cDNA and Gene Sequence of *Manduca sexta* Arylphorin, an Aromatic Amino Acid-rich Larval Serum Protein

HOMOLOGY TO ARTHROPOD HEMOCYANINS*

(Received for publication, June 26, 1989)

Elizabeth Willott‡, Xiao-Yu Wang, and Michael A. Wells§

From the Biochemistry Department, BioSciences West, University of Arizona, Tucson, Arizona 85721

The serum (storage) proteins produced by insect larvae at the end of the feeding cycle are hexameric blood proteins with one or more type of subunits. The cDNA and gene structure of the aromatic amino acid-rich larval serum protein arylphorin from the tobacco hornworm, *Manduca sexta*, has been determined. In *M. sexta* arylphorin there are two subunits α and β, which have 668 and 687 amino acids, respectively, and whose amino acid sequences are 68% identical. The two genes, separated by 7.1 kilobases of chromosomal DNA, are transcribed in the same direction. Based on the alignment of the amino acid sequence, the rate of nucleotide substitution between the two coding regions predicts that the two genes diverged about 100 million years ago. Both genes contain 5 exons and the upstream region contains a sequence, TGATAAAA, which is similar to a sequence found in all other storage protein genes for which information is available. When the National Biomedical Research Foundation protein sequence data base was searched, it was found that the arylphorin subunits showed significant similarity to the arthropod hemocyanins, which are hexameric oxygen-carrying proteins. Based on the alignment of the sequence of *M. sexta* arylphorin and the hemocyanin from the spruce lobster (*Panulirus interruptus*), for which a 3.2 Å structure has been determined, it was observed that the highest concentration of conserved residues were found in those regions of the sequence which are involved in subunit interactions in the hexameric protein. It is suggested that the insect storage proteins and the arthropod hemocyanins have evolved from a common ancestor.

During the final larval stadium of holometabolous insects, a few proteins, called the larval serum proteins or larval storage proteins (LSP), accumulate in large amounts in the hemolymph where they can account for up to 85% (by weight) of the hemolymph protein. The LSPs have molecular masses of about 500 kDa and are composed of six subunits of approximately 72–80 kDa, which may or may not be identical. The proteins have been called storage proteins because amino acids from them are recoverable from many different proteins in the adult (Levenbook and Bauer, 1984) suggesting that the LSPs serve as a store of amino acids for synthesis of adult proteins. The proteins have also been implicated in cuticle sclerotization (Agrawal and Scheller, 1987; Kalifas et al., 1984; König et al., 1986; Webb and Riddiford, 1988).

There are two well characterized classes of LSPs (Levenbook, 1985). One class is rich in aromatic amino acids, generally containing 18–25% phenylalanine and tyrosine, and proteins in this class are called arylphorins (Telfer et al., 1983). Another class is relatively rich in methionine, approximately 6%, and is often more prevalent in females than in males (Tojo et al., 1980; Ryan et al., 1985a). *Manduca sexta* arylphorin has two subunits which by sodium dodecyl sulfate-polyacrylamide gel electrophoresis are approximately 72 and 76 kDa each with a high mannose type carbohydrate chain (Ryan et al., 1985b). Arylphorin is present in both males and females throughout larval development though its concentration in the hemolymph increases greatly in the last larval instar (Kramer et al., 1980; Willott, 1988).

The LSPs represent important model proteins because their genes are highly expressed and are developmentally regulated. In order to investigate gene regulation it is necessary to know the sequence of the genes and mRNAs which code for the proteins. The cDNA and gene structure for the methionine-rich storage protein from *Bombyx mori* have been reported (Sakurai et al., 1988a, 1988b). The genes for the three arylphorin subunits from *Drosophila melanogaster* (Smith et al., 1981; Delaney et al., 1986; Lepesant et al., 1986) and for the arylphorin from *Sacrophaga peregrina* (Matsumoto et al., 1986) have been cloned and some partial sequence data reported. However, to date, the complete sequence for an arylphorin has not been reported.

