Proc. Natl. Acad. Sci. USA
Vol. 86, pp. 6843–6847, September 1989
Biochemistry

Structure and organization of four clustered genes that encode bombyxin, an insulin-related brain secretory peptide of the silkmoth *Bombyx mori*

(insect development/neurosecretory hormone/multi-gene family)

ATSUSHI KAWAKAMI*, MASAFUMI IWAMI*, HIROMICHI NAGASAWA†, AKINORI SUZUKI†, AND HIRONORI ISHIZAKI*‡

*Biological Institute, Faculty of Science, Nagoya University, Nagoya 464–01, Japan; and †Department of Agricultural Chemistry, Faculty of Agriculture, The University of Tokyo, Tokyo 113, Japan

Communicated by Donald F. Steiner, May 23, 1989

ABSTRACT **Four genes encoding bombyxin have been located in a 14-kilobase *Bombyx* genomic DNA segment. All of these genes encode preprobombyxin, the precursor molecule for bombyxin, with the domain organization of signal peptide/B chain/C peptide/A chain. Bombyxins are classified as family A or B according to their sequence homology. Two genes, each belonging to a different family, are closely apposed to form a pair with opposite orientation, presumably forming a regulatory unit for transcription. Genomic Southern blot hybridization suggested that there are many such gene pairs in the *Bombyx* genome. Differences between bombyxin genes and vertebrate insulin-family genes indicate that different mechanisms operate in the evolution of invertebrate and vertebrate insulin-family genes.**

Bombyxin, a 5-kDa brain secretory peptide of the silkmoth *Bombyx mori*, belongs to the insulin family (1, 2). Bombyxin has been purified by using its biological activity of stimulating the prothoracic glands to release ecdyson when applied to another moth, *Samia cynthia ricini* (3), but its physiological function in *Bombyx* is still obscure. Bombyxin is a heterogeneous group of molecular species that differ by only a few amino acid substitutions. Bombyxin molecules are heterodimers of A and B chains whose amino acid sequences are homologous with vertebrate insulin-family peptides (1, 4–6). The A and B chains of bombyxins are linked together by disulfide bonds, exactly as is insulin (7). We have cloned and characterized two distinct bombyxin cDNAs (8) and one bombyxin gene (9). Genomic Southern hybridization analyses using a bombyxin cDNA as probe indicated that multiple copies of bombyxin gene, as many as about 20, existed in the *Bombyx* genome. We now report the cloning and characterization of four bombyxin genes[§] that form a cluster in a 14-kilobase (kb) region of the *Bombyx* genome.

MATERIALS AND METHODS

Cloning and Sequencing of Bombyxin Genes. Clone λ4K105, one of the four clones previously isolated (9), was used in this study. The nucleotide sequence was determined by the dideoxynucleotide chain-termination method (10) with overlapping deletion plasmids created by unidirectional digestion with exonuclease III and mung bean nuclease (11). Nucleotide sequence data were assembled and analyzed by using programs in DNASIS (Hitachi Software Engineering).

Primer Extention. Oligonucleotide primers were chemically synthesized. They are 5′-CGTGTGCACTCTTTGTGGCTGTTG-3′ (family A primer; as is described in detail in *Results*, bombyxin genes are classified into families A and B according to their nucleotide sequence homology) and 5′-ACACATCAGACTGATCACGATTAC-3′ (family B primer). The reaction was done with avian myeloblastosis virus reverse transcriptase (12) using 1.5 μg of *Bombyx* brain poly(A)$^+$ RNA from fifth-instar larvae.

Probes for Northern and Southern Hybridizations. As a probe for detecting family A members in Northern and Southern hybridization analyses, we used bombyxin gene A-1 (positions 505–1045 in figure 1 of ref. 9). For the detection of family B members, bombyxin gene B-1 (positions 370–996, see Fig. 2) was used as probe. The DNA fragments were nick-translated with [α-^{32}P]dCTP to give the specific activities of 1–3 × 10^8 cpm/μg of DNA.

