Volume 256, number 1,2, 215–218 FEB 07708 October 1989

Molecular cloning of an insect pheromone-binding protein

K. Raming, J. Krieger and H. Breer

University Stuttgart-Hohenheim, Institute of Zoophysiology, 7000 Stuttgart 70, FRG

Received 23 August 1989

Clones coding for the pheromone binding protein precursor have been selected from a cDNA library derived from antennae of the male moth, *Antheraea polyphemus*. The deduced protein sequence consists of a signal peptide of 20 amino acid residues and a mature binding protein of 142 amino acid residues. RNA blot hybridization indicated that the mRNA is selectively expressed in the antennae of the male moth.

Pheromone binding protein; cDNA cloning; Nucleotide sequence; Primary structure; RNA blot hybridization

1. INTRODUCTION

Insects, in particular moths, have a remarkable ability to detect extremely low concentrations of sex pheromones [1] and are now considered as interesting models for unravelling transductory mechanisms which translate chemical into electrical signals [2]. The antennae of the male moth, *Antheraea polyphemus*, bear about 55 000 sensory hairs [3]; each sensillum contains the pheromone-sensitive dendrites of 2–3 sensory cells bathing in sensillum lymph, a highly proteinaceous fluid. Two soluble proteins, which are supposed to be involved in pheromone processing, have been detected in lymph fluid, a pheromone degrading esterase and a pheromone binding protein (PBP) [4,5]. The PBP ($M_r \approx 15\,000$) is present in extremely high concentrations (10–20 mM); its interaction with the pheromone molecules gives a binding constant of about 10^{-7} [6]. It has been suggested that PBP may play a multifunctional role in pheromone processing, including solublization of the hydrophobic pheromone molecules in the sensillum lymph, its protection against the agressive esterase and translocation to reach the receptors presumably located in the dendritic membrane ('carrier model') [7] and/or the inactivation of the pheromones ('contact model') [8]. There is as yet no direct experimental proof for the particular functional role of PBP. A precise molecular characterization of the protein may provide some clues to elucidate its functional properties.

Correspondence address: H. Breer, University Stuttgart-Hohenheim, Institute of Zoophysiology, 7000 Stuttgart 70, FRG

2. MATERIALS AND METHODS

2.1. cDNA library

RNA (15μg) was isolated from antennae of 100 newly eclosed male *Antheraea polyphemus* as described by Maniatis et al. [9] and poly(A^+) RNA was obtained by oligo(dT) chromatography [10]. 1 μg of poly(A^+) RNA was used to synthesize an oligo(dT)-primed double-stranded cDNA following the procedure of Gubler and Hoffman [11]. cDNA ≥ 0.8 kb was ligated into the cloning vector λgt10. Approximately 10^5 recombinants were screened with a mixed, ^{32}P-endlabeled 15-residue oligonucleotide probe. Positive clones were plaque-purified and their cDNA inserts subcloned into M13mp18.

2.2. Analysis of cDNA clones

For DNA sequencing, nested deletions were produced by ExoIII/Nuclease S1 treatment. DNA sequencing was performed using the dideoxychain termination method [12] and T7 DNA polymerase.

2.3. Northern blot analyis

Poly(A^+) RNA was separated on 1.2% agarose/formaldehyde gels, transferred onto nylon filters and hybridized with a ^{32}P-labeled DNA-probe [13]. The blots were washed

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/89/$3.50

with 0.1 × SSC, 0.1% SDS for 2 × 30 min at 50°C and subsequently exposed to Fuji RX Safety films.

3. RESULTS AND DISCUSSION

Because of its abundance, the PBP could be readily purified to homogeneity and partial N-terminal sequences have been obtained by gas-phase microsequencing [7,14]. Based on this information, a redundant, mixed 15-residue oligonucleotide probe, designed according to amino acids 3–7 of the N-terminal sequence EIMKN, was synthesized

