Structure and Expression of Gene Coding for Sex-specific Storage Protein of *Bombyx mori*

Hiroshi Sakurai, Tomoko Fujii, Susumu Izumi, and Shiro Tomino

From the Department of Biology, Tokyo Metropolitan University, Setagaya-ku, Tokyo 158, Japan

A major plasma protein termed “SP 1” accumulates in a sex- and stage-specific manner in the larval hemolymph of the silkworm, *Bombyx mori*. We have cloned the genomic sequence coding for SP 1 and analyzed its primary structure. The SP 1 mRNA sequence is encoded by five exons interspersed with four introns. Initiation site for the SP 1 gene transcription was identified at nucleotide level. Sequence homologous to the SV 40 enhancer “core” structure exists in two adjacent locations in the first intron. The 5’ flanking region of the SP 1 gene contains a sequence highly homologous to the putative hormone-receptor complex binding site found in the edestoid-sensitive genes of *Drosophila melanogaster*. Developmental change in the level of the SP 1 mRNA precursor in the fat body faithfully reflects that of SP 1 mRNA, indicating that the biosynthesis of SP 1 in *B. mori* is regulated in a sex- and stage-specific fashion at the level of transcription.

In contrast to those in vertebrate sera, major protein components in insect hemolymph are not only limited in number, but they undergo qualitative as well as quantitative changes during the postembryonic development (1–3). In holometabolous insects, specific proteins termed as “storage proteins” comprise major protein components of the larval hemolymph (2–4). These proteins are synthesized in large quantities by the fat body of actively feeding larvae and released into hemolymph. At the conclusion of the feeding period, however, they are selectively taken up by the fat body cells and stored there in the form of protein granules that are required for the development of adult tissues (3–5).

In the silkworm, *Bombyx mori*, storage proteins are known to occur in two forms, referred to as SP 1 and SP 2, respectively (6). Both proteins are of molecular weights of approximately 500,000, and their native molecules are each composed of six identical subunits with molecular weights around 50,000 (6). SP 1 is characterized by an exceptionally high content of methionine, while the amino acid composition of SP 2 is analogous to the diterpoid storage proteins, being rich in phenylalanine and tyrosine (6–8). Recently, the structural gene coding for SP 1 has been mapped at a site proximal to the tub gene on the 23rd chromosome, whereas the SP 2 gene locates on the 3rd chromosome (9).

The *B. mori* SP 1 provides a unique opportunity for studying mechanisms bringing about the expression of secondary sexual characters in insects, since this protein exhibits the stage-specific sexual dimorphism in hemolymph. Hemolymphs of both sexes of *B. mori* contain nearly equal amounts of SP 1 until the end of the fourth larval instar. In the fifth (the final) instar larvae, however, the amount of SP 1 greatly increases in females, while markedly declines in males (6, 10). Our previous study employing the sex-mosaic individuals of *B. mori* provided evidence that the sex-dependent synthesis of SP 1 is primarily determined by the sex chromosome composition in each fat body cell and is developmentally regulated without participation of any sex-specific humoral factors functioning like sex hormones in vertebrates (10).

The present experiments have been undertaken to investigate further mechanisms underlying the sex-dependent expression of SP 1 in the *B. mori* silkworm and to elucidate molecular processes involved in the expression of secondary sexual characters in insects. For this purpose, we cloned the genomic sequence coding for SP 1 and analyzed its structure. Evidence is also presented that the expression of the SP 1 gene is regulated in a sex- and stage-specific fashion at the level of transcription in the fat body.

**MATERIALS AND METHODS**

**RESULTS AND DISCUSSION**

**Cloning of SP 1 Gene Sequence**—To facilitate cloning of the SP 1 gene, an attempt was made to construct a *B. mori* mini gene library from the DNA fragments enriched for the SP 1 gene sequence. The EcoRI digest of the fat body DNA was fractionated by preparative gel electrophoresis as detailed under “Materials and Methods,” in the Miniprint, by which the SP 1 genomic DNA was enriched by about 20-fold above total genomic DNA. The DNA fragments enriched for the SP 1 gene sequence were ligated with Charon 4A arm DNAs, and the resulting recombinant DNA was introduced into λ phage particles (18). The library was screened for the SP 1 mRNA sequence by plaque hybridization, and four plaques which gave positive hybridization signal were isolated.

