FEBS Letters 356 (1994) 207–211

FEBS 14878

Vitellogenin gene of the silkworm, *Bombyx mori*: structure and sex-dependent expression

Ken-ichi Yano, Mayumi Toriyama Sakurai**, Susumu Izumi, Shiro Tomino*

[a]*Department of Biology, Tokyo Metropolitan University, Minamiohsawa 1-1, Hachioji-shi, Tokyo 192-03, Japan*

Received 13 October 1994; accepted 4 November 1994

Abstract Vitellogenin of *Bombyx mori* is a precursor of major yolk protein synthesized in the female fat body at larval–pupal ecdysis. The gene for *B. mori* vitellogenin is composed of seven exons interspersed by six introns. Developmental profile of the primary transcript of the gene indicated that the biosynthesis of *B. mori* vitellogenin is regulated transcriptionally in a sex- and stage-dependent manner in the fat body. The Arg-X-Arg-Arg sequence, which conforms to the recognition site of mammalian furin, occurs in a region just upstream of the putative proteolytic cleavage site of *B. mori* previtellogenin.

Key words: Vitellogenin; Yolk protein; Genomic sequence; Sex-dependent expression; Transcriptional regulation; *Bombyx mori*

1. Introduction

Vitellogenins, the precursors of major yolk proteins of oviparous animals, are synthesized extraovarially and sequestered by the growing oocytes [1]. Insect vitellogenins, with the exception of yolk proteins of higher Dipterans, are multimeric glycolipoproteins with high molecular weights that are synthesized by the fat bodies of sexually matured females [1,2]. While information has accumulated on the biochemical characteristics and developmental profiles of vitellogenins from a variety of insect species, our knowledge on their gene structures has been rudimentary; the mRNA and genomic sequences have been elucidated for the vitellogenins of very limited insect species that include *Anthonomus grandis* [3], *Bombyx mori* [4] and *Aedes aegypti* [5].

Vitellogenin of the silkworm, *Bombyx mori*, is a tetramer with molecular weight of 440k, composed of each two molecules of non-identical subunits termed 'heavy chain' and 'light chain', respectively [6]. In *B. mori*, vitellogenin is synthesized by the fat body cells of females at about larval–pupal ecdysis and released into hemolymph before being taken up by the developing oocytes [4,7]. Previously, we have cloned and characterized the mRNA sequence for the *B. mori* vitellogenin [4]. By comparing the deduced amino acid sequence with the amino-terminal primary structures of vitellogenin subunits, we demonstrated that two subunits are derived from the proteolytic processing of a precursor protein coded for by a single contiguous mRNA.

In the present study we have cloned the genomic sequence for vitellogenin and analyzed its structure. This paper describes the structure and organization of vitellogenin gene of *B. mori*. S1 nuclease mapping analysis of the primary transcript during larval–pupal development revealed that the vitellogenin synthesis in *B. mori* is regulated in a sex- and stage-dependent manner at the level of gene transcription in the fat body cells. By comparing the deduced primary structures of insect vitellogenins, possible mechanism of proteolytic processing of vitellogenin precursors in insects is also discussed.

2. Materials and methods

2.1. Cloning and sequence analysis of the vitellogenin gene

An EMBL3 genomic library of *Bombyx mori* (Tokai × Asahi) was provided by Dr. Shigetoshi Mori [8]. The library (1.5 × 105 pfu) was screened by the method of plaque hybridization with the ^{32}P-labeled vitellogenin cDNA probe [4], and 3 overlapping clones were isolated. Insert DNA of one of the recombinant phage (λGVL1) was subcloned into pUC18 plasmid vector. Nucleotide sequences of the subclones were determined by the dideoxy chain termination method using a 7-deaza sequencing kit supplied by Takara Shuzo Co. The nucleotide and deduced amino acid sequences were determined by use of a DNASIS sequence analysis system, ver. 7.0 (Hitachi Software Engineering Co.)

