Archives of Insect Biochemistry and Physiology 39:98–108 (1998)

5′ Coding Region of the Follicular Epithelium Yolk Polypeptide 2 cDNA in the Moth, *Plodia interpunctella*, Contains an Extended Coding Region

Paul D. Shirk* and O. P. Perera

Center for Medical, Agricultural and Veterinary Entomology, Agricultural Research Service, U. S. Department of Agriculture, Gainesville, Florida

The 5′ region of YP2 cDNA, a follicular epithelium yolk protein subunit in the moth, *Plodia interpunctella*, shows that the polypeptide contains an extended internal coding region. Partial cDNA clones for YP2 were isolated from a pharate adult female ovarian cDNA expression library in Lambda Zap II by screening with antigen selected YP2 antiserum. The 5′ sequence of the YP2 transcript was determined by 5′ RACE PCR of ovarian mRNA using YP2 sequence-specific nested primers. The combined cDNA and 5′ RACE sequencing showed the YP2 transcript to be 1971 bp in length up to the poly(A) tail with a single open reading frame for a predicted polypeptide of 616 amino acids. Northern analysis showed a single YP2 transcript to be present in ovarian RNA that was approximately 2 kb in length. The predicted amino acid sequence for YP2 from *P. interpunctella* is most closely related to egg specific protein (ESP) from *Bombyx mori* and the partial YP2 sequence from *Galleria mellonella*. YP2 from *P. interpunctella* also is similar to vertebrate lipases and contains a conserved lipid binding region. However, the 5′ coding region of YP2 from *P. interpunctella* contains an in-frame insert of approximately 438 bp that had replaced an approximately 270-bp region as compared with ESP from *B. mori* and YP2 of *G. mellonella*. This suggests that the insert occurred by a recombinational event internal to the YP2 structural gene of *P. interpunctella*. Arch. Insect Biochem. Physiol. 39:98–108, 1998.

Key words: oocyte; ovary; reproduction; lipase; recombination

Abbreviations used: bp = base pair; ESP = egg specific protein; FEYP = follicular epithelium yolk protein; kb = kilobase pair; YP2 = yolk polypeptide 2.

*Correspondence to: Paul D. Shirk, USDA ARS CMAVE, P.O. Box 14565, Gainesville, FL 32604. E-mail: pshirk@gainesville.usda.ufl.edu

The sequence reported in this paper has been deposited in the GenBank database (accession no. AF063014)

Received 10 June 1998; Accepted 22 September 1998

INTRODUCTION

In some Lepidoptera, the follicular epithelial cells that surround the oocyte produce and secrete proteins that become a major component of proteinacious yolk that is packaged in the oocytes during vitellogenesis. These follicular epithelium produced yolk proteins are of two types with either a homo- or hetero-subunit composition. The homo-subunit type, such as paravitellogenin (70 kDa) from *Hyalophora cecropia* (Bast and Telfer, 1976; Telfer et al., 1981) and egg-specific protein (ESP at 72 kDa) from *Bombyx mori* (Ono et al., 1975; Irie and Yamashita, 1983; Zhu et al., 1986; Indrasith et al., 1988a), consist of a single subunit that associates as dimers or tetramers. The hetero-subunit type, such as follicular epithelium yolk protein (FEYP) with subunits YP2 (69 kDa) and YP4 (33 kDa) from *Plodia interpunctella* (Shirk et al., 1984; Bean et al., 1988) and other pyralid moths (Shirk, 1987), have two subunits that associate in dimeric and tetrameric forms. Typically, the FEYPs contribute approximately 25–40% of the proteinacious material in the oocytes and are the first to be degraded among the yolk proteins during embryogenesis. The digestion of ESP in *B. mori* has been shown to be carried out by a highly specific trypsin-like protease that appears early during embryogenesis (Indrasith et al., 1987, 1988b). Proteolytic cleavage of the YP2 subunit of FEYP from *P. interpunctella* has also been observed (Bean et al., 1988).

In *H. cecropia*, *B. mori,* and *P. interpunctella* as well as other Lepidoptera, transplantation of pre-metamorphic ovaries into males has been shown to result in the formation of normal eggs during pharate adult development but without the deposition of vitellin (Telfer, 1954; Yamashita and Irie, 1980; Indrasith et al., 1988a; Zimowska et al., 1995). In an elegant set of experiments, the mature eggs from ovaries that had developed in male *B. mori* were stimulated to parthenogenically initiate embryogenesis. These embryos completed normal embryonic and post-embryonic development even though they lacked vitellin (Yamashita and Irie, 1980). ESP is the predominant protein component of the yolk in eggs developed in males, which suggests that ESP is present in sufficient quantities to provide the protein resources for embryogenesis.

Cloning of cDNA and genomic sequences for ESP from *B. mori* have been reported (Inagaki and Yamashita, 1989; Sato and Yamashita, 1991). The predicted amino acid sequence for ESP showed limited similarity with lipases from vertebrates and included a partially conserved lipid binding site. The proteolytic cleavage sites were present and the amino terminal sequence that is removed from ESP during embryogenesis by the trypsin-like protease showed that it contained two long stretches of polyserine. However, the nature of the function of this protein during embryogenesis was not clear from the derived sequence. We set about to clone the cDNA for YP2 and compare the predicted sequence structures with those of ESP because of the similarities in size, yolk sphere deposition, and embryonic proteolytic cleavage and because there are differences in the subunit binding properties between ESP from *B. mori* and YP2 from *P. interpunctella*. This work shows that YP2 has similarity with ESP and with lipases from vertebrates. However, the amino terminal region of YP2 from *P. interpunctella* has limited similarity with the amino terminal region of ESP because of the presence of a long insert.

MATERIALS AND METHODS

The *Plodia interpunctella* colony was reared according to Silhacek and Miller (1972) in a 16 h light:8 h dark cycle at 30EC and 70% relative humidity. All molecular biological procedures were conducted as described in Sambrook et al. (1989) unless otherwise stated.

cDNA Library Construction

Total RNA was extracted from the ovaries of late pharate adult females using an RNA Extraction Kit (Pharmacia Biotech, Piscataway, NJ). To construct the cDNA library, mRNA was isolated from the total RNA by two cycles of chromatography through an oligo(dT)-cellulose column.

The Library Construction Core Laboratory, Interdisciplinary Center for Biotechnology Research, University of Florida, performed the cDNA library construction. The cDNA library was constructed from poly(A)$^+$ RNA of the ovaries. cDNA synthesis was performed with the Lambda ZAP II cDNA Synthesis Kit (Stratagene, La Jolla, CA) using an oligo (dT) primer with an *Xho*I linker. The cDNA was size-fractionated on a Sephacryl S-400 column with a minimum size cutoff of 0.5 kb. An *Eco*RI adaptor was added to the cDNA, and then the DNA was restriction digested with *Eco*RI and *Xho*I. The cDNA was then ligated into Lambda ZAP II arms that were restriction digested with *Eco*RI and *Xho*I. The ligation was

packaged into lambda phage according to the manufacturer's protocols.

