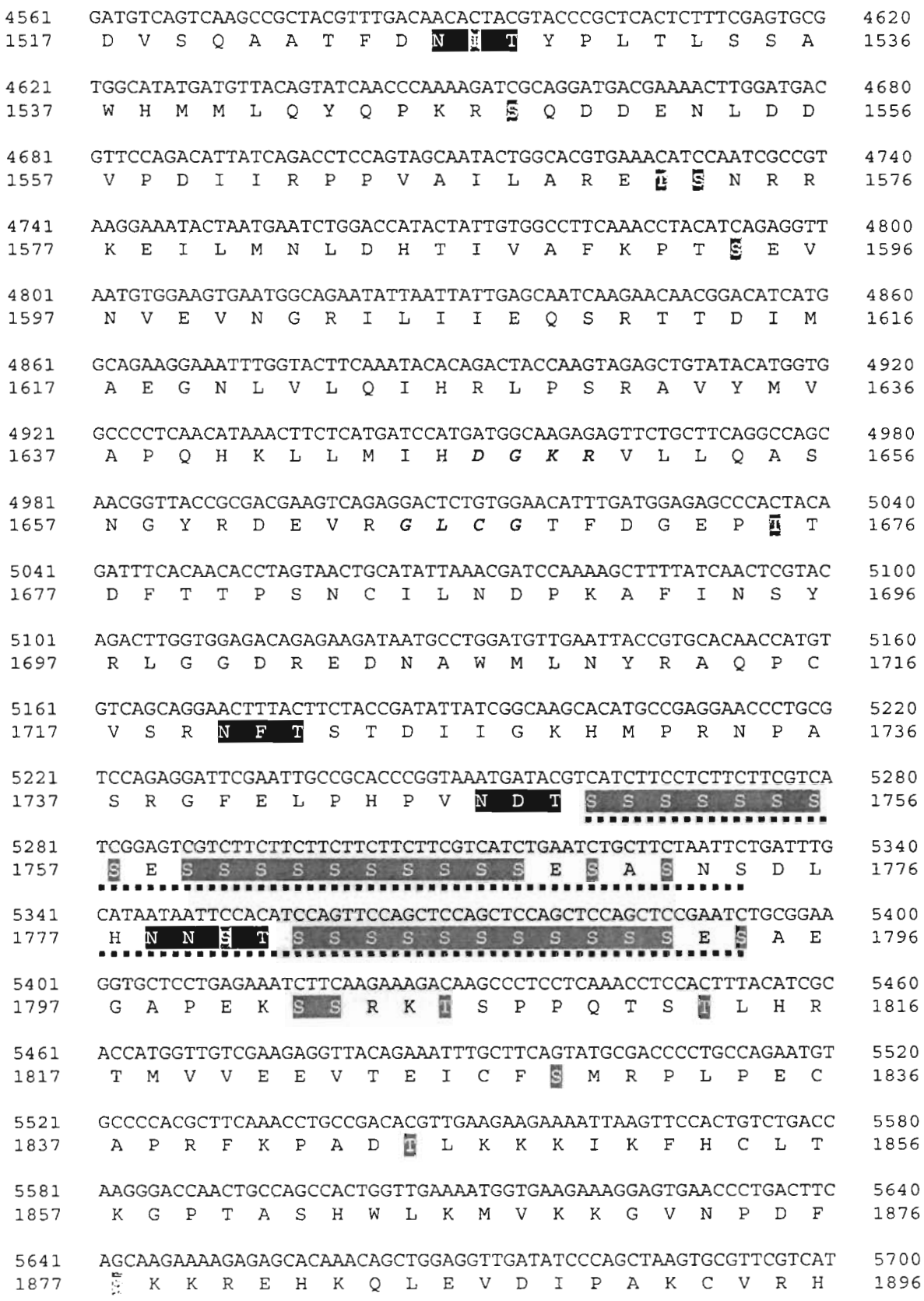
3181 GAGGCACAAAATATTGCTCTAAATATGAAGCTGATTGGGAAAGACTCTAATCCAAGATTA 3240
1057 E A Q N I A L N M K L I G K D S N P R L 1076

3241 GCACAATTGAGCACAAACACATATACTACCAATCACAAGATCACCATCATTAAACCGGCT 3300
1077 A Q L S T N T Y T T N H K I T I I K P A 1096

3301 ATACAAGATGAGGAAGCGAACGAAATTCACGTCCGCCCAGCACGGATGATGAGAGACACG 3360
1097 I Q D E E A N E I H V R P A R M M R D T 1116

3361 TTTGGCAAACTTCAGACTGGAATTGCCCTTAATTTAGAAATAGAAACTGAAGACGAATTT 3420
1117 F G K L Q T G I A L N L E I E T E D E F 1136


3421 TTGGATATGGAGTTCGTTCGCCAGGAGCTTCAGGGCCGCGACTTCACATCTGCTCTCGCT 3480
1137 L D M E F V R Q E L Q G R D F I S A L A 1156
3481 TTTATTTGGGCTCGACGGACTATCAATAACCACAACGTTACTCTGTCCATCGACGAAGAT 3540
1157 F I W A R R T I N N H N V T L S I D E D 1176
3541 CAAAGTACTACCAATGCTGTAAGAATTGAAGGAAAATACGCGTCCAAAATAAACGACGTT 3600
1177 Q S T T N A V R I E G K Y A S K I N D V 1196
3601 GACGCCGGTACTTGGACTGAGTGGAGAAGCAGCAGAGTTGAGCGCAGTGCTAAACGTCAT 3660
1197 D A G T W T E W R S S R V E R S A K R H 1216
3661 CATCGGCCCAGGCCCGAGGCCCAGGAGTACGATAGCAGCGCCAATCCTGCCATTGTAGAT 3720
1217 H R P R P E A Q E Y D S S A N P A I V D 1236
3721 CGCGCTGATAAACTAGAAGAATTCCTCAACAGGGCCAGCGGTAACCGTGTGCATGGACTT 3780
1237 R A D K L E E F L N R A S G N R V H G L 1256
3781 GCAGTTCGGGTAGCTTTCACTGGTTCAAGCGATGCAACATTCGACTTAACTGCCGCACTG 3840
1257 A V R V A F T G S S D A T F D L T A A L 1276
3841 GGTCTTTCCAACGTCAACGGTAGCGCCCGTGCTTTGGTTTCCTATATCTCACAACCAGCC 3900
1277 G L S N V N G S A R A L V S Y I S Q P A 1296
3901 CATCCAGCTGATGTTATGCCAAGAAAAACGGAGTACGATTTGTTCGCCGCTTTATCCATG 3960
1297 H P A D V M P R K T E Y D L F A A L S M 1316
3961 CCTGCACCGCCAATCATTAATTTCAACCAGGCATTGGAATTTGACCCAGACTCTAATCTG 4020
1317 P A P P I I N F N Q A L E F D P D S N L 1336
4021 GATGCTGGTCTTTCGATATTCACAAATGATAAGCCTTCCGGAAACCTTCGTATTAAGGGC 4080
1337 D A G L S I F T N D K P S G N L R I K G 1356
4081 GAACTGCAACAGAGCGAAGAACGAAGAAACGCCATTAGGTCAACTCCTGCAGCTCTTGCG 4140
1357 E L Q Q S E E R R N A I R S T P A A L A 1376
4141 TGCATGAGAGAAATGGCAAATGGCAATAATAACCTTCTGCCATCTTGCCGCAACGCCACA 4200
1377 C M R E M A N G N N N L L P S C R N A T 1396
4201 GAAATGGCTAACAGACTGGATCGCATTCGATTACAAGCGAAGTTTGAAAACCTTTCCGAT 4260
1397 E M A N R L D R I R L Q A K F E N L S D 1416
4261 GACTTGATCAACAATACTTACAAAGCGTACACCTGGATTCGTTATTTTACACAACCATAC 4320
1417 D L I N N T Y K A Y T W I R Y F T Q P Y 1436
4321 GTCACGGAGAATATTGCGCAGGAACAAAACCCAGGAAGATTGAATATTAACGTGGATGTG 4380
1437 V T E N I A Q E Q N P G R L N I N V D V 1456
4381 AATAATGATGGCACTGCTCTAAATGCGTCCGTAGATACAGCGCTTATGAGCATCACATGG 4440
1457 N N D G T A L N A S V D T A L M S I T W 1476
4441 ACGAACATTCGTCTGAATCGGTGGACAAGATCTTTGGTTGAACCCAGCCCGCAAGATACC 4500
1477 T N I R L N R W T R S L V E P S P Q D I 1496
4501 GCACTGGACCGTCTTGCTAGGGAGGCCTTACCACTTTACTACGAGCCGACTTGCGTTCTT 4560
1497 A L D R L A R E A L P L Y Y E P T C V L 1516



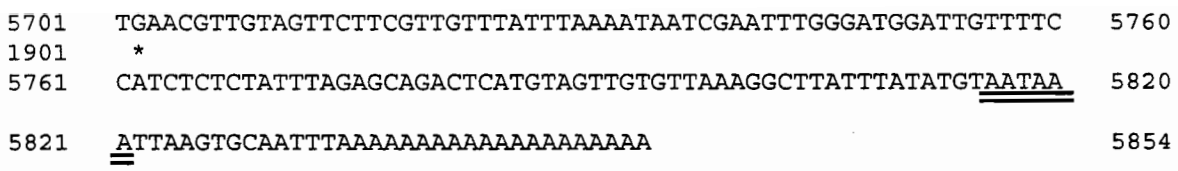


Fig. 4. Nucleotide sequence and deduced amino acid sequence of the Vg cDNA of *P. americana* (GenBank accession number: AB034804). The putative signal peptide is underlined. A consensus polyadenylation signal is double-underlined. The determined 12 N-terminal amino acid residues for 170 kD polypeptide are shown with a bold line. Amino acid sequences with a consensus RXXR for possible cleavage sites are boxed. Clusters of serine residues are underlined with dotted line. Gene specific primer (GSP) is shown with boldface-letters, whereas arrow indicates the 5`-end portion amplified. Possible glycosylation sites (having consensus sequence: NXS/T) are shown with dark-shaded boxes. The potential phosphorylated serine (S), threonine (T) and tyrosine (Y) residues [predicted by using the Net Phos 2.0 computer program (Blom et al., 1999)] are indicated with light-shaded frames. The GLI/CG and DGXR motifs are shown with bold-face-Italic letters. Whereas, asterik indicates the stop codon.

but the sequences were identical among the clones except for a very few polymorphic base substitutions. The longest clone in terms of the 5` end was selected and the sequence was determined. The clone was 3502 bp long, and encoded 1116 amino acids in a single ORF followed by a termination codon, a consensus polyadenylation signal (AATAAA) and a poly A tract (Fig. 4). The deduced amino acid sequence aligned well with the known insect Vg sequences. We thus concluded that the clone was a 3` portion of the *P. americana* Vg cDNA.

The remaining 5` end portion was cloned using the 5`-RACE method. A double-stranded cDNA library was prepared and subjected to PCR with a gene-specific primer (nucleotide positions 2540-2564 in the final clone, Fig. 4) and an adaptor primer. Four clones (each about 2.5 kb in length) were selected, restriction maps were constructed and sequenced. Fig. 3 shows the final restriction map following subcloning and the sequencing strategy. The sequences of four clones matched except for a very few polymorphic base substitutions. Combining the initial clone and the longest 5` end clone provided the sequence of 5854 bp, encoding 1896 amino acids (see Fig. 4).

2.4.3 Structural Analysis of *P. americana* Vg

The common characteristics of insect Vg amino acid sequences were also found in the *P. americana* Vg amino acid sequence (Figs. 4 and 5). Fourteen consensus post-translational cleavage sites, RXXR, a conserved GL/ICG motif (amino acid positions 1665-1668) and five cysteine residues that follow at conserved locations near the C-terminal were present (see Fig. 4). It was noted that a DGXR motif, starting 18 residues upstream of the GL/ICG motif, was conserved among insect Vgs. There was a serine-rich stretch at the C-terminal region, a characteristic which, except for in cockroaches, has been observed only in the mosquito, *Aedes aegypti*.

Twelve N-terminal amino acid residues chemically determined for the 170 kD polypeptide (SIRNDDSSSSSS) matched exactly the deduced amino acid sequence following one of the RXXR cleavage sites (RTRR, amino acid positions 405-408) (Fig.

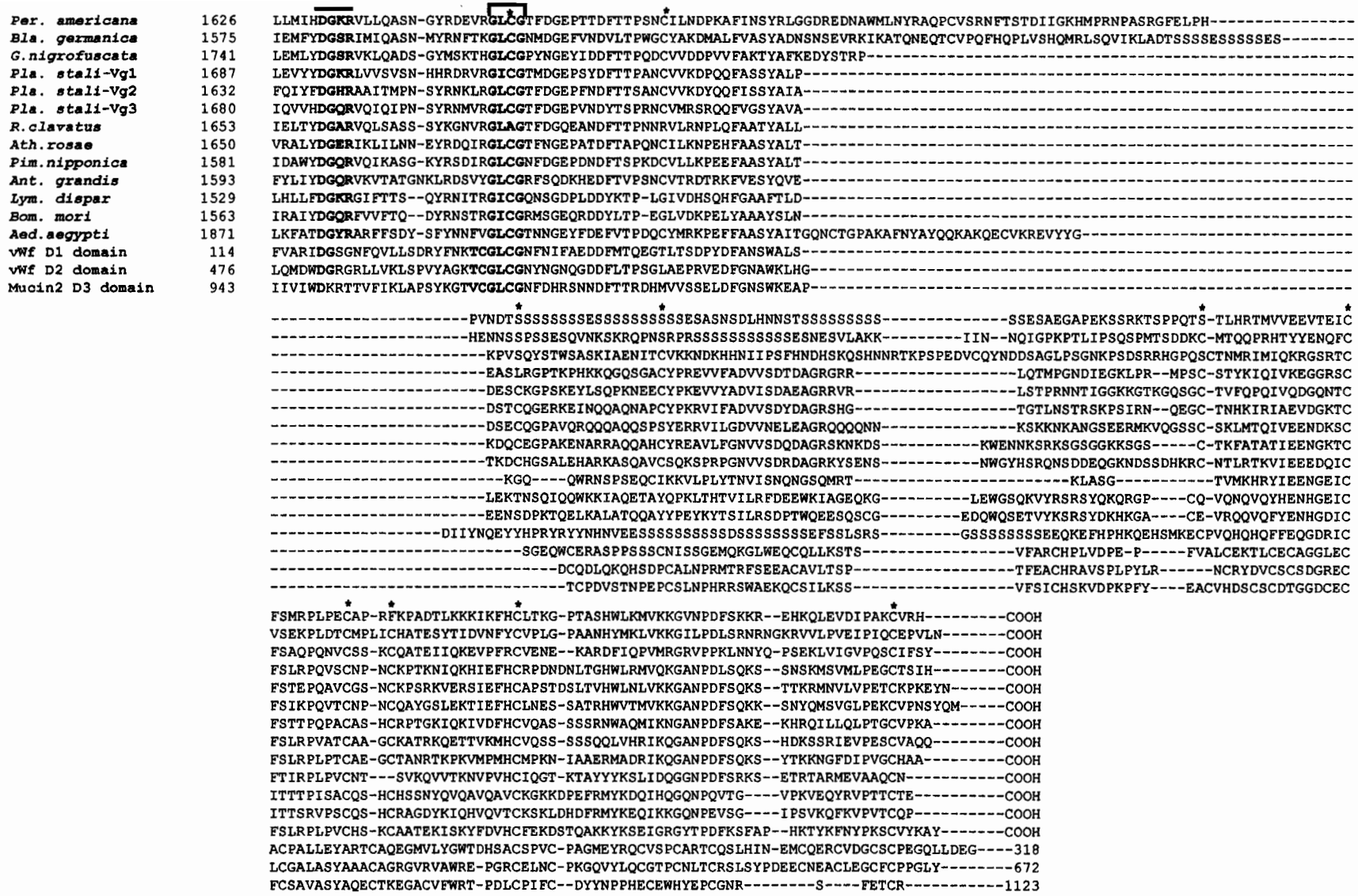


Fig. 5. Alignment of C-terminal portion of insect Vg amino acid sequences, muc2 D3 domain, human vWf D1 and D2 domains. Alignment was done using the Clustal W program. The overline indicates the DGXR motif in insects (corresponding DG residues are seen in muc2 and vWf). The bracket indicates the GL/ICG motif in insects (corresponding TGLCG motif is present in muc2 and vWf). Asterisks show cysteine residues at conserved locations. Numbers are amino acid positions relative to the translation initiation methionine (excluding the signal peptides)

4). We could not, however, chemically determine the N-terminal amino acid sequences for the 100 kD and 50 kD polypeptides.

2.4.4 Expression of the *P. americana* Vg Gene

Northern blot analysis was conducted to determine when the Vg gene was expressed in *P. americana* (Fig. 6). Total RNAs were extracted from adult female (1-, 2-, 3- and 5-day-old) and male (1- and 5-day-old) fat body cells, and ovaries from 5-day-old adult females. Two short fragments of *P. americana* Vg cDNA were used as probes, and essentially the same results were obtained. The transcript of the *P. americana* Vg gene was detected as a single 6.5-kb band. The gene was expressed in the fat body cells of adult females at 2-days posteclosion and thereafter. The gene expression was not detected in 1-day-old adult females, in males, or in ovaries.

2.4.5 Similarities in Vn Antigenicity Among Cockroaches

Egg extracts prepared from each of the ten cockroach species: five of the superfamily Blattoidea (*P. americana*, *P. brunnea*, *P. japonica*, *P. fuliginosa* and *Blatta lateralis*) and five of the superfamily Blaberoidea (*Blabarus sp.*, *Pycnoscelus surinamensis*, *Leucophaea maderae*, *Nauphoeta cinerea* and *Asiablatta kyotensis*), were subjected to SDS-PAGE followed by Western blot analyses using the antisera raised against each of the *P. americana* Vns (Fig. 7). The major yolk proteins separated on the SDS-PAGE gel consisted of one with a large molecular mass (about 170 kD), some with mid-sized molecular masses (90-110 kD) and a few with small molecular masses (50-60 kD) in the species of Blattoidea. On the other hand, the large molecular mass polypeptide was absent in the species of Blaberoidea. Anti-*P. americana* 100 kD polypeptide antiserum reacted with the large and the mid-sized molecular mass polypeptides in the members of Blattoidea. The antiserum against *P. americana* 50 kD polypeptides reacted with the small molecular mass polypeptides of all species belonging to the Blattoidea. Both antisera did not react with the polypeptides of the

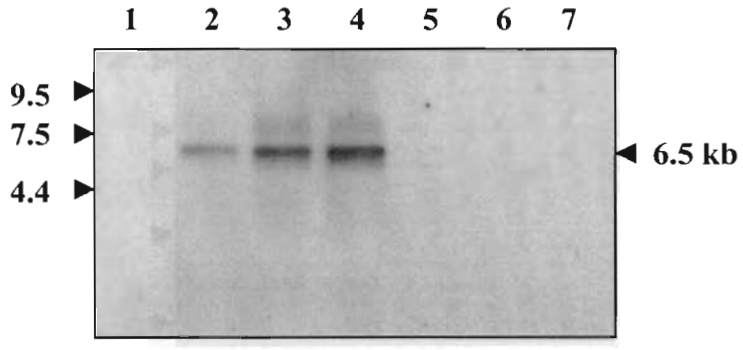


Fig. 6. *P. americana* Vg gene expression detected by Northern blot analysis. The results obtained using a short fragment of *P. americana* Vg cDNA (nucleotide positions 696-1497) as a probe are shown. Total RNAs were prepared from fat body cells of 1-, 2-, 3- and 5-day-old adult females (**lanes 1-4**, respectively), of 1- and 5-day-old adult males (**lanes 5** and **6**, respectively), and from ovaries of 5-day-old adult females (**lane 7**). The mobility of a standard size marker (kb) is shown on the left.

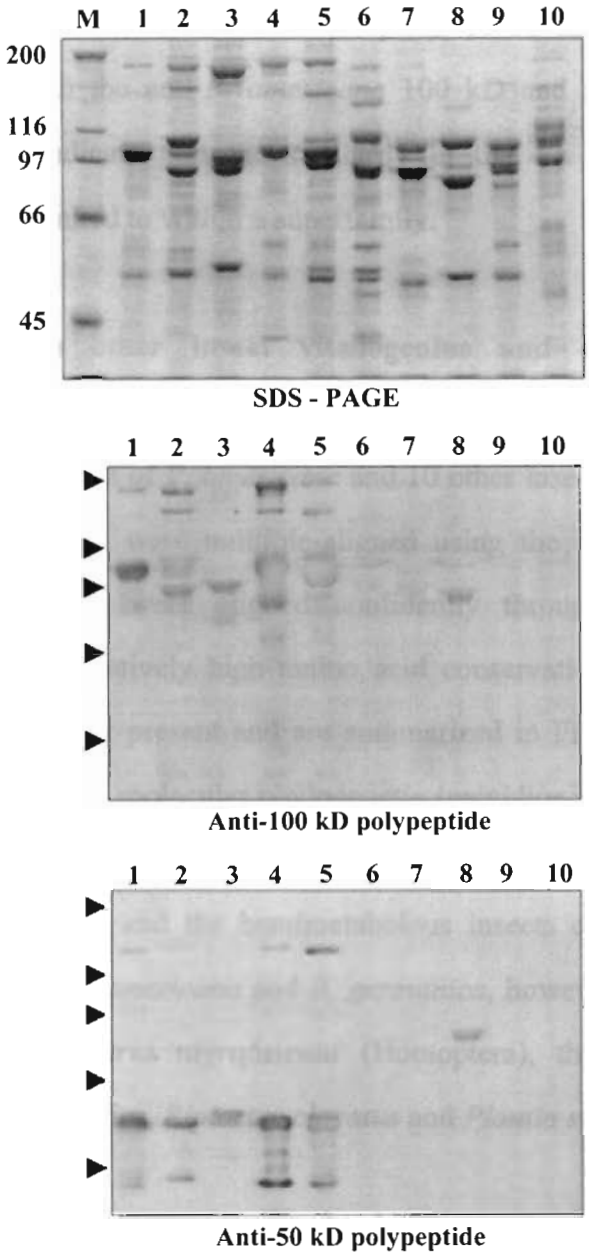


Fig. 7. Similarities in Vn antigenicity among ten cockroaches. An SDS-PAGE gel was stained with Coomasie blue (**Top**). Western blot analyses were performed using anti-*P. americana* 100 kD multisubunits (**Middle**) and anti- *P. americana* 50 kD polypeptide (**Bottom**). Lanes 1-5 represent species of Blattoidea: *P. americana* (**lanes 1**), *P. brunnea* (**lanes 2**), *P. japonica* (**lanes 3**), *Blatta lateralis* (**lanes 4**), and *P. fuliginosa* (**lanes 5**). Lanes 6-10 represent species of Blaberoidea: *Blabarus sp.* (**lanes 6**) *Pycnoscelus surinamensis* (**lanes 7**), *Leucophaea maderae* (**lanes 8**), *Nauphoeta cinerea* (**lanes 9**), and *Asiablatta kyotensis* (**lanes 10**). The left-most lane (M) shows molecular mass markers (kD).

species belonging to Blaberoidea, with one exception. A 90 kD polypeptide in *L. maderae* reacted with both the anti-*P. americana* 100 kD and 50 kD polypeptides antisera. These results indicate that the similarity in the Vn antigenicity among cockroaches is basically limited to within a superfamily.

2.4.6 Comparison with other insect vitellogenins and construction of a phylogenetic tree

The amino acid sequences of *P. americana* and 10 other insect species (excluding the signal peptide sequences) were multiple-aligned using the Clustal W computer program. All the sequences were aligned confidently throughout their lengths. Subdomains I-V, areas of relatively high amino acid conservation, previously noted (Chen et al., 1997) are clearly present and are summarized in Fig. 8. The Clustal W computer program also gave a molecular phylogenetic (neighbor-joining) tree and was found to agree basically with the accepted tree based on morphological characteristics (Fig. 9). The holometabolous and the hemimetabolous insects clustered as separate groups. Two cockroaches, *P. americana* and *B. germanica*, however, formed a cluster with the cicada, *Graptopsaltria nigrofuscata* (Homoptera), the latter thus being separated from the related species, *Riptortus clavatus* and *Plautia stali* (Heteroptera).

2.5 DISCUSSION

The synthesis of insect Vgs and their uptake to developing oocytes have been used as an index of hormonal activity (Wyatt, 1991; Wyatt and Davey, 1996). It was shown in the German cockroach, *Blattella germanica* that a cloned partial Vg cDNA was a powerful tool to study the molecular action of the regulators of vitellogenesis (Comas et al., 1999). In the present study we obtained a complete Vg cDNA of another model species, *Periplaneta americana*. The Vg gene expression was first detected by Northern blot analysis in the fat body cells of 2-day-old adult females, and the

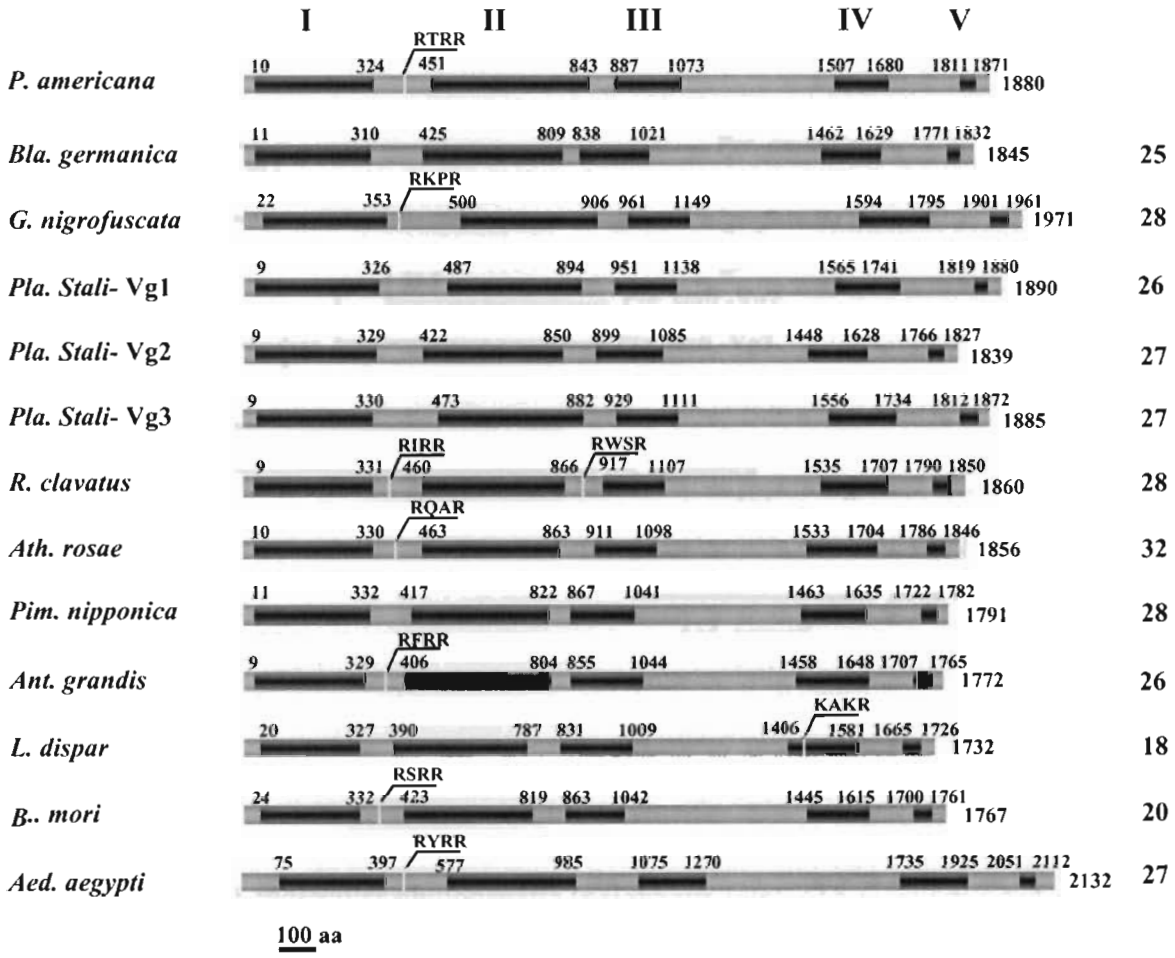


Fig. 8. Comparison of protein primary structures of 11 insect Vgs. Numbers indicate amino acid positions from the N-terminua (excluding the signal peptides). Subdomains I-V (hatched areas) are areas of relatively high amino acid conservation (Chen et al., 1997). Bold letters raised and underlined are basic amino acid sequences representing the dibasic protease recognition sites. Identical amino acid ratios (%) compared to *P. americana* sequence are shown on the right side.

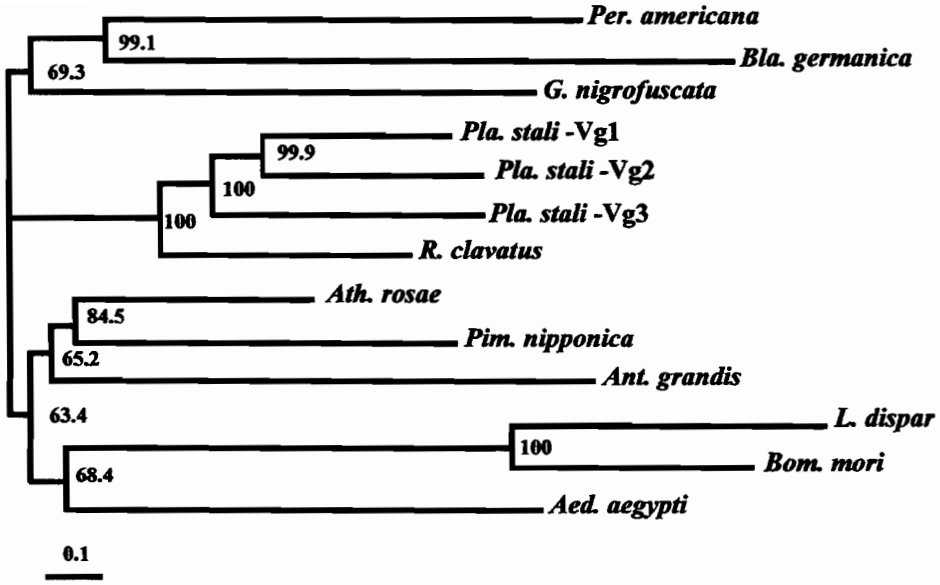


Fig. 9. A molecular phylogenetic (neighbour-joining) tree constructed based on the entire Vg amino acid sequences of 11 insects. Bootstrap values in 1000 replicates are shown as percentage. The scale indicates distance (number of amino acid substitutions per site).

hemolymph Vgs were first detected by Western blot analysis in 4-day-old adult females. The results demonstrate that the *P. americana* Vg cDNA clone will also be a sensitive probe to study hormonal regulation on Vg gene expression. A complete Vg cDNA clone has a further advantage. The neighboring 5' regulatory region in the genomic DNA should be obtained easily by the primer extension method using sequence-specific primers corresponding to a portion near the 5' terminal of the complete Vg cDNA. In contrast, considerably more effort would have to be exerted to obtain the regulatory region by using the partial Vg cDNAs, which generally represent only the 3' portion.

Previous reports showed the Vns of *P. americana* comprised two polypeptides (Vn1/Vg1 and Vn2/Vg2) (Bell, 1970; Engelmann, 1979; Storella et al., 1985; Kim and Lee, 1994). Vn1/Vg1 is composed of four major polypeptides (170, 105, 72 and 78 kD) and Vn2/Vg2 of three major polypeptides (105, 101 and 60 kD). A hypothesis that post-translational processing occurs after incorporation of Vgs into developing oocytes has been proposed (Storella et al., 1985). The complicated reactivities of the antisera against *P. americana* Vns with the polypeptides in egg extracts and hemolymph thus were thought to reflect such post-translational processing. The fact that 14 consensus cleavage sites (RXXR) exist in the *P. americana* Vg amino acid sequence suggests that at least some of those sites are actually used in the post-translational processing. The N-terminal amino acid sequences were determined chemically only for the 170 kD polypeptide and not for the 100 and 50 kD polypeptides. The latter two, each seemingly a single polypeptide band on SDS-PAGE, most probably consist of multiple subunits due to post-translational processing and hence the N-terminal amino acid sequences could not be determined.

In the present study, we obtained only one Vg cDNA molecular species through screening of the cDNA expression library with the antiserum against the 100 kD polypeptide, which is derived from multiple subunits both for Vn1 and Vn2 (Storella et al., 1985). This would be a chance product that would be formed when the conventional method was employed to clone insect Vg cDNAs. Recently Lee et al. (2000a)

developed a method to clone multiple Vg cDNA species simply and rapidly by using PCR. The method should also be applicable to *P. americana* to obtain the other Vg cDNA(s).

The entire (deduced) amino acid sequence of *P. americana* Vg showed extensive similarity with the known insect Vg sequences, showing the common features such as consensus RXXR cleavage sites, the GL/ICG motif and nine cysteine residues that follow at conserved locations. In addition, we noted a DGXR motif located at 17-19 residues upstream of the GL/ICG motif that is conserved in all insect Vg amino acid sequences (Fig. 5). Similar sequences were also found in human von Willebrand factors (vWf) (Baker, 1988) and mucin 2 glycoprotein (Muc 2) (Gum et al., 1994), both of which are members of the Vg gene superfamily. The GL/ICG motif and cysteine residues were necessary for oligomerization of vertebrate Vns (Mayadas and Wagner, 1992; Mouchel et al., 1996). It was also demonstrated that the DG residues located near the cysteine-rich repeat contributed to establish appropriate folding of the LDL receptor protein (Djordjevic et al., 1996). One possible function of insect Vns oligomerized in eggs would be binding lipids, in which inactive ecdysteroids are enclosed. The ecdysteroids are then released as Vns undergo proteolysis during embryogenesis (Hagedorn et al., 1998; Giorgi et al., 1999). The DG residues together with GL/ICG motif and the following cysteine residues might take part in forming the structure necessary for Vns to function properly during embryogenesis.

It has been reported that there were similarities of Vn/Vg among insects not only in the primary structure, but also in the antigenicity. The antigenic similarity of Vns existed beyond the family level among 21 symphytan hymenopteran species (Takadera et al., 1996). In the pyralid moths, however, the antigenic similarity was limited to within the family level (Shirk, 1987). The present results have shown that Vn antigenicity was limited to within the superfamily in cockroaches. The only exception was *Leucophaea maderae* (Blaberoidea), of which the major Vn (90 kD) was antigenically related to that of *P. americana* (Blattoidea), and thus the similarity

extended in this case beyond the superfamily level. The evolutionary relatedness between these two species is thus of considerable interest.

A molecular phylogenetic tree constructed based on the entire Vg amino acid sequences of 11 insects was basically in agreement with that constructed based on comparative morphology. It is noted, however, that one cluster consisted of cockroaches and a cicada (Fig. 9). The 11 species belong to Polyneoptera, Paraneoptera and Oligoneoptera. To determine whether the Vg amino acid sequences can be a supplementary molecular marker reflecting the insect phylogenetic tree, we must obtain many more clones of Vg cDNAs from other species, particularly those from more primitive species belonging to the Paleoptera (such as mayflies and dragonflies) and the Apterygota (such as silver-fishes and bristletails).

2.6 LITERATURE CITED

Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, Struhl K., 1995. Short protocols in molecular biology. 3rd ed. New York: Wiley.

Baker ME., 1988. Invertebrate vitellogenin is homologous to human von Willebrand factor. J Biochem 256: 1059-1061.

Bell WJ., 1970. Demonstration and characterization of two vitellogenic blood proteins in *Periplaneta americana*: an immunochemical analysis. J Insect Physiol 16: 291-299.

Chen JS, Cho WL, Raikhel AS., 1994. Analysis of mosquito vitellogenin cDNA: similarity with vertebrate phosvitins and arthropod serum proteins. J Mol Biol 237: 641-647.

Chen JS, Sappington TW, Raikhel AS., 1997. Extensive sequence conservation among insect, nematode and vertebrate vitellogenins reveals ancient common ancestry. J Mol Evol 44: 440-451.

Chomczynski P, Sacchi N., 1987. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal Biochem 162: 156-159.

Comas D, Piulachs MD, Belles X., 1999. Fast induction of vitellogenin gene expression by juvenile hormone III in the cockroach *Blattella germanica* (L.) (Dictyoptera, Blattellidae). Insect Biochem Mol Biol 29: 821-827.

Comas D, Piulachs MD, Belles X., 2000. The vitellogenin of *Blattella germanica* (L.) (Dictyoptera, Blattellidae): nucleotide sequence of the cDNA and analysis of the protein

structure. EMBL/DDBJ/GenBank Accession number AJ005115.

Djordjevic JT, Bieri S, Smith R, Kroon PA., 1996. A deletion in the first cysteine-rich repeat of the low-density-lipoprotein receptor leads to the formation of multiple misfolded isomers. Eur J Biochem 239: 214-219.

Engelmann F., 1979. Insect vitellogenin: identification, biosynthesis, and role in vitellogenesis. Adv Insect Physiol 14: 49-108.

Giorgi F, Bradley JT, Nordin JH., 1999. Differential vitellin polypeptide processing in insect embryos. Micron 30: 579-596.

Gum JR Jr, Hicks JW, Toribara NW, Siddiki B, Kim YS., 1994. Molecular cloning of human intestinal mucin (*MUC2*) cDNA: identification of the amino terminus and overall sequence similarity to prepro-von Willebrand factor. J Biol Chem 269: 2440-2446.

Hagedorn HH, Maddison DR, Tu Z., 1998. The evolution of vitellogenins, cyclorrhaphan yolk proteins and related molecules. Adv Insect Physiol 27:335-384.

Hatakeyama M, Sawa M, Oishi K., 1990. Ovarian development and vitellogenesis in the sawfly, *Athalia rosae ruficornis* Jakovlev (Hymenoptera, Tenthredinidae). Invert Reprod Dev 17: 237-245.

Hirai M, Watanabe D, Kiyota A, Chinzei Y., 1998. Nucleotide sequence of vitellogenin mRNA in the bean bug, *Riptortus clavatus*: analysis of processing in the fat body and ovary. Insect Biochem Mol Biol 28: 537-547.

Hiremath S, Lehtoma K., 1997a. Complete nucleotide sequence of the vitellogenin mRNA from the gypsy moth: novel arrangement of the subunit encoding regions. Insect Biochem Mol Biol 27: 27-35.

Hiremath S, Lehtoma K., 1997b. Structure of the gypsy moth vitellogenin gene. Arch Insect Biochem Physiol 36: 154-164.

Kageyama Y, Kinoshita T, Umesono Y, Hatakeyama M, Oishi K., 1994. Cloning of cDNA for vitellogenin of *Athalia rosae* (Hymenoptera) and characterization of the vitellogenin gene expression. Insect Biochem Mol Biol 24: 599-605.

Kim HR, Lee SD., 1994. Purification and characterization of vitellin-2 from the ovary of the American cockroach, *Periplaneta americana*. Comp Biochem Physiol 108B: 135-145.

Kim HR, Katagiri C, Nagao E, Chino H., 1992. Purification and characterization of vitellogenin from the American cockroach, *Periplaneta americana*. Comp Biochem Physiol 103B: 963-967.

Lee JM, Hatakeyama M, Oishi K., 2000a. A simple and rapid method for cloning insect vitellogenin cDNAs. Insect Biochem Mol Biol 30: 189-194.

Lee JM, Nishimori Y, Hatakeyama M, Bae T-W, Oishi K., 2000b. Vitellogenin of the cicada *Graptopsaltria nigrofuscata* (Homoptera): analysis of its primary structure. Insect Biochem Mol Biol 30: 1-7.