In this paper we report the complete cDNA and gene sequences for the two subunits of arylphorin from *M. sexta*. In addition we report that the arylphorins, as well as the *B. mori* methionine-rich storage protein, show a remarkable degree of sequence similarity to the arthropod hemocyanins.

MATERIALS AND METHODS

cDNA Library Construction and Screening—A larval fat body cDNA library in Agr11 (Kanost et al., 1988) was screened by differential hybridization (Maniatis et al., 1982) to larval and adult cDNA to select clones expressed mainly or solely in larval fat body. The larval-specific clones were then screened with antisera to arylphorin as described by Cole et al. (1987). From this screening two clones were isolated which each had an approximately 2.4-kilobase insert but different restriction enzyme maps. Arylphorin encoding sequences
**M. sexta Arylphorins**

**Fig. 1. Structure of the M. sexta arylphorin genes and sequencing strategy.** Boxes depict exons and lines introns. Solid lines beneath the gene representation show subclones sequenced left to right and dotted lines subclones sequenced right to left. bp, base pairs.

**Fig. 2. Partial restriction maps of the M. sexta arylphorin cDNAs and sequencing strategy.** Solid lines beneath the gene representation show subclones sequenced left to right and dotted lines subclones sequenced right to left. Restriction enzyme abbreviations are: E, EcoRI; P, PstI; S, SalI; Sca, ScaI; B, BglII.

**Fig. 3. Partial restriction map of a 20.9-kilobase fragment of M. sexta DNA showing the relative chromosomal location and orientation of the two arylphorin genes.** This map was constructed from the two overlapping λ clones (10A and B). S, SalI; E, EcoRI; H, HindIII. The open boxes below λ clone 10A show the EcoRI fragments which hybridized to the α subunit cDNA. kb, kilobases.
were subcloned into pTZ vectors by using convenient DNA restriction sites or by using these after DNA amplification of the Xgtll inserts with the polymerase chain reaction (Oste, 1988), using reverse primers. To obtain suitable length fragments in the DNA, deletions of pTZ clones were constructed as described by Henikoff (1984).  

Isolation of Genomic Clones—The probe, which corresponded to the coding region of the \( \alpha \) subunit of arylphorin, was \( ^{32} \)P-labeled by nick translation (Maniatis et al., 1982), and purified by the method of Henikoff, 1984). A DNA was obtained from positive clones by the plate lysate method (Maniatis et al., 1982), and purified by the method of Benson and Taylor (1984). The clones were further characterized by DNA sequencing using the dideoxy method (Sanger et al., 1980).  

Sequence Data—The National Biomedical Research Foundation protein sequence data base was searched with the FASTP program, and the significance of the similarity was determined with the RDF program (Lipman and Pearson, 1985). The sequences of the two arylphorin subunits were aligned with the PRTALN program (Wilbur and Lipman, 1983), and this alignment was used to determine the rates of nucleotide substitution between the two arylphorin mRNAs using the computer programs of Li et al. (1985). Progressive sequence alignment and construction of phylogenetic trees were performed using the computer programs of Feng and Doolittle (1987).  

**RESULTS AND DISCUSSION**

Fig. 1 shows the organization of the genes for the two subunits of \( M. \) sexta arylphorin and the strategy used to sequence the genes. Fig. 2 presents a partial restriction enzyme map of the cDNAs for the two proteins and the sequencing strategy used. Both genes contain five exons and \( ^{3} \) indicates the translation start site. The box indicates the location of the oligonucleotide used for primer extension analysis.
The numbers refer to the position of the 5' end of the sequence in the gene and are numbered relative to the transcription start site. YP H-BOX refers to the Drosophila yolk protein H-box described by Yan et al. (1987). The B. mori sequence is from Sakurai et al. (1988b); the S. peregrina sequence from Matsumoto et al. (1986); the Drosophila LSP sequences from Delaney et al. (1986).