Screening of the cDNA Library

Affinity purified polyclonal antibodies for YP2 (Bean et al., 1988) were prepared as described by Robinson et al. (1988). The cDNA library was screened with the affinity purified antibodies by the plaque method of Huynh et al. (1985). Positive plaques were rescreened 3 times before excision of the recombinant pBluescript II (SK^+) plasmid according to the manufacturer's protocols.

Amino Acid Sequence Analysis

The FEYP of *P. interpunctella* was purified as described previously (Bean et al., 1988) and resolved by SDS-PAGE. The protein was electroblotted to Trans-Blot membrane (BioRad, Hercules, CA) and the YP2 band was cut from the total membrane. The Protein Chemistry Core Facility, Interdisciplinary Center for Biotechnology Research, University of Florida, determined the NH_2-terminal amino acid sequence of YP2 by automated dansyl-Edman degradation using an online sequencing analyzer (Model 470A, Applied Biosystems, Foster City, CA) as described by Walker (1984). The derived amino acid sequence of YP2 was used to confirm the identity of the putative YP2 cDNA clones.

DNA Sequencing

The DNA Sequencing Core Laboratory, Interdisciplinary Center for Biotechnology Research, University of Florida, performed the DNA sequencing. The sequencing strategy is shown in Figure 1a using initially the M13 primer sites within the pBluescript II (SK+) plasmid and subsequently selected primers identified from the derived sequences. The sequence of the 5′ end of the YP2 transcript was determined by 5′ RACE essentially as directed by the manufacturer (Life Technologies, Gaithersburg, MD). The 5′ RACE utilized the YP2R749 ($^{5'}$GTAGCGTTCAGTCTTGCGTCTT$^{3'}$) reverse primer and the YP2R696 ($^{5'}$TTCGTGGTATTTCTCCTTAGTT$^{3'}$) reverse primer as the primary and nested primers, respectively. The 5′ RACE PCR products were TA cloned into *pCR2.1* as described by the manufacturer (Invitrogen, Carlsbad, CA) and sequenced. The sequencing of all cDNA and 5′ RACE products was accomplished by the Taq DyeDeoxy Terminator (no. 401388) and DyePrimer (no. 401386) Cycle Sequencing protocols developed by Applied Biosystems (a division of Perkin-Elmer Corp., Foster City, CA) using fluorescent-labeled dideoxynucleotides and primers, respectively. The labeled extension products were analyzed on a Model 373A DNA Sequencer (Applied Biosystems).

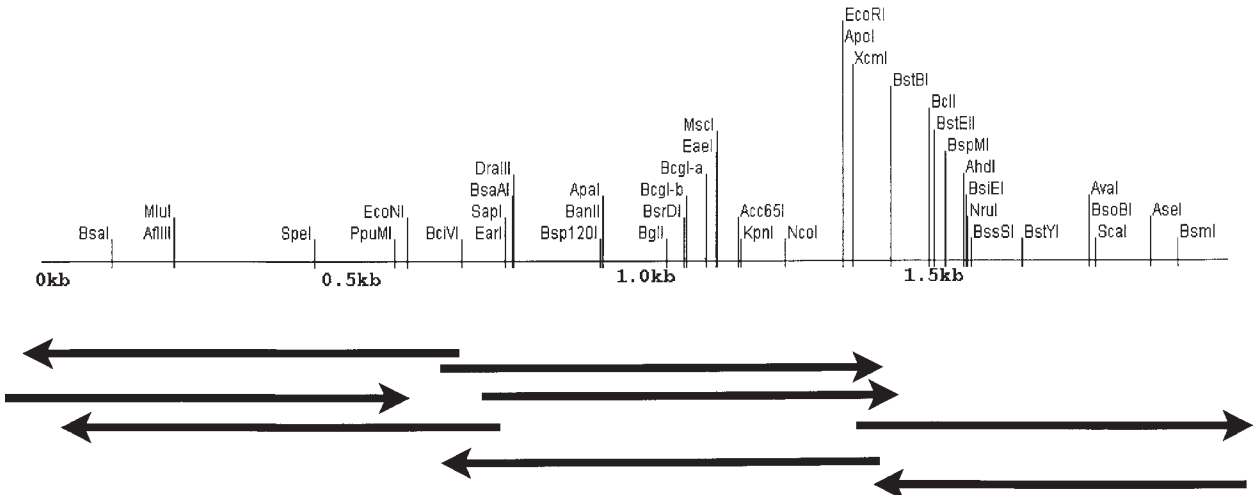
The DNA and the conceptual translation sequences were used to search peptide and nucleotide sequence databases using the BLAST (Altschul et al., 1990) network service. Alignments for the amino acid sequences and the Kyte-Doolittle hydropathy plots (Kyte and Doolittle, 1982) were produced using GeneWorks 2.5.1 software (Oxford Molecular Group, Campbell, CA) and adjusting for known sequence similarities.

Northern Blot Analysis

Total RNA from ovaries and somatic tissues of pharate adults was isolated using TRIZOL reagent as directed by the manufacturer (Life Technologies, Gaithersburg, MD). Northern analysis was carried out using NorthernMax (Ambion, Austin, TX) reagents. Five micrograms of total RNA of each sample was resolved in a 1% formaldehyde-agarose gel. The gel was run for 4 h at 60 V with buffer recirculation. RNA Millenium marker (Ambion, Austin, TX) was included in the gel to determine the size of the target mRNA. At the end of the run the marker lane was cut out, stained with ethidium bromide, and photographed. The RNA was then transferred onto a Nytran (Schleicher & Schuell, Keene, NH) blotting membrane by downward capillary transfer using the rapid transfer buffer of the NorthernMax reagent system. After 1 hour, the transfer apparatus was disassembled, the membrane was washed in the transfer buffer briefly and UV cross linked using a Stratalinker (Strategene, LaJolla, CA).

Prehybridization and hybridization of the Northern blot was carried out at 48°C in the NorthernMax hybridization buffer. The probe was synthesized by amplification of a 696 bp fragment from the YP2 5′ RACE clone. The amplification reaction contained 10 pg of template DNA, 3 μl of alpha ^{32}PdCTP (10 μCi/mmol), 10 pmol each of PiYP2R696 (reverse) and the 5′ RACE universal amplification primer (UAP) primers, 100 nmol each of dATP, dGTP, and dTTP, and 2.5 mM $MgCl_2$ in a 15 μl volume. Cycling parameters were the same as for the amplification step in the 5′ RACE procedure except that only 20 cycles of amplification were performed. Probe was purified from unincorporated nucleotides and primers using a BioGel P60 (100–200 mesh) column. Heat denatured probe was added to the hybridization buffer and incubated overnight.