The genomic DNA inserts were subjected to restriction mapping analysis as shown in Fig. 1. One of these clones contained the 11-kb EcoRI fragment, and three others carried the 9-kb fragments which are identical with respect to the

**The “Materials and Methods” are presented in miniprint at the end of this paper. Miniprint is easily read with the aid of a standard magnifying glass. Full size photocopies are included in the microfilm edition of the Journal that is available from Waverly Press.**

*To whom correspondence should be addressed: Dept. of Biology, Tokyo Metropolitan University, 2-1-1 Fukasawa, Setagaya-ku, Tokyo 158, Japan.*

*This work was supported in part by Grant-in-Aid for Scientific Research (Category B, Project Number 6048021) from the Ministry of Education, Science and Culture, Japan. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.*

*The abbreviations used are: kb, kilobase pair(s); bp, base pair(s); PIPES, pipеразин-N,N'-бис(2-етанесульфонновая кислота).*
FIG. 1. Restriction maps of 9- and 11-kb SP 1 genes. Open bars represent genomic inserts, and shadowed bars indicate SP 1 cDNA. The location of the pBMSP1C1 sequence was determined by Southern hybridization. Position on exons of the pBMSP1C2 cDNA sequence was verified by comparing the nucleotide sequence of cDNA with that of the genomic DNA (23). Polarity of the cloned SP 1 gene was assigned as indicated in the figure. Small arrows indicate directions of sequencing and sizes of the sequenced fragments. Symbols for restriction enzymes are: A, ScaI; B, BglII; D, HindIII; E, EcoRV; F, XbaI; H, HincII; L, Sall; N, NsiI; P, PstI; R, EcoRI; S, SacI; and X, XhoI.

physical maps. Southern hybridization with the SP 1 cDNA proved location of the cDNA sequence on the genomic clone. Polarity of the SP 1 gene was also assigned as indicated in Fig. 1 by comparison with cDNA of locations of restriction sites.

Comparison of physical maps between the 9- and the 11-kb DNA inserts indicated that major portion of two DNA inserts are structurally homologous, except for the 3' flanking segments of the mRNA-coding region. Moreover, partial nucleotide sequence of the 11-kb DNA completely matched with that of the 9-kb DNA (data not shown). This could mean that two DNA fragments might represent duplicate genes coding for identical SP 1 as observed with the genes for dipterous storage proteins (4, 20, 21). Alternatively, they might be in allelic relation, since these clones were isolated from a library which had been constructed from DNA of heterozygous individuals (Tokai X Asahi). Analogous phenomenon has been observed with the sericin genes of B. mori (22).

Exon/Intron Composition of SP 1 Gene—Structure of the 9-kb DNA insert was analyzed in detail. The size and number of exons in SP 1 gene were confirmed by the method of RNase mapping (17) as illustrated in Fig. 2. The SP 1 genomic DNA was digested with BgII and each of the 4- and the 5-kb fragments was inserted into multicloning sites of pSP64 and pSP65. The recombinant plasmids were transcribed by the SP6 RNA polymerase, and RNAs complementary to the SP 1 mRNA (cRNA) were synthesized. Hybridization of cRNAs with poly(A)+ RNA from the male fifth instar larvae (Fig. 2, lane 3) failed to produce any detectable signal on the autoradiogram, since, as described previously (11), the male fifth instar larvae are virtually free of the SP 1 mRNA. When poly(A)+ RNA from the fourth instar larvae was subjected to hybridization analysis (Fig. 2, lanes 1 and 2), electrophoretic patterns of the hybridized fragments were essentially identical with those observed with poly(A)+ RNA from the female fifth instar larvae (Fig. 2, lane 4).

The transcript of the 4-kb DNA, which carries upstream half of the 9-kb SP 1 gene, yielded three protected fragments with 123, 134, and 269 bases in size, respectively (Fig. 2A). The cRNA corresponding to downstream region of the 9-kb DNA protected three bands with 161, 565, and 1077 nucleotides, respectively (Fig. 2B). Since it has been known from the results of preliminary S1 mapping experiments that the BgII restriction site is lying in the 834-bp exon, the 269 and the 565 base fragments seen on radioautograms are most likely to be derived from the 834-bp exon. Recently, an additional SP 1 cDNA clone (pBMSP1C2) carrying the 161- and the 1077-bp exons at 3' proximal region of the SP 1 gene was isolated (Ref. 23, Fig. 1). The results clearly indicate that two exons each with 123 and 134 bp in size exist upstream of the 834-bp third exon and that the 161- and the 1077-bp fragments are the fourth and the fifth exons, respectively.