by about 7.1 kilobases of chromosomal DNA and are transcribed in the same direction. Figs. 4 and 5 present the sequence of the genes for the two proteins. Both genes contain a typical TATA box which begins 30 and 29 nucleotides upstream from the transcription start site in the α and β genes, respectively. The 5'-untranslated regions of the two genes contain a number of homologous sequences. Since the two genes are subject to identical regulation.
Fig. 6 presents the amino acid sequences deduced from the gene and cDNA sequences. Table II presents the amino acid composition of the two subunits derived from the sequence, as well as the composition of the isolated protein, and it can be seen that the agreement is excellent. Each subunit contains a 16-amino acid signal sequence; in these signal sequences 14 of the 16 amino acids are identical. The mature a subunit contains 668 amino acids and a calculated molecular weight of 82,278, while the mature b subunit contains 687 amino acids and has a calculated molecular weight of 82,279. These molecular weights do not agree with the differences noted by sodium dodecyl sulfate-polyacrylamide gel electrophoresis, but we can offer no explanation at present. Each subunit contains two consensus glycosylation sites (NXT/S), although only one appears to be used (Ryan et al., 1985b). The alignment shown in Fig. 6 shows that the mature proteins are identical with only two gaps. Using the protein alignments shown in Fig. 6 to align the cDNA sequences and the computer program of Li et al. (1985), it was determined that there have been 0.208 ± 0.013 substitutions/nonsynonymous site between the coding regions of the two genes. Using a value of 0.9 × 10⁻⁹ substitutions/nonsynonymous site/year (Li et al., 1985), we can estimate that the two genes diverged approximately 100 million years ago.

In the apolipoprotein-III genes from M. sexta and Locusta migratoria a bias toward the use of C or G in the third position of codons was noted (Kanost et al., 1988). In the apolipophorin genes a strong bias for use of C in the third position of codons is found for Phe (84% for the a gene and 91% for the b gene); Tyr (79% for the a gene and 81% for the b gene); Asn (76% for the a gene and 78% for the b gene); and Asp (83% for the a gene and 61% for the b gene). Together TTC (Phe) and TAC (Tyr) account for about 16% of all codon usage, which means that the corresponding tRNAs must be abundant in order to support the high rates of apolipophorin biosynthesis seen in the last larval stage.

When the National Biomedical Research Foundation protein sequence data base was searched, the only sequences found with significant similarity to the M. sexta apolipoporphins were several subunits of various arthropod hemocyanins. The significance of these similarities was evaluated by comparing the sequences of the hemocyanins with 50 randomly shuffled sequences having the same amino acid composition as either the a or b subunits using the RDF program (ktup = 2) (Lipman and Pearson, 1985). For the a subunit the following z values were obtained, where z > 10 is considered significant: b subunit, 105.4; B. mori methionine-rich storage protein, 101.6; Limulus polyphemus (horseshoe crab) hemocyanin, 54.4; Euryphilma californicum (tarantula) hemocyanin subunit D, 54.0 and subunit E, 45.4; Panulirus interruptus (spiny lobster) hemocyanin, 14.7. For the b subunit the following values were obtained: B. mori methionine-rich storage protein, 82.5; L polyphemus hemocyanin, 42.7; E. californicum hemocyanin subunit D, 41.9 and subunit E, 57.1; P. interruptus hemocyanin, 16.1.
**Fig. 7.** Alignment of the amino acid sequences of insect storage proteins and arthropod hemocyanins. SPLB, *P. interruptus* (spiny lobster) hemocyanin; ALPH and BETA, *M. sexta* arylphorin subunits; BMOR, *B. mori* methionine-rich storage protein; HSCB, *L. polyphemus* (horseshoe crab) hemocyanin. The sequences were aligned by the method of Fung and Doolittle (1987) \* indicates identical residues in all sequences; \# indicates identical residues in the spiny lobster hemocyanin and the *M. sexta* arylphorins; \# indicates conservative replacements between the spiny lobster hemocyanin and the *M. sexta* arylphorins. The sequences are underlined.

