This is to certify that the

dissertation entitled

Molecular Characterization of
Mosquito Yolk Proteins

presented by

Jeng-Shong Chen

has been accepted towards fulfillment
of the requirements for

Doctoral degree in Philosophy

Alexander S. Raikhel

Dr. Alexander Raikhel
Major professor

Date June 15, 1995

MOLECULAR CHARACTERIZATION OF MOSQUITO YOLK PROTEINS

By

Jeng-Shong Chen

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Graduate Program in Genetics and Program of Cell Molecular Biology

1995

ABSTRACT

MOLECULAR CHARACTERIZATION OF MOSQUITO YOLK PROTEINS

BY

JENG-SHONG CHEN

In oviparous animals, accumulation of yolk proteins or vitellogenesis is critical for egg development. In mosquitoes, this process is initiated by blood feeding and as a consequence is linked to transmission of numerous mosquito-borne diseases of humans and domestic animals. Therefore, elucidation of biochemical and molecular aspects of mosquito vitellogenesis is important in the search of novel mosquito control strategies. The cDNA of mosquito vitellogenin (Vg), the major yolk protein precursor, was cloned and characterized. Several features of this important molecule have been identified including the location of the Vg subunits in the prepro-Vg, the sequence of the subunit cleavage site, the putative receptor recognition domains and the presence of large polyserine domains previously known only in the Vgs of vertebrates. The cleavage of pro-Vg into two subunits is an essential, rate-limiting step in its biosynthesis in the fat body, an insect analog of the vertebrate liver. There is a conserved sequence, RXRR, for convertases identified in the cleavage site of mosquito Vg. Based on this feature and the sequences of other known convertases, the cDNA encoding a fat body-specific vitellogenic convertase (VC) was cloned. Co-expression of the VC and Vg cDNAs using coupled *in vitro* transcription and translation resulted in

correct processing of the Vg precursor. Thus, a novel Vg-processing convertase has been identified in insects. The cDNA of another yolk protein which is denoted as vitellogenic thiol protease (VTP) has been cloned and analyzed. Similar to Vg, this protein is synthesized and secreted by the mosquito fat body and deposited in developing oocytes. VTP has high sequence similarity to cathepsins B. Biochemical studies revealed that this protein is a proenzyme activated only at the onset of embryonic development in mosquito eggs. Furthermore, evidence suggests that VTP directly participates in embryonic degradation of vitellin, the crystalline form of Vg.

ACKNOWLEDGMENTS

I would like to thank Dr. Alexander Raikhel, my advisor, and my committee, Dr. Susan Conrad, Dr. Donald Jump and Dr. Suzanne Thiem, for their advice, help and guidance throughout the course of this study. I would also like to thank my colleagues, Dr. Avraham Biran, Steven Cassar, Chi-Ju Chen, Dr. Wen-Long Cho, Dr. Kirk Deitsch, Neal Dittmer, Alan Hays, Dr. Marianna Kapitskaya, Dr. Vladimir Kokoza, Dr. Tom Sappington and Dr. Ekaterina Snigirevskaya, who have been very helpful throughout my work in the lab and while completing my dissertation. Dr. Kirk Deitsch deserves particular thanks for his enthusiasm in helping me to master the English language.

Special thanks to Dr. Kirk Deitsch and Neal Dittmer for their friendship and patience. They helped me to learn, understand and explore American culture. They made my stay in Michigan an unforgettable experience.

This dissertation is dedicated to my parents, Chiang-Sung Chen and Yueh-Hwa Wu as well as my wife, Dr. Yueh-Jiang Hwang and my daughter, Emilie Chen. My achievement is owed to their continuous support, encouragement and inspiration.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 5

ABBREVIATIONS

bp ..... base pair

cDNA ..... complementary DNA

DNA ..... deoxyribonucleic acid

kDa ..... kilodalton

kb ..... kilobase pair

Mr ..... molecular weight

mRNA ..... messenger RNA

PAGE ..... polyacrylamide gel electrophoresis

PBM ..... post blood meal

PCR ..... polymerase chain reaction

RNA ..... ribonucleic acid

SDS ..... sodium dodecyl sulfate

UGAL ..... University of Georgia at Athens laboratory strain

VC ..... vitellogenin convertase

VCP ..... vitellogenic carboxypeptidase

VTP ..... vitellogenic thiol protease

Vg ..... vitellogenin

Vn ..... vitellin

YP ..... yolk protein

CHAPTER 1

INTRODUCTION

I. Overview

Mosquitoes are the most important vector insects that transmit numerous diseases of humans, livestock and poultry. Many diseases carried by mosquitoes are fatal. Mosquitoes of the genus *Aedes* transmit several viral diseases including Dengue fever, yellow fever and Japanese encephalitis. Mosquitoes in this genus are also capable of carrying parasites that cause filariasis. *Anopheles* mosquitoes are most known for their transmission of malaria and filariasis parasites (Sturchler, 1989). These mosquito-borne diseases have a tremendous impact and cause great losses to society.

Repressing the mosquito population has proven to be the most effective strategy to prevent and control the outbreak of mosquito-borne diseases, especially those caused by viral pathogens. Due to their high efficiency, low cost and ease of application, insecticides remain the method of choice among available control methods. In the mid-1950's, the World Health Organization organized world-wide malaria mosquito control projects in an attempt to slow down malaria transmission. At that time, large scale application of DDT was the main strategy used. Very promising results were achieved in the beginning of these projects, however, mosquitoes responded by developing high resistance to DDT in a relatively short period of time. Mass use of other insecticides also resulted in development of high resistance in mosquitoes. Many populations and species of mosquitoes have been reported to have developed multiple

insecticide resistance against carbamates, organophosphates and pyrethroids or even cross resistance to different classes of insecticides. In addition, the traditional chemical methods of mosquito control also face concerns about side effects on non-target organisms and the environment.

As a remedy for malaria, chloroquine gave excellent results in killing the protozoan parasites. Yet, high selection pressures on malaria parasites also caused development of resistance to chloroquine and other drugs. Recently, mainly due to the development of insecticide resistance of mosquitoes and drug resistance of pathogens, several mosquito-borne diseases have become serious again in various parts of the world (McClelland, 1992). Although a great deal of effort has been expended toward the development of vaccines against malaria and other mosquito-borne diseases, thus far no significant breakthrough has been achieved. In addition, the full benefits of vaccination of high risk populations may be difficult to achieve due to problems of distribution and high costs (Anderson and May, 1991).

Control of modern mosquito-borne diseases requires the development of new methods against vector insects and pathogens. Rational and efficient control methods that depend upon data from basic research in mosquito biology provide hope to overcome insecticide resistance in mosquito populations. Since blood meals are required for a successful reproductive cycle of mosquitoes, studies of mosquito reproduction can provide useful data for the development of new control strategies. In particular, the rapidly developing molecular tools offer a promising future for the control of mosquito-borne diseases.

II. Reproduction Biology of Female Mosquitoes

A reproductive cycle of anautogenous mosquitoes, including those of the medically important species *Aedes aegypti* and *Anopheles gambiae*, can be

divided by a blood meal into pre-vitellogenesis and vitellogenesis. The period of physiological change from initiation to termination of massive production and accumulation of yolk proteins is denoted as vitellogenesis.

A female mosquito does not take blood and can not produce eggs during pre-vitellogenesis, the first three days after eclosion. A transition period that remodels the synthetic tissues is required for a female to become competent for extremely intensive production of yolk proteins during vitellogenesis. This period is called the preparation stage of pre-vitellogenesis. Subsequently, the female enters a state-of-arrest and is ready to take blood. Once a blood meal is taken, vitellogenesis is initiated (Raikhel, 1992).

The fat body is the tissue responsible for yolk protein synthesis in a vitellogenic female mosquito. The fat body is a tissue made up of metabolic cells and is analogous to the vertebrate liver. During the preparation stage of pre-vitellogenesis, the ribosomal RNA and ploidy levels of fat bodies increase dramatically. These cellular events are under the control of juvenile hormone (Dittman *et al.*, 1989; Raikhel and Lea, 1990). The titer of this hormone reaches a peak two days after eclosion and then slowly declines (Shapiro *et al.*, 1986). At the end of the preparation stage, the cellular machinery is ready for intensive yolk protein synthesis.

The steroid hormone 20-hydroxyecdysone (20E) is important for regulation of the mosquito yolk protein genes. Its titer rises after a blood meal, reaching a peak at 18-24 h post blood meal (PBM) and then falling dramatically (Hagedorn, 1985). Although the kinetics of the titer of 20E match the kinetic patterns of yolk protein gene expression, it has been demonstrated that 20E may regulate the mosquito yolk protein genes in an indirect way. Cyclohexamide studies revealed that protein synthesis is required for 20E to regulate the yolk

protein genes (Deitsch *et al.*, 1995). The complicated cascade of hormonal regulation of mosquito yolk protein gene expression is not totally understood.

Termination of vitellogenesis begins when the titer of 20E has declined to background levels. At this time, the vitellogenic genes are turned off and biosynthetic machinery is degraded (Raikhel, 1986a,b). This stage lasts from 30 to 48 h PBM. A batch of 150 eggs is developed uniformly and laid at the same time. Following egg laying, the female mosquito returns to the physiological stage of pre-vitellogenesis and is ready to take a blood meal for the next reproductive cycle.

III. Expression of mosquito yolk protein genes

Vitellogenin (Vg) is the major yolk protein synthesized in the mosquito fat body during vitellogenesis. There are four genes that code for this protein (Hamblin *et al.*, 1987). Expression of the mosquito Vg genes is tightly regulated by a blood meal. It is initiated after the female mosquito takes blood, and then peaks around 24-27 h PBM. The expression of the Vg genes drops dramatically to background levels around 36 h PBM. The mosquito Vg genes are expressed in a sex-, tissue-, and stage-specific manner. The size of the mosquito Vg mRNA is 6.5 kb and the cDNA of mosquito Vg has been cloned and sequenced (Chen *et al.* 1994). Mosquito pro-Vg, 250 kDa in size, is post-translationally modified in the rough endoplasmic reticulum and the Golgi complex. After modification, mosquito pro-Vg becomes a sulfated phosphoglycolipoprotein (Dhadialla and Raikhel, 1990; Raikhel, 1992). It is further cleaved into small and large subunits of 66 and 200 kDa before secretion (Raikhel and Bose, 1988).

Insect Vgs have a motif containing paired basic residues located at the C-termini of the small subunits that is known to be a convertase recognition site (Trewitt *et al*, 1992; Chen *et al.*, 1994; Yano *et al.*, 1994; Kageyama *et al*, 1994).

Based on the conserved amino acid sequence at the cleavage sites of several insect Vgs, the enzymes responsible for subunit cleavage of these proteins were suggested to be convertases (Chen and Raikhel, unpublished). Both subunits bind each other non-covalently and circulate in the hemolymph. Vg is taken up by an oocyte Vg receptor that belongs to the LDL receptor family (Sappington *et al.*, 1995) and deposited in the yolk body of a growing oocyte.

A yolk protein other than Vg named vitellogenic carboxypeptidase (VCP) has been characterized both biochemically and molecularly (Hays and Raikhel, 1990; Cho *et al.*, 1991; Deitsch and Raikhel, 1993). The mRNA transcribed from the VCP gene is 1.5 kb long and the translated product is 53 kDa. Sequence analysis revealed that it resembles carboxypeptidases (Cho *et al.* 1991). The expression of the VCP gene is similar to that of Vg. It is produced exclusively in the fat body of the vitellogenic female. Following secretion, it is deposited in the growing oocyte. VCP is activated during embryogenesis, however, its role during embryogenesis is not yet clear.

Recently, another mosquito yolk protein precursor has been identified and molecularly characterized. This protein was denoted as vitellogenic thiol protease (VTP) because it belongs to the eukaryotic thiol protease superfamily. Its sequence shows high sequence similarity to cathepsins B (Chen *et al.*, 1995). The size of the VTP mRNA is 1.1 kb and the secreted VTP protein is 44 kDa. The expression of this gene is also similar to Vg as well as VCP. Its synthesis is tightly regulated by a blood meal, and it is exclusively produced and secreted by the vitellogenic female fat body. This pro-enzyme is circulated in the hemolymph and then taken up by oocytes. VTP is activated during embryogenesis where it plays an important role in degrading vitellin (Vn) which provides the building blocks for embryo development.

IV. Significance in studying mosquito reproduction

The study of mosquito reproductive systems is very important because the dispersal, transmission and propagation of pathogens causing mosquito-borne diseases depend upon blood meals and subsequent successful reproduction of the female. This is a major reason why a great deal of effort has been applied to elucidating the molecular basis of mosquito vitellogenesis. In the future, these studies may provide the means to decrease the impact of mosquito-borne diseases on human populations and loss in production of livestock and poultry by interrupting or manipulating mosquito genes that are involved in reproduction.

CHAPTER 2*

ANALYSIS OF MOSQUITO VITELLOGENIN CDNA: SIMILARITY WITH VERTEBRATE PHOSVITINS AND ARTHROPOD SERUM PROTEINS

* Originally published in Journal of Molecular Biology. Reference: Chen, J. S., Cho, W. L. and Raikhel, A. S. 1994. J. Mol. Biol. **237,** 641-647.

ABSTRACT

The cDNA coding for vitellogenin of the mosquito, *Aedes aegypti,* was cloned and sequenced. An immunological analysis of expressed deletions from the 5'-end of the vitellogenin cDNA clones using vitellogenin subunit-specific antibodies showed that the small vitellogenin subunit is located at the N-terminus and the large one at the carboxyl-portion of the pre-pro-vitellogenin. The position of the cleavage between the vitellogenin subunits in the pre-pro-vitellogenin was identified by locating the N-terminus of the large subunit. The cleavage site has a consensus RXRR for the subtilisin processing endoprotease. Mosquito vitellogenin is highly hydrophilic with 17 putative N-linked glycosylation sites and 13 potential tyrosine sulfation sites. In contrast to known invertebrate vitellogenins, mosquito vitellogenin contains three polyserine domains that are similar to those of phosvitins in vertebrate vitellogenins. These polyserine domains, originally presumed to be vertebrate-specific, have several phosphorylation consensus sites in their sequences. Unlike other known vitellogenins, mosquito vitellogenin is rich in aromatic amino acids, tyrosine and phenylalanine, and in this respect is similar to insect serum proteins, arylphorins. This similarity suggests that mosquito vitellogenin may supply aromatic amino acids to the cuticle of rapidly developing embryos.

INTRODUCTION

Vitellogenin (Vg), a protein unique to oviparous animals, is the major yolk protein precursor. It is produced by extraovarian tissues, such as the vertebrate liver or the insect fat body, secreted into the blood or the hemolymph and then specifically accumulated by developing oocytes. Vitellogenin plays an important role in the development of the embryo by providing amino acids, carbohydrates,

lipids and ions (Kunkel & Nordin, 1985, Byrne *et al.*, 1989; Raikhel & Dhadialla, 1992).

With the exception of higher Diptera, Vgs of oviparous animals are phospholipoglycoproteins of high molecular weight which presumably had a common origin and a limited structural conservation throughout their evolution (Wahli, 1988; Byrne *et al.*, 1989). The deduced amino acid sequences have provided invaluable information to the study of the structure-functional properties of Vgs as well as their evolutionary relations. However, genes and cDNAs encoding Vgs have been sequenced only from a few animals. Among them are nematode, *Caenorhabditis elegans* (Spieth *et al.*, 1985), chicken, *Gallus gallus* (van het Schip *et al.*, 1987), frog, *Xenopus laevis* (Gerber-Huber *et al.*, 1987), lamprey, *Ichthyomyzon unicuspis* (Sharrock *et al.*, 1992) and an insect, boll weevil *Anthonomus grandis* (Trewitt *et al.*, 1992). Considering the diversity of oviparous animals, the conclusions based on comparison of such a small number of Vgs have obvious limitations. In addition, most Vg sequences are available only for animals used as experimental models and do not reflect the wide diversity of nutritional and other adaptations existing among oviparous animals. This is particularly true for insects, which comprise the largest and the most diverse group of oviparous animals. It is not clear, for example, if Vgs are used only as a general source of nutrients or whether in addition, these molecules are utilized as a vehicle for passing the specific nutritional values of the parent's food to developing embryos. An analysis of Vgs from insects with different feeding adaptations will permit us to address this question.

The evolutionary success of mosquitoes is attributed to their feeding on vertebrate blood. The latter is also responsible for mosquitoes becoming the most important vectors of numerous pathogens of human and other vertebrates. In mosquitoes, *Aedes aegypti*, vitellogenesis is tightly coupled with blood feeding

and is initiated only after the mosquito female ingests blood (Raikhel, 1992). Previously, we characterized the biosynthesis and the subunit composition of mosquito Vg (Dhadialla & Raikhel, 1990). Mosquito Vg consists of two subunits with M_r = 200,000 and 66,000. Both the Vg subunits are first synthesized as a single precursor. The pre-pro-Vg is extensively co- and post-translationally modified and cleaved before Vg is secreted from the fat body (Dhadialla & Raikhel, 1990). In the mosquito ovary, Vg binds with high affinity (K_d = 2.8 x 10^{-8}M) to a single class of Vg receptors with M_r = 205,000 (Dhadialla *et al.*, 1992).

MATERIALS AND METHODS

Reagents - Analytical grade of chemicals and reagents for immunoassay were purchased from Sigma Chemical Co. (St. Louis, MO) unless stated otherwise. Restriction enzymes, DNA and RNA markers were supplied by Gibco BRL. Sequenase Ver. 2.0 was purchased from United State Biochemical. Both ^{32}P-dATP (3,000 Ci/mmlole) and ^{35}S-dATP (1,000-1,500 Ci/mmole) were obtained from New England Nuclear, DuPont.

Animals - Mosquitoes were reared at 27°C and 85% relative humidity. Larvae were fed on a standard diet (Lea, 1964). Three to five days after eclosion adult females were fed on white rats to initiate vitellogenesis. Mosquitoes were dissected in *Aedes* physiological saline (Hagedorn *et al.*, 1977) at room temperature.

Cloning and Characterization of Mosquito Vg cDNA - The entire mosquito Vg cDNA was obtained by screening a vitellogenic female mosquito fat body cDNA

lambda Zap-II library (Stratagene) with a 1.8 kb *Eco* RI-*Eco* RI fragment of the mosquito Vg A1 gene (Hamblin *et al.*, 1987) (a gift from Dr. H. H. Hagedorn) and the Rapid Amplification of cDNA Ends (RACE) method described by Frohman *et al.* (1988). For the missing 5' end, a 32-mer antisense primer, TTCGCTGCAGACCAGTCCACTGATCGTCCAAT, and a 17-mer oligo(dT) and an *Xba* I site at the 5'-end (dT-Primer) were used for PCR. A 358-bp fragment was specifically amplified after 25 cycles. The 3'-terminus of the Vg cDNA was obtained by a similar method in which a 126 bp fragment was amplified using dT-Primer and a sense 28-mer, CGAATTCTGTCTACAAGGCAT ACTAGGA. All the cDNAs were sequenced from both strands by the dideoxy chain termination method (Sanger *et al.,* 1977). The amino acid sequence was deduced from the Vg cDNA sequence. For microsequencing, vitellogenin was purified from vitellogenic mosquito fat bodies following the methods detailed in Koller *et al.* (1989). Both the large and small Vg subunits were gel-purified and electroblotted onto PVDF membrane.

Northern and Immuno-dot Blot Analyses - Method for total RNA preparation from mosquito fat bodies was described by Bose and Raikhel (1988). Twenty micrograms of total RNA was separated by a 1.2% formaldehyde/agarose gel, transferred to nitrocellulose membrane, and hybridized with ^{32}P-labeled Vg genomic and cDNA probes. For immuno-dot blot analysis, the three cultures of XL1-Blue strain of *E. coli* containing pBluescript, pVG1 and pVG1a, respectively, were grown in LB medium with 10 mM IPTG. The overnight cultures were lysed by 1.5% SDS. The lysates were loaded onto nitrocellulose membrane. Purified mosquito Vg was used as a positive control, while cells transformed by pBluescript without inserts served as a negative control. Polyclonal antibodies specific to the large subunit (Vg-L) (Hays & Raikhel, 1990) and monoclonal

antibodies to the small subunit (Vg-S) (Raikhel *et al.,* 1986) were used to detect the translated products.

DNA Analysis and Alignment - The putative signal peptide of the deduced amino acid sequence was determined according to Kyte and Doolittle (1982). The deduced amino acid sequence was analyzed by FASTA, MOTIFS and GAP programs (University of Wisconsin, Genetics Computer Group software).

RESULTS AND DISCUSSION

The deduced amino acid sequence of the mosquito Vg is presented in Figure 1. This sequence was derived from cDNA clones identified by screening the vitellogenic mosquito fat body cDNA library with a mosquito Vg-A1 genomic probe (Hamblin *et al.*, 1987) and clones generated using the RACE method (Frohman *et al.*, 1988). The complete cDNA sequence has a restriction map similar to that of the coding region of the Vg gene A1 reported by Hamblin *et al.* (1987) and a polyadenylation signal located 22 bp upstream of the poly(A) tail.

In the sequences of these cDNA clones, there were several base substitutions without a change of amino acids (data has been reported to GenBank™). We concluded this finding was not due to PCR or sequencing errors because base substitutions have been identified in clones from both the library screening and PCR. In addition, base substitutions occur only at the third base of a codon. The detected heterogeneity likely reflects allelic variations for the Vg gene A1 in the mosquito population. Similarly, several alleles have been found in the vitellogenin carboxypeptidase gene (VCP) and lysosomal aspartic protease gene (mLAP) in the same population of *A. aegypti* (Deitsch & Raikhel, 1993; Dittmer, Cho & Raikhel, unpublished).

Figure 1. Deduced amino acid sequence of mosquito vitellogenin. The putative signal peptide is in bold print and underlined with a broken line. The amino acid sequences that match the direct N-terminus peptide sequencing of both subunits are underlined. The cleavage site is indicated by an empty arrowhead. Potential N-linked glycosylation sites (N) are marked by a cross sign, while potentially sulfated tyrosines (Y) are labeled by a solid square. The double underlined portions indicate polyserine domains, whereas potential phosphorylated serines (S) are marked with an asterisk. The consensus for the protease of the subtilisin family is labeled by a broken double underline.


MLAKLLLLALAGLTAAYQYENSFKGYNPGYKGYDAGYKGYGYDAGYKYNNQGYSYKNGFEYGYQNAYQAAFYKHRPNVTEFEFSSWMPNYEYVYNVTSKT 100
MTALAELDDQWTGVFTRAYLVIRPKSRDYVVAYVKQPEYAVFNERLPHGYATKFYHDMFKFQPMPMSSKPFGIRYHKGAIKGLYVEKTIPNNEVNILKAW 200
ISQLQVDTRGANLMHSSKPIHPSKNEWNGHYKVMEPLVTGECETHYDVNLIPAYMIQAHKQWVPQGQLRGEDDQFIQVTKTQNFDRCDQRMGYHFGFTGY 300
SDFRPNTNQMGNVASKSLVSYMYLTGNWYNFTIQSSSMINKVAIAPSLVNKEPALVYAQVNMTLNDVHPYDKVPMGPAEDLKVFVDLVYSYNMPSDKKNY 400
VRPGNETSSSSSSSSSSSSSSSSRSSSSSSESVENPKISPVEQYKPLLDKVEKRGNRYRRDLNAIKEKKYYEAYKMDQYRLHRLNDTSSDSSSSDSSSSSS 500
SESKEHRNGTSSYSSSSSSSSSSSSSSESSSYSSSSSSSSESYSISSEEYYYQPTPANFSYAPEAPFLPFFTGYKGYNIFYARNVDAIRSVGKLVEEIASD 600
LENPSNLPKANTMSKFNILTRAIRAMGYEDIYELAQKYFVSQKERQVAQFSDKKFSKRVDAWVTLRDAVAEAGTPSAFKLIFDFIKEKKLRGYEAATVIA 700
SLAQSIRYPTEHLLHEFFLLVTSDVVLHQEYLNATALFAYSNFVNQAHVSNRSAYNYYPVFSFGRLADADYKIIEHKIVPWFAHQLREAVNEGDSVKIQV 800
YIRSLGNLGHPQILSVFEPYLEGTIQITDFQRLAIMVALDNLVIYYPSLARSVLYRAYQNTADVHEVRCAAVHLLMRTDPPADMLQRMAEFTHHDPSLYV 900
RAAVKSAIETAALADDYDEDSKLALNAKAAINFLNPEDVSIQYSFNHIRDYALENLELSYRLHYGEIASNDHRYPSGLFYHLRQNFGGFKKYTSFYYLVS 1000
SMEAFFDIFKKQYNTKYFADYYKSADYSTNYYNFDKYSKYYKQYYYSKDSEYYQKFYGQKKDYYNDKEPFKFTAPRIAKLLNIDAEEAEQLEGQLLFKLF 1100
NGYFFTAFDNQTIENLPHKMRHLFENLEDGYAFDVTKFYQQQDVVLAWPLATGFPFIYTLKAPTVFKFEVDASAKTHPQVYKMPAGHPETENDDFFYMPQ 1200
SINGSVDVNLLYHRMVDAKVGFVTPFDHQRYIAGYQKKLHGYLPFNVELGLDFVKDEYEFEFKFLEPKDDHLLFHMSSWPYTGYKDITDMRPIAENPNAK 1300
IVHDDNQSTKTMEHTFGQDMTGVALRFHAKYDFDLINFQQFWSLIQKNDFVSAVNYPFAYQPYEYHQFNLFYDSQRTHAKSFKFYAYQKFGAPSFEETGP 1400
KHPANRHSYSGNYYESNYAQPFVYSPGSQRRYEQFFRNAASGIRNSFVRYYDFGFEFYAPQYKSEFTFTTAFADSPVDKTSRQLYYFYASPMFPSQSYFK 1500
DIPFSGKQFQFCATATSEFPRVPYLKFSDFDKYYGDASQYFDFLYGESCQGGAHIAVKGKQKQTGKYREYLRFSDVAKACKEQMANGYYQFEECQQAIDQ 1600
AYYYDFYDYAIEYKDVGSVAKNLTNKFYNYFQYAFYPYFESNFFYHGKSNYIKAEFEFAPYGDYYNASFFGPSYAFQVQNYPVFNDYSTYFPYFFKYTFF 1700
PRYQPYYMHRLPSHKPRNRPYYELSNYEQFAIFDRKPQYPSCSFSNDYFYTFDNKKYFYDMGECWHAVMYTVKPDYDFYAQQSHFYNSDFEYKYKNGFEE 1800
YEQFAALARRGSDNQLYFKFLFGDNYIEVFPNNGGVPFVKYNGRPYDISKSNIAHFEYKEGYPSFPFFYAFAYPNKDLEVSFFGGKLKFATDGYRARFFS 1900
DYSFYNNFVGLCGTNNGEYFDEFVTPDQCYMRKPEFFAASYAITGQNCTGPAKAFNYAYQQKAKQECVKREVYYGDIIYNQEYYHPRYRYYNHNVEESSS 2000
SSSSSSSDSSSSSSSSSEFSSLSRSGSSSSSSSSSEEQKEFHPHKQEHSMKECPVQHQHQFFEQGDRICFSLRPLPVCHSKCAATEKISKYFDVHCFEKDST 2100
QAKKYKSEIGRGYTPDFKSFAPHKTYKFNYPKSCVYKAY 2139

To verify that the obtained cDNA clones indeed encode mosquito Vg, the Vg genomic fragment and the entire pVG1 cDNA clone were used as probes to hybridize with vitellogenic mosquito RNA . Both probes recognized a 6.5 kb mRNA (Figure 2). In addition, the 0.5 kb *Eco* RV-*Pst* I fragment from the 5'-end of pVG1 and the 1 kb *Xho* II-*Xho* II fragment of 1.0 kb from the 3'-end of pVG1 were also hybridized to the same mRNA (data not shown). These results suggest that the pVG1 clone represents a non-recombined continuous cDNA encoding mosquito Vg. Immunological analysis using antibodies specific to the Vg subunits was then utilized in order to verify the obtained putative Vg cDNA clones and to locate the position of each Vg subunit in the pre-pro-Vg molecule. The cDNA clone pVG1 and its 5'-end serial deletion derivatives, covering more than half of its length, were expressed in *E. coli*, and in-frame clones were selected by antibodies specific to both Vg subunits. Further analysis was carried out using subunit-specific antibodies. The translated product of an in-frame intact cDNA clone, pVG1, was recognized by antibodies specific to either the small or the large Vg subunit. The in-frame deletions lacking 1/3 or more of the 5'-sequence, such as pVG1a, were recognized only by the antibodies to the large Vg subunit (Figure 3). These experiments clearly demonstrated that the small subunit is located at the N-terminus, and the large subunit occupies the carboxyl-portion of pre-pro-Vg.

The mosquito Vg cDNA has an open reading frame that encodes a 2,139 amino acid peptide with a predicted M_r = 249,387. According to the hydropathy analysis (Kyte & Doolittle, 1982) and the N-terminal sequence of the Vg small subunit, YQYENSFK, a stretch of 16 highly hydrophobic residues is a putative signal peptide (Figure 1). The N-terminus of the large Vg subunit, DLNAIKEKK, starts at the position D_{460}. This places a putative cleavage site of the pro-Vg into two subunits at positions between R_{459} and D_{460} (Figure 1). This cleavage

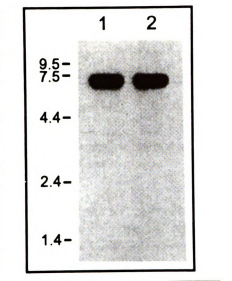


Figure 2. Verification of mosquito vitellogenin cDNA clone by Northern blot analysis. Fifteen micrograms of total RNA from fat bodies of mosquito females, 24 h after a blood meal were used for each lane. pVG1 (lane 1) and the Vg A1 genomic fragment (Hamblin *et al.*, 1987) (lane 2) were used as probes.

site between Vg subunits is positioned between a turn and an alpha helix as predicted by the methods of Chou & Fasman (1978). It is located in the second hydrophilic region and is flanked by two polyserine domains (Figure 4). However, amino acids in the immediate proximity of the cleavage site form a less hydrophilic stretch, L_{461}-I_{464}, which leads the cleavage site inward. The anterior side of the cleavage site contains a consensus cleavage sequence, RXRR (R_{456}-R_{459}) for dibasic processing endoproteases of the subtilisin family (Barr, 1991). The most striking feature common to Vgs of the mosquito and boll weevil (Heilmann *et al.*, 1993) is the conservation of the subunit cleavage site.