A



B

```
     GACTTTTTTCAACAAAATGAAGGCAGTCTGGCTATTATTCTGCCTTGGGCTGGCCCGCGGGTCCATCTTCCGTCGGCACTATCTCCCTCAAATGGAGAGG   100
1                    M  K  A  V  W  L  L  F  C  L  G  L  A  R  G  S  I  F  R  R  H  Y  L  P  Q  M  E  R
     TTCGCCCAGCGACTGGACTCGGTGGAGACCCCACGGGAAAACCCACAGTTGGTCAGGCACAGCATCGTCTACGACCCCACCCCGGCGGCCTCCGGCATC   199
29   F  A  Q  R  L  D  S  V  E  T  P  R  E  N  P  Q  L  V  R  H  S  I  V  Y  D  P  T  P  A  A  S  G  I
     CAGCAAAACTTGGAGGAAATCAGACGCGTCAACGCTGACTACTACTCCCCCAACCAGCTGTATAACAAGAAACACATGAACAGGAAACTATCCAAACCC   298
62   Q  Q  N  L  V  E  I  R  R  V  N  A  D  Y  Y  S  P  N  Q  L  Y  N  K  K  H  M  N  R  K  L  S  K  P
     CTTCAAGATTTCAACGAAGACTTGGAAAATTACGCTGAGGGAGAATACTACCAAATCCCCCAGTATGCTGACCAGGAGCGCGTCTTCACCGTCAACAAG   397
95   L  Q  D  F  N  E  D  L  E  N  Y  A  E  G  E  Y  Y  Q  I  P  Q  Y  A  D  Q  E  R  V  F  T  V  N  K
     AGGCTATCAAGACCGACCCAGGCTGCCATATCTAATGACCAAGATTTGGAACAATTCAGACTAGTCAAGGGCGTGTACTATGCCATTCCGCGGTACGGA   496
128  R  L  S  R  P  T  Q  A  A  I  S  N  D  Q  D  L  E  Q  F  R  L  V  K  G  V  Y  Y  A  I  P  R  Y  G
     GATGAGAAACAGTCTTTCTCTCTCCTGAAGAAGAACACTGAACGCACGGAGACTTATGGAGATGTTGCTGATTGGAGAGGAATCAAATACGCCGTGGGA   595
161  D  E  K  Q  S  F  S  L  L  K  K  N  T  E  R  T  E  T  Y  G  D  V  A  D  W  R  G  I  K  Y  A  V  G
     CCCAACTCCAAAGTCCTCAAGAAGGACATCGAAAGTATCTTCTACGAGGCACAGAAATCCATGAAGCACCCCAGCGAGGAAACTAAGGAGAAATACCAC   694
194  P  N  S  K  V  L  K  K  D  I  E  S  I  F  Y  E  A  Q  K  S  M  K  H  P  S  E  E  T  K  E  K  Y  H
     GAAATGTACGAGTCTGTATCCGCGGAGAACATTGAAGACGCAAGACTGAACGCTACTCAGCTGTTGAACAAATACCAGTACCCTGTTGAAGAGCACGTA   793
227  E  M  Y  E  S  V  S  A  E  N  I  E  D  A  R  L  N  A  T  Q  L  L  N  K  Y  Q  Y  P  V  E  E  H  V
     GTGAGGACAGACGATGGATATTTCCTAACCTTGTTCCGCATCTCAAAGCAGACAGAGAAAGATACTACTGATGAAGTGGTCCAGAAGCCAGTCGTACTG   892
260  V  R  T  D  D  G  Y  F  L  T  L  F  R  I  S  K  Q  T  E  K  D  T  T  D  E  V  V  Q  K  P  V  V  L
     CTAATGCACAGTATGCTCGGAAGCGCCGATGACTGGCTCCTCATGGGCCCAGGTCAGTCCCTCGCCTACCTCCTAGCCGATCAGGGCTACGATGTATGG   991
293  L  M  H  S  M  L  G  S  A  D  D  W  L  L  M  G  P  G  Q  S  L  A  Y  L  L  A  D  Q  G  Y  D  V  W
     CTCGGCAACGCCCGCGGTAACAGATATACCCGTCACCACGTCAACCATCATGCCGCTAAGGCTGATTTCTGGCGGTACAGCAATGACGACATCGCCCTT  1090
326  L  G  N  S  R  G  N  R  Y  T  R  H  H  V  N  H  H  P  A  K  A  D  F  W  R  Y  S  N  D  D  I  A  L
     CACGACCTTCCTGCCATGATCGATTACGCCCTGAAGACGACTGGCCAGCGCAAGCTGTTCTACGTCGGCTACGACCTGGGTACCACTGCCTTCTTCGCC  1189
359  H  D  L  P  A  M  I  D  Y  A  L  K  T  T  G  Q  R  K  L  F  Y  V  G  Y  D  L  G  T  T  A  F  F  A
     TTGGCTTCCACTCGCCCAGAGTACAACAACAAGGTTGCCATGATGTACGCCCTTTCCCCCATGGCGTACATGAGCCACGTCCGCAGCCCGCTGGTCAAG  1288
392  L  A  S  T  R  P  E  N  N  N  K  V  A  M  M  Y  T  L  S  P  M  A  Y  M  S  H  V  R  S  P  L  V  K
     ATGATCGCTCCCGACAGTCCATTCTACAATAACCTCAAGCAGTACCTGAAGGACGGCGAATTCAAACCCAGCAAGGAGTTGGTTTACACGATGGGCGGT  1387
425  M  I  A  P  D  S  P  F  Y  N  N  L  K  Q  Y  L  K  D  G  E  F  K  P  S  K  E  L  V  Y  T  M  G  G
     GAGATGTTGGAGAACGAGATCGGATGCAAGAAGATAGCTTCGAACGTGAACTTCGTGATGTCCGGCATGAACGTGGATAACATGGACGTGAAATCGATC  1486
468  E  M  L  E  N  E  I  G  C  K  K  I  A  S  N  V  N  F  V  M  S  G  M  N  V  D  N  M  D  V  K  S  I
     CGCGTGATCATGGGTCACCTTCCAGCGGGCGGTTCCACGCGGCAGGTGCGCCAGTACGGACAGGCGGTCGCGACCCACGAGTTCAGGATGTACGACTAC  1585
491  R  V  I  M  G  H  L  P  A  G  G  S  T  R  Q  V  R  Q  Y  G  Q  A  V  A  T  N  E  F  R  M  Y  G  D
     GGCTCAGAAGTCAACCAGGAGGTGTACGGGGACCGTGTGCCGCCCGTCTATGACGTCACGAAGATCCGCACCCCGGTGGCGCTGTACTTCAGCGAGCAC  1684
524  G  S  E  V  N  Q  E  V  Y  G  D  R  V  P  P  V  Y  D  V  T  K  I  R  T  P  V  A  L  Y  F  S  E  H
     GACTGGTTGGCGCACCCCAAGGACGTCCTCAGACTGAAGGAGCAGCTCCCCAACGTAACGGAATACTACCAGGTGCCCGAGGAGTACTTCAGCCACATG  1783
557  D  W  L  A  H  P  K  D  V  L  R  L  E  E  Q  L  P  N  V  T  E  Y  Y  Q  V  P  E  E  Y  F  S  H  M
     GACTTCCTCTACTCGCAGAAGGCCCCCGTAGTTGTCTACAAGAATCTCATTAACTCCATCAATAACAATATCCATAAGAATTAATGTGTATAGGATTAT  1882
590  D  F  L  Y  S  Q  E  A  P  V  V  V  H  K  N  L  I  N  S  I  N  N  N  I  H  K  N
     TAGTATTAAAAAACCTCATATACTTAGCATTCATTATTTATCGTTGTTAATAGGTCATTGACGCATAATAAATCCGTCCGATTTTTTTACATTAAAAAA  1981
```

Fig. 1. Restriction map and sequence of the YP2 cDNA. **A:** The YP2 cDNA restriction map for endonucleases with only one restriction site and the sequencing strategy for YP2 cDNA clones. The primers used for the sequencing of the YP2 cDNA clones were initially M13 reverse and M13 forward primers originating in the Bluescript plasmid. Subsequent primers were derived from identified YP2 sequences. **B:** The nucleotide sequence and predicted amino acid sequence for the consensus YP2 cDNA. The nucleotide position is on the right, and the predicted amino acid position is on the left. The single ORF begins at bp 15 of the nucleotide sequence. The predicted amino acid sequence of the open reading frame is shown under the nucleotide sequence in the single-letter amino acid code. The single underlined sequence is identical to the 30 amino acid sequence from the NH_2-terminal of mature YP2. The bold underlined sequence shows the poly(A)$^+$ addition signal beginning at base 1,939. Only six A's of the poly(A) tail are shown.

