Journal of Insect Biotechnology and Sericology 71, 1-15 (2002)

Review

The Silk of Lepidoptera

Robert Fedič, Michal Žurovec and František Sehnal*

Institute of Entomology, Academy of Sciences, and Faculty of Biology, University of South Bohemia, Branišovská 31, 370 05 České Budějovice, Czech Republic

(Received January 31, 2002; Accepted February 4, 2002)

The larvae of most Lepidoptera secrete fibrous proteins from their labial glands, collectively known as silk. Silk is used to construct shelters for the larvae and cocoons in which the larvae pupate. The water insoluble core of the silk filament is produced in the posterior gland section. Its crucial component, the heavy chain fibroin (H-fibroin), is characterized by hierarchical arrangement of species specific amino acid repeats. The *H-fibroin* genes of different Lepidoptera are homologous but their repetitive region has diversified. Amino acid composition, complexity of basic repetitive units, and arrangement of these units into iterated higher order domains determine physical properties of the silk fiber. In most Lepidoptera, the fiber includes two other peptides, light chain fibroin (L-fibroin) and P25 that occurs as two proteins differing by the degree of glycosylation. For the silkworm it has been shown that six H-fibroin, six L-fibroin, and one P25 molecules are assembled into an elementary silk unit. The fiber of tussah silk is different because it lacks the L-fibroin and P25 components; it seems to be composed of H-fibroin dimers. The fiber core in all species is enveloped by several sericins that are derived from two genes expressed in the middle section of the gland. Differential splicing and probably also different extent of glycosylation provide for the diversity of sericins that serve as a glue for fiber adherence. The silk further contains several proteins of smaller size that are not essential for silk function but may provide protection against predators, molds, and microbes. Greatly diversified proteins called seroins, to emphasize their production both in the middle, sericin producing and the posterior, fibroin producing gland sections, were identified as products of one or two genes, depending on the species. Two other genes specifically expressed in the silk glands encode a Kunitz and a Kazal type protease inhibitors that are active on certain fungal and bacterial proteases. The inhibitors are produced before cocoon spinning and constitute very stable cocoon components. Structures of several other small silk proteins await elucidation. Silk produced by a handful of species has been woven into textiles for several millennia, and used as additive into cosmetics and occasionally also remedies. Modern technologies offer the potential to produce for practical use specific silk components such as the protease inhibitors. Creation of recombinant silkworms secreting silk fibers of new quality has become feasible. Industrial spinning of a silk dope prepared under *in vitro* conditions is a challenge for current research.

Key words: *Bombyx*, *Antheraea*, *Galleria*, fibroin, sericin, seroin, protease inhibitors

INTRODUCTION

Silk is one of the few commodities that are directly derived from the insects. Sericulture as agricultural technology dates back more than 5000 years when the domestic silkworm, *Bombyx mori*, first appeared in China. The silkworm remains the major source of the commercial silk, although a handful of wild silkmoth species are also used. With the advancement of science about a century ago, practical considerations as well as the suitability of silk proteins and silk glands as model systems for theoretical studies stimulated extensive research centered around silk. Owing to the effort of several generations of Japanese scientists, the silkworm became a very well known insect species, second only to *Drosophila*. Investigations on the silkworm contributed significantly to the present knowledge of insect physiology and genetics, and led to several discoveries of great general significance. In the realm of chemistry, crystallographic measurements of silk led to understanding of the peptide arrangements in pleated β-sheets (Pauling and Correy, 1953). In biology, ultrastructural studies of silk glands greatly contributed to our understanding of the role of ribosomes in proteosynthesis (Akai, 1965), and biochemical investigations unearthed functional adaptation of isotRNAs to the frequency of isocodons in abundantly expressed genes (Nony *et al.*, 1995). The fibroin gene was among the first eukaryotic genes expressed in an acellular system and analysis of this process fostered understanding of the mechanisms regulating gene expression such as the extent and localization of the regulatory cis-sequences and corresponding trans-elements (reviewed by Hui and Suzuki, 1995).

Modern scientific methods greatly accelerated the expansion of knowledge on silk composition and production. Investigations were performed on various spider silks and the interest in insect silk expanded to

* To whom correspondence should be addressed.
Fax: +420-38-5300354. Tel.: +420-38-5300350.
Email: Sehnal@entu.cas.cz.

several lepidopteran species. Unfortunately, virtually no information is available on the silks of lepidopteran sister group, the Trichoptera, and the more distantly related orders of the panorpoid clade, Mecoptera, Siphonaptera, and Diptera. The results of comparative studies on Lepidoptera proved that they share basic silk components but that the internal structure of the principal proteins is extremely diversified. Due to this diversification, silks produced by some of the wild species acquired propensities superior to those of the commercial silks. Structural analysis of hitherto neglected silk types has become an important trend in modern sericultural research. Another new aspect of current sericultural investigations is the search for silk components that are not essential for the fiber construction but have other functions and properties of practical importance.

Information on new types of silk and finding of hitherto unknown silk components open a new era of sericulture. Advances of biotechnologies allow us to synthesize artificial genes encoding natural or modified low molecular silk components and to express them in various vectors, e.g. bacteria and yeast. Fabrication of dope based on silk proteins has been achieved in a number of host systems and artificial spinning of a recombinant spider silk produced in mammalian cells has been successfully accomplished (Lazaris *et al.*, 2002). Production of recombinant lepidopteran silks with such desired properties as high elasticity has become feasible by the refinement of the technology for silkworm transgenesis. Selection of suitable silk genes is crucial step in the exploitation of all these technologies. The aim of this review is to aid this selection and to foster the use of low molecular silk components.

SILK PRODUCTION AND USE IN LEPIDOPTERA

The ability to produce silk evolved in several insect lineages (Sehnal and Akai, 1990; Craig, 1997). Silk production by labial glands, which is typical for the caterpillars of Lepidoptera, is probably rooted in common ancestors of several endopterygote orders. It occurs in the larvae of Hymenoptera, Mecoptera, Siphonaptera, Diptera (although never in Cyclorrhapha), Trichoptera and Lepidoptera. Silk utilization is mostly restricted to the construction of cocoons where the insects pupate. In some cases the silk is used also for the protection of larvae, attachment of larvae or pupae to a substrate, dispersion of larvae, and building of traps for food collection or capture. The capacity for silk production, its use and properties greatly vary within all listed orders.

Silk production is a groundplan function of the larval labial glands in Amphiesmenoptera (Trichoptera + Lepidoptera) but was lost in some of their lineages (Akai *et al.*, in press). The function of labial gland secretion in such lineages is insufficiently known. We believe that larval labial glands of all Lepidoptera produce homologous proteins that might have evolved to allow polymerization into fibers. However, genes encoding some of these proteins are subject to rapid mutations and the changes in protein structure alter physical properties of the secretion. Its function can rapidly be adapted to new needs and used as glue, silk or a parchment-like film. Information on the heavy-chain fibroin demonstrates the velocity and the extent of such structural changes that were in this case restricted by upholding the ability of fiber polymerization. Comparison of structural changes involved in converting silk to glue or *vice versa*, as seen in several lepidopteran families, may reveal what is actually crucial for the polymerization of the labial gland secretion into fiber.