RNA Preparation and Northern Blot Analysis. The total RNA was "mini-prepared" by AGPC method (13) from various tissue sources of *Bombyx*. Aliquots (10 μg) of total RNA were electrophoresed on formaldehyde-denaturing 1.2% agarose gels (14) and transferred to nylon membrane filters with 20× SSC (3 M NaCl/0.3 M sodium citrate, pH 7.0). After prehybridization, the filters were hybridized at 37°C for 24 hr in 50% (vol/vol) formamide/5× SSC/5× Denhardt's solution (1× Denhardt's solution = 0.02% Ficoll/0.02% polyvinylpyrrolidone/0.02% bovine serum albumin)/0.01% tRNA/0.01% denatured salmon sperm DNA/20 mM sodium phosphate, pH 6.5/0.1% SDS/10% (wt/vol) dextran sulfate and the probes at 1.5 × 10^6 cpm/ml. The filters were washed in 2× SSC/0.1% SDS at 60°C for 3 hr followed by 0.2× SSC/0.1% SDS at 65°C for 1.5 hr.

Genomic DNA Preparation and Southern Blot Analysis. High molecular weight DNA was prepared from *Bombyx* silk glands of fifth-instar larvae as described (15). Aliquots (5 μg) of the genomic DNA were digested with restriction enzymes, separated by electrophoresis on 0.7% agarose gel, and alkali-transferred (16) to the nylon membrane filters. Hybridization was done at 60°C for 36 hr in 6× SSC/10× Denhardt's solution/20 mM sodium phosphate, pH 6.5/0.02% yeast tRNA/0.02% denatured salmon sperm DNA/0.5% SDS/10% dextran sulfate and the probes at 4–5 × 10^6 cpm/ml. The filters were washed as for Northern blot hybridization.

RESULTS

Overall Organization of Four Bombyxin Genes in the *Bombyx* Genomic DNA Insert in Clone λ4K105. Fig. 1 illustrates the restriction map of the *Bombyx* genomic DNA insert in a λ4K105 clone. This DNA hybridized with probes A and B at

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: MIP, molluscan insulin-related peptide.

‡To whom reprint requests should be addressed.

§The sequences reported in this paper have been deposited in the GenBank data base (accession nos. M26068, M26069, and M26070).

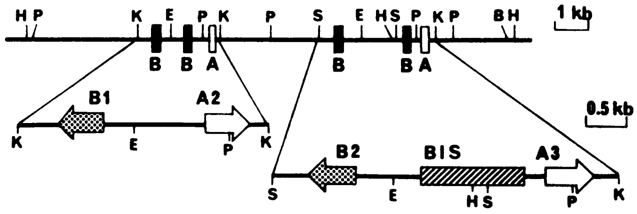

FIG. 1. Restriction map of a *Bombyx* genomic DNA segment in clone λ4K105. Fragments that hybridized to oligonucleotide probes A and B are indicated by open and solid bars, respectively. Sequenced regions are depicted in expanded scale to show four preprobombyxin genes. The open and stippled arrows represent family A (A-2 and A-3) and family B (B-1 and B-2) genes, respectively, and show the direction of transcription. BIS, a transposon-like inserted element; B, *Bam*HI; E, *Eco*RI; H, *Hin*dIII; K, *Kpn* I; P, *Pst* I; S, *Sal* I.

two and four sites, respectively. The hybridizing sites apparently formed two clusters and the nucleotide sequence was determined for these two regions (Fig. 2) to reveal four bombyxin genes (as shown in expanded scale in Fig. 1). As is discussed in the following section, bombyxin genes were classified into two families, termed families A and B. Two genes, each of which belongs to the different families, are paired (genes B-1/A-2 and B-2/A-3) with opposite orientation. Genes B-1 and A-2 are ≈1.1 kb apart and genes B-2 and A-3 are ≈2.3 kb apart. The wider space between B-2 and A-3 is due to the presence of an insertion sequence, designated BIS (≈1.2 kb). The failure of probe A to hybridize with genes B-1 and B-2, as shown in the restriction map, is accounted for by a low degree of the sequence matching between the probe and the A-chain portion of these genes.

