Vol. 189, No. 1, 1992 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
November 30, 1992 Pages 221-226

cDNA CLONING AND SEQUENCE DETERMINATION OF THE PHEROMONE BIOSYNTHESIS ACTIVATING NEUROPEPTIDE OF THE SILKWORM, BOMBYX MORI

Tsuyoshi Kawano, Hiroshi Kataoka, Hiromichi Nagasawa,
Akira Isogai and Akinori Suzuki

Department of Agricultural Chemistry, Faculty of Agriculture,
The University of Tokyo, Bunkyo-ku, Tokyo 113, Japan

Received October 14, 1992

We have identified the cDNAs encoding pheromone biosynthesis activating neuropeptide (PBAN) using PCR technique. The nucleotide sequence showed that the PBAN gene encodes, besides PBAN, diapause hormone and three putative amidated peptides. These four peptides share with PBAN the C-terminal pentapeptide amide which is corresponding to the shortest fragment with pheromonotropic activity. The organization of the PBAN gene is characteristic of several short neuropeptides and has some degree of similarity to that of the gene for the insect neuropeptide FMRFamide. Thus, the PBAN gene products construct a family of structurally related peptides and have various biological functions.

A neuropeptide named pheromone biosynthesis activating neuropeptide (PBAN) promotes the biosynthesis of sex pheromone in several lepidopteran insects (1,2). Recently, PBAN was chemically characterized in corn earworm, Heliothis zea, and in silkmoth, Bombyx mori (3,4). Bombyx-PBAN (Bom-PBAN) is a 33-residue peptide with a C-terminal amide and has 80% sequence homology with Heliothis-PBAN (Hez-PBAN). The C-terminal pentapeptide amide FXPRL-NH_2 is also found in diapause hormone (DH) in B. mori (5) and in the pyrokinin family of peptides with myotropic activity from the locust (6) and the Madeira cockroach (7). In fact, their common C-terminal pentapeptide amide has been proved to show both pheromonotropic and myotropic activities (8).

In this report, we describe the nucleotide sequence of the entire cDNA coding for Bom-PBAN. The deduced amino acid sequence indicated that the PBAN gene products could consist of DH and three other PBAN related peptides, besides PBAN. Interestingly, they share the C-terminal pentapeptide amide FXP(R or K)L-NH_2 possibly with myotropic activity. The present study will be an access to understanding the evolution of genetic structure together with separation of biological activities of these peptides.

MATERIALS AND METHODS

GENERAL PROCEDURE: DNA blot analyses, subcloning of the PCR-amplified DNA into Bleuscript (Stratagene) and preparation of [^{32}P] labeled nucleotides were

0006-291X/92 $4.00

carried out by the standard methods (9). PCR was performed on a DNA thermal cycler (ATTO) utilizing Thermus aquaticus (Taq) polymerase purchased from Takara.

Isolation of Bombyx mRNA: Total cellular RNA was extracted from 2500 brain-suboesophageal ganglion complexes of the female adults of B. mori using the AGPC method (10). Poly(A)$^+$RNA was isolated by oligo(dT)-latex chromatography (Takara).

Synthesis and Design of Oligonucleotides: Oligonucleotides were synthesized on an Applied Biosystem model 380B DNA synthesizer and purified using NENSORB PREP cartridges (Du Pont). Primers 1 and 2 were designed on the basis of the partial amino acid sequences of PBAN. The linker primer is $(dT)_{18}$ followed by XhoI site. Design of anchored primers was performed as described by Loh et al. (11). Other nucleotides for PCR primers were designed according to the nucleotide sequences of amplified DNA. Details are described in Fig. 1.

cDNA Amplification by PCR: The procedure is briefly outlined in Fig. 1. A single-stranded cDNA was generated by using a kit for cDNA synthesis (Stratagene). The resulting cDNA was tailed with dG by terminal deoxynucleotidyl transferase (Takara), and used for PCR. 5'-end of PBAN cDNA was amplified by the "anchored PCR" method (11), and 3'-end amplification of PBAN cDNA was carried out as described by Ohara et al. (12). Details of PCR conditions will be described elsewhere.