Accurate locations of exons in the SP 1 gene were assigned by comparing its nucleotide sequence with that of the SP 1 mRNA. The nucleotide sequence of the SP 1 mRNA was determined by primer extension technique in the presence of dideoxy- and deoxynucleotide triphosphates to yield its cDNA sequence. The primer DNA lying in the third exon was hybridized with total RNA from the female fifth instar larvae, and cDNA was synthesized. The SP 1 mRNA sequence homologous to the genomic sequence occurs at about 630 bp upstream from the 5' splicing site of the third exon. Sequence analysis was likewise repeated using primer DNA derived from the second exon, and it was concluded that the first exon is followed by a 1074-bp intron. Sizes of the first and the second exons are determined to be 123 and 134 bp, respec-
tively, from the results of RNase mapping and direct sequencing of mRNA.

The mRNA-coding sequence of SP 1 gene stretches for a 4.8-kb region on the 9-kb DNA fragment and is separated into five exons interspersed by four introns (Fig. 3A). From the estimated lengths of exons total length of SP 1 mRNA is calculated to be about 2300 nucleotides. The value is well in agreement with that estimated by Northern hybridization of the fat body RNA (11) as well as that calculated from the molecular weight of the SP 1 subunit (6). The splice junctions of the SP 1 gene are summarized in Fig. 3B. Nucleotide sequences at the 5' and the 3' ends of introns are homologous to those reported for other genes in B. mori (22, 24, 25) and the consensus sequence for general splice junction (26).

**Transcription Initiation Site of SP 1 mRNA**—The transcription initiation site of SP 1 mRNA was deduced by direct sequencing of mRNA. The 73-base primer complementary to the sequence in the first exon was hybridized with the fat body RNA, and the primer was elongated along with the mRNA sequence. As depicted in Fig. 4B, the primer DNA was elongated by some 40 nucleotides and terminated in sequence ladders, indicating that the sequence around the cap site of the SP 1 mRNA is AGUGUG and that adenine is the putative transcription initiation residue. The result was further confirmed by S1 nuclease mapping. The DNA fragment labeled at the Sty1 site was cleaved at AluI site within the 5' flanking region of the SP 1 gene, and the resultant 186 base probe DNA was hybridized with RNA. After treatment with varying concentrations of S1 nuclease, protected fragments were electrophoresed in parallel with the mRNA sequence ladders. Among a cluster of fragments seen in radioautogram, the most prominent was one with the terminal "A" residue (Fig. 4A).

Presence of multiple protected fragments probably represents nondiscrete digestion by S1 nuclease at the protected end rather than multiple transcription initiation sites on the SP 1 mRNA. Taken together with these results, the initiation site for the SP 1 gene transcription was identified as adenine at position +1 shown in Fig. 5. This in agreement with the report that the cap sites of most eukaryotic mRNAs are adenine (27). The sequence TTCAGTG (*underline indicates initiation site) in the SP 1 mRNA is highly homologous to the consensus sequence ATCAGTY ("Y" represents pyrimidine) at the cap site of insect mRNAs (28).

**Nucleotide Sequence around Transcription Initiation Site**—TATA box is the most established cis-acting element of genes for accurate transcription (29, 30). In the SP 1 gene sequence TATATATA (TATA box-like structure) is identified at position -31 (Fig. 5). Some eukaryotic class II genes carry CCAAT sequence (CAT box) which may enhance transcription activity at 80–100 nucleotides upstream from the cap site (31). The SP 1 gene, however, lacks such a sequence. The nucleotide sequence homologous to the SV 40 enhancer "core" sequence (32) is detected in two locations in the first intron of the SP 1 gene (Fig. 5).

We have suggested the possibility that cellular factor(s) together with balance in the hemolymph concentration of ecdysone and juvenile hormone participates in the sexually dimorphic expression of SP 1 in B. mori (10). In this connec-

---

**Fig. 3.** Exon/intron composition of SP 1 gene. A, structure of the SP 1 gene. Open bar indicates the 9-kb SP 1 gene. Shadowed bars and bent lines represent exons and introns, respectively. Exon/intron composition of the SP 1 gene was determined as described in the text. Numbers indicate lengths of exons. Symbols for restriction enzymes are same as those indicated in Fig. 1. B, exon/intron boundaries of the SP 1 gene and other class II genes. Boundary sequences in genes for fibroin (24), sericin (22), and chorion proteins (25) are cited from the literature in parentheses, respectively. Capital letters represent exon sequences, and lower case letters represent intervening sequences (IJS). In the consensus sequence, n and y represent any nucleotide and a pyrimidine, respectively.