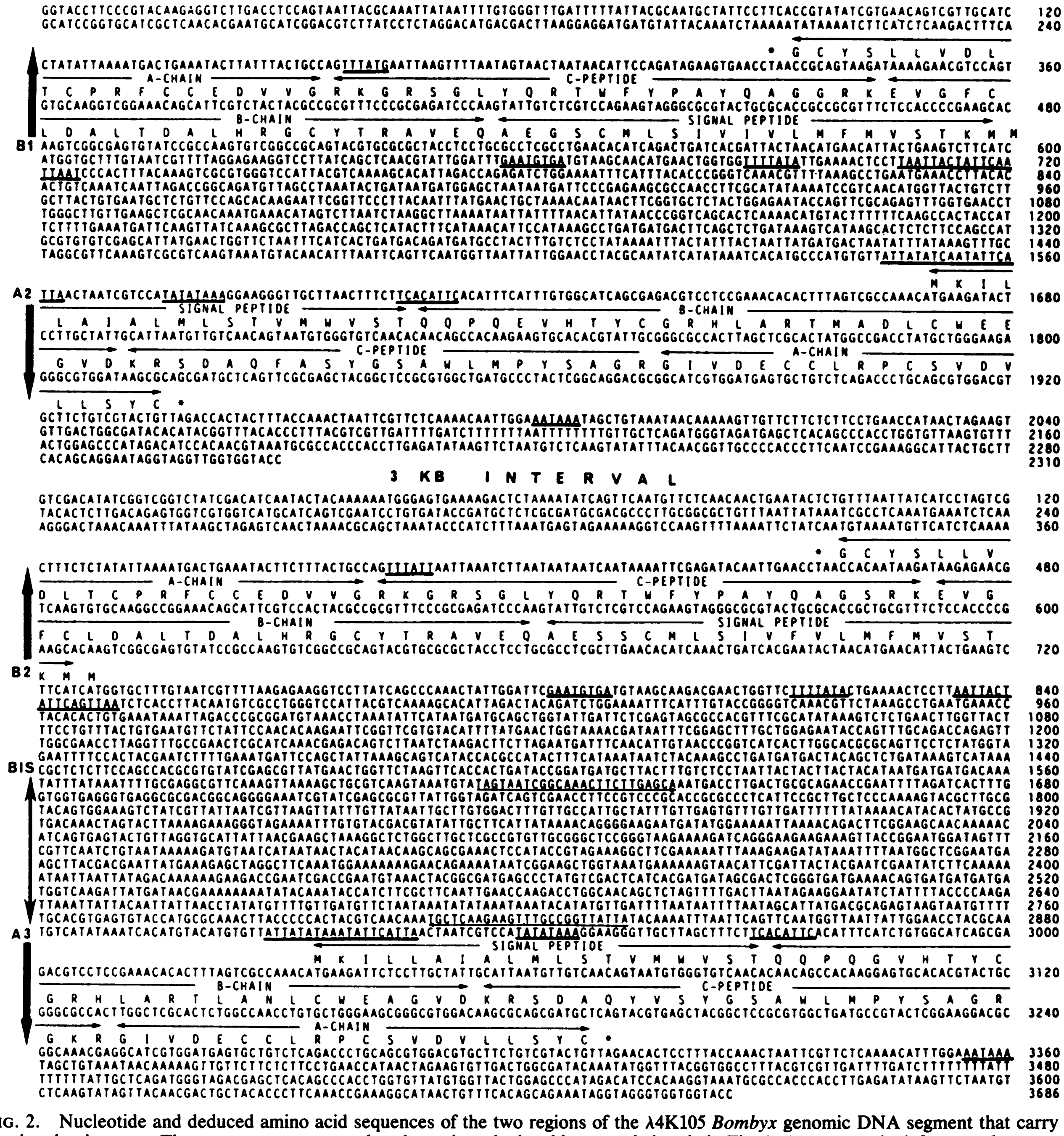

FIG. 2. Nucleotide and deduced amino acid sequences of the two regions of the λ4K105 *Bombyx* genomic DNA segment that carry four preprobombyxin genes. The sequences correspond to the regions depicted in expanded scale in Fig. 1. Arrows on the left cover the respective genes and show the direction of transcription. Stop codons are marked with asterisks. "TATA" boxes, polyadenylylation signals, cap consensus sequences, and A+T-rich sequences are underlined. BIS is a transposon-like inserted sequence with terminal inverted repeats indicated by horizontal arrows below the sequence.