Martin D, Piulachs MD, Comas D, Belles X., 1998. Isolation and sequence of a partial vitellogenin cDNA from the cockroach, *Blattella germanica* (L.) (Dictyoptera,

Blattellidae), and characterization of the vitellogenin gene expression. Arch Insect Biochem Physiol 38: 137-146.

Mayadas TN, Wagner DD.,1992. Vicinal cysteines in the prosequence play a role in von Willebrand factor multimer assembly. Proc Natl Acad Sci USA 89: 3531-3535.

Mouchel N, Trichet V, Betz A, Le Pennec JP, Wolff J., 1996. Characterization of vitellogenin from rainbow trout (*Oncorhynchus mykiss*). Gene 174: 59-64.

Nose Y, Lee JM, Ueno T, Hatakeyama M, Oishi K., 1997. Cloning of cDNA for vitellogenin of the parasitoid wasp, *Pimpla nipponica* (Hymenoptera: Apocrita: Ichneumonidae): vitellogenin primary structure and evolutionary considerations. Insect Biochem Mol Biol 27: 1047-1056.

Raikhel AS, Dhadialla TS., 1992. Accumulation of yolk proteins in insect oocytes. Annu Rev Entomol 37: 217-251.

Romans P, Tu Z, Ke Z, Hagedorn HH., 1995. Analysis of a vitellogenin gene of the mosquito, *Aedes aegypti* and comparisons to vitellogenins from other organisms. Insect Biochem Mol Biol 25: 939-958.

Sambrook J, Fritsch EF, Maniatis T., 1989. Molecular cloning: a laboratory manual. 2nd ed. New York: Cold Spring Harbor Laboratory Press.

Shirk PD., 1987. Comparison of yolk production in seven pyralid moth species. Int J Invert Reprod Dev 11: 173-188.

Storella JR, Wojchowski DM, Kunkel JG., 1985. Structure and embryonic degradation

of two native vitellins in the cockroach, *Periplaneta americana*. Insect Biochem 15: 259-275.

Takadera K, Yamashita M, Hatakeyama M, Oishi K., 1996. Similarities in vitellin antigenicity and vitellogenin mRNA nucleotide sequence among sawflies (Hymenoptera: Symphyta: Tenthredinoidea). J Insect Physiol 42: 417-422.

Thompson JD, Higgins DG, Gibson TJ., 1994. Clustal W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22: 4673-4680.

Trewitt PM, Heilmann LJ, Degrugillier SS, Kumaran AK., 1992. The boll weevil vitellogenin gene: nucleotide sequence, structure, and evolutionary relationship to nematode and vertebrate vitellogenin genes. J Mol Evol 34: 478-492.

Wyatt GR., 1991. Gene regulation in insect reproduction. Invert Reprod Dev 20: 1-35.

Wyatt GR, Davey KG., 1996. Cellular and molecular actions of juvenile hormone. II: roles of juvenile hormone in adult insects. Adv Insect Physiol 26: 1-155.

Yano K, Sakurai MT, Izumi S, Tomino S., 1994a. Vitellogenin gene of the silkworm, *Bombyx mori*: structure and sex-dependent expression. FEBS Lett 256: 207-211.

Yano K, Sakurai MT, Watabe S, Izumi S, Tomino S., 1994b. Structure and expression of mRNA for vitellogenin in *Bombyx mori*. Biochem Biophys Acta 1218: 1-10.

CHAPTER 3

Molecular Evidence for Two Vitellogenin Genes and Processing of Vitellogenins in the American Cockroach, *Periplaneta americana*

3.1 ABSTRACT

The American cockroach, *Periplaneta americana* has two vitellins (Vn1 and Vn2) and corresponding vitellogenins (Vg1 and Vg2). Vns/Vgs were separated on the SDS-PAGE as three major polypeptide bands [170, 100 (multisubunits) and 50 kD] and a minor polypeptide band (150 kD) both in the egg (mature terminal oocyte) extract and in the female hemolymph. We previously cloned one Vg (Vg1) cDNA and showed that the 170 kD polypeptide originated from the C-terminus of the Vg1. In the present study we cloned the other Vg (Vg2) cDNA. It is 5,826 bp long encoding 1,876 amino acid residues (including 16 residues for putative signal peptide) in a single ORF. The deduced amino acid sequences of both Vgs (Vg1 and Vg2) of *P. americana* showed 30% identity. The GL/ICG motif is followed by eight cysteine residues at conserved locations near the C-terminal and the DGXR motif starts 18 residues upstream of the GL/ICG motif. The chemically determined N-terminal amino acid sequences of the 150 kD and of the 50 kD polypeptides matched exactly with each other and with the deduced N-terminal amino acid sequence of the Vg2 cDNA. The pattern of processing in *P. americana* Vns/Vgs is discussed.

3.2 INTRODUCTION

Vitellogenins (Vgs) are precursors of the major egg yolk protein, vitellins (Vns), in many oviparous animals. Vg of most insect species is synthesized in the fat body cells, in tissue-, sex- and stage-specific manners, secreted into the hemolymph and then ultimately taken up by the developing oocytes via recepter-mediated endocytosis (reviewed in: Byrne et al., 1989; Raikhel and Dhadialla, 1992). During these processes Vgs and Vns are modified through cleavage, glycosylation, lipidation, and phosphorylation (reviewed in: Raikhel and Dhadialla, 1992; Hagedorn et al., 1998; Giorgi et al., 1999). After incorporation into the eggs, Vns represent a major component of egg yolk proteins and serve as storage proteins to provide a source of amino acids, carbohydrates, lipids, and phosphates to the developing embryo (reviewed in: Kunkel and Nordin, 1985; Bownes, 1986; Byrne et al., 1989).

Vgs and their genes or cDNAs have been investigated in a wide group of animals, both vertebrates and invertebrates including insects. Most insect Vgs/Vns consist of large (140-190 kD) and small (40-60 kD) subunits. The Vg genes are large and specify a single transcript encoding the Vg precursor protein of approximately 200 kD which undergoes proteolytic processing to generate the large and small subunits (reviewed in: Wyatt, 1991; Raikhel and Dhadialla, 1992). Recent molecular studies have shown that this is a common feature in most insect Vgs so far cloned and analyzed. Exceptions are the Vgs of the higher Hymenoptera (suborder Apocrita) which are not cleaved (Harnish and White, 1982; Nose et al., 1997) before being secreted into the hemolymph, and they are incorporated into developing oocytes as a large molecule.

Vgs and Vns of *P. americana* have been intensively investigated to determine the number and molecular weights of native Vg proteins and their derived subunits (Bell, 1970; Engelmann, 1979; Sams et al., 1980; Harnish and White, 1982; Storella et al., 1985; Kim et al., 1992; Kim and Lee, 1994). Previous studies indicated that there are at least two Vns (Vn1 and Vn2) and corresponding Vgs (Vg1 and Vg2), and each consists of multisubunits in this species (Storella et al., 1985; Kim et al., 1992; Kim and Lee,

1994). In our previous study, we cloned and sequenced the Vg1 cDNA of *P. americana* and suggested, on the basis of the N-terminal amino acid sequence analysis, that there are at least two Vg genes in *P. americana* and each of their products undergoes post-translational processing producing subunits with smaller molecular masses (Tufail et al., 2000). However, the proteolytic processing of the Vg is complicated and was not well clarified.

The aim of the present study was to clone/sequence the second Vg precursor of *P. americana*. It should clarify, on molecular basis, that the component Vn polypeptides, such as 150 and 50 kD, are really originated from this precursor. In fact, the N-terminal amino acid sequences of 150 and 50 kD Vn polypeptides did not match with the deduced amino acid sequence of Vg1 of *P. americana* (Tufail et al., 2000). The present results confirmed our previous proposal (Tufail et al., 2000) and are also consistent with the previous reports (Bell, 1970; Engelmann, 1979; Storella et al., 1985; Kim and Lee, 1994) that showed two electrophoretically and immunologically distinct Vg molecular species in *P. americana*. The deduced amino acid sequence of Vg2 of *P. americana* described herein was compared with that of the Vg1 of this speices (Tufail et al., 2000) and with those of 10 other known insect Vgs.

3.3 MATERIALS AND METHODS

3.3.1 Insects and Tissue Collection

Stock colonies of *Periplaneta americana* were maintained by feeding on artificial diet (MF, Oriental Yeast Corp.) and water, under constant light at 26°C. White roaches were collected daily and kept separately under LD 12:12 at 26°C. The fat body from the adult females (3-5 days-old) was isolated and the poly $(A)^+$ RNA was purified as reported previously (Tufail et al., 2000).

3.3.2 SDS-PAGE

Extracts of terminal oocytes and adult female and male hemolymph samples of *P. americana* were applied on 7% SDS-PAGE as described previously (Tufail et al., 2000). Samples subjected to SDS-PAGE per lane were 0.004 egg equivalents and 0.05µl hemolymph equivalents, respectively.

3.3.3 Immunological Screening of the cDNA Library and Determination of Nucleotide Sequence

A portion of the previously constructed cDNA expression library (approximately 0.8 x 10^5 pfu) from poly(A)$^+$ RNA obtained from female fat body cells (Tufail et al., 2000) was screened immunologically with the anti-*P. americana* 100 kD (multisubunits) antiserum raised previously (Tufail et al., 2000) following the method described in Nose et al. (1997) based on Sambrook et al. (1989). After three successive screenings, the positive clones were excised into pBluescript II SK$^-$. The transformants were stored at -80°C as 50% glycerol stocks.

The selected clones were individually digested with appropriate restriction enzymes and subcloned in pBluescript II SK$^-$. Each clone was directly sequenced using an ABI prism BigDye terminator cycle sequencing ready reaction kit (PE Applied Biosystems) with M13 forward and reverse primers according to the supplier's protocol. The PCR was done using GeneAmp PCR System 2400 under the conditions of 96°C for 1 min for denaturation followed by 25 cycles of 96°C for 10 sec, 50°C for 5 sec, 60°C for 4 min. The PCR products were concentrated by ethanol precipitation and then dissolved in 20 µl of the template suppression reagent supplied with the kit. The products were analyzed using a DNA sequencer (ABI, Prism 310 Genetic Analyser) according to the supplier's protocol.

3.3.4 5' End Amplification of cDNAs

The 5' end of the partial cDNA was obtained by the rapid amplification of the

cDNA ends (RACE) method as described previously by Tufail et al. (2000) based on Frohman et al. (1988). An adaptor-ligated double-stranded cDNA library was constructed from a portion (1 μg) of poly $(A)^+$ RNA using a Marathon cDNA amplification kit (Clontech), and was subjected to PCR with a gene-specific primer corresponding to a region of cloned *P. americana* Vg2 cDNA, and the adaptor primer supplied in the kit. The amplified PCR products (shown in Fig. 2) were separated on 1.2% agarose gels and DNA bands were cloned into the TOPO XL cloning vector (Invitrogen). Clones were cycle sequenced as described above.

3.3.5 N-terminal Amino Acid Sequencing of Vns

Egg (mature terminal oocyte) extracts (0.1-0.25 egg equivalents per lane) were subjected to SDS-PAGE as described by Tufail et al. (2000). The proteins were then transferred to the PVDF membrane (Millipore, Immobilon), stained for 30 sec with Ponceau S (0.2% in 1% acetic acid) and destained in distilled water. The bands corresponding to the respective Vn were cut out and two bands of each polypeptide (150 and 50 kD) were applied to a gas phase amino acid sequencer (Perkin Elmer, 492 Procise).

3.4 RESULTS AND DISCUSSION

3.4.1 Cloning and Sequence Determination of the cDNA for *Periplaneta americana* Vg2

The yolk proteins of *P. americana* consisted of three major polypeptides (170, 100 and 50 kD) and a minor polypeptide (150 kD) on the 7% SDS-PAGE gels (Fig. 1) as described previously (Tufail et al., 2000). A cDNA expression library constructed from poly$(A)^+$ RNA purified from the female fat body cells of *P. americana* was screened immunologically with the anti-*P. americana* 100 kD multisubunits antiserum. After three successive screenings, we obtained four positive clones ranging between 1.8-2.5

kb in length and constructed restriction maps for them. The nucleotide sequences of these four clones were then partially determined at both the 5' and 3' ends (each about 1 kb). Two of the four clones were identical and matched exactly with the *P. americana* Vg1 cDNA cloned and sequenced previously (Tufail et al., 2000), whereas the remaining two were, although identical with each other, different from the Vg1 cDNA. The longer clone in terms of the 5' end was selected and the sequence was determined. The clone was 2,078 bp long encoding 632 amino acids in a single ORF followed by a termination codon, a consensus polyadenylation signal (AATAAA) and a poly A tail (Fig. 4). The deduced amino acid sequence aligned well with the known insect Vg sequences. This suggested that the clone was a 3' portion of a new Vg cDNA of *P. americana*.

The remaining 5' end portion was cloned using the 5'-RACE method. A double-stranded cDNA library was prepared and subjected to PCR with a gene-specific primer (nucleotide positions 3852-3876 in the final clone, Fig. 4) and an adaptor primer. Three clones (each about 4.0 kb in length) were selected and sequenced. The sequences matched among the clones except for a very few polymorphic base substitutions. We termed this cDNA the *P. americana* Vg2. The final subcloning and sequence strategy along with the resriction map of cloned *P. americana* Vg2 cDNA is shown in Fig. 3. whereas, the entire nucleotide sequence and the deduced amino acid sequence obtained after combinig the initial clone and the 5' end portion of the cDNA for *P. americana* Vg2 are shown in Fig. 4. The complete nucleotide sequence was 5,826 bp long. Starting from the initiation codon, which appeared 15 nucleotides downstream from the linker site, a single long ORF encoding 1,876 amino acids followed by a termination codon (TAA) was

found. A consensus polyadenylation signal (AATAAA) was found at 136 nucleotides downstream of the termination codon and a ploy A tail started at 14 nucleotides downstream of the polyadenylation signal.

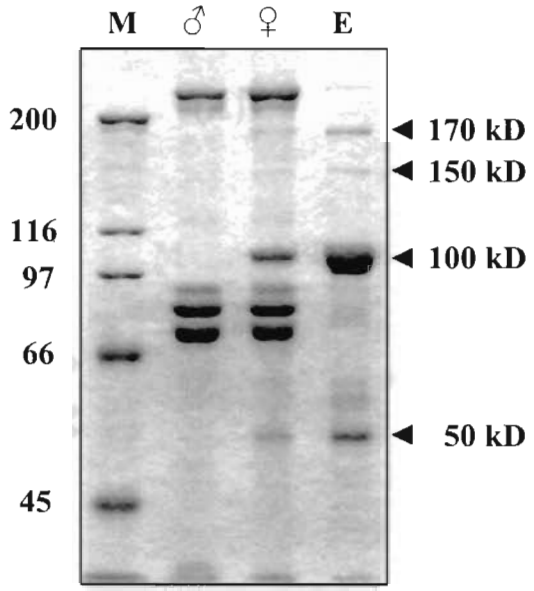


Fig. 1. Identification of *P. americana* Vns/Vgs by SDS-PAGE. Hemolymph from 8 day-old adult males and females, and egg (mature terminal oocyte) extracts (E) were applied. The SDS-PAGE gel (7%) was stained with Coomasie blue. M: molecular weight markers (kD). Arrowheads on the right indicate the four Vn/Vg polypeptides.

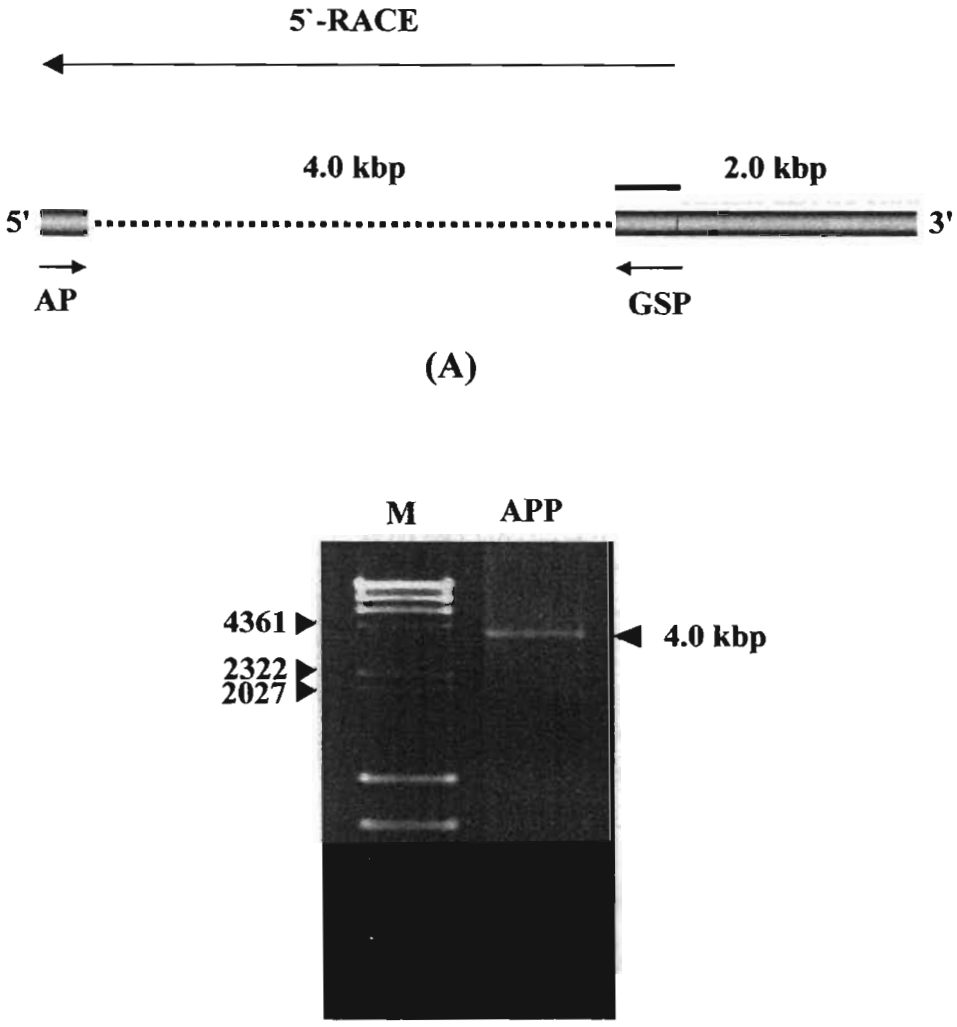


Fig. 2. 5′-end amplification of *P. americana* Vg2 cDNA. (A). A partial cDNA clone (2.0 kbp) has been shown on the 3` region; GSP, Gene specific primer (antisense); AP1, adapter primer; and a bold line is showing the overlapping region to identify the amplified portion (4.0 kbp). (B). M is molecular weight marker; APP is the amplified PCR-product.

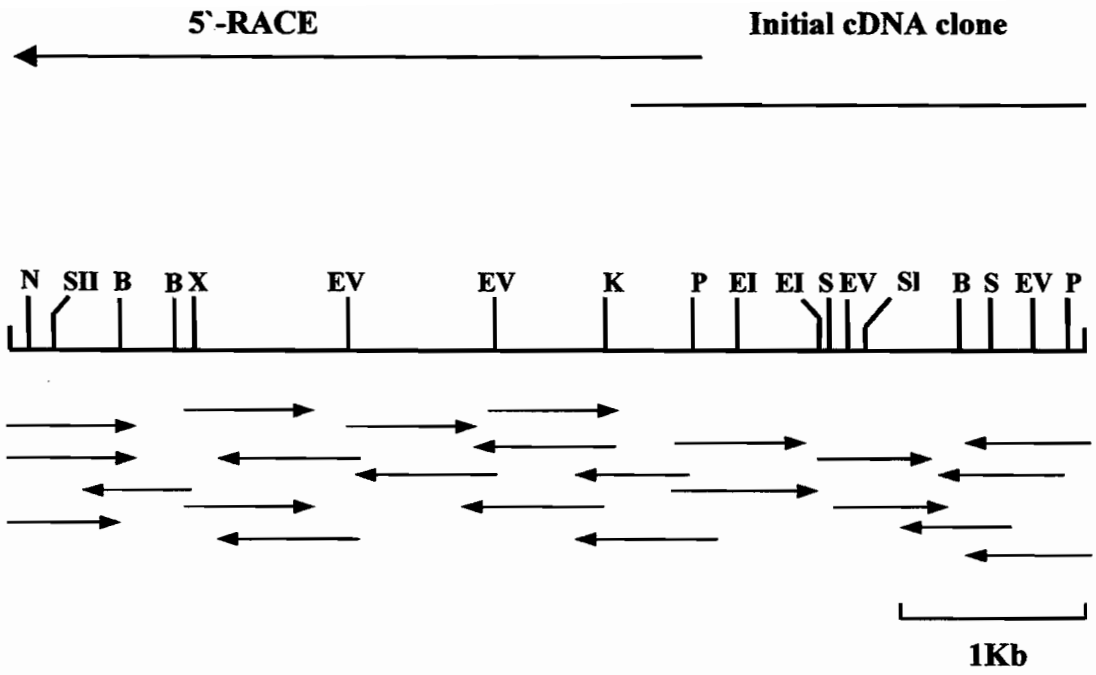
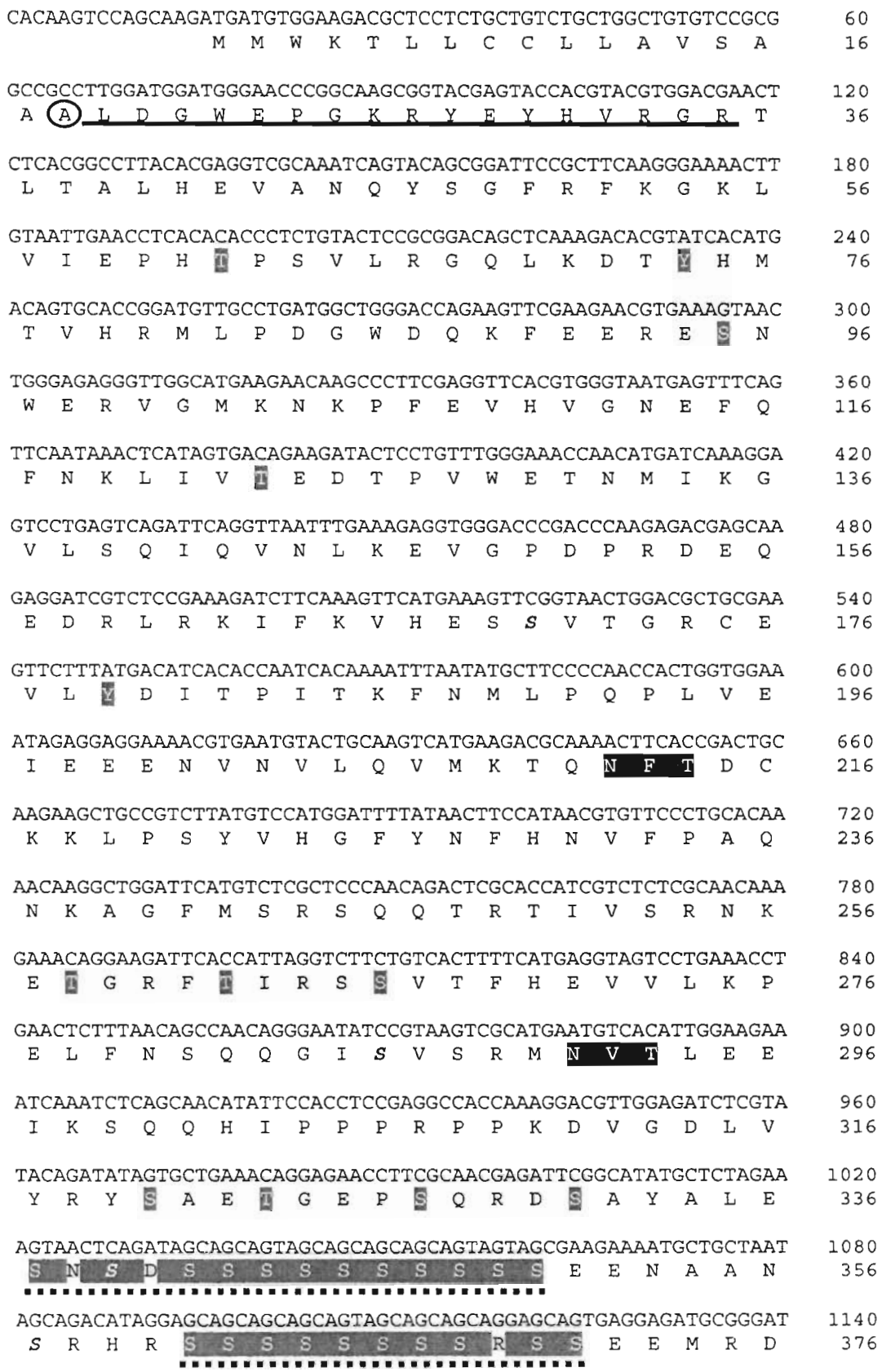
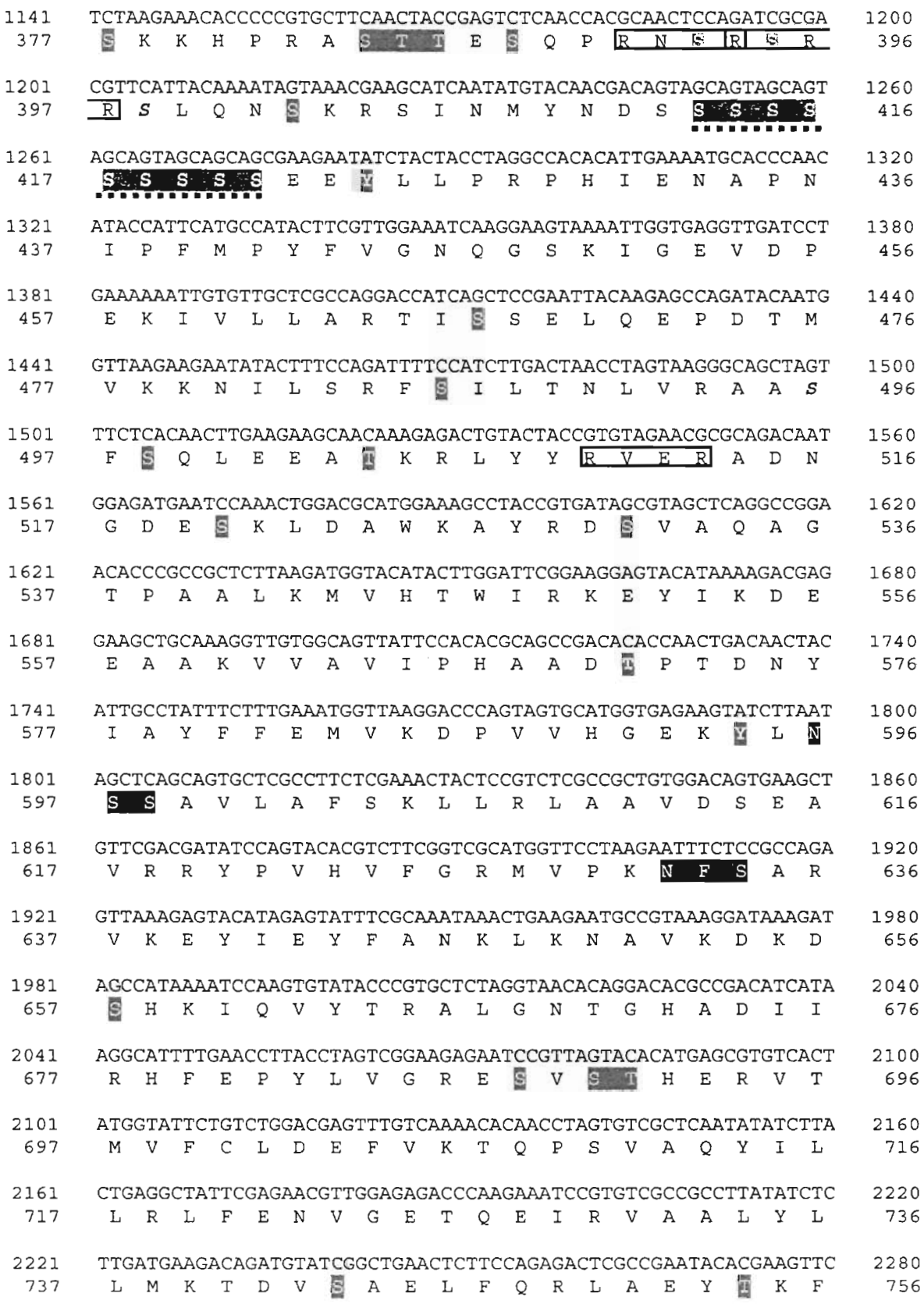


Fig. 3. Restriction map and sequence strategy of the cloned *P. americana* - Vg2 cDNA. N, *Not* I; SII, *Sac*II; B, *Bgl*II; X, *Xba*I; EV, *EcoR*V; K, *Kpn*I; P, *Pst*I; EI, *EcoR*I; S, *Sac*I and SI, *Sal*I.





1141 TCTAAGAAACACCCCCGTGCTTCAACTACCGAGTCTCAACCACGCAACTCCAGATCGCGA 1200
377 S K K H P R A S T T E S Q P R N S R S R 396

1201 CGTTCATTACAAAATAGTAAACGAAGCATCAATATGTACAACGACAGTAGCAGTAGCAGT 1260
397 R S L Q N S K R S I N M Y N D S S S S S 416

1261 AGCAGTAGCAGCAGCGAAGAATATCTACTACCTAGGCCACACATTGAAAATGCACCCAAC 1320
417 S S S S S E E Y L L P R P H I E N A P N 436

1321 ATACCATTCATGCCATACTTCGTTGGAAATCAAGGAAGTAAAATTGGTGAGGTTGATCCT 1380
437 I P F M P Y F V G N Q G S K I G E V D P 456

1381 GAAAAAATTGTGTTGCTCGCCAGGACCATCAGCTCCGAATTACAAGAGCCAGATACAATG 1440
457 E K I V L L A R T I S S E L Q E P D T M 476

1441 GTTAAGAAGAATATACTTTCCAGATTTTCCATCTTGACTAACCTAGTAAGGGCAGCTAGT 1500
477 V K K N I L S R F S I L T N L V R A A S 496

1501 TTCTCACAACTTGAAGAAGCAACAAAGAGACTGTACTACCGTGTAGAACGCGCAGACAAT 1560
497 F S Q L E E A T K R L Y Y R V E R A D N 516

1561 GGAGATGAATCCAAACTGGACGCATGGAAAGCCTACCGTGATAGCGTAGCTCAGGCCGGA 1620
517 G D E S K L D A W K A Y R D S V A Q A G 536

1621 ACACCCGCCGCTCTTAAGATGGTACATACTTGGATTCGGAAGGAGTACATAAAAGACGAG 1680
537 T P A A L K M V H T W I R K E Y I K D E 556

1681 GAAGCTGCAAAGGTTGTGGCAGTTATTCCACACGCAGCCGACACACCAACTGACAACTAC 1740
557 E A A K V V A V I P H A A D T P T D N Y 576

1741 ATTGCCTATTTCTTTGAAATGGTTAAGGACCCAGTAGTGCATGGTGAGAAGTATCTTAAT 1800
577 I A Y F F E M V K D P V V H G E K Y L N 596

1801 AGCTCAGCAGTGCTCGCCTTCTCGAAACTACTCCGTCTCGCCGCTGTGGACAGTGAAGCT 1860
597 S S A V L A F S K L L R L A A V D S E A 616

1861 GTTCGACGATATCCAGTACACGTCTTCGGTCGCATGGTTCCTAAGAATTTCTCCGCCAGA 1920
617 V R R Y P V H V F G R M V P K N F S A R 636

1921 GTTAAAGAGTACATAGAGTATTTCGCAAATAAACTGAAGAATGCCGTAAAGGATAAAGAT 1980
637 V K E Y I E Y F A N K L K N A V K D K D 656

1981 AGCCATAAAATCCAAGTGTATACCCGTGCTCTAGGTAACACAGGACACGCCGACATCATA 2040
657 S H K I Q V Y T R A L G N T G H A D I I 676

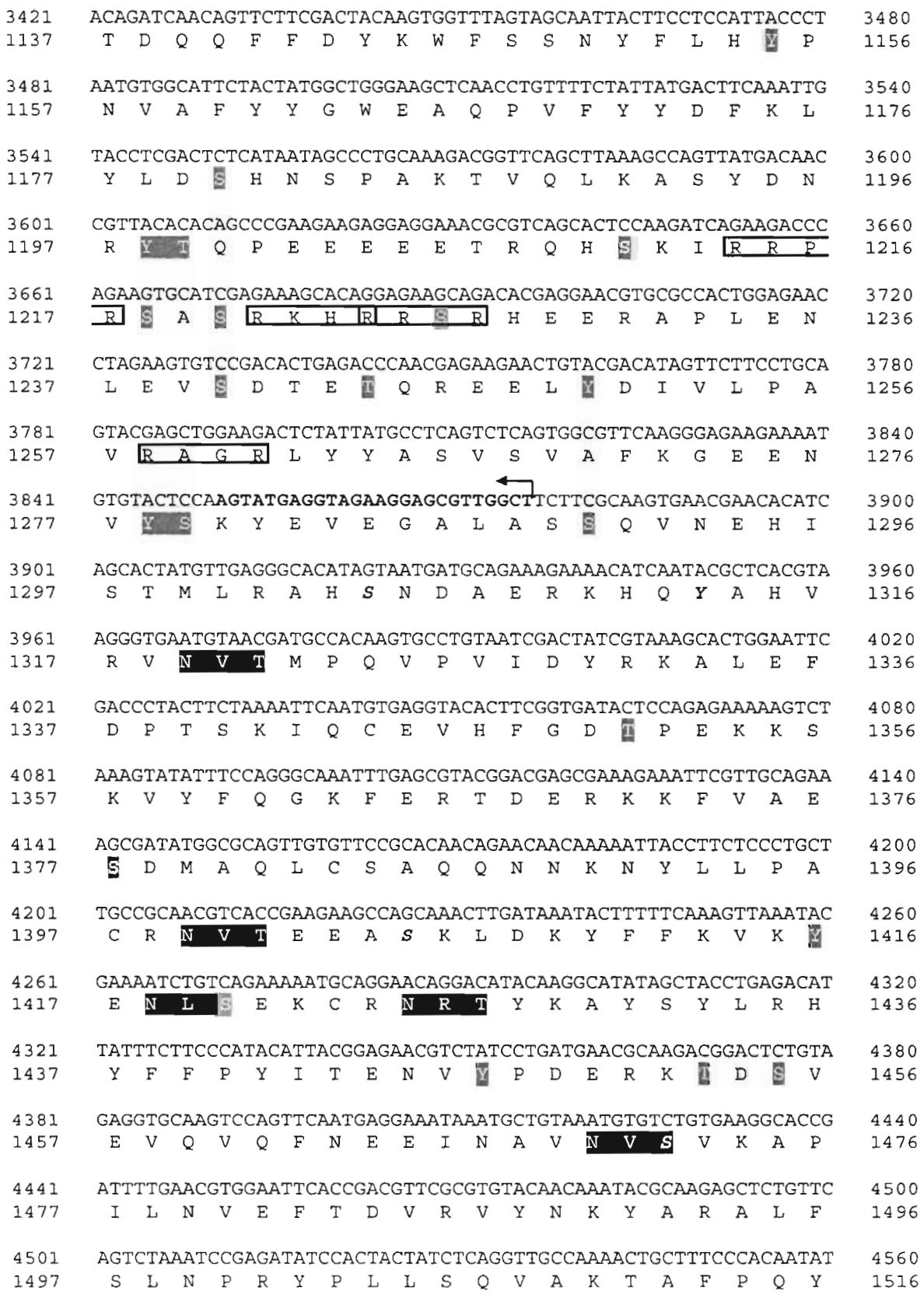
2041 AGGCATTTTGAACCTTACCTAGTCGGAAGAGAATCCGTTAGTACACATGAGCGTGTCACT 2100
677 R H F E P Y L V G R E S V S T H E R V T 696