Interestingly, no such consensus is found at the cleavage site of vertebrate Vgs. In spite of this difference, the cleavage site of lamprey Vg, F_{1305}-K_{1306} is also located in a region of highly polar amino acids (Sharrock *et*

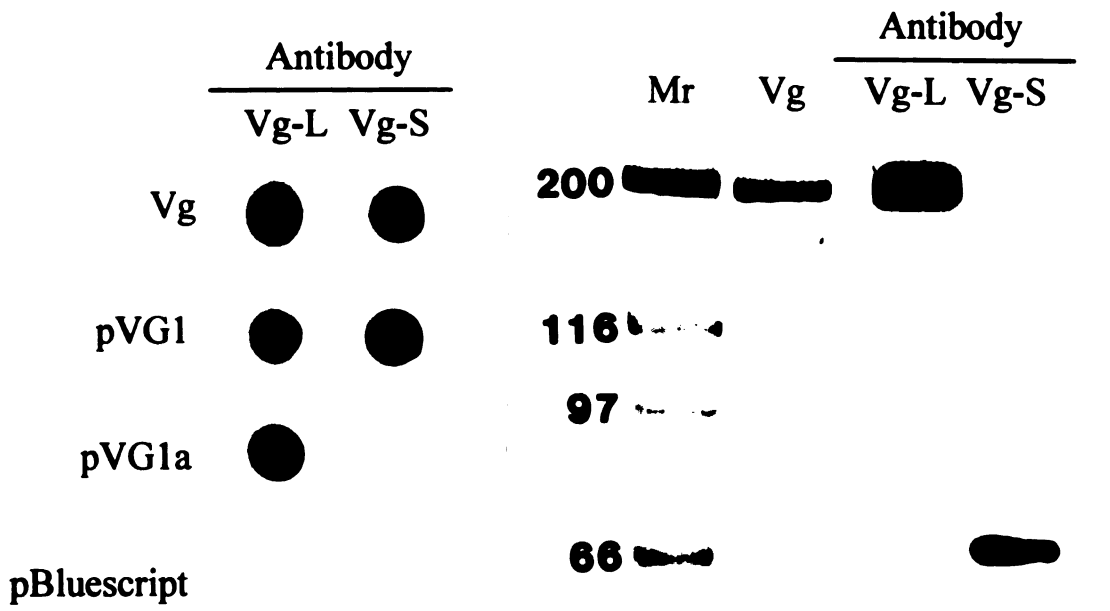


Figure 3. Localization of Vg subunits in pre-pro-Vg sequence by immunoassay. (A) The lysates of XL1-Blue strain of *E. coli* containing pBluescript, pVG1 and pVG1a, respectively, were loaded onto nitrocellulose membrane. Purified mosquito Vg and cells transformed by pBluescript served as controls. Polyclonal antibodies specific to the large subunit (Vg-L) and monoclonal antibodies to the small subunit (Vg-S) were used to detect the translated products. (B) The specificity of antibodies used for the assay in (A). Purified mosquito Vg was electrophoresed on a 6% polyacrylamide gel. A Commassie Blue stained gel (Vg) showed the two Vg subunits (200 and 66 kDa). Western blots were performed using polyclonal antibodies specific to the large subunit (Vg-L) and monoclonal antibodies specific to the small subunit (Vg-S).

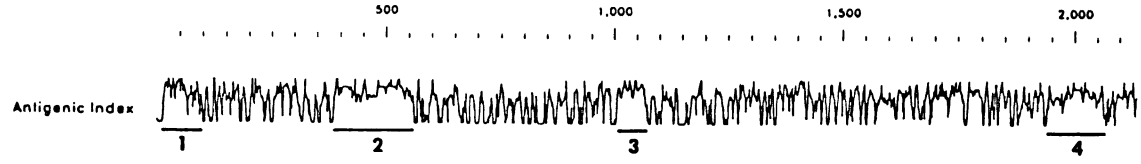


Figure 4. Antigenic index plot of the deduced amino acid sequence of mosquito vitellogenin. This plot was determined by the GCG program. The scale indicates the number of residues. The underlined and numbered regions are the major hydrophilic domains analyzed. The first and the third domains are tyrosine-rich, while the second and fourth are serine-rich domains.

al., 1992). The difference may be due to the fact that vertebrate Vg is not cleaved during its production in the liver (it is cleaved in the oocyte after uptake), while insect Vg is cleaved in the rough endoplasmic reticulum or *cis*-Golgi of the fat body (Dhadialla & Raikhel, 1990). A kallikrein-like enzyme may be involved in the cleavage of vertebrate Vgs (van het Schip *et al.*, 1987). Cathepsin D is also implicated in the cleavage at the carboxyl-terminal end of phosvitin (Byrne *et al.*, 1989).

The deduced amino acid composition of mosquito pre-pro-Vg were compared with Vgs from six other invertebrate and vertebrate species. Mosquito Vg has a higher serine (Ser) content (10.1%) than other invertebrate Vgs, one similar to that of vertebrate Vgs. This is mainly due to the polyserine domains found in mosquito Vg. Byrne *et al.* (1989) pointed out that high Glx (Glu + Gln) content is a feature of most Vgs. However, the Glx content of mosquito Vg is close to the average of eukaryotic proteins reported by Doolittle (1986).

A high content of aromatic amino acids, Tyr and Phe, is another unusual feature of mosquito Vg which sets it apart from the Vgs of other invertebrate and vertebrate animals with known amino acid sequences (Figure 1). In this respect, mosquito Vg is similar to insect larval storage proteins which are hexamers and

Table 1. Similarity between mosquito vitellogenin and arthropod serum proteins

Peptide	Subunit	Species	No. Amino Acids in Overlapped Sequence	Similarity (%)
Arylphorin[a]	a	*Manduca sexta*	716	58.0
Arylphorin[a]	b	*Manduca sexta*	680	61.2
Arylphorin[b]	A4	*Calliphora vicina*	482	60.0
Larval Storage Protein[c]	Sp1	*Bombyx mori*	121	60.3
Larval Storage Protein[d]	Sp2	*Bombyx mori*	57	59.6
AJSP-1[e]		*Trichoplusia ni*	147	64.0
Hemocyanin[f]	a	*Eurypelma californica*	62	79.0
Hemocyanin[f]	e	*Eurypelma californica*	116	62.9

a. Willott *et al.* (1989).; b. Naumann and Scheller (1991).; c. Sakurai *et al.* (1988).; d. Fujii *et al.* (1989).; e. Acidic juvenile hormone suppressible protein-1, a hemocyanin-related protein (Jones *et al.*, 1990).; f. Voit and Feldmaier-Fuchs (1990).

include two types, the aromatic amino acid-rich type (arylphorins) and the Met-rich type (Telfer & Kunkel, 1991). Larval storage hexameric proteins serve as an important source of amino acids for molting and metamorphosing insects (Scheller *et al.*, 1990; Telfer & Kunkel, 1991). The aromatic amino acids in arylphorins are thought to be utilized for the formation of the cuticle and its sclerotization (Scheller *et al.*, 1990). The high content of aromatic amino acids in mosquito Vg may be an important source for cuticle formation of the rapidly developing mosquito embryos. This property of mosquito Vg could be another evolutionary adaptation in the utilization of the protein-rich food for the benefit of the next generation.

In addition to the high content of aromatic amino acids, mosquito Vg shares a limited sequence similarity with several larval storage proteins (Table 1) and other Vgs (Figure 5). The sequence homology between mosquito Vg and arthropod serum proteins suggests that there are regions of these proteins which

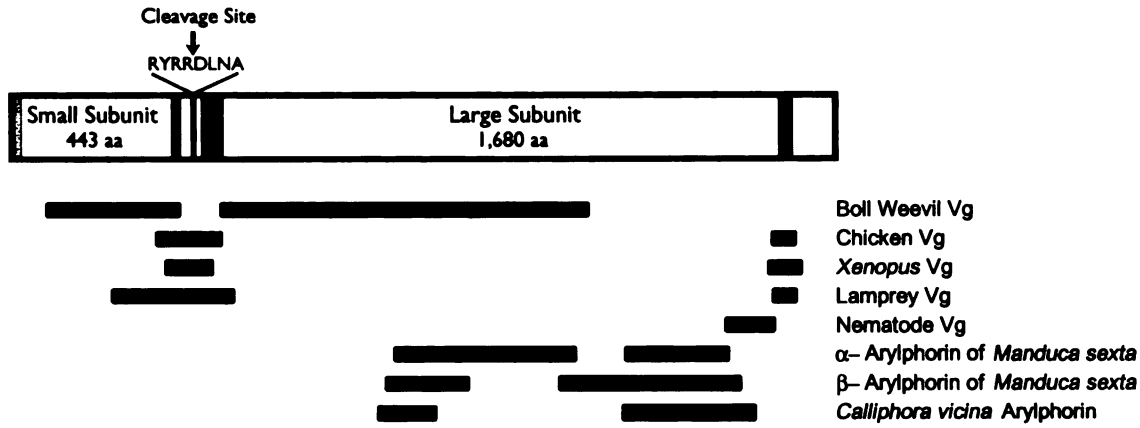


Figure 5. Diagram showing the structure of mosquito prepro-Vg and its similarity to other Vgs and insect larval serum proteins. The empty bar stands for the deduced amino acid sequence of mosquito pre-pro-Vg. The light gray region is the signal peptide (16 residues long) and the dark gray regions are polyserine domains. The black bars indicate the regions in mosquito Vg that have homology to other Vgs and insect larval serum proteins. The alignments of deduced amino acid sequence of mosquito Vg with other proteins were analyzed by the GCG programs (University of Wisconsin Genetics Computer Group Software) according to Devereux *et al.* (1984). The deduced amino acid sequence of mosquito Vg was compared to deduced amino acid sequences translated from GenBank (release 73.0) and SwissProt (release 23.0) databases.

may be important to maintaining their stability in the hemolymph. There is a conservation of several domains among the Vgs of *C. elegans*, *Xenopus* and chicken and vertebrate serum proteins such as apolipoprotein-B, apolipoprotein-E and the von Willebrand factor (Byrne *et al.*, 1989; Perez *et al.*, 1991). Sequence alignment also shows that there is a 69.2 % similarity within a 1,404 amino acid overlap between Vgs of mosquito and boll weevil (Figure 5). Yet, boll weevil Vg neither has serine-rich nor tyrosine-rich regions. There is a deletion of polyserine domains in the cleavage site region of boll weevil Vg when compared to that of mosquito. However, similar to mosquito Vg, the cleavage site of boll weevil Vg is flanked by highly charged residues. As for the comparison of mosquito Vg with nematode Vg (Spieth *et al.*, 1985), there is a 68.7 % similarity in a 115 amino acid overlap region. The similarity of mosquito Vg with the known sequences of vertebrate Vgs is limited to the polyserine domains (Figure 5).

Like the Vgs of all oviparous animals, mosquito Vg is extensively co- and post-translationally modified (Dhadialla & Raikhel, 1990). In the deduced amino acid sequence of pre-pro-Vg, there are 17 potential N-linked glycosylation sites with the consensus, Asn-X-Ser or Asn-X-Thr (Marshall, 1972). The small subunit has 5 potential glycosylation sites and the large one 12 (Figure 1). According to the consensus sulfation rules (Huttner, 1987), mosquito Vg has 13 putative tyrosine sulfation sites; one in the small subunit and 12 in the large one (Figure 1). This is consistent with previous data showing that the large Vg subunit is sulfated more extensively than the small one (Dhadialla & Raikhel, 1990).

Mosquito pre-pro-Vg has three polyserine domains, two flanking the subunit cleavage site and the third located near the pre-pro-Vg carboxyl-terminus (Figures 1 and 5). After cleavage, the small subunit has a polyserine domain at its carboxyl-terminus, while the large subunit has two such domains near both termini (Figures 1 and 5). The consensus sequence for phosphorylation by

casein kinase II (CK-2) identified for vertebrate phosvitins (Kuenzel *et al.*, 1987; Meggio & Pinna, 1988) is present in all three polyserine domains of mosquito Vg (Figure 1). In addition, two and 11 individual potential phosphorylation serine sites are found in the small and large Vg subunits respectively (Figure 1).

Significantly, mosquito Vg has three large polyserine domains which are similar to that of the phosvitin of vertebrate Vgs. The presence of a polyserine domain has previously been considered to be specific to vertebrate Vgs (Wahli, 1988). In contrast to mosquito Vg, there is only one polyserine domain in the phosvitin region, located in the middle of vertebrate Vgs (Wahli, 1988). The evolutionary trend towards the expansion of the serine-encoding capacity of phosvitins from fish to birds has been demonstrated (Byrne *et al.*, 1989). It is unknown why mosquitoes require more phosphoserines in their Vg than other invertebrates studied so far. The phosvitin region of vertebrate Vgs is suggested to be a calcium carrier for embryonic bone formation (Wahli, 1988; Byrne *et al.* 1989). However, the acquisition of polyserine domains by mosquito Vg could be a consequence of blood feeding, as a result of which high levels of phosphates are assimilated by the vitellogenic female mosquito and transferred to the next generation in the form of yolk protein. Analysis of the Vg sequences from other blood-feeding insects, particularly from those with independently evolved blood feeding may provide support for this hypothesis. Interestingly, there are deletions at the places of polyserine domains in the Vg sequence from the herbivore insect, the boll weevil, which is otherwise well aligned with mosquito Vg.

Analyses of hydropathy (Kyte & Doolittle, 1982), the surface probability (Emini *et al.*, 1985) and the antigenic index analyzed by GCG program reveal that within the predominantly hydrophilic pre-pro-Vg sequence, four major highly hydrophilic and charged regions are evident (Figure 4). The first and third hydrophilic regions are notable in their richness in tyrosine and charged

residues. Two other regions, marked 2 and 4 in Figure 4, are serine-rich. These highly hydrophilic regions have a strong propensity to distribute themselves on the surface of the folded protein (Emini *et al.*, 1985).

In order for Vg to be accumulated by developing oocytes it should be specifically recognized by the oocyte receptors. The properties of Vg-receptor interaction appear to be conserved among oviparous animals. In all animals studied so far including the mosquito, the charge of the Vg molecule is critical to its interaction with the receptors. Vitellogenin-receptor interaction can be disrupted by negatively-charged compounds such as suramin (Raikhel and Dhadialla, 1992).

Phosphorylated polyserine domains of Vg may play an important role in its recognition by the oocyte receptors, affecting the charge of the molecule. The dephosphorylation of chicken phosvitin decreased its uptake by oocytes (Miller *et al.*, 1982). Similarly, the binding of mosquito Vg to its receptor was significantly reduced when phosphates were removed (Dhadialla *et al.*, 1992). This is in agreement with the observation that the large Vg subunit, which has more potential phosphorylation sites (Figure 1), has a higher affinity to the Vg receptor than the small Vg subunit (Dhadialla *et al.*, 1992). A charged domain in chicken and *Xenopus* phosvitins that forms an alpha helix is believed to have a functional role in recognition by the oocyte Vg receptor (Byrne *et al.*, 1989). Sharrock *et al.* (1992) proposed that the region of lamprey Vg, which contains most serines and charged residues and is removed during Vg processing in the oocyte, is involved in receptor binding. Similar domains are present within the serine-rich regions in both subunits of mosquito Vg. However, since polyserine domains are not conserved among Vgs, other portions of the molecule should play a role in determining Vg-receptor interaction. The tyrosine-rich domains in mosquito Vg also form highly hydrophilic charged domains which could be involved in

interaction with the Vg receptor as well (Figures 1 and 4). Further studies involving X-ray crystallography and receptor-binding analysis with modified mosquito Vg are needed in order to identify precisely the Vg domains responsible for its binding to the receptor. The present sequence analysis of mosquito Vg has provided a foundation for such studies.

The deduced amino acid sequence of mosquito Vg reveals several unique characteristics of this developmentally important protein. The unique features of mosquito Vg may reflect the adaptations of this molecule for passing the nutritional values of the vertebrate blood to the developing mosquito embryos and thus, contributing to the evolutionary success of mosquitoes. Therefore, in addition to a well-established conservation of Vgs throughout evolution (Wahli, 1988; Byrne *et al.*, 1989), our findings emphasize the possibility of a rapid evolutionary change of Vg in specialized groups.

ACKNOWLEDGMENTS

We thank Dr. Henry H. Hagedorn for the kind gift of the mosquito Vg-A1 genomic probe, Alan R. Hays for purifying mosquito vitellogenin subunits for peptide sequencing and antibody preparation, Drs. Suzanne M. Thiem and Kirk W. Deitsch for their critical reading of this manuscript. This research was supported by a grant from the National Institutes of Health AI-24716 to ASR.

CHAPTER 3

CONSERVATION IN INSECT VITELLOGENINS

ABSTRACT

Vitellogenins (Vgs) are the major yolk proteins that supply nutrition for developing embryos of oviparous animals. Recent efforts in molecular cloning and characterization of Vg cDNAs and genes from several insect species have made it possible to compare insect Vgs with those of other taxa in more detail. Vgs are believed to have evolved from a common ancestor, however, their peptide sequences and functional domains are not highly conserved among species from different phyla. Insect Vgs are evolutionarily distant from other Vgs. It is not only because their sequences show limited similarity to other Vgs but also because they have conserved acidic serine-rich and charged domains flanking a paired basic motif by the cleavage site. These conserved acidic serine-rich and charged domains form β-turns that are essential for recognition of convertase to the paired basic motif. Discovery of these conserved domains and paired basic motifs at the cleavage sites of insect Vgs strongly suggest that the insect Vg processing enzymes may belong to the convertase family. Sequence analyses also indicated that the acidic serine-rich domains of insect Vgs are similar to phosvitin, a vertebrate Vg-specific serine-rich subunit. Unlike vertebrate Vgs that have a single long serine-rich domain, the number and size of acidic serine-rich domains in Vgs vary considerably among insect species.

INTRODUCTION

Vitellogenin (Vg) is a key protein essential for the reproduction of most egg-laying animals because it is the major nutrient source for developing embryos. Apart from higher dipteran insects (Terpstra and AB 1988; Rina and Savakis 1991) and sea urchins (Shyu *et al.* 1987), whose yolk protein sequences are unrelated to any known Vg sequences, vitellogenin is universally found in eggs of oviparous animals (Byrne *et al.* 1989). Vg is a yolk protein precursor synthesized and secreted by extraovarian tissues and circulated in the serum until sequestered by oocytes through a receptor mediated pathway (Byrne *et al.* 1989; Raikhel and Dhadialla 1992). In addition to supplying amino acids for growing embryos, most Vgs carry covalently linked carbohydrates after glycosylation and ions after phosphorylation and sulfation (Kunkel and Nordin 1985; Byrne *et al.* 1989; Dhadialla and Raikhel 1990). Furthermore, circulating lipids and hormones may bind non-covalently to Vgs in the serum, subsequently accumulating together in developing oocytes (Kunkel and Nordin 1985; Bownes *et al.* 1988).

Although Vgs are physiologically similar among animals from different phyla, the secretion and cleavage of Vg precursors are quite different (Figure 1) (Byrne *et al.* 1989; Spieth *et al.* 1991; Raikhel and Dhadialla 1992). Insect Vgs are synthesized in the fat body, the functional equivalent of the vertebrate liver (Byrne *et al.* 1989). In most insects, following glycosylation, phosphorylation and in some cases, sulfation, they are cleaved into small and large subunits (Figure 1) and form large oligomeric molecules before being secreted into the hemolymph (Raikhel and Dhadialla 1992). In contrast, vertebrate Vgs are synthesized and post-translationally modified in the liver and secreted as a precursor into the blood. These precursor are cleaved into three subunits, lipovitellin 1, phosvitin and lipovitellin 2, in a tandem manner in oocytes (Figure

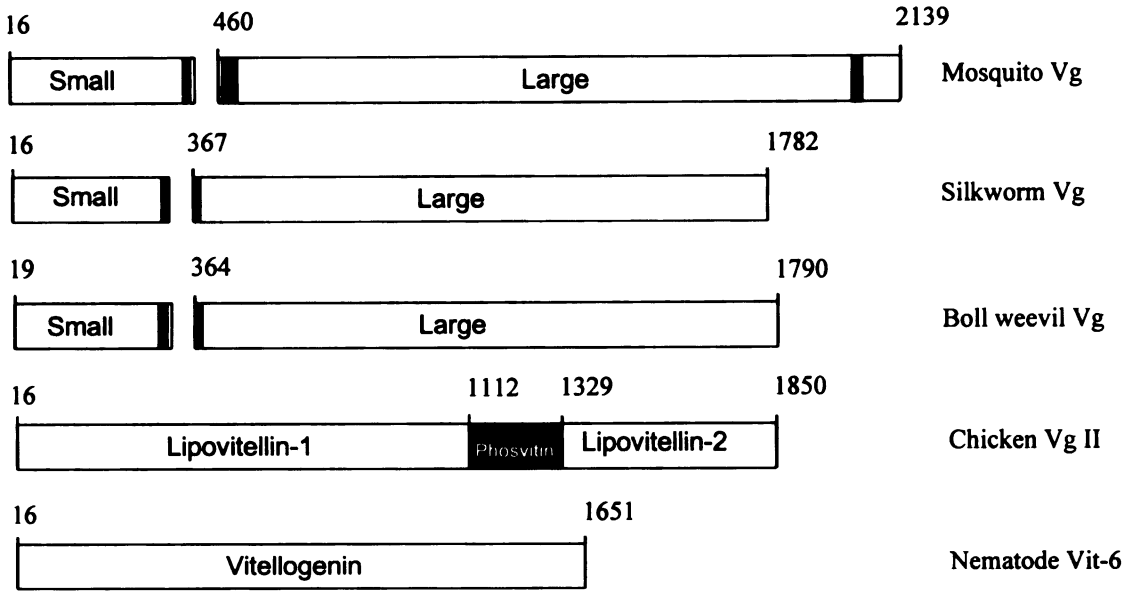


Figure 1. Peptide structures of secreted vitellogenins of insects, chicken and nematode. Insect vitellogenins are secreted as subunits, while vertebrate vitellogenins are precursors. Nematode Vit-6 is cleaved after secretion, however, the cleavage site is unknown and not shown here. A single box represents the precursor, while the separated boxes are subunits. Solid bars symbolize the location and size of a serine-rich domains. The numbers on each box indicate the amino acid numbers of the border residues of each subunit.

1). Among these subunits, phosvitin is unique in that it contains an extremely high percentage of serine residues (Byrne *et al.* 1989). Nematode Vgs are synthesized in the intestine and transferred to the gonad. There are two types of nematode Vgs (Byrne *et al.* 1989). The products of genes vit 2-5 are not modified or cleaved, while the vit-6 product is cleaved into small and large subunits after secretion from the intestine (Sharrock 1984; Spieth *et al.* 1991).

The evolutionary relationship between vertebrate and nematode Vgs at the molecular level including peptide and nucleotide sequences has been intensively reviewed (Nardelli *et al.* 1987; Wahli 1988; Byrne *et al.* 1989; Spieth *et al.* 1991). Recent efforts in molecular cloning and characterization of Vg cDNAs and genes from several insect species (Trewitt *et al.* 1992; Chen *et al.*

1994; Yano *et al.* 1994a,b) have made it possible to compare insect Vgs with those of other taxa in more detail. Based on the gene structure, Trewitt *et al.* (1992) suggested that insect Vgs may share a common ancestor with nematodes and vertebrates. Sequence analyses of insect Vgs reveal that they have multiple acidic serine-rich (ASR) domains that contain polyserine stretches and acidic residues. ASR domain resembles the phosvitin domain of vertebrate Vgs that also contain polyserine stretches. However, not all of phosvitin domains contain acidic residues (Byrne *et al.* 1989). The sequence alignment analysis indicates conservation in the deduced amino acid sequences of insect Vgs (Chen *et al.* 1994; Yano *et al.* 1994b), but that they are less similar to (Chen *et al.* 1994) and evolutionarily distant from (LaFleur, Jr. *et al.* 1995) those of other animals. This paper mainly focuses on the amino acid sequence conservation among known insect Vgs and the evolutionary relationship of polyserine stretches in insect and vertebrate Vgs.

MATERIALS AND METHODS

Amino acid sequences of Vgs were obtained from the GenBank data base and published papers. The source of Vg sequences analyzed in this paper are: boll weevil (M72980; Trewitt *et al.* 1992), chicken (M18060; van het Schip *et al.* 1987), frog (Y00354, Gerber-Huber *et al.* 1987), gypsy moth (L28097; Hiremath *et al.* 1994), lamprey (M88749; Sharrock *et al.* 1992), nematode (X56121 and X56213; Spieth *et al.* 1991), sawfly (Kageyama *et al.* 1994), silkworm (D13160; Yano *et al.* 1994) and yellow fever mosquito (U02548; Chen *et al.* 1994).

Sequence alignment and the secondary structure analyses were performed by using Gap and Peptidestructure programs, respectively, of Genetics Computer Group software (University of Wisconsin, Madison, WI). The

gap weight was 3.00 and the gap length weight 0.10. The window for hydrophobicity analysis (Kyte and Doolittle 1982) was 7 amino acids.

RESULTS AND DISCUSSION

Comparison of Amino Acid Composition and Deduced Amino Acid Sequence of Vitellogenins

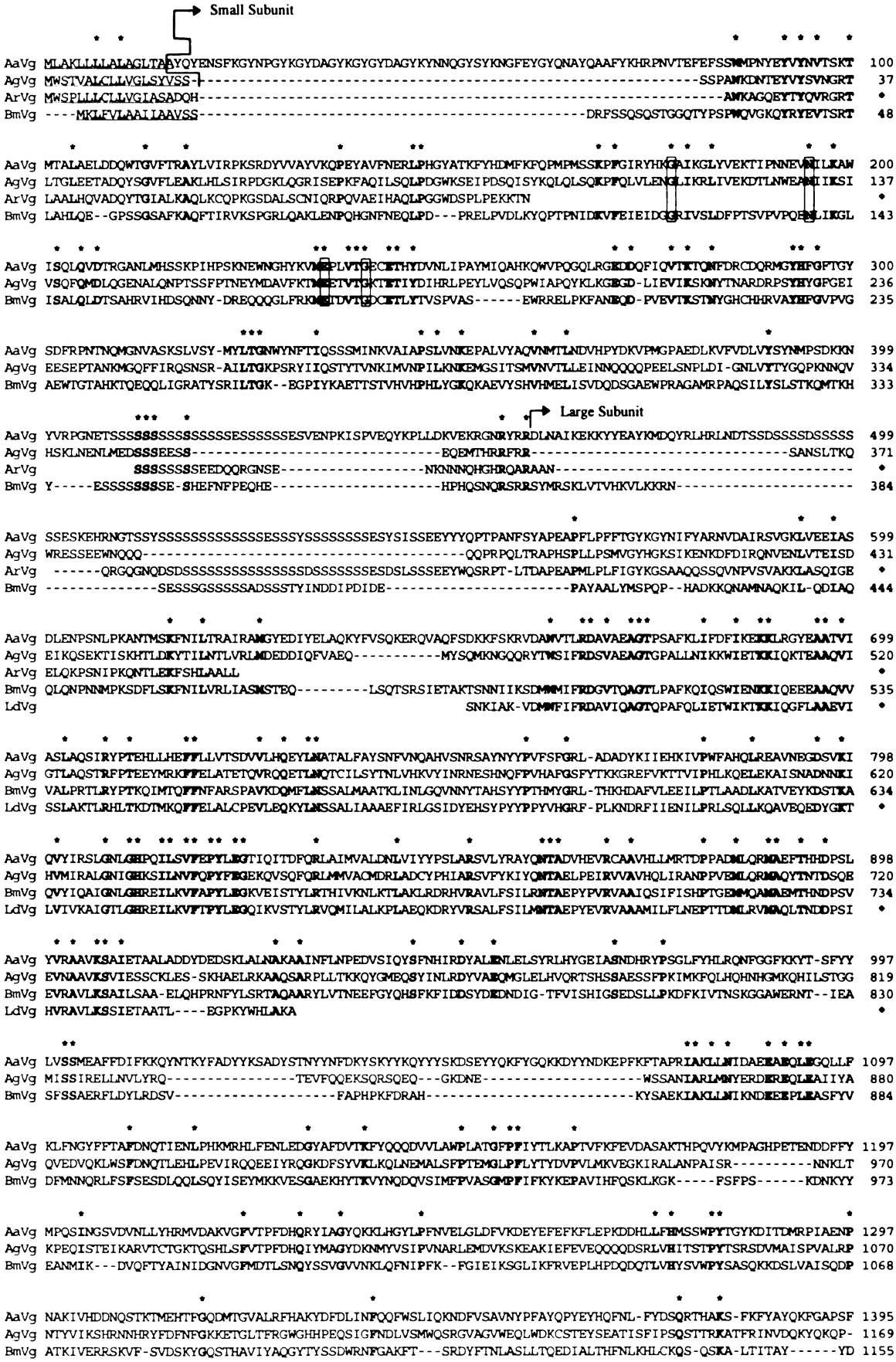
Amino acid composition of several Vgs have been reviewed in Byrne *et al.* (1989) and Trewitt *et al.* (1992). To compare the recently available insect Vg sequences with the previous data, an amino acid comparison analysis was done and listed in Table 1. Comparison of known Vgs with the average amino acid composition of eukaryotic proteins (Doolittle 1986) shows that most amino acids in the Vgs listed are similar to the average except that the amount of serine residues in vertebrate Vgs is almost double that found in the average eukaryotic protein (Table 1). This is due mainly to long stretches of polyserine in the phosvitin domains of vertebrate Vgs (Byrne *et al.* 1989). Some insect Vgs, including that of yellow fever mosquito and silkworm, also have high levels of serine residues (Table 1). In accordance with this feature, domains containing polyserine stretches similar to that of vertebrate phosvitins are found in these insect Vgs. Nevertheless, boll weevil Vg has an average serine content, and with short polyserine stretches (Figure 2 and Table 2).