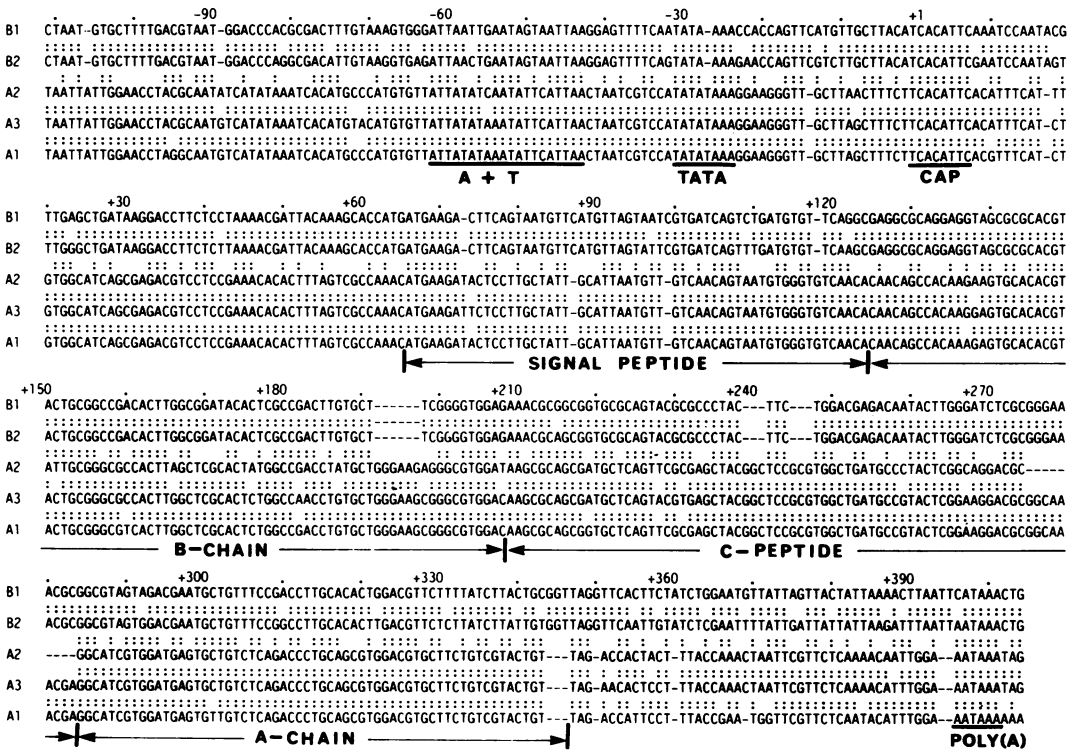

FIG. 3. Alignment of coding and flanking regions of five preprobombyxin genes. Gene A1 has been sequenced (9). Identical nucleotides between the neighboring sequences are indicated by colons.

Structure of Bombyxin Genes and Their Flanking Regions and Classification of Bombyxin Genes into Families A and B. Nucleotide sequences of the two regions containing the bombyxin genes are shown in Fig. 2. Each gene contains two or three candidates for the translation initiation codon but we postulate the initiation sites to be those indicated in Fig. 2 because of the presence of general features for a signal peptide. The open reading frame of all bombyxin genes apparently encodes a preprobombyxin, which is composed of four domains in the order signal peptide/B chain/C peptide/A chain. The structure of preprobombyxins suggests that they are posttranslationally modified to form mature bombyxins through the excision of the C peptide and the conversion of the glutamine at the N terminus of the B chain to pyroglutamate. These inferences have been made (8, 9) for the bombyxin gene and cDNAs characterized.

In Fig. 3 the coding and flanking sequences of the four genes reported here and of the characterized bombyxin gene (9) are aligned. The sequence homology is remarkably high (≈90%) between the upper two sequences and among the lower three sequences, whereas far less homology (≈50%) is seen between these two groups. Based on this sequence comparison, we tentatively classify the bombyxin genes into two families, families A and B. The characterized gene (9) belongs to family A and we now designate it gene A-1. The nucleotide sequence of λBb360 cDNA (8) completely matches the corresponding region of gene A-2. All of the genes lack introns.

A+T-rich, "TATA-box," transcription-start, and polyadenylylation-signal sequences are highly conserved throughout the five genes with respect to their sequences and positions (Fig. 3). The transcription start sites were determined by primer extension (data not shown). Besides these sites, no conserved sequences were found in the flanking regions including sequences outside of the area shown in Fig. 3.

When two gene pairs, B-1/A-2 and B-2/A-3, are compared by Harr plot homology analysis, the homology in the flanking regions extends for a considerable distance in both the 3′ and 5′ directions, except for the BIS insertion (Fig. 4). Thus gene pairs B-1/A-2 and B-2/A-3 form a highly homologous tandem repeat, suggesting that these gene pairs must have been generated by duplication of an original gene pair.