2101 ATGGTATTCTGTCTGGACGAGTTTGTCAAAACACAACCTAGTGTCGCTCAATATATCTTA 2160
697 M V F C L D E F V K T Q P S V A Q Y I L 716

2161 CTGAGGCTATTCGAGAACGTTGGAGAGACCCAAGAAATCCGTGTCGCCGCCTTATATCTC 2220
717 L R L F E N V G E T Q E I R V A A L Y L 736

2221 TTGATGAAGACAGATGTATCGGCTGAACTCTTCCAGAGACTCGCCGAATACACGAAGTTC 2280
737 L M K T D V S A E L F Q R L A E Y T K F 756

```
2281   GACAAGAATCATCAAGTAGTTTCCGCCGTCCAATCAGCCATTCGATCTGCTGCAAAAGTT   2340
 757    D  K  N  H  Q  V  V  S  A  V  Q  S  A  I  R  S  A  A  K  V    776

2341   GAGGGACCGTACAAAAAAGAGACAGCCAAAAATGCTCAGGCAGCAGTCAAAATCCTTAGT   2400
 777    E  G  P  Y  K  K  E  T  A  K  N  A  Q  A  A  V  K  I  L  S    796

2401   TCTAAGCCTTATGATGATTCATATTCCAAGAGTTTCATCCTCAACAACTACAGGAGAGAG   2460
 797    S  K  P  Y  D  D  S  Y  S  K  S  F  I  L  N  N  Y  R  R  E    816

2461   ATTGACGTTGGTTATTCGCGCTTATACAATCAAATCGGAAGTCGAGACAGTTTCATGCCC   2520
 817    I  D  V  G  Y  S  R  L  Y  N  Q  I  G  S  R  D  S  F  M  P    836

2521   AAGTCAGTCTTCTATAAGCTGGTCAATATCATTGACGGTGACAGAGATGATCAAGCTAAG   2580
 837    K  S  V  F  Y  K  L  V  N  I  I  D  G  D  R  D  D  Q  A  K    856

2581   TTTGGGGGAGCAGTGTCAAGTGTAAGAGATGTAATCGACTTCATCAGACAACAGTTTAAG   2640
 857    F  G  G  A  V  S  S  V  R  D  V  I  D  F  I  R  Q  Q  F  K    876

2641   AAAGATGACAGTCAGGATGAGCTTGAAAATTCAAAGTATGCAGAAGACGATGATATCTGG   2700
 877    K  D  D  S  Q  D  E  L  E  N  S  K  Y  A  E  D  D  D  I  W    896

2701   GACTTGAGAGAAATTGCAAATTTATTGGAAATGGAGGAAGAGAATGTCGATCCGCTGGAA   2760
 897    D  L  R  E  I  A  N  L  L  E  M  E  E  E  N  V  D  P  L  E    916

2761   GGAAATGTACATTACGACTACTTCGGCGCTCAGAGATTTTTTACATTAAACAAAACCTCA   2820
 917    G  N  V  H  Y  D  Y  F  G  A  Q  R  F  F  T  L  N  K  T  S    936

2821   TTTGAATTTAGAGAAGAGCTGAAAAAATACTTCAAGAAACCACAAATTACGAATATTAAT   2880
 936    F  E  F  R  E  E  L  K  K  Y  F  K  K  P  Q  I  T  N  I  N    956

2881   AAGTTGTACAACAGAATGGAATTGAAAGTTGGCTATCCTAACGTTATGGGCGTTCCGTTC   2940
 957    K  L  Y  N  R  M  E  L  K  V  G  Y  P  N  V  M  G  V  P  F    976

2941   TTCTTCACATTCAAGAGACCAACACTTGTGAAACTTACAGCTAAGACATTTATCATGCCA   3000
 977    F  F  T  F  K  R  P  T  L  V  K  L  T  A  K  T  F  I  M  P    996

3001   TTGAAACCATGCGATCATGGTAAACCACATAAGTTCCCCAGGATCTTCAACGTTACTTCA   3060
 997    L  K  P  C  D  H  G  K  P  H  K  F  P  R  I  F  N  V  T  S   1016

3061   GATGTCTCATTTGTATATTCCTTCGACATGCATAGTCACATGGGAGTTGTGGCTCCATTT   3120
1017    D  V  S  F  V  Y  S  F  D  M  H  S  H  M  G  V  V  A  P  F   1036

3121   AACAAGAAGGAGTATGTAACAGGTATTCAGAGAAAACATATGATCCAAATTCCATTGAAC   3180
1037    N  K  K  E  Y  V  T  G  I  Q  R  K  H  M  I  Q  I  P  L  N   1056

3181   GTAAGTGTACATGTGAACTTGGATAAGAATAAAGTGGCGGCAGATTTCAAGCCATATTAT   3240
1057    V  S  V  H  V  N  L  D  K  N  K  V  A  A  D  F  K  P  Y  Y   1076

3241   GAAGATAATTTCAAAGTGGCGGAAGCAAGAGGTATTCCATTCACCACCGTGCACGACATC   3300
1077    E  D  N  F  K  V  A  E  A  R  G  I  P  F  T  I  V  H  D  I   1096

3301   AAATCCTTGGTACCGTATGTGGAAGCTGAGCATACTAGTTACATACGCGTGCGTCCAAGT   3360
1097    K  S  L  V  P  Y  V  E  A  E  H  T  S  Y  I  R  V  R  P  S   1116

3361   AAAGCGTACGAGGGCAACTTTGGTAAGAGCGTTGGAATGGTGTATCACTACAACTTCGAA   3420
1117    K  A  Y  E  G  N  F  G  K  S  V  G  M  V  Y  H  Y  N  F  E   1136
```



```
4561  TACGAACCTACCTGCGTCGTCGACTACTCCAAGGTGAACACATTTGATAACAGAACTTAC  4620
1517   Y  E  P  T  C  V  V  D  Y  S  K  V  N  T  F  D  N  R  T  Y   1536

4621  GAGCACGATATGCTGAATGACTGGGTCGAAGTGATGTTTCATAAACCGAGAAGCAAGATC  4680
1537   E  H  D  M  L  N  D  W  V  E  V  M  F  H  K  P  R  S  K  I   1556

4681  TACAAGCAGGTGTCTGTTTCAGCGAAACAGGCCCACAGCAAGATGGTGCTGAAAGTCCTC  4740
1557   Y  K  Q  V  S  V  S  A  K  Q  A  H  S  K  M  V  L  K  V  L   1576

4741  CGTGGAGATGAAGTGAAATTCGAAATGAAACAGCCTCGCGATTCTTCATCGTCTCCAGAG  4800
1577   R  G  D  E  V  K  F  E  M  K  Q  P  R  D  S  S  S  S  P  E   1596

4801  CTGAAAATGAATGATAAAGTTATCGAATACGAAAGAAGTCCTGCATTTATCCACTATAAG  4860
1597   L  K  M  N  D  K  V  I  E  Y  E  R  S  P  A  F  I  H  Y  K   1619

4861  CATGAATTGATTGCAGTAGCATATGCACTACCCAGCAAAGCTCTGCATTTAGATCTACTG  4920
1617   H  E  L  I  A  V  A  Y  A  L  P  S  K  A  L  H  L  D  L  L   1636

4921  AACGATAGCTTGGTGTTTGTTTATGATGGAGAGAGAGTCATGTTGCACGCCGGCAACCAT  4980
1637   N  D  S  L  V  F  V  Y  D  G  E  R  V  M  L  H  A  G  N  H   1656

4981  TATCGCAACCAAGTCAGAGGTCTGTGTGGCACATTCGACGGCGAACCATCCACAGACTTC  5040
1657   Y  R  N  Q  V  R  G  L  C  G  T  F  D  G  E  P  S  T  D  F   1676

5041  AAGGCACCGCAAAACTGCCACGTGAGAGATGTTGAGGACCTAATCCTTGCCTATACACTC  5100
1677   K  A  P  Q  N  C  H  V  R  D  V  E  D  L  I  L  A  Y  T  L   1696

5101  GTAAGGGATTTGGACAGGTCACGATTAAGAGATGAAAACATTTGCGTGAGAGAGGACGTC  5160
1697   V  R  D  L  D  R  S  R  L  R  D  E  N  I  C  V  R  E  D  V   1716

5161  CAGCTCGTCAACTTGACCAACCATCGCCACGCTGAGAAATCTGGCATCAGACCGTATGAT  5220
1717   Q  L  V  N  L  T  N  H  R  H  A  E  K  S  G  I  R  P  Y  D   1736

5221  ATTGATGACGATTCATCTTCTTCCTCATCGTCTTCGTCTTCGTCTTCATCTTCCTCTTCG  5280
1737   I  D  D  D  S  S  S  S  S  S  S  S  S  S  S  S  S  S  S  S   1756

5281  TCGTCCAAATCCAATTCCACTTCGAGCTCAAGTTCGGAATCAAATGAATCCGCTCTTCCC  5340
1757   S  S  K  S  N  S  T  S  S  S  S  S  E  S  N  E  S  A  L  P   1776

5341  AGAGGTGAAAACAAGCTCCATCGCGCACAACAACCCTCACGGAACTGCCACATGCATCTT  5400
1777   R  G  E  N  K  L  H  R  A  Q  Q  P  S  R  N  C  H  M  H  L   1796

5401  CACAGAATAGTGACCCATAACGGAAAACAATGTATCAGTAAGTTGGCCCTGACAGAATGC  5460
1797   H  R  I  V  T  H  N  G  K  Q  C  I  S  K  L  A  L  T  E  C   1816

5461  GCTCCTCTATGTCGCGAGGAATCTCACACCACGAAGACCGAGGCATTTGTATGCTTCCCA  5520
1817   A  P  L  C  R  E  E  S  H  T  T  K  T  E  A  F  V  C  F  P   1836

5521  CCTGGACCAACTGCTGACCATTATACGAAACTTGTCAGGAAGGGAGTGAGTCCAGATTTC  5580
1837   P  G  P  T  A  D  H  Y  T  K  L  V  R  K  G  V  S  P  D  F   1856

5581  AGCAGGAAGACTGATATCGTCAATTTACGTGTAACTATTCCTTCGAGGTGTGTTTCAAAA  5640
1857   S  R  K  T  D  I  V  N  L  R  V  T  I  P  S  R  C  V  S  K   1876

5641  ATTTAATTAABCCTTATGTTGTTGTATAAGATAAGAGATAGGGAATTTATTTGATTTTAA  5700
1877   I  *
```

```
5701    AAGAAATGTTGTAGTTACTATGTATTATTTATTCACAATATTTTTCACAAAATTAACTAG    5760

5761    AAGCATCTATATTTATGTATTTAATAAACGAACTGCAGATTTAAAAAAAAAAAAAAAAAA    5820

5821    AAAAAA                                                          5826
```

Fig. 4. Nucleotide sequence and deduced amino acid sequence of the Vg2 cDNA of *P. americana* (GenBank accession number: AB047401). Nucleotides are numbered on the right of each line. Amino acids are numbered on the left of each line from the translation initiation methionine. The determined 15 N-terminal amino acid residues (both for 150 and 50 kD polypeptide) are shown with a bold-line following the circled "A, alanine" (which is neither included in the 16 residues for the putative signal peptide nor matched with the determined N-terminal residues). GSP is shown with boldface-letters. Arrow indicates the 5`-end portion amplified. Consensus RXXR sequences for possible cleavage sites are boxed. Clusters of serine residues are underlined with a dotted line. A consensus polyadenylation signal is double-underlined. Possible glycosylation sites (having consensus sequence: NXS/T) are indicated with dark-shaded-boxes. The Potential phosphorylated serine (S), threonine (T) and tyrosine (Y) residues [predicted by using the Net Phos 2.0 computer program (Blom et al., 1999)] are shown with light-shaded-frames. The GLI/CG and DGXR motifs are shown with bold-face-Italic letters. Whereas, asterik indicates the stop codon.

3.4.2 Structural Analysis of *P. americana* Vg2

Four clusters of serine residues, including the one near the C-terminal region and 18 putative glycosylation sites were present. Seven putative cleavage sites (RXXR) for the proteases of the subtilisin family (Barr, 1991) were also found. The GL/ICG motif, eight cysteine residues at conserved locations near the C-terminal and the DGXR motif, starting 18 residues upstream of the GL/ICG motif were observed (Fig. 4) similar to the motifs and cysteines found in other insect Vgs (see Romans et al., 1995; Nose et al., 1997; Lee et al., 2000; Comas et al., 2000; Tufail et al., 2000).

Fifteen N-terminal amino acid residues chemically determined each for the 150 kD and 50 kD polypeptides matched completely with each other and both matched the deduced amino acid sequence of Vg2 that followed the putative signal peptide. We were not able to chemically determine the N-terminal amino acid sequences for the 100 kD multisubunits polypeptides.

There was one alanine residue suspended (circled in Fig. 4) which did not appear to belong to the signal peptide and which was not found in the determined N-terminal amino acid sequence in the 150 and 50 kD polypeptides of Vg2. The alanine residue might still be a part of the putative signal peptide. It is more probable, however, that the residue was originally part of the Vg product, but was removed in the mature polypeptide. The removal of similar residues at the N-terminal has been observed in *Bombyx mori* in which four residues were lost in the mature product after cleavage at the RXXR cleavage site (Yano et al., 1994).

3.4.3 Comparison With Vitellogenin Sequences of Other Species

The deduced amino acid sequences (excluding the signal peptides) of Vg2 of *P. americana*, Vg1 of the same cockroach (Tufail et al., 2000), and those of 10 other insect species [*Anthonomus grandis* (Coleoptera): Trewitt et al., 1992; *Aedes aegypti* (Diptera): Chen et al., 1994, Romans et al., 1995; *Bombyx mori* (Lepidoptera): Yano et al., 1994a, b; *Athalia rosae* (Hymenoptera): Kageyama et al., 1994, Nose et al., 1997;

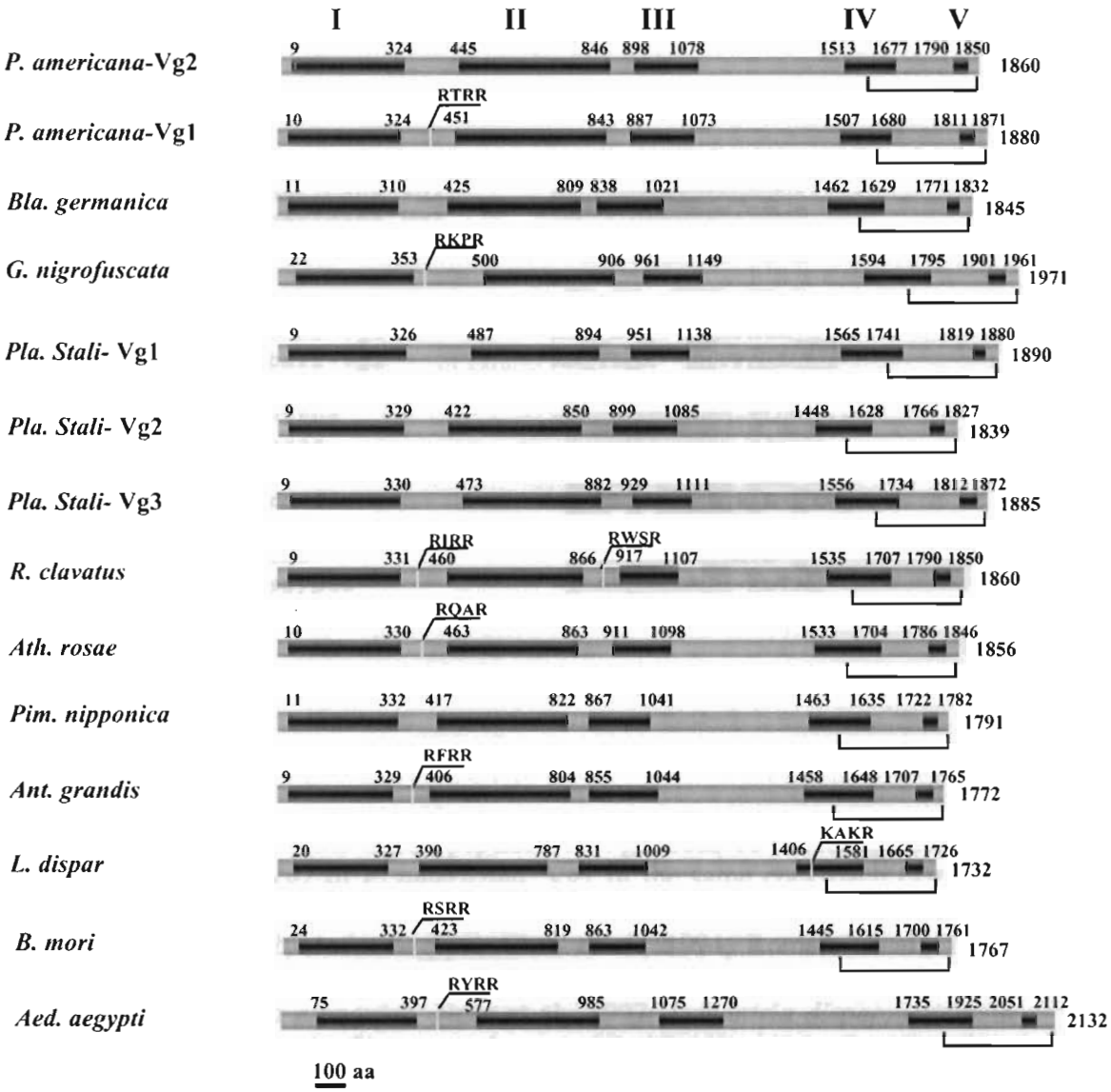


Fig. 5. Comparison of protein primary structures of 11 insect Vgs. Numbers indicate amino acid positions from the N-terminua (excluding the signal peptides). Subdomains I-V (hatched areas) are areas of relatively high amino acid conservation (Chen et al., 1997). Bold letters raised and underlined are basic amino acid sequences representing the dibasic protease recognition sites. Horizontal clamps at the C-termini indicate areas where 10 cysteines are present at conserved locations.

Fig. 6. Alignment of the signal peptides of Vgs from 11 insect species. The signal peptides of Vg2 of *P. americana*, Vg1 of the same cockroach (Tufail et al., 2000) and those from Vgs of 10 other insect species [*Anthonomus grandis*, Trewitt et al., 1992; *Aedes aegypti,* Chen et al., 1994, Romans et al., 1995; *Bombyx mori*, Yano et al., 1994a, b; *Athalia rosae*, Kageyama et al., 1994, Nose et al., 1997; *Pimpla nipponica*, Nose et al., 1997; *Lymantria dispar*, Hiremath and Lehtoma, 1997a, b; *Riptortus clavatus*, Hirai et al., 1998; *Graptopsaltria nigrofuscata*, Lee et al., 2000b; *Plautia stali*, Lee et al., 2000a; *Blattella germanica*, Comas et al. 2000] were aligned and compared.

Table 1. Overall amino acid identity ratios (%) in Vg amino acid sequences of 5 hemimetabolous and 6 holometabolous insect species

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. *P. americana*-Vg1		30	25	28	26	27	27	28	32	28	26	18	20	27
2. *P. americana*-Vg2			27	25	27	26	26	27	29	25	23	19	21	26
3. *Bla. Germanica*				23	24	23	23	22	25	22	20	17	19	23
4. *G. nigrofuscata*					30	29	28	30	33	28	25	22	23	27
5. *Pla. Stali*-Vg1						52	44	43	35	28	26	21	22	30
6. *Pla. Stali*-Vg2							45	42	35	29	25	21	23	29
7. *Pla. Stali*-Vg3								42	34	29	26	22	22	28
8. *R. clavatus*									38	32	29	22	24	29
9. *Ath. Rosae*										40	32	25	27	34
10. *Pim. Nipponica*											29	22	23	29
11. *Ant. Grandis*												22	22	26
12. *L. dispar*													43	23
13. *B. mori*														25
14. *Aed. aegypti*														

All sites with gaps in any sequence were deleted in the analysis.

Pimpla nipponica (Hymenoptera): Nose et al., 1997; *Lymantria dispar* (Lepidoptera): Hiremath and Lehtoma, 1997a, b; *Riptortus clavatus* (Heteroptera): Hirai et al., 1998; *Graptopsaltria nigrofuscata* (Homoptera): Lee et al., 2000b; *Plautia stali* (Heteroptera): Lee et al., 2000a; *Blattella germanica* (Dictyoptera): Comas et al. (2000)] were aligned using the Clustal W computer Program (Thompson et al., 1994) and compared to confirm the common characteristics found in the protein primary structure of insect Vgs (see Fig. 5). The comparison of overall amino acid identity ratio (%) in Vg amino acid sequence of Vg2 of *P. americana* with those of other insect species (Table 1) indicates that the identity ratio was comparatively higher within the sequences/members of the Hemimetabola (like *P. americana* -Vg1: 30%; *B. germanica*: 27%; *G. nigrofuscata* : 25%; *P. stali* -Vg1, -Vg2 and -Vg3 : 26-27%; *R. clavatus* : 27%) than from the members of the Holometabola (like *P. nipponica* : 25%; *A. grandis* : 23%; *L. dispar* : 19%; *B. mori* :21% and *A. aegypti* : 26%) with the exception of that of *A. rosae* Vg where the amino acid identity ratio was also higher (29%) as previously reported (Comas et al., 2000).

The overall amino acid identity between the two Vgs (Vg1 and Vg2) of *P. americana* was unexpectedly low (only 30%), in contrast to the situation in the bean bug, *Plautia stali*, where the ratio was 52% between Vg-1 and Vg-2, 44% between Vg-1 and Vg-3, and 45% between Vg-2 and Vg-3 (Lee et al., 2000).

Moreover, the signal peptide of *P. americana* Vgs (Vg1and Vg2) and those of 10 other insect Vgs were aligned and compared (see Fig. 6). The high conservation of amino acid residues in the primary structure of these signal peptides suggest that thay are orthologous and are originated from a common ancestor signal peptide for insect vitellogenins.

P. americana Vns/Vgs and Their Processing

Previous studies indicated that there are two Vn/Vg molecular species in *P. americana* each of which consists of multisubunits (Bell, 1970; Engelmann, 1979;

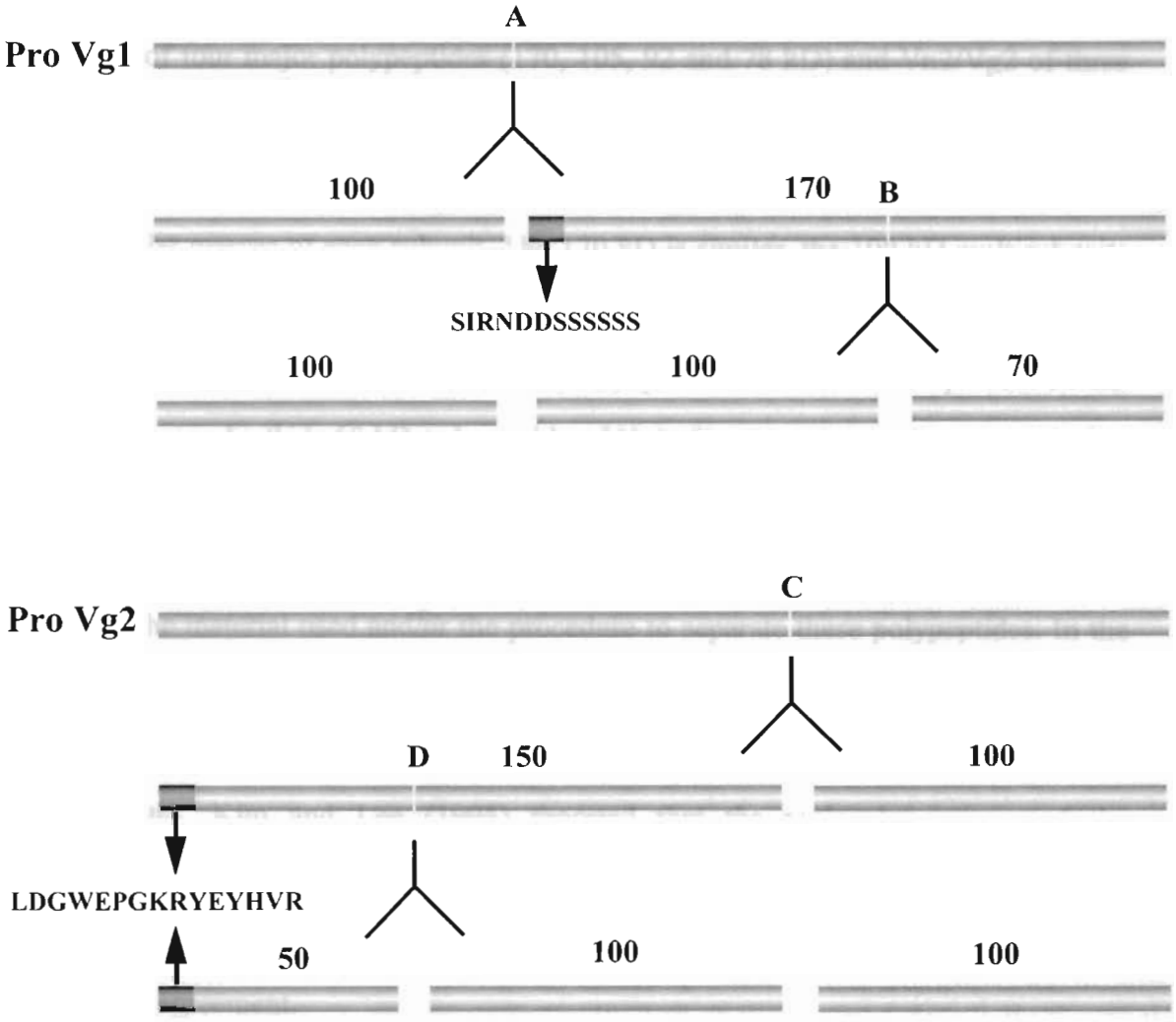


Fig. 7. Diagrammatic representation of the proposed processing patterns of Vgs of *P. americana*. Numbers indicate the molecular masses (kD) of the polypeptides. Letters indicate the cleavage sites. Pro-vitellogenin (Pro Vg) is synthesized and then processed in the fat body before being secreted into the hemolymph. The determined N-terminal amino acid sequences are shown with shaded boxes and arrows.

Storella

et al., 1985; Kim and Lee, 1994). Storella et al. (1985) showed that Vn1/Vg1 was composed of four major polypeptides (170, 105, 92 and 78 kD) and Vn2/Vg2 of three major polypeptides (105, 101, and 60 kD). The present findings are basically consistent with those repoted by Storella et al. (1985), although there are some discrepancies in the molecular masses of each subunit. The 170 kD is similar, the 100 kD multisubunits in the present study most probably represent the mixture of their 105 and 92 kD polypeptides of Vn1 and 105 and 101 kD polypeptides of Vn2, and the present 50 kD polypeptide may be their 60 kD polypeptide of Vn2. The 150 kD polypeptide is similar to their 161 kD "alpha" polypeptide described in Fig. 12 of Storella et al. (1985), which was a combination of the 100 kD and a 50 kD polypeptides. These discrepancies might be due to the material used and/or the procedure to separate these polypeptides. In the present experiments the mature terminal oocyte extracts were used whereas Storella et al. (1985) used whole oothecal extracts which might contain components other than from oocytes. Kim and Lee (1994) reported that the molecular masses of the components of Vn2 were 102, 72 and 42 kD. They used the growing ovary which was separated by gel filteration. The present and these previous findings are, however, basically in agreement.

From the present findings, it is suggested that the processing steps of *P. americana* Vns/Vgs are as summarized in Fig. 7. The Vg1 polypeptide would first be processed completely at site A resulting in the production of 100 and 170 kD polypeptides as suggested also by Storella et al. (1985). This is supported by our previous findings that the determined N-terminal amino acid sequence of the 170 kD polypeptide matched the deduced amino acid sequence of the Vg1 (Tufail et al., 2000). The 170 kD polypeptide would then be cleaved incompletely at site B into 100 and 70 kD polypeptides in the fat body before being secreted into the hemolymph according to the previous study (Storella et al., 1985). However, the 70 kD polypeptide was not visible on the SDS-PAGE gels. The cleavage might rarely occur and most of the 170

kD polypeptide might remain uncleaved. Alternatively, the 70 kD polypeptide might be further cleaved into smaller polypeptides so far undetected on the SDS-PAGE gels. In the case of Vg2 polypeptides, the main processing site is at C resulting in the 150 and 100 kD polypeptides. The Vg2 polypeptide is then further processed at site D resulting in the production of 50 and 100 kD polypeptides in the fat body. The determined N-terminal amino acid sequences of 150 and 50 kD polypeptides demonstrated that they share the same N-terminal amino acid sequences and are thus the N-terminus of the Vg2. This was in contrast to those shown by Storella et al. (1985) in which the 150 kD polypeptide originated from the C-terminus of the Vg2. We previously reported that the 50 kD polypeptide might consist of multisubunits (Tufail et al., 2000). The present findings, however, show that this polypeptide is not a multisubunit but actually is a component of Vg2. The cleavage at site D would be incomplete so that the 150 kD polypeptides remained as a minor component in the hemolymph and egg extracts. The cleavage at the sites A and C should be complete because ~250 kD polypeptide was not formed either in the hemolymph or in the egg extracts. The incomplete cleavage is probably not uncommon as observed in the Vg of *Riptortus clavatus* (Hirai et al., 1998).

The findings of the present study conclusively demonstrated the existence of two Vg genes in *P. americana.* The existence of multiple Vg genes was reported in *Locusta migratoria* (Wyatt et al., 1984), *Aedes aegypti* (Chen et al., 1994; Romans et al., 1995), *Riptortus clavatus* (Hirai et al., 1998) and *Plautia stali* (Lee et al., 2000). It remains unclear why multiple Vg genes exist in many species. Multiple Vg genes may be necessary to provide a large amount of Vn within a short period of egg maturation. A not-mutually-exclusive possibility is that multiple genes, apparently derived from a common ancestor, might have acquired different functions. For instance, the plasma clotting protein in the crayfish share common characteristics with Vgs and yet has unique functions (Hall et al., 1999). Cloning of Vg cDNAs and of genomic DNAs from many more species should prove to be of importance.

LITERATURE CITED

Barr PJ. 1991. Mammalian subtilisins: the long-sought dibasic processing endoproteases. Cell 66: 1-3.

Bell WJ. 1970. Demonstration and characterization of two vitellogenic blood proteins in *Periplaneta americana*: an immunochemical analysis. J Insect Physiol 16: 291-299.

Blom N, Gammeltoft S, Brunak S. 1999. Sequence-and Structure-Based Prediction of Eukaryotic Protein Phosphorylation Sites. J Mol Biol 294 (5): 1351-1362.

Bownes M. 1986. Expression of genes coding for vitellogenin (yolk protein). Annu Rev Entomol 31: 507-531.

Byrne BM, Gruber M, Ab G. 1989. The evolution of egg yolk proteins. Progr Biophys Molec Biol 53: 33-69

Chen JS, Cho WL, Raikhel AS. 1994. Analysis of mosquito vitellogenin cDNA: similarity with vertebrate phosvitins and arthropod serum proteins. J Mol Biol 237: 641-647.

Comas D, Piulachs MD, Belles X. 2000. The vitellogenin of *Blattella germanica* (L.) (Dictyoptera, Blattellidae): nucleotide sequence of the cDNA and analysis of the protein primary structure. Arch. Insect Biochem. Physiol. 36, 154-164.

Engelmann F. 1979. Insect vitellogenin: identification, biosynthesis, and role in vitellogenesis. Adv Insect Physiol 14: 49-108.

Frohman MA, Dush MK, Martin GR. 1988. Rapid production of full length cDNAs from rare transcripts: amplification using a single-gene specific oligonucleotide primer. Proc Natl Acad Sci USA 85: 8998-9002.

Giorgi F, Bradley JT, Nordin JH. 1999. Differential vitellin polypeptide processing in insect embryos. Micron 30: 579-596.

Hagedorn HH, Maddison DR, Tu Z. 1998. The evolution of vitellogenins, cyclorrhaphan yolk proteins and related molecules. Adv Insect Physiol 27: 335-384.

Hall M, Wang R, Van Antwerpen R, Sottrup-Jensen L, Soderhall K. 1999. The crayfish plasma clotting protein: a vitellogenin-related protein responsible for clot formation in crustacean blood. Proc Natl Acad Sci USA 96: 1965-1970. 220: 1-10.

Harnish DG, White BN. 1982. An evolutionary model for the insect vitellins. J Mol Evol 18: 405-413.

Hirai M, Watanabe D, Kiyota A, Chinzei Y. 1998. Nucleotide sequence of vitellogenin mRNA in the bean bug, *Riptortus clavatus*: analysis of processing in the fat body and ovary. Insect Biochem Mol Biol 28: 537-547.

Hiremath S, Lehtoma K, 1997a. Complete nucleotide sequence of the vitellogenin mRNA from the gypsy moth: novel arrangement of the subunit encoding regions. Insect Biochem. Mol. Biol. 27, 27-35.

Hiremath S, Lehtoma K, 1997b. Structure of the gypsy moth vitellogenin gene. Arch. Insect Biochem. Physiol. 36, 154-164.

Kageyama Y, Kinoshita T, Umesono Y, Hatakeyama M, Oishi K. 1994. Cloning of

cDNA for vitellogenin of *Athalia rosae* (Hymenoptera) and characterization of the vitellogenin gene expression. Insect Biochem. Mol. Biol. 24, 599-605.

Kim HR, Katagiri C, Nagao E, Chino H. 1992. Purification and characterization of vitellogenin from the American cockroach, *Periplaneta americana*. Comp Biochem Physiol 103B: 963-967.

Kim HR, Lee SD. 1994. Purification and characterization of vitellin-2 from the ovary of the American cockroach, *Periplaneta americana*. Comp Biochem Physiol 108B: 135-145.

Kunkel JG, Nordin JH. 1985. Yolk proteins. In: Kerkut GA and Gilbert LJ (eds) Comprehensive Insect Physiology, Biochemistry and Pharmacology. Permagon Press, Oxford, pp 83-111.

Lee JM, Hatakeyama M, Oishi K. 2000a. A simple and rapid method for cloning insect vitellogenin cDNAs. Insect Biochem Mol Biol 30: 189-194.

Lee JM, Nishimori Y, Hatakeyama M, Bae T-W, Oishi K. 2000b. Vitellogenin of the cicada *Graptopsaltria nigrofuscata* (Homoptera): analysis of its primary structure. Insect Biochem. Mol. Biol. 30, 1-7.

Nose Y, Lee JM, Ueno T, Hatakeyama M, Oishi K. 1997. Cloning of cDNA for vitellogenin of the parasitoid wasp, *Pimpla nipponica* (Hymenoptera: Apocrita: Ichneumonidae): vitellogenin primary structure and evolutionary considerations. Insect Biochem Mol Biol 27: 1047-1056.

Raikhel AS, Dhadialla TS. 1992. Accumulation of yolk proteins in insect oocytes.

Annu Rev Entomol 37: 217-251.

Romans P, Tu Z, Ke Z, Hagedorn HH. 1995. Analysis of a vitellogenin gene of the mosquito, *Aedes aegypti* and comparisons to vitellogenins from other organisms. Insect Biochem Mol Biol 25: 939-958.

Sambrook J, Fritsch EF, Maniatis T. 1989. Molecular Cloning: a laboratory manual. 2nd ed. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Sams GR, Bell WJ, Weaver RF. 1980. Vitellogenin: its structure, synthesis and processing in the cockroach *Periplaneta americana*. Biochim Biophysica Acta 609: 121-135.

Storella JR, Wojchowski DM, Kunkel JG. 1985. Structure and embryonic degradation of two native vitellins in the cockroach, *Periplaneta americana*. Insect Biochem 15: 259-275.

Thompson JD, Higgins DG, Gibson TJ. 1994. Clustal W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22, 4673-4680.

Tufail M, Lee JM, Hatakeyama M, Oishi K, Takeda M. 2000. Cloning of vitellogenin cDNA of the American cockroach, *Periplaneta americana* (Dictyoptera), and its structural and expression analyses. Arch Insect Biochem Physiol 45: 37-46.

Trewitt PM, Heilmann LJ, Degrugillier SS, Kumaran AK. 1992. The boll weevil vitellogenin gene: nucleotide sequence, structure, and evolutionary relationship to nematode and vertebrate vitellogenin genes. J. Mol. Evol. 34, 478-492.

Wyatt GR. 1991. Gene regulation in insect reproduction. Invert Reprod Dev 20: 1-35.

Wyatt GR, Locke J, Bradfield JY. 1984. The vitellogenin of *Locusta migratoria* and other insects. Adv Invert Reprod 3: 73-80.

Yano K, Sakurai MT, Izumi S, Tomino S. 1994a. Vitellogenin gene of the silkworm, *Bombyx mori*: structure and sex-dependent expression. FEBS Lett. 256, 207-211.

Yano K, Sakurai MT, Watabe S, Izumi S, Tomino S. 1994b. Structure and expression of mRNA for vitellogenin in *Bombyx mori*. Biochem Biophys Acta 1218: 1-10.

CHAPTER 4

Vitellogenin of the cockroach, *Leucophaea maderae*: nucleotide sequence, structure and analysis of processing in the fat body and oocytes

4.1 ABSTRACT

A cDNA encoding vitellogenin (Vg) of the cockroach, *Leucophaea maderae* was cloned and sequenced. The deduced amino acid sequence consisting of 1,913 residues (including 15 residues for a putative signal peptide) was obtained. Amino-terminal sequence analysis demonstrated that the pro-Vg was cleaved into four polypeptide "subunits" following the three consensus RXXR cleavage site sequences, which were secreted as four Vg polypeptides (apparent molecular weights = 112-, 100-, 92- and 55-kD), sequestered, and deposited in the egg as four respective vitellin (Vn) polypeptides. There was, however, an additional 90-kD Vn polypeptide existed in the egg. We show that this polypeptide is a processed product from 92-kD Vn polypeptide. Northern blot analysis of poly $(A)^+$ RNA reveals that mRNA coding for Vg is present only in the female fat body cells but neither in the ovary nor in the male fat body cells. The deduced amino acid sequence contained a serine-rich at the C-terminal region. This stretch occurred also in Vgs of *Periplaneta americana* (Vg1 and Vg2) and *Blattella germanica*. The Vg of *L. maderae* had 26% and 31% homology with those of *P. americana* (Vg1 and Vg2) and *B. germanica*, respectively. Phylogenetic analysis (neighbour-joining) was made using four cockroach Vgs and the tree was compared with other molecular and conventional phylogenetic trees.

4.2 INTRODUCTION

Vitellogenins (Vgs), the yolk protein precursors and vitellins (Vns), the yolk proteins have been studied extensively. They are considered to be the source of nutrition for developing embryo. In insects, the vitellogenesis generally proceeds from the biosynthesis of Vgs, the processing in the fat body, the secretion of the processed Vgs into the hemolymph, the selective uptake by competent oocytes and finally to the utilization of Vns by the developing embryo (Byrne et al., 1989; Raikhel and Dhadialla, 1992; Wyatt, 1991; Izumi et al., 1994; Hagedorn et al., 1998; Sappington and Raikhel, 1998; Giorgi et al., 1999). During post-translational modifications (see above reviews), the primary Vg gene product of most insect species, with a molecular mass of about 200 kD, is cleaved into large (140-190 kD) and small (adout 50 kD) subunits before being secreted into the hemolymph. In bees and wasps (belonging to higher Hymenoptera, Apocrita), the primary Vg gene product, however, is about 180 kD and is secreted without processing. In contrast, the yolk proteins of higher Diptera such as *Drosophila melanogaster* are quite different from other insect Vgs and also are not processed. Vg synthesis in all insect species, except in certain Lepidoptera and Diptera, is regulated at the transcriptional level. The sex-, tissue-, and hormone-mediated developmental specificities of Vg transcription have been reported for many insect species (Bownes, 1986). Vg genes and/or cDNAs provide excellent model systems for studying the molecular basis of gene regulation.

We previously cloned two Vg (Vg1 and Vg2) cDNAs from *P. americana* (Tufail et al., 2000; 2001) and also reported the similarities in Vn-antigenicity among 9 cockroach species belonging to 2 superfamilies: Blattoidea and Blaberiodea, using the antisera raised against the *P. americana* Vns (Tufail et al., 2000). The antigenicity was limited to within the superfamily in cockroaches, except for in *L. maderae* where a 90 kD Vn polypeptide reacted with both the anti-*P. americana* 100 and 50 kD Vn polypeptide antisera. This antigenic relatedness between these two species raised our interest to clone/analyze the Vg cDNA of *L. maderae,* and to compare it with *P.*

americana and other known cockroach Vgs to clarify the phylogenetic relationship at the molecular level.

In the present study, we report on the structural characteristics of Vg of *L. maderae* by cloning a complete cDNA for Vg through simple and rapid method (Lee et al., 2000) and also discuss the processing patterns of this molecule. We also confirm the Vg gene expression in the fat body of vitellogenic females.

4.3 MATERIALS AND METHODS

4.3.1 Animals

Cultures of *L. maderae* were maintained at 26°C under constant light, fed with an artificial diet (MF, Oriental Yeast Corp.) and water. Newly emerged females were collected from stock colonies and kept separately under LD 12:12 at 26°C. The fat body was isolated from 15 females (3-6 days old adult) which was rinsed in phosphate-buffered saline (PBS) (see Sambrook et al., 1989). It was frozen immediately in liquid nitrogen and stored at –80°C for the extraction of the Poly(A)$^+$ RNA.

4.3.2 Construction of an Adaptor-Ligated Double-Stranded cDNA Library

Total RNA (868 μg) was extracted from the female fat body as reported previously (Tufail et al., 2000). Poly(A)$^+$ RNA (17.6 μg) was purified from total RNA using an mRNA purification kit (Amersham-Pharmacia). A total of 1 μg of Poly(A)$^+$ RNA was used to construct an adaptor-ligated double stranded (ds) cDNA library using a marathon cDNA amplification kit (Clontec). The adaptor–ligated ds cDNA library (10 μl) was diluted with Tricine-EDTA (1:50), heat denatured at 94°C for 2 min and was then used for the extension of both the 3'- and 5'- ends of *L. maderae* Vg cDNA as described by Lee et al. (2000).

4.3.3 3`-end Amplification of cDNAs and Sequencing

To amplify the 3' end portion of *L. maderae* Vg cDNA from GL/ICG motif (see Lee et al., 2000; Tufail et al., 2000), ten primers (11-mer each) for this motif (7 for GLCG and 3 for GICG) were prepared. The oligonucleotides for the primers were designed on the basis of the known Vg sequences from other species and are shown in Table 1. The 3' end portion of *L. maderae* Vg was then amplified through RACE-PCR (shown in Fig. 1). An adaptor-ligated ds cDNA library was subjected to PCR (Gene Amp PCR Systems 2400) with a GL/ICG primer and the adaptor primer (marathon cDNA adaptor, Clontech) following the protocols of the supplier. The PCR conditions employed were: 94°C for 1 min for denaturing, followed by 30 cycles of 94°C for 30 s, 55°C for 30 s and 68°C for 3 min.