Several splitting events in the evolution of Lepidoptera gave birth to the seven basal superfamilies before the clade Neolepidoptera was established (Kristensen, 1999). Silk is not known in any of the basal groups (Akai *et al.*, in press) and the function of the secretion of their labial glands is mostly unknown. The larvae of Agathiphagidae mining in conifer seeds may use it to produce the varnish-like lining of the pupal cell in the mined seed. Silk production is common in the larvae of Neolepidoptera that are typical caterpillars with crochet-bearing prolegs that might have evolved to facilitate movement on a silken substrate. The intensity of silk production greatly varies within most neolepidopteran superfamilies and often also within families. The classification of the intensity of silk production presented in Fig. 1 considers the best silk-producers in a given group. In some species, silk production is important for the first instar larvae that use silk threads for dispersal by wind (e.g., in Zygaenoidea, Cossidoidea, Tortricoidea, and Lymantriidae). In most superfamilies, one finds caterpillars that spin silken tubes (Tineoidea, Pyraloidea) and nests in which they hide. The constructions may be very obvious and cover the entire host tree, as in the genus Yponomeuta or in the tent caterpillars of the family Lasiocampidae. Much more silk is used to construct larval hides than to spin cocoons for pupation. Caterpillars of several clades (Gelechioidea, Tortricoidea, Pyraloidea, rarely in Papilionoidea and Noctuoidea) use silk to fold, roll or tie leaves of their food plants into protective shelters, others use silk and foreign materials to construct portable cases (Tineoidea). Silken lining and sealing of burrows in which the caterpillars live and feed is also common (Hepialoidea, Noctuoidea). Some caterpillars use silk spinning as a devise to descent from the canopy where they have fed to the ground where they pupate (Pyraloidea, Geometroidea, Noctuoidea). Caterpillars of most Bombycoidea and Noctuoidea spin large amounts of silk only at the end of larval development and construct dense cocoons. Most Sphingidae and Noctuidae,

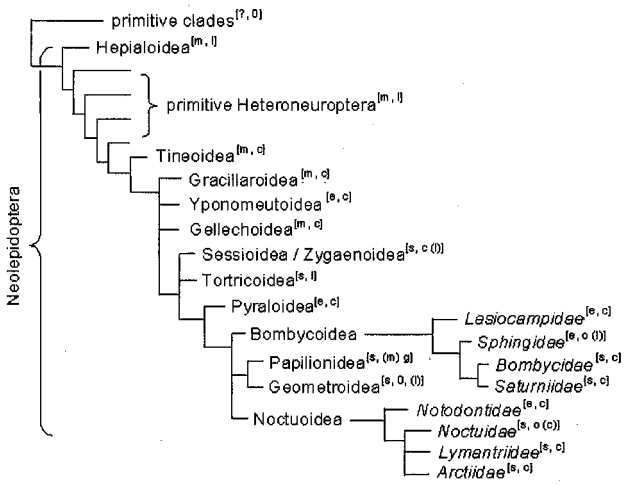


Fig. 1. Overview of major lepidopteran superfamilies and representative families (in italics) of Bombycoidea and Noctuoidea. Spinning activities of early instar larvae are indicated with superscript letters [s] (light spinning), [m] (moderate), and [e] (extensive), and spinning intensity associated with cocoon formation with letters [c] (well-formed cocoon is produced), [l] (loose cocoon or similar shelter), and [g] (silken girdle or similar attachment).

however, lost this capacity and pupate in soil chambers whose walls are smoothed by the secretion of the labial glands (a few species spin loose cocoons). Pupation without a cocoon is also common in Geometroidea but some species produce loose cocoons or attach pupa to substrate by a silken girdle. Pupal girth and exceptionally flimsy cocoons are produced before pupation by some Papilionoidea.

Silk exploitation by man is limited to a small number of species that produce large cocoons (Peigler, 1993). These species belong to families Lasiocampidae, Bombycidae, Saturniidae, and Notodontidae (Fig. 1) whose adults escape from the cocoon through an opening they make with their enzymatic secretions. In most Lepidoptera, the pupating larva bites a hatch in the cocoon wall. Silk is used almost exclusively by the textile industry, to small extent as an additive to cosmetics, and marginally as a source of remedies. The potential of practical use of the non-silky secretions of the labial glands is unknown.

THE SILK GLANDS

The silk of caterpillars is derived from a pair of labial glands that in most other insects serve as salivary glands (cf. review by Akai, 1984). The gland consists of monolayered tubular epithelium that is longitudineally differentiated into functionally distinct posterior (PSG), middle (MSG), and anterior sections. The embryonic development of silk glands has been described in the silkworm as a succession of 30 distinct morphological stages (Nunome, 1937). Silk gland organogenesis is initiated by invagination of a patch of ectoderm in the labial segment, and is completed several days later. Silk proteins are secreted into the lumen before hatching (Ohta *et al.*, 1988). The subsequent growth of the gland epithelium occurs via endomitotic cycles of DNA replication, leading to great cell enlargement (Goldsmith and Kafatos, 1984). In the last larval instar of the silkworm, the silk gland cells are more than a thousand-ploid. They are characterized by well developed rough endoplasmic reticulum, numerous Golgi bodies and secretory vesicles from which primary secretory silk globules are released into the gland lumen (Akai, 1965). The entire circumference of the gland tube is in most Lepidoptera made up of just two cells, rather exceptionally of four cells (Sehnal and Akai, 1990).

The intensity of silk production during larval development is different in different taxa (Fig. 1). Caterpillars of many species, such as the waxmoth, spin nearly continuously to construct protective tubes in which they hide, eventually spinning cocoons in which they pupate. Just in the last larval instar of the waxmoth, silk produced prior to cocoon spinning amounts to more than 20% of the pupal weight, while the cocoon silk represents less than 7% of the pupal weight (Jindra and Sehnal, 1989). In the silkworm and silkmoths, spinning is minimal until the last larval instar when half of the biomass is invested into the cocoon.

Irrespectively of the mode of spinning, the silk gland cells undergo a cycle of secretory activity in each larval instar (Prudhomme *et al.*, 1985; Sehnal and Akai, 1990). The cycle consists of four overlapping phases characterized by cell growth, silk production, cessation of the production, and regression of the proteosynthetic machinery. In the silkworm, DNA synthesis and cell enlargement seem to occur exclusively in the growth phase at the beginning of each instar, when the cells also accumulate rRNAs and tRNAs. The secretory phase commences when the content of DNA attains a maximum for a given instar, while the RNA levels continue to increase. Activities of the enzymes involved in amino acid metabolism and proteosynthesis reach their maxima. A decline of the RNA and later also of the protein levels mark the termination of silk synthesis when the larva enters apolysis. An increase of enzymes destroying silk in the secretory globules within the cells as well as in the gland lumen signify silk gland regression. The contents of rRNA, tRNAs and the mRNAs encoding silk proteins decline and the glands are functionally silent throughout the molting period. Their function is resumed by the onset of a new growth phase after ecdysis into the next instar.

Functional phases of waxmoth silk glands are slightly different, being adjusted to considerable silk production in course of all larval instars (Sehnal *et al.*, 1983). DNA

synthesis ensues shortly after each ecdysis and continues until the end of the instar. At the beginning of the instar, silk glands pass through a preparatory phase when the rate of RNA synthesis is high but silk production is still low. In the following accumulation phase, which occupies most of the instar, the RNA synthesis is maintained at a moderate level and proteosynthesis rapidly rises to a maximum. The cessation of proteosynthesis overlaps with the regression phase marked by increased activity of proteolytic enzymes and occurrence of autophagic vacuoles, like in the silkworm. However, the contents of mRNAs for some of the silk proteins do not drop during molting as markedly as in the silkworm.

Functional fluctuations are similar in all three silk gland sections but their extent and the nature of secretory products are different. PSG is specialized for the secretion of proteins that constitute the core of the silk fiber (typically H-fibroin, L-fibroin, and P-25), while sericins that provide fiber coating are derived from MSG. MSG is also the major production site of additional silk proteins, such as seroins and protease inhibitors. The role of the anterior silk gland section is little known. Water withdrawal and other modifications of the dope in this section may be crucial for the conversion of liquid silk into solid fiber. In some species, silk thread is supplemented with carotenoid pigments in the anterior silk gland section.

