Characterization of the P25 silk gene and associated insertion elements in 
*Galleria mellonella*

Changsong Yang a,1, Xiuyin Teng a, Michal Zurovec a, Klaus Scheller b, Frantisek Sehnal a,*

a Institute of Entomology, Academy of Sciences, and Faculty of Biological Sciences, University of South Bohemia, 370 05 Ceske Budejovice, Czech Republic  
b Department of Zoology, University of Wurzburg, 8700 Wurzburg, Germany

Received 2 September 1997; accepted 16 December 1997; Received by B. Dujon

Abstract

Insect silk genes attract attention by their precise territorial and developmental regulations and extremely high expression rates. Our present investigations demonstrated that the P25 silk gene of Galleria mellonella is down-regulated by ecdysteroid hormones. The gene was identified within 5217 nucleotides (nt) of two genomic clones. In contrast to other silk genes, Galleria P25 lacks the canonical TATA box. Transcription is initiated within a region of three nucleotides that lie at the end of a capsite initiator sequence ACAGT and about 90 nt downstream from a CAAT box. A stretch of 32 nt with a core sequence CTTTT was detected in the 5′ region of Galleria P25 as well as in the presumptive regulatory regions of all other silk genes that are expressed in the posterior silk gland. However, consensus sequences reported for the regulatory regions of Bombyx silk genes are not obvious in Galleria P25. The coding sequence of this gene includes 654 nt, is interrupted by 4 introns, and ends in position +3369; a potential polyadenylation signal starts at +4382. The gene contains 3 copies of a short interspersed nuclear element (SINE), which are located in the upstream region (−833 to −579) and in the first (+542 to +840) and second (+2259 to +2556) introns. The repeat, which was named Gm1, occurs in some other Galleria genes and exhibits homology to Bm1 SINE of the silkworm and to a similar element of a spider. Another insertion of at least 150 nt and with loosely defined borders is present in the 3′ untranslated region (UTR) of Galleria P25. It includes a box (+3453 to +3552) of 99 nt that is tentatively called Lep1 because it was disclosed also in some other Lepidoptera. Lep1 seems to represent the core region of insertion elements that occur in the genomes of lepidopteran insects in various species specific and region specific modifications. © 1998 Elsevier Science B.V.

Keywords: DNA repeats; Ecdysone; Fibroin; Genomic evolution; SINE

1. Introduction

The salivary glands of some insect larvae produce proteinaceous threads known as silk. In the larvae of lepidopteran insects, the posterior section of such silk glands secretes a core of the silk fibre, whereas the middle section provides the core with a sticky coating made of several sericin proteins. For two distantly related lepidopteran species, the waxmoth, Galleria mellonella, and the silkworm, Bombyx mori, it has been shown that the core is composed of three types of proteins known as heavy-chain fibroin, light-chain fibroin, and P25 (Tanaka et al., 1993; Zurovec et al., 1995). Proper assemblage of these components provides the strength, resistance, and contractibility of the fibre.

The expression of silk genes is restricted to appropriate sections of the silk glands and fluctuates during development in dependence on nutrition supply and in response to hormones. DNA sequences indispensable for this spatial and temporal specificity of expression were studied in Bombyx silk genes. Most work was done with the *Fib-H* gene, which encodes heavy chain fibroin, and with the P25 gene. These two genes share a number of short, presumably regulatory motifs in their 5′ flanking regions (Couble et al., 1985). Y. Suzuki and co-workers
Galleria mellonella L., reared on a semiartificial diet

2.1. Effect of 20-hydroxyecdysone on P25 expression

All experiments were performed with the wax moth, Galleria mellonella L., reared on a semiartificial diet (Sehnal, 1966). Fully grown Galleria larvae were anaesthetized in water and ligated with a thread across mesothorax; body region anterior to the ligation was cut off. The removal of head and prothorax deprived the insects of ecdysteroids and other hormones and caused a cessation of development. The abdomens, which survived for several weeks, were taken for experiments 3 days after ligation. Experimental specimens were each injected with 1.6 or 2.4μg of the natural ecdysteroid hormone, 20-hydroxyecdysone, which was dissolved in 2 or 3μl of 8% ethanol; controls were injected with the solvent alone. Silk glands were dissected 24 h after the injection. RNA extractions and detection of the P25 mRNA with a cDNA probe (Zurovec et al., 1998) were done as described previously (Yang et al., 1996).