The PCR products were separated on 1.2% low melting agarose gels and the amplified DNA bands of approximately 1.0 kb (Fig. 1) corresponding to the 3' portion of the gene were cloned into the TOPO TA cloning kit (Invitrogen). The clones were then directly sequenced as reported previously (Tufail et al., 2001). The obtained nucleotide sequences were analyzed and the deduced amino acid sequences having the GL/ICG motif were checked for their homology with Vgs of other insects using the FASTA homology search on the DDBJ database.

4.3.4 5`-end Amplification of cDNAs

To extend toward the 5'-end of the gene, a gene-specific primer corresponding to the 5' end of the initial sequence containing the GLI/CG motif coding sequence of *L. maderae* Vg cDNA (nucleotide positions 5102 to 51025) was prepared. An adaptor-ligated ds cDNA library was subjected to PCR with a gene-specific primer and the adaptor primer (Clontech) following the protocols of the supplier. The PCR conditions employed were: 94°C for 1 min for denaturing, followed by 36 cycles of 94°C for 30 s and 68°C for 8 min. The amplified cDNAs (Fig. 2) were purified, cloned and sequenced as described previously (Tufail et al., 2001).

Table 1. Oligonucleotides used as primers for the GLI/CG motif prepared on the basis of known sequences of other insects to obtain the 3`-end portion of *L. maderae* Vg cDNA

Primer name	Sequence (5`→3`)
GLCG-P1	GGACTCTGTGG
GLCG-P2	GGTCTGTGTGG
GLCG-P3	GGTCTCTGCGG
GLCG-P4	GGACTGTGTGG
GLCG-P5	GGTCTCTGTGG
GLCG-P6	GGGCTCTGCGG
GLCG-P7	GGDCTSTGYGG
GICG-P8	GGTATTTGCGG
GICG-P9	GGCATCTGTGG
GICG-P10	GGYATYTGYGG

Where as D = A+T+G, S = G+C, and Y = C+T and P with numeral represents the number of the primer

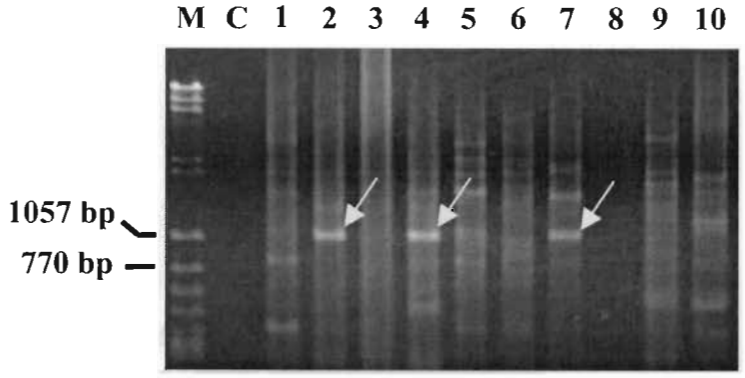


Fig. 1. The 3` RACE-PCR products as identified on agarose gel electrophoresis. An adaptor–ligated double stranded cDNA constructed from poly $(A)^+$ RNA prepared from females (3-6 days old adult) fat body cells of *L. maderae* was subjected to PCR with each of the primers# P1-P10 (see materials and methods) and with the adaptor sequence for lanes 1-10. M: molecular size marker; C: negative control (reaction mixture only). The products cloned and sequenced are indicated by arrowheads.

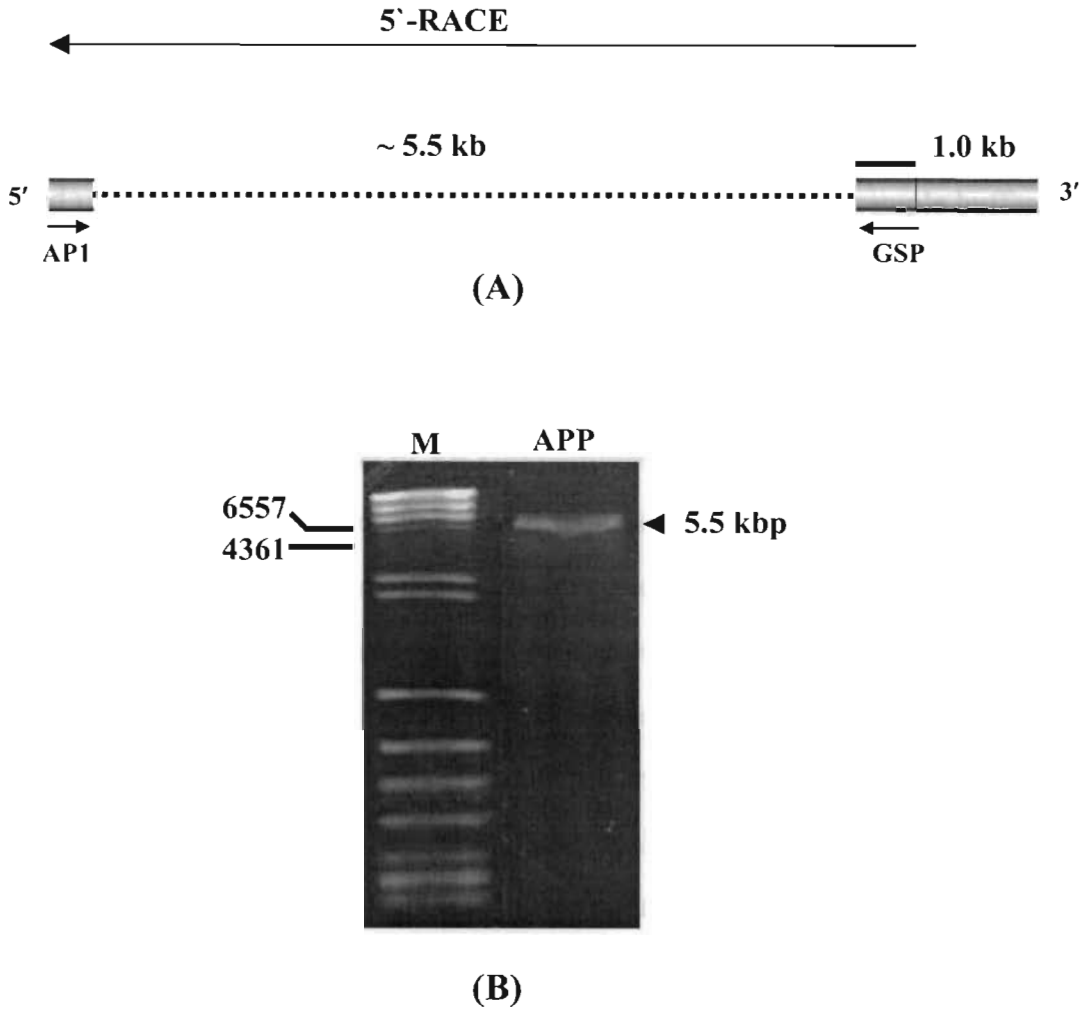


Fig. 2. 5′-end amplification of *L. maderae* Vg cDNA. A gene specific primer (GSP) was prepared corresponding to the 5`-end portion of the initial clone (1.0 kbp) shown on the 3` region (A), and was used with the adapter primer (AP1) to amplify the missing portion through RACE-PCR (see materials and methods). The amplified PCR-product (APP) was separated and identified on 1.2% agarose gel (B). Whereas, M is the molecular weight marker, and a bold line (B) is representing the overlaping region to identify the amplified portion (~5.5 kbp).

4.3.5 SDS-PAGE

SDS-PAGE was carried out as described previously (Tufail et al., 2000). Eggs (mature terminal oocytes) were isolated from the females and homogenized in the sample buffer (1 egg/600 µl) as described previously (Tufail et al., 2000). Hemolymph was collected with a micro pipette after the foreleg was amputated or the cervix was punctured using a fine pair of forceps. It was diluted at 1:50 with the sample buffer. A sample of 2 µl of egg extracts or 5 µl of the hemolymph was applied per lane on 7% polyacrylamide gels.

4.3.6 Amino-terminal Sequences of Vn Polypeptides

To determine the N-terminal amino acid sequence of Vn polypeptides, each egg (mature terminal oocyte) was homogenized in 200 µl of sample buffer and subjected to 5% SDS-PAGE (2-15 µl per lane), as mentioned above. The polypeptides so obtained were transferred to the PVDF membrane (Millipore, Immobilon) and stained with Ponceau S (0.2% in 1% acetic acid). The bands corresponding to the respective Vn were cut out and subjected to protein sequence analysis. Amino-terminal sequences were determined by Edman degradation with a gas phase amino acid sequencer (Perkin Elmer, 492 Procise).

4.3.7 Northern Blot Hybridization Analysis

Northern blot hybridization analysis for the total RNA extracted from the male and female fat body, and from the ovaries of *L. maderae* was conducted as reported previously (Tufail et al., 2000). Briefly the total RNA was applied (5 µg per lane) on 1% agarose gels containing formaldehyde. The gels were then transferred to Hybond N+ membranes (Amersham-Pharmacia). Probes were prepared using short fragments (nucleotide positions: 102-873; 4024-5168, see Figs. 3 and 4) of cloned Vg cDNA of *L. maderae*. DNA fragment (50 ng) was labeled using a Gene image random prime labeling kit (Amersham-Pharmacia). Hybridization and subsequent washes were

performed as described previously (Tufail et al., 2000), whereas signals were detected using a Gene image CDP-star detection module (Amersham-Pharmacia) following the protocols of the supplier.

4.3.8 Comparison with Vitellogenin Amino Acid Sequences of Other Cockroach Species and Phylogenetic Inference

The entire amino acid sequences (excluding the signal peptide sequences) of 4 Vgs including the two of *P. americana* (Vg1 and Vg2), one of *B. germanica* (Comas et al., 2000) and one of *L. maderae* were multiple-aligned and compared using the Clustal W computer program (Thompson et al., 1994). A molecular phylogenetic (neighbour-joining) tree was constructed using the MEGA version 2.1 (Kumar et al., 2001).

For comparison, we also prepared a phylogenetic tree, using the same computer program, based on the DNA sequences of the mitochondrial 12S rRNA genes of three cockroach species. The sequences used were those of *Rhyparobia* (=*Leucophaea*) *maderae* (accession no: U17826), *P. americana* (accession no: U17805) and *Blattella vaga* (accession no: U17776). All these sequences have been reported previously by Kambhampati (1995). For comparison we tried to use the same cockroach species, but unfortunately the sequence data for the mitochondrial 12S rRNA gene of *B. germanica* was not available, instead we used *B. vaga* which is biologically very closely related with the latter (Stay B., personal communication).

4.4 RESULTS AND DISCUSSION

4.4.1 Cloning and Sequence Analysis of the cDNA for *L. maderae* Vg

The *L. maderae* Vg cDNA was cloned using the simple and rapid method of cloning insect Vg cDNAs (Lee et al., 2000). Using 10 primers for the GL/ICG motif (Table 1) and the adaptor primer, 3' end portion was amplified through RACE-PCR.

The three amplified bands (about 1.0 kb in length), one each from the GLI/CG primer-P2, -P4 and -P7, were sequenced and analyzed. We analyzed three independent clones from each batch of the above three primers. The nucleotide sequences obtained from all clones (from three batches) were similar except for a very few polymorphic base substitutions, whereas the deduced amino acid sequences were exactly identical and all showed only the GLCG motif. The FASTA homology search on the DDBJ batabase showed that all had homology with Vgs of other insects. This suggested that the cloned cDNAs were all the 3' part of the same Vg cDNA of *L. maderae* and that each contained a sequence of 997 bp (nucleotide positions: 4924-5920) that encoded 276 amino acid residues (amino acid positions: 1638-1913) in a single open reading frame followed by a termination codon (TAA) and a consensus polyadenylation signal (AATAAA) (Fig. 4). One of the clones was thus selected, a gene-specific primer was prepared (see Fig. 4), and used with an adaptor ligated ds cDNA library, in addition to the adaptor primer. Then 5` portion was extended following RACE-method which was subcloned and sequenced. The final subcloning and sequencing strategy alongwith the final restriction map is shown in Fig.3.

The complete deduced amino acid sequence obtained from two overlapping RACE-PCR fragments was of 1,913 residues (Fig. 4) and agreed exactly with the pre-pro Vg molecular weight of 190 kD as demonstrated by Don-Wheeler and Engelmann (1997). Harmon (1994) who cloned a cDNA encoding the C-terminal portion of *L. maderae* Vg also reported that the size of *L. maderae* Vg mRNA was 5,400 nucleotides that encodes, theoretically, a protein of about 180 kD. A comparison between the size of Vg mRNA (5,400 nucleotides) reported by Harmon (1994) and the size of Vg cDNA (5,920 nucleotides) reported in the present study suggested that the hybridization signals he detected, to estimate the size of the protein or the protein primary product, were not the complete messages encoding *L. maderae* Vg and, thus, a smaller size of mRNA (only 5400 nucleotides) and lower weight of the product (only 180 kD) were estimated.

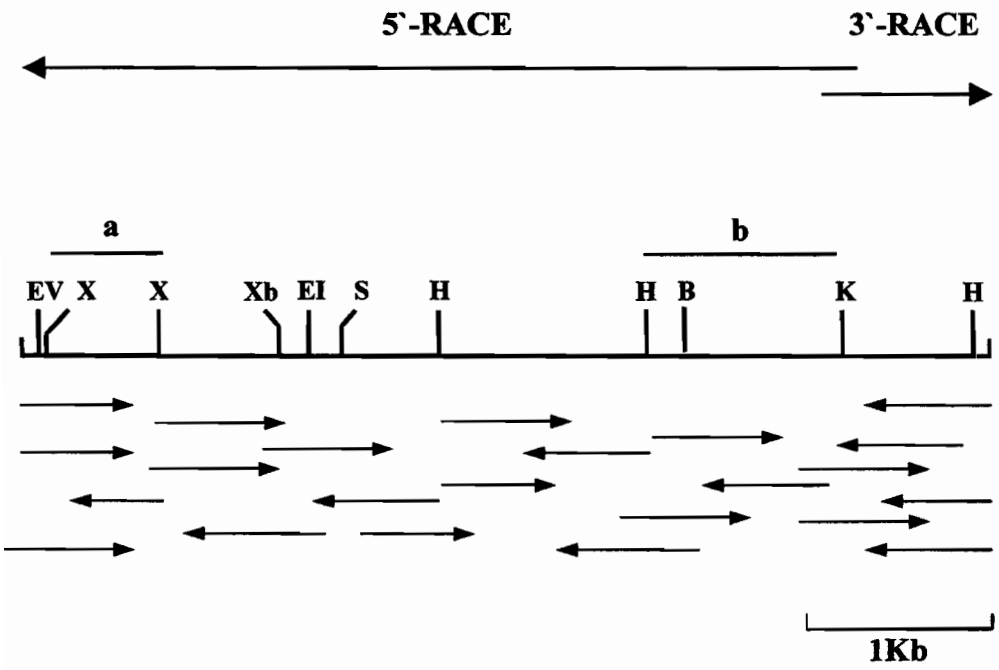
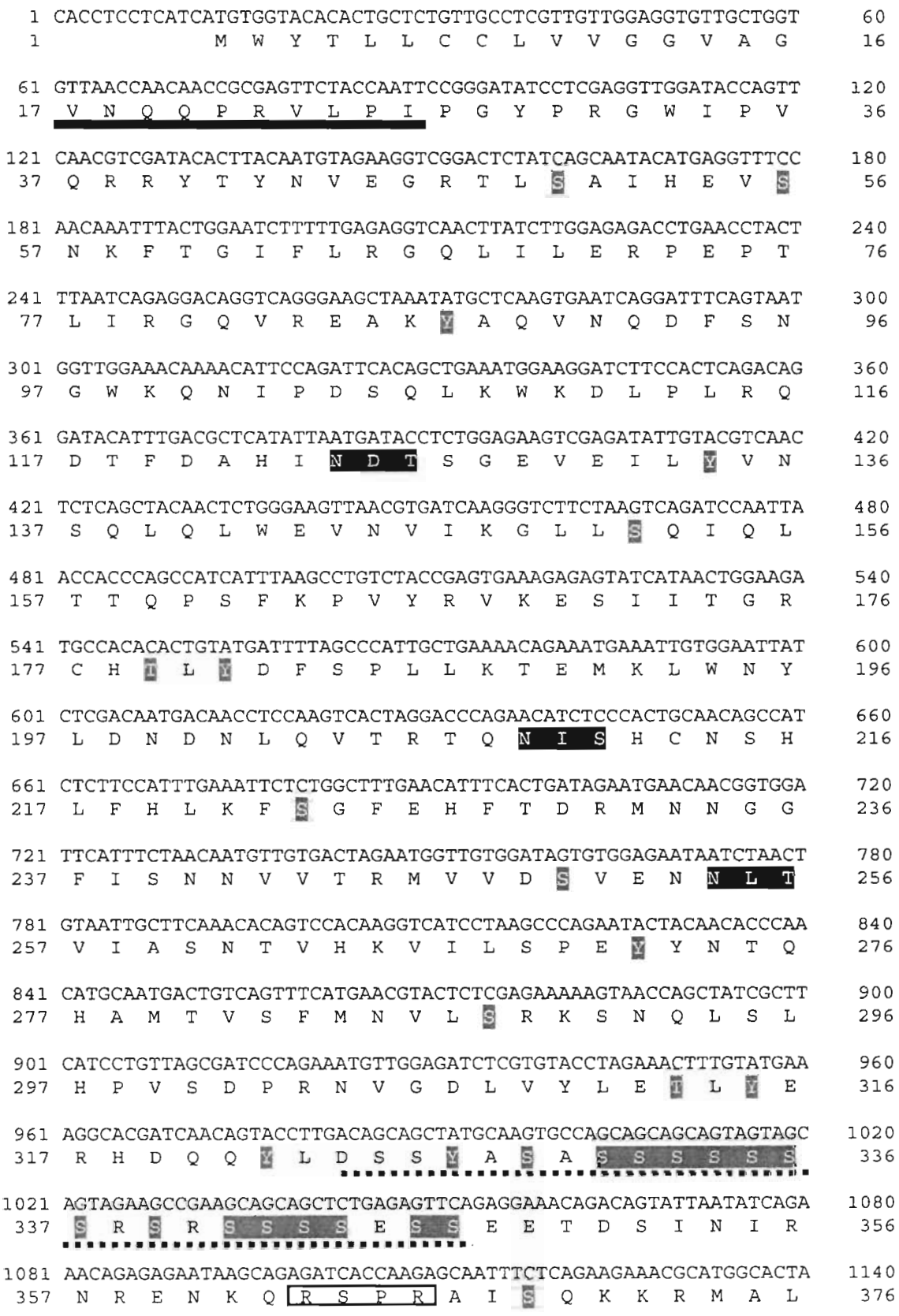
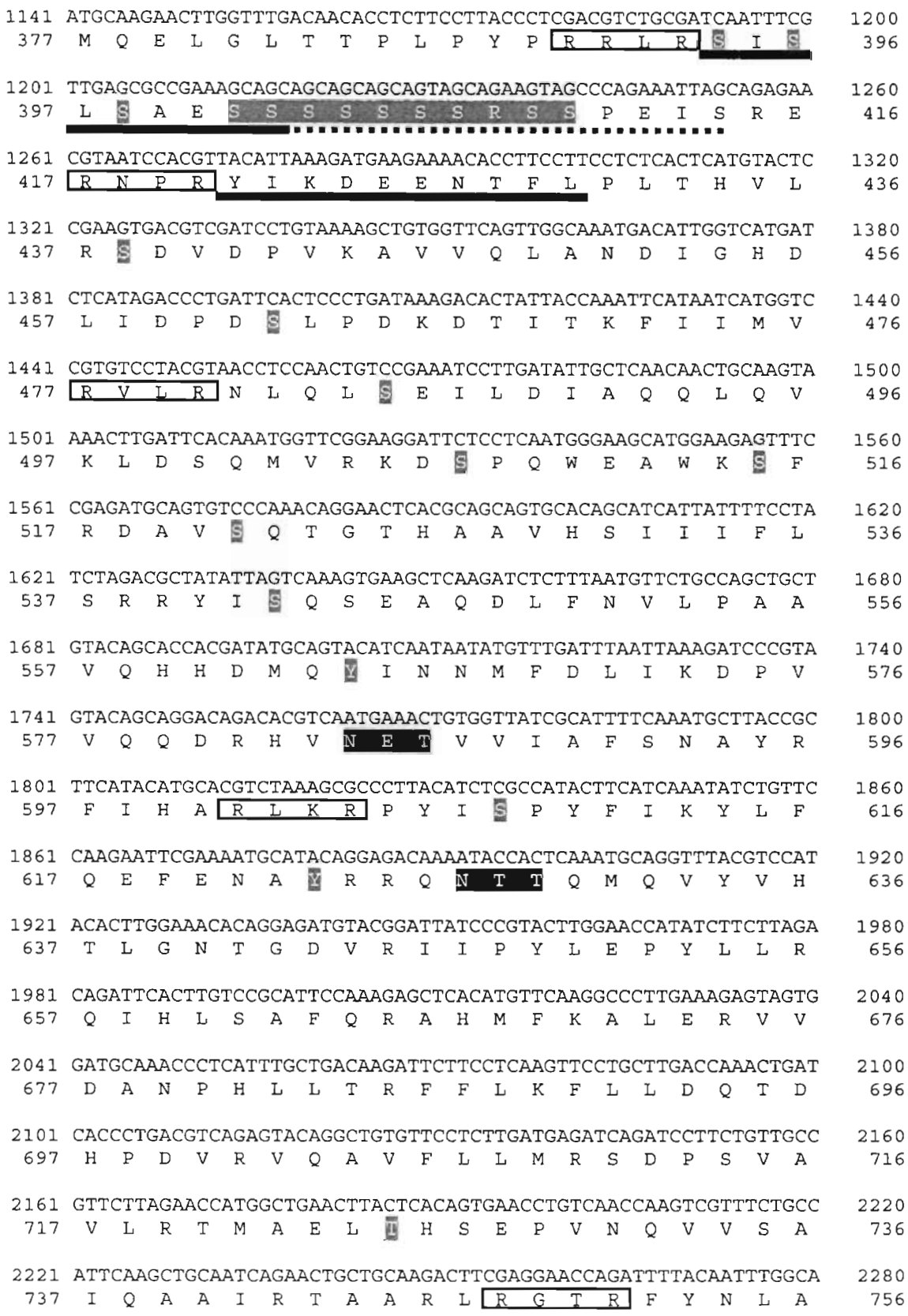


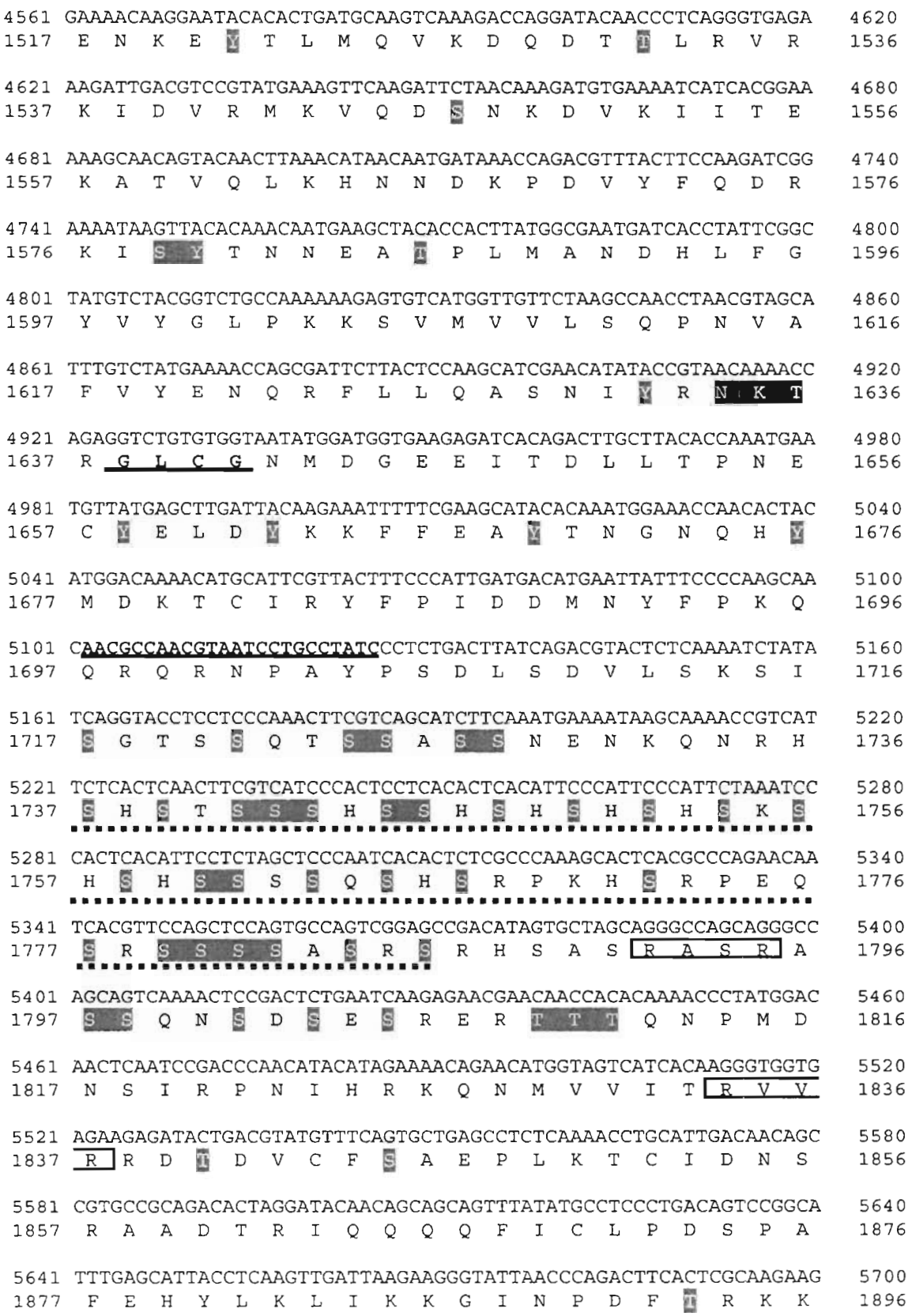
Fig. 3. Restriction map and sequence strategy of the cloned *L. maderae* Vg cDNA and the two probes (a and b) used for Northern blot analysis. EV, *EcoR*V; X, *Xho*I; Xb, *Xba*I; EI, *EcoR*I; S, *Sac*I; H, *Hind*III; B, *BamH*I and K, *Kpn*I.


1 CACCTCCTCATCATGTGGTACACACTGCTCTGTTGCCTCGTTGTTGGAGGTGTTGCTGGT 60
1 M W Y T L L C C L V V G G V A G 16
61 GTTAACCAACAACCGCGAGTTCTACCAATTCCGGGATATCCTCGAGGTTGGATACCAGTT 120
17 V N Q Q P R V L P I P G Y P R G W I P V 36
121 CAACGTCGATACACTTACAATGTAGAAGGTCGGACTCTATCAGCAATACATGAGGTTTCC 180
37 Q R R Y T Y N V E G R T L S A I H E V S 56
181 AACAAATTTACTGGAATCTTTTTGAGAGGTCAACTTATCTTGGAGAGACCTGAACCTACT 240
57 N K F T G I F L R G Q L I L E R P E P T 76
241 TTAATCAGAGGACAGGTCAGGGAAGCTAAATATGCTCAAGTGAATCAGGATTTCAGTAAT 300
77 L I R G Q V R E A K Y A Q V N Q D F S N 96
301 GGTTGGAAACAAAACATTCCAGATTCACAGCTGAAATGGAAGGATCTTCCACTCAGACAG 360
97 G W K Q N I P D S Q L K W K D L P L R Q 116
361 GATACATTTGACGCTCATATTAATGATACCTCTGGAGAAGTCGAGATATTGTACGTCAAC 420
117 D T F D A H I N D T S G E V E I L Y V N 136
421 TCTCAGCTACAACTCTGGGAAGTTAACGTGATCAAGGGTCTTCTAAGTCAGATCCAATTA 480
137 S Q L Q L W E V N V I K G L L S Q I Q L 156
481 ACCACCCAGCCATCATTTAAGCCTGTCTACCGAGTGAAAGAGAGTATCATAACTGGAAGA 540
157 T T Q P S F K P V Y R V K E S I I T G R 176
541 TGCCACACACTGTATGATTTTAGCCCATTGCTGAAAACAGAAATGAAATTGTGGAATTAT 600
177 C H T L Y D F S P L L K T E M K L W N Y 196
601 CTCGACAATGACAACCTCCAAGTCACTAGGACCCAGAACATCTCCCACTGCAACAGCCAT 660
197 L D N D N L Q V T R T Q N I S H C N S H 216
661 CTCTTCCATTTGAAATTCTCTGGCTTTGAACATTTCACTGATAGAATGAACAACGGTGGA 720
217 L F H L K F S G F E H F T D R M N N G G 236
721 TTCATTTCTAACAATGTTGTGACTAGAATGGTTGTGGATAGTGTGGAGAATAATCTAACT 780
237 F I S N N V V T R M V V D S V E N N L T 256
781 GTAATTGCTTCAAACACAGTCCACAAGGTCATCCTAAGCCCAGAATACTACAACACCCAA 840
257 V I A S N T V H K V I L S P E Y Y N T Q 276
841 CATGCAATGACTGTCAGTTTCATGAACGTACTCTCGAGAAAAAGTAACCAGCTATCGCTT 900
277 H A M T V S F M N V L S R K S N Q L S L 296
901 CATCCTGTTAGCGATCCCAGAAATGTTGGAGATCTCGTGTACCTAGAAACTTTGTATGAA 960
297 H P V S D P R N V G D L V Y L E T L Y E 316
961 AGGCACGATCAACAGTACCTTGACAGCAGCTATGCAAGTGCCAGCAGCAGCAGTAGTAGC 1020
317 R H D Q Q Y L D S S Y A S A S S S S S S 336
1021 AGTAGAAGCCGAAGCAGCAGCTCTGAGAGTTCAGAGGAAACAGACAGTATTAATATCAGA 1080
337 S R S R S S S S E S S E E T D S I N I R 356
1081 AACAGAGAGAATAAGCAGAGATCACCAAGAGCAATTTCTCAGAAGAAACGCATGGCACTA 1140
357 N R E N K Q R S P R A I S Q K K R M A L 376


1141 ATGCAAGAACTTGGTTTGACAACACCTCTTCCTTACCCTCGACGTCTGCGATCAATTTCG 1200
377 M Q E L G L T T P L P Y P R R L R S I S 396
1201 TTGAGCGCCGAAAGCAGCAGCAGCAGCAGTAGCAGAAGTAGCCCAGAAATTAGCAGAGAA 1260
397 L S A E S S S S S S S S R S S P E I S R E 416
1261 CGTAATCCACGTTACATTAAAGATGAAGAAAACACCTTCCTTCCTCTCACTCATGTACTC 1320
417 R N P R Y I K D E E N T F L P L T H V L 436
1321 CGAAGTGACGTCGATCCTGTAAAAGCTGTGGTTCAGTTGGCAAATGACATTGGTCATGAT 1380
437 R S D V D P V K A V V Q L A N D I G H D 456
1381 CTCATAGACCCTGATTCACTCCCTGATAAAGACACTATTACCAAATTCATAATCATGGTC 1440
457 L I D P D S L P D K D T I T K F I I M V 476
1441 CGTGTCCTACGTAACCTCCAACTGTCCGAAATCCTTGATATTGCTCAACAACTGCAAGTA 1500
477 R V L R N L Q L S E I L D I A Q Q L Q V 496
1501 AAACTTGATTCACAAATGGTTCGGAAGGATTCTCCTCAATGGGAAGCATGGAAGAGTTTC 1560
497 K L D S Q M V R K D S P Q W E A W K S F 516
1561 CGAGATGCAGTGTCCCAAACAGGAACTCACGCAGCAGTGCACAGCATCATTATTTTCCTA 1620
517 R D A V S Q T G T H A A V H S I I I F L 536
1621 TCTAGACGCTATATTAGTCAAAGTGAAGCTCAAGATCTCTTTAATGTTCTGCCAGCTGCT 1680
537 S R R Y I S Q S E A Q D L F N V L P A A 556
1681 GTACAGCACCACGATATGCAGTACATCAATAATATGTTTGATTTAATTAAAGATCCCGTA 1740
557 V Q H H D M Q Y I N N M F D L I K D P V 576
1741 GTACAGCAGGACAGACACGTCAATGAAACTGTGGTTATCGCATTTTCAAATGCTTACCGC 1800
577 V Q Q D R H V N E T V V I A F S N A Y R 596
1801 TTCATACATGCACGTCTAAAGCGCCCTTACATCTCGCCATACTTCATCAAATATCTGTTC 1860
597 F I H A R L K R P Y I S P Y F I K Y L F 616
1861 CAAGAATTCGAAAATGCATACAGGAGACAAAATACCACTCAAATGCAGGTTTACGTCCAT 1920
617 Q E F E N A Y R R Q N T T Q M Q V Y V H 636
1921 ACACTTGGAAACACAGGAGATGTACGGATTATCCCGTACTTGGAACCATATCTTCTTAGA 1980
637 T L G N T G D V R I I P Y L E P Y L L R 656
1981 CAGATTCACTTGTCCGCATTCCAAAGAGCTCACATGTTCAAGGCCCTTGAAAGAGTAGTG 2040
657 Q I H L S A F Q R A H M F K A L E R V V 676
2041 GATGCAAACCCTCATTTGCTGACAAGATTCTTCCTCAAGTTCCTGCTTGACCAAACTGAT 2100
677 D A N P H L L T R F F L K F L L D Q T D 696
2101 CACCCTGACGTCAGAGTACAGGCTGTGTTCCTCTTGATGAGATCAGATCCTTCTGTTGCC 2160
697 H P D V R V Q A V F L L M R S D P S V A 716
2161 GTTCTTAGAACCATGGCTGAACTTACTCACAGTGAACCTGTCAACCAAGTCGTTTCTGCC 2220
717 V L R T M A E L T H S E P V N Q V V S A 736
2221 ATTCAAGCTGCAATCAGAACTGCTGCAAGACTTCGAGGAACCAGATTTTACAATTTGGCA 2280
737 I Q A A I R T A A R L R G T R F Y N L A 756

```
2281 TTCAAAGCACAGACTGTTGTGAATCTGCTGAGCGATAAGAACTTGGATGTGTCATACTCC 2340
 757  F  K  A  Q  T  V  V  N  L  L  S  D  K  N  L  D  V  S  Y  S   776

2341 AAGAATTATATGCTAGATCAAGAAGCTAGGGAATATAACCTCGATTTCCAACTCTTCTAC 2400
 777  K  N  Y  M  L  D  Q  E  A  R  E  Y  N  L  D  F  Q  L  F  Y   796

2401 GAACAAATTGGAAGTCAAGACAATCTATTACCAAAATCTGCTCTCCTTGACATTTTCTCT 2460
 797  E  Q  I  G  S  Q  D  N  L  L  P  K  S  A  L  L  D  I  F  S   816

2461 TATGTTGGAGGTGCAAAATCTGACCACCAAACTGGATACACAGTATCAAGTATTGACAAA 2520
 817  Y  V  G  G  A  K  S  D  H  Q  T  G  Y  T  V  S  S  I  D  K   836

2521 GTACTCAATGATATTCAACTGCAATTCAAGAACTTCACACAGCAAGGATGGACACAACAA 2580
 837  V  L  N  D  I  Q  L  Q  F  K  N  F  T  Q  Q  G  W  T  Q  Q   856

2581 CAGATGAGGGATGATTTAACGAAGCTTGTAGAAGGAAACATTCAATACCAAGTCTTGGGT 2640
 857  Q  M  R  D  D  L  T  K  L  V  E  G  N  I  Q  Y  Q  V  L  G   876

2641 GTTCAGAGGTTCTGGCCATTTGACCAAGATAGCATTAAAAGCATTCCTAATGTAATACAA 2700
 877  V  Q  R  F  W  P  F  D  Q  D  S  I  K  S  I  P  N  V  I  Q   896

2701 AAATTCGTCAAAGACTACAGGGAAGTTAAGTCATTCAACCTCACAAAGTTCTTCACCACC 2760
 897  K  F  V  K  D  Y  R  E  V  K  S  F  N  L  T  K  F  F  T  T   916

2761 TCCACTGGAATCTACGGTTTCCCAACCGTCATGGGCTTCCCAGGAGTATACACACTGCAT 2820
 917  S  T  G  I  Y  G  F  P  T  V  M  G  F  P  G  V  Y  T  L  H   936

2821 ACCCCATCACTGTGGAAAGCAGATGGTGAATTAAAAGTGACTACTGTGCCAGATCTTGAA 2880
 937  T  P  S  L  W  K  A  D  G  E  L  K  V  T  T  V  P  D  L  E   956

2881 CAAAATCCACATTATCTCCCAGGAATCGTGGATGTACAACTGAGAGTTCGTCCTCTGTAT 2940
 957  Q  N  P  H  Y  L  P  G  I  V  D  V  Q  L  R  V  R  P  L  Y   976

2941 GCTGCCAAACTTCAGAGCAAACTGAGCGTCATCACACCTTTCAATGATATGAGGTACACT 3000
 977  A  A  K  L  Q  S  K  L  S  V  I  T  P  F  N  D  M  R  Y  T   996

3001 GCTGGTGTCAACAGGCACTTCCAATTACACGTTCCCTTCCAAGTGAAAATTCACGCCGAA 3060
 997  A  G  V  N  R  H  F  Q  L  H  V  P  F  Q  V  K  I  H  A  E  1016

3061 ATGAACTACAACAATATGAATGATGACAAAAATAATAACCTCTACGCTATCAGAGCCAAT 3120
1017  M  N  Y  N  N  M  N  D  D  K  N  N  N  L  Y  A  I  R  A  N  1036

3121 GTTAAGCAATATGACAGTGACAAGGACCACAGAGTACTATACATGAGTTCTATTCCTTTC 3180
1037  V  K  Q  Y  D  S  D  K  D  H  R  V  L  Y  M  S  S  I  P  F  1056

3181 ACAACCATACATGACATTAGGAGTTTGAACCCTGATTCGAAGGACGACGATTTCCAGATT 3240
1057  T  T  I  H  D  I  R  S  L  N  P  D  S  K  D  D  D  F  Q  I  1076

3241 CTTCATGTACGACCAATGAAGAAGTATAACAGAGATTATGGACAGGATGTCGGACAAGCA 3300
1077  L  H  V  R  P  M  K  K  Y  N  R  D  Y  G  Q  D  V  G  Q  A  1096

3301 GTCAAGCTACAGTATGAAACTGAGGACGATTACGTGGACTTGAAATTGAGCAACAAAAAC 3360
1097  V  K  L  Q  Y  E  T  E  D  D  Y  V  D  L  K  L  S  N  K  N  1116

3361 ATATGGCTCTCAAATGCCCTTTCTCCACTGCCTGGATGGATGACATCACAAATTATCTAC 3420
1117  I  W  L  S  N  A  L  S  P  L  P  G  W  M  T  S  Q  I  I  Y  1136
```

```
3421 AGGAAACTCGGCGTTACCTATGCCACCAGACAATGCACAAACAATGTCATAGAGCTTTCA 3480
1137  R  K  L  G  V  T  Y  A  T  R  Q  C  T  N  N  V  I  E  L  S  1156

3481 GCAGTGCTTGCAAATGACCAACAAAACAACCAATACCCTAATACTCAGAATGATGATGGA 3540
1157  A  V  L  A  N  D  Q  Q  N  N  Q  Y  P  N  T  Q  N  D  D  G  1176

3541 CATTCAGCAAGGAAACACAAGGCCCGCAGGACAAGAAGTGCAAGGAAAGATGATCGTCAA 3600
1177  H  S  A  R  K  H  K  A  R  R  T  R  S  A  R  K  D  D  R  Q  1196

3601 TCAAGTGGGGAAAGATCAGACTCGAACCCTGCAATTCCCTCTGACACAAAGCCTGACAGT 3660
1197  S  S  G  E  R  S  D  S  N  P  A  I  P  S  D  T  K  P  D  S  1216

3661 GATGCCAGAAGACAACAATACCTTCGTGCAGCACGAGAACAAGTTAGAAACAATGCAAGC 3720
1217  D  A  R  R  Q  Q  Y  L  R  A  A  R  E  Q  V  R  N  N  A  S  1236

3721 TCTTACGTTCTTGACTTGGGAGTGAACTTCAAAGGACAGAATCCAGCTTACATAGTATTC 3780
1237  S  Y  V  L  D  L  G  V  N  F  K  G  Q  N  P  A  Y  I  V  F  1256

3781 ACTGGAGCATATGCAAAATCCCTTGTAAATGGAAATTCAAATCATTTGGTTTTCTACAAT 3840
1257  T  G  A  Y  A  K  S  L  V  N  G  N  S  N  H  L  V  F  Y  N  1276

3841 CAACAGTTTCTCAAACCTGAAGACAACAAACAGGTTTGTTTGAGTGCTAACATCATGAAA 3900
1277  Q  Q  F  L  K  P  E  D  N  K  Q  V  C  L  S  A  N  I  M  K  1296

3901 CCACAAATGCCACTCAACAACTACGACGATGCTCTACAATCTGACCCAACATCACAAGTG 3960
1297  P  Q  M  P  L  N  N  Y  D  D  A  L  Q  S  D  P  T  S  Q  V  1316

3961 AGAATGATATTGAATGCTGGAAACAAATGCCAAGAAGGAAGTGGACAAGCTACTGTAGAG 4020
1317  R  M  I  L  N  A  G  N  K  C  Q  E  G  S  G  Q  A  T  V  E  1336

4021 GAAAGCTTCAAAGAACTAAGGAGTATGAAAAATTCATCAAAGACTGGGCATTGGCTAGGG 4080
1337  E  S  F  K  E  L  R  S  M  K  N  S  S  K  T  G  H  W  L  G  1356

4081 AATGTCAGAATGACATGGACAAGTACCGCAATTATCTCCTCCGAGCACTGCCAGAACGTC 4140
1357  N  V  R  M  T  W  T  S  T  A  I  I  S  S  E  H  C  Q  N  V  1376

4141 ACTTACCGTGCCGACGATCTCAAGGATTATACATTCAGAGCCATCTATGATAACAAGTTA 4200
1377  T  Y  R  A  D  D  L  K  D  Y  T  F  R  A  I  Y  D  N  K  L  1396

4201 CCCGATTTTGTCAAAGAGAGGTTGTATCAGGCATATGCTCTTCTACGTAACAGACTTCAC 4260
1397  P  D  F  V  K  E  R  L  Y  Q  A  Y  A  L  L  R  N  R  L  H  1416

4261 CGACATGTTTCAGAGGATCCTTTCAAGATTAAAGCCAACAGTGGTCAGTTGGACCTCTCA 4320
1417  R  H  V  S  E  D  P  F  K  I  K  A  N  S  G  Q  L  D  L  S  1436

4321 GTACAACTCAACAATGTCAGCAAAGTTTTCAACTTGACTCTGGAATCGGCTCTTGGAGAA 4380
1437  V  Q  L  N  N  V  S  K  V  F  N  L  T  L  E  S  A  L  G  E  1456

4381 TCAAGATTTATCAATGTACCAGTCCACGACTGGGCCGGCAACATGTTGAGCGTAAACCCA 4440
1457  S  R  F  I  N  V  P  V  H  D  W  A  G  N  M  L  S  V  N  P  1476

4441 AGGACTTCAATTGCAGAACGACTTGCCCAGTACGAACTTCCTCTCTATAACAACCCAACA 4500
1477  R  T  S  I  A  E  R  L  A  Q  Y  E  L  P  L  Y  N  N  P  T  1496

4501 TGTGCTCTGGATAACAGTGCTATTAACACTTTTGACAATTTAACAATCTACAACCGCTTT 4560
1497  C  A  L  D  N  S  A  I  N  T  F  D  N  L  T  I  Y  N  R  F  1516
```