2.2. Determination of transcription initiation site

Total RNA was prepared from the fat body by the acid–guanidinium–phenol–chloroform method [9]. Poly(A)$^+$ RNA was isolated from total RNA by chromatography on oligo(dT) cellulose [10].

An oligonucleotide complementary to a region (+30 to +49) of vitellogenin mRNA was synthesized and 5′ end-labeled with ^{32}P. The probe DNA was hybridized with 25 μg of poly(A)$^+$ RNA from the female fat body at larval–pupal ecdysis in a solution containing 40 mM PIPES, pH 6.4, 1 mM EDTA, 0.4 M NaCl, 80% formamide at 37°C for 16 h. RNA sequence was determined by the primer extension method in the presence of dideoxynucleotides using rous associated virus 2 reverse transcriptase as described previously [11]. Products of the reverse transcription were separated on a 8% polyacrylamide gel containing 7 M urea and detected by autoradiography.

2.3. Detection of primary transcript of vitellogenin gene

The 5′ end of *Ssp*I site within the first intron of vitellogenin gene (+173) was labeled with ^{32}P. The labeled DNA was digested with *Hin*fI at position −99 to yield a DNA probe with 272 nucleotide long. This DNA probe was hybridized with 100 μg of total RNA prepared from fat body as described above. The DNA–RNA hybrids were digested with S1 nuclease as described previously [11]. Protected fragments were separated on a 4% polyacrylamide gel containing 7 M urea and detected by autoradiography.

3. Results and discussion

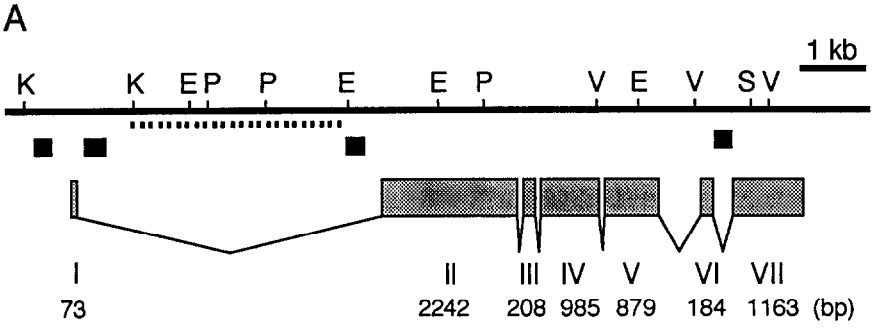
3.1. Isolation and characterization of the vitellogenin gene

Three genomic clones bearing the vitellogenin gene sequence were isolated from an EMBL3 library containing fat body DNA fragments of *B. mori* by screening the library with the

*Corresponding author. Fax: (81) (426) 77-2559.