The insect Vgs sequenced so far do not have the third ASR domain as found in the yellow fever mosquito (Figures 1 and 2, Table 2). Silkworm Vg has a serine content almost as high as that of mosquito (Table 1), however, unlike mosquito Vg which has three long ASR domains, it has only two short ones that flank its cleavage site (Figure 2). Most serine residues are scattered throughout the peptide (Yano *et al.* 1994a).

Table 1. Deduced amino acid composition of vitellogenins

Amino Acid	Amino Acid Composition (%)											
	Vitellogenins										Alpha Subunit of Arylphorin in Hornworm[k]	Average Amino Acid Composion in Eukaryotes[l]
	Mosquito[a]	Silkworm[b]	Boll Weevil[c]	Shrimp[d]	Nematode[e]	Leech[f]	Lamprey[g]	Kilifish[h]	Frog[i]	Chicken[j]		
Ala	7.2	6.8	4.8	10.8	4.8	4.5	8.7	8.7	8.0	7.2	4.6	7.5
Arg	3.6	4.7	5.0	4.7	6.2	4.5	5.6	4.9	5.0	5.9	3.4	4.3
Asn+Asp	10.8	9.8	10.2	6.2	8.2	10.4	7.0	7.5	8.6	8.6	12.0	10.8
Cys	0.9	0.8	1.1	1.3	1.3	ND[m]	1.4	1.6	1.7	2.0	0.0	1.7
Gln+Glu	10.5	12.9	16.3	11.4	17.1	11.0	10.7	9.6	12.3	9.7	10.5	9.7
Gly	4.6	4.5	4.3	7.1	2.8	5.4	4.9	3.8	4.8	4.6	4.1	6.9
His	2.9	3.4	2.4	2.5	2.1	2.6	2.6	1.9	3.0	2.6	2.6	2.4
Ile	3.2	5.3	5.7	6.7	5.5	6.3	4.4	6.0	5.2	5.4	4.3	6.1
Leu	5.3	6.5	7.4	7.4	8.8	11.0	8.8	9.3	7.5	7.7	6.6	8.9
Lys	6.9	6.5	7.3	6.3	7.7	6.7	6.7	7.6	7.7	6.8	7.6	6.5
Met	1.8	2.2	2.7	2.2	2.1	1.3	2.4	2.6	2.6	2.5	2.5	2.0
Phe	8.1	4.4	3.5	3.8	4.4	3.8	3.0	3.2	3.6	2.8	9.0	4.3
Pro	5.0	4.5	4.1	5.4	4.0	5.9	4.6	5.6	4.2	4.8	6.6	4.6
Ser	**10.1**	**9.8**	**7.7**	**7.4**	**7.4**	**7.8**	**14.4**	**12.3**	**11.1**	**14.0**	**5.3**	**7.1**
Thr	3.7	6.1	6.0	6.1	6.2	6.5	4.3	5.6	4.9	4.9	2.0	5.8
Trp	0.6	0.9	0.9	0.0	0.9	ND	0.8	0.6	0.7	1.0	0.9	1.4
Tyr	10.1	5.3	4.1	3.2	3.3	3.7	2.6	2.8	2.9	3.0	10.9	3.8
Val	4.9	5.9	6.8	7.8	7.3	8.7	6.9	7.5	6.1	6.6	7.3	6.1

[a] Chen et al. (1994); [b] Yano et al. (1994); [c] Trewitt et al. (1992); [d] Data from an animo acid analysis (Tom et al. 1992); [e] Vit-5 gene (Spieth et al. 1985); [f] Data from an amino acid analysis (Baert et al. 1991); [g] Sharrock et al. (1992); [h] LaFleur et al. (1995); [i] A2 gene (Gerber-Huber et al. 1987); [j] van het Schip et al. (1987); [k] Willott et al. (1989); [l] Doolittle (1986); [m] Not detected


Small Subunit
Large Subunit
AaVg
AgVg
ArVg
BmVg
LdVg

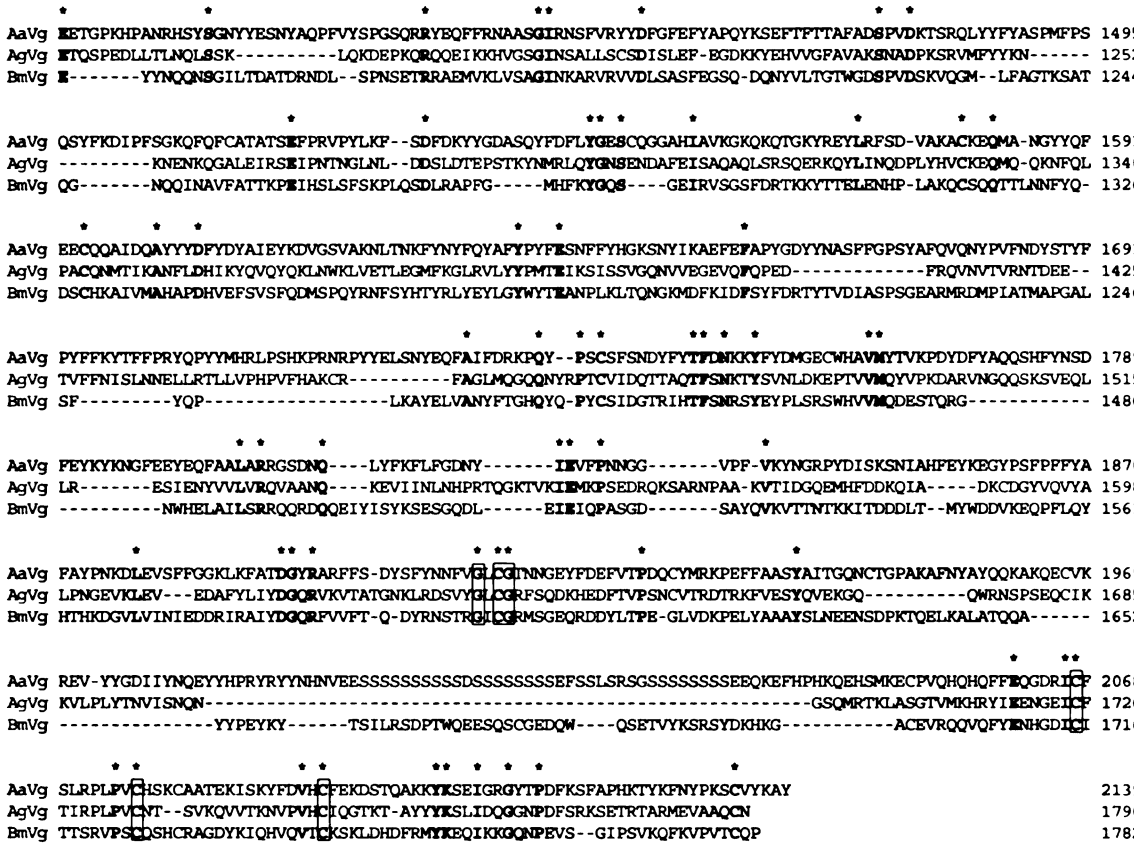


Figure 2. Comparison of amino acid sequences of several insect vitellogenins. The conserved residues among insect Vgs are in bold print and marked by an asterisk. A box represents residues that are also conserved in Vgs of nematode, chicken and frog (Spieth *et al.* 1991). A diamond symbol indicates that the numbers of these sequences are unknown because only partial sequences are available. The abbreviations are: AaVg, vitellogenin of yellow fever mosquito, *Aedes aegypti* (Chen *et al.* 1994); AgVg, vitellogenin of boll weevil, *Anthonomus grandis* (Heilmann *et al.* 1993); ArVg, vitellogenin of sawfly, *Athalia rosae* (Kageyama *et al.* 1994); BmVg, vitellogenin of silkworm, *Bombyx mori* (Yano *et al.* 1994b); LdVg, vitellogenin of gypsy moth, *Lymantria dispar* (Hiremath *et al.* 1994).

Table 2. Percentage of identities and similarities in aligned amino acid sequences of insect vitellogenins

	Similarity (%)			
between Species	Signal Peptide	Small Subunit	Large Subunit	Total
mosquito / boll weevil	43.8	52.1	49.6	49.9
mosquito / silkworm	75.0	49.0	50.3	50.7
boll weevil / silkworm	40.0	51.1	47.5	48.1

In addition to serine, yellow fever mosquito Vg is notable for its richness in the aromatic amino acids phenylalanine and tyrosine (Chen *et al.*, 1994). The percentage is about two- to three-fold higher than in other Vgs, and is similar to that of arylphorin from the hornworm, *Manduca sexta* (Willott *et al.* 1989) (Table 1). The sequence of the large subunit of yellow fever mosquito Vg shows similarity to arthropod serum storage proteins including arylphorins and hemocyanins (Chen *et al.* 1994). The richness of both mosquito Vg and arthropod serum storage proteins in aromatic amino acids also reflects this similarity (Table 1). Accordingly, boll weevil and silkworm Vgs, which are not rich in aromatic amino acids, show no similarity to these arthropod serum proteins. In contrast to the observation that vertebrate and nematode Vgs share similarity to the mammalian apolipoprotein B and the von Willebrand factor (Baker 1988a,b), insect Vgs show no sequence similarity to these mammalian serum proteins (Chen *et al.* 1994).

The percentage of alanine residues in boll weevil Vg is much lower than average of eukaryotic proteins. This is similar to the Vgs of nematode and leech.

The total content of glutamic acid and glutamine (Glx) in boll weevil and nematode Vgs are much higher than other Vgs. The functional implications of these differences are still not clear (Byrne *et al.* 1989; Trewitt *et al.* 1992).

Peptide sequences of Vgs are not highly conserved probably because Vg functions as an egg storage protein that supplies nutrition for embryos. Yet, limited sequence conservation was identified among different phyla. Among invertebrate Vgs, the amino acid sequence of nematode Vgs most closely resembles that of vertebrate Vgs, but lacks the phosvitin domain (Nardelli *et al.* 1987; Wahli 1988; Byrne *et al.* 1989; Spieth *et al.* 1991). Insect Vgs show limited sequence similarity to other oviparous animals (Trewitt *et al.* 1992; Chen *et al.* 1994; LaFleur, Jr. *et al.* 1995). Although the overall sequence identity among insect Vgs is not high (Table 3), Yano *et al.* (1994b) reported that there is a sequence conservation near the N-terminus of the large subunit of insect Vgs. Nevertheless, the precursor of these known insect Vgs can be entirely aligned (Figure 2) and sequence conservation (the bold print in Figure 2) is not only limited to the N-terminus of the large subunit.

There is a 20-residue segment conserved between nematode and vertebrate Vgs (Spieth *et al.* 1991). In addition, many cysteine residues that may form disulfide bonds are found conserved among Vgs of these animals (Spieth *et al.* 1991). These authors therefore suggested that the tertiary structures of these Vgs have been maintained through evolutionary time. Interestingly, the deduced amino acid sequences of insect Vgs have neither the 20-residue segment nor high conservation in cysteine residues as identified by Spieth *et al.* (1991). Very limited sequence conservation, located near either termini of the Vg precursors, was found when the results of Spieth *et al.* (1991) were compared with insect Vgs (boxed residues in Figure 2). There are only four such conserved cysteine residues identified in insect Vgs, and all are located at the C-terminus of the Vg

Table 3. Comparison of number, size and location of serine-rich domains in insect vitellogenins

Animal	No. of ASR Domains	Residues in an ASR Domain[a] (% Ser, % TCX codons)			
		I	II	III	Total
Mosquito	3	28 (78.6, 90.9)	64 (68.8, 86.4)	40 (75.0, 80.0)	132 (72.8, 85.4)
Boll weevil	2	20 (30.0, 50.0)	5 (40.0, 50.0)	—	25 (32.0, 50.0)
Silkworm	2	13 (69.2, 100)	16 (75.0, 66.7)	—	29 (72.4, 81.0)
Sawfly[b]	2	11 (72.7, 54.5)	40 (72.5, 44.8)	—	51 (72.5, 51.4)

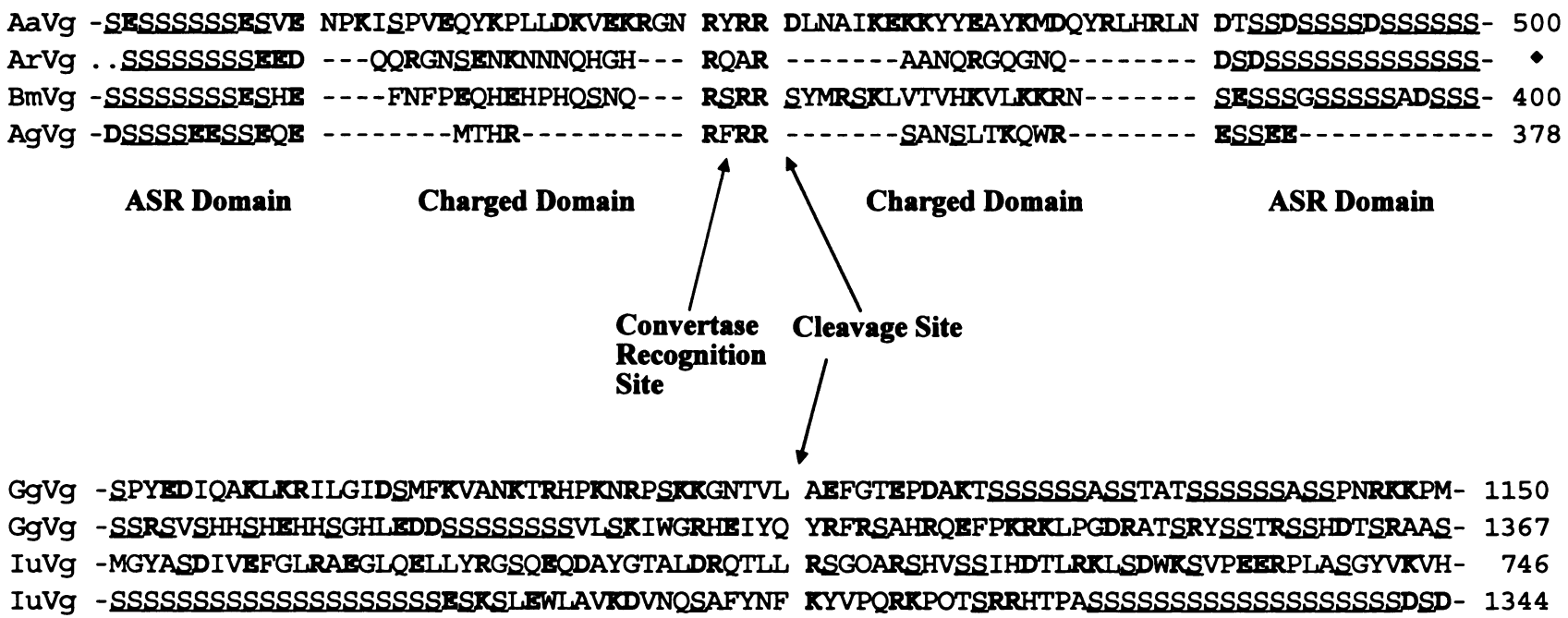
[a] Acidic serine-rich domain
[b] Not completely sequenced. Data listed are based on sequence available. The first ASR domain is truncated.

precursors (Figure 2). Sequence alignment and comparison of conserved cysteine residues among known Vg sequences suggest that only a partial tertiary structure of insect Vg precursors has been retained from the common Vg ancestors.

Unique Domains Are Conserved around the Cleavage Site of Insect Vitellogenins

Insect Vgs are different from other Vgs because they are cleaved before being secreted by the fat body. This is also true of their cleavage sites and the domains flanking this cleavage site. Just as the ways of Vg cleavage vary in animals from different phyla, there also is no universal sequence around the cleavage sites of Vgs. Although the sequence at the N-termini of the large subunits are not conserved, an RXRR or an RXXR motif, which is recognized by eukaryotic convertases (Barr 1991), was identified at the N-terminal side of the cleavage site, i.e., the C-termini of the small subunit of insect Vgs (Figure 3).

Figure 3. Sequence of domains flanking the cleavage site of several vitellogenins. The abbreviations and references used in this figure are as in Figure 2 except that GgVg and luVg are vitellogenins of chicken (van het Schip *et al.* 1987) and lamprey (Sharrock *et al.* 1992), respectively. Acidic serine-rich domains are abbreviated as ASR domains. Each region was determined by sequence similarity and are separated by a space between each region. Both ASR domains of yellow fever mosquito and the second such a domain of sawfly shown in this figure are only partially presented. For GgVg and luVg, the upper sequences are those found around the first cleavage sites, while the lower ones are those found around the second sites. Charged amino acid residues are shown in bold print, while serine residues are underlined. Dots and diamond symbols indicate the sequence and the sequence number, respectively, are unknown because the complete sequence of sawfly vitellogenin is not available.


AaVg -SESSSSSSESVE NPKISPVEQYKPLLDKVEKRGN RYRR DLNAIKEKKYYEAYKMDQYRLHRLN DTSSDSSSSDSSSSSS- 500
ArVg ..SSSSSSSSEED ---QQRGNSENKNNNQHGH--- RQAR -------AANQRGQGNQ-------- DSDSSSSSSSSSSSSS- ◆
BmVg -SSSSSSSSESHE ----FNFPEQHEHPHQSNQ--- RSRR SYMRSKLVTVHKVLKKRN------- SESSSGSSSSSADSSS- 400
AgVg -DSSSSEESSEQE --------MTHR---------- RFRR -------SANSLTKQWR-------- ESSEE------------ 378
ASR Domain
Charged Domain
Charged Domain
ASR Domain
Convertase Recognition Site
Cleavage Site
GgVg -SPYEDIQAKLKRILGIDSMFKVANKTRHPKNRPSKKGNTVL AEFGTEPDAKTSSSSSSASSTATSSSSSSASSPNRKKPM- 1150
GgVg -SSRSVSHHSHEHHSGHLEDDSSSSSSSSVLSKIWGRHEIYQ YRFRSAHRQEFPKRKLPGDRATSRYSSTRSSHDTSRAAS- 1367
IuVg -MGYASDIVEFGLRAEGLQELLYRGSQEQDAYGTALDRQTLL RSGOARSHVSSIHDTLRKLSDWKSVPEERPLASGYVKVH- 746
IuVg -SSSSSSSSSSSSSSSSSSSSESKSLEWLAVKDVNQSAFYNF KYVPQRKPOTSRRHTPASSSSSSSSSSSSSSSSSSSDSD- 1344

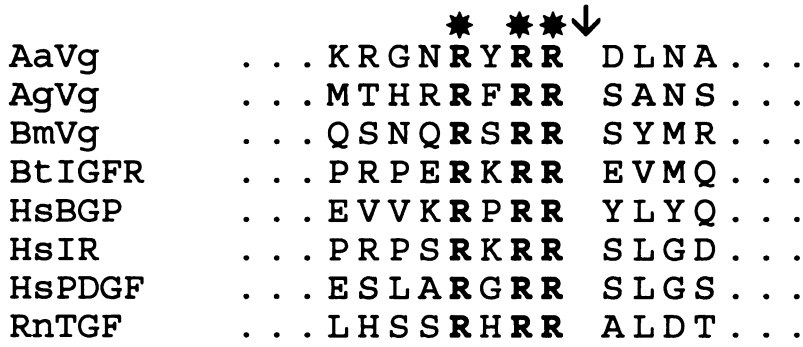


Figure 4. Comparison of amino acid sequences around the cleavage site of several proteolytic processed proteins. The abbreviations and references used in this figure are as in Figure 2 except for BtIGFR, *Bos taurus* insulin-like growth factor-1 receptor beta subunit (cattle) (Sneyers *et al.* 1991), HsBGP, *H. sapiens* Gla protein (Kiefer *et al.*, 1990), HsIR, *Homo sapiens* insulin receptor (human) (Ebina *et al.*, 1985), HsPDGF, *H. sapiens* platelet-derived growth factor B chain (Collins *et al.*, 1985) and RnTGF, *Rattus norvegicus* transforming growth factor-beta 1 (rat) (Qian *et al.*, 1990). The consensus for the subtilisin-like convertase is in bold and marked with the asterisk. The arrowheads indicate the cleavage sites.

This cleavage site motif is identical to that found in precursors of some peptide hormones and some hormone receptors (Figure 4).

The sequence alignment of insect Vgs indicates that the regions surrounding the cleavage sites are not similar at the residue level, but rather are conserved at the domain level (Figures 2 and 3). The domain arrangement shows conservation in a manner as ASR domains flanking two charged domains that enclose an RXRR motif or an RXXR motif and a cleavage site (Figure 3). A derivative form was observed in boll weevil Vg in which ASR and charged domains are short and many serine residues are either deleted or substituted by polar and charged residues (Figures 2 and 3). Both ASR and the charged domains may be important for cleavage of insect Vg precursors because they form β-turn structures (Figure 3). Rholam *et al.* (1986) observed that not all the

paired basic motifs in precursors are cleaved. These authors found only the paired basic sites that are located inside or next to β-turns are cleaved *in vivo*. Brakch *et al.* (1993) used mutated analogs to determine the significance of β-turns to cleavage of prohormones. These authors concluded that the β-turn is a key that optimized recognition of convertases. All paired basic motifs in insect Vgs are situated in or next to β-turns (Figure 3). These results suggest that the insect Vg processing enzyme responsible for subunit cleavage may belong to a family of eukaryotic convertases (Heilmann *et al.* 1993; Chen *et al.* 1994; Yano *et al.* 1994b). In addition, ASR and charged domains are highly hydrophilic (Figure 5). They have a strong tendency to locate themselves outward. Furthermore, these regions form turns. They are very flexible and are usually exposed on the surface of a folded protein (Chothia *et al.* 1989). Such an outward flexible structure may facilitate segmental adaptation of a peptide and favor recognition of convertase to the paired basic motif (Brakch *et al.* 1993). These unique structures are consistent with insect-specificity of pathways of Vg cleavage in the fat bodies (Dhadialla and Raikhel 1990). The cleavage of the pro-Vg is essential for secretion because mosquito Vg precursor can not be secreted from mosquito fat bodies (Raikhel and Dhadialla, 1990). Therefore, ASR and charged domains with the RXRR motif of insect Vgs may be a result of co-evolution with the processing enzymes, convertases.

Comparison of sequences around the known cleavage sites of vertebrate Vgs indicates that sequences at the subunit borders of chicken (van het Schip *et al.* 1987), frog (Gerber-Huber *et al.* 1987) and lamprey (Sharrock *et al.* 1992) Vgs do not have a doublet basic motif for convertases (Figure 3). The precursor of vertebrate Vgs are cleaved by cathepsins D in oocytes (Retzek *et al.* 1992; Yoshizaki and Yonezawa 1994) rather than in the extraovarian tissue as in insects. Amino acid sequences surrounding the cleavage sites of vertebrate Vgs

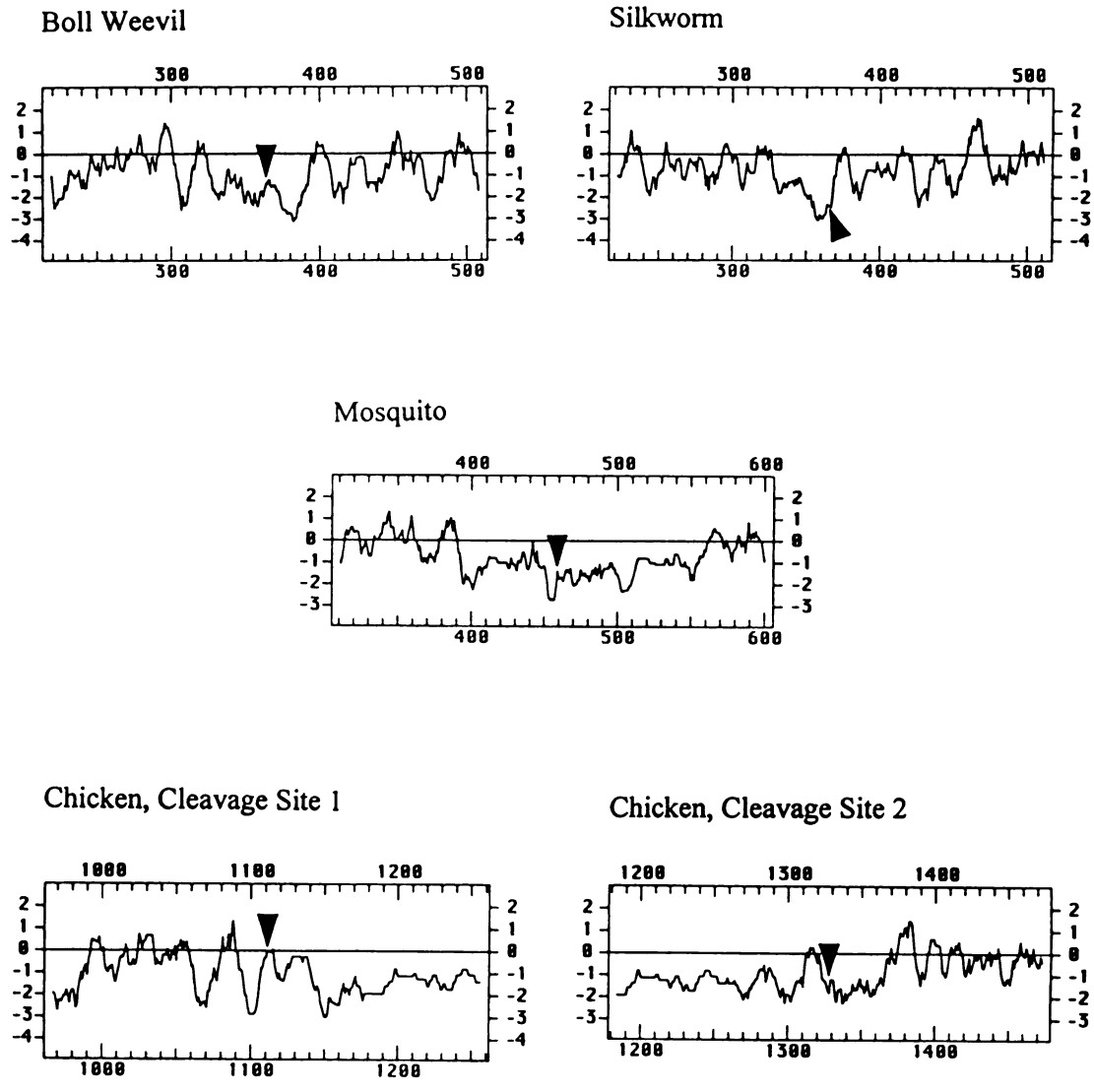


Figure 5. Hydrophobicity analysis of amino acid sequences around the cleavage site of insects and chicken Vgs. Solid arrowheads indicate the cleavage sites. These data were analyzed by Kyte and Doolittle (1982) methods by Peptidestructure program of GCG software.

are not similar to each other (Figure 3). Hydrophobicity analysis of these sequences also reflect this fact because not all of these domains have a strong hydrophilic tendency to push the cleavage site onto the surface (Figure 5). Some of these sequences also contain charged amino acids and polyserine stretches like insect Vgs. Unlike insect Vgs, however, there is only one serine-rich domain near each cleavage site of chicken Vg but none around the first cleavage site of lamprey Vg (Figure 3). It is interesting that the second cleavage site of lamprey Vg has a similar domain arrangement with that of insects except for the paired basic motif (Figure 3). However, these similarities may not reflect functional significance in cleavage of vertebrate Vgs because the domains flanking the cleavage site are not conserved in vertebrate Vgs. These charged domains in vertebrate Vgs have been identified as necessary for recognition and uptake by the Vg receptor in the oocyte membrane (Byrne *et al.* 1989).

Similar to insect Vgs, the product of vit-6 gene of Nematode is also cleaved into two subunits (Sharrock 1984). The cleavage site and the processing enzyme have yet not been determined. Interestingly there is an RXKR paired basic motif located by a β-turn which forms a putative recognition site for convertases. Although there is no adjacent ASR and charged domains as of insect Vgs, this site may be recognized by convertase. Further investigations are required to test whether nematode Vg is cleaved by convertases.