4561 GAAAACAAGGAATACACACTGATGCAAGTCAAAGACCAGGATACAACCCTCAGGGTGAGA 4620
1517 E N K E Y T L M Q V K D Q D T T L R V R 1536
4621 AAGATTGACGTCCGTATGAAAGTTCAAGATTCTAACAAAGATGTGAAAATCATCACGGAA 4680
1537 K I D V R M K V Q D S N K D V K I I T E 1556
4681 AAAGCAACAGTACAACTTAAACATAACAATGATAAACCAGACGTTTACTTCCAAGATCGG 4740
1557 K A T V Q L K H N N D K P D V Y F Q D R 1576
4741 AAAATAAGTTACACAAACAATGAAGCTACACCACTTATGGCGAATGATCACCTATTCGGC 4800
1576 K I S Y T N N E A T P L M A N D H L F G 1596
4801 TATGTCTACGGTCTGCCAAAAAAGAGTGTCATGGTTGTTCTAAGCCAACCTAACGTAGCA 4860
1597 Y V Y G L P K K S V M V V L S Q P N V A 1616
4861 TTTGTCTATGAAAACCAGCGATTCTTACTCCAAGCATCGAACATATACCGTAACAAAACC 4920
1617 F V Y E N Q R F L L Q A S N I Y R N K T 1636
4921 AGAGGTCTGTGTGGTAATATGGATGGTGAAGAGATCACAGACTTGCTTACACCAAATGAA 4980
1637 R G L C G N M D G E E I T D L L T P N E 1656
4981 TGTTATGAGCTTGATTACAAGAAATTTTTCGAAGCATACACAAATGGAAACCAACACTAC 5040
1657 C Y E L D Y K K F F E A Y T N G N Q H Y 1676
5041 ATGGACAAAACATGCATTCGTTACTTTCCCATTGATGACATGAATTATTTCCCCAAGCAA 5100
1677 M D K T C I R Y F P I D D M N Y F P K Q 1696
5101 CAACGCCAACGTAATCCTGCCTATCCCTCTGACTTATCAGACGTACTCTCAAAATCTATA 5160
1697 Q R Q R N P A Y P S D L S D V L S K S I 1716
5161 TCAGGTACCTCCTCCCAAACTTCGTCAGCATCTTCAAATGAAAATAAGCAAAACCGTCAT 5220
1717 S G T S S Q T S S A S S N E N K Q N R H 1736
5221 TCTCACTCAACTTCGTCATCCCACTCCTCACACTCACATTCCCATTCCCATTCTAAATCC 5280
1737 S H S T S S S H S S H S H S H S H S K S 1756
5281 CACTCACATTCCTCTAGCTCCCAATCACACTCTCGCCCAAAGCACTCACGCCCAGAACAA 5340
1757 H S H S S S S Q S H S R P K H S R P E Q 1776
5341 TCACGTTCCAGCTCCAGTGCCAGTCGGAGCCGACATAGTGCTAGCAGGGCCAGCAGGGCC 5400
1777 S R S S S S A S R S R H S A S R A S R A 1796
5401 AGCAGTCAAAACTCCGACTCTGAATCAAGAGAACGAACAACCACACAAAACCCTATGGAC 5460
1797 S S Q N S D S E S R E R T T T Q N P M D 1816
5461 AACTCAATCCGACCCAACATACATAGAAAACAGAACATGGTAGTCATCACAAGGGTGGTG 5520
1817 N S I R P N I H R K Q N M V V I T R V V 1836
5521 AGAAGAGATACTGACGTATGTTTCAGTGCTGAGCCTCTCAAAACCTGCATTGACAACAGC 5580
1837 R R D T D V C F S A E P L K T C I D N S 1856
5581 CGTGCCGCAGACACTAGGATACAACAGCAGCAGTTTATATGCCTCCCTGACAGTCCGGCA 5640
1857 R A A D T R I Q Q Q Q F I C L P D S P A 1876
5641 TTTGAGCATTACCTCAAGTTGATTAAGAAGGGTATTAACCCAGACTTCACTCGCAAGAAG 5700
1877 F E H Y L K L I K K G I N P D F T R K K 1896

```
5701 AATTTCGTACAGCTGGAGGTCAAGATTCCAACCAAATGCATTAAGAGTCAATAAATACAA   5760
1897  N  F  V  Q  L  E  V  K  I  P  T  K  C  I  K  S  Q  *         1913

5761 GCTTTCCATCCAGCTGCCTCTTTAATGATAATGACAAAGCACGTTACATGATGATTCACC   5820

5821 TGTATAAATTAATGAAACAACTGTAACTTTTATTTATATATTTAATGAATAAACTGAACT   5880

5881 GCAAATTGCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA                       5920
```

Fig. 4. Nucleotide sequence and deduced amino acid sequence of the vitellogenin cDNA of *L. maderae*. Consensus RXXR sequences for possible cleavage sites are boxed and clusters of serine residues are dot-underlined. The GLI/CG motif and gene specific primer are shown with underlined, boldface letters. Possible glycosylation sites are shown with dark-shaded boxes. Possible phosphorylated serine (S), threonine (T) and tyrosine (Y) residues as predicted using the NetPhos 2.0 Computer Program are shown with light-shaded frames. A consensus polyadenylation signal is double underlined. Asterik indicates the stop codon. The chemically determined N-terminal amino acid sequences for the 55-, 100-, 92-, 90- and 112-kD Vn polypeptides are underlined with a bold line. The sequence for the 55-kD Vn polypeptide is found following the putative signal peptide, that the 100-kD polypeptide is found following the RRLR cleavage site. The 92- and 90-kD polypeptides share sequences and are found following the RNPR cleavage site (see text), whereas, that for the 112-kD polypeptide is present following the RRTR cleavage site. This sequence has been submitted to GenBank and assigned the accession number: AB052640.

The analysis of the deduced amino acid sequence revealed that the first 15 amino acids would correspond to a signal peptide as predicted by using the SignalP VI.I computer

program (Nielsen et al., 1997). There were 11 putative cleavage sites showing the RXXR consensus sequence (see Sappington and Raikhel, 1998) and 13 putative glycosylation sites. Moreover, the putative phosphorylated serine, threonine and tyrosine residues have also been observed in the amino acid sequence of *L. maderae* Vg (Fig. 1), similar to that of *P. americana* (Tufail, et al., 2001), which shows that the Vg molecule described herein is highly phosphorylated, as was demonstrated previously (Della-Cioppa and Engelmann, 1987; Don-Wheeler and Engelmann, 1991; Don-Wheeler and Engelmann, 1997). In addition, the existence of certain remarkable motifs, such as polyserine domains, three in the present sequence (two near the N-terminus and one long fused stretch near the C-terminus), the GL/ICG motif (amino acid positions: 1638-1641) (see Fig. 4), and the cysteine residues at conserved locations near the C-terminal, similar to those of present in other insect Vgs (see Chen et al., 1997; Lee et al., 2000; Tufail et al., 2000; Comas et al., 2000) was noted in the amino acid sequence of *L. maderae* Vg, whereas the DGXR motif was absent in this insect species. Indeed, the DGXR motif was conserved in all insect Vgs examined (Tufail et al., 2000) but its absence in *L. maderae* Vg is unclear.

4.4.2 The Vns/Vgs of *L. maderae*

The yolk proteins of *L. maderae* were identified through SDS-PAGE by comparing the polypeptide profiles of adult male and female hemolymph and of egg (terminal mature oocyte) extracts. SDS-PAGE (Fig. 5) revealed four Vn polypeptides (with apparent molecular weights: 112-, 100-, 92- and 55-kD) corresponding to the four Vg polypeptides, whereas these polypeptides were absent in the male hemolymph. The Vn/Vg polypeptides

so detected were suggested to be yolk proteins. There was, however, an additional

90-kD polypeptide found in the egg extracts. These results obtained were in agreement with those reported previously (Engelmann, 1979; Della-Cioppa and Engelmann, 1987; Don-Wheeler and Engelmann, 1991; Don-Wheeler and Engelmann, 1997).

4.4.3 Amino-terminal Sequences of Vn Polypeptides and Processing Patterns of *L. maderae* Vgs/Vns

The N-terminal amino acid sequence chemically determined each for the 112-, 100-, 92-, 90- and 55-kD Vn polypeptide matched completely with the deduced sequence from the clone (see Figs. 4 and 6). These sequences were SARKDDRQSSGE (amino acid position: 1189-1200), following the RRTR cleavage site, SISLSAESS (amino acid position: 394-402), following the RRLR cleavage site and YIKDEE (amino acid position: 421-425), following the RNPR cleavage site for the 112-, 100-, and 92-kD Vn polypeptides respectively (Figs. 4 and 6) where the post-translational cleavage occurs in the fat body cells. The sequence for the 55-kD Vn polypeptide was VNQQPRVLPI and found the following putative signal peptide. The determined N-terminal amino acid sequence of the 90-kD Vn polypeptide (YIKDEENTFLPL) was identical to that of the 92-kD polypeptide and was also found following the RNPR cleavage site (amino acid position: 421-433) (Figs. 4 and 6) which demonstrates that the 92-kD Vn polypeptide is further processed to a 90-kD fragment after incorporation into the oocyte, and that the 90-kD fragment must come from the cleavage of the 92-kD polypeptide at the C-terminus. In other words the missing 2-kD must be the C-terminus of the 92-kD Vn polypeptide, but the sequence at the cleavage site is not known. It does not appear, however, that an RXXR motif (see Fig. 4) is in the appropriate vicinity of where the C-terminal cleavage site must be, arguing that a different type of enzyme is involved in the oocyte than the dibasic proteases that recognize the RXXR consensus cleavage site sequences.

Fig. 6 summarizes the proposed processing steps of *L. maderae* Vgs/Vns. The Vg molecule is first cleaved completely at site A into 155 and 112 kD polypeptides as

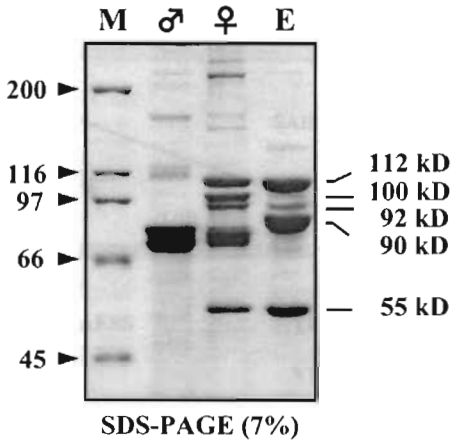


Fig. 5. SDS-PAGE analysis of Vns/Vgs from adult male and female hemolymph, and from the egg (terminal mature oocytes) extracts (E) of *L. maderae*. The gel was stained with Coomassie blue. M is molecular weight markers (kD). Whereas identified Vns/Vgs are indicated by lines on the right.

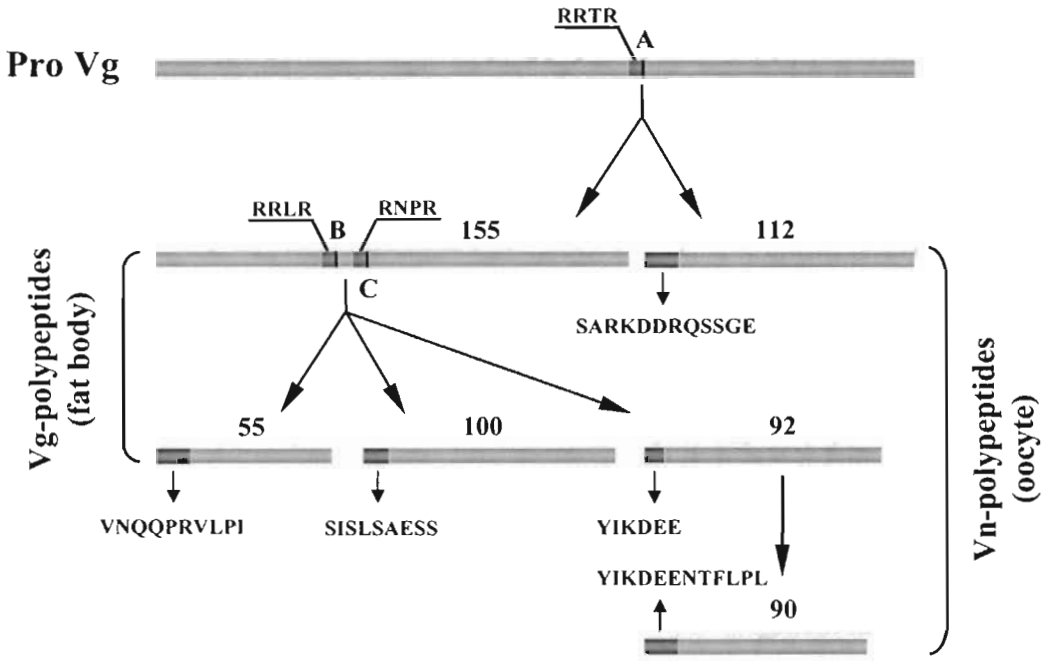


Fig. 6. Diagramatic representation of the proposed processing patterns of Vg/Vn in the fat body and oocyte of *L. maderae*. Numbers indicate the molecular sizes (kD) of the polypeptides. Letters indicate the cleavage sites on the amino acid sequence. Pro-vitellogenin (Pro Vg) is cleaved into four polypeptides (112, 55, 100 and 92 kD) in the fat body before being secreted into the hemolymph (see text). The Vn in the oocyte is further processed and results in the production of an additional 90 kD Vn polypeptide. The determined N-terminal amino acid sequences of the Vn polypeptides are shown with shaded boxes and arrows, whereas, the determined cleavage sites having RXXR consensus sequence are shown with shaded areas just before the cleavage sites and are raised and under-lined.

suggested by Della-Cioppa and Engelmann (1987), and Don-Wheeler and Engelmann (1997). The 155 kD polypeptide is then further processed at sites B and C resulting in the production of 55, 100 and 92 kD polypeptides (Fig. 6) in the fat body before being secreted into the hemolymph. The 155 kD polypeptide, however, was not observed on the SDS-PAGE gels. The latter would be a short-lived polypeptide, as suggested by Don-Wheeler and Engelmann (1997), and processed soon after its appearance. It is to be noted that the complete cleavage at site B would first produce a 55 kD polypeptide and then the incomplete cleavage at site C gives two more polypeptides (100 and 92 kD). The incomplete cleavage of Vg molecules appears to be common in cockroaches, as was also observed in Vgs of *P. americana* (Tufail, et al., 2001). The 92-kD Vn polypeptide is then further processed to 90-kD Vn polypeptide after incorporation into the oocyte. The matching of N-terminal amino acid sequences of both (92- and 90-kD) Vn polypeptides (Figs. 4 and 6) demonstrates that the 90-kD fragment must come from the cleavage of the 92-kD polypeptide at the C-terminus. However, nothing is known about the cleaved C-terminal (2-kD) portion of the 92-kD polypeptide. Moreover, the 100-kD fragment may be incompletely processed at cleavage site C in the egg, and the 92-kD fragment may be incompletely processed to 90-kD at the C-terminal site. This would account for the 100- and 92-kD Vn polypeptides being minor components of the egg extracts (see Fig. 2) as was also reported by Don-Wheeler and Engelmann (1991), and Don-Wheeler and Engelmann (1997).

4.4.4 Expression of the Vitellogenin Gene

Northern blot analysis was carried out to study the tissue and gender specificity of the Vg gene expression using the total RNAs extracted from the fat body of adult males and females (5 days old) and from the ovaries of adult females (10 days old). A short fragment of *L. maderae* Vg cDNA was flourescein-labled and used as a probe. The *L. maderae* Vg

gene was detected as a single 6.6 kb band (Fig. 7). The gene was expressed in the

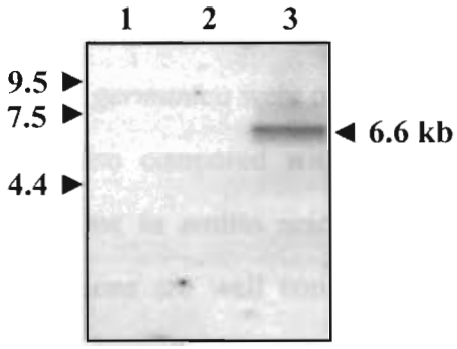


Fig. 7. Vitellogenin gene expression in *L. maderae* as detected by Northern blot hybridization analysis. The results were obtained by using a short fragment of *L. maderae* Vg cDNA (nucleotide positions 4024-5168) as a probe against total RNA from 5-days-old adult male (Lane 1) and female fat bodies (Lane 3) and from ovaries of 10-days-old adult females (Lane 2). The mobility and size (kb) of a standard RNA marker are shown on the left.

female fat body cells only. No trace of the hybridization signal was detected from male fat body or from ovaries, despite the amount of total RNAs for them being twice as much as that of the female fat body RNAs.

4.4.5 Comparison with Vitellogenin Sequences of other Cockroach Species and an Estimated Molecular Phylogenetic Tree

The Vg amino acid sequences (excluding the signal peptides) of *L. maderae, P. americana* (Vg1 and Vg2) and *B. germanica* were multiple-aligned using the Clustal W computer program, and were also compared with other insect Vgs (Fig. 8). The comparison of polyserine domains in amino acid sequences from three cockroach species suggests that their positions are well conserved at both termini (Fig. 8) in contrast to the other insect Vgs where these domains are either absent altogether or present only at the N-termini except for that of *Aedes aegypti* (see Fig. 8). Moreover, a highly conserved GL/ICG motif existed in all Vg sequences. The only difference from the others was that the DGXR motif was absent in *L. maderae* Vg. The comparison of the overall amino acid identity ratio (%) in the Vg amino acid sequence of *L. maderae* with those of other cockroach species revealed that the identity ratio was high within the members of the same superfamily (such as *B. germanica* : 31%) than with the members of the other superfamily (such as *P. americana* -Vg1 and -Vg2 : 26%), which reveals that the percentage of identity is directly proportional to the phylogenetic relationship and vice versa.

The present homology analysis based on 4 Vg amino acid sequences of 3 cockroach species has provided some phylogenetic inferences (Fig.9, A-left). In this analysis we observe that the two Vgs (Vg1 and vg2) of *P. americana* representing Blattidae (Blattoidea) cluster together, whereas the Vgs from two species of Blaberoidea, *L. maderae* and *B. germanica* representing Blaberidae and Blattellidae, respectively, cluster

together. The present results, thus, reveal, on a molecular basis, that *L. maderae* and *B.*

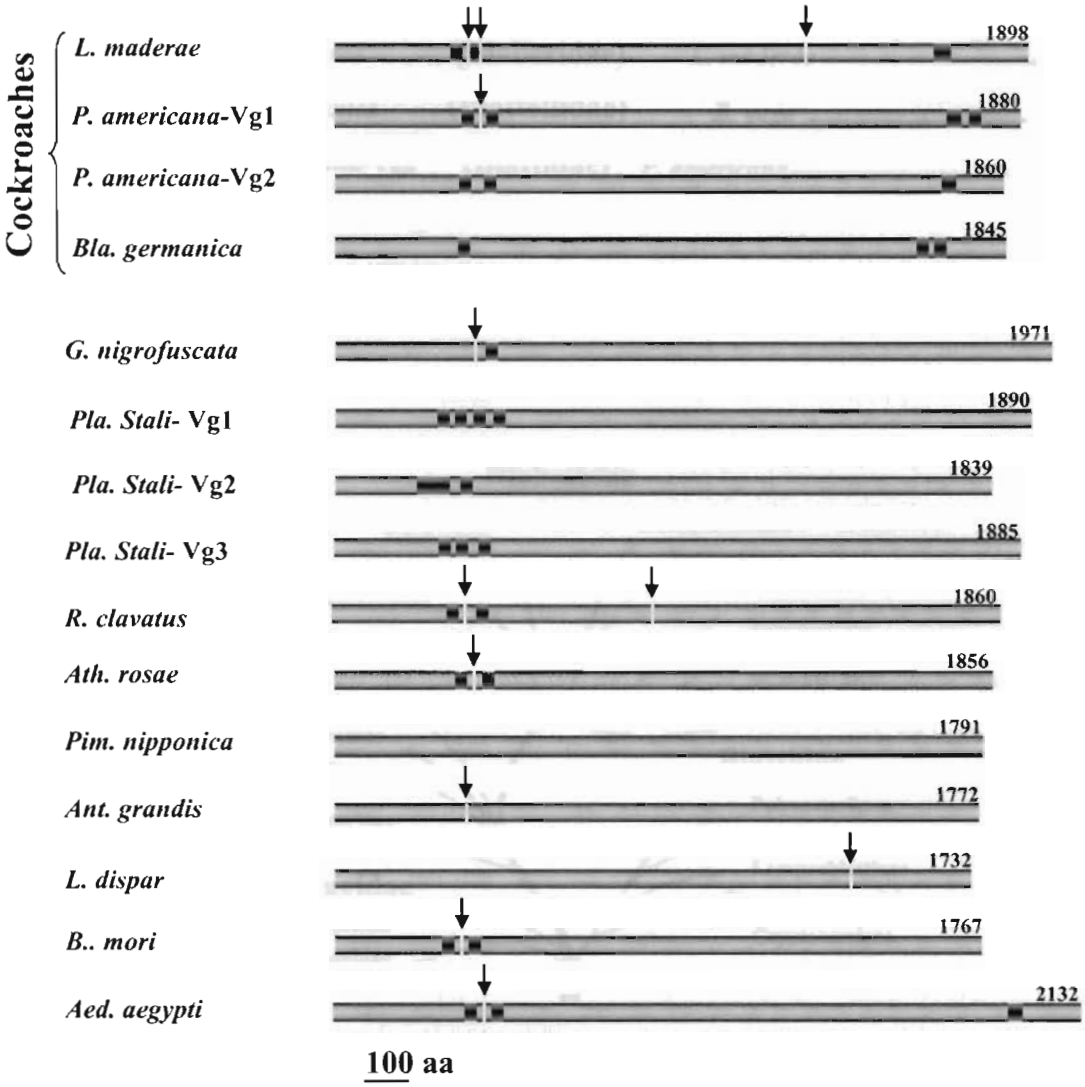
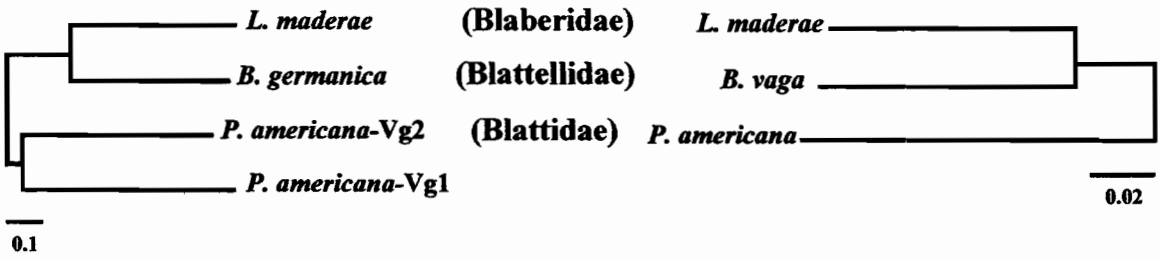
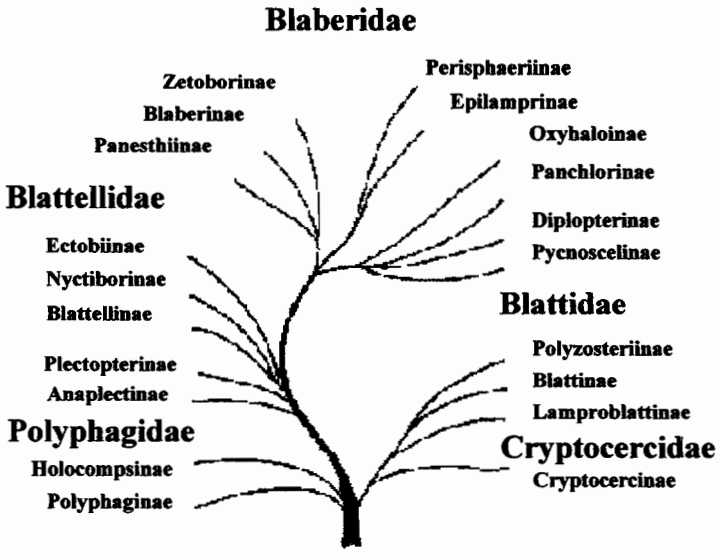


Fig. 8. The deduced amino acid sequences (excluding the siganl peptides) of Vgs from three cockroach species [*L. maderae,* the present study, *Periplaneta americana* Vg1 and Vg2 (Tufail et al., 2000, 2001), *Blattella germanica* (Comas et al., 2000)] and those of 9 other insect species [*Anthonomus grandis*, Trewitt et al., 1992; *Aedes aegypti*, Chen et al., 1994, Romans et al., 1995; *Bombyx mori*, Yano et al., 1994a, b; *Athalia rosae*, Kageyama et al., 1994, Nose et al., 1997; *Pimpla nipponica*, Nose et al., 1997; *Lymantria dispar*, Hiremath and Lehtoma, 1997a, b; *Riptortus clavatus*, Hirai et al., 1998; *Graptopsaltria nigrofuscata*, Lee et al., 2000b; *Plautia stali*, Lee et al., 2000a;] (most of the references are given in chapter 2) were multiple aligned to compare the polyserine stretches. The polyserine stretches are shown with black segments. The arrows and white lines are the determined cleavage sites following a consensus RXXR cleavage site sequence.



A



B

Fig. 9. A comparison of cockroach phylogenies based on the amino acid sequence of Vgs (**A-left**) [a phylogenetic (neighbour-joining) tree constructed based on the entire Vg amino acid sequences of three cockroach species and *L. dispar*, used as an outgroup.]; the sequence of mitochondrial 12S rRNA gene (**A-right**) [a phylogenetic tree constructed based on DNA sequence of mitochondrial 12S rRNA genes of three cockroach species and *M. religiosa*, as an outgroup. The sequence data were taken from the GeneBank data base (see text) and were reported by Kambhampati (1995).] and the morpho-ethological characteristics (**B**), a most widely suggested phylogeny of cockroaches by Mckittrick (1964). Bootstrap values in 1000 replicates are shown as percentage.

germanica are members of Blaberoidea and appear as the sister-group of *P. americana* (Blattoidea), as proposed in the phylogeny by McKittrick (1964), based on comparative morphology (Fig. 9, B). Moreover, the present phylogenetic results also agree with those based on the DNA sequences of mitochondrial 12S rRNA genes of *L. maderae, P. americana* and *B. vaga* (Fig. 9, A-right).

Both molecular trees were, thus, quite similar to each other and were basically in agreement with the most widely suggested phylogeny of cockroaches proposed by McKittrick (1964), based on morpho-ethological characteristics (Fig. 9). Our neighbour-joining tree based on Vg amino acid sequences (Fig. 9), admittedly incomplete at present, was limited to only a few species, *L. maderae* (Blaberidae), *B. germanica* (Blattellidae) and *P. americana* (Blattidae). To clarify the phylogenetic relationships of cockroaches at the molecular level, Vg cDNAs from other cockroach species of the same families, representing different subfamilies, and of other families, particularly those from polyphagidae and cryptocercidae needs to be cloned.

Moreover, we have also cloned and sequenced another Vg cDNA of *L. maderae* (see Fig. 10) which has stretches of amino acid sequences different from the one reported previously (Tufail and Takeda, 2002). The complete nucleotide sequence of this Vg (which we named as Vgb) consists of 5915 bp which encodes a deduced amino acid sequence of 1911 amino acid residues (including a putative signal peptide sequence) in a single open reading frame (Fig. 10). The difference of amino acid sequence stretches was confirmed by alignment of the deduced amino acid sequence of this Vg (Vgb) of *L. maderae* and that of reported previously from this cockroach species (Tufail and Takeda, 2002) and is shown in Fig. 11. The reason why this happened is not clear at the moment. One possibility that we happened to detect mutations which caused by not a single base substitution but more extensive one such as addition/deletion of a base or addition of a base at one position followed by deletion of a base at another position in the base sequence and vice versa. To confirm this situation the complete base sequences of the both *L. maderae* Vgs (Vga, the one

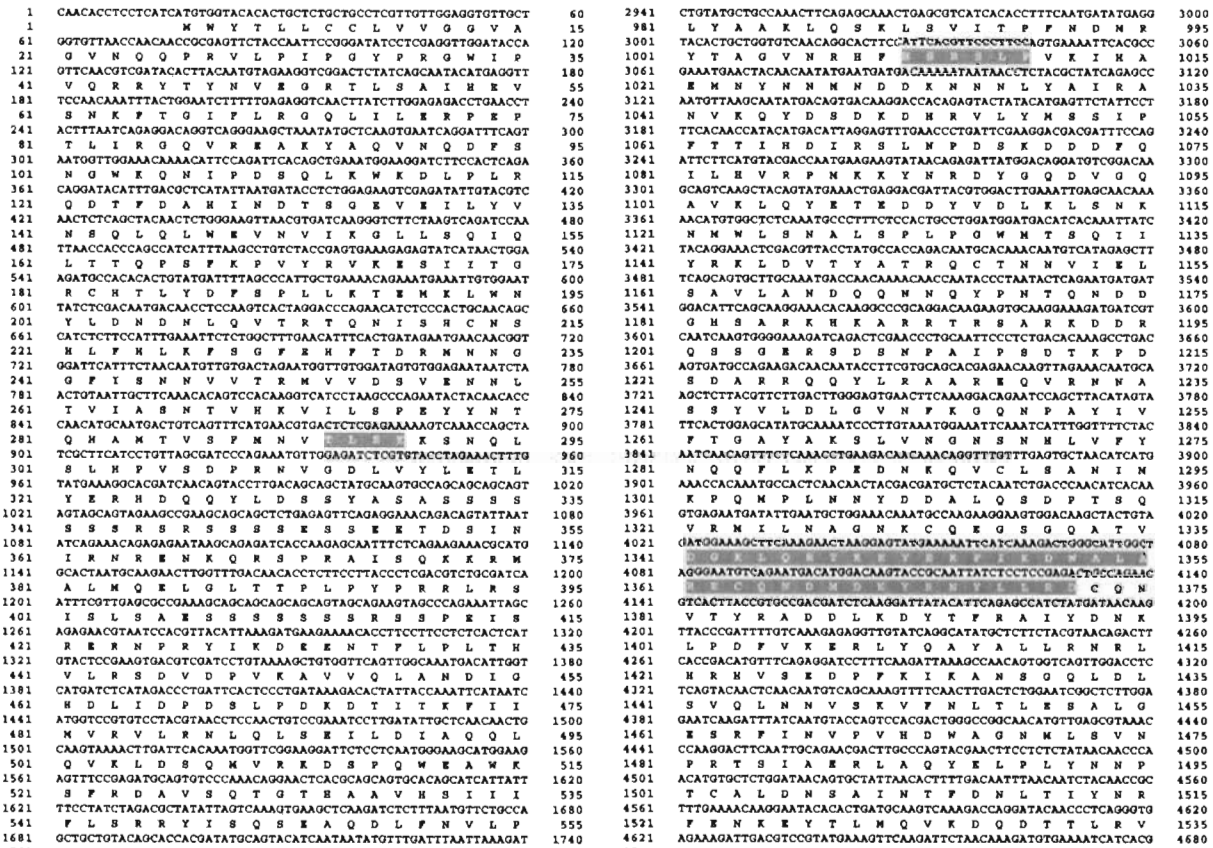


Fig. 10. Nucleotide sequence and deduced amino acid sequence of the other Vg cDNA of *L. maderae.* The amino acid sequence of this Vg has stetches of amino acid sequences which are different from the one reported earlier in this chapter (also see Tufail and Takeda, 2002) and are indicated with the shaded frames.

```
Lm-Vga          MWYTLLCCLVVGGVAGVNQQPRVLPIPGYPRGWIPVQRRYTYNVEGRTLSAIHEVSNKFT
Lm-Vgb          MWYTLLCCLVVGGVAGVNQQPRVLPIPGYPRGWIPVQRRYTYNVEGRTLSAIHEVSNKFT
                ************************************************************

Lm-Vga          GIFLRGQLILERPEPTLIRGQVREAKYAQVNQDFSNGWKQNIPDSQLKWKDLPLRQDTFD
Lm-Vgb          GIFLRGQLILERPEPTLIRGQVREAKYAQVNQDFSNGWKQNIPDSQLKWKDLPLRQDTFD
                ************************************************************

Lm-Vga          AHINDTSGEVEILYVNSQLQLWEVNVIKGLLSQIQLTTQPSFKPVYRVKESIITGRCHTL
Lm-Vgb          AHINDTSGEVEILYVNSQLQLWEVNVIKGLLSQIQLTTQPSFKPVYRVKESIITGRCHTL
                ************************************************************

Lm-Vga          YDFSPLLKTEMKLWNYLDNDNLQVTRTQNISHCNSHLFHLKFSGFEHFTDRMNNGGFISN
Lm-Vgb          YDFSPLLKTEMKLWNYLDNDNLQVTRTQNISHCNSHLFHLKFSGFEHFTDRMNNGGFISN
                ************************************************************

Lm-Vga          NVVTRMVVDSVENNLTVIASNTVHKVILSPEYYNTQHAMTVSFMNVLS-FKSNQLSLHPV
Lm-Vgb          NVVTRMVVDSVENNLTVIASNTVHKVILSPEYYNTQHAMTVSFMNVTLEKKSNQLSLHPV
                ********************************************   :**********

Lm-Vga          SDPRNVGDLVYLETLYERHDQQYLDSSYASASSSSSSSRSRSSSSESSEETDSINIRNRE
Lm-Vgb          SDPRNVGDLVYLETLYERHDQQYLDSSYASASSSSSSSRSRSSSSESSEETDSINIRNRE
                ************************************************************

Lm-Vga          NKQRSPRAISQKKRMALMQELGLTTPLPYPRRLRSISLSAESSSSSSSRSSPEISRERNP
Lm-Vgb          NKQRSPRAISQKKRMALMQELGLTTPLPYPRRLRSISLSAESSSSSSSRSSPEISRERNP
                ************************************************************

Lm-Vga          RYIKDEENTFLPLTHVLRSDVDPVKAVVQLANDIGHDLIDPDSLPDKDTITKFIIMVRVL
Lm-Vgb          RYIKDEENTFLPLTHVLRSDVDPVKAVVQLANDIGHDLIDPDSLPDKDTITKFIIMVRVL
                ************************************************************

Lm-Vga          RNLQLSEILDIAQQLQVKLDSQMVRKDSPQWEAWKSFRDAVSQTGTHAAVHSIIIFLSRR
Lm-Vgb          RNLQLSEILDIAQQLQVKLDSQMVRKDSPQWEAWKSFRDAVSQTGTHAAVHSIIIFLSRR
                ************************************************************

Lm-Vga          YISQSEAQDLFNVLPAAVQHHDMQYINNMFDLIKDPVVQQDRHVNETVVIAFSNAYRFIH
Lm-Vgb          YISQSEAQDLFNVLPAAVQHHDMQYINNMFDLIKDPVVQQDRHVNETVVIAFSNAYRFIH
                ************************************************************

Lm-Vga          ARLKRPYISPYFIKYLFQEFENAYRRQNTTQMQVYVH-LGNTGDVRIIPYLEPYLLRQIH
Lm-Vgb          ARLKRPYISPYFIKYLFQEFENAYRRQNTTQMQVYVH-LGNTGDVRIIPYLEPYLLRQIH
                ************************************:***********************

Lm-Vga          LSAFQRAHMFKALERVVDANPHLLTRFFLKFLLDQTDHPDVRVQAVFLLMRSDPSVAVLR
Lm-Vgb          LSAFQRAHMFKALERVVDANPHLLTRFFLKFLLDQTDHPDVRVQAVFLLMRSDPSVAVLR
                ************************************************************

Lm-Vga          TMAELTHSEPVNQVVSAIQAAIRTAARLRGTRFYNLAFKAQTVVNLLSDKNLDVSYSKNY
Lm-Vgb          TMAELTHSEPVNQVVSAIQAAIRTAARLRGTRFYNLAFKAQTVVNLLSDKNLDVSYSKNY
                ************************************************************

Lm-Vga          MLDQEAREYNLDFQLFYEQIGSQDNLLPKSALLDIFSYVGGAKSDHQTGYTVSSIDKVLN
Lm-Vgb          MLDQEAREYNLDFQLFYEQIGSQDNLLPKSALLDIFSYVGGAKSDHQTGYTVSSIDKVLN
                ************************************************************

Lm-Vga          DIQLQFKNFTQQGWTQQQMRDDLTKLVEGNIQYQVLGVQRFWPFDQDSIKSIPNVIQKFV
Lm-Vgb          DIQLQFKNFTQQGWTQQQMRDDLTKLVEGNIQYQVLGVQRFWPFDQDSIKSIPNVIQKFV
                ************************************************************

Lm-Vga          KDYREVKSFNLTKFFTTSTGIYGFPTVMGFPGVYTLHTPSLWKADGELKVTTVPDLEQNP
Lm-Vgb          KDYREVKSFNLTKFFTTSTGIYGFPTVMGFPGVYTLHTPSLWKADGELKVTTVPDLEQNP
                ************************************************************

Lm-Vga          HYLPGIVDVQLRVRPLYAAKLQSKLSVITPFNDMRYTAGVNRHFQLHVPFQYKIHAEMNY
Lm-Vgb          HYLPGIVDVQLRVRPLYAAKLQSKLSVITPFNDMRYTAGVNRHFHSRS-LEYKIHAEMNY
                ************************************************: :  : **********

Lm-Vga          NNMNDDKNNNLYAIRANVKQYDSDKDHRVLYMSSIPFTTIHDIRSLNPDSKDDDFQILHV
Lm-Vgb          NNMNDDKNNNLYAIRANVKQYDSDKDHRVLYMSSIPFTTIHDIRSLNPDSKDDDFQILHV
                ************************************************************
```