Nucleotide Sequence Analysis: PCR products were subcloned into Bluescript and screened by DNA blot analyses. The resulting positive clones were sequenced by the chain termination method (13) using a BcaBEST DNA sequencing kit (Takara). The sequences of several region were determined with synthetic primers. In order to eliminate artifacts due to low fidelity of Taq polymerase (14), more than 10 independent clones were analyzed for each PCR product.

RESULTS

To obtain the cDNA encoding Bom-PBAN, we utilized a PCR-assisted cDNA cloning strategy (11,12), which is briefly illustrated in Fig. 1. We first adopted the anchored PCR with degenerated primers 1, 2 and "anchored primers", because PBAN is a very short peptide and includes serine, leucine and arginine residues corresponding to highly degenerated codons. Unfortunately, an amplified DNA fragment (87 bp) was much shorter than we had expected. However, the deduced amino acid sequence from DNA sequence of the fragment was identical to N-terminal 10 residues of PBAN (Leu^1-Ala^{10}). We next tried to amplify 3'-end cDNA using one-sided PCR strategy with primers 3 and 4 which were designed on the basis of the DNA sequence of the anchored PCR product as shown in Fig. 1. Interestingly, the nucleotide sequence of the 3'-end cDNAs were separated into two types. The deduced amino acid sequence of one type was completely identical to amino acid sequence of PBAN (Asp^{10}-Leu^{33}). In order to figure out the entire structure of the PBAN cDNA, we tried to amplify 5'-end cDNA including the open-reading frame by the anchored PCR method again (Fig. 1). The amino acid sequence deduced from the DNA sequence of the amplified fragment included the entire amino acid sequence of PBAN, as expected. The nucleotide sequences obtained from 5'- and 3'-end cDNA amplification were

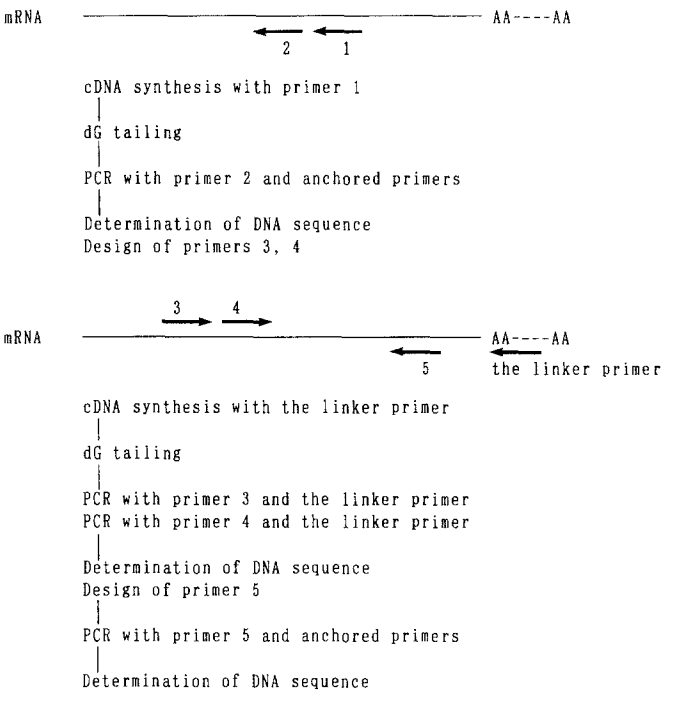

Fig. 1. Schematic outline of procedures of cDNA amplification. Primers 1 and 2 are 3'-CTRGGNCTYCTYTACCTY-5', 3'-CTRGTYCTYTACATRGTYGG-5' which were derived from possible cDNA sequences corresponding to the parts of amino acid sequences of Asp^{18}-Glu^{23} and Asp^{11}-Pro^{17} of PBAN, respectively (where, N=A, C, G, or T; R=A or G; Y=C or T). Primer 3, 5'-CTCGGAAGGCGGCTCTCTGA-3'; Primer 4, 5'-GACATGCCTGCTACGCCAGC-3'; Primer 5, 3'-GTAAACTCCAGAGCTAGCTGAGG-5'; the primer-linker, 3'-TTTTTTTTTTTTTTTTTTGAGCTC-5'. Anchored primers are mixture of two nucleotides 5'-GCATGCGCGCGGCCGCGGAGGCCCCCCCCCCCCCCCC-3' and 5'-GCATGCGCGCGGCCGCGGAGGCCC-3' at a molar ratio 1:24.

combined with the internal cDNA sequence to yield the full-length PBAN cDNA sequence (Fig. 2-a).