Evolution of the Serine-rich Domains in Vitellogenins

The most significant similarity between vertebrate and insect Vgs are the domains that contain polyserine stretches. ASR domains in insect Vgs have many acidic amino acid residues similar to some vertebrate phosvitin domains. Both the polyserine stretch itself (Meggio and Pinna 1988) and a serine with an acidic residue at +3, +4 or +5 sites (Kuenzel *et al.* 1987) are good substrates for

casein kinase-II catalyzed phosphorylation. Phosphorylation may play an important role in maintaining a special configuration that helps recognition of Vg by the oocyte Vg receptor. Dephosphorylation studies in the chicken phosvitin domain and yellow fever mosquito Vg decreased their uptake by oocytes (Miller *et al.* 1992; Dhadialla *et al.* 1992). Sharrock *et al.* (1992) suggested that phosphorylated polyserine stretches in lamprey Vg may be involved in receptor recognition/uptake because most of the polyserine stretches are removed after uptake of oocytes. Nevertheless negatively charged phosphorylated polyserines should not be the recognition sites because the Vg receptor belongs to LDL receptor family that have negatively charged ligand binding sites (Bujo *et al.* 1994). It is possible that the repellent forces, caused by the interaction between phosphorylated polyserine domains of Vgs and ligand binding sites of Vg receptors, somehow rotate Vg and facilitate the recognition sites on the surface of Vgs recognized by their receptors. The exceptional third ASR domain near the C-terminus of the large subunit of yellow fever mosquito Vg and the short ASR domains in boll weevil Vg may be results of evolution of their folding and of co-evolution with their oocyte Vg receptors.

It is not clear why mosquito and silkworm Vgs are richer in serine residues than other invertebrates. Chen *et al.* (1994) hypothesized that yellow fever mosquito may take the advantage of their highly phosphorylated food and pass phosphates through polyserine stretches in Vg to their embryos. However the silkworm does not ingest phosphate-rich food, so there is no comparable need for long and additional phosphorylatable polyserine stretches as in yellow fever mosquito Vg. One possibility is that it is used in silk production by newly hatched silkworm larvae because serine residues are one of the major components of silk fibroin in silkworm (Yamaguchi *et al.* 1989).

Alignment of the Vg sequences of chicken, frog and nematode suggested that the phosvitin domains arose after the separation of vertebrates and nematodes during the Precambrian period (Wahli 1988). The polyserine stretches within these domains have expanded through evolutionary time (Wahli 1988). Although polyserine stretches may have similar functions in insect and vertebrate Vgs, it is likely that insect and vertebrate Vgs have evolved polyserine stretches independently. Insects have no direct phylogenetic stem ancestor with vertebrates (Barrett *et al.* 1986). Since nematode Vgs do not have serine-rich domains, one may hypothesize that the common ancestor of Coelomata obtained polyserine stretches after the divergence from Acoelemata, and only vertebrates and insects maintain these domains. However, based on the amino acid components (Table 1), this seems improbable. Although the sequence data of shrimps and leeches Vgs are not available, it is unlikely that they have polyserines. Both Crustacea, that were evolved from a same stem ancestor of Insecta, and leeches, that were evolved from Protostomia, a stem ancestor of Arthropoda and Annelida (Barrett *et al.* 1986), have a serine content similar to the average of eukaryotes as nematode Vgs (Table 1). In addition, the codon usage (Table 2), the size and the location of serine residues in serine-rich domains found in Vg precursors of insects and vertebrates are very different (Figure 1). More than 50 % of serine residues in vertebrate Vgs are encoded by AGY codons (LaFleur, Jr. *et al.* 1995), while most serine residues in these domains of insect Vgs are encoded by TCX codons (Table 2). All the known vertebrate Vgs have a single long serine-rich domain, that becomes phosvitin subunit after cleavage, contains many polyserines. In contrast, insects have multiple ASR domains. These domains are eventually distributed to both small and large subunits after cleavage. Unlike vertebrates, ASR domains in insect Vgs are not processed as subunits as are the phosvitins of vertebrates (Chen *et*

al. 1994) (Figures 1 and 3), and probably retain their polyserine stretches given that there is no obvious size decrease after internalization by the oocyte (Raikhel and Bose 1988; Raikhel and Dhadialla 1992). The size of ASR domains of insect Vg precursors vary among insect species. Although it is not possible to compare the hierarchy of insects from different orders, it is likely that the more evolutionarily advanced the insect, the shorter the polyserine domain near the cleavage site of Vg (Figure 3). For example, boll weevil is a highly evolved Coleopteran insect. Its ASR domains are much shorter than other insects (Figure 3 and Table 3). In contrast, mosquito and sawfly are more ancient species in Diptera and Hymenoptera, respectively. Their Vgs maintain long polyserine stretches. It is possible that the Vg of the common insect ancestor obtained long ASR domains near the junction of the cleavage site. However, the sizes, numbers and locations of ASR domains have become diverse with the evolution of insects. Especially most of the polyserine stretches were somehow lost or substituted during the evolution of these more advanced insects. These observations support the argument that polyserines in insects and vertebrates are the result of convergence evolution.

ACKNOWLEDGMENT

We thank Drs. Kirk Deitsch and Tom Sappington for their invaluable suggestions to this manuscript. This work was supported by NIH grant AI-24716 to ASR.

CHAPTER 4

MOLECULAR CLONING AND FUNCTIONAL EXPRESSION OF MOSQUITO VITELLOGENIN CONVERTASE

ABSTRACT

A cDNA encoding mosquito vitellogenin convertase (VC) was cloned, characterized and functionally expressed. Biosynthesis of mosquito vitellogenin requires cleavage by VC because pro-vitellogenin cannot be secreted from the fat body. The transcript for mosquito VC is 4.2 kb. This gene is tightly regulated by a blood meal and expressed during early vitellogenesis. The VC cDNA clone has an open reading frame encoding a protein 115 kDa in size. This protein increases to 140 kDa after post-translational modification. The mosquito VC gene is exclusively expressed in the vitellogenic fat body of female mosquitoes. When the VC cDNA was co-expressed *in vitro* with the mosquito vitellogenin cDNA, it produced cleaved pro-vitellogenin. Thus, the newly found convertase is indeed a functional fat-body-specific enzyme for pro-vitellogenin cleavage. Discovery of mosquito VC is significant in that members of the convertase family were not previously known to play a role in the cleavage of insect pro-vitellogenins.

INTRODUCTION

Processing of pro-vitellogenins (pro-Vg), precursors of the major nutrient source supporting embryonic development in oviparous animals, varies greatly in

different phyla (Byrne *et al.*, 1989). Insect Vg precursors are cleaved into two subunits in the fat body before secretion (Dhadialla and Raikhel, 1990; Raikhel and Dhadialla, 1992). In contrast, vertebrate Vgs are secreted as intact molecules and cleaved in the oocyte (Byrne *et al.*, 1989). Among nematode Vgs, most of the products of Vg genes are not cleaved. The Vit-6 protein is the only one that is cleaved into two subunits after its secretion from the intestine (Sharrock, 1984).

Mosquito Vg has been characterized molecularly (Chen *et al.*, 1994) and biochemically (Dhadialla and Raikhel, 1990). Expression of the mosquito Vg genes is tightly regulated by a blood meal. These genes are expressed exclusively in the fat body of vitellogenic females (Raikhel and Dhadialla, 1992). Mosquito pro-Vg is 250 kDa in size. It is further cleaved into small and large subunits, which are subsequently post-translationally modified and become 66 and 200 kDa, respectively. Both subunits form a large oligomeric sulfated, phosphoglycolipoprotein and are secreted into the hemolymph (Dhadialla and Raikhel, 1990).

Cleavage of insect pro-Vg is crucial because pro-Vg can not be secreted from the fat body (Dhadialla and Raikhel, 1990). In spite of progress in characterizing pathways of biosynthesis of insect Vgs, enzymes that cleave insect pro-Vgs have not been identified. In contrast, cathepsins D have been reported to be the pro-Vg cleaving enzyme of vertebrates (Retzek *et al.* 1992; Yoshizaki and Yonezawa 1994).

Subtilisin-like convertases have been proposed to be the putative insect pro-Vg processing enzymes because paired basic motifs recognized by these enzymes have been identified at the cleavage sites of insect Vgs (Heilmann *et al.*, 1993; Chen *et al.*, 1994; Yano *et al.*, 1994). Furthermore, adjacent β-turns that are required for a paired basic motif recognized by convertases (Brakch *et*

al., 1989, 1993) have been identified at the cleavage sites of insect Vgs (Chen and Raikhel, unpublished). To elucidate the role that convertase plays in mosquito vitellogenesis, we have cloned and characterized a convertase cDNA from a mosquito vitellogenic female fat body cDNA library. This cDNA encodes a convertase that has a novel biological function in subunit cleavage of insect pro-Vgs. This enzyme is denoted as vitellogenin convertase (VC).

MATERIALS AND METHODS

Cloning and characterization of the mosquito VC cDNA - A mosquito VC cDNA fragment was obtained by PCR. Total RNA was isolated from fat bodies of 1-18 h post-blood-fed females and used as the template for reverse transcription and PCR. A pair of fully degenerate primers were designed based on the sequence of the catalytic domains of *Drosophila* dKLIP-1 (Hayflick *et al.*, 1992), dfurin1 (Roebroek *et al.*, 1991) and dfurin2 (Roebroek *et al.*, 1992). Both primers had an *Xba* I site at their 5'-ends. The sequence of the first primer was: GTTCTAGACA(CT)GG(ACGT)AC(ACGT)CG(ACGT)TG(CT)GC, while the second one was GTTCTAGAG(ACGT)GG(ACGT)GC(ACGT)GA(ACGT)GC(AC GT)GA. The PCR-amplified cDNA was subcloned into pUC119 and confirmed by hybridization to the cDNA coding for the catalytic domain of dKLIP-1 (Hayflick *et al.*, 1992). This PCR-product was used to screen a female mosquito vitellogenic fat body cDNA library. The conditions for reverse transcription, PCR amplification and cDNA cloning were as in Chen *et al.* (1994). The enzymatic method of Sanger *et al.* (1977) was utilized for DNA sequencing. Sequence alignment was performed by using the Gap program of the Genetics Computer Group software (University of Wisconsin, Madison, WI). The gap weight was 3.00 and the gap length weight 0.10.

Northern and slot blot analyses - Poly(A^+) RNA was prepared from total RNA of female mosquito fat bodies using Biomag Oligo d(T) magnetic beads (PerSeptive Diagnostics). RNAs were either separated by electrophoresis in 1.2% agarose formaldehyde gels and transferred to a nitrocellulose membrane or loaded onto a nitrocellulose membrane with a slot blotter. These blots were hybridized to a ^{32}P-labeled 766-bp *Hind* III-*Pst* I fragment of the VC cDNA encoding the pre-pro-portion of VC. As controls, actin (Deitsch *et al.*, 1995) and Vg (Chen *et al.*, 1994) cDNAs from *A. aegypti* were used. The conditions of hybridization at high stringency were used according to Sambrook *et al.* (1989).

Functional expression of mosquito VC and Vg - The cDNAs of VC and Vg were expressed in the TNT Coupled Reticulocyte Lysate gene expression system (Promega). The entire cDNA of VC was cloned in pBluescript (Stratagene) under the control of the T7 promoter (pVC). Due to the size limitation of this expression system, the 5'-end 3.7 kb *Xba* 1-*Bam* H1 fragment of the Vg cDNA encoding the entire small and the N-terminal half of the large subunit was subcloned to pGEM 7Z(+) (Promega) under the control of the T7 promoter (pTVg). Plasmid DNA was purified using the Wizard Minipreps DNA Purification System (Promega). One microgram of purified plasmid DNA was used for expression. For co-expression, 0.1 μg of pTVg and 1.0 μg of pVC were used for each reaction. Canine pancreas microsomal membranes (Promega) were added to monitor post-translational modification. ^{35}S-methionine was used to label the translated products. All protocols were followed as described by the manufacturer. Ten co-expression reactions were combined to increase the yield from immunoprecipitation. Monoclonal antibodies against the Vg small subunit (Raikhel *et al.*, 1986) were used to immunoprecipitate Vg using the method of Hays and Raikhel (1990). Sodium dodecyl sulfate polyacrylamide gel

electrophoresis was done to separate expressed products. Proteins were stained with Coomassie Brilliant Blue R-250 and were processed for fluorography (Hays and Raikhel, 1990).

RESULTS AND DISCUSSION

Cloning and analysis of the mosquito VC cDNA

A 537 bp cDNA fragment was amplified from total RNA prepared from fat bodies of vitellogenic female mosquitoes by reverse transcription and PCR. This fragment was sequenced. It shows high similarity to the sequence of the catalytic domains of *Drosophila* convertases dKLIP-1 (Hayflick *et al.*, 1992), dfurin1 (Roebroek *et al.*, 1991) and dfurin2 (Roebroek *et al.*, 1992). This fragment was used as a probe to screen a cDNA library constructed from the fat body of vitellogenic mosquito females. Although positive clones were not abundant, six cDNA clones were isolated from several large-scale screenings. One of the cDNA clones was sequenced from both strands. This cDNA clone is 3,735 bp long. It has a putative polyadenylation signal, AATAAA, located 24 bp upstream of the poly(A) tail (Figure 1).

To verify that this cDNA clone contains an open reading frame, it was expressed in an *in vitro* coupled transcription and translation system under control of the T7 promoter. The translated product, 115 kDa, shown in Figure 2 revealed that this cDNA clone encodes an open reading frame, with a sequence similar to convertases, containing a translation initiation methionine residue. This protein also has a signal peptide for post-translational modification because addition of canine pancreas microsomal membranes caused a shift of this translated product to 140 kDa (Figure 2).

Figure 1. Nucleotide and deduced amino acid sequences of vitellogenin convertase from the mosquito, *Aedes aegypti*. For the nucleotide sequence, the non-coding regions are shown in lower case while the coding region is capitalized. The polyadenylation signal is in lower case and double underlined. The deduced amino acid sequence of this pre-pro-enzyme is displayed in capitalized letters. The codons encoding putative translation initiation methionine residues are underlined by dots. The putative transmembrane domains are boxed. A repeat of a putative paired basic motif for autocleavage is in bold and underlined. The potential autocleavage site is indicated by an arrowhead. The cysteine residues of the cysteine-related motif are squared and in bold. The amino acids corresponding to active sites of convertases are circled and in bold.

```
gcctaatgagttcaacagccggcaaaagctgcacagaagaccaccaatgaagcttgaaga   60
cctgaacggagctgccggaagtggcggcacaggtggggaagttatgttcagtgacgtcag  120
atggacagttgaacccggtagcgttcgaagtggcgaaggtcgccgaaaaagtgccaaatg  180
tagcaatcgcttgtcgcgaaagatcgtcaagggagctgaagaagtgatggcggagagctc  240
                                      M  P  C  P  A  A  S  R  R    9
gaccgaagcattccagtcctgcccattgccgacgATGCCTTGTCCAGCGGCGTCCAGAAG  300
  L  S  G  S  G  N  S  A  E  Q  S  N  S  S  S  S  S  S  S  R   29
GTTGAGTGGGAGTGGTAACTCAGCGGAACAGTCGAACAGCAGCAGCAGCAGCAGTAGTAG  360
  R  R  S  R  S  S  S  S  S  R  L  W  C  W  H  R  S  S  P  A   49
GCGGCGATCTAGGTCGTCGTCATCCAGCAGGTTGTGGTGCTGGCACCGGTCTTCACCGGC  420
  L  V  L  Y  I  V  G  V  V  L  V  G  A  I  S  T  V  Y  C  D   69
GTTAGTGTTGTACATAGTAGGCGTAGTGTTAGTAGGTGCAATTTCCACGGTTTACTGTGA  480
  A  V  S  S  G  V  S  S  S  S  S  S  S  V  S  S  N  S  N  P   89
TGCAGTTAGTAGTGGGGTCAGCAGCAGCAGCAGTAGTAGTGTTAGTAGTAATAGCAATCC  540
  N  Q  R  G  G  D  N  V  V  V  D  G  G  Q  H  L  V  S  G  E  109
TAATCAAAGAGGTGGTGATAACGTCGTCGTGGACGGGGGGCAACATCTGGTCAGTGGCGA  600
  S  N  A  V  E  Q  D  D  A  F  G  S  G  S  V  G  A  C  E  D  129
GAGCAACGCGGTGGAACAGGACGACGCGTTCGGCAGTGGGAGTGTGGGTGCGTGCGAGGA  660
  H  Y  E  G  S  G  H  Y  T  H  H  W  A  V  H  I  P  E  G  G  149
TCATTACGAGGGCAGCGGACACTACACACACCACTGGGCGGTGCACATTCCGGAGGGCGG  720
  G  E  T  A  E  Q  V  A  D  E  H  G  F  I  N  H  G  K  I  F  169
CGGCGAAACGGCGGAGCAGGTCGCCGACGAGCACGGGTTCATCAATCATGGCAAGATATT  780
  D  G  Y  Y  H  F  E  H  R  H  L  Q  K  R  S  L  N  P  S  G  189
CGACGGGTACTACCACTTCGAGCACCGGCACCTGCAGAAGCGATCCCTTAATCCTTCGGG  840
  H  H  Q  R  R  L  D  G  D  D  R  D  R  W  A  K  Q  Q  R  A  209
GCATCACCAGCGGCGACTGGACGGTGACGATCGGGATCGATGGGCGAAACAGCAGCGAGC  900
  K  R  R  P  K  R↓ D  F  R  P  L  K  S  P  Y  T  I  Q  L  N  229
CAAACGGCGCCCGAAGCGGGACTTTCGACCCCTCAAGAGCCCCTACACGATACAGCTGAA  960
  D  P  K  W  G  E  M  W  Y  L  N  R  G  N  G  L  D  M  N  V  249
CGACCCCAAGTGGGGCGAGATGTGGTACCTGAATCGAGGAAACGGCCTAGACATGAATGT 1020
  I  P  A  C  K  E  G  V  T  G  K  G  V  V  V  T  I  L  D  D  269
GATACCAGCCTGCAAGGAAGGCGTGACCGGAAAGGGCGTCGTGGTGACGATTCTGGACGA 1080
  G  L  E  S  D  H  P  D  L  E  H  N  Y  D  P  K  A  S  Y  D  289
CGGTCTGGAATCGGACCATCCTGATCTGGAACATAACTACGACCCCAAAGCATCGTACGA 1140
  V  N  G  N  D  G  D  P  M  P  H  C  D  L  T  D  S  N  R  H  309
CGTGAACGGAAACGACGGTGACCCGATGCCACACTGCGACCTTACGGACTCGAACCGACA 1200
  G  T  R  C  A  G  E  V  A  A  T  A  N  N  S  K  C  A  V  G  329
CGGGACTCGGTGTGCCGGTGAGGTCGCTGCCACGGCTAATAATTCCAAGTGCGCCGTGGG 1260
  I  A  Y  G  A  R  V  G  G  V  R  M  L  D  G  D  V  T  D  V  349
CATTGCGTACGGAGCCCGGGTCGGTGGTGTCCGGATGCTCGACGGGGACGTTACGGACGT 1320
  V  E  A  K  S  L  G  L  N  S  Q  H  I  D  I  Y  S  A  S  W  369
GGTAGAAGCGAAATCGCTCGGTCTGAACTCGCAACATATCGACATCTACAGTGCCTCGTG 1380
  G  P  D  D  D  G  K  T  V  D  G  P  G  D  M  A  T  R  A  F  389
GGGACCGGACGACGACGGGAAAACCGTTGATGGTCCGGGGGACATGGCGACGCGGGCCTT 1440
  I  E  G  V  R  K  G  R  G  G  K  G  S  I  F  I  W  A  S  G  409
CATCGAAGGCGTCCGGAAGGGACGTGGCGGCAAGGGTTCCATTTTCATCTGGGCCTCGGG 1500
```

Figure 1 Cont'd

```
  (N) G  G  R  E  H  D  N  C  N  C  D  G  Y  T  N  S  I  W  T  429
AAACGGTGGCCGGGAACACGACAACTGCAATTGCGACGGGTACACCAACTCCATCTGGAC 1560
  L  S  I  S  S  A  S  Q  E  G  L  V  P  W  F  S  E  M  C  S  449
CCTGTCGATTTCTAGCGCCTCGCAGGAAGGATTGGTGCCTTGGTTTTCGGAAATGTGCAG 1620
  S  T  L  A  T  T  Y  S  S  G  N  T  N  E  K  Q  V  I  T  T  469
CTCTACGCTGGCGACCACTTATAGCAGCGGTAACACCAATGAAAAACAGGTGATCACAAC 1680
  D  L  H  H  S  C  T  S  S  H  T  G  T  (S) A  S  A  P  L  A  489
CGATTTACATCACTCCTGTACGTCATCGCACACGGGAACGTCGGCTTCGGCTCCACTTGC 1740
  A  G  I  A  A  L  V  L  E  A  N  P  N  L  T  W  R  D  L  Q  509
AGCCGGAATAGCTGCTCTCGTCCTGGAGGCGAATCCTAACCTCACGTGGCGAGATCTACA 1800
  H  I  V  V  R  T  A  K  P  G  N  L  K  D  P  T  W  S  K  N  529
GCACATAGTGGTCCGTACGGCTAAACCAGGCAATCTGAAGGATCCAACGTGGTCAAAAAA 1860
  G  V  G  R  R  V  S  H  S  F  G  Y  G  L  M  D  A  A  A  M  549
CGGCGTCGGTCGACGAGTATCTCACTCGTTTGGCTACGGGCTGATGGACGCCGCAGCAAT 1920
  V  K  L  A  R  T  W  K  T  V  P  E  Q  Q  I  C  E  I  N  A  569
GCACTGCAAGGGCGTCAACTATCTGGAGCATGTGCAAGCGAAAATAACACTGACCTCCCA 1980
  P  H  L  D  K  Q  I  P  P  R  T  K  V  T  L  Q  L  V  V  E  589
GGTGAAGCTGGCACGTACATGGAAAACCGTTCCGGAGCAGCAAATCTGCGAAATCAACGC 2040
  H  C  K  G  V  N  Y  L  E  H  V  Q  A  K  I  T  L  T  S  Q  609
ACCCCATCTGGACAAGCAAATTCCACCTCGCACCAAAGTCACCCTACAGCTCGTGGTGGA 2100
  R  R  G  D  I  Q  I  F  L  T  S  P  S  G  T  R  V  T  L  L  629
GCGACGGGGCGATATTCAAATATTCCTAACCTCGCCGTCCGGCACCCGAGTCACGCTGCT 2160
  T  P  R  S  H  D  L  S  R  S  G  F  N  Q  W  P  F  M  S  V  649
GACACCTAGATCCCATGATTTATCTCGTTCCGGTTTCAATCAGTGGCCTTTCATGTCGGT 2220
  H  T  W  G  E  A  P  H  G  T  W  Q  L  E  I  H  N  E  G  R  669
GCACACATGGGGCGAAGCGCCGCACGGGACCTGGCAGCTGGAGATCCACAACGAAGGCCG 2280
  L  L  A  Q  I  T  H  W  N  L  I  F  Y  G  T  E  T  P  A  Q  689
TCTATTAGCTCAAATAACGCACTGGAATCTGATCTTCTACGGTACGGAAACGCCGGCTCA 2340
  P  D  D  P  V  R  L  G  K  P  G  F  S  D  S  N  Y  G  G  E  709
ACCGGATGACCCGGTCCGACTGGGGAAGCCCGGCTTCAGTGATAGCAACTATGGCGGTGA 2400
  I  E  H  N  S  L  E  F  D  N  G  I  T  S  D  Q  W  R  N  M  729
GATCGAACACAACTCGCTGGAGTTCGACAACGGCATCACCAGCGACCAGTGGAGGAATAT 2460
  Q  Q  I  G  E  S  H  V  D  V  P  A  D  G  F  E  R  R  D  G  749
GCAGCAGATCGGTGAAAGCCACGTGGACGTCCCAGCGGACGGCTTCGAGCGACGTGACGG 2520
  L  G  C  A  A  Y  S  S  G  A  L  C  I  D  C  G  S  S  S  Y  769
CCTCGGGTGTGCTGCATATTCCTCCGGAGCCTTGTGCATAGATTGCGGCTCATCCTCATA 2580
  L  Y  K  G  R  C  Y  S  T  C  P  D  S  T  F  P  S  D  A  V  789
CTTGTACAAAGGCCGCTGCTACAGTACATGCCCGGATTCGACCTTCCCGTCGGACGCGGT 2640
  P  S  G  P  P  N  N  D  D  E  A  A  I  K  E  S  A  F  I  E  809
CCCAAGTGGGCCACCGAACAACGACGATGAAGCTGCCATTAAAGAAAGTGCCTTCATCGA 2700
  Q  Q  P  L  P  K  G  S  L  R  R  S  T  A  A  A  I  A  F  Q  829
GCAGCAACCGCTACCGAAAGGATCATTACGACGCAGCACTGCGGCAGCTATCGCTTTCCA 2760
  L  V  D  D  N  S  N  N  L  D  N  D  A  L  L  V  P  V  A  K  849
GCTGGTCGACGATAATTCCAACAATTTGGACAACGACGCACTACTGGTGCCAGTGGCAAA 2820
  P  S  P  Q  E  R  D  T  Q  P  H  E  Q  Q  E  Q  D  Q  E  S  869
ACCATCACCGCAAGAACGTGACACTCAACCTCACGAGCAGCAGGAGCAGGATCAGGAAAG 2880
```

Figure 1 Cont'd

```
  R  L  C  I  Q  C  H  P  T  C  L  K  C  F  G  P  D  E  F  E  889
CCGTTTGTGCATTCAGTGTCATCCCACGTGCCTCAAGTGTTTCGGCCCGGACGAATTTGA 2940
  C  T  E  C  Q  P  L  F  A  F  V  V  D  G  D  N  D  V  N  A  909
GTGCACCGAGTGTCAGCCATTATTTGCCTTCGTCGTCGACGGCGACAATGATGTCAATGC 3000
  S  H  P  S  S  D  G  G  L  P  N  Q  R  R  H  C  V  S  I  T  929
GTCGCATCCCAGTTCGGATGGGGGTTTGCCAAACCAACGACGTCACTGCGTCTCTATCAC 3060
  G  K  Q  R  S  G  G  T  N  K  S  R  P  N  A  T  L  S  Y  Q  949
CGGAAAACAACGCTCGGGCGGAACAAATAAATCCCGGCCTAATGCTACATTATCCTACCA 3120
  D  D  S  Q  K  S  D  K  S  W  Q  S  F  D  Y  L  I  M  I  T  969
GGACGATAGCCAAAAGTCCGATAAGAGCTGGCAATCGTTCGACTACCTCATTATGATAAC 3180
  A  V  G  A  T  L  V  V  A  F  A  V  I  Y  L  M  W  R  R  C  989
CGCCGTAGGAGCGACACTGGTAGTGGCCTTTGCGGTCATCTATCTGATGTGGCGGCGTTG 3240
  F  R  G  V  L  L  G  S  G  V  G  S  G  G  T  A  D  Y  Q  Y 1009
CTTCAGGGGTGTCCTGTTGGGTTCCGGGGTAGGTTCTGGTGGCACCGCCGATTACCAGTA 3300
  D  R  V  R  T  D  E  A  N  A  G  S  E  P  S  Y  A  E  L  V 1029
CGACCGGGTGAGAACAGACGAAGCCAATGCCGGCAGTGAACCGTCCTACGCCGAACTGGT 3360
  R  D  E  I  D  D  I  L  G  E  S  S  S  D  E  S  G  G  S  F 1049
TCGAGACGAGATTGATGACATTCTGGGCGAAAGCTCCAGCGACGAGTCGGGCGGCAGTTT 3420
  M  A  T  R  I  I  A  P  L  E  R                            1060
TATGGCCACCCGAATCATTGCACCACTGGAACGAtagagatgtggtgtcattagcatggg 3460
cccgaatcagcgaaagcaccattcgacactccatgtcatgtggcaaatggtgctaacact 3520
ccggcaaagatgggtaggagagcattaagtagtatgtagttgattggatttctagtagtt 3580
acaggtaggaaggtgattcagagaaatggatgagcacaattgaactaaaatgaaccagtg 3640
ttaagatgaagcttttaaactagtaaaataaggtaacaaacaaaaagtgattagaataaa 3700
gtagataaatattttattaaaaaaaaaaaaaaaa                          3735
```

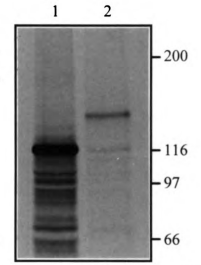


Figure 2. Analysis of ^{35}S-methionine-labeled products of *in vitro* expression of a mosquito VC cDNA by a coupled transcription and translation system. Lane 1, total translated products; lane 2, processing of the translated polypeptide in the presence of canine pancreatic microsomal membranes. Translation products were analyzed by SDS-PAGE in 7.0 % gels under reducing conditions and fluorography. The molecular mass standards in order of decreasing molecular weight (in kDa) were phosphorylase B, bovine serum albumin, ovalbumin, carbonic anhydrase, soybean trypsin inhibitor and lysozyme.

Features of the deduced amino acid sequence of mosquito VC

There are four methionine residues identified at the beginning of the open reading frame of VC (Figure 1). These putative translation initiation sites show limited similarity to the optimal context of the eukaryotic translation initiation site (Kozak, 1986) and to the *Drosophila* consensus translation start site (Cavener, 1987). To determine the translation initiation site of VC, molecular weights of translated products from these four methionine residues were estimated and compared with the size of the *in vitro* translated product without post-translational modification. The fourth methionine residue was determined to be the initiation residue because the predicted size of the protein translated from this residue is similar to that of the protein from the *in vitro* gene expression shown in Figure 2.

The deduced amino acid sequence of VC exhibits high similarity to other members of the eukaryotic convertase family (Figure 3). Sequence alignment indicates that all four active sites of convertases are conserved in VC at Asp^{268}, His^{309}, Asn^{410}, and Ser^{483} (Figure 1). This suggests that VC may form a structure similar to the model of the catalytic domain of human furin constructed by Siezen *et al.* (1994). VC as well as *Drosophila* convertases have the primary structure of the furin-like convertase including a prepropeptide, a catalytic domain, a cysteine-rich domain, a C-terminal transmembrane domain and a cytoplasmic domain (Figure 3).