Silk production culminates in the last larval instar before and at the initial phases of cocoon spinning. Thereafter the glands regress in response to increased ecdysteroid titer (Chinzei, 1975a), similarly as during larval molts but at pupation the regression passes over into apoptosis and complete silk gland degeneration. Solubilization and resorption of silk residues in the gland lumen, fragmentation of the nuclei, engulfing of organelles in autophagosomes and autolysosomes, and phagocytosis of large chops of silk gland cells by the hemocytes cause silk gland destruction (Chinzei, 1975b). The glands are histolyzed during early pupal development and their breakdown products are apparently used in the development of imaginal organs and egg formation.

COMPOSITION OF THE SILK CORE

Shimura *et al.* (1976), Gamo *et al.* (1977) and other Japanese researchers found that the silk filament derived from the secretion of *B. mori* PSG contained two dominant components, which could be separated by reductive cleavage of disulphide bonds. The large product was called heavy chain fibroin (H-fibroin) and the small one light chain fibroin (L-fibroin). The composition of the silk core was elucidated in more detail when the methods of protein separation were refined and molecular biology techniques allowed identification and sequencing of appropriate genes. It was then recognized that the silk filament contained an additional component designated P25. The *H-fibroin* gene was first partly characterized by Suzuki and Brown (1972), the *L-fibroin* gene by Kimura *et al.* (1985), and the *P25* gene by Couble *et al.* (1983). Homologous genes were identified in *G. mellonella* by Žurovec *et al.* (1992, 1995). Since *G. mellonella* and *B. mori* are evolutionary distant (Fig. 1), the formation of silk filament from H-fibroin, L-fibroin, and P25 apparently represents a fairly general situation in Lepidoptera. Recent finding of L-fibroin and P25 homologues in *Dendrolimus spectabilis* and *Papilio xuthus*, which represent the families Lasiocampidae and Papilionidae, respectively (Fig. 1), supports this assumption (Tanaka and Mizuno, 2001).

There was initially a contradiction about the identity of L-fibroin and P25 (Kimura *et al.*, 1985). The peptide and nucleic acid analysis (Chevillard *et al.*, 1986; Yamaguchi *et al.*, 1989) and immunological studies (Tanaka *et al.*, 1993; Žurovec *et al.*, 1995) proved unequivocally that they are two distinct silk components derived from unrelated genes. They are present in single copies per genome and each generates a single primary translation product. H-fibroin is a very long polypeptide composed for the most part of reiterated peptidic motifs. For *B. mori* it has been shown that it is associated with a much smaller L-fibroin via a disulfide bond and this assembly is crucial for the intracellular transportation of nascent silk (Takei *et al.*, 1987). L-fibroin occurs in the silk in two forms that differ by the extent of glycosylation. Protein P25 was classified as a chaperonin-like compound allowing the cellular transport and the secretion of the highly insoluble H-fibroin (Tanaka *et al.*, 1999). Recently it has been proved that H-fibroin, L-fibroin, and P25 are assembled into an elementary secretory unit in the ratio 6:6:1 (Inoue *et al.*, 2000, 2001).

Identical building blocks are obviously used in the silk of *D. spectabilis*, representing a different family, and in *G. mellonella* and *P. xuthus*, which belong to different superfamiles than *B. mori* (Fig. 1). Surprisingly the tussah silk, produced by Saturniidae, which are close relatives of Bombycidae, is different. It lacks L-fibroin and P25 and is apparently made up of H-fibroin homodimers (Tamura *et al.*, 1987; Takeda and Mizuno, 2001).

Light-chain fibroins

The structure of the *L-fibroin* gene was elucidated only in *B. mori* (Kikuchi *et al.*, 1992). The gene consists of 13472 nt and its coding region is interrupted with six introns. The 5'-flanking region contains a canonical TATA box and a CAAT sequence and exhibits similarities to the 5' end of the *H-fibroin* gene. Both genes include three specific sequences of 18-30 nt and several common elements of 8-10 nt, six of which coincide with the

bindings sites of homeodomain proteins (Hui *et al.*, 1990). The first exon of the *L-fibroin* gene contains a 41-bp non-coding sequence and a 36-bp coding sequence (Yamaguchi *et al.*, 1989). The first intron is very large (8145 bp) and accounts for about 60% of the total gene length. It contains several interspersed repetitive elements that are for the most part related to *Bm1* (Adams *et al.*, 1986) and *BmX* (Wilson *et al.*, 1988). A single-copy *L-fibroin* gene is also present in *G. mellonella* (Žurovec *et al.*, 1995).

The gene of *B. mori* is transcribed into a single mRNA (1180 nt) and homologous mRNAs of similar length were found in *B. mandarina*, *D. spectabilis*, and *P. xuthus* (Tanaka and Mizuno, 2001). *G. mellonella* has one L-fibroin mRNA of similar size (1192 nt) and a shorter one (1104 nt) that is truncated at the non-translated 3' terminus. The distal 3' part of the full-size mRNA, which is lacking in the truncated one, contains stretches of 33 nt and 38 nt that are present (67% and 76% identity, respectively) in the *L-fibroin* mRNA of *B. mori*.

The deduced sequences of L-fibroin in *B. mori* and *B. mandarina* include 262 amino acid residues of which 261 are identical. L-fibroin homologues of *G. mellonella*, *D. spectabilis*, and *P. xuthus* differ by a few deletions or insertions but the total number of residues is similar, with a minimum 257 in *P. xuthus* and a maximum 267 in *G. mellonella* (Fig. 2). The homologues are about 50% identical, including conserved positions of all three cysteines. Yamaguchi *et al.* (1989) concluded from their data on *B. mori* L-fibroin that two cysteine residues form an intramolecular disulphide bridge while the third one binds with H-chain fibroin. It is worth noting that the L-fibroins of *G. mellonella*, *D. spectabilis*, and *P. xuthus*, representing different superfamilies, are more alike than the L-fibroins of *D. spectabilis* and *B. mori* that are of the same superfamily. Similar correlations are found in some other silk genes, indicating that their diversification in *B. mori* might have been accelerated by artificial selection.

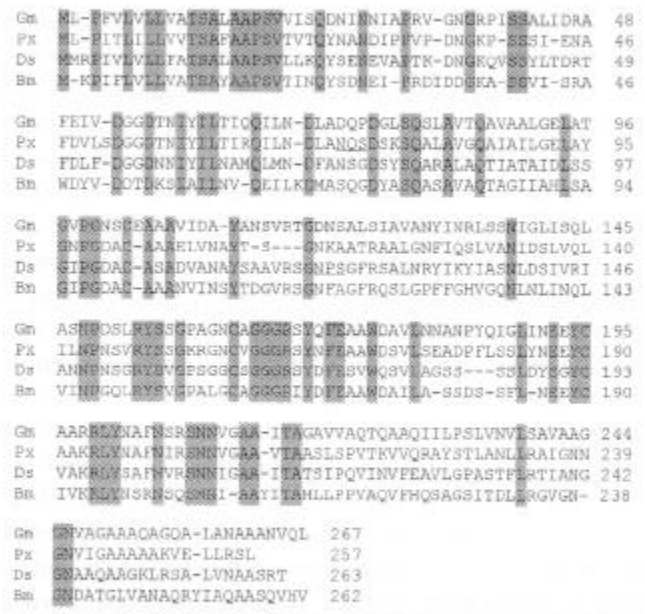


Fig. 2. Alignment of L-fibroins of *Galleria mellonella* (Gm), *Papilio xuthus* (Px), *Dendrolimus spectabilis* (Ds), and *B. mori* (Bm), modified after Tanaka and Mizuno (2001). Identical residues are highlighted and potential N-glycosylation sites are underlined.