***Present address:* Pharmaceutical Research Center, Meiji Seika Kaisha Ltd., Yokohama 222, Japan.

SSDI 0014-5793(94)01265-2



B

```
                                                    GGTACCACAAGCACCGGT  -721
CACTGTTCTCGTCGAACCCGTCGCTTGCGACGAAGGGCTCGACTTGCGAATTAACCCATA  -661
GACACAGCCCACTGAGTTTCTTGCCGGATCTTCTCAGTGGGTCGCGTTTCCGATCCGGTG  -601
GTAGATTCTGCGAAGCACTGCTCTTGCCAGGGCCAGTGTTAACAACACTCCGGTTTGAAC  -541
CCCGTGAGCTCACCTACACACGTTAGGGTGAAGCTGAAATGGCCTCTCAAGGCTATCAGC  -481
ATATGTAGGAATAAAAAGCAATTTTGTCCATATTCCTATAATTCGAAATTCCAGTTTTTT  -421
TTTTATTGCCTTTGTAGGCAGACGGGCATACGGCCCACCTGATGGTGAGTGGTTACCGTC  -361
GCCCATGGACTTCAGCAATGCCAGGGCAGAGCCATCCCGCTGCCTACGCGCTTAACACTC  -301
TCCACACGCCTCGTTTGAAGAAGGACATGTCATAGCGCTCGGTAAACACCGTGGAGGGGA  -241
GCTGCATTCCATAGCCGGATGGTACGCGATCTTTCGCGTTCTAATTTTTATTAAAAGCGA  -181
ATATTCACGCTTGATTTAACTCCTCGCCCCAGTTCCCTCAAATTTCAGTTTCCACAAAGA  -121
CAAAGGAATAAGAAAATGATTCGTTACATTGTATTTTATCTGCGTTTACTATCTAGGGTC   -61
ACAGTTCTGGGACACGTGTACCCTCCCTATATAAAGGGGGTGACCTGGCCAGAGATGTTC    -1
AGATCCTTCGGAGCTCAGCAGACTGACAGCGGAGCTAGTACAATGAAGCTATTCGTCTTG    60
                                                  M  K  L  F  V  L
GCGGCAATTATAGGTAAGATAATTCCTTATTACACAATCACTGCGTCTTCTTATTTCGAA   120
 A  A  I  I
TGTGCAACTGCAATTGCGTTTTTGATGTATGTTTATTTGTAATATGCAAGAATATTTCTT   180
TTTTGAACCACAATAAACAAATAAGCGTTCAATACAATATAAATATTTGTAATTGTAATC   240
GCTTCAACGGCAGTCGTGCACAGAAATAAATTATTTATTATAGTAATTGAGATATTATGT   300
ATAACACAAATGCATTATTTAAATAATTAACGATCTAACGCGATTTATTAATTTTTAAAA   360
TAATTTTCGAGGTAAAAAATCGAAACTCTTAGTACTGACTGATATGAAATCAATAACAAG   420
CGAAATTATATTTAAAATGCATGTAATGTAAATTTTGCGTGTAATGTAAATTAGATTTAT   480
TTTTTATTGCTTAGTTGGATGGACGAGCTCACAGCCCACCTGGTGTTAAATGGTTACTGG   540
AGCCCATAGACATCTGCAATGTAAATGCGCCACCCACCATGAGATATAAGTTCTAAGGTG   600
TCAGTATAGTTACAGCGGCTGCCCTACCCTTTAGACCGAAACGCATTACTGCTTCACGGC   660
AGAAATAGGCAGGGTGAATAGAATGCCTAATAGTTGTCTTGAGATAGTCTTGAGGGAACG   720
CAGGCTACCTACCTACCTGTTCTTTATATCGTCCGTCCCCGAAAAACTAACTCATTTAAA   780
```

Fig. 1. Structure of the vitellogenin gene. (A) Schematic diagram of the vitellogenin gene. Solid bar indicates restriction map of λGVL1. Shadowed boxes with Roman numerals and bent lines represent exons and introns, respectively. Arabic numerals under the shadowed boxes are the length of exons in base pairs. Closed boxes indicate the Bml repeats. Unsequenced region in the first intron is indicated by dotted line under the restriction map. Symbols for restriction enzymes are: E, *Eco*RI; K, *Kpn*I; P, *Pst*I; S, *Sal*I; V, *Eco*RV. (B) Nucleotide sequence around the transcription initiation site of the vitellogenin gene. Nucleotides are numbered from the putative transcription initiation site (+1). Boxed region represents first exon in which predicted amino acid sequence is also shown. TATA box is shown by double underline. Arrows represent inverted repeats. Dotted underlines indicate tandem repeat sequences. The sequence homologous to binding site of *D. melanogaster* double-sex protein is shown by wavy underline. Solid underlines indicate Bml repeats.