At the end of hybridization, the probe was removed and the filter was washed twice with NorthernMax low stringency wash buffer at room temperature and then twice with high stringency wash buffer at 48°C for 15 min each. An autoradiogram of the blot was made by exposing XOMAT AR film (Eastman Kodak, Rochester, NY) for 6 hours to visualize the signal.

RESULTS

Isolation and Characterization of cDNA and the Predicted Amino Acid Sequences of YP2

In order to provide a confirmation of the cloned isolates, the first thirty amino acids of mature YP2 isolated from 4-h-old eggs were determined. The YP2 NH_2-terminal sequence, SIFRRHYLPQMERFAQRLDSVETPRENPQL, shared no similarity with any known sequences.

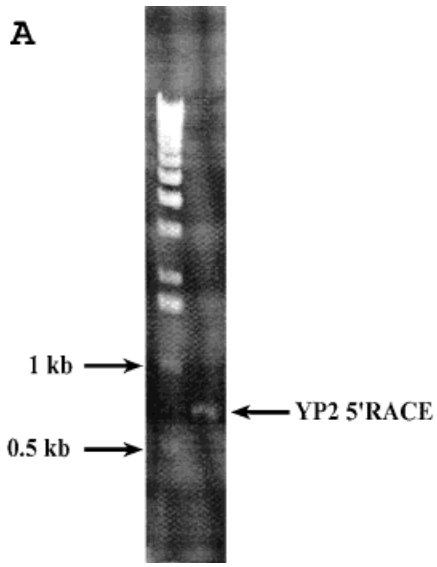
Six different positive clones of cDNA were isolated through the immunological screening of approximately 500 recombinant phage from the pharate adult ovarian cDNA library with antigen-selected YP2 antiserum. The recovery rate for YP2 clones was about 1% from this library, which suggests that the YP2 transcript is present in the ovary at about 1% of the total transcript. Following the excision of the recombinant Bluescript phagemid, each of the positive phagemids was restriction digested with *Eco*RI and *Xho*I and were found to contain an insert of approximately 2 kb (data not shown). After DNA sequencing, the 30 amino acid sequence derived from the NH_2-terminal sequencing of the mature YP2 was used to confirm that all of the clones contained YP2 cDNA. However, none of the cDNA clones isolated from the Lambda Zap II library were found to contain a full-length copy of the YP2 transcript (Fig. 2B).

In order to determine the sequence of the 5′ end of the YP2 transcript, ovarian mRNA was used to perform 5′ RACE PCR as described to produce the 5′ cDNA sequence (Fig. 2A). The YP2 5'RACE products were cloned into *pCR2.1* and sequenced. The two longest cDNA clones, YP-211AR and YP211R, were found to be lacking or incorrect in the first 12 bp of the full-length sequence (Fig. 2B). The first 12 base pairs of the 5′ end, including the T of the start codon, were evidently replaced by the *Eco*R1 linker during construction of the Lambda Zap II library. As can be seen from some the sequence of some of the cDNA clones, the reverse transcriptase terminated early leaving truncated sequences.

The complete cDNA sequence was 1,971 bp in length with a single open reading frame of 1,848 bp (bp 13 to 1,860) that encoded for a 616 amino acid polypeptide (Fig. 1B). Northern analysis showed there to be a single transcript of approximately 2 kb in length in ovarian RNA from vitellogenic late pharate adult females, and that it was not present in non-ovarian tissues of the abdomens from these females (Fig. 3). The 30 amino acid sequence derived from the NH_2-terminal sequencing of YP2 had total identity with amino acids 16 to 46 of the predicted amino acid sequence from the cDNA. The 3′ noncoding region of the cDNA was 102 bp long and contained one AATAAA poly(A)$^+$ addition signal beginning at bp 1,945, which is upstream from the start of the 3′ poly(A)$^+$ tail at base 1,971.

Sequence Similarity Between YP2 and Lepidoptera Yolk Proteins

The YP2 cDNA sequence and the predicted amino acid sequence were BLAST searched against the available gene and protein banks. The greatest sequence similarity was shared with the ESP gene from *Bombyx mori* (Inagaki and Yamashita, 1989; Sato and Yamashita, 1991) and the partial YP2 gene from *Galleria mellonella* (Accession no. U69881) (Fig. 4). Although most of the sequences were similar, the alignment of these three conceptual gene products showed a striking divergence in the 5′ translated region for the YP2 sequence. Over the first 40 amino acids, which includes the hydrophobic leader sequence from aa 1 to aa 18, the predicted amino acid sequence for ESP from *B. mori* and YP2 from *P. interpunctella* had 13 identities (33% identity) and 6 conservative substitutions (50% similarity). The partial sequence for YP2 from *G. mellonella* terminates before this region. However, for the sequences from aa 38 to aa 179 in YP2 from *P. interpunctella* and aa 40 to aa 125 in ESP from *B. mori* there was only 8% identity. A similar comparison between YP2 from *P. interpunctella* and aa 1 to aa 79 of the partial YP2 from *G. mellonella* had 12% similarity. In the subsequent 437 amino acids of YP2, there were 223 amino acid residues (51%) that were identical between the two sequences with an additional 54 conservative substitutions (63% similarity overall). Thus, the sequence data suggest that there has been a recombinational event that left a longer DNA sequence in the 5′ coding region of the YP2 gene of *P. interpunctella* than is present in the closely related genes of either ESP in *B. mori* or of YP2 in *G. mellonella* (see insertion in the alignment of Fig. 4A).