The secreted L-fibroin of *B. mori* contains 244 amino acids, with acetylated N-terminal serine, and its deduced size is 25,800 kDa (Yamaguchi *et al.*, 1989). N-terminal sequencing of *G. mellonella* silk proteins showed that there are two isoforms of L-fibroin, one beginning with Ser as the L-fibroin of *B. mori*, and another one extended at the N-terminus by Ala-Pro dipeptide. It was proposed that signal peptide cleavage occurs proximally to this dipeptide that is removed secondarily, presumably by amino dipeptidase IV, in a portion of the L-fibroin molecules (Žurovec *et al.*, 1995). The L-fibroins of *D. spectabilis* and *P. xuthus* each contain one potential N-glycosylation site and the L-fibroin of *P. xuthus* is actually glycosylated (Tanaka and Mizuno, 2001).

P25 proteins

The discovery of the P25 glycoprotein was based on the analysis of cDNAs prepared from the *B. mori* PSG (Couble *et al.*, 1983). The *P25* gene, present in a single copy per genome, is nearly 3500 nt long and includes five exons with 1173 nt (Couble *et al.*, 1985; Chevillard *et al.*, 1986a,b). The 5' flanking sequence exhibits numerous similarities to the 5' end of the *H-fibroin* gene, including similar spacing of a TATA and a CAAT box. Similar *P25* gene organization, but without the consensus sequences reported for the regulatory regions of the *Bombyx* silk genes, was found in *G. mellonella* (Yang *et al.*, 1998). *Galleria P25* lacks the canonical TATA box and 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. The coding sequence of *G. mellonella P25* 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, which are located in the upstream region and in the first and second introns. Another insertion of 150 nt with loosely defined borders is present in the 3' non-translated region.

A single *P25* mRNA of about 1.3 kb was found in *B. mori*, *B. mandarina*, and *P. xuthus* (Tanaka and Mizuno, 2001). The homologous mRNA of *G. mellonella* is around 2 kb large, due to extension at the non-translated 3' end (Žurovec *et al.*, 1998; Yang *et al.*, 1998). A much longer 3' extension occurs in one of the two P25 mRNAs of *D. spectabilis*, whose sizes are 4.4 kb (major) and 1.2 kb

(minor), respectively (Tanaka and Mizuno, 2001). The putative translation products are of similar length in all species, varying from 215 residues in *P. xuthus* to 219 in *D. spectabilis* (Fig. 3). The proteins of *B. mori* and *B. mandarina* differ in a single amino acid, those of the other species share about 50% of the residue positions, including those of 8 cysteines. The signal peptide cleavage occurs in *G. mellonella* and *Bombyx* between alanine and glycine of the AlaGlyPro tripeptide; the site of cleavage in the other species remains to be established. In contrast to other silk proteins, the signal peptide and the start of the secreted peptide show considerable species-specific differences.

P25 proteins of the *Bombyx* species contain three, and those of the other species two N-glycosylation sites (Tanaka and Mizuno, 2001). One of the sites is in identical position in all species, and one in *Bombyx* and *D. spectabilis*; the remaining N-glycosylation sites are in different positions (Fig. 3). P25 seems to bind mannose-rich N-glycans and its apparent size increases to 27 and 30 kDa in *B. mori* (Tanaka *et al.*, 1993), 29 and 30 kDa in *G. mellonella* (Žurovec *et al.*, 1998a), and similarly in the other species (Tanaka and Mizuno, 2001).

Heavy-chain fibroin

H-fibroin is the major, very large protein of lepidopteran silk fiber. The H-fibroin gene of *B. mori* is about 16 kb long (Lizardi and Brown, 1975) and is composed of one very short and one very long exons, and an intron. Identical gene organization was found in *B. mandarina* (Kusuda *et al.*, 1986), *A. yamamai* (Tamura *et al.*, 1987), *A. pernyi* (Sezutsu and Yukuhiro, 2000), and *G. mellonella* (Žurovec, unpublished). The upstream region is AT rich, with frequent concatenations of up to 7 A or T nucleotides.

Fig. 3. Alignment of P25 proteins of *Galleria mellonella* (Gm), *Papilio xuthus* (Px), *Dendrolimus spectabilis* (Ds), and *B. mori* (Bm) taken from Tanaka and Mizuno (2001) and modified. Identical residues are highlighted and potential N-glycosylation sites are underlined.

Several regulatory sequences required for transcription initiation and enhancement, and regulating tissue specificity of the gene expression, were identified with the aid of cell-free transcription systems (reviewed by Hui and Suzuki, 1995). Interspecies homology is obvious downstream from the position about -50, including the presence of a TATA box at -30 to -24 (Fig. 4). *G. mellonella* acquired, probably by coincidental mutations, another TATA box at -49 to -43. The standard TATA box seems to be used in *G. mellonella* at the beginning of the last instar when transcription begins in the usual +1 position, whereas in course of the instar the start of transcription shifts to -18, indicating involvement of the upstream box. Transcription of the H-fibroin gene at the beginning of the last larval instar of *B. mori* is restricted to the anterior region of PSG, eventually spreading to the posterior region (Obara and Suzuki, 1988). It is possible that similar shift in the topology of H-fibroin production in *G. mellonella* is associated with the change in the TATA box use and transcription start.

Homology between H-fibroin genes of different species is very clear through the first exon, including the exon/intron boundary, at the boundary between intron and the second exon, and in about 250 nt at the beginning and 200 nt at the end of this exon (Fig. 4). The length of the intron and its internal sequence, as well as the length and the internal sequence of the second exon are species specific. The intron sequence and the far upstream sequence diversified even between closely related *B. mori* and *B. mandarina* whose *H-fibroin* genes are otherwise nearly identical (Kusuda *et al.*, 1986).

More than 95% of the coding region of the *H-fibroin* is composed of repeats encoding reiterated peptidic motifs. The uniformity of DNA repeats is strengthened by the preferential use of certain isocodons (Yukuhiro *et al.*, 1997; Mita *et al.*, 1988). The codon-choice pattern is species specific and is correlated with skewed cellular contents of the isoaccepting tRNAs (Garel *et al.*, 1970; Suzuki and Brown, 1972). Unequal crossing over and misreplication drive concerted evolution of the DNA repeats and lead to their homogenization (Smith, 1976; Dover, 1982). This process promotes spreading of accidental mutations that do not impair function throughout the repeats. Mutations improving the function are favored by natural selection and become "fixed". Rapid accumulation of neutral or favorable mutations undoubtedly drove rapid diversification of the repetitive part of the H-fibroin gene. Differences in the repetitive H-fibroin region between *B. mori* and *A. pernyi* demonstrate that the region diversified beyond recognition within the Bombycoidea superfamily.

H-fibroin is the principal silk core protein and the structure of the reiterated peptidic motifs determines silk fiber properties. The over-all size of fibroin varies from

200-250 kDa in the examined saturniids (Tamura and Kubota, 1989; Sezutsu and Yukuhiro, 2000; Datta *et al.*, 2001) to 391 kDa in *B. mori* (Zhou *et al.*, 2000), and about 500 kDa in *G. mellonella* (Zurovec and Sehnal, unpublished). This diversity in size is due to differences in the large central protein region that is composed of reiterated motifs. The terminal non-repetitive sequences are of similar length and apparently homologous (Fig. 5). Very alike are these parts of the gene between *B. mori* and *B. mandarina* (Kusuda *et al.*, 1986) and between *A. pernyi* and *A. yamamai* (Tamura *et al.*, 1987). At the N-terminus, the homology extends for about 100 amino acid residues (Fig. 5A). The signal sequence, which is encoded by the first exon, is nearly identical in all analyzed species. The non-repetitive sequence encoded by the second exon contains similar spacing of more than 20 residues arranged in conserved groups. The N-terminus of the *Antheraea* H-fibroin is conspicuous by a deletion of 20 residues. High proportion and conservation of charged amino acids, 10 basic and 13 acidic in *G. mellonella*, 10 and 11 in *B. mori*, and 10 and 12 in *A. pernyi*, respectively, suggest that the electrostatic charge may play a role in fibroin assembling.