^{32}P-labeled vitellogenin cDNA probe. Since restriction maps of all the isolated clones overlapped each other, one of these clones (λGVL1) was subjected to sequence analysis. The nucleotide sequence of the genomic fragment was determined and compared with that of the vitellogenin cDNA. The sequences of the first exon and 5′ proximal region of the second exon were

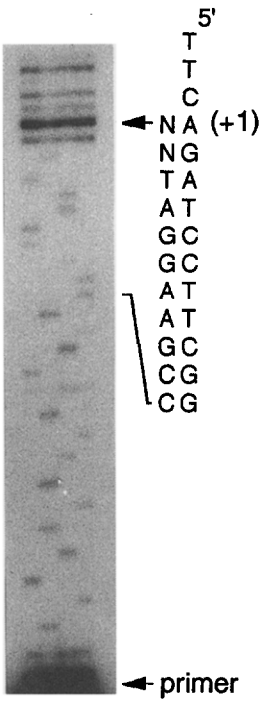


Fig. 2. Determination of transcription initiation site of the vitellogenin gene.Sequencing of the vitellogenin mRNA was carried out by primer extension method using the 5′ end-labeled oligonucleotide complementary to the +30 to +49 region of first exon. The cDNA sequence in parallel with the complementary genomic sequence is indicated to the right of the figure. Arrow at position +1 indicates cap site.

determined by S1 nuclease mapping and direct sequencing of vitellogenin mRNA.

The exon/intron organization of the *B. mori* vitellogenin gene is shown in Fig. 1A. Fig. 1B depicts partial nucleotide sequence of the gene covering the 5′ upstream region of transcription initiation site through a part of the first intron. (A 3.5 kb-region in the first intron was not seqenced as indicated in Fig. 1A.) The nucleotide sequence of the vitellogenin gene will appear in the DDBJ, GSDB, EMBL and NCBI databases under the accession numbers D30732 and D30733. The vitellogenin gene stretches over 12 kb region of chromosomal DNA and is composed of seven exons interspersed by six introns. The result of Southern blot hybridization analysis of genomic DNA prepared from both sexes of pupae suggested that the vitellogenin gene of *B. mori* is a single copy gene and is located on an autosome (data not shown). Nucleotide sequences homologous to Bm1 element, a highly reiterated retrotransposon-like sequence of *B. mori* genome [12], are present in the 5′ flanking region, first intron and sixth intron of the gene.

The initiation site of transcription was determined by direct sequencing of mRNA as shown in Fig. 2. Several DNA bands across all four sequencing lanes are seen at the end of the sequencing ladder. Since the adenine at position +1 gave the strongest signal, the adenine residue indicated by arrow in Fig. 2 was tentatively assigned as the transcription initiation site. Other signals appear to be the products resulting from 'snap-back' elongation. Nucleotide sequence around this tentative transcription inititation site is compatible with the consensus sequence predicted for the transcription initiation site of RNA polymerase II [13]. A typical TATA box occurs 33 bp upstream of the transcription initiation site. The nucleotide sequence homologous to the binding site of double sex protein [14] which regulates sex-dependent expression of yolk protein gene in *Drosophila melanogaster* [15] is detected at position −95. However, it is unknown at present whether homologs of double-sex gene exist in *B. mori*. A pair of inverted repeat structure with each 13 bp in length is detectable at positions −710 to −678. Furthermore, two 10 bp repeats in same orientation overlap the inverted repeat sequences.

3.2. Expression of vitellogenin gene

In the previous study, we reported that the vitellogenin mRNA is detectable in the fat body of females at about a period of larval–pupal ecdysis to early pupal stage [4]. To confirm the level at which the vitellogenin synthesis is regulated, the amount of primary transcript of the gene in the fat body was measured during larval to pupal development. As shown in Fig. 3, S1 nuclease protection analysis of the fat body RNA proved that the primary transcript of vitellogenin gene is detectable in the fat body of females on and after day-2 of spinning stage through pupal stage. The maximal level of the transcript was seen in females on the day of larval–pupal ecdysis, whereas no trace of the transcript was detected in males. The result confirms that sex- and stage-dependent synthesis of vitellogenin is regulated at a level of transcription in the fat body.