Fig. 2-a. shows the 794-nucleotide sequence of the cDNA for Bom-PBAN and its deduced amino acid sequence. We could find an open-reading frame (192 amino acid residues); in contrast with the mature PBAN consisting of 33 residues. In 5'-nontranslated region, a transcriptional consensus (15), ATCACTT, is located at the 5' end, and a translational consensus sequence (16), AACAACAAAA, is situated upstream of the first initiation codon ATG. The initial methionine is in frame with the peptide sequence of PBAN, and precedes a short sequence composed of 23 amino acid residues which has the typical feature of eukaryotic signal peptides (17). A termination codon TAA is found in frame 212 nucleotides upstream of the poly(A)tail addition site. The 3'-nontranslated region consists of 172 nucleotides, excluding the poly(A) tail, and three sequences that match the consensus AATAAA site for poly(A) addition are found in this region. Fig. 2-b shows the putative proteolytic sites and the resulting peptides in the precursor. The precursor contains two distinct insect neuropeptides (DH and PBAN) and three amino acid sequences flanked by potential endoproteolytic cleavage sites (18), single and dibasic

a

```
                                          -1 1
-40  ATCACTTCGCCCTCCAACCACTGAAGGGACAACAACAAAA ATG TAT AAA ACC AAC ATT GTT TTC AAC GTT TTA GCT TTG
                                             Met Tyr Lys Thr Asn Ile Val Phe Asn Val Leu Ala Leu

 40  GCA TTG TTC AGT ATT TTC TTC GCG AGT TGC ACG GAT ATG AAG GAT GAA AGC GAC AGA GGA GCT CAC AGT
     Ala Leu Phe Ser Ile Phe Phe Ala Ser Cys Thr Asp Met Lys Asp Glu Ser Asp Arg Gly Ala His Ser

109  GAG CGG GGC GCT CTC TGG TTC GGC CCC AGA CTC GGG AAG CGA TCA ATG AAG CCA TCC ACT GAA GAT AAC
     Glu Arg Gly Ala Leu Trp Phe Gly Pro Arg Leu Gly Lys Arg Ser Met Lys Pro Ser Thr Glu Asp Asn

178  AGG CAA ACC TTC CTG AGG CTG CTC GAG GCG GCT GAT GCC CTC AAA TTT TAT TAC GAC CAG CTA CCT TAC
     Arg Gln Thr Phe Leu Arg Leu Leu Glu Ala Ala Asp Ala Leu Lys Phe Tyr Tyr Asp Gln Leu Pro Tyr

247  GAG AGG CAA GCC GAT GAA CCG GAA ACC AAA GTA ACA AAG AAG ATC ATC TTC ACC CCC AAA CTC GGG AGG
     Glu Arg Gln Ala Asp Glu Pro Glu Thr Lys Val Thr Lys Lys Ile Ile Phe Thr Pro Lys Leu Gly Arg

316  AGC GTC GCC AAA CCC CAG ACG CAT GAA AGC CTC GAA TTC ATC CCC CGG CTC GGA AGG CGG CTC TCT GAG
     Ser Val Ala Lys Pro Gln Thr His Glu Ser Leu Glu Phe Ile Pro Arg Leu Gly Arg Arg Leu Ser Glu

385  GAC ATG CCT GCT ACG CCA GCT GAC CAG GAA ATG TAC CAA CCT GAC CCC GAA GAA ATG GAG TCA AGA ACA
     Asp Met Pro Ala Thr Pro Ala Asp Gln Glu Met Tyr Gln Pro Asp Pro Glu Glu Met Glu Ser Arg Thr

454  AGA TAC TTC TCG CCC AGG CTG GGG CGC ACC ATG AGC TTT TCG CCC AGA CTG GGA AGG GAG CTT TCG TAC
     Arg Tyr Phe Ser Pro Arg Leu Gly Arg Thr Met Ser Phe Ser Pro Arg Leu Gly Arg Glu Leu Ser Tyr

523  GAT TAC CCT ACA AAA TAT AGG GTT GCC AGA AGC GTT AAC AAG ACA ATG GAC AAC TAAACGAATTATGGTCCG
     Asp Tyr Pro Thr Lys Tyr Arg Val Ala Arg Ser Val Asn Lys Thr Met Asp Asn ***

595  CTTGAGGTACCTCATTTGAGGTCTCGATCGACTCCGACGAACGGTTACGGGTAAACGGCGACAATGTTAATGTTTTGGACGAAACAATTG
685  TTAATTAATAAATTCATGTGATTTTGTAATTGTAATTTATAAGTGAATAAAAAAATAAACTATTTAAAAT(polyA)
```