The C-terminus of the H-fibroin in *B. mori* and *G. mellonella* is characterized by the presence of 3 cysteines that are located in the proximity of eight basic residues (Fig. 5B). For *B. mori* it has been shown that the cysteine most

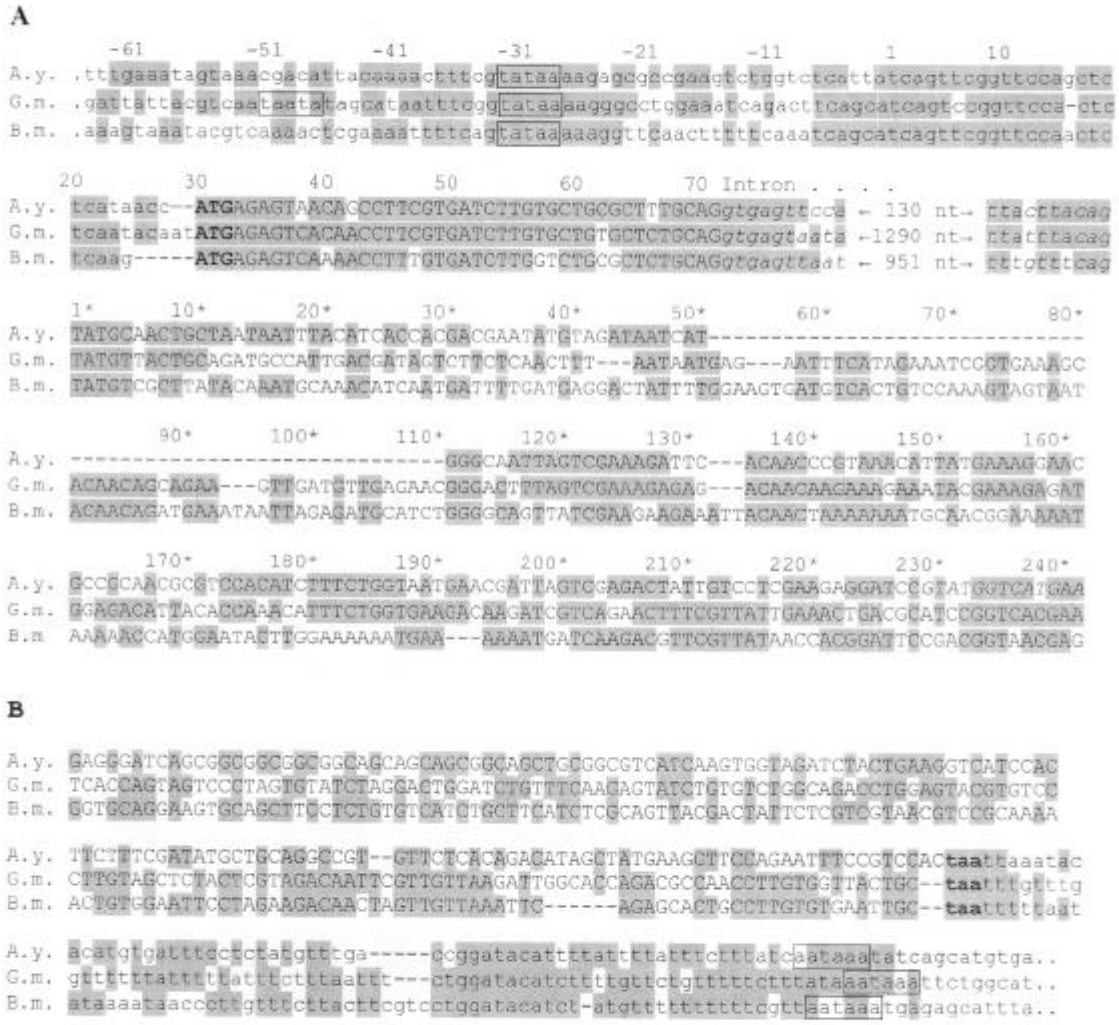


Fig. 4. Terminal parts of the *H-fibroin* genes aligned to maximize homologies (nucleotides present in identical positions in at least two genes are shown on gray background). The non-coding parts are typed in the lower, and the coding parts in the upper case letters. A. y., *A. yamamai* (GenBank entry X05578, Tamura *et al.*, 1987) except the last 13 residues (italics) that represent the closely related gene of *A. pernyi* (AF083333, Suzutsu *et al.*, 2000,); G. m., *G. mellonella* (AF095239 and AF095240); B. m., *B. mori* (V00094, Tsujimoto and Suzuki, 1984, and AF226688, Zhou *et al.*, 2000). (A) The 5' end with the upstream flanking region. The TATA motifs are boxed and the initiation codons printed in bold. Sequences of the upstream region and the first exon are numbered from the transcription start that is used in all species (transcription in G. m. may also begin at the -18 position). The partial intron sequences (italics) are not numbered, and the following coding sequences are numbered from the start of the 2nd exon to emphasize conserved nucleotide positions. (B) The *H-fibroin* 3' ends aligned to match the positions of the termination codon (in bold). The polyadenylation signals are boxed.

distant from the protein end forms a disulfide linkage to the L-fibroin, while the more proximal and the terminal cysteines form an intramolecular disulfide bond (Tanaka *et al.*, 1999). Conservation of the cysteines in the L-fibroin (Fig. 2) and H-fibroin (Fig. 5B) of *G. mellonella* indicates that they may be connected in the same way as in *B. mori*. By contrast, the C-terminus of *Antheraea* H-fibroin is short and includes only 3 basic residues (Tamura *et al.*, 1987). It does contain three cysteines but the general alignment to the H-fibroin C-termini of *B. mori* and *G. mellonella* is ambiguous. The C-terminus of *Antheraea* H-fibroin apparently diversified from the consensus in connection with the loss of L-fibroin (Tanaka and Mizuno, 2001). *Antheraea* H-fibroin is probably made up of homodimers that are held together by a disulfide bridge in which one of the cysteines is engaged (Tamura *et al.*, 1987).

The major central part of the H-fibroin protein is composed of amino acid repeats that are arranged into higher order iterations and eventually into 11 to 12 large domains. Available information allows us to recognize three types of such arrangements that possibly represent Pyraloidea, Bombycidae, and Saturniidae. These H-fibroin types differ by the ratios of major amino acids and by their arrangement into peptidic repeats. The H-fibroin of Pyraloidea, which are represented by *G. mellonella*, are distinguished by the abundance of residues with bulky side chains, i.e. serine, leucine, isoleucine, and valine, and by the relatively high representation of proline. The H-fibroins of the *Bombyx* and *Antheraea* species are dominated by glycine and alanine, are fairly rich in serine, and from the H-fibroins of pyraloids are distinguished by higher proportion (5.3%) of tyrosine. The ratio of glycine to alanine is about 3:2 in *Bombyx* and reversed in *Antheraea*.

The structure, length, and hierarchical arrangement of the repeated peptidic motifs are very different in the three H-fibroin types. In the pyralid moth *G. mellonella*, the repetitive region is made up of 3 kinds of long repeats that are extremely conserved in both length and composition. These repeats, which were designated **A** (63 residues), $\mathbf{B}_1$ (43 residues), and $\mathbf{B}_2$ (18 residues) are assembled into higher order repetitions $\mathbf{AB}_1$ and $\mathbf{AB}_2$, and these into eleven large domains $[\mathbf{AB}_1\mathbf{AB}_1\mathbf{AB}_1\mathbf{AB}_2(\mathbf{AB}_2)\mathbf{AB}_2]_{11}$ that build up more than 95% of the H-fibroin molecule (Fig. 6).