The molecular process involved in the female-specific expression of yolk protein gene has been well established in *D. melanogaster* [15,16]. We have previously demonstrated that the female-specific synthesis of storage protein 1 and vitellogenin in *B. mori* is primarily determined by the sex chromosome composition of the fat body cells and is regulated without participation of any sex-specific humoral factors [7]. However, little has been

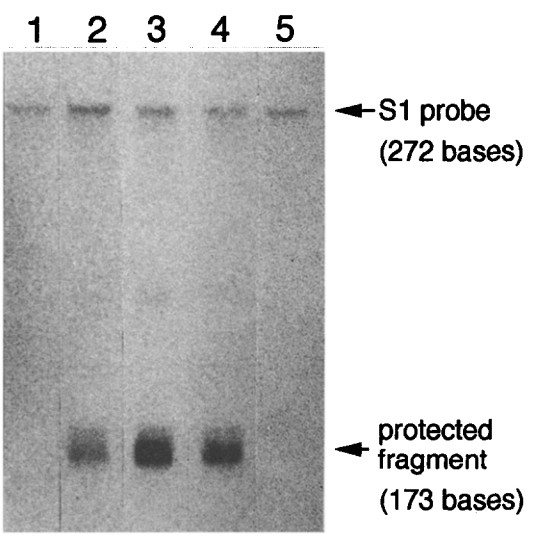
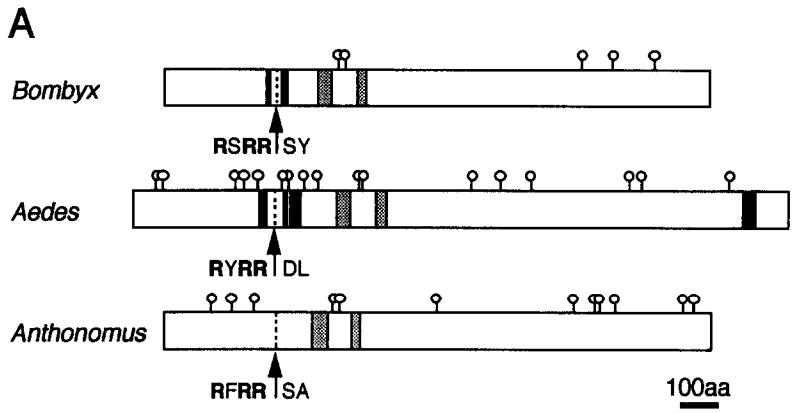


Fig. 3. S1 nuclease mapping analysis of the precursor of vitellogenin mRNA. The *Ssp*I site (position +173) within the first intron was end-labeled, and *Hin*fI/*Ssp*I fragment (position −99 to +173) was prepared. The labeled DNA was hybridized with each 100 μg of total fat body RNA prepared from females (lanes 1–4) or males (lane 5) at developmental stages as indicated below, and digested with S1 nuclease. Protected fragments were electrophoresed and visualized by autoradiography. Developmental stages are: lane 1 = day-5 fifth larval instar; lane 2 = day-2 spinning; lanes 3 and 5 = day of larval-pupal ecdysis; lane 4 = day-2 pupae.



B

```
                  *  **  *  ***  *   *   *  **    ***  *   *    *
1 Bombyx      SDMWMIFRDGVTQAGTLPAFKQIQSWIENKKIQEEEAAQVVVALPRTLR 543
2 Aedes       VDAWVTLRDAVAEAGTPSAFKLIFDFIKEKKLRGYEAATVIASLAQSIR 707
3 Anthonomus  RYTWSIFRDSVAEAGTGPALLNIKKWIETKKIQKTEAAQVIGTLAQSTR 528

   **      **       *  *  **                 *       *    *
1 YPTKQIMTQFFNFARSPAVKDQMFLNSSALMAATKLINLGQVNNYT-AHSYYPTHMYGRL- 602
2 YPTEHLLHEFFLLVTSDVVLHQEYLNATALFAYSNFVNQAHVSNRS-AYNYYPVFSFGRLA 767
3 FPTEEYMRKFFELATETQVRQQETLNQTCILSYTNLVHKVYI-NRNESHNQFPVHAFGSFY 588

                 *     *  *     *  *  *  *   ** **  ** ** ** **
1 THKHDAFVLEEILPTLAADL-KATVEYKDSTKAQVYIQAIGNLGHREILKVFAPYLEGKVE 662
2 DADYK-IIEHKIVPWFAHQLREAVNEG-DSVKIQVYIRSLGNLGHPQILSVFEPYLEGTIQ 826
3 TKKGREFVKTTVIPHLKQELEKAISNA-DNNKIHVMIRALGNIGHKSILNVFQPYFEGEKQ 648
```