Hydropathy analysis (Kyte and Doolittle, 1982) of the VC deduced amino acid sequence shows that VC does not have a hydrophobic signal peptide following the putative initiation methionine residue as most convertases do. However, there is an internal hydrophobic domain identified near the N-terminus between residues 48 and 66 (Figure 1). This is similar to *Drosophila* convertases dKLIP-1 (Hayflick *et al.*, 1992), dfurin1 (Roebroek *et al.*, 1991) and dfurin2 (Roebroek *et al.*, 1992). Such internal hydrophobic domains still function

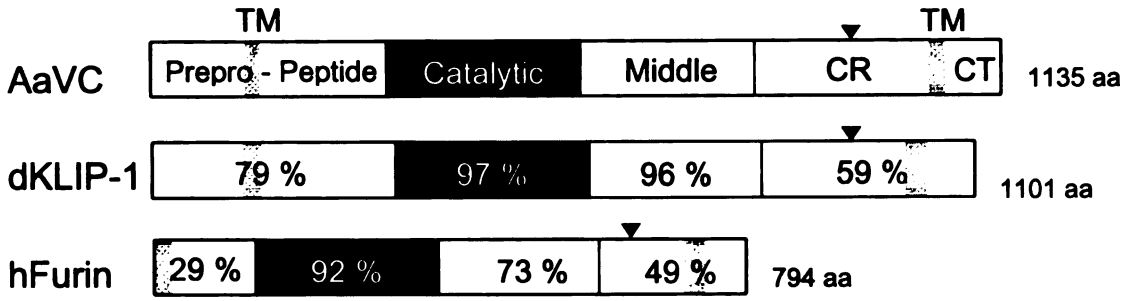


Figure 3. Schematic representation of mosquito VC, Drosophila dKLIP-1, and human furin convertases. Amino acid sequences were aligned on the basis of the maximal amino acid similarity. The percentage shown on each domain of dKLIP-1 and furin stands for the similarity of these domains to mosquito VC. A triangle indicates the location of the cysteine-related motif. Abbreviations used in this figure are: CRR, cysteine-rich region; CT, cytoplasmic domain, and TM, transmembrane domain.

as signal peptides because mosquito VC (Figure 2) as well as *Drosophila* dKLIP-1 (Hayflick *et al.*, 1992) have been demonstrated to be post-translationally modified. Molloy *et al.* (1994) demonstrated that human furin is targeted to the rough endoplasmic reticulum and eventually localized in the *trans*-Golgi network (TGN). Its C-terminal hydrophobic domain serves as a retention signal and an anchor for the TGN. The C-terminal transmembrane domain of VC is located between residues 966 and 986 (Figure 1).

Roebroek *et al.* (1992) suggested that convertases, found both in vertebrates and invertebrates, are likely able to autoactivate themselves because they have the conserved putative recognition motifs of convertases located at the junction between their pro-peptide and enzyme regions. The autoactivation motif of mosquito VC, however, is unique in that it is composed of double repeats, $RAKR^{211}$-$RPKR^{215}$ (Figure 1).

Most convertases have a cysteine-rich domain containing one or more cysteine-related motifs, $CX_2CX_3CX_2CX_{5-7}CX_2CX_{10-15}CX_{3-5}C$ (Nakagawa *et al.*, 1993; Roebroek *et al.*, 1992; Smit *et al.*, 1994). Mosquito VC also has such a motif in its cysteine-rich domain, however, this mosquito motif lacks the last two cysteine residues (Figure 1). Hatsuzawa *et al.* (1992) concluded that the cysteine-rich domain containing the cysteine-related motif is not required for enzyme activity of furin. Although Smit *et al.* (1994) suggested that the cysteine-related motif may play a role in protein targeting or protein stabilization, the function of this motif remains unclear (Roebroek *et al.*, 1992).

Tissue distribution and expression of the mosquito VC gene

A cDNA fragment encoding the prepro-peptide region of mosquito VC was labeled as a probe to hybridize to mRNA prepared from heads, thoraxes, midguts, ovaries, and pre-vitellogenic and vitellogenic fat bodies of female mosquitoes. The size of the VC mRNA is 4.2 kb. It is expressed exclusively in fat bodies of vitellogenic adult female mosquitoes (Figure 4).

Study of the kinetics of gene expression of mosquito VC indicated that VC is expressed during early vitellogenesis. The levels of VC transcript increases in accordance with that of Vg, but decreases to background levels before 24 h post blood meal (Figure 5), the peak of Vg gene expression. The kinetics of VC gene expression also matches the observation that pro-Vg is only identified at the early stage of vitellogenesis (Dhadialla and Raikhel, 1990). Although expression of the VC gene drops before Vg mRNA peaks (Figure 5), enough VC may accumulate for cleaving pro-Vg throughout vitellogenesis because the production peak of ribosomes is between 16-18 hours post blood meal (Hotchkin and Fallon, 1987).

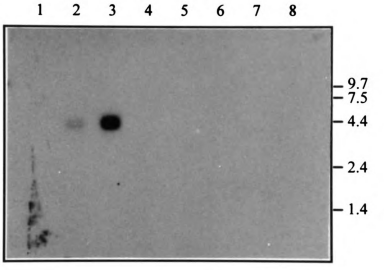


Figure 4. Northern blot analysis of gene expression of mosquito VC. The mRNA prepared from different tissues and stages of female mosquitoes were separated by a 1.2% agarose gel and transferred to a nitrocellulose membrane. This membrane was hybridized with a ^{32}P-labeled 766-bp *Hind* III-*Pst* I fragment of the VC cDNA, encoding the pre-pro-portion of VC. Lane 1, pre-vitellogenic fat body; Lane 2, vitellogenic fat body, 6 h post blood meal (PBM); Lane 3, vitellogenic fat body, 12 h PBM; Lane 4, vitellogenic fat body, 24 h PBM; Lanes 5-8: tissues of female mosquitoes 12 h PBM; Lane 5, heads; Lane 6, thoraxes; Lane 7, midguts; Lane 8, ovaries.

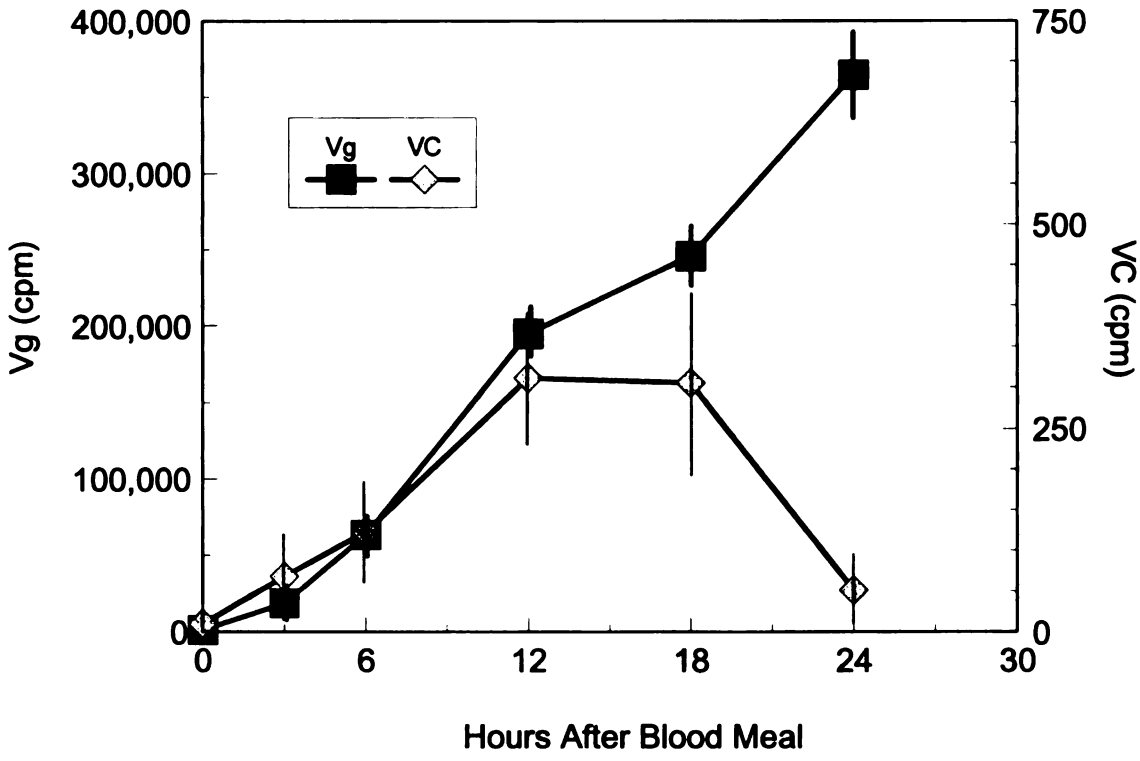


Figure 5. Kinetics of gene expression of mosquito VC and Vg. Slot blot analysis was used to determine mRNAs of VC, Vg and actin transcribed at different stages before and after a blood meal. The blots were hybridized with ^{32}P-labeled probes of VC, Vg and actin. Hybridization was quantified using a liquid scintillation counter. Count from VC and Vg were normalized using actin as an internal control. Vertical lines indicate the standard error of the mean of three repetitions.

Functional expression of mosquito VC

To demonstrate that VC is a pro-Vg cleaving enzyme, a cDNA encoding mosquito Vg, the putative substrate of VC, was initially expressed in a baculovirus gene expression system. However, most of the over-expressed products were degraded in the host insect cells (data not shown). Alternatively, both the Vg and VC cDNAs were co-expressed in an *in vitro* coupled transcription and translation gene expression system. Preliminary experiments showed that the entire Vg cDNA was too large to be expressed sufficiently. To improve Vg expression, pTVg, a truncated 3.7 kb mosquito Vg cDNA, was used. This pTVg contains a cDNA fragment encoding the Vg signal peptide, the small subunit, the cleavage site and N-terminal half of the large subunit of mosquito Vg. It can be expressed by a coupled transcription and translation system and modified in the presence of canine microsomal membranes. The size of the TVg protein is 139 kDa. It matches the molecular weight predicted from the deduced amino acid sequence. This protein becomes 160 kDa after post-translational modification with canine microsomal membranes (Figure 6). These translated Vg products were verified by immunoprecipitation using monoclonal antibodies specific to the small subunit of mosquito Vg (Figure 6).

The small subunit of mosquito Vg, 66 kDa, which is not present in the expression reaction of TVg alone (Figure 6, Lanes 1 and 2), was immunoprecipitated from the co-expression products of TVg and VC (Lane 3 of Figure 6, Lane 3). This indicates that VC is functional in cleaving TVg (Figure 6). Such data is in accordance with the observation that furin-like convertases recognize and cleave the RX(R/K)R motif (Hosaka *et al.*, 1991; Hatsuzawa *et al.*, 1992) because mosquito Vg, the substrate of VC, has an RXRR motif at its cleavage site (Chen *et al.*, 1994).

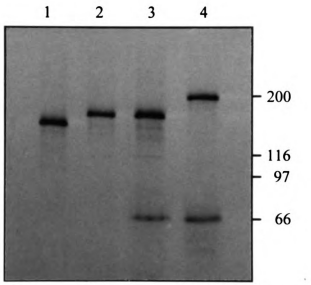


Figure 6. *In vitro* proteolytic processing of pro-Vg by VC. The cDNA of pTVg was expressed in a coupled transcription and translation system. To demonstrate the proteolytic processing activity of VC, pTVg was co-expressed with pVC. The expression products were immunoprecipitated by monoclonal antibody specific to the small subunit of mosquito Vg and analyzed by SDS-PAGE. Lane 1, pTVg; Lane 2, pTVg with canine microsomal membranes; Lane 3, pTVg and pVC with canine microsomal membranes; Lane 4, mosquito Vg secreted from the cultured fat body. Markers used are as in Figure 2.

Although the result shown in Figure 6 has demonstrated that VC is a functional convertase and can catalyze pro-Vg subunit cleavage, the activity of VC was not high. This may be because the truncated Vg did not form a recognizable configuration for VC. It is also possible that *in vitro* expressed mosquito VC does not have a conformational properties identical to endogenous VC. Scheele (1984) pointed out that difficulty in demonstrating significant biological activity of an *in vitro* translated protein in canine microsomal vesicles may be due to formation of inappropriate sets of disulfide bonds.

Mosquito VC is functionally different from *Drosophila* convertases

In *Drosophila*, dKLIP-1 (Hayflick *et al.*, 1992) and dfurin2 (Roebroek *et al.*, 1992, 1995) have been functionally expressed and shown to cleave murine pro-β-nervous growth factor and pro-von Willebrand factor, respectively. However, substrates of *Drosophila* convertases have not been identified. Although *Drosophila* convertases are very similar to mosquito VC in sequence and domain structures, these convertases must have substrates other than insect pro-Vgs because the higher dipteran insects like *Drosophila* do not have Vg genes (Byrne *et al.*, 1989). These insects, instead, produce small yolk proteins. These yolk proteins are not cleaved (Byrne *et al.*, 1989) and do not share any sequence similarity to insect Vgs (Chen *et al.*, 1994).

Dfurin1 was localized in the embryo (Roebroek *et al.*, 1991), while Dfurin2 was identified in all stages (Roebroek *et al.*, 1992). Another *Drosophila* convertase, dKLIP-1, was detected in adult cortical regions of the central nervous system and the fat body. Both dKLIP-1 and dfurin2 transcripts are also found in growing oocytes. They are temporally and spatially expressed during embryogenesis (Hayflick *et al.*, 1992; Roebroek *et al.*, 1995). Based on its tissue-specific expression, mosquito VC obviously has a biological function different

than that of *Drosophila* convertases. It seems that these *Drosophila* convertases play roles in embryo development and the central nervous system. Mosquitoes may have additional convertases with functions similar to these *Drosophila* convertases.

Discovery of mosquito VC is remarkable in filling the gap in the pathway of insect Vg biosynthesis because this long-sought-after enzyme was difficult to identify. This work is also significant in that members of the convertase family were unknown previously for processing insect pro-Vgs. Conservation of the paired basic motif at the cleavage site of insect pro-Vgs suggests that insect VCs co-evolved with insect Vgs. Characterization of this novel convertase reveals that these enzymes are applied by nature for more diverse biological functions than those that currently acknowledged.

ACKNOWLEDGMENT

We thank Drs. G. Tomas and W. J. M. Van de Ven for their generous gifts of *Drosophila* convertase probes, Dr. S. M. Thiem for advice in baculovirus expression system, Drs. A. Biran and K. W. Deitsch for critical suggestions to this manuscript. This work was supported by NIH grant AI-24716 to ASR.

CHAPTER 5*

MOSQUITO CATHEPSIN B-LIKE THIOL PROTEASE INVOLVED IN EMBRYONIC DEGRADATION OF VITELLIN IS PRODUCED AS A LATENT EXTRAOVARIAN PRECURSOR

*The research described in this chapter was done in collaboration with Alan R. Hays and Ekaterina S. Snigirevskaya

ABSTRACT

A novel member of the thiol protease superfamily has been identified in the yellow fever mosquito, *Aedes aegypti*. It is synthesized and secreted as a latent proenzyme in a sex-, stage- and tissue-specific manner by the fat body, an insect metabolic tissue, of female mosquitoes during vitellogenesis in response to blood feeding. The secreted, hemolymph form of the enzyme is a large molecule, likely a hexamer, consisting of 44-kDa subunits. The deduced amino acid sequence of this 44-kDa precursor shares high similarity with cathepsin B but not with other mammalian cathepsins. We have named this mosquito enzyme vitellogenic thiol protease (VTP). The size of VTP decreases to 42 kDa after internalization by oocytes where it is localized in organelles of the endocytotic pathway. In mature yolk bodies it is located in the matrix surrounding the crystalline yolk protein, vitellin. At the onset of embryogenesis VTP is further processed to 33 kDa. Embryonic extract containing the 33-kDa VTP was active toward benzoyloxycarbonyl-Arg-Arg-para-nitroanilide, a cathepsin B-specific substrate, and degraded vitellogenin, the vitellin precursor. Both of these enzymatic activities were prevented by E-64, a thiol protease inhibitor.

INTRODUCTION

Cathepsin B is a thiol (cysteine) protease with both endopeptidase and peptidyldipeptidase activities. Due to its broad specificity, cathepsin B plays a key role in intracellular protein catabolism in the lysosomal system (Bond and Butler, 1987). Cathepsin B is well characterized both enzymatically and molecularly (Takio *et al.*, 1983; Chan *et al.*, 1986; Musil *et al.*, 1991; Béchet *et al.*, 1991; Qian *et al.*, 1991; Rowan *et al.*, 1992; Mach *et al.*, 1994a). The

mammalian cathepsin B is implicated in tumor invasion and metastasis (Poole *et al.*, 1978; Sloane *et al.*, 1981). Tumor-specific cathepsin B is secreted by malignant cells as a latent high molecular weight precursor which is presumably activated at cell contacts (Mort and Recklies, 1986; Sloane *et al.*, 1990).

In addition, cathepsins B, as well as the related cathepsins L, have been identified in numerous parasitic protozoa and helminths, including important pathogens of human and domestic animals (Cazzulo *et al*, 1989; Rosenthal *et al.*, 1989; Klinkert *et al.,* 1989; Cox *et al.,* 1990; Pratt *et al.,* 1990, 1992; Merckelbach *et al.*, unpublished, GenBank accession number X70968, 1993; Robertson and Coombs, 1993). In the blood-sucking bug, *Rhodnius prolixus*, cathepsin B is the major gut proteolytic enzyme (Terra *et al.*, 1988). In all these organisms, cathepsins B and L are believed to be involved in degradation of host hemoglobin.

In insects and other arthropods, cathepsins B and L also are involved in key developmental processes. In the flesh fly, *Sarcophaga peregrina*, hemocytes produce the extracellular form of a cathepsin B-like enzyme which participates in decomposition of the larval fat body during metamorphosis (Kurata *et al.*, 1992a,b; Takahashi *et al.*, 1993). Cathepsins B and L have been implicated in degradation of yolk proteins during embryonic development (Medina *et al.*, 1988; Medina and Vallejo, 1989; Fagotto, 1990a,b; Takahashi *et al.*, 1993; Yamamoto and Takahashi, 1993; Yamamoto *et al.*, 1994b).

Understanding of developmental mechanisms in the mosquito is important because this insect transmits the most devastating of vector-borne human diseases including malaria, lymphatic filariasis, Dengue fever and many others. Little is known, however, about the process of yolk protein degradation in the mosquito embryo. We found previously that during vitellogenesis, the female fat body, a metabolic tissue analogous to the vertebrate liver, synthesizes and

secretes a latent proenzyme of a serine carboxypeptidase which is homologous to yeast carboxypeptidase Y (Hays and Raikhel, 1990; Cho *et al.*, 1991). This 53-kDa proenzyme, named vitellogenic carboxypeptidase (VCP), is specifically accumulated by developing oocytes and deposited in yolk bodies. Although we demonstrated that VCP is activated during embryogenesis, its role remains unknown.

In this paper we report the discovery of an unusual cathepsin B-like thiol protease from the mosquito, *Aedes aegypti*. Similar to VCP, it is produced by the fat body of vitellogenic female mosquitoes. Its synthesis in the fat body is initiated by blood feeding. Cloning and sequencing of the cDNA encoding this mosquito protein revealed a high similarity to vertebrate cathepsin B. Due to its unusual nature, we have named this mosquito enzyme vitellogenic thiol protease (VTP). We report here that the mosquito cathepsin B-like enzyme, VTP, is secreted by the fat body as a latent proenzyme of a size similar to that of the latent tumor cathepsin B (44-kDa). It is accumulated by developing oocytes where it is stored in yolk bodies. At the onset of embryogenesis, VTP is processed to 33 kDa, a size similar to that of the activated single-chain mammalian cathepsin B. Furthermore, we demonstrate that embryo extract containing the 33-kDa VTP was active against benzoyloxycarbonyl-Arg-Arg-para-nitroanilide (Z-Arg-Arg-*p*NA), a cathepsin B-specific substrate, and degraded vitellogenin (Vg), the yolk protein precursor. Both of these enzymatic activities were prevented by E-64, a thiol protease inhibitor.

MATERIALS AND METHODS

Reagents - All analytical grade chemicals and protease inhibitors were purchased from Sigma Chemical and Calbiochem, respectively, unless stated

otherwise. [^{32}P]-dATP (3,000 Ci/mmole) and [^{35}S]-dATP were obtained from DuPont NEN, while [^{35}S]-methionine (1,120 Ci/mmole) was from ICN Radiochemicals. DEAE-Sepharose was purchased from Pharmacia. Bio-Rad was the source for protein assay reagents, econo-Pac desalting columns and molecular weight standards for electrophoresis. Safety Solve II scintillation cocktail was supplied by Research Product International. Paraformaldehyde, glutaraldehyde and LR White were obtained from Polyscience, and protein-A-colloidal gold conjugates from E-Y Laboratories.

Animals - Mosquitoes, *Aedes aegypti*, were reared as previously described (Hays and Raikhel, 1990). Three to five days after eclosion adult females were fed on white rats to initiate vitellogenesis. Mosquitoes were dissected in TES buffered *Aedes* physiological saline (TAPS) (Hagedorn *et al.*, 1977) at room temperature.

Protein Preparation and Electrophoresis - Fat body, ovarian and embryonic proteins were prepared as previously described (Hays and Raikhel, 1990) in B3 buffer containing 25 mM (2-((tris-(hydroxymethyl)-methyl)-amino)-ethanesulfonic acid) (TES), pH 7.5, 150 mM NaCl, 10 % glycerol, 10 mM ethylene-diamine-tetraacetic acid (EDTA). Unless otherwise noted, all solutions contained the following inhibitors: 1 mM 4-(2-aminoethyl)-benzenesulfonylfluoride, HCl (AEBSF) (Calbiochem), 1 mM benzyl sulfonyl fluoride (PMSF) (Boehringer-Mannheim), 5 mM ε-amino-*n*-caproic acid (ACA), 1 mM benzamidine (BAM); 10 mM EDTA; 10 μg/ml aprotinin, and 2 μg/ml each antipain, leupeptin, pepstatin (Boehringer-Mannheim) and chymostatin. Fat body-secreted proteins were collected in TAPS supplemented with aprotinin, to which the remainder of the inhibitors were added upon harvesting.

Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) was done using either a 10 % or 12 % straight gel by the method of Laemmli (1970) in vertical slab gels. Proteins were visualized by staining with either Coomassie Brilliant Blue R-250 or silver, or were processed for fluorography.

Non-reducing native polyacrylamide gel electrophoresis was performed using 3.5-25% acrylamide gradient gels crosslinked with 2.6 % methylene bis-acrylamide (Ornstein, 1964; Davis, 1964; Sappington *et al.*, 1995). The native PAGE sample buffer contained 125 mM Tris-HCl, pH 6.8, 12 % glycerol and 15 mM Bromophenol Blue. Gels were run for 18-21 h at 125 volts at 4°C.

***In vitro* Protein Labeling** - *In vitro* culture of fat body and ovaries was as previously described (Hays and Raikhel, 1990). Synthesized and secreted fat body proteins were radiolabeled by incubating 3 female or 6 male fat bodies at 27°C in 50 ml of TES buffered culture media modified (Dhadialla and Raikhel, 1990) to contain the appropriate concentration of [^{35}S]-methionine (pulse media). If only pulse media was needed, fat bodies were incubated for 3 h.

Ovary-synthesized proteins were labeled *in vitro* by placing five pairs of ovaries (18 h post blood meal (PBM)) into 100 ml TES buffered culture media (Koller *et al.*, 1989) with [^{35}S]-methionine replacing unlabeled methionine. The ovaries were removed after 2.5 h, rinsed 3 times in APS and frozen at -80°C.

Antibody Production - Anti-VTP polyclonal antibodies were produced in female New Zealand white rabbits. Ovaries dissected from *A. aegypti* 28 h after blood feeding were processed and subjected to DEAE-Sepharose anion-exchange chromatography (Koller *et al.*, 1989). Proteins in the Vn-free unbound fraction were separated by preparative SDS-PAGE. A band containing VTP (42 kDa) was excised and electroeluted using an ISCO model 1750 Electrophoretic

Concentrator according to recommendations of the manufacturer. The elute was used as the source of antigen to produce antibodies to this peptide (VTP pAB). Rabbits were injected with antigen/adjuvant (TiterMax, CytRx) two times at a one month interval. Serum proteins from the immunized rabbit were precipitated with ammonium sulfate at 35% of saturation and reconstituted at 5 mg/ml. Polyclonal antibodies against VCP and the Vg large subunit were already available (Hays and Raikhel, 1990).

Amino acid sequence analysis - The proteins, prepared as described above for antibody production, were electroblotted onto PVDF membrane (0.2 μm pore size; Bio-Rad). The 42-kDa band was excised and sent to Harvard Microchem (Harvard University) for tryptic digestion and amino acid sequence analysis.

Immunoblots and Immunoprecipitation - Proteins from SDS-PAGE and native PAGE gels were transferred to a nitrocellulose membrane (Schleicher & Schuell) as previously described (Sappington *et al.*, 1995) and detected with the ECL Western blotting detection system (Amersham). Prior to immunoprecipitation, Vn or Vg were partially removed from ovarian extracts or fat body secretions, respectively, using a suspension of DEAE-Sepharose CL-6B as described previously by Hays and Raikhel (1990). After removal of Vn or Vg, VTP was immunoprecipitated with anti-VTP antibodies. Protein A-Sepharose was added as a precipitating agent (Hays and Raikhel, 1990). After immunoprecipitation, the resulting pellet was washed and used in radioimmunoassay or SDS-PAGE and fluorography experiments (Hays and Raikhel, 1990).

Cloning of VTP cDNA - The probe used for screening a cDNA library for VTP was generated by PCR. RNA was isolated from fat bodies of 24 h post-blood-

fed female mosquitoes. cDNA was synthesized from the fat body mRNA and used as the template for PCR. An antisense degenerate primer was constructed from a portion of an internal peptide sequence, IMEEIFIN, obtained from microsequencing a tryptic fragment of the 42-kDa ovarian VTP. This primer had an *Xba* I adapter site at its 5'-end. The sequence of the first primer was: GTTCTAGAT(ACT)ATGGA(AG)GA(AG)AT(ACT)TT(CT)AT(ACT)AA(CT)G. The second primer was a 17-mer d(T) with an *Xba* I site located at its 5'-end. The resulting PCR product was used to screen a mosquito vitellogenic fat body cDNA library which was constructed at the *Eco* RI site of the Lambda-Zap II vector (Stratagene). The conditions for reverse transcription, PCR amplification and cDNA cloning were as in Chen *et al.* (1994). The method of enzymatic sequencing utilized was that reported by Sanger *et al.* (1977).

Isolation of RNA and Northern Hybridization - Total RNA was prepared using the guanidine isothiocyanate method (Chirgwin *et al.*, 1979; Chomczynski and Sacchi, 1987; Dhadialla and Raikhel, 1990). For Northern blot analyses, total RNA and RNA markers (Life Technologies) was separated by electrophoresis in 1.2% agarose formaldehyde gels. The RNA was transferred to a nitrocellulose membrane and hybridized to ^{32}P-labeled probes. The conditions of hybridization at high stringency were according to Sambrook *et al.* (1989). A 255 bp *Eco* RI-*Hinc* II fragment that is located at the 5'-end of the VTP cDNA was used as a probe. As a control for RNA loading, cytoplasmic actin gene of *A. aegypti* was used to probe the mosquito actin mRNA. Cloning of mosquito actin cDNA was previously described (Deitsch *et al.*, 1995).

***In vitro* transcription and translation** - For *in vitro* transcription and translation, the entire cDNA of VTP was subcloned into pGEM 7Z(+) (Promega) under the

control of the SP6 promoter. Plasmid DNA was purified by the Wizard Minipreps DNA Purification System (Promega). Two micrograms of purified plasmid DNA were used for transcription and translation in the SP6 TNT Coupled Reticulocyte Lysate system (Promega). Canine microsomal membranes (Promega) were added to the same system to monitor post-translational modification. ^{35}S-labeled methionine was used to label the translated product. All protocols were followed as described by the manufacturer.

DNA Analysis and Alignment - The putative signal peptide in the deduced amino acid sequence was determined according to Kyte and Doolittle (1982). The deduced amino acid sequence was analyzed by FastA, Motifs and Gap programs of Genetics Computer Group software (University of Wisconsin, WI).

Gold Immunolabeling - For immunocytochemical examination, tissues were fixed with 2% formaldehyde and 0.1% glutaraldehyde in 0.05M Na-phosphate buffer, pH 7.4 for 45-60 min at room temperature. The specimens were then blocked with 100 mM glycine in the same buffer for 1 h at room temperature. After dehydration in ethanol, the tissues were embedded in LR-White resin. The ultra-thin sections were prepared using an Ultracut E microtome (Reichert-Jung) and were mounted on formvar-coated or uncoated nickel grids. The labeling was carried out on drops of reagents as previously described (Raikhel, 1987). Double labeling was performed according to the two-side method (Bendayan, 1982). The controls included substitution of primary antibody by non-immune rabbit IgG and the omission of primary antibodies followed by normal staining procedures. After immunolabeling, the sections were stained with uranyl acetate and lead citrate and examined with a JEOL 100CX II transmission electron microscope.