The "crystalline" domains of *B. mori* H-fibroin are largely built up from Gly-X dipeptides in which X is Ala in 65% of cases, Ser in 23%, and Tyr in 9% (Zhou *et al.*, 2001). Each domain is composed of subdomains containing reiterations of GAGAGS motif (Mita *et al.*, 1994) and terminated with GAAS. Two to eight subdomains are assembled into a domain that is flanked by highly conserved "non-crystaline" sequence of 42-44 amino acids. While the length of the "crystalline" domains varies from 147 to 596 residues, the structure of the "non-crystalline" spacer sequence is highly conserved both at the protein and DNA levels (Zhou *et al.*, 2000).

The basic units of the H-fibroin of *Antheraea* silkmoths are 21 to 40 residues long. Each consists of a track of 4-14 alanines followed by one of 6 different types of non-crystalline motifs that are rich in G, Y, and S (Sezutsu and Yukuhiro, 2000). Four types of these units are reiterated in each of the 12 regular internal domains and in one (*A. pernyi*) or three (*A. yamamai*) incomplete terminal domains (Fig. 7). The total number of repeated units is 78 in *A. pernyi* and 77 (including 2 that lack the polyalanine block) in *A. yamamai*. The reiterated units are diversified between the two species and their arrangements into domains are also different. The 4th unit from the N-terminus, and the 4th unit from the C-terminus of the *A. yamamai* H-fibroin has

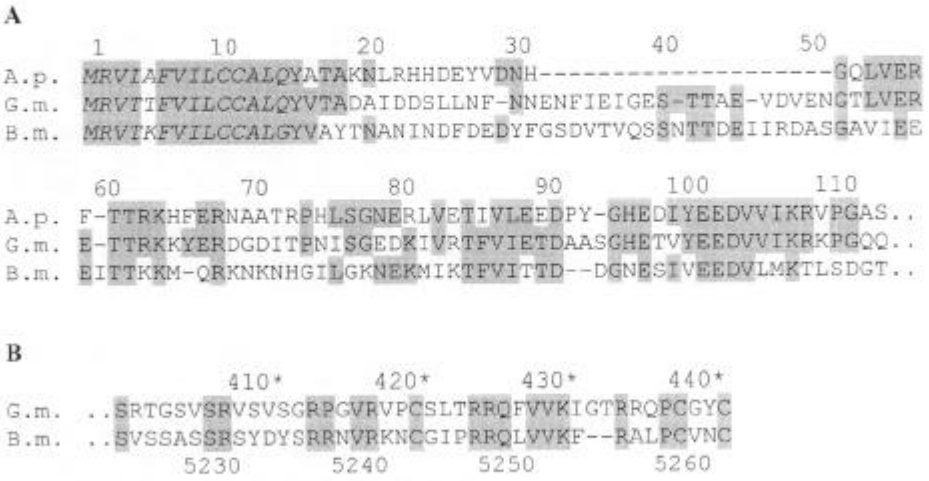


Fig. 5. The non-repetitive terminal sequences of H-fibroins. (A) The N-termini of the H-fibroins of *A. pernyi* (A. p.), *G. mellonella* (G. m.), and *B. mori* (B. m.). Signal sequence, encoded by the first exon, is typed in italics. (B) Extreme C-termini of *G. mellonella* (numbering according to Žurovec and Sehnal, unpublished) and *B. mori* (B. m., numbered according to Zhou *et al.*, 2000).

lost the polyalanine block. In contrast to the repeats, the unique N-terminal sequence, and the two atypical motifs with the polyalanine blocks at the C-terminus are highly conserved (Fig. 7).

The presence and composition of repeated motifs in the H-fibroins is consistent with the detection of crystallites in lepidopteran silk. The measurements of X-ray diffraction revealed that the silks contain peptidic chains aligned into pleated β-sheets which are stacked into crystallites (Fig. 8). It was found that the silk of *B. mori* occurs in at least three crystalline conformations (Valuzzi and Gido, 1997). That with monoclinic unit cells based of the protein chain folding into parallel zigzag β-pleated sheets is regarded as a standard conformation of lepidopteran silk (Warwicker, 1960). On the basis of comparative studies, the silk of Lepidoptera was divided into four classes (Lucas and Rudall, 1968). The silk of *B. mori* was classified as group I, which is characterized by dense molecule packing (intersheet distance 9.3 Å) in which glycine is involved (Marsh *et al.*, 1955a). The silks of *A. pernyi* and *G. mellonella* were allotted to group beta 3 with 10.6 Å intersheet packing that is consistent with high alanine content (Marsh *et al.*, 1955b). The structures of H-fibroins in *Bombyx* (Fig. 6) and *Antheraea* (Fig. 7) are fully consistent with this crystallographic classification. The crystalline core of H-fibroin is dominated in *B. mori* by the GX repeats (Zhou *et al.*, 2000) and in *A. pernyi* it contains polyalanine blocks (Sezutsu and Yukuhiro, 2000). Crystallites similar to those made of by the alanine stretches could be formed in *G. mellonella* H-fibroin by the SSAASAA (AA) repeats (Fig. 6), indicating that different peptidic motifs assume similar crystallite conformations. On the other hand, rapid evolutionary changes of H-fibroin composition led to divergence of the X-ray pattern within superfamilies, such as Bombycoidea (*Bombyx* versus *Antheraea*), and even within families. For example, in the family Notodontidae the silk of *Anaphe molneyi* was classified as group 2a and that of *Thaumatopoea pityocampa* as group 4 (Lucas and Rudall, 1968).

SILK FIBER COATING

Sericins, which are produced exclusively in MSG, are a family of at least six glycoproteins functioning to glue the