Fig. 4. Structural comparison of previtellogenins in three insect species. (A) Schematic representation of previtellogenins of *Bombyx mori*, *Aedes aegypti* and *Anthonomus grandis*. Serine-clusters are represented by filled boxes. Dotted lines with arrows indicate proteolytic cleavage sites. Amino acid sequences around the putative cleavage sites are shown in capital letters. Shadowed boxes indicate the regions conserved among three species. Vertical lines with small circle represent possible asparagine-linked glycosylation sites. (B) Sequence alignment of the conserved regions. Amino acid residues conserved among three species are indicated by asterisks. Gaps introduced in the sequences to optimize the alignment are represented by dashes. Possible glycosylation sites are underlined.

known on the mechanism underlying sexual differentiation in insects with ZW-type sex chromosomes. The vitellogenin synthesizing system of *B. mori* will certainly provide an attractive model system for studying the problem.

3.3. Structural characteristics of insect vitellogenins

Combined exon sequences of the *B. mori* vitellogenin gene from the first ATG (+43) to termination codon in the seventh exon encodes a polypeptide with 1782 amino acid residues. As reported previously [4], no remarkable similarity in overall primary structure was noted between *B. mori* vitellogenin and vertebrate [17,18] or nematode [19] vitellogenins nor between three insect vitellogenins thus far sequenced [3–5]. However, highly conserved sequences were found in the primary structures of insect vitellogenins in two locations where over 40% of amino acids are conserved among three species (Fig. 4). Moreover, two possible asparagine-linked glycosylation sites are present in each vitellogenin molecule in a region between two conserved sequences (Fig. 4). Other glycosylation sites are detected in various locations. It is possible that glycosylation of these conserved regions might have some roles in the function of insect vitellogenins.

It has been known in several insect species that mature vitellogenins are composed of multisubunits which are derived from one or two precursor proteins by proteolytic processing [20–22]. In *B. mori*, two subunits of the vitellogenin are derived from a primary translation product [4]. The sequence R-S-R-R occurs just upstream of the putative cleavage site of *B. mori* previtellogenin. Analogous sequences (R-Y-R-R and R-F-R-R) are also present in the deduced previtellogenin sequences of *A. aegypti* [5] and *A. grandis*, respectively [23]. The R-X-R-R sequence has been predicted for the consensus sequence of recognition site by mammalian furin, a proprotein converting enzyme structurally related to bacterial subtilisin [24,25]. In addition, it has also been shown in *B. mori* that the prepeptides of prothoracicotropic hormone [26], bombyxin [27] and diapause hormone [28] are cleaved to yield mature peptides at sites comprised of pairs of basic amino acid residues, the structures which conform to the recognition site of mammalian subtilisin-related endoprotease [29]. In view of these observations, it is highly likely that proteolytic processing system analogous to those functioning in mammalian cells also exists in insects. Studies on the processing of previtellogenin are in progress in our laboratory. Preliminary experiments are suggesting the occurrence of furin-homologue in *B. mori* (Yano, unpublished observation). Understanding the molecular processes of vitellogenin maturation may lay foundation for elucidating general mechanisms of precursor-processing in insects.