2.2. Preparation and screening of a genomic library

High-molecular-weight DNA was prepared from newly ecdyised last instar Galleria larvae that had a low fat content and nearly empty guts. DNA was partially digested with Sau3A; fragments of 15–20 kb were isolated by centrifugation in a sucrose gradient and ligated to XhoI half-site arms of LambdaGEM12 vector (Promega). Phages assembled with the Packagene system (Promega) were introduced into the LE392 strain of Escherichia coli. Transformed cells were screened with the 5′ (328 bp) and 3′ (322 bp) fragments of Galleria P25 cDNA (Zurovec et al., 1998). The probes were digoxigenin- or 32P-labeled with the Random Primed DNA Labelling Kit (USB). Two genomic clones, designated GGE1 and GGE2, were found by screening 4.0 × 105 plaques with the 5′ and 3′ cDNA fragments, respectively.

2.3. Sequencing, primer extension, and Southern analysis

Standard methods of DNA analysis were used (Sambrook et al., 1989). The genomic clones were mapped with restriction and subcloned; three sets of nested subclones were generated with unidirectional deletion method (Erase-a-Base System, Promega) and the full sequence of the 5′ end of the cDNA was determined. Over the years, they have identified the silkworm P25 as a critical gene involved in the control of ecdysteroid hormone expression. The sequence was usually read from two overlapping DNA motifs. This perspective prompted us to elucidate the full sequence of Galleria P25 gene. We report here that this gene has a similar structure to the Bombyx silk genes. In experiments described in this paper, we verified that expression of the newly identified P25 gene in Galleria is regulated by ecdysteroid hormones. This result and other data on the insects of ecdysteroids and other hormones and sequenced as depicted in Fig. 2. Single- or double-stranded templates were sequenced by the dye dideoxy chain termination reaction using [α-35S]dATP or [γ-32P]ATP, T3 or T7 primers, and the Sequenase Version 2.0 DNA Sequencing Kit (Amersham). Specifically designed primers were used to confirm the junctions of particular subclones. The sequence was usually read from two overlapping subclones of one DNA strand and, in cases of doubt, also from subclones of the opposite DNA strand. For Southern analysis, aliquots of 5 μg genomic DNA from Galleria larvae were digested with XhoI, Sall, and PviI, respectively, and electrophoresed on 0.8% agarose gel. The blots on nylon membrane were probed with the 32P-radio labeled 5′ fragment (328 bp) of Galleria P25 cDNA. Relative positions of HindIII-digested λ phage DNA fragments were used as size markers. A 22-nt synthetic oligonucleotide complementary to the region +68 to +89 of the gene (Fig 3) was employed for primer extension analysis. Using T4 polynucleotide kinase, the primer was 5′-labeled with [γ-32P]ATP, and a dose of 5 × 104 cpm was hybridized
3. Results and discussion

3.1. Down-regulation of \( P25 \) by 20-hydroxyecdysone

In a previous study, we found that the silk gland content of \( P25 \) mRNA decreased at the last larval molt, increased during the intermolt period, and dropped before the pupal ecdysis of \textit{Galleria} (Zurovec et al., 1998). This pattern indicated that the surges of ecdysteroids, which precede and evoke ecdyses (Sehnal et al., 1981), are causally associated with the declines of \( P25 \) mRNA.

To prove that high ecdysteroid concentrations cause a \( P25 \) turn-off, we prepared isolated abdomens that are devoid of these hormones. The content of \( P25 \) mRNA in the silk glands of such abdomens was gradually reduced, obviously as a consequence of insufficient nutrient supply, but 3 days after ligation it was still relatively high (Fig. 1, lane 1). Injections of 20E in physiological doses caused a dramatic decrease in mRNA (Fig. 1, lanes 2 and 3), whereas injection of the solvent had no appreciable effect on the gradual decline of the mRNA content (Fig. 1, lane 4). These results were consistent with our previous observations on the ecdysteroid effects on other silk genes of \textit{Galleria} (Yang et al., 1995, 1996). The source of endogenous ecdysteroids was removed by ligating off the head and thorax of fully grown larvae. Relative content of \( P25 \) mRNA in the glands was established 3 days after ligation when the insects were injected (lane 1) and 24 h after the injection of 1.6 \( \mu \)g 20E (lane 2), 2.4 \( \mu \)g 20E (lane 3), and the solvent alone (lane 4), respectively.