The BIS insertion has 25-base-pair inverted repeats at both ends (Fig. 2) that are flanked by 2-base-pair (TA) target-site duplications, indicating that this insertion may be a transposable element (17). A GenBank homology search (release 58.0) failed to find sequence homology between BIS and any insertion or transposon sequence thus far identified.

Amino Acid Sequence Comparisons of Preprobombyxins, Bombyxins, and Other Insulin-Family Peptides. The sequences of preprobombyxins deduced from family A and B genes, a preprobombyxin deduced from the λBb204 cDNA (8), and native mature bombyxins (1, 4, 6) are aligned in Fig. 5 with the sequences of human insulin and the molluscan

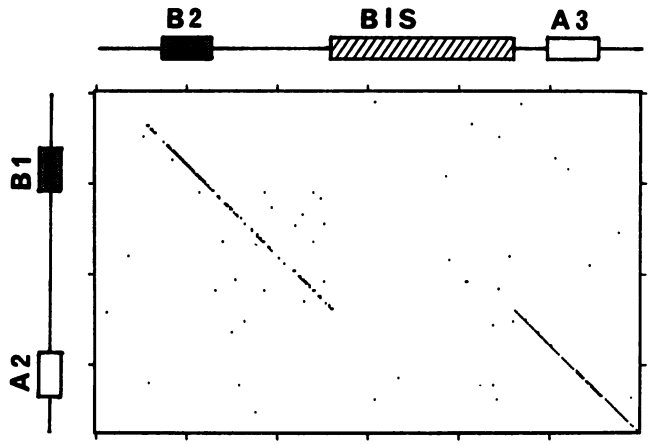

FIG. 4. Harr plot dot matrix for the two segments carrying gene pairs B-1/A-2 and B-2/A-3. DNAS 1 (V 3.0) of DNASIS (Hitachi Software Engineering) was used. The check-size/match-base number is 9/9.

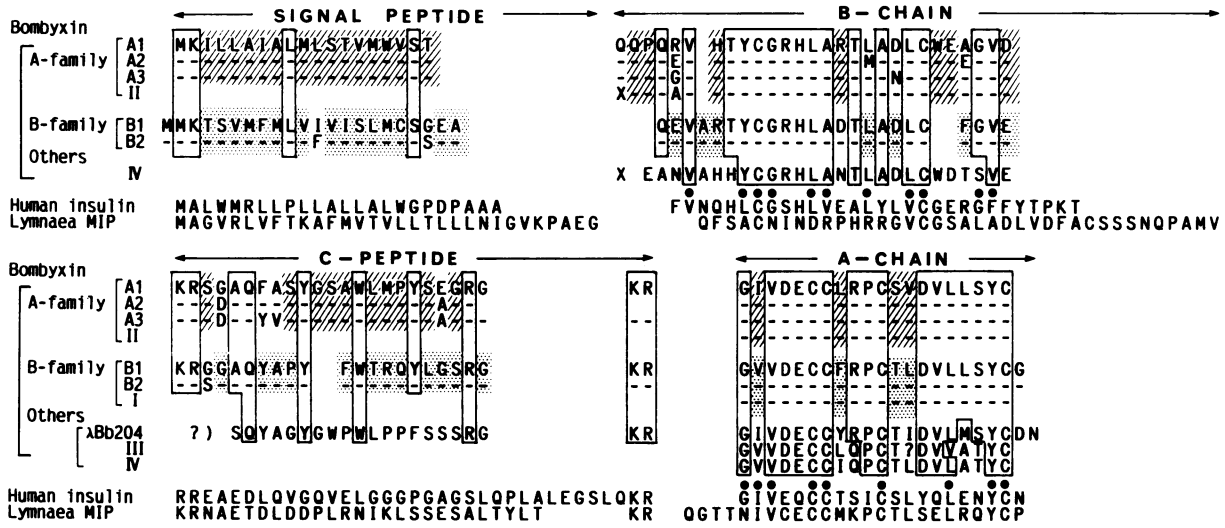

FIG. 5. Amino acid sequence comparison between preprobombyxins, bombyxins, and other insulin-family prepropeptides. Preprobombyxins A-2, A-3, B-1, and B-2 were deduced from the genes characterized in the present study. Preprobombyxin A-1 was deduced from the gene (9) and λBb204 is a part of the putative preprobombyxin deduced from the λBb204 cDNA (8). Bombyxin-I through -IV are mature peptides purified from *Bombyx* heads (1, 4, 6). Prepropeptides of human insulin (18) and molluscan insulin-related peptide (19) are also aligned. Gaps were introduced for maximum alignment. Homologous residues throughout bombyxin families A and B are boxed and the boxes are extended where identity was observed to other bombyxin molecules that do not belong to these families. Residues that are conserved only within bombyxin families A or B are indicated by stippled and hatched areas, respectively. The residues important for the formation of insulin tertiary structure (5), which are conserved or conservatively substituted in bombyxins, are indicated by the circles above the insulin sequence. X, pyroglutamate. The single-letter amino acid code was used.