References

[1] Hagedorn, H.H. and Kunkel, J.G. (1979) Annu. Rev. Entomol. 24, 475–505.
[2] Kunkel, J.G. and Nordin, J.H. (1985) in: Comprehensive Insect Physiology, Biochemistry and Pharmacology (Kerkut, G.A. and Gilbert, L.I. eds) vol. 1, pp. 83–111, Pergamon Press, Oxford.
[3] Trewitt, P.M., Heilmann, L.J., Degrugillier, S.S. and Kumaran, A.K. (1992) J. Mol. Evol. 34, 478–492.
[4] Yano, K.-i., Sakurai, M.T., Watabe, S., Izumi, S. and Tomino, S. (1994) Biochim. Biophys. Acta 1218, 1–10.
[5] Chen, J.-S., Cho, W.-L. and Raikhel, A.S. (1994) J. Mol. Biol. 237, 641–647.
[6] Izumi, S., Tomino, S. and Chino, H. (1980) Insect Biochem. 10, 199–208.
[7] Mine, E., Izumi, S., Katsuki, M. and Tomino, S. (1983) Dev. Biol. 97, 329–337.
[8] Mori, S., Izumi, S. and Tomino, S. (1991) J. Mol. Biol. 218, 7–12.
[9] Chomczynski, P. and Sacchi, N. (1987) Anal. Biochem. 162, 156–159.
[10] Aviv, H. and Leder, P. (1972) Proc. Natl. Acad. Sci. USA 69, 1408–1412.
[11] Sakurai, H., Fujii, T., Izumi, S. and Tomino, S. (1988) J. Biol. Chem. 263, 7876–7880.
[12] Adams, D.S., Eickbush, T.H., Herrera, R.J. and Lizardi, P.M. (1986) J. Mol. Biol. 187, 465–478.
[13] Bucher, P. (1990) J. Mol. Biol. 212, 563–578.
[14] Burtis, K.C., Coschigano, K.T., Baker, B.S. and Wensink, P.C. (1991) EMBO J. 10, 2577–2582.
[15] Coschigano, K.T. and Wensink, P.C. (1993) Genes Dev. 7, 42–54.
[16] Hodgkin, J. (1989) Cell 56, 905–906.
[17] van het Schip, F.D., Samallo, J., Broos, J., Ophuis, J., Mojet, M., Gruber, M. and Ab, G. (1987) J. Mol. Biol. 196, 245–260.
[18] Gerber-Huber, S., Nardelli, D., Haefliger, J.-A., Cooper, D.N., Givel, F., Germond, J.-E., Engel, J., Green, N.M. and Wahli, W. (1987) Nucleic Acids Res. 15, 4737–4760.
[19] Spieth, J., Denison, K., Zucker, E. and Blumenthal, T. (1985) Nucleic Acids Res. 13, 7129–7138.
[20] Chen, T.T., Strahlendorf, P.W. and Wyatt, G.R. (1978) J. Biol. Chem. 253, 5325–5331.
[21] Wojchowski, D.M., Parsons, P., Nordin, J.H. and Kunkel, J.G. (1986) Dev. Biol. 116, 422–430.
[22] Dhadialla, T.S. and Raikhel, A.S. (1990) J. Biol. Chem. 265, 9924–9933.
[23] Heilmann, L.J., Trewitt, P.M. and Kumaran, A.K. (1993) Arch. Insect Biochem. Physiol. 23, 125–134.
[24] Hosaka, M., Nagahama, M., Kim, W.-S, Watanabe, T., Hatsuzawa, K. Ikemizu, J., Murakami, K. and Nakayama, K. (1991) J. Biol. Chem. 266, 12127–12130.
[25] Barr, P.J. (1991) Cell 66, 1–3.
[26] Kawakami, A., Kataoka, H., Oka, T., Mizoguchi, A., Kimura, M.K., Adachi, T., Iwami, M., Nagasawa, H., Suzuki, A. and Ishizaki, H. (1990) Science 247, 1333–1335.
[27] Kawakami, A., Iwami, M., Nagasawa, H., Suzuki, A. and Ishizaki, H. (1989) Proc. Natl. Acad. Sci. USA 86, 6843–6847.
[28] Sato, Y., Oguchi, M., Menjo, N., Imai, K., Saito, H., Ikeda, M., Isobe, M. and Yamashita, O. (1993) Proc. Natl. Acad. Sci. USA 90, 3251–3255.
[29] Halban, P.A. and Irminger, J.-C. (1994) Biochem. J. 299, 1–18.