T A T
Nucleotide probe: 5′-GTTCTTCATGATCTC-3′
T

and employed to screen a cDNA library derived from poly(A^+) RNA from the antennae of freshly eclosed males of *Antheraea polyphemus*. About 10^5 recombinant clones in λgt10 were screened and 19 plaques formed by recombinant phages

```
                                                1 GGATTAGTGAACTGGGTATTCTTGTAAAAGATAGAG

 37  ATG TTG AGA AAA ATT TCG TTG TTG TTG CTG CCG GTG TTC GTG GCA
-21  Met Leu Arg Lys Ile Ser Leu Leu Leu Leu Pro Val Phe Val Ala

 82  ATA AAC CTA GTA CAT TCT TCG CCA GAG ATC ATG AAG AAT TTA AGC
 -6  Ile Asn Leu Val His Ser Ser Pro Glu Ile Met Lys Asn Leu Ser

127  AAT AAT TTT GGT AAA GCT ATG GAT CAG TGT AAG GAC GAG CTT AGT
 10  Asn Asn Phe Gly Lys Ala Met Asp Gln Cys Lys Asp Glu Leu Ser

172  CTC CCT GAT TCC GTT GTC GCC GAT TTG TAC AAC TTT TGG AAG GAT
 25  Leu Pro Asp Ser Val Val Ala Asp Leu Tyr Asn Phe Trp Lys Asp

217  GAC TAT GTG ATG ACC GAT AGG TTA GCA GGA TGT GCC ATA AAC TGT
 40  Asp Tyr Val Met Thr Asp Arg Leu Ala Gly Cys Ala Ile Asn Cys

262  TTG GCC ACC AAG CTA GAT GTA GTC GAT CCT GAT GGA AAT CTC CAC
 55  Leu Ala Thr Lys Leu Asp Val Val Asp Pro Asp Gly Asn Leu His

307  CAT GGA AAC GCA AAG GAC TTC GCG ATG AAG CAT GGA GCT GAT GAA
 70  His Gly Asn Ala Lys Asp Phe Ala Met Lys His Gly Ala Asp Glu

352  ACC ATG GCG CAG CAA CTG GTG GAT ATT ATA CAC GGA TGT GAA AAG
 85  Thr Met Ala Gln Gln Leu Val Asp Ile Ile His Gly Cys Glu Lys

397  TCT GCT CCA CCT AAT GAC GAT AAA TGT ATG AAG ACT ATA GAT GTT
100  Ser Ala Pro Pro Asn Asp Asp Lys Cys Met Lys Thr Ile Asp Val

442  GCG ATG TGT TTC AAA AAA GAG ATC CAT AAG CTG AAC TGG GTT CCT
115  Ala Met Cys Phe Lys Lys Glu Ile His Lys Leu Asn Trp Val Pro

487  AAC ATG GAC CTT GTA ATA GGC GAG GTC TTA GCT GAA GTT TAG ATT
130  Asn Met Asp Leu Val Ile Gly Glu Val Leu Ala Glu Val END

532  GTGTCGACACTTAATAGCGGTCTCAGACTCCATGACATCTTGACGAAAAGACCTTAAAT
591  AAAATATTTTGAATTTAGGTTGTACCAAAAATTTTAAATACCTATCAAAATTAAATTCA
650  GCTCAATATTTTAATGATCATTAAATCTAAAATTACATTGTAGTATATTTACTAATAAA
709  ATAACAAAAGAATTTAGTACAATCAATAACTTAACTAGCCCAATGATATCGTATATACA
768  TTTAGCACATTGAACATTTTTATAAATCTACAAATCGTCGTTACCGAGCTTGAAAGTTG
827  GCTCCGTCAAATGATAATATTAATTTTGATAGTAATTTCAATCCAAATAATAAGCGAGC
886  TTCGTTTTGGTAATATTTTGTTTTAATTTTTTATGAATATTTATATTATATATATTATA
945  TATATTTTATGAATATTTGTGATAAAGAAAAAATTTACATTTGTATAAATTTGTACGAT
1004 TTATAAAAAAGTTGAAGTTGACTTTCTTCATGATTTACTTTCTTTTCTCTTATATTTGT
1063 ATACAAATTAGTTATTGTCTATGCACCTTAACATAGTATACTTTGCTTTCTATATACTT
1122 TTAGCACAGATAAACCTATATTCTTGTATTCTTGAAATAAAGCGTTTCAAATG-poly A
```