**Fig. 8.** Phylogenetic tree showing relationship between arthropod hemocyanins and insect storage proteins. SPLB, *P. interruptus* (spiny lobster) hemocyanin; ALPH and BETA, *M. sexta* arylphorin subunits; BMOR, *B. mori* methionine-rich storage protein; TARD and TARE, *E. californium* (tarantula) subunits D and E; HSCB, *L. polyphemus* (horseshoe crab) hemocyanin. The tree is based on the alignment shown in Fig. 7.
The highest concentration of identical residues or conservative protein, and hemocyanin subunits from sextu
the storage proteins and the hemocyanins have similar sub-units and form hexamers. When the sequences of the
storage proteins were obtained. While the construction of the protein sequences was obtained. While the construction
of the hexameric motif has been shown to be common feature in the arthropods, although at this time there is no evidence to suggest why a hexameric protein is advantageous.

Acknowledgments—We thank Dr. Ken Cole for preparation of the larval-specific library, Drs. Don Frohlich, Eric Hanneman, John Law, and Michael Kanost for helpful comments during preparation of the manuscript and Mary Medina for animal care.

Note Added in Proof—After this paper was accepted for publication, the sequence of the arylphorin from Bombyx mori was reported (Fujii, T., Sakurai, H., Isumi, S., and Tomino, S. (1989) J. Biol. Chem. 264, 11020–11025). Progressive sequence alignment of the B. mori arylphorin with the proteins listed in Table III gave the following results for percent identity: ALPH (66.3), BMOR (29.8), TARD (25.0), TARE (25.2), HSCB (24.0), SPLB (26.1). Thus, the B. mori arylphorin is more closely related to the M. sexta arylphorin subunits than it is to the B. mori methionine-rich storage protein. The average rate of substitutions per nonsynonymous between the M. sexta arylphorin subunits and the B. mori arylphorin is 0.259 ± 0.014, suggesting that the B. mori and M. sexta genes diverged about 140 million years ago. On the other hand, the rate of nonsynonymous substitution between the B. mori methionine-rich storage protein and the B. mori arylphorin (0.80 ± 0.038) and the M. sexta arylphorin subunits (a = 0.748 ± 0.036; b = 0.747 ± 0.033) suggest that the arylphorin and methionine-rich protein genes diverged about 425 million years ago.

Also appearing after this paper was accepted was a paper showing that a monoclonal antibody prepared against tarantula hemocyanin cross-reacted with the arylphorin from the blowfly Calliphora vicina (Markl, J., and Winter, S. (1989) J. Exp. Biol. 140, 1645–1655). This result further supports the suggestion that the insect storage proteins and arthropod hemocyanins are related.

REFERENCES
Agrawal, O. P., and Scheller, K. (1987) Biol. Chem. Hoppe Seyler 368, 572
Beres, S. A., and Taylor, R. K. (1984) BioTechniques 2, 126–127
Cole, K. D., Frederico, J. M., and Neff, H. D. (1987) J. Mol. Biol. 199, 81–90
Delaney, S. J., Smith, D. F., McClelland, S., Sunkel, C., and Glover, D. M. (1986) J. Mol. Biol. 199, 1–11
Dente, L., Cesareni, G., and Cortese, R. (1984) Nucleic Acids Res. 11, 1645–1655
Fujii, T., and Doolittle, R. (1987) J. Mol. Biol. 251, 1–15
Gaykema, W., Hol, W. G. J., Vereijken, J. M., Soeter, N. M., Bak, H. J., and Beintema, J. J. (1989) Nature 309, 23–29
Henikoff, S. (1984) Gene (Amst.) 28, 551–559
Kalinin, A. D., Marmaras, V. J., and Christodoulou, C. (1984) Roux's Arch. Dev. Biol. 194, 37–43
Kanost, M. R., Prasad, S. V., and Wells, M. A. (1989) J. Biol. Chem. 264, 965–972
König, M., Agrawal, O. P., Schenkki, H., and Scheller, K. (1986) Roux's Arch. Dev. Biol. 195, 279–283
Kramer, S. J., Mundall, E., and Law, J. H. (1980) Insect Biochem. 10, 279–286
Levenbook, M., Schenkki, H., and Scheller, K. (1986) Roux's Arch. Dev. Biol. 195, 279–283
Linzen, B., Soeter, N. M., Riggs, A. F., Schneider, H. J., Schartau, W., Moore, M. D., Yokata, E., Behrens, P. Q., Nakashima, H., Takagi, T., Nemoto, T., Vereijken, J. M., Bak, H. J., Beintema, J.