**Fig. 4.** Determination of transcription initiation site of SP 1 mRNA. A, a 186-nucleotide-long AluI/StyI (position -74 to +113) fragment was labeled with 32P and made single-stranded. This DNA was hybridized with total RNA from the female fifth instar larvae and treated with S1 nuclease. The concentrations of S1 nuclease used were 25 (lane 1), 100 (lane 2), and 400 units/ml (lane 3), respectively. The protected fragments were electrophoresed on the 8% acrylamide, 7 M urea sequence gel, and radioactivity was detected by autoradiography. The arrow to the left indicates the transcription initiation site. B, a labeled 73-nucleotide RsaI/StyI (position +40 to +113) single-stranded DNA was hybridized with the same RNA as in A and extended by reverse transcriptase in the presence of dideoxy- and deoxynucleotide triphosphates as described under "Materials and Methods." The extended cDNAs were analyzed by gel electrophoresis as above. cDNA sequence and corresponding mRNA sequence are indicated to the right of figure.
the nascent SP 1 of the signal sequence for secretion is clearly demonstrated by comparison of the deduced amino acid sequence with the amino-terminal primary structure of SP 1. Hence, putative signal peptide cleavage site is assigned between Ala15 and Ser16 as indicated in Fig. 5. In the deduced primary structure of SP 1 mRNA sequence together with the amino-terminal primary structure of SP 1.

**Sex and Stage Dependence of SP 1 mRNA Precursor—** During the postembryonic development of *B. mori*, SP 1 accumulates in hemolymph in a sex- and stage-specific fashion (6, 10). Our previous study demonstrated that the sexual dimorphism in the SP 1 expression is regulated at the level of mRNA (11). To confirm the specific step at which the cellular level of SP 1 mRNA is regulated, the amount of precursor RNA for SP 1 mRNA in the fat body was measured by the method of intron-labeled S1 mapping.

As is evident from Fig. 6, the SP 1 precursor RNA was detectable in the fat body RNA of the male as well as the female fourth instar larva. In the fifth instar larvae, the amount of the SP 1 precursor RNA markedly increases in both sexes, while it is scarcely detectable in males. The precursor RNA reappears in the fat body of both sexes at the early pupal stage. The result of S1 nuclease mapping is consistent with that of RNA blot analysis (11), indicating that the male-specific expression of SP 1 is regulated at the intron level.
homologous to the cis-acting regulatory elements found in other class II genes are also recognized in the region upstream of the SP 1 structural gene. Although the physiological significance of these structures in the SP 1 gene remains obscure, regulatory mechanisms bringing about the sex- and stage-specific transcription of SP 1 gene can be explained, at least in part, by analyzing interactions between the regulatory region of SP 1 gene and cellular factors.

Acknowledgments—We thank Drs. Hiroshi Kataoka and Susumu Maeda, Zoecom Research Institute, Palo Alto, CA, for sequencing the amino-terminal primary structure of SP 1. We also thank Dr. Tohio Fukazawa, Keio University Medical School, for critically reading the manuscript.