The products were analyzed on a sequencing gel (8% polyacrylamide, 7 M urea) with the aid of G, A, T, and C ladders of appropriate genomic subclone that was sequenced with the same primer.

![Fig. 1. Effect of 20-hydroxyecdysone (20E) on \textit{Galleria} \( P25 \) expression.](image)

![Fig. 2. Structure and sequencing strategy of \textit{Galleria} \( P25 \) gene. (A) Partial restriction map of two genomic clones (GGE1 and GGE2); horizontal arrows indicate the direction and extent of sequencing. (B) Schematic chart of gene organization; exons (E1–E5) are indicated by hatched boxes, the Gm1 elements by boxes with vertical lines.](image)
Fig. 3. Nucleotide sequence of Galleria P25 gene with the 5'-flanking region and the deduced amino acid sequence. Nucleotides are numbered (thick arrow = +1) from the central position of three nucleotides that were identified as alternative transcription starts (Fig. 4). The CAAT box is underlined and the potential polyadenylation signal double-underlined, the introns are shown in lower-case letters, and the stop codon is marked with asterisks. The repetitive elements Gm1 in 5'∞UTR and the 1st and 2nd introns, and the insertion element Lep1 in 3'∞UTR are underlined. The complementary oligonucleotide primer used for the primer extension experiment is overlined. The beginning of sequenced cDNA is earmarked with a diamond, and the end with a triangle. The gene sequence has EMBL Accession Number AF009677.

hoped that a comparison of Galleria P25 gene with the Bombyx silk genes would reveal cis-motifs involved in this regulation.

3.2. Isolation of Galleria P25 gene

Several genomic clones were found to hybridize with our P25 cDNA probes. On the basis of restriction mapping and Southern blot analysis, we chose to sequence appropriate parts of two overlapping clones GGE1 and GGE2, which included the entire P25 gene. The restriction map and sequencing strategy of the clones are shown in Fig. 2. A comparison of the established Bombyx silk genes would reveal cis-motifs involved in their regulation. sequence of 5217 nt (Fig. 3) with the P25 cDNA sequence (Zurovec et al., 1998) confirmed that these sequences corresponded to the same gene. The previously reported cDNA sequence began in position +34 of our gene numbering, included all coding exons, and terminated at +3739. The cDNA and the genomic sequences differed only in 3 nt; two at silent codon positions (C in place of T at +1043 and G in place of C at +3258), and one (T instead of A at +3374) in the 3' untranslated
3.3. The structure of the Galleria P25 gene

Comparison of the genomic sequence with the P25 cDNA (Zurovec et al., 1998) disclosed that the P25 gene was interrupted by four introns. The donor and acceptor sequences of all exon-intron junctions conformed to the GT/AG rule (Brehmneh et al., 1978). Exon 1 (101 nt) encoded the 5'-UTR and the signal peptide; exon 2 (223 nt) the N-terminus of the secreted protein; exon 3 (136 nt) and exon 4 (110 nt) the internal part of the protein; and exon 5 encoded protein C-terminus and contained 3' UTR. The exon-intron boundary domains and the length of exons of Galleria P25 gene are almost identical as described for the Bombyx P25 (Chevillard et al., 1986). The Galleria gene, however, contains much longer non-coding sequences than the Bombyx counterpart. Extended lengths of the first and second introns of Galleria P25 are due to insertions of a SINE motif, which is also present in the upstream region (see Section 3.6). A different insertion element extends the length of the 3'-UTR (see Section 3.6). P25 genes of Galleria and Bombyx also seem to differ in their regulatory regions.