Enzyme Assay - All enzymatic reactions were performed in duplicate in 100 mM MES (2-(N-morpholino)-ethanesulfonic acid) buffer (pH 5.5) containing 1 mM PMSF, 1 mM AEBSF, 5 μM pepstatin and 2.5 mM EDTA in duplicate. A mixture of embryos from 12-48 h after egg deposition were extracted in B3 buffer with 1 mM PMSF, 1 mM AEBSF and 5 μM pepstatin. Each reaction received 38 μg total protein of this extract. Controls were done in the above mixture only. Treatment received either chymostatin or E-64 (*trans*-epoxysuccinly-L-leucylamido-(4-guanidino)butane) in addition to the above inhibitors at the following concentrations: 0.3, 0.6, 1.2, or 2.5 mM and preincubated with the controls for 20 min at 37°C. Enzymatic assays were initiated by the addition to the reaction mixture of 81.1 mM Z-Arg-Arg-*p*NA (benzoyloxycarbonyl-Arg-Arg-*para*-nitroanilide) (Bachem Biosciences) in 100 % methanol to yield a final concentration of 1.6 mM substrate and 2 % methanol. Data was collected by measuring the absorbance at 405 nm with a spectrophotometer after 2.5 h. A standard curve of para-nitroaniline (*p*NA) was done and regressed by the linear least squares method. This curve was used to calculate the amount of *p*NA liberated by enzymatic activity. In experiments to determine the pH optimum of the enzyme, a range of pH from 3.5 to 7.5 was used.

RESULTS

Identification of a novel yolk protein precursor from the mosquito fat body

Analysis of proteins which are synthesized and secreted by the vitellogenic mosquito fat body revealed a polypeptide with M_r = 44,000 in addition to vitellogenin (Vg) and vitellogenic carboxypeptidase (VCP) (Hays and Raikhel, 1990). A polypeptide of 42-kDa was present among proteins extracted from vitellogenic ovaries or newly-laid eggs (Hays and Raikhel, 1990). We

produced polyclonal antibodies against the 42-kDa polypeptide gel-purified from the ovarian source. Immunoblot analysis showed that in addition to the 42-kDa ovarian polypeptide, these antibodies recognized a single polypeptide with M_r = 44,000 from vitellogenic fat bodies and their secretions, thus confirming the immunological identity of these two polypeptides. The immunoreactive polypeptides were not present in previtellogenic female fat bodies or males (Figure 1). Immunoprecipitation of radiolabeled proteins with anti-42-kDa antibodies demonstrated that this polypeptide is produced exclusively by the fat bodies of vitellogenic females (not shown).

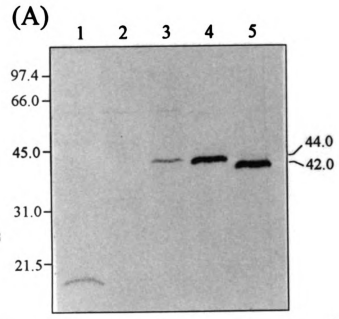
Amino acid sequences, obtained by microsequencing tryptic fragments of the 42-kDa polypeptide, were analyzed for protein similarity. A computer search showed that a 25 amino acid fragment shared 79% identity with human cathepsin B (not shown).

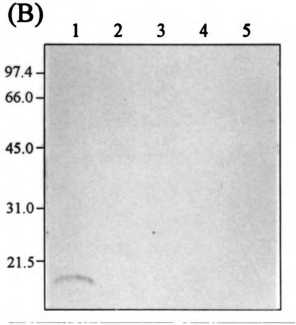
Cloning and analysis of the cDNA encoding the 44-kDa yolk protein precursor, vitellogenic thiol protease.

A 398 bp cDNA fragment was obtained from amplification of vitellogenic fat body RNA by the polymerase chain reaction technique (PCR) using degenerate primers derived from the 42-kDa polypeptide microsequences. Both ends of the fragment matched the PCR primer sequences, and an open reading frame showed high similarity to vertebrate cathepsin B (not shown). This PCR fragment was subcloned into pUC119 at the *Xba* I site and was used as a probe to screen the cDNA library prepared from fat bodies of vitellogenic female mosquitoes.

The twenty-five longest cDNA clones were chosen from numerous positive clones. All of these cDNA clones were about 1.1 kb in size and had identical restriction maps. One cDNA clone of 1,111 bp was sequenced from

Figure 1. Presence of vitellogenic thiol protease (VTP) in tissues of the mosquito. A. Immunoblot analysis of mosquito tissue extracts using polyclonal antibodies to the ovarian 42-kDa form of VTP. B. The same using pre-immune serum. Lane 1, whole male; lane 2, previtellogenic female fat body; lane 3, vitellogenic female fat body, 20 h PBM; lane 4, *in vitro* secretion from vitellogenic fat bodies, 20 h PBM; lane 5, ovary, 24 h PBM. SDS-PAGE using a 12 % gel was performed under reducing conditions. The molecular standards in order of decreasing molecular weight (in kDa) were phosphorylase B, bovine serum albumin, ovalbumin, carbonic anhydrase, soybean trypsin inhibitor and lysozyme.


(A)
1
2
3
4
5
97.4
66.0
45.0
31.0
21.5
44.0
42.0


(B)
1
2
3
4
5
97.4
66.0
45.0
31.0
21.5

both strands. This cDNA clone has an open reading frame of 1,026 nucleotides with a putative polyadenylation signal, aataaa, located 14 bp upstream of the poly(A) tail (Figure 2). It encodes a pre-pro-protein of 342 amino acid residues with a predicted M_r of 37,900. The two sequences determined by direct microsequencing of the ovarian 42-kDa VTP are both found in its deduced amino acid sequence (Figure 2). Hydropathy analysis (Kyte and Doolittle, 1982) predicts that the protein is hydrophilic with a strongly hydrophobic putative signal peptide of 19 residues led by a methionine residue at its N-terminus (Figure 2). It is a positively charged protein with a predicted isoelectric point (pI) of 7.64. However, there are three potential N-linked glycosylation sites (Marshall, 1972) and six potential phosphorylation sites (Kuenzel *et al.*, 1987) and the fully processed protein could carry a significantly different net charge (Figure 2).

To verify the identity of the 1.1-kb cDNA clone, it was expressed using a coupled *in vitro* transcription/translation system (Figure 3). A single polypeptide, 38 kDa, was immunoprecipitated from the translation reaction by the anti-VTP antibodies. Furthermore, the size of the expressed polypeptide increased from 38 to 44 kDa in the presence of canine microsomal membranes, a size identical to that of the 44-kDa VTP secreted by the fat body. The 1.1-kb cDNA clone was also used as a probe for Northern hybridization. It hybridized strongly only to 1.1-kb mRNA from the fat body of vitellogenic female mosquitoes (see below). Taken together, these data confirmed that the 1.1 kb cDNA encodes the 44-kDa fat body precursor of the ovarian 42-kDa VTP.

Mosquito 44-kDa yolk protein precursor (VTP) is a proenzyme homologous to cathepsin B

The deduced amino acid sequence of VTP exhibits high similarity to the family of eukaryotic thiol proteases (EC 3.4.22.-) (Dufour, 1988). It shows a

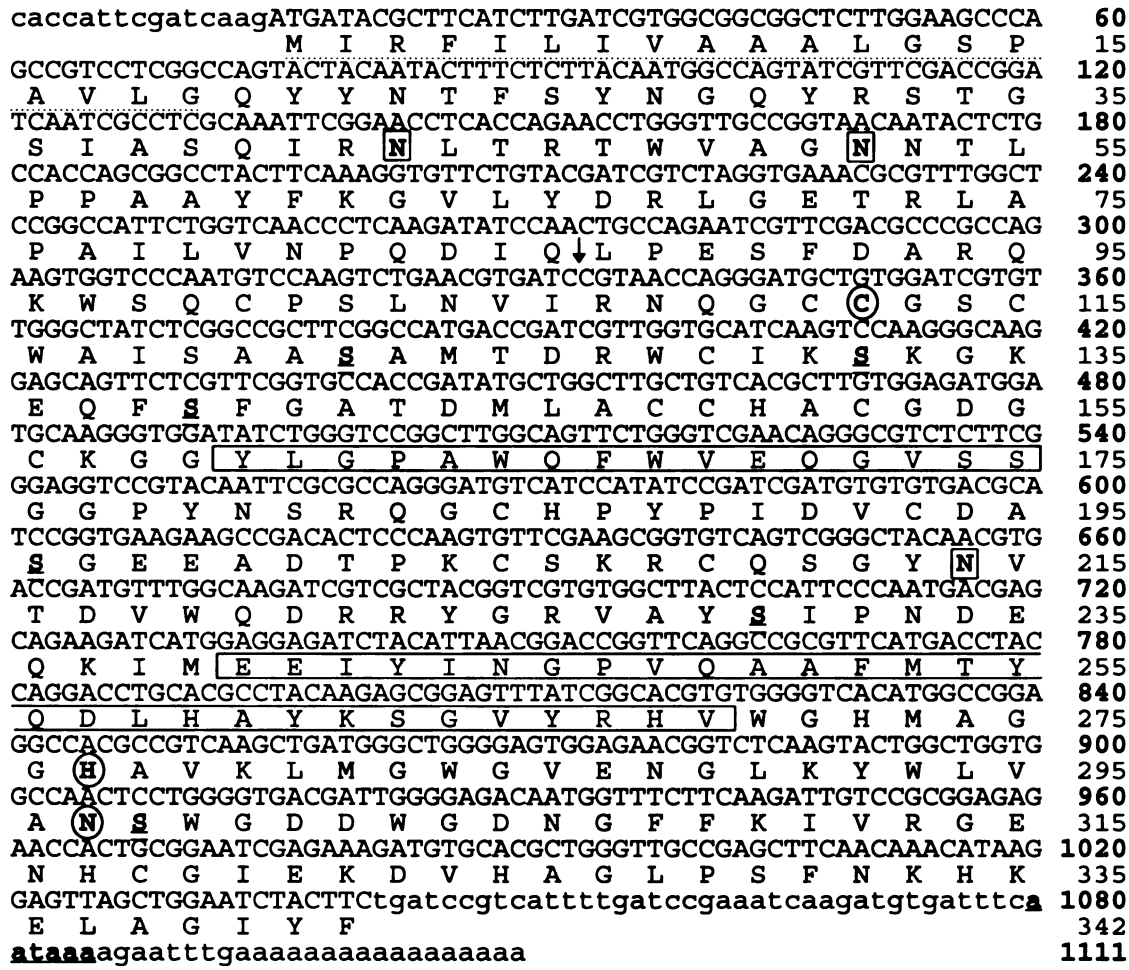


Figure 2. Nucleotide and deduced amino acid sequences of vitellogenic thiol protease from the mosquito, *Aedes aegypti*. For nucleotide sequence, the non-coding regions are shown in lower case while the coding region is capitalized. The polyadenylation signal is in lower case, bold print and underlined. The deduced amino acid sequence of the pre-pro-enzyme is displayed in capitalized letters. Both nucleotides and deduced amino acids are numbered with nucleotides in bold. A putative signal peptide is marked by a dotted underline. The deduced amino acid sequences matching those obtained by direct peptide sequencing are boxed. A potential pro-enzyme cleavage site is indicated by an arrowhead. Potential N-linked glycosylation sites (N) are marked by squares and are in bold. Potential phosphorylated serines (S) are marked by a double underline and are in bold. The amino acids corresponding to active sites of cathepsin B are circled and bold.

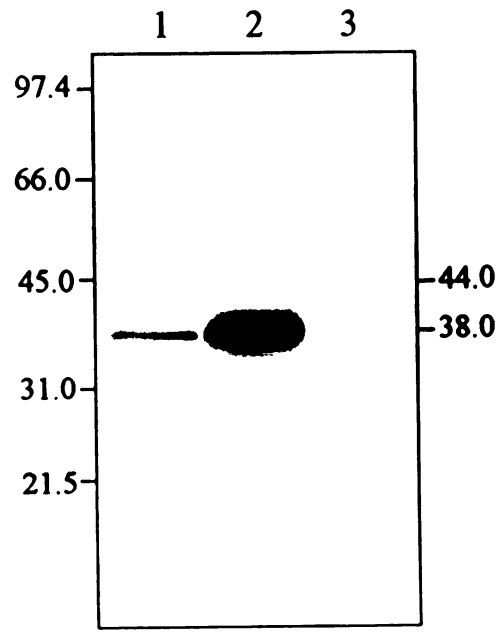


Figure 3. Analysis of ^{35}S-methionine-labeled products of *in vitro* expression of a mosquito VTP cDNA by a coupled transcription and translation system (Promega). Lane 1, immunoprecipitation of a translated product using polyclonal antibodies to ovarian 42-kDa form of VTP; lane 2, total translated products; lane 3, processing of the translated polypeptide in the presence of canine pancreatic microsomal membrane preparation (Promega). Translation products were analyzed by SDS-PAGE in 12.5% gels under reducing conditions and fluorography. The molecular standards are as in Figure 1.

particularly high similarity to mammalian cathepsins B and to cathepsin B-like proteases of invertebrates, but only limited similarity to other cathepsins (Figure 4 and Table 1) (Chan *et al.,* 1986; Meloun *et al.,* 1988; Klinkert *et al.,* 1989; Cox *et al.,* 1990; Pratt *et al.,* 1990; Ray and McKerrow, 1992; Merckelbach *et al.*, unpublished; Takahashi *et al.*, 1993). Multiple alignment of homologous thiol proteases predicts that a putative start of the mature VTP is at position Leu 87 (Figs 2 and 4). In contrast to mature enzymes (Figure 4), the pro-peptide portion of VTP shows only very limited similarity to those of cathepsins B (not shown). Interestingly, the VTP pro-peptide has two putative glycosylation sites compared to only one in pro-enzyme portions of mammalian cathepsins B and none in cathepsin B-like proteases of invertebrates (Figure 2 and Table 2).

The sequence alignment indicates that all three active sites of cathepsins B are conserved in VTP at positions Cys 112, His 277 and Asn 297 (Figure 4). The positions of cysteines involved in five putative disulfide bonds are also conserved in VTP out of six characteristic for thiol proteases (Musil *et al.*, 1991).

Mosquito VTP is a fat body-specific protein produced in vitellogenic females in response to a blood meal.

Northern blot analyses confirmed data previously obtained at the protein level demonstrating that the VTP gene is expressed exclusively in the fat body of vitellogenic female mosquitoes and that its expression is initiated in response to a blood meal (Figure 5). When the entire VTP cDNA clone was used as a probe, in addition to a 1.1-kb transcript specific to vitellogenic fat bodies, trace hybridization to a slightly larger mRNA band was observed in all other tested tissues of both sexes (not shown). This trace hybridization was likely due to partial similarity between VTP and thiol proteases of lysosomal origin. When a 255-bp *Eco* RI-*Hinc* II fragment from the 5'-end of the VTP cDNA clone, that

Figure 4. Alignment of the deduced amino acid sequence of mosquito VTP with other thiol proteases. The consensus residues of enzyme active sites are marked by asterisks (*). Conserved amino acids are marked by black boxes. Paired numbers under the sequences denote the locations of cysteine residues that form a disulfide bond. Abbreviations are: AaVTP, *Aedes aegypti* VTP; HsCtB, human cathepsin B, *Homo sapiens* (Chan *et al.,* 1986); RnCtB, rat cathepsin B, *Rattus norvegicus* (Chan *et al.*, 1986); BtCtB, bovine cathepsin B, *Bos taurus* (Meloun *et al.,* 1988); MmCtB, mouse cathepsin B, *Mus musculus* (Chan *et al.*, 1986); Sp29K, 29-kDa protease of the blowfly, *Sarcophaga peregrina* (Takahashi *et al.,* 1993); SjCtB, trematode cathepsin B, *Schistosoma japonicum* (Merckelbach *et al.*, unpublished, GenBank accession number X70968); SmCyP, trematode cysteine protease, *S. mansoni* (Klinkert *et al.,* 1989); HcCy1, nematode gut thiol protease 1, *Haemonchus contortus* (Cox *et al.,* 1990); HcCy2, nematode thiol protease 2, *H. contortus* (Pratt *et al.*, 1990); CeCy1, nematode gut-specific cysteine protease, *Caenorhabditis elegans* (Ray and McKerrow, 1992). The descent rank of similarity to mosquito VTP was determined by *Initn* index of FastA of Genetics Computer Group software.

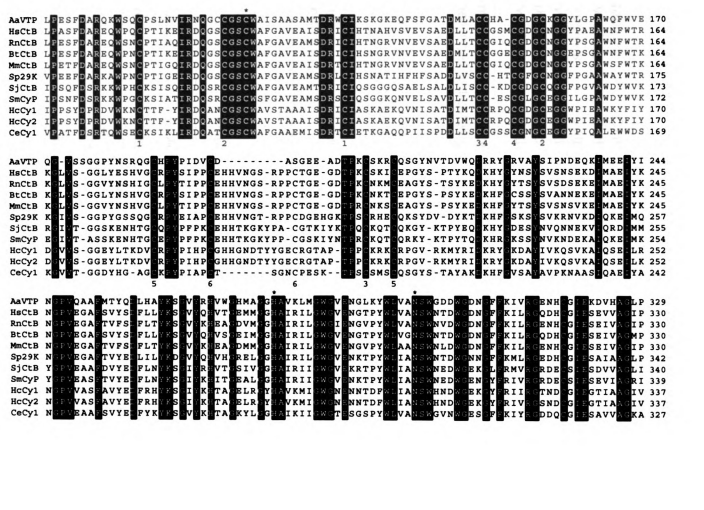

AaVTP
HaCtB
RnCtB
BtCtB
MmCtB
Sp29K
SjCtB
SmCyP
HcCy1
HcCy2
CeCy1

Table 1. Homology between mosquito thiol protease, cathepsins and papain

Species	Protease	*Initn* Index[a]
Homo sapiens	Cathepsin B	851
Rat norvegicus	Cathepsin H	260
Rat norvegicus	Cathepsin C	209
Rat norvegicus	Cathepsin S	173
Carica papaya	Papain	172
Rat norvegicus	Cathepsin L	165

[a] Determined by FastA program of Genetics Computer Group (GCG) software. A higher index indicates greater similarity.

Table 2. Number of putative N-linked glycosylation sites in several mammalian and invertebrate cathepsins B

		No. of N-linked Glycosylation Sites	
Animal	Protease	Pro-peptide	Active Enzyme
Aedes aegypti	Vitellogenic Thiol Protease	2	1
Homo sapiens	Cathepsin B	1	2
Rattus norvegicus	Cathepsin B	1	1
Bos taurus	Cathepsin B	1	3
Mus musculus	Cathepsin B	1	2
Sarcophaga peregrina	29 *kDa* Protease	0	3
Schistosoma joponica	Cathepsin B	0	1
Schistosoma mansoni	Cysteine Protease	0	2
Haemonchus contortus	Thiol Protease 1	0	4
Haemonchus contortus	Gut Thiol Protease 2	0	4
Caenorhabditis elegans	Gut-Specific Cysteine Protease	0	0

encodes the pre-pro-portion of VTP, was used as a probe, it hybridized only to the 1.1-kb transcript from vitellogenic fat bodies (Figure 5). In contrast, expression of actin, a house-keeping gene, was relatively constant in all tested tissues (Figure 5).

The secretory activity of fat bodies with respect to production of yolk protein precursors was evaluated by labeling fat bodies in the presence of ^{35}S-methionine for 1 h and the chase media secretions collected for analysis. VTP secretion was monitored by anti-VTP antibodies. As a control, anti-VCP antibodies were used to trace secretion of this yolk protein precursor. These analyses showed that the kinetics of VTP secretion by the fat body are similar to those of VCP (Figure 6). Both yolk protein precursors were detected at 4 h after initiation of vitellogenesis by a blood meal. Their synthesis and secretion increased rapidly to a maximum at 24 h PBM, declined to a very low level by 36 h and reached background level by 48 h (Figure 6).

VTP is co-localized with other yolk protein precursors in fat body cell's secretory and in oocyte's endocytotic organelles

Immunocytochemistry at the electron microscopical level was used to localize VTP in trophocytes of the vitellogenic fat body utilizing anti-VTP antibodies and protein A-colloidal gold. VTP was localized in organelles of the secretory pathway: the Golgi complex and secretory granules (not shown). No labeling was detected in lysosomes of fat body cells (not shown). When double immunolabeling was performed, utilizing antibodies for Vg and VTP and protein A-colloidal gold particles of two different sizes, VTP was co-localized with Vg in the Golgi complex and secretory granules of fat body trophocytes (Figure 7).

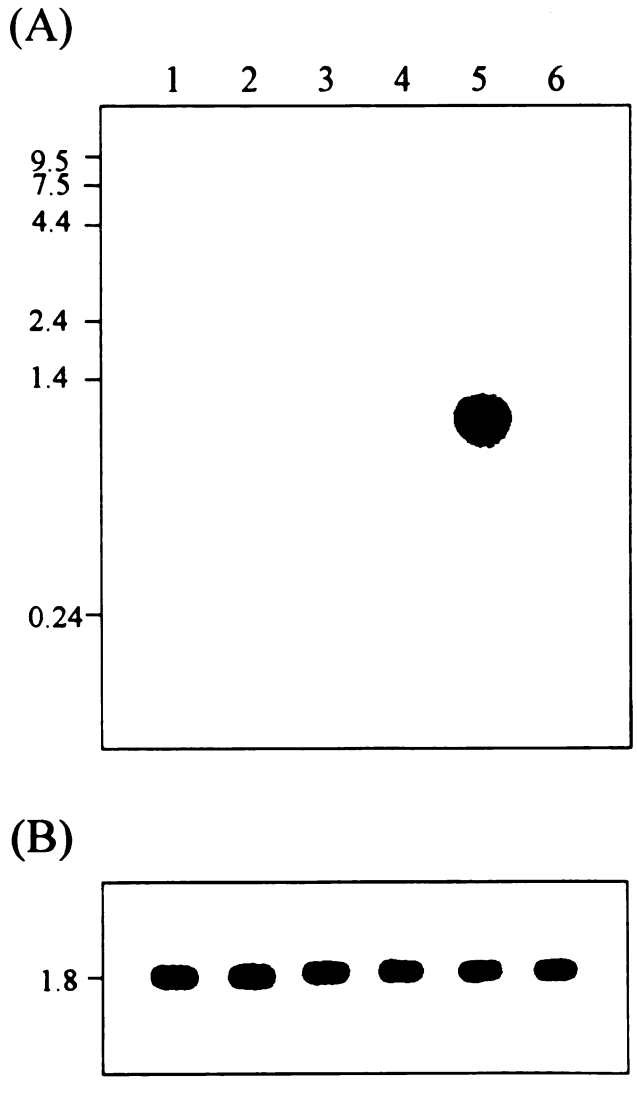


Figure 5. Northern blot analysis of gene expression of mosquito VTP and actin. Ten micrograms of total RNA from different sexes, tissues and stages were separated by a 1.2% agarose gel and transferred to a nitrocellulose membrane. (A) This membrane was hybridized with a [^{32}P]-labeled 255-bp *Eco* RI-*Hinc* II fragment of the VTP cDNA, encoding the pre-pro-portion of VTP. (B) After stripping, the membrane was rehybridized with a mosquito actin cDNA probe (Deitsch *et al.*, 1995). Lane 1, whole male; lane 2, vitellogenic female midgut, 24 h PBM; lane 3, vitellogenic female ovary, 24 h PBM; lane 4, pre-vitellogenic female fat body; lane 5, vitellogenic female fat body, 24 h PBM; lane 6, post-vitellogenic female fat body, 48 h PBM. RNA markers (Life Technologies) are shown to the left in (A).

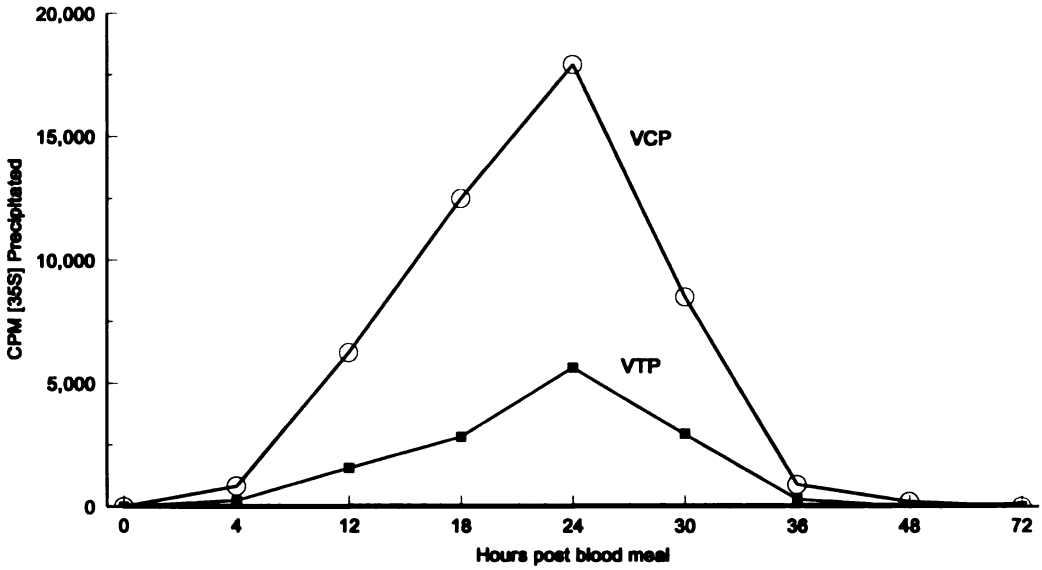
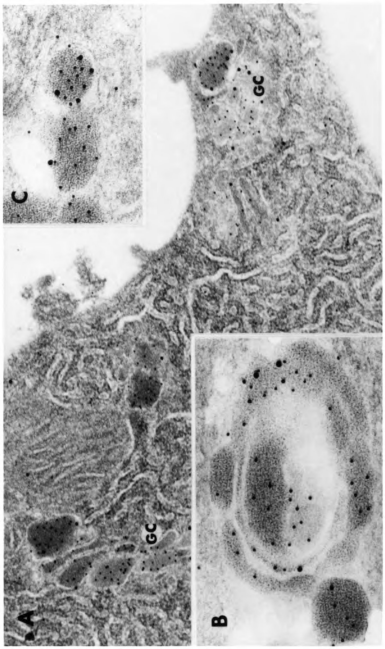


Figure 6. Time course of VTP secretion by fat bodies of female mosquitoes. Abdominal wall with adhering fat body of vitellogenic females at various stages of vitellogenesis, initiated by a blood meal, were incubated in culture medium in the presence of [^{35}S]methionine for 1 h and chased for 1 h in isotope-free medium. Chase media samples were immunoprecipitated using polyclonal antibodies to 42-kDa VTP or VCP and analyzed as described in Material and Methods. Data are expressed as CPM of [^{35}S]methionine-labeled protein per h from 3 fat bodies.

Figure 7. Co-localization of VTP and vitellogenin (Vg) in Golgi complexes (GC) and secretory granules of fat body cells from a vitellogenic female mosquito. VTP was localized with polyclonal antibodies to 42-kDa VTP followed by protein A-colloidal gold (15 nm); Vg was localized with polyclonal antibodies to the large Vg subunit and protein A-colloidal gold (10 nm). (A) fat body cell; (B) Golgi complex; (C) secretory granules. A is 50,000X; B and C are 100,000X.


A
GC
GC
B
C

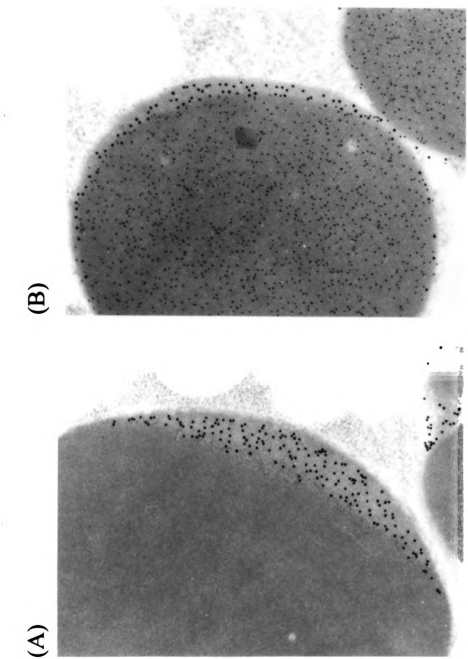
In vitellogenic ovarian follicles, which consist of the oocyte and nurse cells surrounded by follicle cells, VTP was present only in the oocyte's coated vesicles, endosomes and yolk bodies (not shown). Simultaneous labeling for VTP and Vg demonstrated that these proteins were mixed together in coated vesicles and endosomes (not shown). However, localization of VTP was dramatically different in mature yolk bodies where it was distributed as a narrow layer on the surface of the crystalline vitellin (Vn), a storage form of Vg (Figure 8A). This area of the yolk body was visible as a non-crystalline matrix separating the crystalline yolk from the yolk body membrane. Double immunolocalization showed that in mature yolk bodies this non-crystalline matrix is free of Vn, while the crystalline yolk is always free of VTP (Figure 8B). The distribution of VTP in mature yolk bodies was similar in oocytes at the peak of endocytosis (24 h PBM) and in those after termination of yolk accumulation (48 h PBM).

Correlation of changes in the native and subunit composition of VTP in the egg with embryonic development

The processing of VTP in oocytes and eggs was monitored by immunoblot analysis (Figure 9). At the peak of yolk protein uptake by oocytes (24 h PBM) (Koller *et al.*, 1989), the size of internalized VTP was 42 kDa. In fully grown oocytes which have completed yolk protein accumulation and have nearly completed choriogenesis (48 h PBM) (Raikhel and Lea, 1991), the size of VTP was slightly reduced to 41 kDa. As embryogenesis is initiated in newly-laid eggs, the molecular mass of VTP becomes 33 kDa. It remained unchanged until the end of embryogenesis, 94-96 h after oviposition. At this stage, only a trace of the 33-kDa band was visible. No immunopositive VTP bands were detected in extracts of newly hatched first instar larvae (Figure 9).