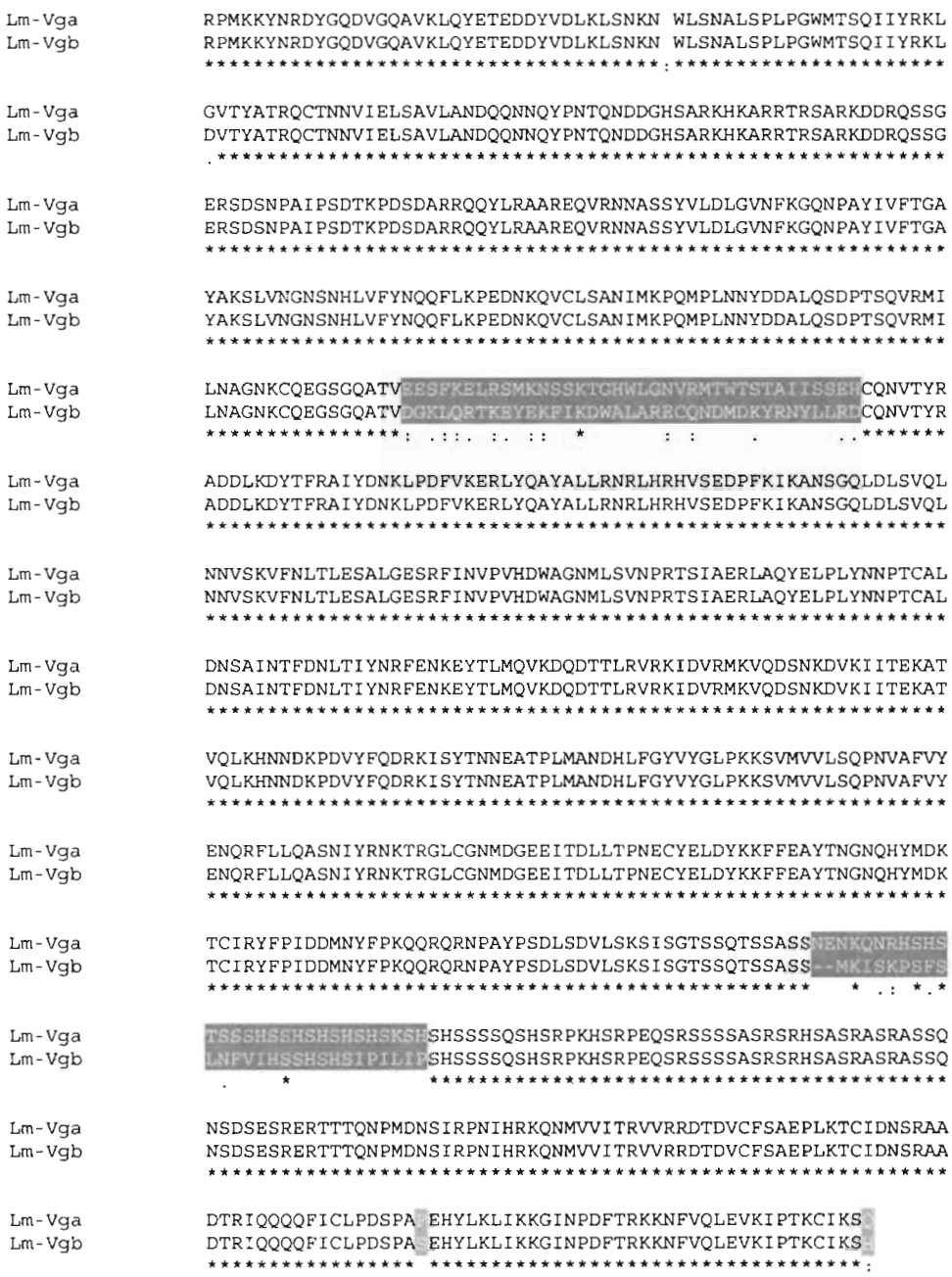


Fig. 11. The alignment of deduced amino acid sequences of two *L. maderae* Vgs [(Lm-Vga, previously reported one (see text); Lm-Vgb, another Vg)] to compare the protein primary structure of both *L. maderae* Vg clones using the Clustal W computer program. The comparison shows that Lm-Vgb has stretches (four in number) of amino acid sequences which are different from that of Lm-Vga and are shown with dark-shaded-frames. The difference of a single amino acid residue is shown with light-shaded-frames.

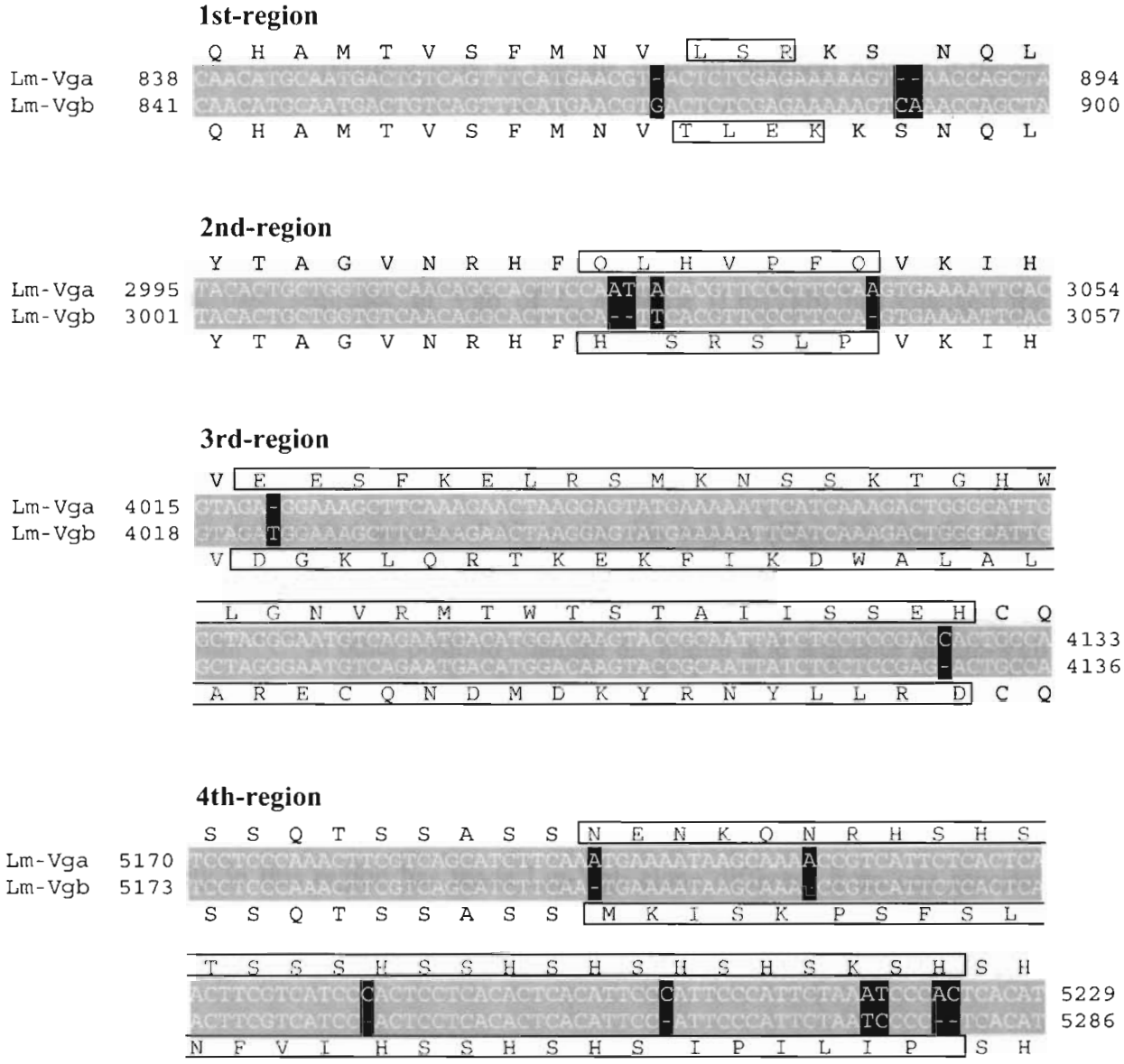


Fig. 12. The comparison of base sequences corresponding to 4 regions (having different amino acid sequence stretches) from two *L. maderae* Vgs [(Lm-Vga, previously reported one; Lm-Vgb, another Vg)] to detect mutations. The base sequences of Lm-Vga (upper) and Lm-Vgb (lower) are indicated with light-shaded frames, whereas, the amino acid sequences deduced are shown up and down of each sequence respectively. Numbers indicate the base sequence positions from the N-termini. The amino acid sequence stretches different in both *L. maderae* Vgs are shown with boxes. The mutations (any addition/deletion or substitution of a base) in the base sequence is show with a dark-shaded frame.

reported previously and Vgb, the present one) were multiple aligned. The comparison of base sequences corresponding only to regions (four in number) having different amino acid sequence stretches is shown in Fig. 12. The base sequence analysis in the first region of both *L. maderae* Vgs (see Fig. 12) shows that additin of three base residues (G "guanine" at one position and CA "cytosine and adenine" at other position) in the base sequence of Vgb of *L. maderae* (Lm-Vgb) has changed the deduced amino acid sequence in this region. In the second region, however the deletion of three bases (AT "adenine and thymine" at one position, and A "adenine" at other position) has been observed from the base sequence of Lm-Vgb. Also we note that thymine "T" has substituted adenine "A" in the this region (Fig. 12). In the third, relatively long region, the addition of thymine "T" at one position followed by the deletion of cytosine "C" at other position changed the amino acid sequence of Lm-Vgb in this region (Fig. 12). The comparison of base sequences in the fourth region of both *L. maderae* Vgs shows that a relatively more extensive deletion of base residues (six in total) and a substitution of adenine "A" and thymine "T" with thymine "T" and cytosine "C" respectively in this region has made Lm-Vgb clone different from one reported previously from the this cockroach species (see Fig. 12).

4.5 LITERATURE CITED

Bownes, M., 1986. Expression of genes coding for vitellogenin (yolk protein). Annu. Rev. Entomol. 31, 507-531.

Byrne, B. M., Gruber, M., Ab, G., 1989. The evolution of egg yolk proteins. Progr. Biophys. Molec. Biol. 53, 33-69.

Chen, J. S., Sapington, T. W., Raikhel, A. S., 1997. Extensive sequence conservation among insect, nematode and vertebrate vitellogenins reveal s ancient common ancestry. J. Mol. Evol. 44, 440-451.

Comas, D., Piulachs, M. D., Belles, X., 2000. The vitellogenin of *Blattella germanica* (L.) (Dictyoptera, Blattellidae): nucleotide sequence of the cDNA and analysis of the protein primary structure. Arch. Insect Biochem. Physiol. 36, 154-164.

Della-Cioppa, G., Engelmann, F., 1987. The vitellogenin of *Leucophaea maderae*: synthesis as a large phosphorylated precursor. Insect Biochem. 17, 401-405.

Don-Wheeler, G., Engelmann, F., 1991. The female- and male-produced vitellogenins of *Leucophae maderae*. J. Insect Physiol. 12, 869-882.

Don-Wheeler, G., Engelmann, F., 1997. The biosynthesis and processing of vitellogenin in the fat bodies of female and male of the cockroach, *Leucophaea maderae*. Insect Biochem. Mol. Biol. 27, 901-918.

Engelmann, F., 1979. Insect vitellogenin: identification, biosynthesis, and the role in vitellogenesis. Adv. Insect Physiol. 14, 49-108.

Giorgi, F., Bradley, J. T., Nordin, J. H., 1999. Differential vitellin polypeptide processing in insect embryos. Micron. 30, 579-596.

Hagedorn, H. H., Maddison, D. R., Tu, Z., 1998. The evolution of vitellogenins, cyclorrhaphan yolk proteins and related molecules. Adv. Insect Physiol. 27, 335-384.

Harmon, M. A., 1994. Cloning and characterization of a cDNA encoding the C-terminal of *Leucophaea maderae* vitellogenin. Dissertation Abstracts International. 55 (2), 257-258 B.

Izumi, S., Yano, K., Yamamoto, Y., Takahashi, S.Y., 1994. Yolk proteins from insect eggs: structure, biosynthesis and programmed degradation during embryogenesis. J. Insect Physiol. 40, 735-746.

Kambhampati, S., 1995. A phylogeny of cockroaches and related insects based on DNA sequences of mitochondrial ribosomal RNA genes. Proc. Natl. Acad. Sci. USA. 92, 2017-2020.

Kumar, S., Koichiro, T., Ingrid, B. J., Masatoshi, N., 2001. *MEGA2*: Molecular Evolutionary Genetics Analysis software, Arizona State University, USA.

Lee, J. M., Hatakeyama, M., Oishi, K., 2000. A simple and rapid method for cloning insect vitellogenin cDNAs. Insect Biochem. Mol. Biol. 30, 189-194.

Mckittrick, F. A., 1964. Evolutionary studies of cockroaches. Cornell University Agricultural experiment Station, Memoir. 389, 1-197.

Nielsen, H., Engelbrecht, J., Brunak, S., von Heijne G., 1997. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Prot. Eng. 10, 1-6.

Raikhel, A. S., Dhadialla, T. S., 1992. Accumulation of yolk proteins in insect oocytes. Annu. Rev. Entomol. 37, 217-251.

Sambrook, J., Fritsch, E. F., Maniatis, T., 1989. Molecular Cloning: A Laboratory Manual. 2nd ed. Cold Spring Harbor Laboratory Press, New York.

Sappington, T. W., Raikhel, A. S., 1998. Molecular characteristics of insect vitellogenins and vitellogenin receptors. Insect Biochem. Molec. Biol. 28, 277-300.

Thompson, J. D., Higgins, D. G., Gibson, T. J., 1994. Clustal W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22, 4673-4680.

Tufail, M., Hatakeyama, M., Takeda, M., 2001. Molecular evidence for two vitellogenin genes and processing of vitellogenins in the American cockroach, *Periplaneta americana*. Arch. Insect Biochem. Physiol. 48, 72-80.

Tufail, M., Lee, J. M., Hatakeyama, M., Oishi, K., Takeda, M., 2000. Cloning of vitellogenin cDNA of the American cockroach, *Periplaneta americana* (Dictyoptera), and its structural and expression analyses. Arch. Insect Biochem. Physiol. 45, 37-46.

Wyatt, G. R., 1991. Gene regulation in insect reproduction. Invert Reprod Dev. 20: 1-35.

CHAPTER 5

Molecular cloning and characterization of the cockroach vitellogenin receptor: a new LDLR gene family member with only 6 repeats in the second ligand binding domain

5.1 ABSTRACT

Insect vitellogenin and yolk protein receptors (VgR/YPR) are recently discovered members of the low-density lipoprotein receptor (LDLR) superfamily, and are unique in having two domains of ligand binding repeats (class A repeats) with 5 repeats in the first domain and 8 in the second domain. We now report the discovery of a new insect VgR having only 6 repeats in the second ligand binding domain from previtellogenic ovaries of the cockroach, *Periplaneta americana*. This novel insect VgR consists of 1709 amino acid residues and shares a significant homology with other LDLR family members, and particularly with Vg/YP receptors reported from the mosquito and *Drosophila*. The cytoplasmic tail of this five-domain receptor contains a leucine-isoleucine internalization signal similar to that of the mosquito VgR, a motif that seems to be common in insect Vg/YP receptors (di-leucine in *Drosophila* YPR), unlike the tight-turn-tyrisine motif (NPXY) of other members of LDLR superfamily. Phylogenetic analysis (neighbour-joining) shows that the mosquito VgR and YPR of *Drosophila* have closer ancestory than from cockroach VgR.

5.2 INTRODUCTION

Receptors that transport vitellogenin (Vg), the major yolk protein precursor, into oocytes are of vital importance to all oviparous animals because they mediate a key step of oocyte development, a prerequisite to reproduction. The Vg is synthesized, in most insects, extraovarially in the fat body under hormonal control and transported via hemolymph to the ovary, where it is then taken up selectively by the oocytes. Vg receptors (VgRs) are localized in coated pits on the surface of growth competent oocytes (Roehrkasten and Ferenz, 1986; Roehrkasten et al., 1989; Hafer et al., 1992; Sappington et al., 1995; Sappington and Raikhel, 1998) which bind the yolk protein precursor and carry it into cells by receptor mediated endocytosis (reviewed by Byrne et al., 1989; Raikhel and Dhadhialla, 1992; Sappington and Raikhel, 1998). Once sequestered by growing oocytes, the Vg is sent to yolk bodies, where it is processed for storage and thus providing the main nutritional reserves necessary for the embryo development.

Molecular characterization of VgRs from different species, of both vertebrates, such as those of chickens, frog, rainbow trout (Bujo et al., 1994; Okabayashi et al., 1996; Part et al., 1998 and Davail et al., 1988) and invertebrates, such as that of the nematode (Grant and Hirsh, 1999), the mosquito (Sappington et al., 1996) and a yolk protein receptor (YPR) of the *Drosophila* (Schonbaum et al, 1995) demonstrated that they all are the members of the low density lipoprotein receptor (LDLR) gene superfamily. The deduced amino acid sequences of the known VgRs revealed that these receptors, like other LDLR family members, contain five distict domains: clusters of cysteine-rich repeats constituting the ligand-binding domain, epidermal growth factor (EGF)-like repeats, repeats containing a YWTD motif that are proposed to form a B-propeller domain (Springer, 1998), a tansmembrane domain anchoring the receptor in the plasma membrane, and a cytoplasmic domain containing at least one copy of the NPXY sequence (di-leucine or leucine-isoleucine motif in insect VgR/YPRs)which is involved in receptor internalization via coated pits (Goldstein et al., 1985; Chen et al.,

1990). The insect VgR/YPRs are, however, double the size, containing two clusters of ligand-binding domain, instead of one like in the vertebrate VgRs and the classical LDLRs (see review: Sappington and Raikhel, 1998).

The LDLR-family of receptors mediate endocytosis and lysosomal targeting of a diverse array of macromolecules and influence directly or indirectly most, if not all, physiological processes including reproduction, development and nutrition and many pathophysiological processes. The VgRs are particularly involved in reproduction. The chicken VgR, for example, has been shown to import very low density lipoprotein (VLDL), riboflavin-binding protein, and alpha$_2$-macdoglobulin into growing oocytes (Stifani et al., 1990; Mac Lachlan et al., 1994). Indeed, mutations that abrogate expression of the VgR result in non-egg laying hens. In *Drosophila melanogaster*, the YPR is encoded by the gene *yolkless* (*yl*) (Schonbaum et al., 1995). Female sterility occurs in insects having genetic deficiency of *yl* (Schonbaum et al., 1995). Analysis of oocytes produced by yl^{-1} females shows drastic reduction in numbers of coated pits and coated vesicles and very little protenacious yolk (DiMario and Mahowald, 1987). Although, VgRs play a critical role in the mediation of oocyte growth, and are the promisng target for future novel pest control strategies, much less is known about these rceptors, paticularly in insects, as compared to their ligands, the Vgs (see Tufail et a., 2001 for references).

We have recently characterized the molecular structures of two Vg molecules from the American cockroach, *Perilaneta americana* (Tufail et al., 2000 and 2001). We report now the cDNA cloning and structural analysis of their counterpart, the receptor, from this cockroach species, the first insect VgR to be reported with a second ligand binding domain of only 6 repeats. This new VgR is clearly a homolog of the other insect Vg/YP receptors, vertebrate VgRs, and LDL receptors. The deduced amino acid sequence of *P. americana* VgR was multiple aligned with 7 other members of the family and a phylogenetic tree (neighbour-joining) was constructed.

5.3 MATERIALS AND METHODS

5.3.1 Animals

Cultures of *P. americana* were maintained in the laboratory as described previously (Tufail et al., 2000). Newly emerged females were collected from stock colonies and kept separately under constant light conditions at 26°C. The ovaries were collected from 3-10 days old adult females dissected in phosphate-buffered saline (PBS). The organs were frozen immediately in liquid nitrogen and stored at –80°C.

5.3.2 Preparation of mRNA and construction of an adaptor-ligated double-stranded cDNA library

Total RNA was extracted from the vitellogenic ovaries as reported previously (Tufail et al., 2000). Poly $(A)^+$ RNA was purified from total RNA by using an mRNA purification kit (Amersham-Pharmacia). A total of 1 μg Poly $(A)^+$ RNA was used to construct an adaptor-ligated double stranded (ds) cDNA library using a marathon cDNA amplification kit (Clontec). The adaptor-ligated ds cDNA library was then used for amplification of both the 5'- and 3'- ends through RACE-PCR as previously described (Tufail and Takeda, 2002).

5.3.3 Amplification of cDNA ends and Sequencing

A cDNA fragment encoding the N-terminus of *P. ameicana* VgR was first amplified from the ds cDNA library prepared from previtellogenic ovaries using the degenerate primers and the adaptor primer (marathon cDNA adaptor, Clontech). The degenerate primers used were: 5`-CCAATCTGACCA-3`, 5`-CCAGTCGGACCA-3`, 5`-CCAGTCAGACCA 3` and 5`-CCA(AG)TC(ACGT)GACCA-3`, and were designed from a conserved region (WSDW) of an EGF precursor homology domain of *Aedes aegypti* VgR (Sappington et al., 1996) and the YPR of *D. melanogaster* (Schonbaum et al., 1995). The amplified ~1.4 kbp PCR products (Shown in Fig. 1) were cloned into

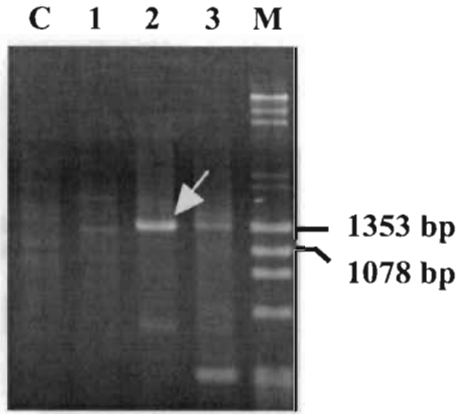


Fig. 1. Amplification of the N-terminal portion of *P. americana* VgR cDNA. A fragment of cDNA encoding N-terminus of *P. americana* VgR was amplified through RACE-PCR by using the three degenerate primers (antisense) (prepared from a conserved region of an EGF precursor homology domain of known insect VgR/YPRs) each with an adaptor primer (a sense primer) and an adaptor–ligated double stranded cDNA constructed from mRNA prepared from vitellogenic ovaries of 3-10 old adult female. The three amplified PCR products were applied one on each lane (1-3) of the agarose gel respectively. Lane M: molecular size marker (bP); Lane C: negative control (reaction mixture only). The product cloned and sequenced is indicated by an arrowhead.

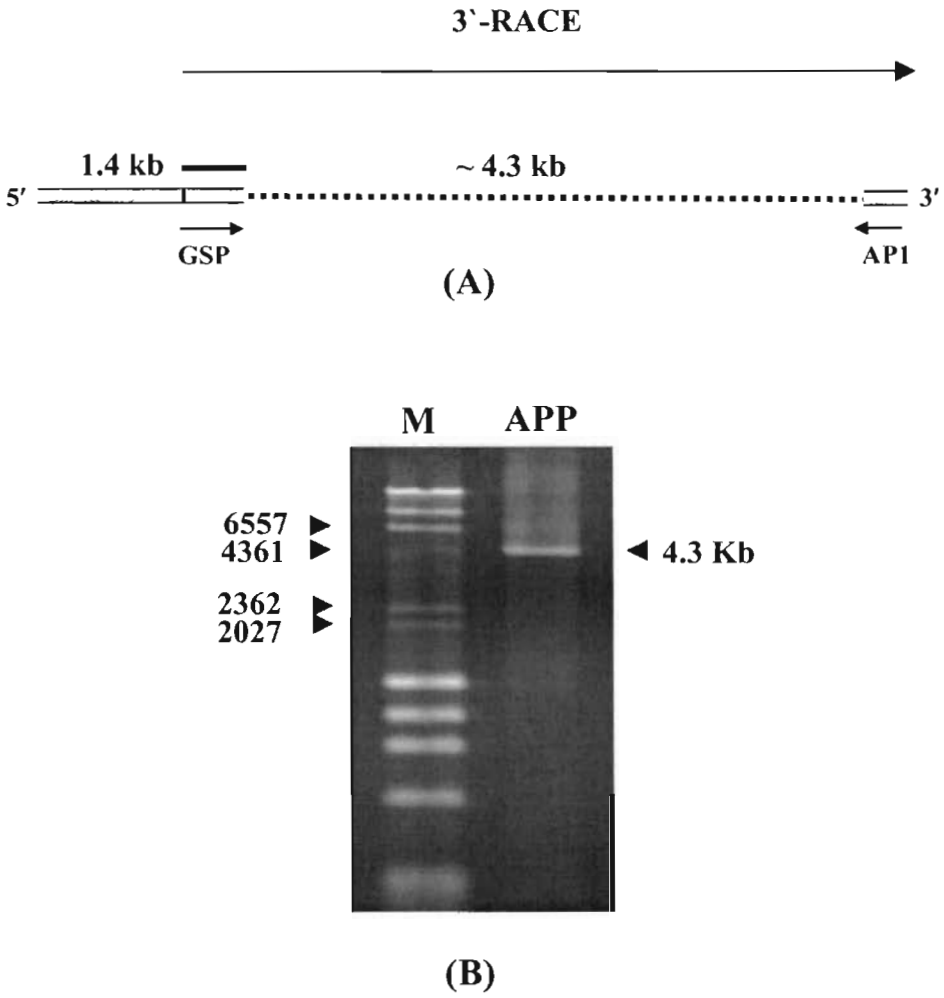


Fig. 2. 3′-end amplification of *P. americana* VgR cDNA. The 3`-end portion of *P. americana* VgR cDNA was obtained through RACE-PCR by using a gene specific primer (GSP) designed from the 3`-terminal portion of the initial cDNA clone (1.4 kb) shown on the 5` region (A), the adaptor primer (AP1), and the adaptor-ligated double stranded cDNA library contructed from the mRNA prepared from the vitellogenic ovaries of 3-10 days old adult females. The amplified PCR product (APP) (4.3 kb) (shown in B) was purified, cloned and sequenced. The amplified fragment was identified by an overlapping region and has been shown with a bold line. M is the molecular weight marker.

the TOPO TA cloning kit (Invitrogen) and sequenced from both sides as previously described (Tufail et al., 2001). The deduced amino acid sequences were analyzed using DNAS-Mac v3.6 software and were checked for their similarity with LDL receptors using FASTA homology search on DDBJ database.

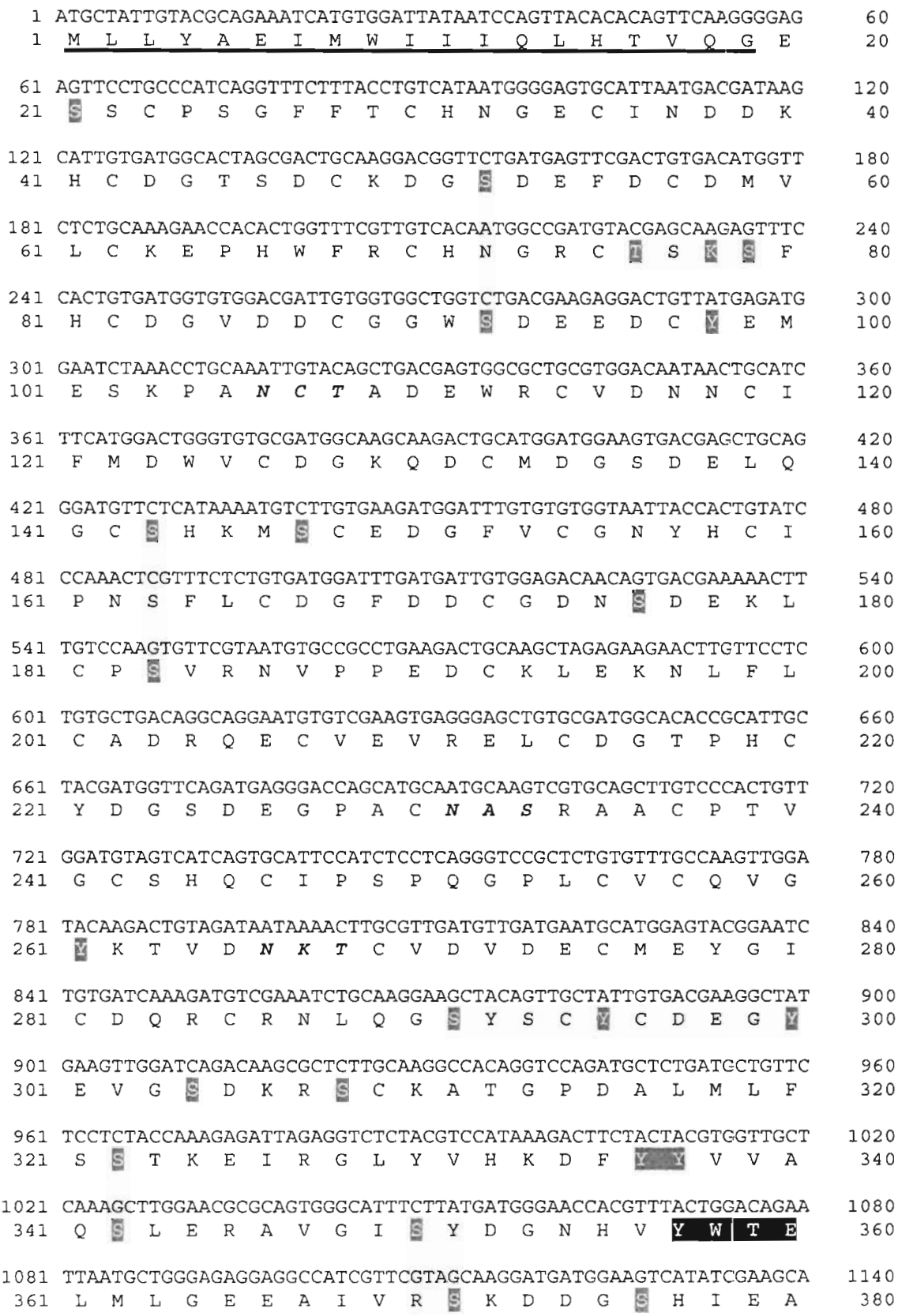
The remaining portion of VgR cDNA was obtained by 3 RACE, using a gene-specific primer (as a sense primer) corresponding to 3' end of the initial 5 end cDNA fragment coding *P. americana* VgR. An adaptor-ligated ds cDNA library was subjected to PCR with a gene-specific primer and the adaptor primer (Clontech) following the instructions of the manufacturer. The PCR conditions employed were: 94 °C for 1 min for denaturing, followed by 36 cycles of 94 °C for 30 s and 68 °C for 8 min. The amplified cDNAs (shown in Fig. 2) were cloned in TOPO XL cloning kit (Invitrogen) and sequenced in both directions.

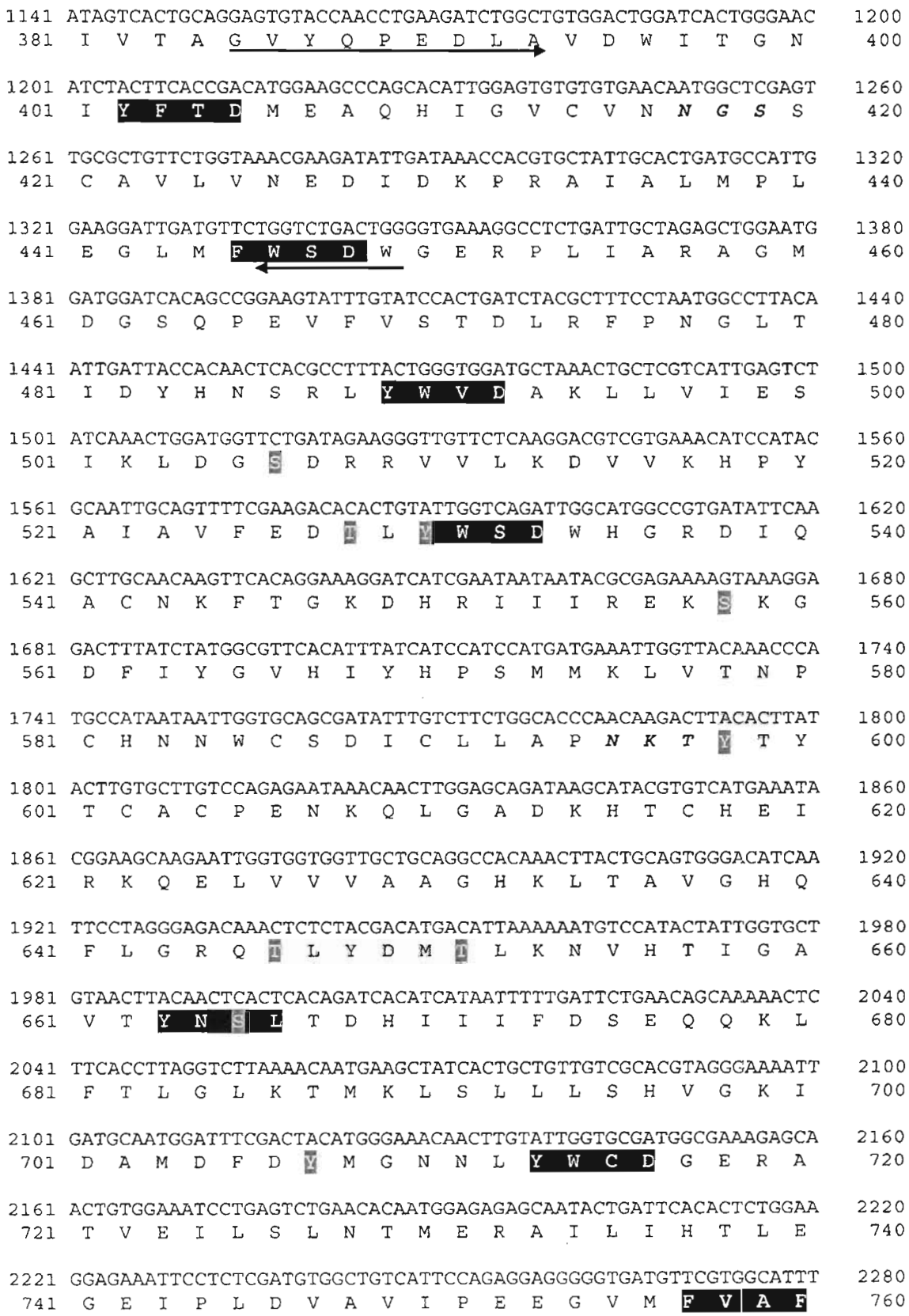
5.4 RESULTS AND DISCUSSION

In the present study, we have cloned the complete cDNA encoding VgR of *P. americana* and studied its primary structure. In order to know the complete structure, we first attempted to clone a partial trnscript encoding the N-terminus of this rceptor. For this purpose, four degenerate primers were designed and synthesized as described in Materials and Methods. Using these primers and an adaptor primer (clontec), a ~1.4 kb PCR fragment was obtained from an adaptor ligated ds cDNA library constructed from the mRNA of vitellogenic ovaries, subcloned into a pCR2.1-TOPO vector (Invetrogen), and analyzed. Five cDNA inserts of amplified PCR product were sequenced and analyzed. The cDNA inserts sequenced were all of 1347 nucleotides long and each encoded 449 amino acid residues in a single open reading frame following the initiation methionine (ATG) (Fig.3). Identities of the fragments were confirmed by sequencing in both directions. The sequence analysis revealed that the amplified clones were all derived from a single transcript that shared the sequence

similarities with LDL receptor family members, of both vertebrates and invertebrates. The remaining 3'-end part of *P. americana* VgR was then cloned with 3'-RACE-PCR method. To obtain the 3'-end part, a gene-specific primer was prepared and used with the adaptor ligated ds cDNA library, in addition to the adaptor primer. The 3'-end amplified PCR product yielded a sequence that was 4258 bp long which encoded 1325 amino acid residues in a single open reading frame following by a termination codon (TGA) and a consensus polyadenylation signal (AATAAA) (Fig. 3). The complete sequencing of two overlapping RACE-PCR fragments of 5'- and 3'-end of *P. americana* VgR thus defined an open reading frame of 5411 bp coding for a protein of 1709 residues, including a putative signal peptide (Fig. 3). The sequencing strategy (overall one) of 5`- and 3`-end RACE-PCR fragments alongwith the restriction map is provided in Fig. 4.