REFERENCES
1. Thomson, J. A. (1975) Adv. Insect Physiol. 11, 321-398
2. Wyatt, W. C., and Jones, M. L. (1975) Annu. Rev. Biochem. 44, 779-817
3. Levensklokk, L. (1966) in Comprehensive Insect Physiology, Biochemistry, and Pharmacology, ed. (Kerkut, G. A. and Gilbert, I. L., ed), Vol. 10, pp. 197-346, Pergamon Press, Oxford
4. Berber, T., and Kornberg, W. H. (1981) Experientia 37, 102-110
5. Telfer, W. H., Keim, P. S., and Law, J. H. (1978) Insect Biochem. 13, 601-613
6. Tan, S., Nagata, M., and Kobayashi, M. (1980) Insect Biochem. 10, 289-303
7. Munn, E. A., Feinstein, A., and Greville, G. D. (1967) Biochem. J. 102, 641-649
8. Munn, E. A., and Greville, G. D. (1969) J. Insect Physiol. 15, 1939-1950
9. Shumada, T., Kobayashi, M., and Yoshikawa, N. (1983) J. Biochem. Soc. Jpn. 54, 464-469
10. Mina, S., Isumi, S., Katsuki, M., and Temino, S. (1983) Dev. Biol. 97, 329-367
11. Isumi, S., Sakurai, H., Fujii, T., Ikeda, W., and Tomino, S. (1988) Biochim. Biophys. Acta 948, 181-188
12. Ohshima, Y., and Suzuki, Y. (1977) Proc. Natl. Acad. Sci. U. S. A. 74, 4689-4692
13. Izumo, S., Tojo, S., and Tomino, S. (1980) Insect Biochem. 10, 429-434
14. Azuma, Y., and Leader, P. (1972) Proc. Natl. Acad. Sci. U. S. A. 69, 1408-1412
15. O'Farrell, P. R., Craven, L., and Byas, C. (1974) Biochemistry 13, 247-256
16. Nagata, M., Fujii, T., and Tomino, S. (1980) J. Virol. 37, 346-356
17. Becker, A., and Gold, M. (1975) Proc. Natl. Acad. Sci. U. S. A. 72, 581-585
18. Sangar, F., Nicklen, S., and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. U. S. A. 74, 5462-5467
19. Tahara, T., Kuroiwa, A., Obinata, M., and Natori, S. (1984) J. Mol. Biol. 174, 19-29
20. Schenk, H., Keijara-Lepeast, J., Bereu, F., Moreau, J., Scheller, K., Bregere, F., and Lepeast, J.-A. (1985) EMBO J. 4, 2983-2990
21. Okae, N., Ishihara, E., and Suzuki, Y. (1982) J. Biol. Chem. 257, 1519-1525
22. Fujii, T., Sakurai, H., Isumi, S., and Tomino, S. (1986) Zoos. Sci. 3, 1023
23. Tsujimoto, Y., and Suzuki, Y. (1979) Cell 18, 591-600
24. Edres, W. D., and Eckstein, F. (1980) Nucleic Acids Res. 8, 1843-1856
25. Mount, S. M. (1982) Nucleic Acids Res. 10, 469-472
26. Bussling, M., Portmann, R., Irvine, J. C., and Birmatel, M. L. (1980) Nucleic Acids Res. 8, 957-966
27. Hultmark, D., Klemz, B., and Gahrng, W. J. (1986) Cell 44, 429-438
28. Brehnath, R., and Chambon, P. (1981) Annu. Rev. Biochem. 50, 349-392
29. Gannon, F., O'Hare, K., Perchin, D., Le Panne, J. P., Benoist, C., Cochet, M., Brehnath, R., Kornberg, A., Gampich, A., Cunl., and Chambon, P. (1979) Nature 284, 432-434
30. Benoist, C., O'Hare, K., Brehnath, R., and Chambon, P. (1980) Nucleic Acids Res. 8, 127-142
31. Khouyr, G., and Plass, F. (1983) Cell 33, 313-314
32. Mayer, R., Schirle, M., Amin, J., Heppner, A., Haentmann, J., and Voellmy, R. (1986) EMBO J. 5, 1667-1673
33. Harghuchile, O., Santer, M., Steets, A., and Mies, R. J. (1978) Cell 13, 679-671
34. Kozak, M. (1978) Cell 15, 1109-1123
35. Purhons, D., and Halvorson, H. O. (1983) J. Mol. Biol. 167, 391-409

Supplemental Material to
Structure and Expression of Gene Coding for Sex-specific Storage Protein of Bombyx mori
by
Hisajuki Sakurai, Tomoko Fujii, Sumaru Ishii, and Shito Tomino

MATERIALS AND METHODS
Nuclei—Larvae of a commercial strain of the silkworm, Bombyx mori (Takifugu) were obtained from Nippon Shuppan Co. Ltd. (Toyama) and reared on silkworm leaves or an artificial diet (Kakaido Shirkou Shoyu Ltd., Tokyo 120-4002).

Preparation of Poly(A)+ RNA 
A RNA sample from a mid-larval instar was prepared by the method reported by Melton et al. (1984). The RNA sample was hybridized with the 32P-labeled recombinant phages as described below.

Hydroxyapatite—Preparation of Poly(A)+ RNA
The isolates were obtained from fat body adults dissected from mature 1-day-old males. After hybridization, the protected RNAs were identified by autoradiography.

RNA Sequencing—The samples were sequenced by denaturing polyacrylamide gel electrophoresis (PAGE) and autoradiography as described above.

Biochemical Analysis—The samples were sequenced by denaturing PAGE and labeled with 32P-labeled recombinant phages as described above.

Biochemical Analysis—The samples were sequenced by polyacrylamide gel electrophoresis, autoradiography, and sequencing as described above.

Biochemical Analysis—The samples were sequenced by polyacrylamide gel electrophoresis, autoradiography, and sequencing as described above.

Biochemical Analysis—The samples were sequenced by polyacrylamide gel electrophoresis, autoradiography, and sequencing as described above.