3.4. Features of the 5'-flanking region

No canonical TATA box sequence was found around position –30 of the 5'-flanking region (Fig. 3). Two typical TATA recognition sequences TATAAAA, however, were evident at –588 to –584 and –575 to –571 positions from the major tsp. A similar situation is known from some other genes, for example a gene encoding human thrombin receptor (Schmidt et al., 1996). Since the relevance of a TATA box at such a distance from the tsp has not been proven, we regard
the Galleria P25 gene as TATA-less. This is common in the 'housekeeping' genes (Azizkhan et al., 1993)—about one-half of known Drosophila promoters are TATA-less (Arkhipova, 1995)—but the absence of a TATA box in a silk gene is exceptional. The 5' flanking region of Galleria P25 contains, however, a CAAT box (Fig. 3) that is located in the usual position around 80 nt upstream from the tsp (Breathnach and Chambon, 1981).

Using sequence alignment, we did not detect in Galleria P25 any of the cis-elements responsible for the spatial and temporal control of Bombyx P25 and Fib-H expression (Nony et al., 1995; Takiya et al., 1997). However, juxtaposing 5' flanking region of the Galleria P25 gene to the sequences of other silk genes revealed in position —76 to —45 a putative element of 32 nt (Fig. 5). The element was over 65% identical with a sequence present in the 5' flanking region of Galleria Fib-H (unpublished data). The core motif GTCTTTT and identical spacing of a few other nucleotides were also revealed in the upstream regions of Bombyx genes P25 (Coublet et al., 1985) and light chain fibroin (Fib-L) (Kikuchi et al., 1992), as well as in the intron of Fib-H (Hut et al., 1990b). Since all these genes are specifically expressed in the posterior section of silk glands, the motif may play some role in the territorial regulation of gene activity.

A search for transcription factor recognition elements using the MatInspector (Quandt et al., 1995) revealed a number of putative binding sites for homeoproteins (data not shown). Binding of several homeoproteins to regulatory silk gene sequences was demonstrated in Bombyx (Hut et al., 1990a; Mach et al., 1995, etc.), but the established cis-elements differed from the putative homeoprotein-binding sites that were found in Galleria P25 in both their sequences and positions.

3.5. Short interspersed nuclear element

The genomes of higher eukaryotes contain multiple copies of short interspersed nuclear elements (SINEs) (Davidson et al., 1973). A repetitive SINE-type element present in the genomes of lepidopteran insects is widespread. A search for transcription factor recognition elements using the MatInspector (Quandt et al., 1995) revealed a number of putative binding sites for homeoproteins (data not shown). Binding of several homeoproteins to regulatory silk gene sequences was demonstrated in Bombyx (Hut et al., 1990a; Mach et al., 1995, etc.), but the established cis-elements differed from the putative homeoprotein-binding sites that we found in Galleria P25 in both their sequences and positions.

Fig. 5. A 32-nt sequence motif identified in the 5' flanking regions of Galleria genes P25 (GP25), and Fib-H (GFib-H), in the 5' flanking regions of Bombyx P25 and (twice) Fib-L genes, and in the intron of Bombyx Fib-H. Multiple alignment was performed using Clustal method in the DNASTAR program. Numbers in parentheses indicate nt positions.
stream was detected in the first intron of Fib-L (Kikuchi et al., 1992), in a randomly amplified DNA sequence (Abe et al., 1995), in the fourth intron of the xanthin dehydrogenase gene (Nimura et al., 1996), and in the third intron of the xanthin dehydrogenase gene (GenBank Accession No. AB005911; Komoto et al., 1997). The

<table>
<thead>
<tr>
<th>Fig. 6. Multiple alignment of Gm1 SINE elements in the upstream region (UP), first intron (Intron 1) and second intron (Intron 2) of P25, in the upstream region of Lhp76, and in intron 6 of Lhp82 Galleria genes. Sequences of the repeats Bm1 from the genome of Bombyx mori and Nc1 from the spider Nephila clavipes are also juxtaposed.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>UP</strong></td>
</tr>
<tr>
<td>Gm1SINE</td>
</tr>
<tr>
<td>Lhp76</td>
</tr>
<tr>
<td>P25</td>
</tr>
<tr>
<td>Bm1</td>
</tr>
<tr>
<td>Nc1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Fig. 7. Multiple alignment of the Lep1 box and adjacent sequences detected in Galleria P25 3' UTR (Gen P25) with Bombyx sequences present in the storage protein 2 gene (Bm SP2), a spacer in bombyxin locus (Bm Space), randomly amplified DNA (Bm RAPD), Fib-L gene (Bm Fib-L), xanthin dehydrogenase gene (Bm XDH), and cuticular protein 12 gene (Bm Cut12).</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Bm SP2</strong></td>
</tr>
<tr>
<td><strong>Bm XDH</strong></td>
</tr>
<tr>
<td><strong>Bm Cut12</strong></td>
</tr>
</tbody>
</table>

---

Data from: C. Yang et al. / Gene 209 (1998) 157–165
beyond about 10 nt upstream and 50 nt downstream of the box.