B

```
                660     |   650     |   640     |   630     |   620     |   610     |
YP2R1       <ACATTCAACA AAATGAAGGC AGTCTGGCTA TTATTCTGCC TTGGGCTGGC CCGCGGGTCC

               680       |670        |660        |650        |640        |630        |
YP2R22      <ACATTCAACA AAATGAAGGC AGTCTGGCTA TTATTCTGCC TTGGGCTGGC CCGCGGGTCC

              680         670         660         650         640         630         |
YP2R20      <ACATTCAACA AAATGAAGGC AGTCTGGCTA TTATTCTGCC TTGGGtTGGC CCGCGGGTCC

                    650|        640|        630|        620|        610|        600|
YP211AR       <ATTCggCA cgAgGAAGGC AGTCTGGCTA TTATTCTGCC TTGGGCTGGC CCGCGGGTCC

                    |630        |620        |610        |600        |590        |
YP211R        <ATTCggCA cgAgGAAGGC AGTCTGGCTA TTATTCTGCC TTGGGCTGGC CCGCGGGTCC

                100       |     110   |     120   |     130   |     140   |     150   |
YP213R          ATTCggC- --AcG-AGGC AGTCTGGCTA TTATTCTGCC TTGGntTGGC CCGCGGGTCC>

                    |  10       |  20       |  30       |  40       |  50       |
YP211RA         TTCggCA cgAgGAAGGC AGTCTGGCTA TTATTCTGCC TTGGGnTGGC CCGCnGGTCC>

                    |     10    |     20    |     30    |     40    |     50    |
YP2F1             CA cgAgGAAGGC AGTCTGGCTA TTATTCTGCC TTGGGCTGGC CCGCGGGTCC>

                    |           |   110     |   120     |   130     |   140     |
YP230R              |      GcAcG- AG-CTcGtTA ngATcCTGCC TTGGGCTGGg CcnCGGGTCC>

                    |           | 90        | 100       | 110       | 120       |
YP227R              |         GGC tG-CaGGCTA TTATTCTGCC TTGGGCTGGC CCGCGGGTCC>

                    |           |       90  |      100  |         110         120
YP220R              |           C gGgCT-GCag gaATTC-GgC acGaG--GGC CCGCGGGTCC>

                    |           |           |    630    |         620|        610|
YP220CR             |           |           |  <ATTC-GgC acGaG--GGC CCGCGGGTCC
```

Fig. 2. 5′ RACE products for YP2 from ovaian mRNA. **A:** The 5′ RACE reaction utilized the YP2R749 ($^{5'}$GTAGCGTTCAGTCTTGCGTCTT$^{3'}$) reverse primer and the YP2R696 ($^{5'}$TTCGTGGTATTTCTCCTTAGTT$^{3'}$) reverse primer as the primary and nested primers, respectively. The size marker is on the left with the bracketing sizes markers designated. **B:** Comparison of sequence from 5′ RACE products with sequences from eight cDNA clones. The sequences for the 5'RACE and cDNA clones were essentially identical following the 3′ base shown. Uppercase letters, identical matches; lowercase letters, mismatches; –, insertions; <, direction of sequencing for the clone. YP2R1 = sequence of 5′ RACE product clone 1; YP2R22 = sequence of 5′ RACE product clone 22; YP2R20 = sequence of 5′ RACE product clone 20; YP211AR = sequence of cDNA clone 11AR; YP211R = sequence of cDNA clone 11R; YP213R = sequence of cDNA clone 13R; YP211RA = sequence of cDNA clone 11RA; YP2F1 = sequence of cDNA clone F1; YP230R = sequence of cDNA clone 30R; YP227R = sequence of cDNA clone 27R; YP220R = sequence of cDNA clone 20R; YP220CR = sequence of cDNA clone 20CR.

The predicted amino acid sequence of YP2 from *P. interpunctella* also had considerable similarity with vertebrate lipases as was observed for ESP from *B. mori* (Sato and Yamashita, 1991). Over the 423 amino acid sequence downstream from the first protease digestion site (aa 176) of YP2 from *P. interpunctella*, YP2 had 134 identities (31%) plus 61 conservative substitutions (46% similarity) when compared with gastric lipase of human (Bodmer et al., 1987). However, there were two stretches that showed very high similarity between gastric lipase from human and YP2 from *P. interpunctella*: from aa 98 to 121 there was 87% similarity (74% identity) and from aa 149 to 186, that included a putative lipid binding region, there was 76% similarity (57% identity). The putative lipid binding region of gastric lipase (bold characters in Fig. 4) had 6 of 10 amino acids that were identical. In contrast to the moderate level of similarity between YP2 and the lipases from vertebrates, YP2 had less than 10% amino acid identities with the three YPs of *D. melanogaster,* and the putative lipid binding region of all three YPs of *D. melanogaster* were positioned downstream of its location in YP2, ESP and the vertebrate lipases (Fig. 4).

The physical similarity of YP2 and ESP following the first protease digestion site (after aa 176 for YP2 and aa 132 for ESP) with the vertebrate lipases was also apparent from a comparison of the hydropathy profiles (Fig. 5). A comparison of the Kyte-Doolittle plots for YP2 from *P. interpunctella* and ESP from *B. mori* showed that both proteins had nine major hydrophobic regions of similar size and position within the lipase region of the proteins. The similarity in predicted hydropathy was also shared with the partial YP2 sequence from *G. mellonella* (data not shown). The hydropathy plots

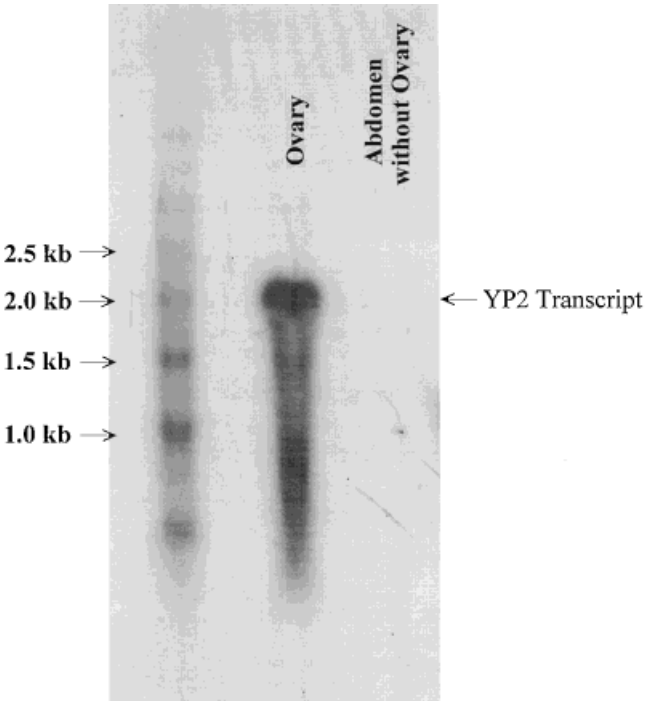


Fig. 3. Northern analysis of YP2 transcript in RNA from late pharate adult females. Five micrograms of total RNA isolated from vitellogenic ovaries of late pharate adult females and abdomens of late pharate adult females that had the ovaries removed were resolved in denaturing gel conditions and then blotted to a membrane. The blot was hybridized to labeled probe made from a 5′ RACE clone as describe. The position of the bracketing size markers is shown on the left.

were also very similar to that of the human gastric lipase. With the sequences centered on the lipid binding region, five of the nine hydrophobic regions apparent in YP2 and ESP aligned with those in human gastric lipase (Fig. 5).