b

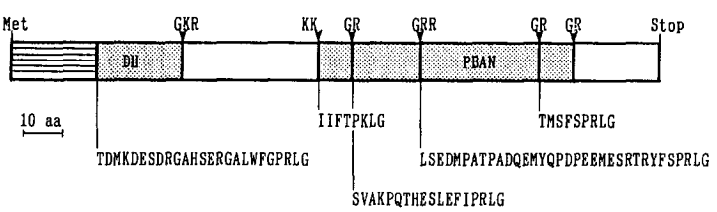

Fig. 2. Nucleotide sequence of cDNA encoding Bom-PBAN, its deduced amino acid sequence and constitution of the precursor.
a) Nucleotide sequence of the cDNA for Bom-PBAN and its deduced amino acid sequence. Nucleotides are numbered from the presumed initiator ATG. Boxes show putative polyadenylylation signals. The amino acid sequence of PBAN is underlined. Arrows indicate primers 3, 4 and 5 used for PCR.
b) Schematic representation of the precursor peptide of Bom-PBAN. The protein has a 23-residue signal peptide (horizontal lines). Possible processing signals (18) are shown by arrowheads. The resulting products after putative proteolytic cleavages are represented as boxes. The stippled boxes indicate DH, PBAN and three putative peptides which have the similar sequence to the C-terminal pentapeptide of PBAN. Their amino acid sequences including a glycine residue at the C-terminus are listed.

residues. Each of them ends in a glycine residue which donates a C-terminal amide (19). They share the sequence FXP(R or K)LG (X=G, T, I or S).

DISCUSSION

In this study, we have obtained the cDNA coding for Bom-PBAN utilizing a PCR-assisted cDNA cloning strategy. The strategy adopted in this experiment may useful for obtaining cDNA corresponding to low-abundance mRNA, especially in the case of short peptides.

In vertebrate and invertebrate species, the genes for short neuropeptides often encode multiple copies of structurally related peptides (20,21), including the case of insect neuropeptide FMRFamide (22).

Interestingly, the Bom-PBAN gene encodes ,besides PBAN, DH which is known to be another neuropeptide involved in reproduction in B. mori (5). Furthermore, DH shares with PBAN the C-terminal pentapeptide amide FXPRL-NH_2 (where, X=G or S), which shows myotropic activity (8). We propose that the Bom-PBAN gene might code for, besides DH and PBAN, three insect neuropeptides IIFTPKL-NH_2, SVAKPQTHESLEFIPRL-NH_2 and TMSFSPRL-NH_2. These three peptides are predicted to show pheromonotropic, myotropic and/or DH activities, because they also share with PBAN and DH the C-terminal pentapeptide amide FXP(R or K)L-NH_2. Considering the structural similarity, the PBAN gene products and the pyrokinin family peptides constitute a superfamily, although the pyrokinin family has not yet been identified in the lepidopteran insects. The Bom-PBAN gene may carry five insect neuropeptides amidated at the C-terminus, indicating the organization of the Bom-PBAN gene is somewhat similar to that of the gene for the insect neuropeptide FMRFamide (22), which encodes many repeat of such peptides as containing a C-terminal tetrapeptide FMRF-NH_2. In lepidopteran insects, however, the PBAN gene is the first example of the genes encoding multiple copies of structurally related peptides in one transcript.

Quite recently, the nucleotide sequence of the genomic DNA for Hez-PBAN was reported by Davis et al. (23). They indicated that the initial methionine corresponding to the putative initiator GTG preceded the proposed open-reading frame (70 amino acid residues). Based on the comparison between our result and their proposal, the reported Hez-PBAN gene may not contain the entire structure of the PBAN gene.