insulin-related peptide (MIP, ref. 19). The homology is high within family A or family B but is low between families A and B. Bombyxin-II appears to belong to family A, whereas bombyxin-I seems to belong to family B, although these assignments are preliminary until the genes are characterized and the entire nucleotide sequences are compared. The amino acid sequences of bombyxin-III and -IV and of the putative partial preprobombyxin deduced from the λBb204 cDNA are quite different from other bombyxins and cannot be assigned to either family A or B, suggesting that still other bombyxin families remain to be defined. A very low degree of homology is observed between bombyxins and *Lymnaea* MIP. This large difference between insect and molluscan peptides suggests extensive evolutionary diversification of invertebrate insulin-related peptides.

When amino acid sequences of family A and B bombyxins were compared with the sequence of insulin, striking similarities were found between residues that contribute to the tertiary structure of insulin and corresponding residues in bombyxins, as was discussed (5) mainly for bombyxin-II. Thus, all residues responsible for the hydrophobic core formation are conserved or substituted by hydrophobic residues (A2, A3, A16, A19, B2, B6, B11, B12, B15, B18, and B24; taking the N termini of the A and B chains of insulin as position 1); A1, B8, and B23 are glycine, which contributes to the main chain conformation; all cysteines are conserved. Thus it seems highly probable that family A and B bombyxins have an insulin-like globular structure.

Bombyxin Family A and B Genes Are Expressed in Brain. To study the tissue-specific expression of family A and B bombyxin genes, Northern blot hybridization of RNA from various *Bombyx* tissues from day-0 fifth-instar larvae and freshly ecdysed pupae was performed using genes A-1 and B-1 as probe. Preliminary hybridization experiments using these probes and subcloned family A and B bombyxin DNAs had shown that these probes hybridized specifically to their respective family DNAs under stringent conditions. RNA was isolated from subesophageal ganglion, fat body, silk gland, brain, malpighian tubule, ovary, and testis. Northern blots of these RNAs were made and hybridized to probes A and B. Brain was the only tissue examined that contained a transcript (≈0.6 kb) that hybridized to both probes, indicating that family A and B genes are specifically expressed only in the brain.

Numerous Copies of Family A and B Genes Exist in the *Bombyx* Genome. As shown in Fig. 6, multiple bands appeared when *Bombyx* genomic DNA digested with various restriction enzymes was hybridized with bombyxin genes A-1 and B-1 as probes, indicating that many copies of family A and B bombyxin genes exist in the *Bombyx* genome. The 1-kb *Hin*dIII fragment detected by the family A probe contained a copy of bombyxin gene A-1 and the 2.5-kb *Sal* I fragment revealed by the family B probe carried a copy of bombyxin gene B-1. By using the band intensity of these fragments, we estimated the number of copies of family A and B genes to be

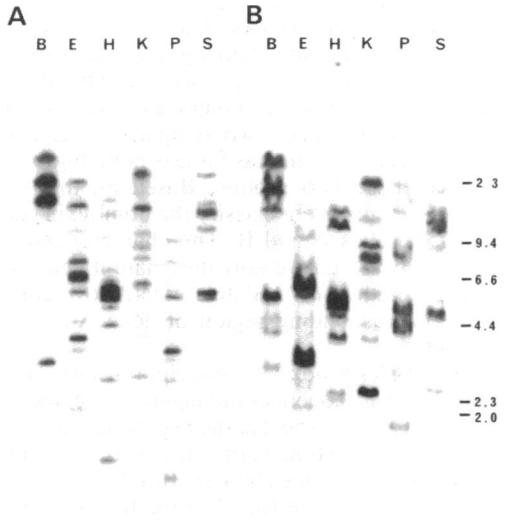

FIG. 6. Southern genomic blot analysis. *Bombyx* genomic DNAs digested with various restriction enzymes as indicated by lane labels (abbreviations are as in Fig. 1) were probed by bombyxin genes A-1 (*A*) and B-1 (*B*).