DISCUSSION

A full-length cDNA sequence for the transcript of the YP2 subunit of the FEYP in *P. interpunctella* was determined. The sequence was derived from partial cDNA clones isolated from a vitellogenic pharate adult ovarian cDNA expression library and from 5′ RACE analysis. A single transcript of approximately 2 kb for this gene was detected in vitellogenic ovaries. The cDNA sequence contained a single open reading frame that encoded for a predicted protein of 616 amino acid residues. The predicted amino acid sequence had an apparent hydrophobic leader sequence of fifteen amino acid residues. To confirm the identity of the cDNA clones, the first 30 amino acid residues of the mature YP2 polypeptide isolated from the egg were sequenced and matched with the predicted amino acid sequence for YP2. The amino acids from the mature YP2 polypeptide had total identity with the sequence immediately following the hydrophobic leader of the predicted YP2 sequence.

The predicted YP2 amino acid sequence from *P. interpunctella* shared considerable similarity with the ESP from *B. mori* (Inagaki and Yamashita, 1989; Sato and Yamashita, 1991) and with the partial sequence for YP2 from *G. mellonella* (Accession no. U69881). The similarity extended to most of the sequence except for a region of the NH_2-terminus. The predicted sequences of YP2 from *P. interpunctella* and ESP from *B. mori* shared conserved post-translational protease cleavage sites at the end of the hydrophobic leader (amino acid residues 16–18; ...Ser-Ile-Phe...; bold-underlined marked residues in Fig. 4). In addition to the cleavage of the hydrophobic leader, ESP is also cleaved at two additional sites, between amino acids 132 and 133 and amino acids 228 and 229, by a trypsin-like protease during embryogenesis (Indrasith et al., 1987, 1988b; Inagaki and Yamashita, 1989). Cleavage by the ESP-specific protease results in the production of a 55 and 36 kDa hydrolytic fragment (Indrasith et al., 1988b). The double hydrolytic digestion of YP2 into two smaller fragments was also observed in embryos of *P. interpunctella* (Bean et al., 1988). In the aligned sequences of YP2 from *P. interpunctella* and *G. mellonella* and ESP from *B. mori,* arginine or lysine residues, which are sites of trypsin cleavage, are present in the YP2s within two residues of these two known trypsin-like protease cleavage sites of ESP. This suggests that these residues in YP2 of *P. interpunctella* and *G. mellonella* are probably conserved and that a similar protein specific trypsin-like protease acts to produce hydrolytic fragments from YP2 during embryogenesis.

ESP was previously shown to contain a consensus lipid binding site that is common to vertebrate lipases (Sato and Yamashita, 1991). YP2 from both *P. interpunctella* (Fig. 4B; 6 of 10 residues) and *G. mellonella* (Fig. 4B; 7 of 10 residues) also showed some conservation of the lipid binding site in human gastric lipase (Bodmer et al., 1987) although not as much as ESP (Fig. 4B; 8 of 10 residues). In addition, a large section of the sequences for these three polypeptides showed similarity with the vertebrate lipases (Fig. 4A). YP2 from *P. interpunctella* had 30% overall identities with human gastric lipase and from amino

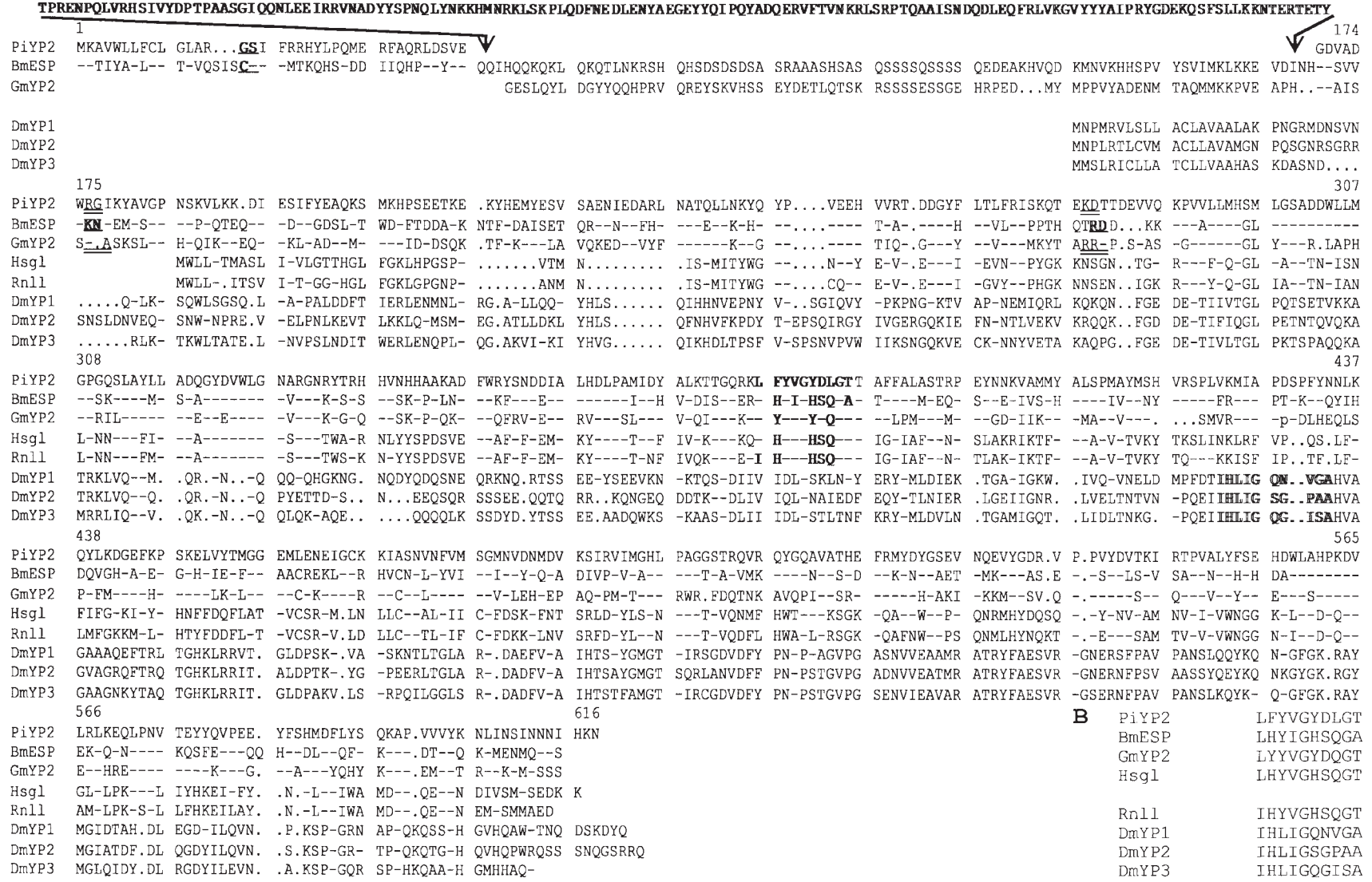


Fig. 4. **A:** The alignment of the predicted amino acid sequence of YP2 from *P. interpunctella* and related proteins. The excerpted sequence arrowed into the YP2 sequence of *P. interpunctella* shows a large non-aligned region with ESP from *B. mori* and YP2 from *G. mellonella*. The bold and underlined residues show the known sites of protease cleavage of ESP from *B. mori* and YP2 from *P. interpunctella*. The double underlined sequences show potential protease cleavage sites in YP2 from both *P. interpunctella* and *G. mellonella*. The bold residues show a conserved lipid binding site. **B:** Alignment of lipid binding sites from insects and vertebrates. . = insertions; - = identical amino acids between amino acid residues in YP2 and other amino acid sequences; BmESP: *B. mori* egg specific protein (Inagaki and Yamashita, 1989; Sato and Yamashita, 1991); DmYP1: *D. melanogaster* YP1 (Yan et al., 1987); DmYP2: *D. melanogaster* YP2 (Yan et al., 1987); DmYP3: *D. melanogaster* YP3 (Yan et al., 1987); GmYP2: *G. mellonella* YP2 (Accession no. U69881); Hsgl: human gastric lipase (Bodmer et al., 1987), PiYP2: *P. interpunctella* YP2; Rnll: rat lingual lipase (Docherty et al., 1985).