Native PAGE and immunoblot analysis revealed that in the hemolymph VTP exists as a high molecular weight molecule (Figure 10). Its apparent size

Figure 8. Immunolocalization of VTP in mature yolk bodies of developing oocytes. A, anti-VTP polyclonal antibodies followed by protein A-colloidal gold (15 nm); B, double labeling for VTP as in (A) followed by polyclonal antibodies to a large Vg subunit and protein A-colloidal gold (10 nm). Note that VTP, marked with 15 nm colloidal gold particles, is localized in the non-crystalline matrix of yolk bodies. A and B are 50,000X.


(B)
(A)

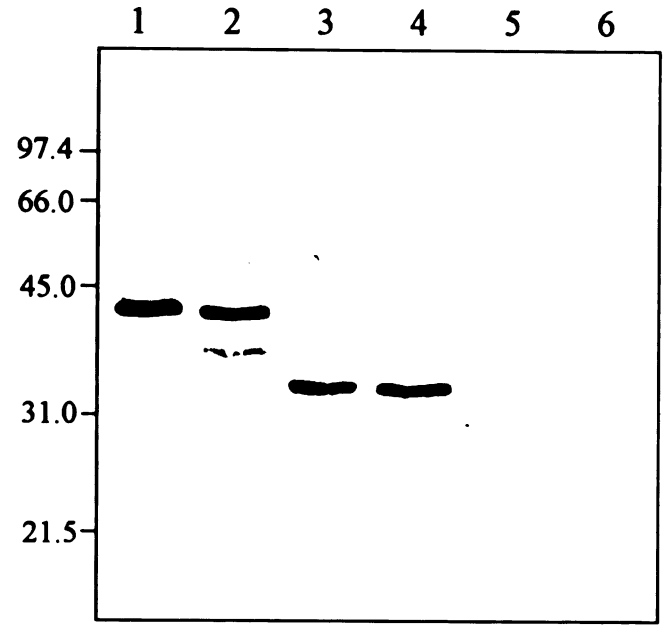


Figure 9. Processing of VTP in the ovary and embryos. Lane 1, ovaries, at the peak of Vg uptake, 24 h PBM; lane 2, ovaries with eggs nearly complete in their development, 48 h PBM; lane 3, 0- to 3-h postoviposition eggs at the onset of embryonic development; lane 4, 48-h postoviposition eggs during mid-embryogenesis; lane 5, 96-h postoviposition eggs at the end of embryogenesis; lane 6, newly hatched first-instar larva. Proteins were resolved by SDS-PAGE under reducing conditions in 12% gels. Proteins transferred onto a nitrocellulose membrane were probed by polyclonal antibodies against 42-kDa VTP. The molecular mass standards are as in Figure 1.

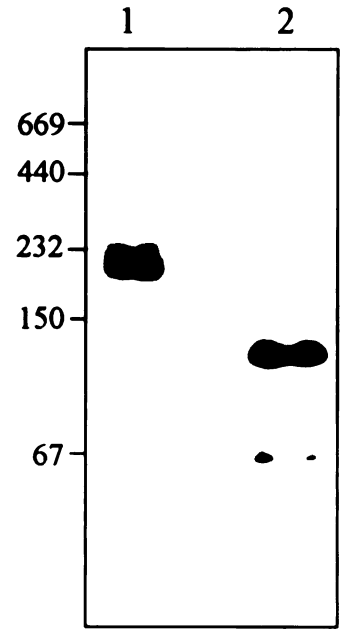


Figure 10. Native molecular mass of hemolymph and embryonic forms of VTP. Lane 1, *in vitro* secretion from vitellogenic fat bodies 20 h PBM; lane 2, 48-h postoviposition eggs during mid-embryogenesis. Proteins were separated in a 3.5 - 25% gradient gel and transferred to nitrocellulose membrane. The blot was processed using polyclonal antibodies to 42-kDa VTP. The molecular standards in order of decreasing molecular mass (in kDa) were thyroglobulin, ferritin, catalase, lactate dehydrogenase and bovine serum albumin (Pharmacia).

varied slightly depending on gel conditions, but averaged a value of 236±7 (SE) kDa. At the onset of embryonic development in the egg, when VTP is processed to its 33-kDa form, its native size decreased to 132±4 kDa. There was also a minor immunopositive band of 67 kDa in the same preparation (Figure 10).

Mosquito VTP is activated in embryos and is involved in degradation of the major yolk protein, vitellin

At the onset of embryonic development, the processing of VTP to its 33-kDa form coincides with the beginning of Vn degradation (Hays and Raikhel, unpublished). To evaluate whether VTP processing results in its enzymatic activation, we used Z-Arg-Arg-*p*NA, a substrate which exhibits specificity to cathepsin B (Barrett and Kirschke, 1981). Secretions from vitellogenic fat bodies, containing 44-kDa VTP, did not exhibit appreciable cathepsin B enzymatic activity (not shown). Protein extracts from embryos containing 33-kDa VTP had significant cathepsin B activity (Figure 11). Enzymatic activity was optimal at pH 5.5 (not shown). This enzymatic activity, associated with embryonic extracts containing the 33-kDa VTP, was highly sensitive to E-64. It showed partial sensitivity to chymostatin (Figure 11).

Next, we determined whether the cathepsin B activity detected in mosquito embryos is linked to Vn degradation. Purified Vg, labeled with [^{35}S]methionine, was degraded by embryonic protein extracts containing the 33-kDa VTP (Figure 12). Degradation of Vg was entirely inhibited by addition of E-64 to the protein mixture. Protease inhibitors specific to serine-, aspartic- and metallo-proteinases were not effective in inhibiting Vg degradation.

We tested whether pro-VTP can be processed and activated under acidic conditions as can mammalian cathepsins B (Rowan *et al.*, 1992; Mach *et al.*, 1994b). The secretory 44-kDa and ovarian 42-kDa forms of VTP were incubated under different acidic pH conditions (4.0 and 5.5). After treatment with acidic pH, the 44-kDa pro-VTP neither catalyzed the substrate Z-Arg-Arg-*p*NA (Figure 13A)

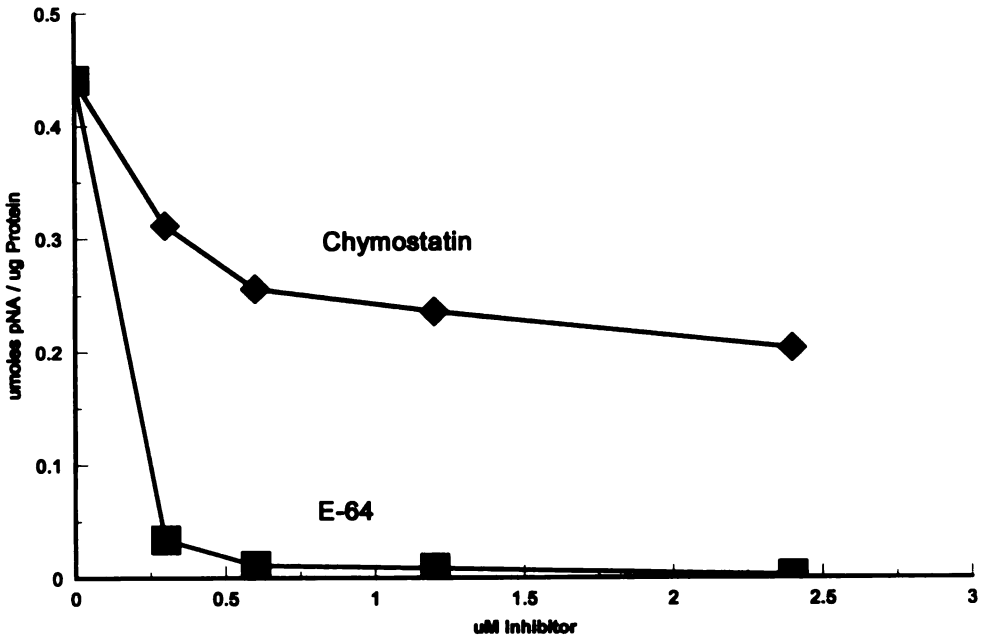


Figure 11. Differential effect of inhibitors on enzymatic activity of VTP. The embryonic extract from 12 to 48 h postoviposition eggs, containing 33-kDa VTP, was examined for enzymatic activity with Z-Arg-Arg-*p*NA as described in Materials and Methods. The reaction shown was performed for 60 min at pH 5.5. The inhibitors, E-64 and chymostatin, were added to the protein mixture at the final concentration indicated on the axis and preincubated for 20 min at 37°C prior to the addition of the enzymatic substrate. The control reaction was preincubated without inhibitor. The activity was expressed as mM of *p*NA released from the substrate Z-Arg-Arg-*p*NA per mg of protein.

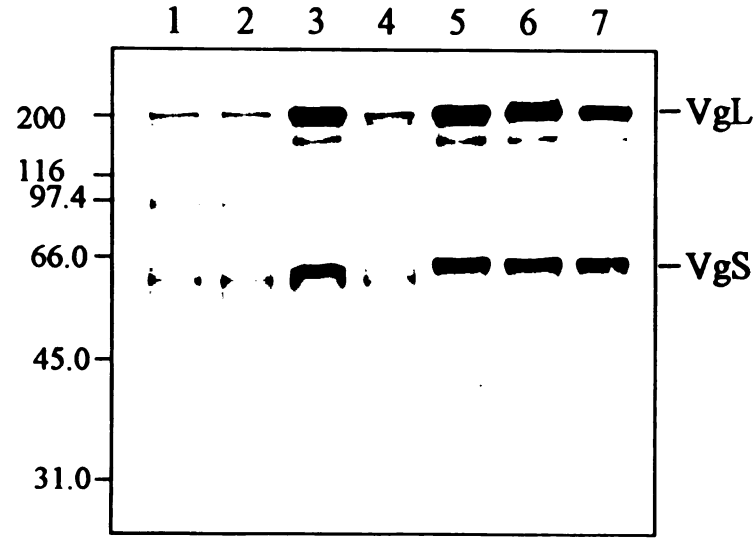
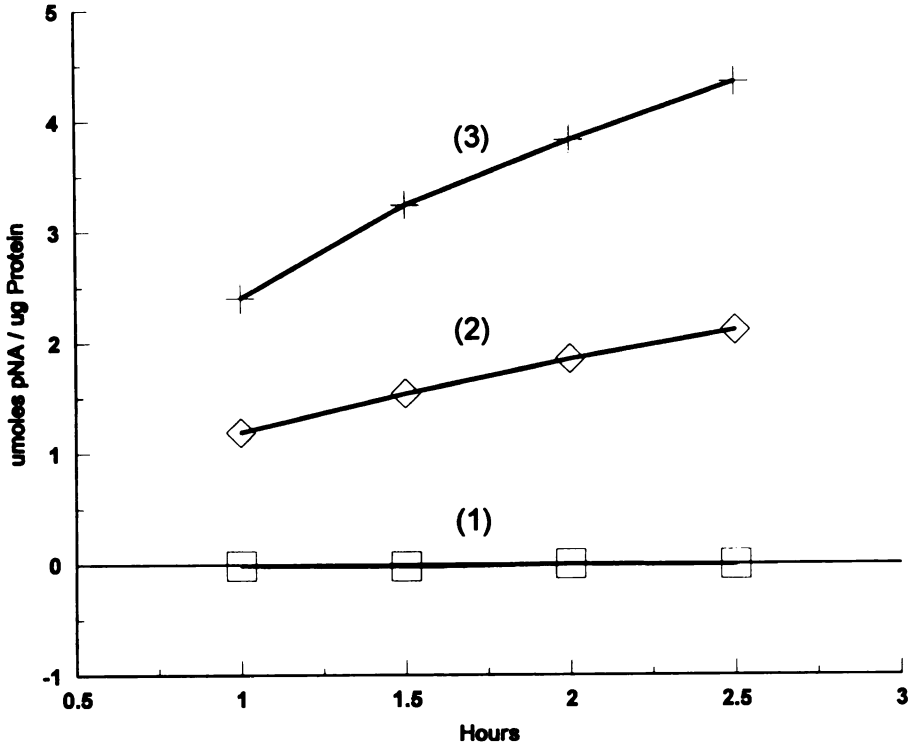


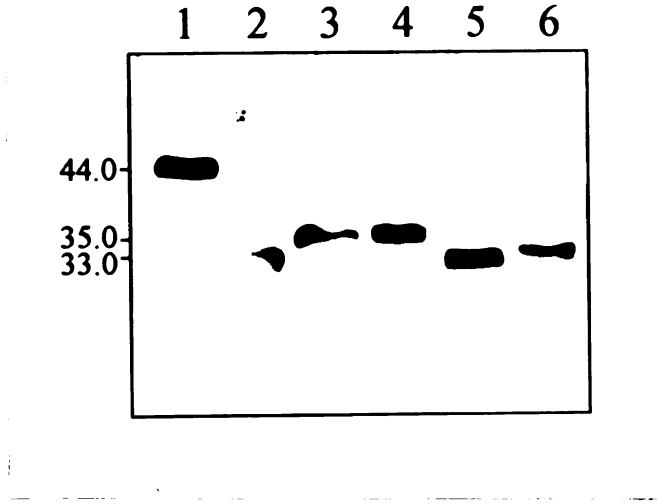
Figure 12. Degradation of vitellogenin by embryonic extract (12 to 48 h postoviposition eggs) containing 33-kDa VTP. Purified [^{35}S]methionine-labeled Vg was incubated with the embryonic extract at pH 4.0 and resolved by SDS-PAGE in a 10 % gel under reducing conditions. The molecular mass standards are as in Figure 1. Lane 1, the reaction was performed in the presence of serine protease inhibitors, 1 mM PMSF and 1 mM AEBSF; lane 2, in the presence of 5 mM pepstatin, an inhibitor of aspartic proteases; lane 3, in the presence of 2.5 mM E-64, an inhibitor of thiol proteases; lane 4, PMSF, AEBSF and pepstatin; lane 5, all listed protease inhibitors; lane 6, incubation of Vg without the embryonic extract, at pH 4.0; lane 7, Vg sample without any treatment. The 200-kDa large Vg subunit (VgL) and a 66-kDa small subunit (VgS) are seen.

Figure 13. Activation of mosquito VTP under acidic conditions. (**A**) Protein extracts supplemented with protease inhibitors (1 mM PMSF, 1 mM AEBSF, 5 mM pepstatin and 2.5 mM EDTA) were incubated for the indicated times under the acidic (pH 4. 0 or 5.5) or control conditions (pH 7.5). Enzymatic activity was measured using Z-Arg-Arg-*p*NA substrate. Activity is expressed as in Figure 11. Line 1, fat body secretions; line 2, protein extracts from ovaries after the termination of yolk accumulation, 48 h PBM; line 3, protein extract from embryos, 12 to 48 h oviposition eggs. (**B**) Immunoblot analysis of activated VTP peptides. Immunoblot was performed as in Figure 1. Lane 1, fat body secretions, under the control conditions at pH 7.5; lane 2, embryonic extract under the control conditions; lane 3, fat body secretions under the acidic conditions at pH 4.0; lane 4, fat body secretions under acidic condition at pH 5.5, corresponds to the enzymatic reaction (1) in 13A; lane 5, protein extract from vitellogenic ovaries under acidic conditions at pH 5.5, corresponds to the enzymatic reaction (2) in 13A; lane 6, protein extract from embryos, 12 to 48 h postoviposition eggs, under acidic conditions at pH 5.5, corresponds to the enzymatic reaction (3) in 13A.

(A)


5
4
3
2
1
0
-1
umoles pNA / ug Protein
(3)
(2)
(1)
0.5
1
1.5
2
2.5
3
Hours

(B)


1 2 3 4 5 6
44.0
35.0
33.0

nor degraded [^{35}S]methionine-labeled Vg (not shown). Immunoblot analysis showed that the 44-kDa proVTP was processed to a 35-kDa peptide but not to a 33-kDa one (Figure 13B). Unlike the 44-kDa pro-VTP, the ovarian extract containing the 42-kDa VTP exhibited cathepsin B activity after treatment with acidic pH (Figure 13A). Immunoblot showed that it was processed to a 33-kDa form (Figure 13B). Processing of VTP was equally efficient at pH 4.0 and 5.5 and was not effected by the presence of protease inhibitors specific to serine-, aspartic- and metallo-proteinases (Figure 13B).

DISCUSSION

In this paper we report an unusual form of cathepsin B-like thiol protease from the mosquito, *Aedes aegypti*. Cloning and analysis of the cDNA encoding VTP demonstrated its high similarity to mammalian cathepsins B and invertebrate cathepsin B-like proteases, but not to other related cathepsins (Table 1 and Figure 4) (Chan *et al.,* 1986; Meloun *et al.,* 1988; Klinkert *et al.,* 1989; Cox *et al.,* 1990; Pratt *et al.,* 1990; Ray and McKerrow, 1992; Merckelbach *et al.*, unpublished; Takahashi *et al.*, 1993). The predicted start of mature VTP at Leu 87, which is similar to mammalian cathepsins B, the conserved active sites, and conserved positions of cysteines determining five putative disulfide bonds (Figure 4) indicates that the folding of VTP and cathepsins B are similar. This conserved structure of the mature VTP suggests that it is a functional enzyme.

The mosquito VTP also bears similarities to mammalian cathepsins B in its size. It is secreted as a proenzyme of 44 kDa that is similar in size to the latent pro-cathepsin B secreted by malignant cells (Mort and Recklies, 1986; Mach *et al.*, 1992). The 33-kDa embryonic form of VTP is similar in size to the active single-chain cathepsin B (Mach *et al.*, 1992).

The presence of the 33-kDa embryonic form of VTP is correlated with activity characteristic of cathepsin B (Barrett and Kirschke, 1981). This activity is highly expressed with the cathepsin B-specific substrate, Z-Arg-Arg-*p*NA; it is sensitive to E-64 but less so to chymostatin.

The unique feature of this mosquito cathepsin B-like enzyme is that it is synthesized and secreted exclusively by the fat body of vitellogenic female mosquitoes as a latent, high molecular size precursor consisting of five or six subunits with molecular mass of 44 kDa. Its synthesis in the female fat body is initiated by blood feeding and the kinetics of its secretion by the vitellogenic fat body are similar to those of the yolk protein precursors, Vg (Cho and Raikhel, 1992) and VCP (Cho *et al.*, 1991). The site of origin of VTP was additionally confirmed by its co-localization with Vg in the Golgi complexes and secretory granules of vitellogenic fat body cells.

The link between the 44-kDa fat body-secreted, hemolymph VTP and the 42-kDa ovarian VTP was firmly established because in all immunological analyses involving the fat body we successfully used the antibodies produced against the ovarian form of VTP. In addition, cloning of the VTP cDNA from the female mosquito fat body cDNA library was based on the amino acid sequences obtained from the purified 42-kDa ovarian VTP. In spite of the presence of large amounts of VTP in the ovary, neither its mRNA nor synthesis were detected there. Thus, the fat body was the only source of the 44-kDa polypeptide as a precursor of the ovarian 42-kDa polypeptide which was likely accumulated in the ovary by endocytosis from the hemolymph similar to the other yolk protein precursors, Vg and VCP (Raikhel and Dhadialla, 1992). Immunolocalization of VTP in the ovary confirmed that this yolk protein precursor is internalized via the endocytotic pathway by developing oocytes. Interestingly, VTP was segregated from crystalline Vn in mature yolk bodies, an oocyte's accumulative endocytotic

organelle, being present only in the non-crystalline matrix surrounding Vn. Here, VTP is mixed with VCP which is also located in the non-crystalline matrix of mature yolk bodies (Snigirevskaya and Raikhel, unpublished data). Co-localization of both these pro-enzymes in the matrix surrounding crystalline Vn likely ensures their rapid activation at the onset of embryonic development when yolk bodies undergo acidification. The latter event has been documented for both insects and vertebrates (Nordin *et al.*, 1991; Fagotto and Maxfield, 1994).

The fat body-secreted, hemolymph form of VTP is a stable latent pro-enzyme which, unlike mammalian cathepsins B (Rowan *et al.*, 1992; Mach *et al.*, 1994b), cannot be activated by acidic pH alone. The size of the hemolymph pro-VTP was reduced after treatment with acidic pH to 35-kDa (not 33-kDa), but this reduction was not sufficient for enzyme activation. Clearly, an additional step or steps are required for activation of the latent hemolymph pro-VTP. Pro-VTP did not activate itself in an *in vitro* transcription/translation system (Figure 3) as was reported for *S. mansoni* cathepsin B-like cysteine protease (Felleisen and Klinkert, 1990). Stability of the hemolymph pro-VTP as a latent proenzyme prior to its internalization by developing oocytes is physiologically important. It may be additionally enhanced by the presence of two glycosylation sites in its pro-enzyme portion. The presence of such sites is unique to VTP in comparison to other cathepsin B-like proteases of invertebrates which function as digestive or lysosomal enzymes (Table 2).

The estimation of native molecular weight suggests that in the hemolymph, VTP exists as a hexamer or a pentamer of 44-kDa subunits (Figure 10). A hexameric structure is characteristic for insect hemolymph proteins (Telfer and Kunkel, 1991). This feature of VTP may also increase its stability in the hemolymph. Alternatively, it may be important for recognition of VTP by oocyte receptors. At the onset of embryonic development, when VTP is

processed to the active 33-kDa form, its native size is reduced to 132±4 kDa which likely corresponds to a tetramer of 33-kDa subunits. The presence of a minor 67-kDa band indicates that at least some of 33-kDa VTP are in a dimer form (Figure 10).

After treatment with acidic pH, the ovarian extract containing the 42-kDa VTP exhibits cathepsin B activity. Significantly, under these conditions the ovarian 42-kDa VTP is processed to the 33-kDa polypeptide indicating its activation. This suggests that pro-VTP undergoes the first step of processing after its internalization by oocytes. Although the nature of this step is unclear, the removal of 2 kDa from each VTP subunit apparently renders pro-VTP capable of activation by acidic pH after the onset of embryonic development. Although the activation reactions were performed using crude extracts, it is unlikely that the proteolytic processing of VTP by treatment with acidic pH is a function of another protease because both the 44-kDa hemolymph and the 42-kDa ovarian VTP were processed to 35 and 33 kDa in the presence of protease inhibitors. Autoactivation under acidic conditions was reported for mammalian cathepsins B (Rowan *et al.*, 1992; Mach *et al.*, 1993) and this is a possible mechanism for activation of the 42-kDa VTP in the mosquito embryo.

Our experiments utilizing ^{35}S-labeled Vg as a substrate suggest that VTP likely plays a key role in degradation of this major yolk protein in mosquito embryos. Only the embryonic extract containing the 33-kDa VTP degraded Vg and this degradation was blocked by E-64. Future tests with purified VTP should be performed to confirm this role of VTP.

Proteases similar to cathepsins B and L have been shown to be active during embryonic development in a number of arthropods (Medina *et al.*, 1988; Medina and Vallejo, 1989; Fagotto, 1990a,b; Ribola *et al.*, 1993; Takahashi *et al.*, 1993; Yamamoto and Takahashi, 1993; Yamamoto *et al.*, 1994b). The best

studied is the cathepsin L-like cysteine protease from the silkworm, *Bombyx mori* (Takahashi *et al.*, 1993; Yamamoto and Takahashi, 1993; Yamamoto *et al.*, 1994a,b). It is produced as a latent proenzyme of 47 kDa which is activated under acidic conditions to a 39-kDa form, corresponding to the active enzyme present in embryos (Takahashi *et al.*, 1993). In addition to enzymatic properties of this embryonic enzyme consistent with a cathepsin L-like cysteine protease, its identity was further confirmed by its high sequence similarity to mammalian cathepsin L (Yamamoto *et al.*, 1994a). Importantly, direct action of *Bombyx* cathepsin L-like cysteine protease on vitellin was demonstrated (Yamamoto *et al.*, 1994b). The cDNA encoding the pro-enzyme of this *Bombyx* cathepsin L-like cysteine protease was cloned from the ovarian cDNA library (Yamamoto *et al.*, 1994a). Immunological analyses showed that it is produced in the ovary by follicular cells and is deposited in developing oocytes (Yamamoto *et al.*, 1994b).

In *Drosophila*, cathepsin B-like protease is presumed to be maternal in origin, however its precise origin is not known (Medina and Vallejo, 1989). The unique feature of the mosquito embryonic cathepsin B-like protease that we have described in this paper is its production by an extraovarian tissue as a yolk protein precursor.

In oviparous animals, extraovarian tissues play an important role in egg maturation by producing yolk protein precursors which serve as a major nutritional source for developing embryos. The liver of oviparous vertebrates and the fat body of insects produce large quantities of yolk protein precursors which are internalized by developing oocytes and deposited in yolk bodies. The most abundant of these yolk protein precursors is Vg which is a large glycophosholipoprotein conserved throughout the evolution of invertebrate and vertebrate oviparous animals (Wahli, 1988; Byrne *et al.*, 1989; Raikhel and Dhadialla, 1992; Chen *et al.*, 1994).

In the mosquito, however, the proenzymes VTP and VCP are produced by the fat body as yolk protein precursors. The discovery that enzymes which are involved in embryonic degradation of yolk protein are themselves produced as yolk protein precursors by an extraovarian tissue is a new biological phenomenon previously unknown in insects or other oviparous animals. Moreover, the finding that one of these enzymes is a cathepsin B sheds light on yet another important mode of utilization of this key protease.

ACKNOWLEDGMENTS

We thank Drs. T. Sappington, K. Deitsch and A. Biran for critical reading of the manuscript. This work was supported by NIH grants AI-24716 and AI-32154 to ASR.

CHAPTER 6

SUMMARY AND FUTURE RESEARCH PROSPECTS

SUMMARY

1. Cloning and Sequence Analysis of the Vg cDNA

Cloning and sequencing of a cDNA coding for mosquito Vg have provided tremendous molecular data for characterizing the physiological significance of this yolk protein precursor. This cDNA encodes a pro-Vg that has a hydrophobic signal located at the N-terminus, followed by a small subunit in the middle and a large subunit at the C-terminus. The cleavage site between the two subunits of mosquito pro-Vg has a consensus RXRR recognized by enzymes of the convertase family. This paired basic motif is found at the cleavage sites of insect Vgs but not Vgs of other taxa. Mosquito Vg is a highly hydrophilic protein with 17 putative N-linked glycosylation sites and 13 potential tyrosine sulfation sites. In contrast to other Vgs, mosquito Vg has three polyserine domains that are similar to the phosvitin subunits of vertebrate Vgs. These hydrophilic polyserine domains, originally presumed to be vertebrate-specific, have several putative phosphorylation sites in their sequences. Mosquito Vg is also unique in that it contains two highly hydrophilic tyrosine-rich domains. Unlike other Vgs, mosquito Vg is rich in the aromatic amino acids tyrosine and phenylalanine. Its sequence also shows similarity to insect serum proteins, arylphorins. Such similarity suggests that mosquito vitellogenin may supply aromatic amino acids to the cuticle of rapidly developing embryos.

2. Conservation in Insect Vgs

Sequences of Vgs, the major yolk proteins that supply nutrition for developing embryos of oviparous animals, have been aligned and analyzed, and are believed to have evolved from a common ancestor. Before insect Vg sequences were available, it was generally accepted that all Vg sequences are conserved because the sequence of nematode Vg can be aligned very well with that of chicken and *Xenopus* except that nematode Vg does not have a serine-rich phosvitin domain. Phosvitin, therefore, was believed to be vertebrate-specific. However, molecular cloning and characterization of Vg cDNAs and genes from several insect species have revised this hypothesis. Insect Vgs are different from other Vgs not only because their sequences have limited similarity but also because they have conserved acidic serine-rich and charged domains flanking a paired basic motif by their cleavage sites. These conserved acidic serine-rich and charged domains form β-turns that are essential for recognition of the putative paired basic motif by convertases. These conserved domains and paired basic motifs at the cleavage sites of insect Vgs reveals that the insect Vg processing enzymes belong to the convertase family. Sequence analyses also indicate that the acidic serine-rich domains of insect Vgs are similar to phosvitin, a vertebrate Vg-specific serine-rich subunit. Unlike vertebrate Vgs that have a single long serine-rich domain, the number and size of acidic serine-rich domains in insect Vgs vary considerably among species.

3. Molecular Cloning and Functional Expression of the VC cDNA

The mosquito VC cDNA was cloned from a vitellogenic female fat body cDNA library. Analysis of the deduced amino acid sequence indicated that mosquito VC has high similarity in sequence and domain structure to the convertase family and especially to *Drosophila* convertases dKLIP-1 and dfurin1.

VC has all four conserved catalytic sites found in convertases. The open reading frame of this cDNA encodes a protein 115 kDa in size. VC is post-translationally modified to 140 kDa. The VC cDNA was co-expressed with a truncated mosquito Vg cDNA in a coupled transcription and translation gene expression system. The result indicated that VC is the processing enzyme that cleaves pro-vitellogenin into small and large subunits. VC is the first convertase identified playing a role in insect pro-Vg cleavage. Discovery of this long-sought-after enzyme has filled the gap in the characterization of insect Vg biosynthesis.

4. Cloning and Characterization of VTP

Sequence analysis of mosquito VTP revealed that this yolk protein precursor is a thiol protease. It is synthesized and secreted exclusively in a sex-, stage- and tissue-specific manner as a latent proenzyme by fat bodies of female mosquitoes during vitellogenesis in response to blood feeding. Such a gene expression pattern is very similar to that of Vg and VCP. The deduced amino acid sequence of this 44-kDa precursor shares high similarity with cathepsins B. The size of VTP shifts from 44 kDa to 42 kDa after internalization by oocytes from the hemolymph. VTP is located in the matrix surrounding vitellin in mature yolk bodies. This yolk protein precursor is further processed to 33 kDa at the onset of embryogenesis. Embryonic extract containing the 33-kDa VTP was active toward benzoyloxycarbonyl-Arg-Arg-para-nitroanilide, a cathepsin B-specific substrate, and degraded vitellin. Both of these enzymatic activities were prevented by E-64, a thiol protease inhibitor. These data indicate that VTP is an enzyme responsible for vitellin degradation during embryogenesis.