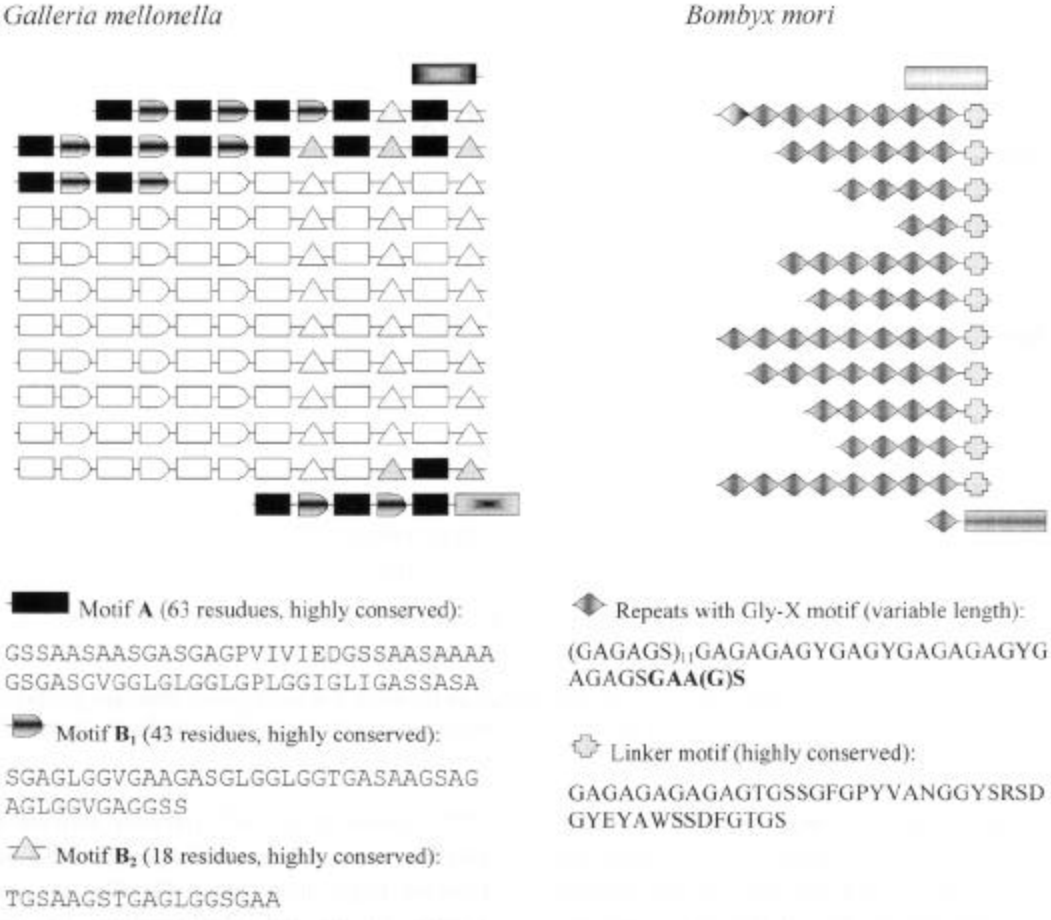
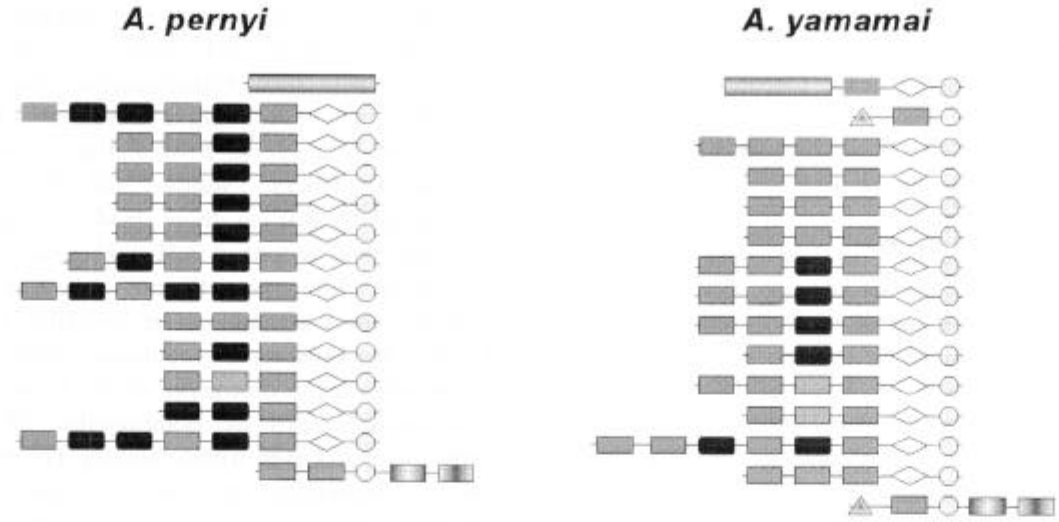


Fig. 6. Arrangements of the repetitive H-fibroin domains in *G. mellonella* (based on GenBank records AF095239 and AF095240) and *B. mori* (Zhou *et al.*, 2001). Only the 5' and 3' parts of *G. mellonella* gene were sequenced; the proposed internal structure (empty symbols) is based on restriction analysis.



N-terminal non-repetitive region

Ap MRV**I**AFVILCCALQ
YATA**K**NL**R**HHDEYVDNHG**Q**LVERFTTRKH**F**ERNAATRPHLSGNERLVETIVLEE
DPYGHEDIYEEDVVINRVPGASSSAAAASSASAGSGQTIIVERQASHGAGGAAG
AAAGAAAGSSAR**RGGG**FYETH**N**SYSSYGSGSSS

Ay MRV**T**AFVILCCALQ
YATA**N**NL**H**HHDEYVDNHG**G**LVERFTTRKH**Y**ERNAATRPHLSGNERLVETIVLEE
DPYGHEDIYEEDVVINRVPGASSSAAAASSASAGSGQTIIVERQASHGAGGAAG
AAAGAAAGSSAR**GGSG**FYETH**D**SYSSYGSGSSS

Central repetitive region

A. pernyi

AAAAAAAAAAAAAGSGAGGS(A,V,R)GGGYGWGDGGYSDS
AAAAAAAAAAAAASGAGGA..(GGY)$_n$GGYGSDS
AAAAAAAAAAAAAGSGAGGRGD.....GGYGSGSS
AAAAAAAAAAAAA.RRGHDRAAGS

A. yamamai

AAAAAAAAAAA(AA) GSGAGG(R) S(R,V)AGGYGGGDGGYGSGS(S)
AAAAAAAAAAA(AA)SGSGGSGGYGGYGSDS
AAAAAAAAAAAA(A)GSGAGG(VGGGYG)RGDG(S)GYGSGSS
AAAAAAAAAAAA(A)RRAGHDRAAGS
SGGGYSWDYSSYGSES

C-terminal non-repetitive motifs

Ap AAAAAAAAAAGAGASR**Q**VGIYGTDDGFVLDGGYDSEGS
Ay AAAAAAAAAAGAGASR**P**VGIYGTDDGFVLDGGYDSEGS

Ap AAAAAAAAAAAASSSGRSTEGHPLLSICCRPCSH**S**HSYEASRISVH
Ay AAAAAAAAAAAASSSGRSTEGHPLLSICCRPCSH**R**HSYEASRISVH

Fig. 7. Sequences of the non-repetitive N- and C-termini (residues by which the two species differ are printed bold) and arrangements of the repetitive domains in the H-fibroins of *A. pernyi* (sequence taken from GenBank record AF364332) and *A. yamamai* (GenBank AF325500).

fibroin threads. The sericins of *B. mori* range in size from 65 to 400 kDa (Sinohara, 1979; Gamo, 1982), account for 50% of the protein synthesized in the MSG during the last larval instar, and represent 20-30% of the cocoon protein (Suzuki, 1977). High serine content (16-42%) and large proportion of other hydrophilic amino acids (Komatsu, 1975; Gamo *et al.*, 1977) render sericins soluble in hot alkaline water. It is generally assumed that most of the silk proteins larger than about 40 kDa are sericins, but the identity of the individual proteins has not been fully elucidated. The multitude of sericins is due to alternative splicing (Michaille *et al.*, 1986; Tripoulas and Samols,

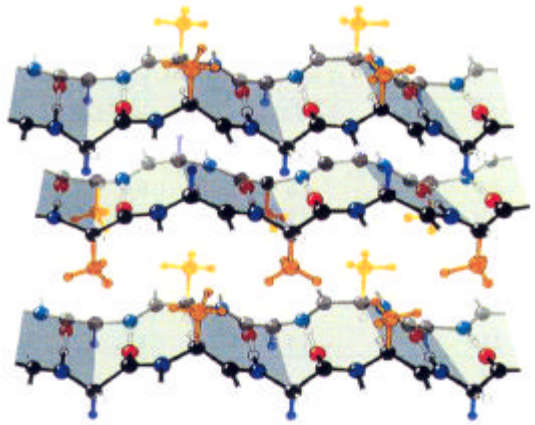


Fig. 8. Arrangement of the reiterated GAGAS motifs in *B. mori* H-fibroin in pleated β-sheets and their stacking into a crystallite (the upper layer is held together by interactions between hydrogens of the glycines and the lower layer by the methyl side chains of alanines).