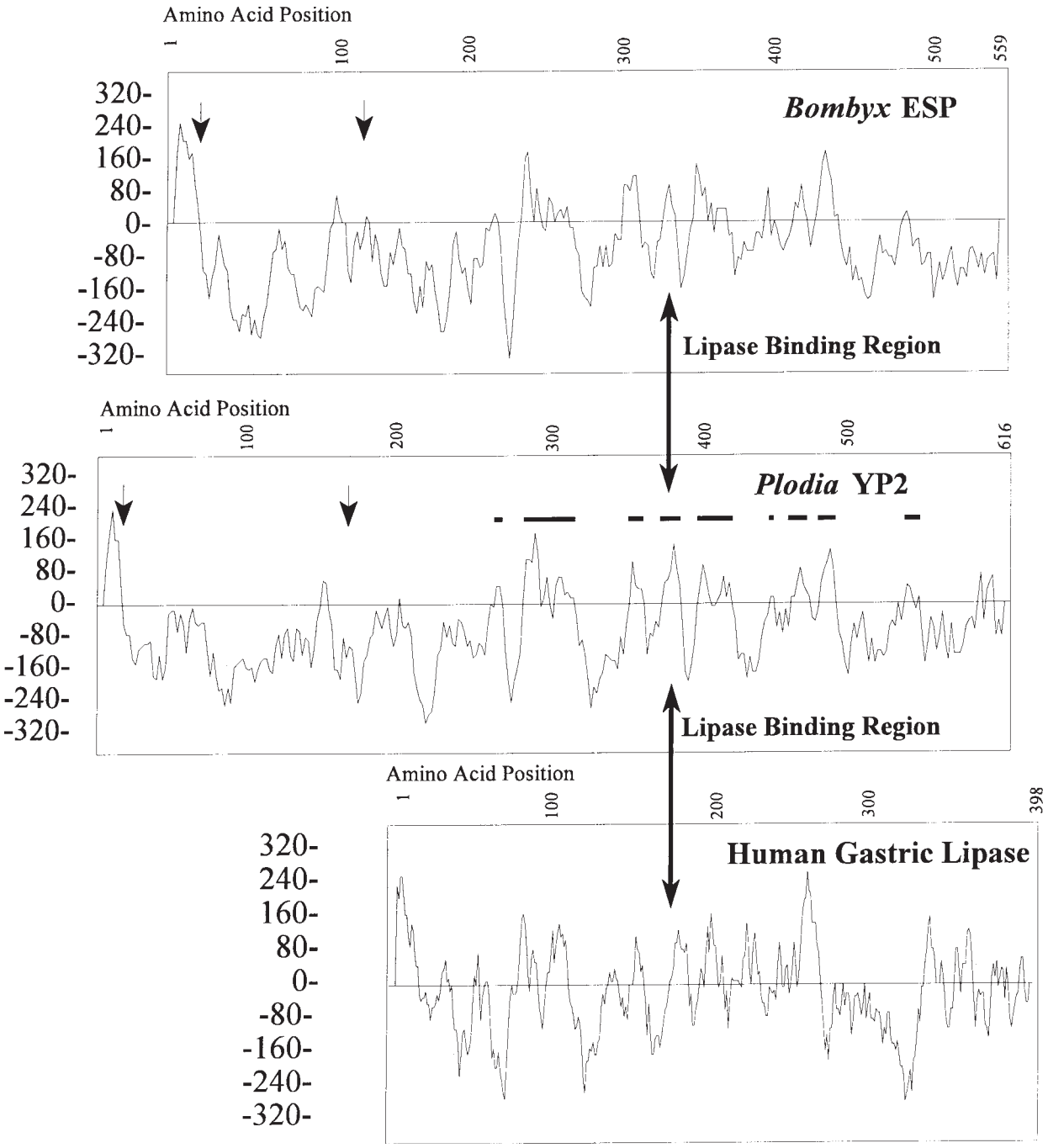


Fig. 5. The hydropathy plots show structural similarity between YP2 from *P. interpunctella*, ESP from *B. mori,* and gastric lipase from humans. The position of the hydrophobic regions shared by *Plodia* YP2 and *Bombyx* ESP is marked by solid horizontal lines in the YP2 plot. The position of the conserved lipase binding region that is present in all three polypeptides is shown by a vertical double-headed arrow. The distance for each amino acid residue was made equal between the plots.

acid residues 298 to 397, which included the lipid binding site, had 49% identities. However, when compared with the YPs of *D. melanogaster*, which also was found to have a consensus lipid binding site (Terpstra and Ab, 1988; see review Bownes, 1992), the alignment showed that the lipid binding site, (2 of 10 residue identities with human gastric lipase) was further downstream from its location in the lipases and the lepidopteran follicle yolk proteins.

The amino terminal region of YP2 from *P. interpunctella,* between amino acid residues 38 to 179, shows it has little if any similarity with the amino terminal sequences of ESP from *B. mori* or the incomplete sequences of YP2 from *G. mellonella.* In this region, the sequence of YP2 from *P. interpunctella* was 42 amino acids longer and had less than 8% identity with either ESP from *B. mori* or YP2 from *G. mellonella* while in the regions flanking this region, the level of similarity was 50%. This suggests that this portion of the YP2 sequence of *P. interpunctella* is unrelated to the sequences surrounding it, or that it has had a very different rate of evolutionary change than the rest of the sequence. The lack of similarity within this region is also striking because the long stretches of serines that are distinctive of ESP from *B. mori* and YP2 from *G. mellonella* are not present in YP2 from *P. interpunctella*. How the removal of the polyserine stretches affects the functional activity of YP2 is not clear because embryonic growth proceeds normally.

Considering that the pyralid family, which includes both *G. mellonella* and *P. interpunctella*, and the bombycid family diverged approximately 40 million years ago (Labandeira, 1994), there should be greater similarity between the YP2s of the two pyralid moths, but the alignments showed that ESP from *B. mori* and YP2 from *G. mellonella* were more closely related than YP2 from *P. interpunctella* over this region. One possible explanation for this divergence would be a recombinational event that replaced the ancestral sequence of YP2 with a different sequence that maintained the reading frame within the 5′ coding region of the YP2 gene. The likelihood that this change in the sequence is the result of recombination could be assessed by determining the sequence for YP2 in more closely related pyralid moths than *G. mellonella*. Cladistic analysis of the Pyralidae shows that the Galleriinae subfamily, which includes *G. mellonella*, is basal to the Phycitinae subfamily, which includes *P. interpunctella* (Solis and Mitter, 1992). Sequence analysis of YP2 from other members of the Phycitinae, Epipaschiinae, and Pyralinae should clarify when such a recombinational event could have occurred.