The analysis of the deduced amino acid sequence revealed that the putative signal peptide is located at the N-terminus of the pre-VgR (predicted by using the SignalP VI.I computer program, Neilsen et al., 1997), and a probable cleavage site is located between residues 19 and 20 (Fig. 3). The deduced amino acid sequence, after removing the signal peptide, predicted a protein of 196.65 kD. The analysis of the mature protein revealed that it is, like that of the mosquito VgR (Sappington et al., 1996) and a YPR of *Drosophila* (Schonbaum et al, 1995), also a member of the LDLR gene superfamily (Fig. 5). There were five modular elements characterized in this protein: I) ligand binding repeats (class A repeats), II) EGF-precursor like repeats (class B), III) repeats containing a YWTD motif IV) a hydrophobic tansmembrane domain, and V) a cytoplasmic domain which contains a leucine-isoleucine internalization signal (amino acid position: 1701-1708), a putative alternative signal required for internalization in clathrin-coated pits which was also found in VgR of mosquito. Interestingly, this receptor does not contain the crucial tyrosine residue (Fig. 6A) in the highly conserved internalization signal (FDNPVY), which was also absent in the recently identified VgR of mosquito and the YPR of *Drosophila*. There is a stretch of 31 amino acid residues


1 ATGCTATTGTACGCAGAAATCATGTGGATTATAATCCAGTTACACACAGTTCAAGGGGAG 60
1 M L L Y A E I M W I I I Q L H T V Q G E 20
61 AGTTCCTGCCCATCAGGTTTCTTTACCTGTCATAATGGGGAGTGCATTAATGACGATAAG 120
21 S S C P S G F F T C H N G E C I N D D K 40
121 CATTGTGATGGCACTAGCGACTGCAAGGACGGTTCTGATGAGTTCGACTGTGACATGGTT 180
41 H C D G T S D C K D G S D E F D C D M V 60
181 CTCTGCAAAGAACCACACTGGTTTCGTTGTCACAATGGCCGATGTACGAGCAAGAGTTTC 240
61 L C K E P H W F R C H N G R C T S K S F 80
241 CACTGTGATGGTGTGGACGATTGTGGTGGCTGGTCTGACGAAGAGGACTGTTATGAGATG 300
81 H C D G V D D C G G W S D E E D C Y E M 100
301 GAATCTAAACCTGCAAATTGTACAGCTGACGAGTGGCGCTGCGTGGACAATAACTGCATC 360
101 E S K P A N C T A D E W R C V D N N C I 120
361 TTCATGGACTGGGTGTGCGATGGCAAGCAAGACTGCATGGATGGAAGTGACGAGCTGCAG 420
121 F M D W V C D G K Q D C M D G S D E L Q 140
421 GGATGTTCTCATAAAATGTCTTGTGAAGATGGATTTGTGTGTGGTAATTACCACTGTATC 480
141 G C S H K M S C E D G F V C G N Y H C I 160
481 CCAAACTCGTTTCTCTGTGATGGATTTGATGATTGTGGAGACAACAGTGACGAAAAACTT 540
161 P N S F L C D G F D D C G D N S D E K L 180
541 TGTCCAAGTGTTCGTAATGTGCCGCCTGAAGACTGCAAGCTAGAGAAGAACTTGTTCCTC 600
181 C P S V R N V P P E D C K L E K N L F L 200
601 TGTGCTGACAGGCAGGAATGTGTCGAAGTGAGGGAGCTGTGCGATGGCACACCGCATTGC 660
201 C A D R Q E C V E V R E L C D G T P H C 220
661 TACGATGGTTCAGATGAGGGACCAGCATGCAATGCAAGTCGTGCAGCTTGTCCCACTGTT 720
221 Y D G S D E G P A C N A S R A A C P T V 240
721 GGATGTAGTCATCAGTGCATTCCATCTCCTCAGGGTCCGCTCTGTGTTTGCCAAGTTGGA 780
241 G C S H Q C I P S P Q G P L C V C Q V G 260
781 TACAAGACTGTAGATAATAAAACTTGCGTTGATGTTGATGAATGCATGGAGTACGGAATC 840
261 Y K T V D N K T C V D V D E C M E Y G I 280
841 TGTGATCAAAGATGTCGAAATCTGCAAGGAAGCTACAGTTGCTATTGTGACGAAGGCTAT 900
281 C D Q R C R N L Q G S Y S C Y C D E G Y 300
901 GAAGTTGGATCAGACAAGCGCTCTTGCAAGGCCACAGGTCCAGATGCTCTGATGCTGTTC 960
301 E V G S D K R S C K A T G P D A L M L F 320
961 TCCTCTACCAAAGAGATTAGAGGTCTCTACGTCCATAAAGACTTCTACTACGTGGTTGCT 1020
321 S S T K E I R G L Y V H K D F Y Y V V A 340
1021 CAAAGCTTGGAACGCGCAGTGGGCATTTCTTATGATGGGAACCACGTTTACTGGACAGAA 1080
341 Q S L E R A V G I S Y D G N H V Y W T E 360
1081 TTAATGCTGGGAGAGGAGGCCATCGTTCGTAGCAAGGATGATGGAAGTCATATCGAAGCA 1140
361 L M L G E E A I V R S K D D G S H I E A 380


1141 ATAGTCACTGCAGGAGTGTACCAACCTGAAGATCTGGCTGTGGACTGGATCACTGGGAAC 1200
381 I V T A G V Y Q P E D L A V D W I T G N 400
1201 ATCTACTTCACCGACATGGAAGCCCAGCACATTGGAGTGTGTGTGAACAATGGCTCGAGT 1260
401 I Y F T D M E A Q H I G V C V N N G S S 420
1261 TGCGCTGTTCTGGTAAACGAAGATATTGATAAACCACGTGCTATTGCACTGATGCCATTG 1320
421 C A V L V N E D I D K P R A I A L M P L 440
1321 GAAGGATTGATGTTCTGGTCTGACTGGGGTGAAAGGCCTCTGATTGCTAGAGCTGGAATG 1380
441 E G L M F W S D W G E R P L I A R A G M 460
1381 GATGGATCACAGCCGGAAGTATTTGTATCCACTGATCTACGCTTTCCTAATGGCCTTACA 1440
461 D G S Q P E V F V S T D L R F P N G L T 480
1441 ATTGATTACCACAACTCACGCCTTTACTGGGTGGATGCTAAACTGCTCGTCATTGAGTCT 1500
481 I D Y H N S R L Y W V D A K L L V I E S 500
1501 ATCAAACTGGATGGTTCTGATAGAAGGGTTGTTCTCAAGGACGTCGTGAAACATCCATAC 1560
501 I K L D G S D R R V V L K D V V K H P Y 520
1561 GCAATTGCAGTTTTCGAAGACACACTGTATTGGTCAGATTGGCATGGCCGTGATATTCAA 1620
521 A I A V F E D T L Y W S D W H G R D I Q 540
1621 GCTTGCAACAAGTTCACAGGAAAGGATCATCGAATAATAATACGCGAGAAAAGTAAAGGA 1680
541 A C N K F T G K D H R I I I R E K S K G 560
1681 GACTTTATCTATGGCGTTCACATTTATCATCCATCCATGATGAAATTGGTTACAAACCCA 1740
561 D F I Y G V H I Y H P S M M K L V T N P 580
1741 TGCCATAATAATTGGTGCAGCGATATTTGTCTTCTGGCACCCAACAAGACTTACACTTAT 1800
581 C H N N W C S D I C L L A P N K T Y T Y 600
1801 ACTTGTGCTTGTCCAGAGAATAAACAACTTGGAGCAGATAAGCATACGTGTCATGAAATA 1860
601 T C A C P E N K Q L G A D K H T C H E I 620
1861 CGGAAGCAAGAATTGGTGGTGGTTGCTGCAGGCCACAAACTTACTGCAGTGGGACATCAA 1920
621 R K Q E L V V V A A G H K L T A V G H Q 640
1921 TTCCTAGGGAGACAAACTCTCTACGACATGACATTAAAAAATGTCCATACTATTGGTGCT 1980
641 F L G R Q T L Y D M T L K N V H T I G A 660
1981 GTAACTTACAACTCACTCACAGATCACATCATAATTTTTGATTCTGAACAGCAAAAACTC 2040
661 V T Y N S L T D H I I I F D S E Q Q K L 680
2041 TTCACCTTAGGTCTTAAAACAATGAAGCTATCACTGCTGTTGTCGCACGTAGGGAAAATT 2100
681 F T L G L K T M K L S L L L S H V G K I 700
2101 GATGCAATGGATTTCGACTACATGGGAAACAACTTGTATTGGTGCGATGGCGAAAGAGCA 2160
701 D A M D F D Y M G N N L Y W C D G E R A 720
2161 ACTGTGGAAATCCTGAGTCTGAACACAATGGAGAGAGCAATACTGATTCACACTCTGGAA 2220
721 T V E I L S L N T M E R A I L I H T L E 740
2221 GGAGAAATTCCTCTCGATGTGGCTGTCATTCCAGAGGAGGGGGTGATGTTCGTGGCATTT 2280
741 G E I P L D V A V I P E E G V M F V A F 760

```
2281 TCCCGGCACGTTATTGGTGATGGACCCCACATTGATCGCATGAATATGGATGGCAGAGGA 2340
 761  S  R  H  V  I  G  D  G  P  H  I  D  R  M  N  M  D  G  R  G   780

2341 GCGCACACACATGTCATTGAAACCAGTCTGGACGGCCCAATTATTTCACTGTTTTATGAC 2400
 781  A  H  T  H  V  I  E  T  S  L  D  G  P  I  I  S  L  F  Y  D   800

2401 AGTGATTTGCACAGAGTGTTCTGGACTGACCCCAACAACGAGGAAATAGGCAGTGCAGCA 2460
 801  S  D  L  H  R  V  F  W  T  D  P  N  N  E  E  I  G  S  A  A   820

2461 GCTGATGGTATGGATCAACATGTATTCAGAAGTGATGTAGAAGGATCTCCAATTGATATT 2520
 821  A  D  G  M  D  Q  H  V  F  R  S  D  V  E  G  S  P  I  D  I   840

2521 GCATCGGTGGGCCGAGATATGTTCTGGACTATGTGGGCCCACCCTTACCTGTATTGGGCC 2580
 841  A  S  V  G  R  D  M  F  W  T  M  W  A  H  P  Y  L  Y  W  A   860

2581 AGTAAATTCAACAGTCAAAGTAGAATGAAGAGGTTGTTGCTTGATGTGGAAGACTCAGAT 2640
 861  S  K  F  N  S  Q  S  R  M  K  R  L  L  L  D  V  E  D  S  D   880

2641 AAACTGCCACTAGTGGCTGTACGAGGTGTTAGAGCACAACCTGATCACCCGTGTCATAAA 2700
 881  K  L  P  L  V  A  V  R  G  V  R  A  Q  P  D  H  P  C  H  K   900

2701 AATAATGGAGGCTGCAGGGATGGTAGTGATGAGTATTACTGCTTTGAAGAAGAATGCAAC 2760
 901  N  N  G  G  C  R  D  G  S  D  E  Y  Y  C  F  E  E  E  C  N   920

2761 GAAGATTTGCAGTTCAAGTGTAGAACTGGGGACTGCATAGTCAAGAGCTGGTATTGTGAT 2820
 921  E  D  L  Q  F  K  C  R  T  G  D  C  I  V  K  S  W  Y  C  D   940

2821 GGCTCAAAGGATTGTGAAGATGGTTCAGACGAAGAGAATTGTGAAGAAGTTACATGCGAA 2880
 941  G  S  K  D  C  E  D  G  S  D  E  E  N  C  E  E  V  T  C  E   960

2881 CCTTCAGCATTTAAATGTGCCCTGGGACAGTGCATTCCTGAAGAGTGGGTGTGTGATGGT 2940
 961  P  S  A  F  K  C  A  L  G  Q  C  I  P  E  E  W  V  C  D  G   980

2941 CAGTCGGACTGTGTTGATGATACGGACGAGCAAAACTGTGCTCCTCCGACTTGTGGTCCT 3000
 981  Q  S  D  C  V  D  D  T  D  E  Q  N  C  A  P  P  T  C  G  P  1000

3001 GGAGCATTTTCTTGTGGAAATGGTCGCTGTATTGATCAGACTCTTCTCTGCAACAACGTG 3060
1001  G  A  F  S  C  G  N  G  R  C  I  D  Q  T  L  L  C  N  N  V  1020

3061 GATGACTGTGGTGATAGATCGGATGAAGATCCTTGTAGAAAACCTGCAAATGAAGAGGAG 3120
1021  D  D  C  G  D  R  S  D  E  D  P  C  R  K  P  A  N  E  E  E  1040

3121 GAACGCCTTTCTGTAATATTATGCAAGGAGGGAGAGTATACCTGCCACCCTCACGGAAAA 3180
1041  E  R  L  S  V  I  L  C  K  E  G  E  Y  T  C  H  P  H  G  K  1060

3181 AATGTTACAATCTGCTTGCCTTCCTCAGGAAGGTGCAATGGCACGGCTGAGTGCCCTCTT 3240
1061  N  V  T  I  C  L  P  S  S  G  R  C  N  G  T  A  E  C  P  L  1080

3241 GGAGATGATGAGAGAGGTTGTGGCTGCCAAGATTTTCAGTTCACATGTTACAATGGAAAA 3300
1081  G  D  D  E  R  G  C  G  C  Q  D  F  Q  F  T  C  Y  N  G  K  1100

3301 TGTATTCCATCGGAATGGGTGTGTGACGGGATCAATGACTGTGGTGATGGATCGGATGAA 3360
1101  C  I  P  S  E  W  V  C  D  G  I  N  D  C  G  D  G  S  D  E  1120

3361 AACAACGCACGCTGTCAGCTGCCATCTTCAGTTGGGACGCCAGGACCTTGCACTGATTAT 3420
1121  N  N  A  R  C  Q  L  P  S  S  V  G  I  P  G  P  C  T  D  Y  1140
```

```
3421 GCTTGCAATGACGGGCAATGCATCTCGTTGAGTTTAGCGTGTAATAACAAGAGGAATTGT 3480
1141  A  C  N  D  G  Q  C  I  S  L  S  L  A  C  N  N  K  R  N  C  1160

3481 GAAGATGGCTCTGATGAAGGCGGTCAGTGCGATATTGCCTGCAATGCAAAGTCTCCCTGT 3540
1161  E  D  G  S  D  E  G  G  Q  C  D  I  A  C  N  A  K  S  P  C  1180

3541 GATCAAATTTGCCAGCCTACCCTCGCGGCCCAAGATGCAACTGTTCACAAGGGCTATGTA 3600
1181  D  Q  I  C  Q  P  T  L  A  A  Q  D  A  T  V  H  K  G  Y  V  1200

3601 CTGAGCAGTGACGGAGCAAAATGTGGCGACATAGACGAGTGCGAGATTGGTGGTGCGTGT 3660
1201  L  S  S  D  G  A  K  C  G  D  I  D  E  C  E  I  G  G  A  C  1220

3661 GCACAGGTGTGCCACAACACCCGCGGAAGCTTCTCGTGTTCTTGTCATCCTGGTTTCCAG 3720
1221  A  Q  V  C  H  N  T  R  G  S  F  S  C  S  C  H  P  G  F  Q  1240

3721 CTTCGATCAGATCACGTGTCCTGTAAAGCATTAGGGGAACCAATGCAGTTCATATTTTCC 3780
1241  L  R  S  D  H  V  S  C  K  A  L  G  E  P  M  Q  F  I  F  S  1260

3781 GCTGGCAACCAGATCAGGAAAGTATCCCACGAACTCCGCTTCACTGATGTTGTGTACCCC 3840
1261  A  G  N  Q  I  R  K  V  S  H  E  L  R  F  T  D  V  V  Y  P  1280

3841 GAAGCTGAGTTGAAGGTTACAGGACTCGATGTGGATTCAGCATCAAATGAAGTTTACTGG 3900
1281  E  A  E  L  K  V  T  G  L  D  V  D  S  A  S  N  E  V  Y  W  1300

3901 AGCACAGATGTAACGAGCACAATTTACAGACTGTCACTGCGTGGAGGGGAGAAGGCGTAT 3960
1301  S  T  D  V  T  S  T  I  Y  R  L  S  L  R  G  G  E  K  A  Y  1320

3961 GCTACAGGAATTGGCACTCCAGGCGACATTGCTGTGGACTGGATCTCTCGGAATGTGTAT 4020
1321  A  T  G  I  G  T  P  G  D  I  A  V  D  W  I  S  R  N  V  Y  1340

4021 TACGTCGATAAAAGTACGCCCCAAGCTATAAGAGCGTGCAATTTGGATGAACATCGTTGT 4080
1341  Y  V  D  K  S  T  P  Q  A  I  R  A  C  N  L  D  E  H  R  C  1360

4081 GCTAAAGTGCTCGTCATCGAACATGGATTTTCTGTGCCAAAGATTGCAGTAGATCCTATT 4140
1361  A  K  V  L  V  I  E  H  G  F  S  V  P  K  I  A  V  D  P  I  1380

4141 GCTGGATTCATATTTTGGCCAGAAGTGACGAAGTGGGTTTTTGGAGAGCCCTCCACTGAT 4200
1381  A  G  F  I  F  W  P  E  V  T  K  W  V  F  G  E  P  S  T  D  1400

4201 TTGTTCCGTTCTGAGTTAACTGGTCGACACAAAACGATGATAGAGACGAACAATTTAATG 4260
1401  L  F  R  S  E  L  T  G  R  H  K  T  M  I  E  T  N  N  L  M  1420

4261 CTGGTGAATGGACTGACTTTGGATATTGTGCGGCAGAGGATATACTTTGCTGATCAGCAC 4320
1421  L  V  N  G  L  T  L  D  I  V  R  Q  R  I  Y  F  A  D  Q  H  1440

4321 AAGCGGACAATTGAGTGTATGGATTACAATGGAGAGGATAGGCACATCATCGTACACAAT 4380
1441  K  R  T  I  E  C  M  D  Y  N  G  E  D  R  H  I  I  V  H  N  1460

4381 GAGCATGTTCAAAACCCGATTGACATGGCGTTATTCGAGGGAACATTGTACTGGTTGACA 4440
1461  E  H  V  Q  N  P  I  D  M  A  L  F  E  G  T  L  Y  W  L  T  1480

4441 GCTGGAACAGGTCAGTTAACCAGCTACAAACTGTACGGTCCACATGAAAGGCGAATTGGA 4500
1481  A  G  T  G  Q  L  T  S  Y  K  L  Y  G  P  H  E  R  R  I  G  1500

4501 AAACTTCAGTTGTACATTTACAGTTCAGATCAGTTTACAATCTTGCAGCAAGCTATACAG 4560
1501  K  L  Q  L  Y  I  Y  S  S  D  Q  F  T  I  L  Q  Q  A  I  Q  1520
```

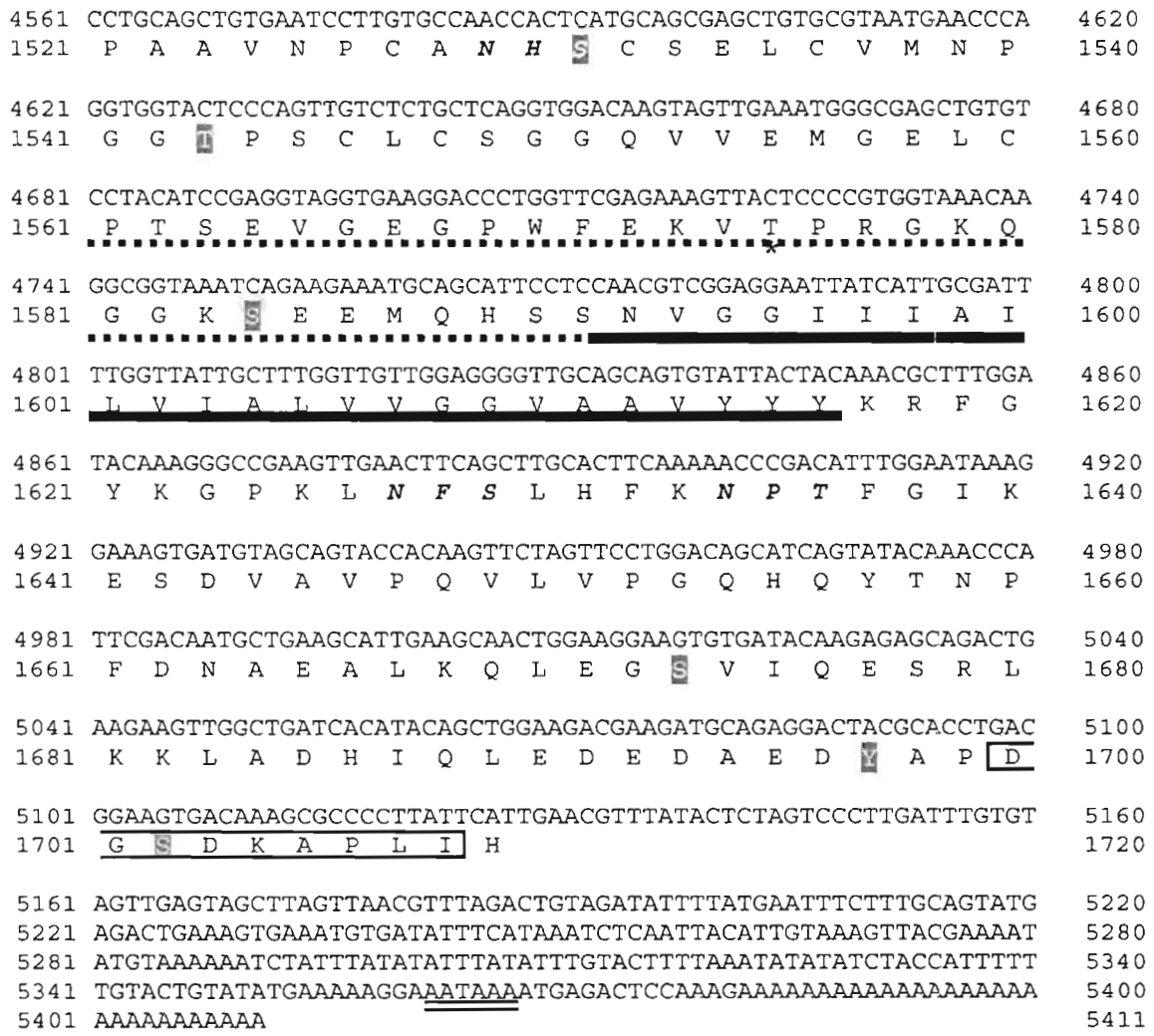


Fig. 3. Nucleotide sequence and deduced amino acid sequence of *P. americana* VgR cDNA. The amino acid sequence of 1709 residues has been derived from an ovarian cDNA of 5411 nucleotides in a single open reading frame. Sequences used for the degenerate primer (prepared from a conserved region (WSDW) of an EGF precursor homology domain of *Aedes aegypti* VgR and the YPR of *Drosophila melanogaster*) and the gene specific primer are shown with arrows. Possible glycosylation sites are shown with boldface-Italic-letters. Possible phosphorylated serine (S), threonine (T) and tyrosine (Y) residues as predicted by using the NetPhos 2.0 Computer Program (Blom et al., 1999) are shown with light-shaded frames. YWXD or potentially related sequences present in the C repeats (YWTD B-propeller domain) are shown with dark-shaded boxes. Potential *O*-linked sugar domain is dotted under-lined, whereas a transmembrane helix is shown with a bold line. The putative leucine-isoleucine internalization signal is boxed The putative signal peptide is underlined, whereas the consensus polyadenylation signal is double underlined. This sequence has been submitted to GenBank and assigned the accession number: AB077047.

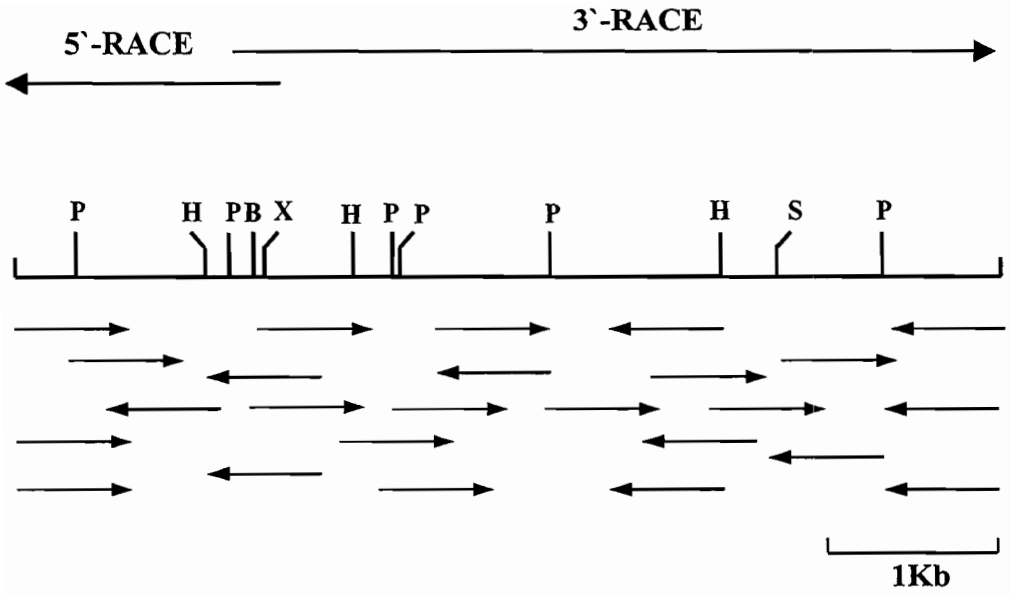


Fig.4. Restriction map and sequence strategy of the cloned *P. americana* VgR cDNA. P, *Pst*I; H, *Hind*III; B, *Bgl*II; X, *Xho*I and S, *Sal*I.

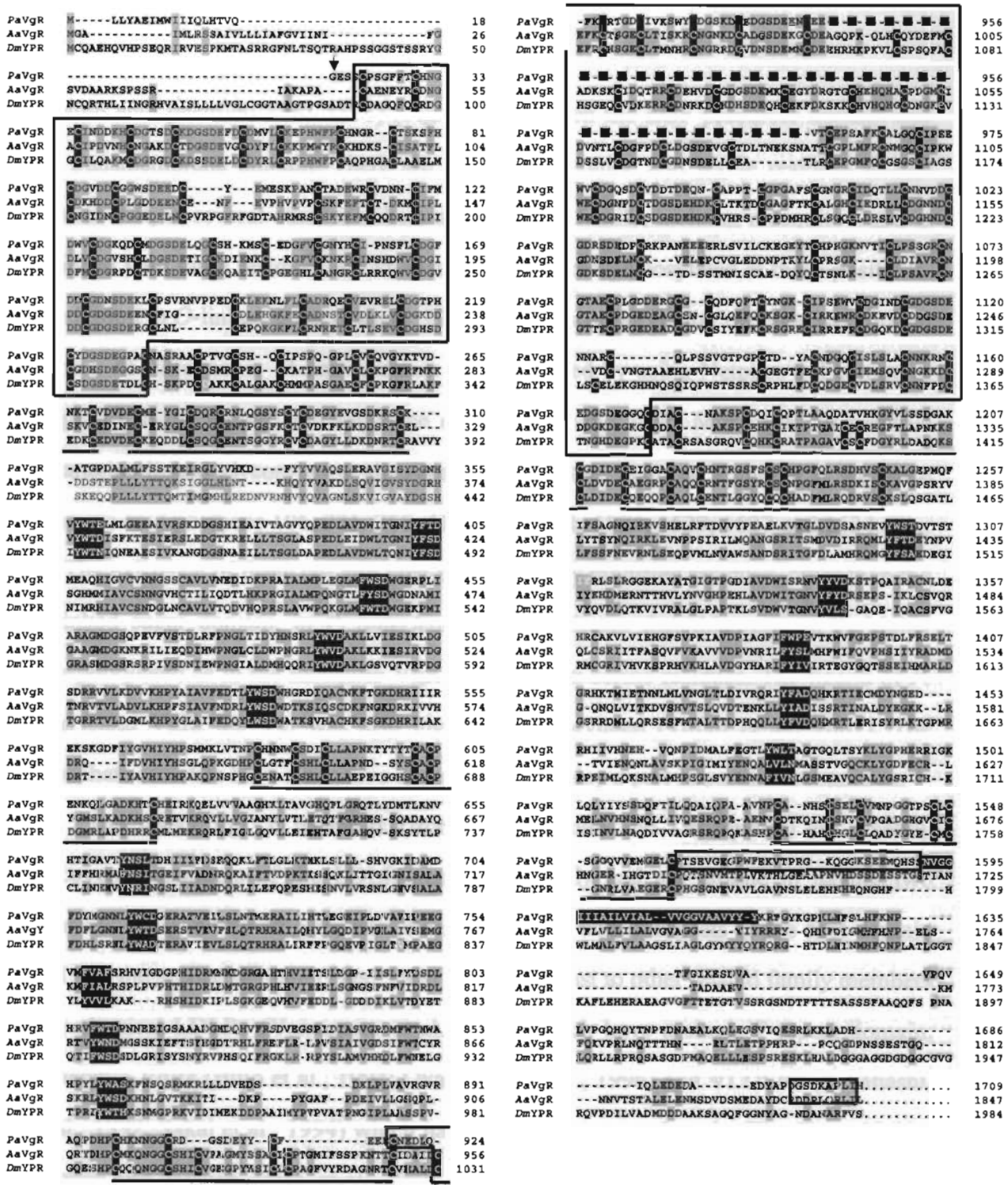
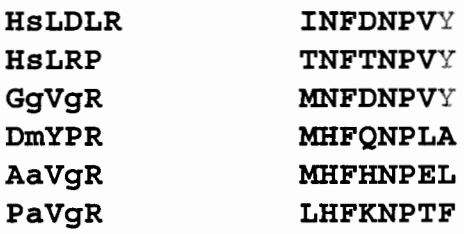


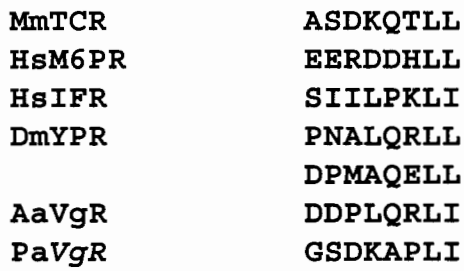
Fig. 5. The comparison of deduced amino acid sequence of *P.americana* VgR (*Pa*VgR) with those of *Aedes aegypti* VgR (*Aa*VgR) (Sappington et al 1996) and *Drosophila melanogaster* YPR (*Dm*YPR) (Schonbaum et al., 1995). The cysteine-residues are indicated with dark-shaded boxes. The two clusters (domains) of ligand binding repeats (class A repeats) are boxed. The epidermal growth factor (EGF)-like repeats (class B repeats) are underlined. YWXD or potentially related sequences present in the C repeats (YWTD B-propeller domain) are shown with dark-shaded frames. The identical residues are shown with the light-shaded frames. The putative *O*-linked sugar domain of *Pa*VgR is boxed, whereas the transmembrane helix is indicated with a dark-shaded frame. The putative signal peptide cleavage site of *Pa*VgR is shown with an arrow, whereas, the leucine-isoleucine internalization signal is shown with a box. The two missing repeats from the second ligand-binding domain of *Pa*VgR are indicated with a dotted-line.

(amino acid position: 1561-1591) (Fig. 3) in between the EGF precursor domain and the transmembrane domain. There are only six threonine and serine residues in this domain that, in several other members of the LDLR family, incluing mosquito VgR, is enriched with these residues and known to be *O*-linked glycosylated at these positions. In addition to *O*-linked sugar domain, 77 potential phosphorylation residues were found (serine: 45, threonone: 14 and tyrosine: 18) (predicted by using Netphos 2.0 computer program) in the amino acid sequence of *P. americana* VgR. Moreover, 10 putative asparagine-linked glycosylation sites are present in the deduced amino acid sequence at amino acid positions 106, 231, 266, 417, 595, 1061, 1074, 1529, 1627 and 1634 (Fig. 3). The existence of these co- and post-translationally modified residues, other than the *O*-linked sugar domain, in *P. americana* VgR indicates that this receptor is highly phosphorylated as in the mosquito VgR especially on serine residues. The ligand of *P. americana* VgR, the Vg, was also highly phosphorylated (Tufail et al., 2001). The phosphate moieties have a negative charge and may play a part in partner recognition.

The insect Vg/YP receptors (from the mosquito and *Drosophila*) are recently discovered memders of the LDLR gene superfamily and are unique in having two domains of ligand binding repeats (class A repeats) with 5 repeats in the first domain and 8 in the second domain (Figs. 5 and 7), in constrast to other related family members like classical LDLRs (Yamamoto et al., 1986) which have a single 7-repeat domain, vertebrate VgRs (Bujo et al., 1994; Okabayashi et al., 1996) and VLDLRs (Takahashi, et al., 1992; Sakai et al., 1994) which have a single 8-repeat domain (Fig. 7), and LRPs (human and chicken, Herz, et al., 1988; Nimpf et al., 1994) and megalins (Saito et al., 1994; Hjalm et al., 1996) which have four domains of 2-7, 8, 10 and 11 repeats. The cDNA coding for VgR from *P. americana* reveals that this receptor, like Vg/YP receptors of the mosquito and *Drosophila*, also contains two ligand binding domains but differs significantly from them in having only 6 repeats in the second ligand binding domain. In other words, the 2 repeats were missing from the second ligand binding domain of *P. americana* VgR. The computer-assisted sequence alignment



A



B

Fig. 6. Comparisons of alternative internalization signals found in the cytoplasmic domains of receptors. A) Tight-tyrosine internalization signal motif in human LDLR (HsLDLR) (Yamamoto et al., 1984), human low density related proteins (HsLRP) (Herz et al., 1988), chicken VgR (GgVgR) (Bujo et al., 1994), *Drosophila* YPR (DmYPR) (Schonbaum et al., 1995), mosquito VgR (*Aa*VgR) (Sappington et al., 1996), and the American cockroach VgR (PaVgR) (the present report). The tyrosine critical for internalization (Y) is missing in the insect Vg/YPRs. B) Di-leucine/leucine-isoleucine signal motif in mouse T-cell antigen receptor (TCR) (Letourneur and Klausner,1992), human cation-dependent mannose 6-phosphate receptor (HsM6PR) (Johnson and Kornfled, 1992), human interferon-g-receptor (HsIFR) (Farrarand Schreiber, 1993), and insect Vg/YPRs.

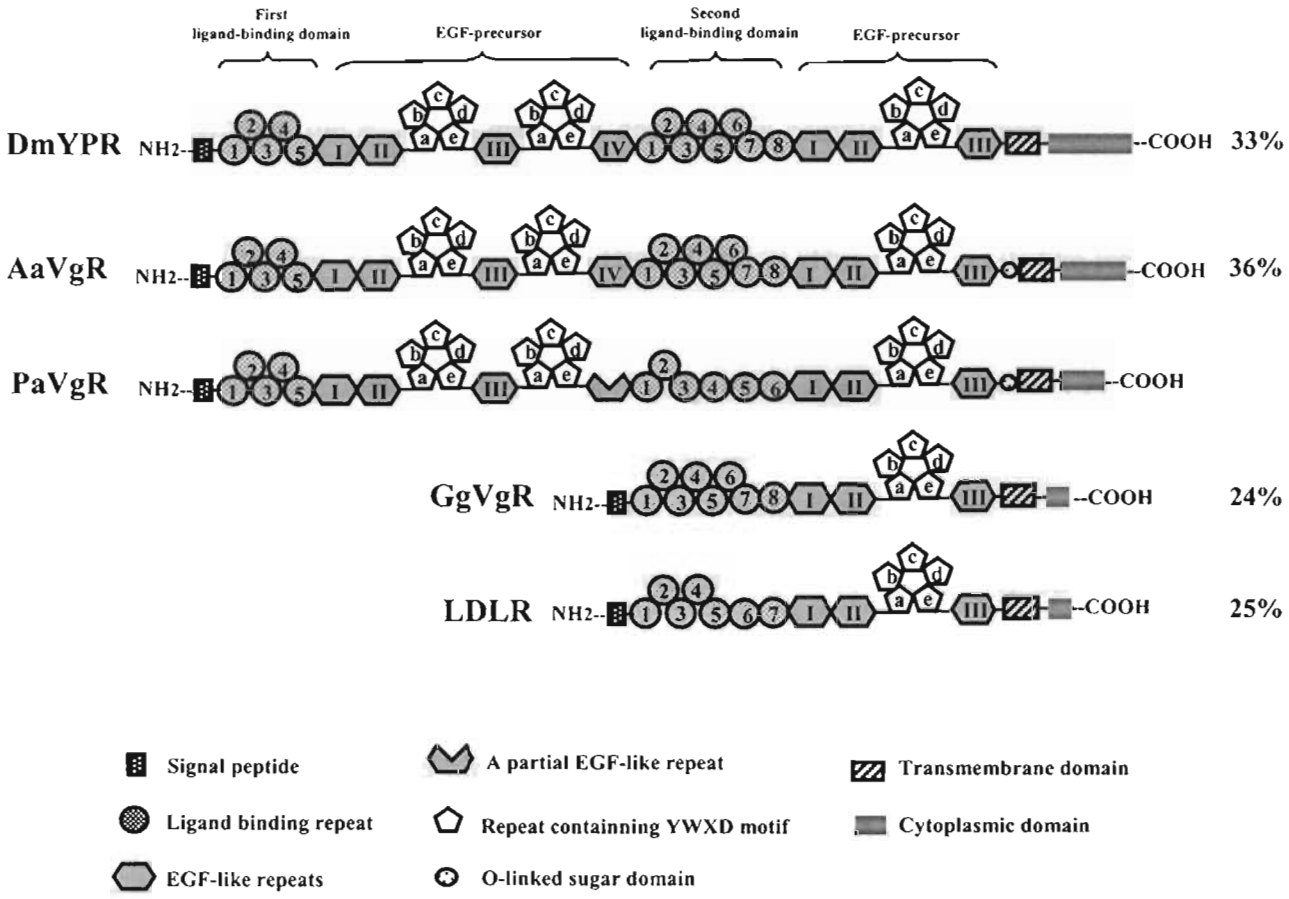


Fig. 7. Schematic comparison of PaVgR with those of AaVgR, DmYPR, chicken VgR (GgVgR) (Bujo et al., 1994) and the human LDLR (Yamamoto et al., 1986) to align the number and arrangement of different motifs/modules. The cysteine-rich repeats in the ligand binding domains are shown with numbers 1-8. The cysteine-rich repeats in the EGF-precursor domains are indicated with Roman numerals I-IV, whereas, the repeats containing YWXD motif are shown with letters a-e. The overall amino acid identity ratios (%) compared to PaVgR are shown on the right side.

(Clustal W) with mosquito VgR and *Drosophila* YPR revealed that repeats 2 and 3 were missing from this transcript (Fig. 5 and 7). Recently, a cDNA encoding ApoER2 was cloned from the mouse brain and was different by possessing only 5 ligand binding repeats as compared to the human ApoER2 which contains 7 repeats (Brandes et al., 2001). Furthermore, the *P. americana* VgR contains six complete and a paritial EGF-like repeats (class B repeats) as compared to the seven complete EGF-like repeats of the mosquito VgR and YPR of *Drosophila* (figs. 5 and 7). This partial EGF-like repeat exists between the last repeat containing the Y/FWXD motif and the first repeat of the second ligand binding domain and consists of only 17 residues (having only 3 cystein residues) in contrast to a complete repeat of 38 residues (having 6 cystein residues) at this position in Vg/YP receptors of the mosquito and *Drosophila*

Taken together, these structural differences establish that this newly discovered member of the LDLR family is a novel insect VgR harbouring only 6 repeats in the second ligand binding domain. However, why *P. americana* VgR possesses only 6-repeats in the second ligand-binding domain is unclear. One possibility that the primordial VgRs (like that of *P. americana* from hemimetabola) may have essentially a second-ligand binding domain of only 6-repeats and that the advanced Vg/YP receptors (like that of the mosquito and *Drosophila* which represent holometabola) may have acquired 2 additional repeats in the second ligand binding domain during evolution. The other possibility which seems more attractive that the second ligand binding domain of *P. americana* VgR might be subjected to an event of alternative splicing which has made this receptor different from those of other insect Vg/YP receptors, as was observed in ApoER2 of the mouse (Brandes et al., 2001).