In conclusion, the PBAN gene products, besides affecting sexual behavior and diapause, may be related to many biological process in B. mori. We hope further study of the gene products will provide an insight into the molecular evolution of insect neuropeptides.

ACKNOWLEDGMENTS

We are grateful to Dr. H. Fugo of Tokyo University of Agriculture and Technology for supplying us the silkmoth, Bombyx mori. We appreciate fruitful technical advice and discussion from Drs. S. Tomino and S. Izumi of Tokyo Metropolitan University. This work was partly supported by Grants-in-Aid for Scientific Research (No. 01060004) from the Ministry of Education, Science and Culture of Japan, by a fund for Japan Society for the Promotion of Science and by a special coordination fund for promoting science and technology in the basic research core system by the Science and Technology Agency (STA).

REFERENCES

1. Raina, A.K. and Klun, J.A (1984) Science **225**, 351-533.
2. Ando, T., Arima, R., Uchiyama, M., Nagasawa, H., Inoue, T. and Suzuki, A. (1988) Agric. Biol. Chem. **52**, 881-883.

3. Raina, A.K., Jaffe, H., Kempe, T.G., Keim, P., Blancher, R.W., Fales, H.M., Klun, J.A., Ridgway, R.L. and Hayes, D.K. (1989) Science **244**, 796-798.
4. Kitamura, A., Nagasawa, H., Kataoka, H., Inoue, T., Ando, T. and Suzuki, A. (1989) Biochem. Biophys. Res. Commun. **163**, 520-526.
5. Imai, K., Konno, T., Nakazawa, Y., Komiya, T., Isobe, M., Koga, K., Goto, T., Yaginuma, T., Sakakibara, K., Hasegawa, K. and Yamashita, O. (1991) Proc. Japan Acad. **67**, 98-101.
6. Schoofs, L., Holman, G.M., Hayes, T.K., Nachman, R.J. and DeLoof, A. (1990) Insect Biochm. **20**, 479-484.
7. Holman, G.M., Nachman, R.J., Schoofs, L., Hayes, T.K., Wright, M.S. and DeLoof, A. (1991) Insect Biochem. **21**, 107-112.
8. Kuniyoshi, H., Nagasawa, H., Ando, T., Suzuki, A., Nachman, R.J. and Holman G.M. (1992) Biosci. Biotech. Biochem. **56**, 167-168.
9. Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular Cloning (2nd ed.) Cold Spring Harbor Laboratory, New York.
10. Chomczynski, P. and Sacchi, N. (1987) Anal. Biochem. **162**, 156-159.
11. Loh, E.Y., Elliott, J.F., Cwirla, S., Lanier, L.L and Davis, M.M. (1989) Science **243**, 217-220.
12. Ohara, O., Dorit, R.L. and Gilbert, W. (1989) Proc. Natl. Acad. Sci. USA **86**, 5673-5677.
13. Sanger, F., Nichlen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA. **74**, 5463-5467.
14. Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuhci, R., Horn, G.T., Mullis, K.B. and Erlich, H.A. (1988) Science **239**, 487-489.
15. Hultmark, D., Klemenz, R and Gehring, W.J. (1986) Cell **44**, 429-438.
16. Douglas, R.C. (1987) Nuc. Acids Res. **15**, 1353-1361.
17. Von Heijne, G. (1986) Nuc. Acids Res. **14**, 4683-4690.
18. Fisher, J.M. and Scheller, R.H. (1988) J. Biol. Chem. **263**, 16515-16518.
19. Bradbury, A.F., Finnie, M.D.A. and Smyth, D.G. (1982) Nature **298**, 686-688.
20. Masaharu, N., Furutani, Y., Takahashi, H., Toyosato, M., Hirose, T., Inayama, S., Nakanishi, S. and Numa, S. (1982) Nature **295**, 202-206.
21. Lechan, R.M., Wu, P., Jackson, I.M.D., Wolf, H., Cooperman, S., Mandel, G. and Goodman, R.H. (1986) Science **231**, 159-161.
22. Schneider, L.E. and Taghert, P.H. (1988) Proc. Natl. Acad. Sci. USA **85**, 1993-1997.
23. Davis, M.B., Nakharia, V.N., Henry, J., Kemre, T.G. and Raina, A.K. (1992) Proc. Natl. Acad. Sci. USA **89**, 142-146.