The source of the extended 5′ sequence in YP2 from *P. interpunctella* is not known, and neither the DNA sequence nor the predicted translation sequences from the three reading frames of this insert showed any significant similarity to known sequences in the GenBank or the Berkeley Drosophila Genome Project databases. When this divergence in the 5′ sequence occurred amongst the pyralid moths can be determined by sequencing this gene from other species belonging to the Phycitinae, Epipaschiinae, and Pyralinae subfamilies that are more closely related to *P. interpunctella* than is *G. mellonella*. Assessing the impact of the extended 5′ sequence on the function of the amino terminal sequence of YP2 in the embryos of *P. interpunctella* will be tested by production of the peptide in an expression vector system, which may help to elucidate the cellular requirements that lead to the maternal contribution and embryonic production of this peptide.

ACKNOWLEDGMENTS

The authors thank Miriam Hemphill for her technical assistance and Drs. A. Chen, A. M. Handler, and A. K. Kumaran for their comments on this manuscript. Mention of proprietary products does not constitute an endorsement by the USDA. The sequence reported in this paper has been deposited in the Genbank database (accession no. AF063014).

LITERATURE CITED

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990): Basic local alignment search tool. J Mol Biol 215:403–410.

Bast RE, Telfer WH (1976): Follicle cell protein synthesis and its contribution to the yolk of the *cecropia* moth oocyte. Dev Biol 52:83–97.

Bean DW, Shirk PD, Brookes VJ (1988): Characterization of yolk proteins from the eggs of the Indianmeal moth, *Plodia interpunctella.* Insect Biochem 18:199–210.

Bodmer MW, Angal S, Yarranton G, Harris TJR, Lyons A, King DJ, Pieroni G, Riviere C, Verger R, Lowe PA (1987): Molecular cloning of a human gastric lipase and expression of the enzyme in yeast. Biochim Biophys Acta 909:237–244.

Bownes M (1992): Why is there sequence similarity between insect yolk proteins and vertebrate lipases? J Lipid Res 33:777–790.

Docherty AJP, Bodmer MW, Angal S, Verger R, Lowe PA, Lyons A, Emtage JS, Harris TJR (1985): Molecular cloning and nucleotide sequence of rat lingual lipase cDNA. Nucleic Acids Res 13:1891–1903.

Huynh TV, Young RA, Davis RW (1985): Constructing and screening cDNA libraries in λgt10 and λgt11. In:

Glover DM (ed): DNA cloning: a practical approach, vol. 1. Oxford: IRL Press, pp 49–78.

Inagaki S, Yamashita O (1989): Complete amino acid sequence of *Bombyx* egg-specific protein deduced from cDNA clone. Arch Insect Biochem Physiol 10:131–139.

Indrasith LS, Furusawa T, Shikata M, Yamashita O (1987): Limited degradation of vitellin and egg-specific protein in *Bombyx* eggs during embryogenesis. Insect Biochem 17:539–545.

Indrasith LS, Sasaki T, Yaginuma T Yamashita O (1988a): The occurrence of a premature form of egg-specific protein in vitellogenic follicles of *Bombyx mori*. J Comp Physiol B 158:1–7.

Indrasith LS, Sasaki T, Yamashita O (1988b): A unique protease responsible for selective degradation of a yolk protein in *Bombyx mori*: purification, characterization, and cleavage profile. J Biol Chem 263:1045–1051.

Irie K, Yamashita O (1983): Egg-specific protein in the silkworm, *Bombyx mori*: purification, properties, localization and titre changes during oogenesis and embryogenesis. Insect Biochem 13:71–80.

Kyte J, Doolittle KR (1982): A simple method for displaying the hydropathic character of a protein. J Mol Biol 157:105–132.

Labandeira C (1994): A compendium of fossil insect families. Milwaukee Public Museum Contributions In: Biology and geology, no. 88. Milwaukee, WI: Milwaukee Public Museum.

Ono S, Nagayama H, Shimura K (1975): The occurrence and synthesis of female- and egg-specific proteins in the silkworm, *Bombyx mori*. Insect Biochem 5:313–329.

Robinson PA, Anderson BH, Loviny TLF (1988): Nitrocellulose-bound antigen repeatedly used for the affinity purification of specific polyclonal antibodies for screening DNA expression libraries. J Immunol Methods 108: 115–122.

Sambrook J, Fritsch EF, Maniatis T (1989): Molecular cloning: a laboratory manual, 2nd ed., vols. 1–3. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Sato Y, Yamashita O (1991): Structure and expression of a gene coding for egg-specific protein in the silkworm, *Bombyx mori*. Insect Biochem 21:495–505.

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

Shirk PD, Bean D, Millemann AM, Brookes VJ (1984): Identification, synthesis, and characterization of the yolk polypeptides of *Plodia interpunctella*. J Exp Zool 232:87–98.

Silhacek DL, Miller GL (1972): Growth and development of the Indian meal moth *Plodia interpunctella* Lepidoptera: Phycitidae), under laboratory mass-rearing conditions. Ann Entomol Soc Am 65:1084–1087.

Solis MA, Mitter C (1992): Review and preliminary phylogenetic analysis of the subfamilies of the Pyralidae (sensu stricto) (Lepidoptera: Pyraloidea). Syst Ent 17:79–90.

Telfer WH (1954): Immunological studies of insect metamorphosis II. The role of a sex-limited blood protein in egg formation by the cecropia silkworm. J Gen Physiol 37:539–558.

Telfer WH, Rubenstein E, Pan ML (1981): How the ovary makes yolk in *Hyalophora*. In: Sehnal F, Zabza A, Menn JJ, Cymborowski B (eds): Regulation of insect development and behavior. Wroclaw: Wroclaw Technical University, pp 637–654.

Terpstra P, Ab G (1988): Homology of Drosophila yolk proteins and the triacylglycerol lipase family. J Mol Biol 202:663–665.

Walker JM (1984): The dansyl-Edman method for peptide sequencing. In: Walker JM (ed): Methods in molecular biology: proteins. Clifton, UK: Humana Press, pp 213–219.

Yamashita O, Irie K (1980): Larval hatching from vitellogenin-deficient eggs developed in male hosts of the silkworm. Nature 283:385–386.

Yan YL, Kunert CJ, Postlethwait JH (1987): Sequence homologies among the three yolk polypeptide (*Yp*) genes in *Drosophila melanogaster*. Nucleic Acids Res 15:67–85.

Zhu J, Indrasith LS, Yamashita O (1986): Characterization of vitellin, egg-specific protein and 30 kDa protein from *Bombyx* eggs, and their fates during oogenesis and embryogenesis. Biochim Biophys Acta 882:427–436.

Zimowska G, Shirk PD, Silhacek DL, Shaaya E (1995): Vitellin and formation of yolk spheres in vitellogenic follicles of the moth, *Plodia interpunctella*. Arch Insect Biochem Physiol 29:71–85.