Fig.1. Nucleotide sequence of the cDNA encoding the pheromone binding protein from *Antheraea polyphemus*. Nucleotide residues are numbered in the 5′ to 3′ direction. The number of the nucleotide residue at the left end of each line is given. The sequence shown is followed on its 3′ end by a stretch of adenosine residues (not shown). The putative polyadenylation signal sequence, AATAAA, is underlined. The deduced amino acid sequence is displayed below the corresponding nucleotide sequence. Amino acid no. 1 is assigned to the first amino acid residue of the mature pheromone binding protein. The amino acid residues in the signal peptide are indicated by negative numbers and underlined.

hybridized with the oligonucleotide probe. Four clones were plaque-purified and their cDNA inserts isolated. One of the inserts (λAPO-3) was subcloned into the M13mp18 plasmid vector and sequenced.

The complete sequence of clone λAPO-3 is shown in fig.1. It comprises 1186 nucleotides including a predicted start codon, ATG, and a stop codon, TAG, at nucleotides 526–528. The open reading frame of 489 bases is flanked by 36 bases on the 5′ side and 658 bases at the 3′ end. The 3′ untranslated region is followed by a stretch of adenosine residues (not shown in fig.1), presumably copied from the mRNA polyadenosine tail during reverse transcription. Beginning 18 bases 5′ to the polyadenosine stretch, there is a consensus polyadenylation signal AATAAA [15]. Thus the insert appears to contain the complete 3′ untranslated region of the mRNA.

The open reading frame putatively encodes a 163 amino acid protein with a predicted molecular mass of 18 142 Da; the deduced amino acid sequence of the PBP is displayed below the corresponding nucleotide sequence in fig.1. The predicted initiator, methionine, which is the first ATG triplet that appears 4 codons downstream of the in frame stop codon (TAA) at position 25–27. Applying the algorithm for assigning a 'score' to potential signal peptidase cleavage sites [16] to the first 37 amino acids gave a maximum score of 6.66 for cleavage after amino acid 21, as expected; there was no other score higher than 2.15. The stretch of 20 amino acid residues has the characteristic features of a signal peptide common to membrane-associated and secretory proteins [16]. These features are a stretch of highly hydrophobic amino acid residues followed by a hydrophilic residue (asparagine) and a residue with a short side chain located at the putative cleavage site (serine). The region immediately following the presumed leader sequence corresponds to the amino-terminal sequence of the mature protein and almost precisely matches the amino acids determined by protein sequencing [7,14]. The deduced structure of the mature protein consists of 142 amino acid residues giving a calculated molecular mass of 15 783 Da and a theoretical isoelectric point of pH 4.43. These values are quite consistent with the results obtained for the PBP from *Antheraea polyphemus* analysed by

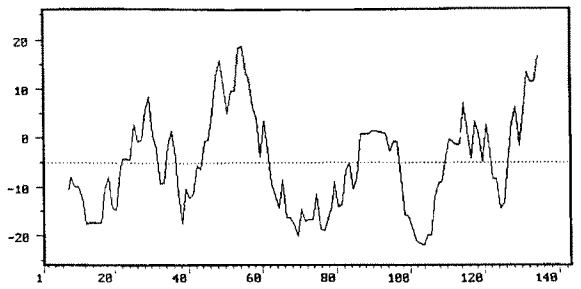

Fig.2. Analysis of the hydropathic properties of the moth pheromone binding protein. Amino acid residues are plotted against their hydropathy index values. Positive values indicate hydrophobicity and negative values hydrophilicity. The window was 15 residues.

micropolyacrylamide gel electrophoresis ($M_r \approx$ 14 400) and isoelectrofocussing on ultrathin layer (p*I* of 4.7) [5].

The amino acid sequence of the pheromone binding protein precursor was analysed for local hydrophobicity according to the procedure of Kyte and Doolittle [17]. Several hydrophobic domains emerged (fig.2); none of them was long enough for a membrane-spanning region. Two major hydrophobic stretches were detected between residues 42–57 and 133–142; three less strongly

