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

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

*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 extracellularly and sequestered by the growing oocytes [1]. Insect vitellogenins, with the exception of yolk proteins of higher Dipterans, are multimeric glycoproteins 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 x Asahi) was provided by Dr. Shigetoshi Mori [8]. The library (1.5 x 105 pfu) was screened by the method of plaque hybridization with the 32P-labeled vitellogenin cDNA probe [4], and 3 overlapping clones were isolated. Insert DNA of one of the recombinant phage (AGVL1) 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-guanidium-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 32P. 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 Sp6 site within the first intron of vitellogenin gene (+173) was labeled with 32P. The labeled DNA was digested with Hinfl 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
Fig. 1. Structure of the vitellogenin gene. (A) Schematic diagram of the vitellogenin gene. Solid bar indicates restriction map of IGVL1. 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, EcoRI; K, KpnI; P, PstI; S, Sall; V, EcoRV. (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.

32P-labeled vitellogenin cDNA probe. Since restriction maps of all the isolated clones overlapped each other, one of these clones (IGVL1) 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.

The initiation site of transcription was determined by S1 nuclease mapping and direct sequencing of vitellogenin mRNA.

Fig. 3. S1 nuclease mapping analysis of the precursor of vitellogenin mRNA. The SstI site (position +173) within the first intron was end-labeled, and HinfI/SstI 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-1 fifth larval instar; lane 2 = day-2 pupae; lanes 3 and 5 = day of larval-pupal ecysis; lane 4 = day-2 pupae.
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.

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 endopeptidase [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