12 and 10, respectively. When the corresponding restriction lanes are compared, in many cases the bands that hybridized to the family A probe also hybridized to the family B probe. Since the family A and B probes did not cross-react, we suspect that the family A and B genes are localized close to one another, possibly forming pairs similar to the gene pairs B-1/A-2 and B-2/A-3. Furthermore, the fact that three intense bands were observed between 10 and 30 kb in the lanes containing *Bam*HI-digested DNA suggests that most of the bombyxin gene copies are clustered in three regions.

DISCUSSION

The presence of multiple gene copies that encode bombyxin in the *Bombyx* genome has been suggested (8) by genomic Southern hybridization using a bombyxin cDNA as probe. The present study has verified this observation by demonstrating that a *Bombyx* DNA segment carries four bombyxin genes. The results of the Northern hybridization experiments and the structural features of all genes characterized in this study suggest that it is probable that all four genes are expressed. The deduced amino acid sequences of bombyxins differ somewhat from one another. Thus, we conclude that the bombyxin peptide polymorphism (1, 2, 6) results from heterogeneity at the genetic level (i.e., multiple copies in the *Bombyx* genome). Multiple bombyxin gene copies in the *Bombyx* genome and the lack of introns are in sharp contrast to vertebrate insulin-family genes that exist in one or two copies per haploid genome and have two or three introns (20). These differences suggest that evolutionary mechanisms for insulin-family genes in invertebrates and vertebrates differ significantly.

Remarkably, bombyxin genes are arranged in pairs in which the component genes belong to different families and are oriented in opposite directions. Genomic Southern hybridization experiments using genes A-1 and B-1 as probes have indicated that such gene pairs seem to occur frequently in the *Bombyx* genome. A similar gene-pair organization has been extensively studied for the moth chorion protein superfamily (21, 22). The chorion genes in each pair are coordinately expressed and the expression of various gene pairs is regulated in a developmental-stage-specific manner (23, 24). Coordinate expression of similar gene pairs has also been suggested for *Drosophila* genes that encode the salivary gland glue protein (25) and cuticular protein (26). Thus, it is probable that a bombyxin gene pair forms a unit for coordinate expression and various gene pairs are expressed in a temporally specific manner during the course of development.

Although bombyxin has a clear-cut prothoracicotropic function when applied to *Samia* (3), its function in *Bombyx* is still unknown. Interestingly, bombyxin has been detected in the *Bombyx* ovary and in embryos throughout embryonic development (27), in contrast to the *Bombyx* prothoracicotropic hormone, which is detectable only at the later stages of embryogenesis (28). It seems, therefore, likely that bombyxin might be involved in the control of oogenesis and embryogenesis, possibly through the regulation of cell proliferation and differentiation. Insulin and insulin-like growth factors control vertebrate cell proliferation and differentiation in a variety of tissues and developmental stages, including oocyte maturation (29) and embryogenesis (30). Detailed studies on the differential expression of bombyxin gene pairs might provide clues to the function of bombyxin.

We thank T. Takeda and M. Yamamoto (The Institute of Medical Science, The University of Tokyo) for preparing the oligonucleotide probes and primers. We also thank K. Soma and I. Kubo for technical assistance. This work was supported by grants from the Ministry of Education, Science and Culture of Japan, from The Ishida Foundation, and from The Mitsubishi Foundation.