### Table III

Percent identities (above rule) and distance scores (below rule) calculated from progressively aligned sequences

| ALPH | BETa | BMOR | TARE | TARD | HSCB | SPLB |
|------|------|------|------|------|------|------|
| ALPH | 68.3 | 32.9 | 27.6 | 27.6 | 27.5 | 25.9 | 25.4 |
| BETa | 29   | 32.5 | 27.1 | 25.4 | 24.2 | 24.6 |
| BMOR | 95   | 95   | 27.0 | 24.8 | 26.8 | 27.4 |
| TARE | 124  | 124  | 129  | 54.2 | 53.3 | 30.7 |
| TARD | 124  | 125  | 151  | 47   | 53.3 | 31.2 |
| HSCB | 132  | 132  | 129  | 50   | 46   | 31.3 |
| SPLB | 117  | 119  | 121  | 109  | 97   | 102 |

Note: Percent identities (above rule) and distance scores (below rule) calculated from progressively aligned sequences.
M. sexta Arylphorins

J. Volbeda, A., Gaykema, W. P. J., and Hol, W. G. J. (1985) Science 229, 519–524
Lipman, D. J., and Pearson, W. R. (1985) Science 227, 1435–1441
Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Springs Harbor, NY
Matsumoto, N., Nakanishi, Y., and Natori, S. (1986) Nucleic Acids Res. 14, 2685–2698
Oste, C. (1988) BioTechniques 6, 162–167
Ryan, R. O., Keim, P. W., Wells, M. A., and Law, J. H. (1985a) J. Biol. Chem. 260, 782–787
Ryan, R. O., Anderson, D. R., Grimes, W. J., and Law, J. H. (1985b) Arch. Biochem. Biophys. 243, 115–124
Sakurai, H., Fujii, T., Izumi, S., and Tomino, S. (1988a) J. Biol. Chem. 263, 7876–7880
Sakurai, H., Fujii, T., Izumi, S., and Tomino, S. (1988b) Nucleic Acids Res. 16, 7717–7718
Sanger, F., Nicklen, S., and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. U. S. A. 74, 5463–5467
Smith, D. F., McClelland, A., White, B. N., Addison, C. F., and Glover, D. M. (1981) Cell 23, 441–449
Southern, E. M. (1975) J. Mol. Biol. 98, 503–517
Telfer, W. H., and Massey, H. C., Jr. (1987) in Molecular Entomology (Law, J. H., ed) pp. 305–314, Alan R. Liss, Inc., New York
Telfer, W. H., Keim, P. S., and Law, J. H. (1983) Insect Biochem. 13, 601–613
Tojo, S., Nagata, M., and Kobayashi, M. (1986) Insect Biochem. 10, 289–303
Wang, X.-Y., Cole, K. D., and Law, J. H. (1989) Gene (Amst.) 80, 259–268
Webb, B. A., and Riddiford, L. M. (1988) Dev. Biol. 130, 671–681
Wilbur, W. J., and Lipman, D. J. (1983) Proc. Natl. Acad. Sci. U. S. A. 80, 726–730
Williams, J. G., and Mason, P. J. (1985) in Nucleic Acid Hybridization: A Practical Approach (Hames, B. D., and Higgins, S. J., eds) pp. 139–160, IRL Press, Oxford
Wiltott, E. (1988) Manduca sexta Fat Body during the Larval-Pupal Transformation: A Structural and Biochemical Study. Ph.D. dissertation, University of Arizona
Yan, Y. L., Kunert, C. J., and Postlethwait, J. H. (1987) Nucleic Acids Res. 15, 67–85