FUTURE RESEARCH PROSPECTS

1. Determining the Tertiary Structure of Mosquito Vg

Analyses of the deduced amino acid sequence and amino acid content of mosquito Vg revealed several unique characteristics of this developmentally important protein. To understand the pathways of biosynthesis and accumulation of mosquito Vg at the molecular level, determination of the tertiary structure of this yolk protein is required. A three-dimension model of Vg will reveal the functional and the structural elements of this yolk protein precursor, e.g., the recognition sites of Vg for the oocyte Vg receptor. Because the deduced amino acid sequence alone is not reliable in predicting tertiary structure, X-ray crystallography of the crystallized protein is necessary to establish the functional structure (Sharrock *et al.*, 1992).

The structure of crystalline lipovitellin from the lamprey has been solved by X-ray diffraction (Raag *et al.*, 1988). Using this data along with the deduced amino acid sequence of lamprey Vg (Sharrock *et al.*, 1992), these authors were able to predict a more concise model of lamprey Vg structure than previously predicted from either data source alone (Sharrock *et al.*, 1992). Sample preparation of mosquito Vg for X-ray diffraction may follow Raag *et al.* (1988). However, it should be noted that not every Vg can be crystallized, for example, *Xenopus* Vg can not (Sharrock *et al.*, 1992). This may due to the fact that *Xenopus* Vgs are encoded by a multi-gene family (Byrne *et al.*, 1989). The heterogeneity in peptide sequences may disrupt crystal formation (Sharrock *et al.*, 1992). Such a technical problem can be avoid by expressing a single mosquito Vg cDNA in a baculovirus or yeast gene expression system. Mosquito Vg has been introduced into a baculovirus genome and expressed in SF-21 insect cells. However, most of the over-expressed mosquito Vg was degraded. To obtain a large amount of Vg for crystallization, the best expression condition

should be worked out, e. g., try to express in other cell lines. Depending on the purpose, the mosquito Vg cDNA can be expressed with or without mosquito VC.

One potential problem for crystallization of mosquito Vg is glycosylation. Protein glycosylation is known as a potential negative factor for protein crystallization (Sharrock *et al.*, 1992). Special conditions should be tried for preparation and crystallization of mosquito Vg because this yolk protein precursor is heavily glycosylated (Dhadialla and Raikhel, 1990; Chen *et al.*, 1994).

2. The Role that Polyserine Domains play in Substrate Recognition of Mosquito VC and Ligand Binding of Oocyte Vg Receptors

Convertases that recognize paired basic motifs are known for cleavage of polypeptides and proproteins (Barr 1991). Rholam *et al.* (1986) found that not all the putative 53 paired basic sites in 20 prohormones examined are cleaved by convertases. These authors, therefore, analyzed the secondary structures around these cleavage sites and concluded that β-turns or Ω-loops situated in or next to the motif are required for recognition by convertases (Rholam *et al.*, 1986; Bek and Berry, 1990). This conclusion has been confirmed by using synthesized peptides as substrates for convertases (Brakch *et al.*, 1993).

Alignment of the sequences near the insect Vg subunit cleavage sites indicates a paired basic motif at the insect pro-Vg cleavage site is enclosed within a highly charged domain that is flanked by two polyserine domains. It is not yet clear what the physiological functions of polyserine domains of insect Vgs are. However, polyserine domains are predicted to form critical β-turns which are usually located on the surface of globular proteins (Ring *et al.*, 1992). According to their high hydrophilicity and β-turn formation, these domains are loose in structure and flexible. Polyserine domains provide strong hydrophilic forces that

may pull these domains and their flanking peptides up to the surface of the folded Vg. Such a feature is physiologically important because it may form surface structures for recognition by proteases, receptors and antibodies. Based on this property, Chen *et al.* (1994) proposed that these domains may play a role in recognition of mosquito Vg by oocyte Vg receptors. In addition, they may also be involved in helping convertase recognize the paired basic motif at the cleavage site of mosquito Vg.

To analyze the possibility that polyserine domains are involved in substrate recognition by convertases and oocyte Vg receptors, the sequence that codes for the polyserine domains can be manipulated by sequence replacement, such as site-directed mutagenesis, insertion and deletion. This question will be answered by examining the cleavage of mutant pro-Vgs which are co-expressed with VC cDNA in a baculovirus or yeast gene expression system. The expressed mutant Vgs can also be isolated and their affinity to the mosquito oocyte Vg receptor tested.

Experiments such as moving the RXRR motif of the mosquito Vg cleavage site to other parts of Vg, with or without an adjacent β-turn structure, can be done to characterize mosquito VC. To test the importance of charged domains with polyserine domains for VC substrate recognition, the third polyserine domain of mosquito Vg should be a good candidate. The third polyserine domain also forms a β-turn, however unlike the first and the second polyserine domains, there is no strongly charged domain nearby (Chen *et al.*, 1994). The accessibility of mutated Vgs with the RXRR motif moved to either the N- or C-termini of the third polyserine domain may test the necessity of the adjacent charged domain.

3. Autocleavage and Autoactivation of VTP

One of the unique properties of cathepsins B is autoactivation, especially under acidic conditions (Rowan *et al.*, 1992; Mach *et al.*, 1993). Acidifying the culture media containing proteins secreted from cultured vitellogenic fat bodies causes the molecular weight of VTP to decrease from 44 to 35 kDa. It is not clear whether this is autocleavage or cleavage by other enzymes. It also remains unclear which part of VTP is removed. This cleavage does not activate VTP because the 35-kDa VTP has no enzyme activity. In contrast, acidified ovarian extracts that contain the 42-kDa form of VTP exhibit cathepsin B activity. The ovarian 42-kDa VTP is processed to 33 kDa which is the same size as the active embryonic VTP. Again, the nature of this processing is unclear. Microsequencing the N-termini of the ovarian 42-kDa, embryonic 33-kDa VTPs and these *in vitro* processed 33 and 35-kDa VTPs may help determine which parts of VTP are removed.

To characterize the autocleavage properties of VTP, monitoring size changes of the purified ovarian 42 kDa and the hemolymph 44 kDa VTPs under acidic conditions can be done. If the 33 and 35 kDa proteins are not obtained from the 42 and 44 kDa VTPs, respectively, then there should be other pathways involved in the cleavage and activation of VTP. A potential VTP processing enzyme is mosquito vitellogenic carboxypeptidase (VCP). This enzyme has a very similar biosynthetic pathway to VTP and is accumulated with VTP in the yolk bodies of oocytes. Furthermore, VCP is also activated during embryogenesis. Although the substrate spectrum of VCP remains unclear, purified 44-kDa or 42-kDa VTP can be incubated with the purified active form of VCP. Monitoring the size changes and the enzyme activities of these VTPs will test this hypothesis.

It is not known what the nature of the 2-kDa reduction of the 44-kDa VTP after uptake by the oocyte is. The removal of the 2-kDa appears to be necessary for subsequent VTP activation. It is possible that this 2-kDa decrease is due to deglycosylation because VTP has three potential N-linked glycosylation sites. Glycosylation is known for protecting proteins from proteolytic attack. These polysaccharides may shield the active sites of VTP such that autoactivation of VTP is inhibited. Although endoglycosidase (Endo H) can be used to remove N-linked glycosylation of VTP, it may remove polysaccharide groups from all three glycosylation sites. Such treatment may not be useful for comparing with the 42-kDa protein because the 42-kDa VTP may still be glycosylated. To test the hypothesis that N-linked glycosylation may be involved in preventing activation of the 44-kDa VTP, sequences encoding putative N-linked glycosylation sites can be replaced by others. VTP mutants with single, double or all glycosylation sites removed can be expressed and secreted by baculovirus or yeast. Monitoring the autoactivation of these deglycosylated mutants under acidic conditions will test the role that N-linked glycosylation plays in activation inhibition. This experiment will also pinpoint which N-linked glycosylation sites are involved in inhibition of VTP autoactivation.

LIST OF REFERENCES

REFERENCES

Anderson, R. M. ,and May, R. M. (1991) *Infectious Diseases of Humans*, Oxford University Press. Oxford, UK.

Baert, J. L., Britel, M., Slomianny, M. C., Delbart, C., Fournet, B., Sautiere, P., and Malecha, J. (1991) *Eur. J. Biochem.* **201,** 91-198.

Baker, M. E. (1988a) *Biochem. J.* **255,** 1057-1060.

Baker, M. E. (1988b) *Biochem. J.* **256,** 1059-1063.

Barr, P. J. (1991) *Cell* **66,** 1-3.

Barrett, J. M., Abramoff, P., Kumaran, A. K., and Millington, W. F. (1986) *Biology*. Prentice-Hall, Englewood Cliffs, NJ

Barrett, A. J., and Kirschke, H. (1981) *Methods Enzymol.* **80,** 535-561.

Béchet, D. M., Ferrara, M. J., Mordier, S. B., Roux, M. P., Deval, C. D., and Obled, A. (1991) *J. Biol. Chem.* **266,** 14104-14112.

Bek, E., and Berry, R. (1990) Biochemistry **29,** 178-183.

Bendayan, M. (1982) *J. Histochem. Cytochem.* **30,** 691-696.

Bond, J. S., and Butler, P. E. (1987) *Annu. Rev. Biochem.* **56,** 333-364.

Bownes, M., Shirras, A., Blair, M., Collins, J., and Coulson, A. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 1554-1557.

Brakch, N., Boussetta, H. Rholam, M., and Cohen, P. (1989) *J. Biol. Chem.* **264,** 15912-15916.

Brakch, N., Rholam, M., Boussetta, H., and Cohen, P. (1993) *Biochemistry* **32,** 4925-4930.

Bujo, H., Hermann, M., Kaderli, M. O., Jacobsen, L., Sugawara, S., Nimpf, J., Yamamoto, T., and Schneider, W. J. (1994) *EMBO J.* **13,** 5165-5175.

Byrne, B. M., Gruber, M., and AB, G. (1989) *Prog. Biophys. Mol. Biol.* **53,** 33-69.

Cavener, D. (1987) *Nucleic Acids Res.* **15,** 1353-1361.

Cazzulo, J. J., Couso, R., Raimondi, A., Wernstedt, C., and Hellman, U. (1989) *Mol. Biochem. Parasitol.* **33,** 33-42.

Chan, S. J., San Segundo, B., McCormick, M. B., and Steiner, D. F. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 7721-7725.

Chen, J. S., Cho, W. L., and Raikhel, A. S. (1994) J. Mol. Biol. **237,** 641-647.

Chen, J. S., Hays, A. R., Snigirevskaya, E. S., and Raikhel, A. S. (1995) *Submitted.*

Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J., and Rutter, W. J. (1979) *Biochemistry* **18,** 5294-5299.

Cho, W. L., Deitsch, K. W., and Raikhel, A. S. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 10821-10824.

Cho, W. L., and Raikhel, A. S. (1992) *J. Biol. Chem.* **267,** 21823-21829.

Chomczynski, P., and Sacchi, N. (1987) *Anal. Biochem.* **162,** 156-159.

Chothia, C., Lesk, A., Trammontano, A., Levitt, M., Smith, G. S., Air, G., Sheriff, S., Padlan, E., Davies, D., Tulip, W., Colman, P., Spinelli, S., Alzari, P., and Poljak. R, (1989) *Nature* **342,** 877-883.

Chou, P. Y., and Fasman, G. D. (1978) *Ann. Rev. Biochem.* **47,** 251-276.

Collins, T., Ginsburg, D., Boss, J. M., Orkin, S. H., and Pober, J. S. (1985) *Nature* **316,** 748-750.

Cox, G. N., Pratt, D., Hageman, R., and Boisvenue, R. J. (1990) *Mol. Biochem. Parasitol.* **41,** 25-34.

Davis, B. J. (1964) *Ann. N. Y. Acad. Sci.* **121,** 404-427.

Deitsch, K. W., and Raikhel, A. S. (1993) *Insect Mol. Biol.* **2,** 205-213.

Deitsch, K. W., Chen, J. S., and Raikhel, A. S. (1995) *Insect Biochem. Mol. Biol.* **25,** 449-454.

Devereux, J., Haeberli, P., and Smithies, O. (1984) *Nucleic Acids Res.* **12,** 387-395.

Dhadialla, T. S., and Raikhel, A. S. (1990) *J. Biol. Chem.* **265,** 9924-9933.

Dhadialla, T. S., Hays, A. R., and Raikhel, A. S. (1992) *Insect Biochem. Mol. Biol.* **22,** 803-816.

Dittman, F., Kogan, P. H., and Hagedorn, H. H. (1989) *Arch. Insect Biochem. Physiol.* **12,** 133-143.

Doolittle, R. F. (1986) *Of Urfs and Orfs: a primer on how to analyze derived amino acid sequences.* University Science Books, Mill Valley, CA.

Dufour, E. (1988) *Biochimie* **70,** 1335-1342

Ebina, Y., Ellis, L., Jarnagin, K., Edery, M., Graf, L., Clauser, E., Ou, J. H., Masiarz, F., Kan, Y. W., Goldfine, I. D., Roth, R. A., and Rutter, W. J. (1985) *Cell* **40,** 747-758.

Emini, E. A., Hughes, J. V., Perlow, D. S., and Boger, J. (1985) *J. Virol.* **55,** 836-839.

Fagotto, F. (1990a) *Arch. Insect Biochem Physiol.* **14,** 217-235.

Fagotto, F. (1990b) *Arch. Insect Biochem Physiol.* **14,** 237-252.

Fagotto, F., and Maxfield, F. R. (1994) *J. Cell Biol.* **125,** 1047-1056.

Felleisen, R., and Klinkert, M. Q. (1990) *EMBO J.* **9,** 371-377.

Frohman, M. A., Dush, M. K., and Martin, G. R. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 8998-9002.

Fujii, T., Sakurai, H., Izumi, S., and Tomino, S. (1989) *J. Biol. Chem.* **264,** 11020-11025.

Gerber-Huber, S., Nardelli, D., Haefliger, J. A., Cooper, D. N., Givel, F., Germond, J. E., Engel, J., Green, N. M., and Wahli, W. (1987) *Nucleic Acid Res.* **15,** 4737-4760.

Hagedorn, H. H. (1985) In *Comprehensive Insect Physiology, Biochemistry and Pharmacology*, (Kerkut, G. A., and Gilbert, L. I., eds). Vol. 8, pp. 205-261. Pergamon Press, Oxford.

Hagedorn, H. H., Turner, S., Hagedorn, E. A., Pontecorvo, D., Greenbaum, P., Pfeiffer, D., Wheelock, D., and Flanagan, T. R. (1977) *J. Insect Physiol.* **23,** 203-206.

Hamblin, M. T., Marx, J. L., Wolfner, M. F., and Hagedorn, H. H. (1987) *Mem. Inst. Oswaldo Cruz.* **82 (Suppl. III),** 109-114.

Hatsuzawa, K., Murakami, K., and Nakayama, K. (1992) *J. Biochem.* **111,** 296-301.

Hayflick, J. S., Wolfgang, W. J., Forte, M. A., and Tomas, G. (1992) *J. Neurosci.* **12,** 705-717.

Hays, A. R., and Raikhel, A. S. (1990) *Roux's Arch Dev. Biol.* **199,** 114-121.

Heilmann, L. J., Trewitt, P. M., and Kumaran, A. K. (1993) *Archs. Insect Biochem. Physiol.* **23,** 125-134.

Hiremath, S., Lehtoma, K., and Nagarajan, M. (1994) *J. Insect Physiol.* **40,** 813-821.

Hosaka, M., Nagahama, M., Kim, W. S., Watanabe, T., Hatsuzawa, K., Ikemizu, J., Murakami, K., and Nakayama, K. (1991) *J. Biol. Chem.* **266,** 12127-12130.

Hotchkin, P. G., and Fallon, A. M. (1987) *Biochim. Biophys. Acta* **924,** 352-359.

Huttner, W. B. (1987) *Trends Biochem. Sci.* **12,** 361-363.

Jones, G., Brown, N., Manczak, M., Hiremath, S., and Kafatos, F. C. (1990) *J. Biol. Chem.* **265,** 8596-8602.

Kageyama, Y., Kinoshita, T., Umesono, Y., Hatakeyama, M., and Oishi, K. (1994) *Insect Biochem. Mol. Biol.* **24,** 599-605.

Kiefer, M. C., Saphire, A. C. S., Bauer, D. M., and Barr, P. J. (1990) *Nucleic Acid Res.* **18,** 1909.

Klinkert, M. Q., Felleisen, R., Link, G., Ruppel, A. and Beck, E. (1989) *Mol. Biochem. Parasitol.*, **33,** 113-122.

Koller, C. N., Dhadialla, T. S., and Raikhel, A. S. (1989) *Insect Biochem.* **19,** 693-702.

Kozak, M. (1986) *Cell* **44,** 283-292.

Kuenzel, E. A., Mulligan, J. A., Sommercorn, J., and Kreb, E. G. (1987) *J. Biol. Chem.* **262,** 9136-9140.

Kunkel, J. G., and Nordin, J. H. (1985) In *Comprehensive Insect Physiology, Biochemistry and Pharmacology*, (Kerkut, G. A., and Gilbert, L. I., eds). Vol. 1, pp. 83-111. Pergamon Press, Oxford.

Kurata, S., Saito, H., and Natori, S. (1992a) *Dev. Biol.* **153,** 115-121.

Kurata, S., Saito, H., and Natori, S. (1992b) *Eur. J. Biochem.* **204,** 911-914.

Kyte, J., and Doolittle, R. F. (1982) *J. Mol. Biol.* **157,** 105-132.

Laemmli, U. K. (1970) *Nautre* **227,** 680-685.

LaFleur,Jr. G. J., Byrne, B. M., Kanungo, J., Nelson, L. D., Greenberg, R. M., and Wallace, R. A. (1995) *J. Mol. Evol.* (in press)

Mach, L., Schwihla, H., Stüwe, K., Rowan, A. D., Mort, J. S., and Glössl, J. (1993) Biochem. J. **293,** 437-442.

Mach, L., Mort, J. S., and Glössl, J. (1994a) *J. Biol. Chem.* **269,** 13030-13035.

Mach, L., Mort, J. S., and Glössl, J. (1994b) *J. Biol. Chem.* **269,** 13036-13040.

Mach, L., Schwihla, H., Stüwe, K., Rowan, A. D., Mort, J. S., and Glössl, J. (1993) *Biochem. J.* **293,** 437-442.

Mach, L., Stüwe, K., Hagen, A., Ballaun, C., and Glössl, J. (1992) *Biochem. J.* **282,** 577-582.

Marshall, R. D. (1972) *Annu. Rev. Biochem.* **41,** 673-702.

McClelland, G. A. H. (1992) *Medical Entomology: an Ecological Perspective*, University of California Press, Davis, California, USA.

Medina, M., and Valljo, C. G. (1989) *Dev. Growth Differ.* **31,** 241-247.

Medina, M., Leon, P., and Vallejo, C. G. (1988) *Arch. Biochem. Biophy.* **263,** 355-363.

Meggio, F., and Pinna, L. A. (1988) *Biochim. Biophys. Acta* **971,** 227-231.

Meloun, B., Baudyš, M., Pohl, J., Pavlik, M., and Kostka, V. (1988) *J. Biol. Chem.* **263,** 9087-9093.

Miller, M. S., Benore-Parsones, M., and White, H. B. (1982) *J. Biol. Chem.* **257,** 6818-6824.

Molloy, S. S., Thomas, L., VanSlyke, J. K., Stenberg, P. E., and Thomas, G. (1994) *EMBO J.* **13,** 18-33.

Mort, J. S., and Recklies, A. D. (1986) *Biochem. J.* **233,** 57-63.

Musil, D., Zucic, D., Turk, D., Engh, R. A., Mayr, I., Huber, R., Popovic, T., Turk, V., Towatari, T., Katunuma, N., and Bode, W. (1991) *EMBO J.* **10,** 2321-2330.

Nakagawa, T., Murakami, K., and Nakayama, K. (1993) *FEBS Lett.* **327,** 165-171.

Nardelli, D., Gerber-Huber, S., van het Schip, F. D., Gruber, M., AB, G., and Wahli, W. (1987) *Biochemistry* **26,** 6397-6402.

Naumann, N., and Scheller, K. (1991) *Biochem. Biophys. Res. Commun.* **177,** 963-972.

Nordin, J. H., Beaudoin, E. L., and Liu, X. (1991) *Arch. Insect Biochem. Physiol.* **18,** 177-192.

Ornstein, L. (1964) *Ann. N. Y. Acad. Sci.* **121,** 321-349.

Perez, L. E., Fenton, M. J., and Callard, I. P. (1991) *Comp. Biochem. Physiol.* **100B,** 821-826.

Poole, A. R., Tiltman, K. J., Recklies, A. D., and Stoker, T. A. M. (1978) *Nature* **273,** 545-547.

Pratt, D., Boisvenue, R. J., and Cox, G. N. (1992) *Mol. Biochem. Parasitol.* **56,** 39-48.

Pratt, D., Cox, G. N., Milhausen, M. J., and Boisvenue, R. J. (1990) *Mol. Biochem. Parasitol.* **43,** 181-192.

Qian, S. W., Kondaiah, P., Roberts, A. B., and Sporn, M. B. (1990) *Nucleic Acid Res.* **18,** 3059.

Qian, F., Frankfater, A., Chan, S. J., and Steiner, D. F. (1991) *DNA Cell Biol.* **10,** 159-168.

Raag, R., Appelt, K., Xuong, N. H., and Banaszak, L. (1988) *J. Mol. Biol.* **200,** 553-569.

Raikhel, A. S. (1986a) *J. Insect Physiol.* **32,** 597-604.

Raikhel, A. S. (1986b) *Tissue Cell* **18,**125-142.

Raikhel, A. S. (1992) *Adv. Disease Vector Res.* **9,** 1-39.

Raikhel, A. S., and Bose, S. G. (1988) *Insect Biochem.* **18,** 565-576.

Raikhel, A. S., and Dhadialla, T. S. (1992) *Ann. Rev. Entomol.* **37,** 217-251.

Raikhel, A. S., and Lea, A. O. (1990) *Gen. Comp. Endocrinol.* **77,** 423-434.

Raikhel, A. S., Pratt, L. H., and Lea, A. O. (1986) *J. Insect Physiol.* **32,** 879-890.

Raikhel, A. S. (1987) *Tissue Cell* **19,** 515-29.

Raikhel, A. S., and Dhadialla, T. S. (1992) *Ann. Rev. Entomol.* **37,** 217-251.

Raikhel, A. S., and Lea, A. O. (1991) *Tissue Cell* **23,** 577-591.

Ray, C., and McKerrow, J. H. (1992) *Mol. Biochem. Parasitol.* **51,** 239-250.

Retzek, H., Steyrer, E., Sanders, E. J., Nimpf, J., and Schneider, W. J. (1992) *DNA Cell Biol.* **11,** 661-672.

Retzek, H., Steyrer, E., Sanders, E. J., Nimpf, J., and Schneider, W. J. (1992) *DNA Cell Biol.* **11,** 661-672.

Rholam, M., Nicolas, P., and Cohen, P. (1986) *FEBS Lett.* **207,** 1-6.

Ribolla, P. E., Daffre, S., and de Bianchi, A. G. (1993) *Insect Biochem. Mol. Biol.* **23,** 217-223.

Rina, M., and Savakis, C. (1991) *Genetics* **127,** 769-780.

Ring, C., Kneller, D., Langridge, R., and Cohen, F. 1992. *J. Mol. Biol.* **224,** 685-699.

Robertson, C. D., and Coombs, G. H. (1993) *Mol. Biochem. Parasitol.*, **62,** 271-280.

Roebroek, A. J. M., Ayoubi, T. A. Y., Creemers, J. W. M., Pauli, I. G. L., and Van de Ven, W. J. M. (1995) *DNA Cell Biol.* **14,** 223-234.

Roebroek, A. J. M., Creemers, J. W. M., Pauli, I. G. L., Kurzik-Dumke, U., Rentrop, M., Gateff, E. A. F., Leunissen, J. A. M., and Van de Ven, W. J. M. (1992) *J. Biol. Chem.* **267,** 17208-17215.

Roebroek, A. J. M., Pauli, I. G. L., Zhang, Y., and Van de Ven, W. J. M. (1991) *FEBS Lett.* **289,** 133-137.

Rosenthal, P. J., McKerrow, J. H., Rasnick, D., and Leech, J. H. (1989) *Mol. Biochem. Parasitol.* **35,** 177-184.

Rowan, A. R., Mason, P., Mach, L., and Mort, J. S. (1992) *J. Biol. Chem.* **267,** 15993-15999.

Sakurai, H., Fujii, T., Isumi, S., and Tomino, S. (1988) *J. Biol. Chem.* **263,** 7876-7880.

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual,* 2nd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Sanger, F, Nicklen, S., and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463-5467.

Sappington, T. W., Hays, A. R., and Raikhel, A. S. (1995) *Insect Biochem. Mol. Biol. In press.*

Scheele, G. (1984) *Methods Enzymol.* **96,** 94-111.

Scheller, K., Fischer, B., and Schenkel, H. (1990) Molecular properties, function and developmentally regulated biosynthesis of arylphorin in *Calliphora vicina*. In *Molecular Insect Science,* (Hagedorn, H. H. *et al.* eds) pp.155-162. Plenum Press, New York.

Shapiro, A. B., Wheelock, G. D., Hagedorn, H. H., Baker, F. C., Tsai, L. W., and Schooley, D. A. (1986) *J. Insect Physiol.* **32,** 867-877.

Sharrock, W. J. (1984) *J. Mol. Biol.* **174,** 419-431.

Sharrock, W. J., Rosenwasser. T. A., Gould, J., Knott, J., Hussey, D., Gordon, J. I., and Banaszak, L. 1992. *J. Mol. Biol.* **226,** 903-907.

Shyu, A. B., Blumenthal, T., and Raff, R. A. (1987) *Nucleic Acids Res.* **15,**10405-10417.

Siezen, R. J., Creemers, J. M. S., and Van de Ven, W. J. M. (1994) *Eur. J. Biochem.* **222,** 255-266.

Sloane, B. F., Dunn, J. R., and Honn, K. V. (1981) *Science* **212,** 1151-1153.

Sloane, B. F., Moin, K., Krepela, E., and Rozhin, J. (1990) *Cancer Metastasis Rev.* **9,** 333-352.

Smit, A. B., Spijker, S., Nagle, G. T., Knock, S. L., Kurosky, A., and Geraerts, W. P. M. (1994) *FEBS Lett.* **343,** 27-31.

Sneyers, M., Kettmann, R., Massart, S., Renaville, R., Burny, A., and Portetelle, D. (1991) *DNA Seq.* **1,** 405-406.

Spieth, J., Denison, K., Zucker, E., and Blumenthal, T. (1985) *Nucleic Acid Res.* **13,** 7129-7137.

Spieth, J., Nettleton, M., Zucker-Aprison, E., Lea, K., and Blumenthal, T. (1991) *J. Mol. Evol.* **32,** 429-438.

Sturchler, D. (1989) *Parasitol. Today* **5,** 39-40.

Takahashi, N., Kurata, S., and Natori, S. (1993) *FEBS Lett.* **334,** 153-157.

Takio, K., Towatari, T., Katunuma, N., Teller, D. C., and Titani, K. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 3666-3670.

Telfer, W. H., and Kunkel, J. G. (1991) *Ann. Rev. Entomol.* **36,** 205-228.

Terpstra, P., and AB, G. (1988) *J. Mol. Biol.* **202,** 663-665.

Terra,W.R., Ferreira,C., and Garcia,E.S. (1988) *Insect Biochem.*, **18,** 423-434.

Tom, M., Fingerman, M., Hayes, T. K., Johnson, V., Kerner, B., and Lubzens, E. (1992) *Comp. Biochem. Physiol.* **102B,** 483-490.

Trewitt, P. M., Heilmann, L. J., Degrugillier, S. S., and Kumaran, A. K. (1992) *J. Mol. Evol.* **34,** 478-492.

van het Schip, F., Samallo, J., Broos, J., Ophuis, J., Mojet, M., Gruber, M., and AB, G. (1987) *J. Mol. Biol.* **196,** 245-260.

Voit, R., and Feldmaier-Fuchs, G. (1990) *J. Biol. Chem.* **265,** 19447-19452.

Wahli, W. (1988) *Trends Genet.* **4,** 227-232.

Willott, E., Wang, X, Y., and Wells. M. A. (1989) *J. Biol. Chem.* **264,** 19052-19059.

Yamaguchi, K., Kikuchi, Y., Takagi, T., Kikuchi, A., Oyama, F., Shimura, K., and Mizuno, S. (1989) *J. Mol. Biol.* **210,** 127-139.

Yamamoto, Y., and Takahashi, S. Y. (1993) *Comp. Biochem. Physiol.* **106B,** 35-45.

Yamamoto, Y., Takimoto, K., Izumi, S., Toriyama-Sakurai, M., Kageyama, T., and Takahashi, S. Y. (1994a) *J. Biochem.* **116,** 1330-1335.

Yamamoto, Y., Zhao, X., Suzuki, A. C., and Takahashi, S. Y. (1994b) *J. Insect Physiol.* **40,** 447-454.

Yano, K., Sakurai, M. T., Izumi, S., and Tomino, S. (1994) *FEBS Lett.* **356,** 207-211.

Yano, K., Sakurai, M. T., Watabe, S., Izumi, S., and Tomino, S. (1994a) *Biochim. Biophys. Acta* **1218,** 1-10.

Yoshizaki, N., and Yonezawa, S. (1994) *Develop. Growth Differ.* **36,** 299-306.