1. Nagasawa, H., Kataoka, H., Isogai, A., Tamura, S., Suzuki, A., Mizoguchi, A., Fujiwara, Y., Suzuki, A., Takahashi, S. Y. & Ishizaki, H. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 5840–5843.
2. Ishizaki, H. & Suzuki, A. (1988) *Horm. Metab. Res.* **20**, 426–429.
3. Ishizaki, H., Mizoguchi, A., Fujishita, M., O'oka, H., Kataoka, H., Isogai, A., Nagasawa, H., Tamura, S. & Suzuki, A. (1983) *Dev. Growth Differ.* **25**, 593–600.
4. Nagasawa, H., Kataoka, H., Isogai, A., Tamura, S., Suzuki, A., Ishizaki, H., Mizuguchi, A., Fujiwara, Y. & Suzuki, A. (1984) *Science* **226**, 1344–1345.
5. Jhoti, H., McLeod, A. L., Blundell, T. L., Ishizaki, H., Nagasawa, H. & Suzuki, A. (1987) *FEBS Lett.* **219**, 419–425.
6. Maruyama, K., Hietter, H., Nagasawa, H., Isogai, A., Tamura, S., Suzuki, A. & Ishizaki, H. (1988) *Agric. Biol. Chem.* **52**, 3035–3041.
7. Nagasawa, H., Maruyama, K., Sato, B., Hietter, H., Kataoka, H., Isogai, A., Tamura, S., Ishizaki, H., Senba, T. & Suzuki, A. (1988) in *Peptide Chemistry 1987*, eds. Senba, T. & Sakakibara, S. (Protein Res. Found., Osaka, Japan), pp. 123–126.
8. Adachi, T., Takiya, S., Suzuki, Y., Iwami, M., Kawakami, A., Takahashi, S. Y., Ishizaki, H., Nagasawa, H. & Suzuki, A. (1989) *J. Biol. Chem.* **264**, 7681–7685.
9. Iwami, M., Kawakami, A., Ishizaki, H., Takahashi, S. Y., Adachi, T., Suzuki, Y., Nagasawa, H. & Suzuki, A. (1989) *Dev. Growth Differ.* **31**, 31–37.
10. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
11. Henikoff, S. (1984) *Gene* **28**, 351–359.
12. Calzone, F. J., Britten, R. J. & Davidson, E. H. (1987) *Methods Enzymol.* **152**, 611–632.
13. Chomczynski, P. & Sacchi, N. (1987) *Anal. Biochem.* **162**, 156–159.
14. Lehrach, H., Diamond, D., Wozney, J. M. & Boedtker, H. (1977) *Biochemistry* **16**, 4743–4751.
15. Weeks, D. P., Beerman, N. & Griffith, O. M. (1986) *Anal. Biochem.* **152**, 376–386.
16. Reed, K. C. & Mann, D. A. (1985) *Nucleic Acids Res.* **13**, 7207–7221.
17. Rubin, G. M. (1983) in *Mobile Genetic Elements*, ed. Shapiro, J. A. (Academic, New York), pp. 329–361, 1431–1440.
18. Bell, G. I., Swain, W. F., Pictet, R., Cordell, B., Goodman, H. M. & Rutter, W. J. (1979) *Nature (London)* **282**, 525–527.
19. Smit, A. B., Vreugdenhil, E., Ebberink, R. H. M., Geraerts, W. P. M., Klootwijk, J. & Joosse, J. (1988) *Nature (London)* **331**, 535–538.
20. Steiner, D. F., Chen, S. J., Welsh, J. M. & Kwok, S. C. M. (1985) *Annu. Rev. Genet.* **19**, 463–484.
21. Goldsmith, M. R. & Kafatos, F. C. (1984) *Annu. Rev. Genet.* **18**, 443–487.
22. Kafatos, F. C., Mitsialis, S. A., Nguyen, H. T., Spoerel, N., Tsitilou, S. G. & Mazur, G. D. (1987) in *Development As an Evolutionary Process*, eds. Raff, R. A. & Raff, E. C. (Liss, New York), pp. 161–178.
23. Spoerel, N., Nguyen, H. T. & Kafatos, F. C. (1986) *J. Mol. Biol.* **190**, 23–35.
24. Jones, C. W. & Kafatos, F. C. (1980) *Cell* **22**, 855–867.
25. Mayerowitz, E. M. & Hogness, D. S. (1982) *Cell* **28**, 165–176.
26. Snyder, M., Hunkapiller, M., Yuen, D., Silvert, D., Fristrom, J. & Davidson, N. (1982) *Cell* **29**, 1027–1040.
27. Fugo, H., Hua, J., Nakajima, M., Nagasawa, H. & Suzuki, A. (1987) *J. Insect Physiol.* **33**, 243–248.
28. Chen, J. H., Fugo, H., Nakajima, M., Nagasawa, H. & Suzuki, A. (1987) *J. Insect Physiol.* **33**, 407–411.
29. El-Etr, M., Schorderet-Slatkine, S. & Baulieu, E.-E. (1979) *Science* **205**, 1397–1399.
30. Froesch, E. R., Schmid, Chr., Schwander, J. & Zapf, J. (1985) *Annu. Rev. Physiol.* **47**, 443–467.