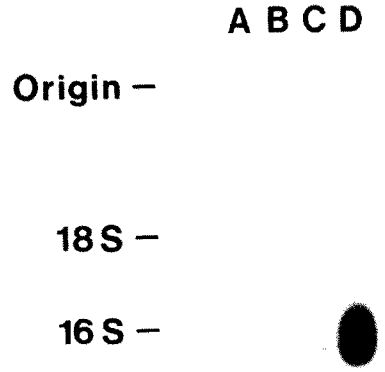

Fig.3. Autoradiogram of blot hybridization analysis of RNA from *Antheraea polyphemus*: (A) muscle poly(A$^+$), (B) female antennae poly(A$^+$), (C) male antennae poly(A$^+$), (D) male antennae poly(A$^+$). The amounts of RNA in lane A and C were 5 μg, in B and D 2 μg. The hybridization probes used derived from clone λAPO-1 labeled by random priming with [α-^{32}P]dATP. Autoradiography was performed at -70°C for 24 h with an intensifying screen. The size markers were *E. coli* rRNAs.

hydrophobic regions between 23–34, 85–97 and 111–118. It remains to be determined if some of these hydrophobic domains are involved in forming binding pockets for the hydrophobic pheromone molecules.

In order to assay the tissue distribution of the mRNA for PBP and to estimate the size of the transcript encoding the PBP, Northern blot analysis was performed using various RNA preparations. As can be seen in fig.3, significant hybridization signals could only be detected with poly(A^+) RNA from antennae of the male moths. The hybridizing RNA, putative mRNA for PBP, was found to comprise a size of about 1.4 kb; thus, the mRNA is about 200 nucleotides larger than the cloned cDNA. This discrepancy probably reflects a more extented polyadenylation in vivo. These results indicate that the pheromone binding protein appears to be expressed selectively in antennae of the male moth. The question if PBP is produced by a particular cell type in the male antennae will be approached by in situ hybridization techniques.

Deciphering the structure of several other clones will elucidate if there is a microheterogeneity of PBP, which could not be resolved by the protein analysis, but may be important for detecting different pheromones. Furthermore, unravelling the primary structure of PBPs from closely related and phylogenetically more distant species will provide information on sequence conservation which may offer some clues for detecting domains of functional importance.

Acknowledgements: This work was supported by the Deutsche Forschungsgemeinschaft. The authors are indepted to Professor Kaissling for encouragement and for generously supplying cocoons of *Antheraea polyphemus*. The support by Dr Wieczorek is greatly acknowledged.

REFERENCES

[1] Schneider, D., Kasang, G. and Kaissling, K.E. (1968) Naturwissenschaften 55, 395.
[2] Kaissling, K.E. (1986) Annu. Rev. Neurosci. 9, 121–145.
[3] Boeckh, J., Kaissling, K.E. and Schneider, D. (1960) Zool. J. Anat. 78, 559–584.
[4] Vogt, R.G. and Riddiford, L.M. (1981) Nature 293, 161–163.
[5] Klein, U. (1987) Insect Biochem. 17, 1193–1204.
[6] De Kramer, J.J. and Hemberger, J. (1987) in: Pheromone Biochemistry (Prestwich, G.D. and Blomquist, G.J. eds) pp. 433–472, Academic, Orlando, FL.
[7] Vogt, R.G. (1987) in: Pheromone Biochemistry, (Prestwich, G.D. and Blomquist, G.J. eds) pp. 385–431, Academic, Orlando, FL.
[8] Kaissling, K.E., Klein, U., De Kramer, J.J., Keil, T.A., Kanaujia, S. and Hemberger, J. (1985) in: Molecular Basis of Nerve Activity (Changeux, J.P., Hucho, F., Maelicke, A. and Neumann, E. eds), pp. 173–183, De Gruyter Berlin.
[9] Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
[10] Aviv, H. and Leder, P. (1972) Proc. Natl. Acad. Sci. USA 69, 1408–1412.
[11] Gubler, V. and Hoffman, B.J. (1983) Gene 25, 263–269.
[12] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[13] Feinberg, A. and Vogelstein, B. (1984) Anal. Biochem. 137, 266–267.
[14] Klein, U., Kehl, M., Hemberger, J. and Günzel, A. (1987) Chem. Senses 12, 211.
[15] Birnstiel, M.L., Busslinger, M. and Strub, K. (1985) Cell 41, 349–359.
[16] Von Heyne, G. (1986) Nucleic Acids Res. 14, 4683–4690.
[17] Kyte, J. and Doolittle, R.F. (1982) J. Mol. Biol. 157, 105–132.