Only about 50% homology within the box consensus and occasional stretches of matching nucleotides downstream of the box (Fig. 7) were found in the 5' UTR of Bombyx P25 (Chevillard et al., 1986), as well as in the presently identified Lepl insertion in the 3' UTR of Galleria P25 (Fig. 3).

A somewhat different group of elements lacking the first third of the Lepl box was disclosed in the moth Hyalophora cecropia. All detected sequences included only two copies of the TGTT motif that was present four times in the Lepl consensus (Fig. 7). In the sequence between the attacin pseudogene and the functional gene for basic attacin (Sun et al., 1991), and in the 5' UTR of the gene for cuticle protein 12 (Binger and Willis, 1994), the TGTT motif is preceded by seven or eight TGTC repeats (Fig. 7). By contrast, unique sequences terminated with just one TGTC motif precede the Lepl homology in the introns of the genes for cecropin (Gudmundsson et al., 1991) (Fig. 7) and for cuticle protein 66 (Lampe and Willis, 1994) (not shown).

All Hyalophora sequences with the Lepl-like box are over 80% identical also in about 80 nt downstream of the box (Fig. 7).

The TGTC repeats were recognized in the Hyalophora attacin locus by Sun et al. (1991) as part of a 1264-nt sequence ACATCTA... TAGACAT (underlined parts are inverted repeats). The authors suggested that the direct repeat ACAT was the target site of an insertion. This is consistent with our assumption that the Lepl box is a small internal part of a larger inserted element. It seems that a mobile element common to Lepidoptera was modified during evolution to diverse forms. For a majority of the elements that we compared, Lepl was the most conserved part. However, a stretch of cca 140 nt, which is shared by the Hyalophora elements, does not include the first third of Lepl.

The alignments in Fig. 7 show that Lepl-type insertions detected in Bombyx could be classified into several categories that differ in the degree of box homology and particularly in the diversification of the regions flanking the box. A single Lepl-type insertion presently identified in Galleria represents a related type, whereas the insertions known from Hyalophora are of two other kinds. It seems that insertions are residues of an ancient mobile element. Some of their modifications, such as the lack of the first third of Lepl box, are taxon-specific, whereas others are found side by side within a single genome.

4. Conclusions

(1) The expression of Galleria silk gene P25 is down-regulated by molt-inducing concentrations of ecdysteroid hormones.

(2) The gene, which was isolated as two overlapping genomic clones, is present in one copy per genome, includes about 4440 nt and is interrupted by four introns.

(3) No canonical TATA box sequence is located at appropriate position of the gene. The 5'-flanking region contains a 32-nt sequence that is partly conserved in all genes that are expressed in posterior silk gland section.

(4) DNA repeat Gm1 of 300 nt, which is homologous to SINE-like element Bm1 of the silkworm and to a spider repetitive element, was disclosed in the upstream region, intron 1 and intron 2 of the P25 gene, and in several other Galleria genes.

(5) An apparent insertion detected in the 3' UTR of Galleria P25 gene contains a conserved Lepl box of cca 100 nt that is widespread in lepidopteran insects. It represents the core region of a DNA element that occurs in Lepidoptera in species-specific modifications.

Acknowledgement

Our study was supported by grant 204/96/1100 of the Grant Agency of the Czech Republic, and by Volkswagenstiftung grant I68 514.

References


Binger, L.C., Willis, J.H., 1986. Identification of the gene, which was isolated as two overlapping genomic clones, is present in one copy per genome, includes about 4440 nt and is interrupted by four introns.

Bucher, P., 1990. Weight matrix descriptions of four eukaryotic RNA
Takaya, S., Hui, C.-C., Suzuki, Y., 1990a. A contribution of the promoter and its surrounding regions to the preferential transcription of the fibroin gene in posterior silk gland extracts. EMBO J. 9, 489–496.