Although, *P. americana* VgR is different from other insect Vg/YP receptors in number of ligand-binding repeats but still shows high homology with them. Compared to the mosquito VgR and YPR of *Drosophila*, the primary sequence of *P. americana* VgR shows 36% and 33% overall amino acid identity to their corresponding gene products (Fig. 5 and 7). Though the chicken VgR (Bujo et al., 1994) and LDLR

(Yamamoto et al., 1986) are half of the size of insect Vg/YP receptors, both were showing significant similarity in modular arrangement with the three insect Vg/YP receptors (Fig. 7) and were showing 24% and 22% identity in the primary protein with that of *P. americana* VgR (data not shown). The cockroach VgR is more close to the mosquito VgR than to *Drosopila* YPR, chicken VgR and LDLR, not only in having high identity (36%) in the primary sequence but also in having a putative *O*-linked sugar domain between the last EGF-like repeat and the transmembrane domain, the consensus location for such regions in the LDLR gene family members. The presence of *O*-glycosylation sites in these two insect VgRs, which are often found in somatic cell-specific members of the LDLR family (Bujo et al., 1995), is surprising. The function of this domain is not known, however, it is thought that this sugar domain keep the LDLR extended. If this is true, then why some members of this family need extension is still a mystery.

The cytoplasmic tail of *P. americana* VgR, like the mosquito VgR, contains a leucine-isoleucine (LI) internalization signal (Fig. 6B), a recently identified signal required for internalization via clathrin coated pits in many receptors (see review: Sappington and Raikhel, 1998). The presence of leucine-isoleucine internalization signal in the three insect Vg/YP receptors (di-leucine in *Drosophila*YPR) (Fig. 6B) suggests that these receptors are more strict in harbouring this signal than a tight-turn-tyrosine signal (NPXY) found in the majority of LDLR family receptors. Although insect Vg/YP receptors have a sequence motif similar to the NPXY motif in their cytoplasmic tails (Fig. 6A) but the crucial tyrosine (Davis et al., 1987) necessary for internalization is missing (Fig. 6A). In *P. americana* VgR, however, the crucial tyrosine (Y) is replaced by phenylalanine (F) which had been shown previously to substitute effectively for the tyrosine in the LDL receptor (Davis et al., 1987).

Since the cloning of the human LDLR (Yamamoto et al., 1984), the prototype of LDLR family, an increasing number of related proteins have been discovered. These can be divided in receptors with a single ligand binding domain and EGF-precursor

domain, such as the classical LDLRs (Yamamoto et al., 1986), vertebrate VgRs (Bujo et al., 1994; Okabayashi et al., 1996), Nematode VgR (Grant and Hirsh, 19999 and very LDLRs (Takahashi, et al., 1992; Sakai et al., 1994) (Fig. 7) and receptors possessing multiple clusters of these two domains, such as the LDLR-related proteins (human and chicken, Herz, et al., 1988; Nimpf et al., 1994), megalins (Saito et al., 1994; Hjalm et al., 1996) and insect Vg/YP receptors (Sappington et al., 1996; Schonbaum et al, 1995). The *P. americana* VgR like other insect Vg/YP receptors also belongs to group having multiple clusters. Our homology analysis (Fig. 8) based on amino acid sequences of the 8 LDLR family members (5 representing the group having a single cluster, 3 (from insects) representing the group having multiple clusters) indicates that the subfamilies with a single cluster and those with multiple clusters evolved independently from a common ancestor with a single cluster of ligand binding and EGF-precursor domains and is in agreement with those reported previously (Sappinngton and Raikhel, 1998b). Moreover, we observe that the mosquito VgR and YPR of *Drosophila* make a single clade which points for their more common ancestors than with the cockroach VgR.

The binding of LDLR family members with multiple ligands having functional diversity is intriguing. The chicken VgR, for example, recognizes at least eight different ligands (Hiesberger et al., 1995; Jacobsen et al., 1995). LDLR-related protein binds more than 20 different ligands (Strickland et al., 1995; Kounnas et al., 1996). The mammalian very LDL receptor has been shown to bind apoE specifically, whereas apoB, a ligand of the LDL receptor, does not interact with this receptor (Takahashi et al., 1992). In insects, the structurally very similar receptors for Vg/YP from the mosquito (Sappington et al., 1996) and *Drosophila* (Schonbaum et al, 1995) recognize quite unrelated ligands. The mosquito Vg is homologus to insect Vgs (including the cockroach Vgs) (see Tufail et al., 2001) whereas the mosquito VgR is more structurally close to *Drosophila* YPR which recognizes the Yps, a product homologus to lipoprotein lipases (Bownes et al., 1988.) which is recognized by LDLR-related protein that mediates binding of B-VLDL and chylomicron remnants (Beisiegel, 1996). Thus,

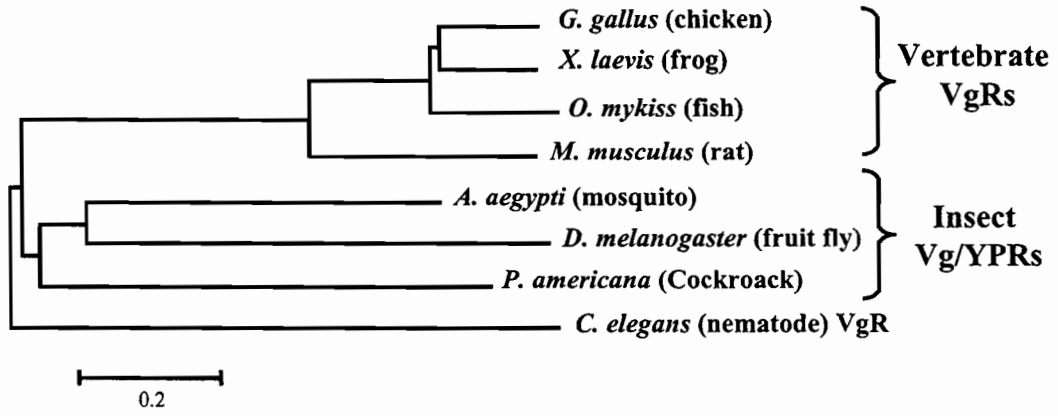


Fig. 8. Phylogenetic relationships among 8 LDLR family members. A multiple alignment of amino acid sequences was performed using Clustal W (Thompson et al., 1994) and used an input for a neighbour-joining tree construction program (MEGA version 2, Kumar et al.,2001). The scale indicates distance (number of amino acid substitutions per site.

how insect Vg/YP receptors interact with their ligands and how the ligand specificity is determined are the next challenges.

We already know the molecular structures of Vgs, the ligand, from *P. americana* (Tufail et al., 2000 and 2001). Now, after cloning and characterizing the complete structure of their counterpart, the receptor, we are in a position to gain better understanding of the ligand-binding interactions, which is not only of scientific interest but also will help in exploring the potential manipulation of insect receptor-endocytotic machinery for vitellogenesis.

We propose that the loss of 2 repeats from the second ligand-binding domain of *P. americana* VgR should be of evolutionary importance. The mouse ApoER2 with 5 ligand-binding repeats, for example, binds only to β-VLDL and reelin but not recognize α–macroglobulin which binds to the avian homologue of ApoER2 harboring 8 ligand binding repeats (Brandes et al., 2001). If this is true than the loss or addition of ligand binding repeats in VgR might result in a complex functional pattern of its products in insects and other species.

Next, during cloning of VgR, we also found a partial clone encoding receptor tyrosine kinase (RTK) from the previtellogenic ovaries of *P. americana*. The remining portion of the RTK was obtained through 5`-RACE-PCR. The complete cDNA for RTK was thus obtained from two overlapping RACE-PCR fragments of 3'- and 5' end and was of 4128 residues which encoded a deduced amino acid sequenece of 1294 residues long including 22 residues for the putative signal peptide (Fig. 9). The deduced amino acid sequence was aligned well with other members of the RTK family of receptors. The comparison of *P. americana* RTK with that of the human receptor protein tyrosine kinase (TIE2) (Ziegler et al., 1993) is shown in Fig. 10. In human, the receptor protein tyrosine kinase is specifically expressed in developing vascular endothelial cells and is believed to constitute the earliest endothelial cell lineage marker, and may regulate the endothelial cell proliferation, differentiation and proper patterning during blood vessel formation. The role of *P. americana* RTK in

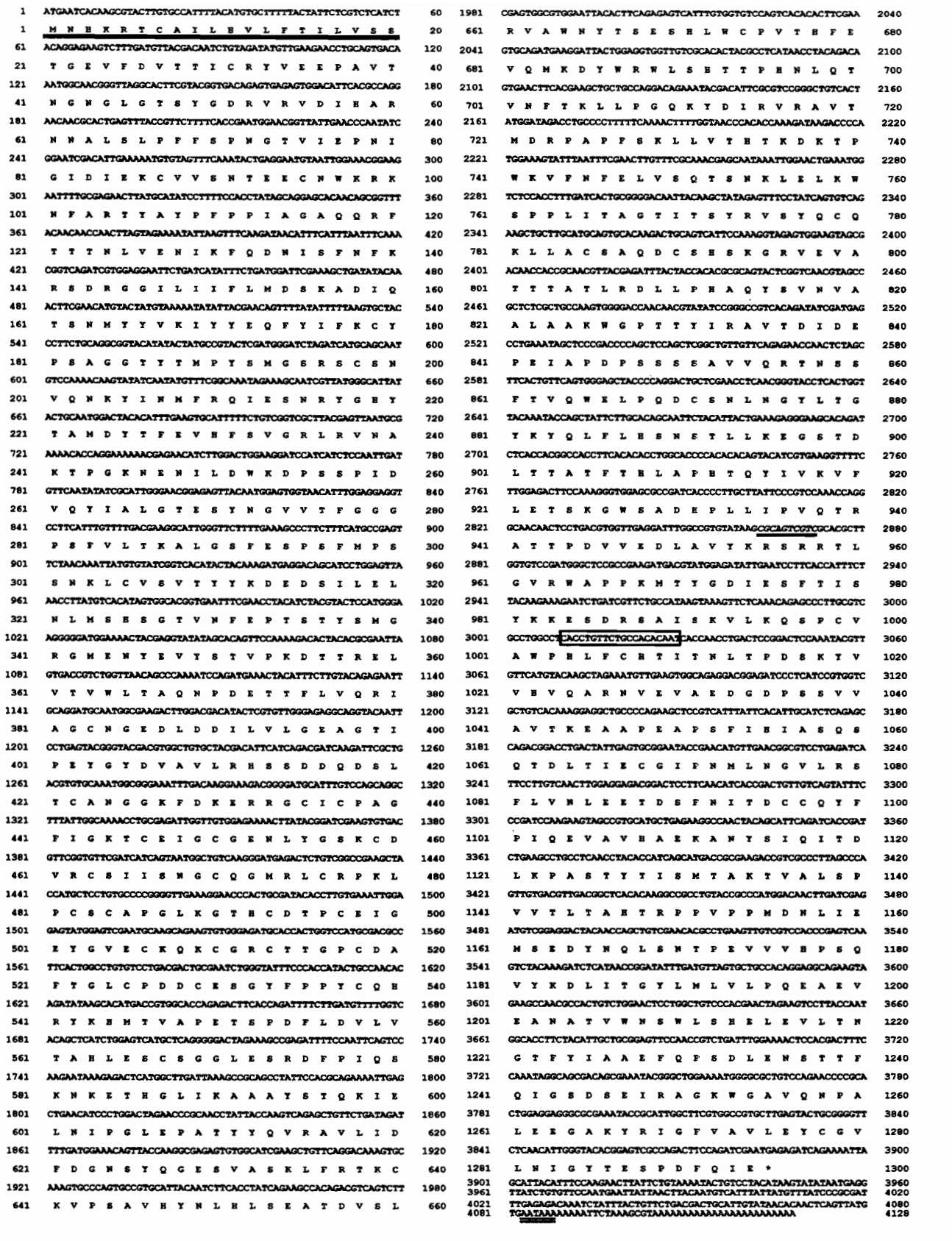


Fig. 9. Nucleotide sequence and deduced amino acid sequence of the receptor tyrosine kinase (RTK) cloned from previtellogenic ovaries of the American cockroach, *P. americana*. The putative signal peptide is shown with a bold underline. A consensus polyadenylation signal is double underlined. The initial primer is shown with underlined-Italic letters. The gene-specific primer is shown with a box. Whereas, asterik indicates the stop codon

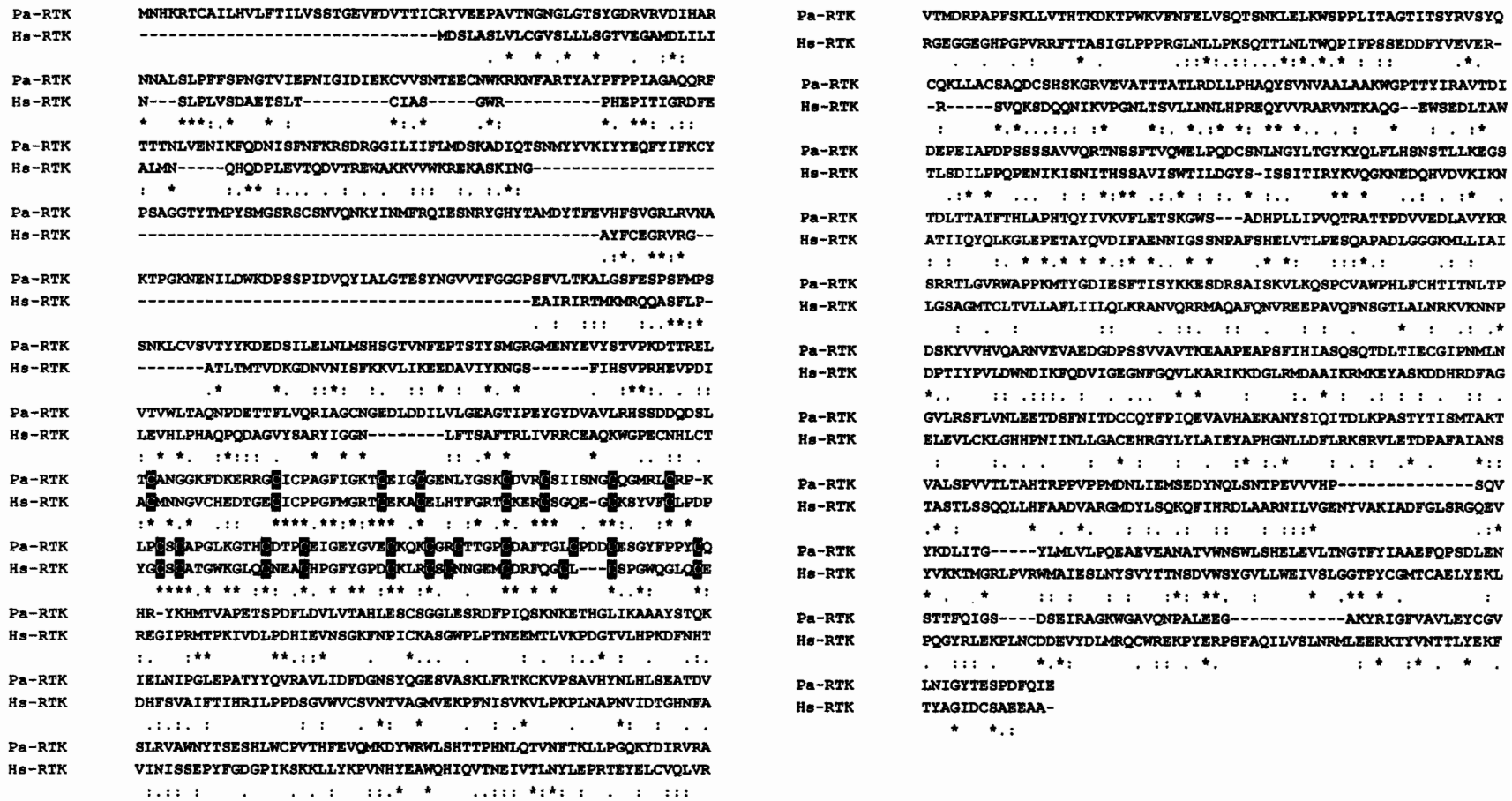


Fig. 10. The alignment of the deduced amino acid sequence of *P. americana* RTK (Pa-RTK) with that of the human receptor protein tyrosine kinase (Hs-RTK) using the Clustal W computer program. The deduced amino acid sequence of Hs-RTK was obtained from the data base (Accession no: Q02763). *, residues are identical in both sequence; ., residues are functionally conserved. The cysteine residues in the EGF-like domains are shown with dark-shaded frames.

previtellogenic ovaries, however, is not clear at the moment. One possibility is that RTK may be involved in cell proliferation in the developing ovaries, and especillay the follicle cells.

5.5 LITERATURE CITED

Beisiegel, U., 1996. New aspects on the role of plasma lipases in lipoprotein catabolism and atherosclerosis. Atherosclerosis, 124, 1-8.

Bownes, M., Shirras, A., Blair, M., Collins, J., Coulson, A., 1988. Evidence that insect embryogenesis is regulated by ecdysteroids released from yolk proteins. Proc. Natl. Acad. Sci. USA. 85, 1554-1557.

Brandes, C., Kahr, L., Stockinger, W., Hiesberger, T., Schneider, W. J., Nimpf, J., 2001. Alternative Splicing in the Ligand Binding Domain of Mouse ApoE Receptor-2 Produces Receptor Variants Binding Reelin but Not Alpha-2Macroglobulin. J. Biol. Chem. 276, 22160-22169.

Bujo, H., Yamamoto, T., Hayashi, K., Hermann, M., Nimpf, J., Schneider, W. J., 1995. Proc. Mutant oocytic LDL receptor gene family member causes atherosclerosis and female sterility. Natl. Acad. Sci. USA. 92, 9905-9909.

Byrne, B. M., Gruber, M., Ab, G., 1989. The evolution of egg yolk proteins. Progr Biophys Molec Biol 53: 33-69

Chen, W. J., Goldstein, J. L., Brown, M. S., 1990. NPXY, a sequence often found in cytoplasmic tails, is required for coated pit-mediated internalization of the low density lipoprotein receptor. J Biol Chem. 265: 3116-3123.

Davail, B., Pakdel, F., Bujo, H., Perazzolo, L. M., Waclawek, M., Schneider, W. J., Le Menn, F., 1988. Evolution of oogenesis: the receptor for vitellogenin from the rainbow trout. J Lipid Res 39: 1929-1937.

David, S. R., Matthew, G., Brandy, C., Debra, M. H., Suzanne, S., Danielle, S., 2001. Yolk protein endocytosis by oocytes in *Drosophila melanogaster*: immunofluorescent localization of clathrin, adaptin and the yolk protein receptor. 47, 715-723.

Davis, C. G., van Driel, I. R., Russell, D. W., Brown, M. S., Goldstein, J. L., 1987. The Low Lipoprotein Receptor IDENTIFICATION OF AMINO ACIDS IN CYTOPLASMIC DOMAIN REQUIRED FOR RAPID ENDOCYTOSIS. J. Biol. Chem. 262, 4075-4082.

DiMario, P. J., Mahowald, A. P., 1987. Female sterile (1) yolkless: a recessive female sterile mutation in *Drosophila melanogester* with depressed numbersof coated pits and coated vesicles within the developing oocytes. J Cell Biol 105: 199-206.

Goldstein, J. L., Brown, M. S., Anderson, R. G. W., Russell, D. W., Schneider, W. J., 1985. Receptor-mediated endocytosis: concepts emerging from the LDL receptor system. Ann. Rev. Cell Biol. 1, 1-39.

Grant, B., Hirsh, D., 1999. Receptor-mediated Endocytosis in the *Caenorhabditis elagans* Oocyte. Mol Biol Cell 10, 4311-4326.

Hafer, J., Fisher, A., Ferenz, H. J., 1992. Identification of the yolk receptor in oocytes of *Nereis virens* (Annelida, *polychaeta*). J Comp B 162, 148-152.

Herz, J., Hamann, U., Rogne, S., Myklebost, O., Gausepohl, H., Stanley, K. k., 1988. Surface location and high affinity for calcium of a 500-kd liver membrane protein closely related to the LDL-receptor suggest a physiological role as lipoprotein receptor. EMBO J. 4111-4127.

Hiesberger, T., Hermann, M., Jacobsen, L., NovaK, S., Hodits, R. A., Bujo, H., 1995. The chicken oocyte receptor for yolk precursors as a model for studying the action of receptor-associated protein and lactoferrin. J. Biol. Chem. 270, 18219-18226.

Hjalm, G., Murray, E., Crumley, G., Harazim, W., Lundgren, S., Onyango, I., 1996. Cloning and sequencing of human gp330, a Ca+-binding receptor with potential intracellular signaling properties. Eur. J. Biochem. 239, 132-137.

Jacobsen, L., Hermann, M., Vieira, P. M., Schneider, W. J., Nimpf, J., 1995. The chicken oocyte for lipoprotein deposition recognizes alpha2-macroglobulin. J. Biol. Chem. 270, 6468-6475.

Kounnas, M. Z., Church, F. C., Argraves, W. S., Strickland, D. K., 1996. Cellular internalization and degradation of antithrombin III-thrombin, heparin cofactor II-thrombin, and a_1-antitrypsin-trypsin complexes is mediated by the low density lipoprotein receptor-related protein. J. Biol. Chem. 271, 6523-6529.

Mac Lachlan, I., Nimpf, J., Schneider, W. J., 1994. Avian riboflavin binding protein binds to lipoprotein receptors in association with vitellogenin. J. Biol. Chem. 269, 24127-24132.

Mukherjee, S., Ghosh, R. N., Maxfield, F. R., 1997. Endosysis. Physiol. Rev. 77, 759-803.

Nielsen, H., Engelbrecht, J., Brunak, S., von Heijne G., 1997. Identificati of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Prot.Eng. 10, 1-6.

Nimpf, J., Stifani, S., Bilous, P. T., Schneider, W. J., 1994. The somatic cell specific LDL receptor-related protein of the chicken: close kinship to mammalian LDL receptor gene family members. J. Biol. Chem. 269, 212-219.

Okabayashi, K., Shoji, H., Nakamura, T., Hashimoto, O., Asashima, M., Sugino, H., 1996. cDNA cloning and expression of the *Xenopus laevis* vitellogenin receptor. Biochem. Biophys. Res. Commun. 224, 406-413.

Part, F., Coward, K., Sumpter, J. P., Tyler, C. R., 1998. Molecular characterization and expression of two ovarian lipoprotein receptors in the rainbow trout *Oncorhynchus mykiss*. Biol. Reprod. 58, 1146-1153.

Raikhel, A. S., Dhadialla, T. S., 1992. Accumulation of yolk proteins in insect oocytes. Annu. Rev. Entomol. 37, 217-251.

Roehrkasten, A., Ferenz, H. J., 1986. Properties of the vitellogenin receptor of isolated locust oocyte membranes.Int J Invert. Reprod. Dev. 10, 133-142.

Roehrkasten, A., Ferenz, H. J., Buschmann-Gebhardt, B., Hafer, J., 1989. Isolation of the vitellogenin-binding protein from locust ovaries. Arch. Insect Biochem. Physiol 10, 141-149.

Saito, a., Pietromonaco, S., Loo, A. K.-C, Farquhar, M. G., 1994. Complete cloning and sequencing of rat gp330/`megalin`, a distinctive member of the low density lipoprotein receptor gene family. Proc. Natl. Acad. Sci. USA. 91, 9725-9729.

Sakai, J., Hirano, A., Takahashi, H., Miura, Y., Ishii, H., Suzuki, H., Kawarabayasi, Y., Yamamoto, T., 1994. Structure, chromosome location and expression of the human very low density lipoprotein receptor gene. J. Biol. Chem. 269, 2173-2182.

Sappington, T. W., Hays, A. R., Raikhel, A. S., 1995. Mosquito vitellogenin receptor-purification, developmental and biochemical characterzation. Insect Biochem. Mol. Biol. 25, 807-817.

Sappington, T. W., Kokoza, V. A., Raikhel, A. S., 1996. Molecular characterization of the mosquito vitellogenin receptor reveals unexpected high homology to the *Drosophila* yolk protein receptor. Proc. Natl. Acad. Sci. USA. 93, 8934-8939.

Sappington, T. W., Raikhel, A. S., 1998a. Molecular characteristics of insect vitellogenins and vitellogenin receptors. Insect Biochem. Mo.l Biol. 28, 277-300.

Sappington, T. W., Raikhel, A. S., 1998b. Ligand-binding domains in vitellogenin receptors and other LDL-receptor family members share a common ancestral ordering of cysteine-rich repeats. J. Molec. Evol. 46, 476-487.

Schneider, W. J., 1996. Vitellogenin receptors: oocyte specific members of the low-density lipoprotein receptor supergene family. Int. Rev. Cytol. 166, 103 137.

Schonbaum, C. P., Lee, S., Mahowald, A. P., 1995. The *Drosophila* yolkless gene encodes a vitellogenin receptor belonging to the low density lipoprotein receptor superfamily. Proc. Natl. Acad. Sci. USA. 92, 1485-1489.

Springer, T. A., 1998. An extracellular beta-propeller module predicted in lipoprotein and scavenger receptors, tyrosine kinases, epidermal geowth factor precursor and extracellular matrix components. J. Mol. Biol. 283, 837-862.

Stifani, S., Barber, D. L., Nimpf, J., Schneider, W. J., 1990. A single chicken oocyte plasma membrane protein mediates uptake of very low density lipoprotein and vitellogenin. Proc. Natl. Acad. Sci. USA. 87, 1955-1959.

Strickland, D. K., Kounas, M. Z., Argraves, W. S., 1995. LDL receptor-related protein: a multiligand receptor for lipoprotein and proteinase catabolis. FASEB J. 9, 890-898.

Takahashi, S., S. Kawarabayasi, T. Nakai, J. Sakai, T. Yamamoto, 1992. Rabbit low density lipoprotein receptor-like protein with distinct ligand specificity. Proc. Natl. Acad. Sci. USA. 89, 9252-9256.

Tufail, M., Hatakeyama, M., Takeda, M., 2001. Molecular evidence for two vitellogenin genes and processing of vitellogenins in the American cockroach, *Periplaneta americana*. Arch. Insect Biochem. Physiol. 48, 72-80.

Tufail, M., Lee, J. M., Hatakeyama, M., Oishi, K., Takeda, M., 2000. Cloning of vitellogenin cDNA of the American cockroach, *Periplaneta americana* (Dictyoptera), and its structural and expression analyses. Arch. Insect Biochem. Physiol. 45, 37-46

Tufail, M., Takeda, M., 2002. Vitellogenin of the cockroach, *Leucophaea maderae*: nucleotide sequence, structure and analysis of processing in the fat body and oocytes. Insect Biochem Mol Biol. 32, 117-124.

Yamamoto, T., Bishop, R. W., Brown, M. S., Goldstein, G. L., Russell, D. W., 1986. Deletion in cysteine-rich region of LDL receptor impedes transport to cell surface in WHHL rabit. Science. 232, 1230-1237.

Yamamoto, T., Davis, C. G., Brown, M. S., Schneider, W. J., Casey, M. L., Golgstein, J. L., 1984. The human LDL receptor: a cysteine rich protein with multiple Alu sequences in its mRNA. Cell. 39, 27-38.

Ziegler, S. F., Bird, T. A., Schneringer, J. A., Schooley, K. A., Baum, P. R., 1993. Molecular cloning and characterization of a novel receptor protein tyrosine kinase from human placenta. Oncogene. 8, 663-670.

CHAPTER 6

GENERAL SUMMARY

In the present study, we report on structural characteristics of Vgs by cloning three complete Vg cDNAs, the two (Vg1 and Vg2) being that of the American cockroach, *Periplaneta americana* (Tufail et al., 2000 and 2001) whereas, the one being that of the Madeira cockroach, *Leucophaea maderae* (Tufail and Takeda, 2002). We show the use of cloned Vg cDNAs as a probe in Northern blot analyses to assess the tissue- and sex-specific expression of Vg gene in both the cockroach species, and also demonstrate that the Vg is a conserved molecule and can be used as a molecular marker to indicate the phylogenetic relationships. We also clarify, on molecular basis, the processing patterns of cloned Vgs from both the cockroach species. We also present the similarity in Vn-antigenicity among 10 cockroach species.

Moreover, we also found another cDNA coding for Vg from *L. maderae* which has stretches of amino acid sequences different from the one reprted previously (Tufail and Takeda, 2002).

Next, we report for cloning of VgR from previtellogenic ovaries of the American cockroach, *P. americana*. We demonstrate that this VgR is a novel LDLR family member with only six repeats in the second ligand binding domain. Moreover, we also report for cloning of a cDNA that encodes receptor tyrosine kinase (RTK) from the previtellogenic ovaries of the American cockroach, *periplaneta americana*.

For cloning of *P. americana* Vgs, a cDNA expression library was constructed from the mRNA purified from the fat body cells of the vitellogenic females. The library was screened, using the anti-*P. americana* 100 kD multiple Vn subunits antiserum. The clones obtained were sequenced and analyzed. The sequence analysis revealed that all the clones were derived from a single Vg molecular species. The complete Vg cDNA obtained was 5854 bp long, which encoded 1896 amino acids (including a signal peptide) in a single open reading frame followed by termination codon (TGA) and a

consensus polyadenylation signal (AATAAA). The entire amino acid sequence deduced was confidently aligned with known insect Vgs. The existence of the motif GL/ICG and the cystein residues was noted at coserved locations near the C-terminal. Moreover, a DGXR motif was observed 13-15 residues upstream of a GL/ICG motif in all the insect Vgs.

The Vns/Vgs of *P. americana* are consisted of three major polypeptides (170, 100 and 50 kD) and a minor polypeptide (150 kD) both in the egg (terminal mature oocyte) extract and in the female hemolymph on the SDS-PAGE. The N-terminal amino acid sequences of these Vn/Vg polypeptides (except for that of 100 kD) were determined and compared with the amino acid sequence deduced from the clone. The determined N-terminal amino acid sequence for 170 kD polypeptide was matched with an amino acid sequence deduced from Vg cDNA following a consensus cleavage site (RTRR), indicating the occurrence of post-translational cleavage in fat body cells as in most insect Vgs, whereas this was not the case with 150 and 50 kD polypeptides. These results indicated that the cloned cDNA encoded one Vg (Vg1), and that the 170 kD polypeptide was a product of Vg1 molecule, whereas the 150 and 50 kD polypeptides were the products of the other Vg (Vg2) molecular species and that 100 kD polypeptide was a contaminant from multi-subunits. These results thus implied the existence of two Vg genes in *P. americana*. Our immunobloting results were also showing the existence of two different antigenic Vg molecules. We, thus, concluded on the bases of the N-terminal amino acid sequence analysis and the Western blot analysis, that there are at least two Vg genes in *P. americana* and each of their products undergoes post-translational processing producing subunits with smaller molecular mases.

To clarify the above results, on molecular basis, the efforts were then made to clone the second Vg precursor of *P. americana*. To achieve the purpose, the cDNA expression library was screened again using the anti-*P.americana* 100 kD polypeptide antiserum which contains multiple Vn subunits. The extensive screening of the library yielded clones which encoded two Vg cDNAs (Vg1 and Vg2). It is to be noted that here

we screened approximately double the phages previously used. The complete nucleotide sequence obtained for the second Vg (Vg2) cDNA was 5,826 bp long, which encoded 1,876 amino acids (including 16 residues for a putative signal peptide) in a single ORF followed by a termination codon (TAA) and a consensus polyadenylation signal (AATAAA). The deduced amino acid sequene had 30% amino acid identity with that of the first Vg molecule, and was showing all the characteristic features found in other insect Vgs.

The chemically determined N-terminal amino acid sequences for 150 and 50 kD Vn polypeptides matched exactly with each other and with the deduced N-terminal amino acid sequence of the Vg2 cDNA. We were, however, unable to determine the N-terminal amino acid sequence for the 100 kD Vn polypeptide, because the latter seems to be a single band on SDS-PAGE but it consists of multiple Vn subunits due to post-translational processing.

The present results thus demonstrate, on molecular basis, that (1) there exist two Vg (Vg1 and Vg2) genes in *P. americana*, (2) the 170 kD polypeptide and one of the 100 kD multisubunits are originated from the Vg1 molecule (3) the 150 kD polypeptide, one of the 100 kD multisubunits and the 50 kD polypeptide are originated from the other Vg (Vg2) molecular species.

Northern blot analysis clearly demonstrates that the synthesis of Vg is limited only to the fat body cells of adult females. No trace of Vg mRNA was detected in the adult male fat bodies nor in the ovaries. Moreover, the Vg gene starts expressing in two days old adult female fat body cells, whereas hemolymph Vg was first detected by immunobloting in 4/5 days old adult females, two days after the gene expression. Thus, the present results correspond well with the data on protein synthesis, indicating the Vg gene is regulated at the transcriptional level under these conditions.

A phylogenetic (neighbour-joining) tree based on Vg sequences of 11 insect species was in agreement with those constructed previousely based on the morphological characteristics and the molecular markers.

The similarity in Vn- antigenicities among 10 cockroach species, belonging to 2 superfamilies, using the anti-*P. americana* Vn-antisera shows that the Vn antigenicities are at least very similar within the members of the same superfamily. The only exception was that of *Leucophaea maderae* (Blaberoidea), where a 90 kD Vn polypeptide was antigenically related to that of *P. americana* (Blattoidea). This antigenic relatedness between these two species raised our interest to clone/analyze the Vg cDNA of *L. maderae,* and to compare it with *P. americana* and other known cockroach Vgs to clarify the phylogenetic relationship at the molecular level.

The cloning of *L. maderae* Vg cDNA was done through RACE-PCR. The complete nucleotide sequence obtained was of 5920 bp which encoded a deduced amino acid sequence of 1913 residues in a single open reading frame. We note that certain characteristics common to insect Vgs, for example a GLI/CG motif and a number of cysteine residues at conserved locations near the C-terminal are present. The deduced amino acid sequence contained a serine-rich at the C-terminal region. This stretch occurred also in Vgs of *P. americana* (Vg1 and Vg2) and *Blatella germanica.* The Vg of *L. maderae* had 26% and 31% homology with those of *P. americana* (Vg1 and Vg2) and *B. germanica*, respectively. The only difference to other insect Vgs was that the DGXR motif was absent in *L. maderae* Vg. There exist 11 consensus cleavage site sequences (RXXR), and at least three of them are actually used for post-translational cleavage in the fat body.

Amino-terminal sequence analysis demonstrates that *L. maderae* pro-Vg was cleaved into four polypeptides following the three consensus RXXR cleavage site sequences, which were secreted as four Vg polypeptides (apparent molecular weights = 112, 100, 92 and 55 kD), sequestered, and deposited in the egg as four respective Vn polypeptides. There was, however, an additional 90 kD Vn polypeptide existed in the egg. The present results show that this polypeptide is a processed product from 92 kD Vn polypeptide.

Northern blot analysis of poly $(A)^+$ RNA reveals that the Vg gene is expressed

only in the female fat body cells. No trace of hybidization signal was detected from male fat body or from ovaries, despite the amount of total RNAs for them being twice as much as that of the female fat body RNAs.

A phylogenetic (neighbour-joining) analysis based on four Vg sequences from three cockroach species gave a tree that was similar to that constructed based on the DNA sequences of mitochondrial 12S rRNA genes, and was also in agreement with the most widely suggested phylogeny of cockroaches proposed by McKittrick (1964), based on the morpho-ethological characteristics.

Moreover, we have also cloned and sequenced another Vg cDNA of *L. maderae* which has stretches of amino acid sequences different from the one reported previously (Tufail and Takeda, 2002). The complete nucleotide sequence of this Vg (which we named as Vgb) consists of 5915 bp which encodes a deduced amino acid sequence of 1911 amino acid residues (including a putative signal peptide sequence) in a single open reading frame. The comparison of base sequences of both *L. maderae* Vgs revealed that the difference was due to mutations (addition/deletion of a base (s) or addition of a base (s) at one position followed by deletion of a base (s) at another position and vice versa) in the base sequence of Vgb (the other Vg) which made its amino acid sequence different from the one reported previously (Tufail and Takeda, 2002). The both *L. maderae* Vgs were, however, showing 96% similarity in the protein primary structure.

Next, the VgR cDNA was cloned through RACE-PCR using an adaptor-ligated double stranded cDNA constructed from the mRNA purified from previtellogenic ovaries of the American cockroach, *P. americana*. The complete cDNA for *P.americana* VgR was obtained from two overlapping RACE-PCR fragments of 5'- and 3'-end and was of 5411 residues which encoded a deduced amino acid sequence of 1709 residues long including 19 rsidues for a putative signal peptide. The deduced amino acid sequence analysis shows that *P. americana* VgR is a novel LDLR family member with only 6 repeats in the second ligand binding domain. This novel VgR

shares a significant homology with chicken, frog and rainbow trout VgRs, and particularly with mosquito VgR and a YPR of *Drosophila* which are also members of the LDLR gene superfamily. The cytoplasmic tail of *P. americana* VgR contains a leucine-isoleucine internalization signal, a motif that seems to be common in insect VgRs/YPRs (di-leucine in *Drosophila* YPR), unlike the tight-turn-tyrisine motif (NPXY) of other members of LDLR gene family. The phylogenetic (neighbour-joining) analysis reveals that the mosquito VgR and YPR of *Drosophila* are more closely related than with the cockroach VgR.

During cloning of *P. americana* VgR, we also found a partial cDNA clone encoding C-terminus of receptor tyrosine kinase (RTK) from previtellogenic ovaries of the American cockroach, *P. americana*. The remaining 5`-end portion of RTK was then cloned through 5` RACE-PCR. The complete sequencing of two overlapping RACE-PCR fragments (both 3-` and 5-end)` of *P. americana* RTK thus specified an open reading frame of 4128 residues which encoded a deduced amino acid sequenece of 1294 residues long including 22 residues for the putative signal peptide. The entire deduced amino acid sequence was aligned well with other members of the RTK family of receptors. We are, however, not clear about the role of RTK in the previtellogenic ovary at the moment. One possibility is that RTK, like other family members, may be involved in cell proliferation, especially follicle cells in the developing ovary.

The present study reports on cloning and characterization of cockroach Vgs and VgR which has led not only to understanding the structures, expressions/regulations and processings (in Vgs) of these proteins, but also has provided a wealth of new insights concerning the evolutionary relationships among these proteins. In addition, the present breakthrough at molecular level will help to elucidate the Vg/VgR binding interactions, and to explore the potential manipulation of insect receptor-endocytotic machinery for incorporation of a foreign molecule, the Vg, into the insect oocyte.

ACKNOWLEDGEMENTS

I would like to express my sincere and warm gratitude to my professor Dr. Makio Takeda for his valuable advice, repeated reading of the manuscript, keen understanding, strong support and precious time during my study under his kind personality that allowed me to work in a comfortable environment. His encouragement and support in all matters other than the academic life are beyond the words of thanks and appreciation. His kind support and care made my stay in Japan a pleasant experience.

My thanks are due to professor Dr. C. Naito, Dr. K. Takasu and late Ms Akie Yasuhara for their kindness and support in many ways during the course of this study.

My cordial thanks are due to professor Dr. Kugao Oishi for providing me a chance to learn about different techniques related to my project in his laboratory. His continuouse support and encouragement other than the research works shall also be ever remembered. My thanks are also due to Drs. M. Hatakeyama and Jae Min Lee for their guidance and moral support. I am also indebted to all fellow students of the Laboratory of Developmental Biology, Kobe University for their kindness and cooperation.

I am also thankfull to all fellow students of the Laboratory of Insect Science, Kobe University for their support and cooperation.

I am extremely thankfull to Drs. Y. Fukami and Kenichi Sato of the Laboratory of Molecular Bioloy, Kobe University, for amino acid analysis of polypeptides.

I am thankfull to the Ministry of Education, Science, Sports and Culture of Japan (MONBUSHO) for supporting my stay in Japan.

Finally, I am thankfull to the officials of the University of Agriculture, Faisalabad, the University Grants Commission, Islamabad and the Ministery of Education of Pakistan to allow me to avail the opportunity to study in Japan.