1986; Couble *et al.*, 1987; Hamada *et al.*, 1987). Frequent allelic polymorphisms of sericins, and possibly also different degree of glycosylation, add to the intricacies of sericin protein analysis. It is very likely that all the transcripts are products of just two genes, *sericin 1* (*Ser1*) and *sericin 2* (*Ser2*) that were identified genetically, isolated and analyzed. It was shown that their expression and splicing in MSG is topologicaly and temporaly regulated, providing for sequential layering of different sericins around the silk thread (Couble *et al.*, 1987).

The *Ser1* gene, which corresponds to the genetically localized *scr* locus (Doira, 1983), was cloned and partially sequenced by Okamoto *et al.* (1982), Michaille *et al.* (1986) and Garel *et al.* (1997). This single-copy gene of 23 kb contains 9 exons and its upstream region exhibits similarities with the *H-fibroin* gene, for example in location of the CAAT box (Okamoto *et al.*, 1982). Four distinct mature mRNAs (2.8/2.9 kb, 4.0 kb, 9.0 kb and 10.5 kb) that have identical 5' and 3' termini originate by alternative splicing from the primary transcript of *Ser1* (Michaille *et al.*, 1986). The established coding sequence revealed that exons 1, 2, 7, 8 and 9 are shared by all mRNAs while the exons 3, 4, 5 and 6 are used selectively. The *Ser1C* mRNA includes all exons (Garel *et al.*, 1997). The largest exon 6, which spreads over about 6 kb, consists of tandem arrays of 114-bp that encode iterated peptidic sequence SRTSGGTSTYGYSSSHRGGSVSSTGSSSNTDSSTKNAG. The alternative exons 3, 4 and 5, as well as the terminal exons 1, 2 and 9 do not contain repeats.

The *Ser2* gene, corresponding to *ser2* genetic locus (Gamo, 1982), is highly polymorphic, and 5 haplotypes were found among two European strains of *B. mori* (Michaille *et al.*, 1990a). *Ser2* includes 4 exons and is typically about 12 kb long. One of the identified alleles, however, contains a retrotransposon of 4.4 kb in an intron and the gene occupies more than 16 kb. Two mRNAs, one 3.1 kb and the other 5.0, 5.4 or 6.4 kb, depending on the allele, are generated by differential splicing (Michaille *et al.*, 1990b). The N-terminal half of the gene includes repeats of 45 bp that are translated into iterated peptidic motif (alternative residues are shown in parentheses) RS(P/Q)S(H/D/Y)(K/R/T)DTEK(V/A)KPN(D/G). The C-terminal part of the gene is non-repetitive, contains about 18% serine and a large proportion of hydrophilic amino acids.

Two sericin cDNAs, *MG-1* and *MG2*, were cloned and partially sequenced in *G. mellonella* (Zurovec *et al.*, 1992). Their products do not show any obvious similarity with the *B. mori* sericins and the ratios of the dominant amino acids in the known sequences are remarkably different (Table 1). Unlike the *sericin* genes of *B. mori*, the *MG-1* and *MG2* genes are similar and probably originated from a common ancestral gene. The *MG-1* gene yields four major mRNAs (1.9 kb, 4.2 kb, 7 kb and ca 10 kb) and *MG2* two mRNAs (3.4 kb and 5.2 kb). Their sizes (Table 1) and developmental expression profiles are remarkably similar to those of the *Ser1* and *Ser2* genes of *B. mori* (Yang *et al.*, 1995b). However, the evolutionary relationship and homology between these genes and the *MG-1* and *MG2* genes remains to be found.

THE NON-STRUCTURAL SILK COMPONENTS

Silk analyses by electrophoresis disclosed presence of several minor polypeptides (Sasaki and Noda, 1973; Shimura *et al.*, 1976). The cocoons of Lepidoptera are not attacked by fungi, bacteria, mites and insects (Akai, 1997). Pupae inside the cocoons rarely succumb to disease or fall victim to predation and often develop in the safety of the cocoon for many months. We thus assume that some of the small proteins possess antimicrobial properties and others deter predators from the feeding on cocoons and the pupae inside. In search for such agents, we subjected to N-terminal sequencing some of the small silk proteins previously identified in the silkworm and the waxmoth (Kodrík, 1992). The results revealed that silk of the two species contained several components that were apparently homologous and these were selected for further work. RT-PCR was employed to search for the corresponding mRNAs in the silk glands. This approach led to the discovery of two distinct classes of non-structural silk proteins, the seroins and the protease inhibitors.

Seroins are present in the silk as several proteins of different size. In *G. mellonella* they are derived from a single gene that is expressed both in the posterior and the middle silk gland sections (Zurovec *et al.*, 1998b). The

Table 1. Approximate sizes of *sericin* mRNAs[a] and representation of dominant amino acids in the sericins[b] of *B. mori* (B. m.), *G. mellonella* (G. m.) and *A. yamamai* (A. y.)

	B. m. Ser1	B. m. Ser2	G. m. MG1	G. m. MG2	A. y. Ser2
mRNA	2.8/2.9; 4.0; 9.0; 10.5 kb	3.1; 5.0/5.4 kb	1.9; 3.2; 4.2; >10 kb	3.4; 5.2 kb	Not established
Ser	38	18	52	56.5	38
Glu	1.6	15	0.2	1.6	6.4
Asp	6	9	0.9	0.9	4.7
Lys	2.1	14	0.2	0.3	3.4
Arg	4.7	7	–	0.9	2.1
Asn	8.6	5.6	7.9	15.0	10.2
Gly	13.4	6	22.6	8.6	8.1

[a] The sizes of mRNAs in kb as published by Michaille *et al.* (1986; 1990b) for *B. mori* and by Žurovec *et al.* (1992) for *G. mellonella*. Two values separated by slash indicate presence of two alleles with different transcript sizes.

[b] Representation of individual residues is expressed in % of the total. Data on *B. mori* Ser1 concern Ser1C and were taken from Garel *et al.* (1997). The values for *B. mori* Ser2 were calculated from the C-teminal part of the *Ser2* gene sequence obtained from the SilkBase (Mita and Kanda, unpublished), those for *G. mellonella* from the summation of deduced amino acid sequences of the N- and C-termini of *MG1* and *MG2*, respectively (Žurovec and Sulitková, unpublished), and the data for *A. yamamai* from the C-terminus of *A. yamamai Ser2* gene (Fedič, unpublished).

putative translation product consists of 167 amino acid residues of which 17 constitute the signal peptide (Fig. 9A). The secreted product is N-glycosylated to 22.5 and 23 kDa proteins that are deposited in the silk spun by early instar larvae as well as in the cocoon silk. *G. mellonella* seroin contains proline-rich repeats (Žurovec *et al.*, 1998b) and by this feature it resembles one class of antibacterial peptides (cf. Hoffmann *et al.*, 1994). Two seroins (9.9 and 10.3 kDa), which have the N-terminal sequences and the spacing of a few amino acids at the C-termini similar to the seroin of *G. mellonella*, occur in the cocoon silk of *B. mori* (Nirmala *et al.*, 2001b). They are products of two separate genes whose expression seems to be limited to the last larval instar. The genes are expressed in the middle, and to a small extent also in the posterior silk gland sections. The *seroin 1* and less conspicuously the *seroin 2* mRNAs accumulate in course of the last larval instar and reach maxima in the silk glands of post-spinning larvae.

The cocoon silks of *B. mori* and *G. mellonella* further contain two silk protease inhibitors (SPI 1 and SPI 2) that inhibit bacterial and fungal proteinases (subtilisin, proteinase K and pronase) and may therefore provide protection against microbial degradation (Nirmala *et al.*, 2001a, b). The *SPI* genes are expressed in the middle silk gland section (*BmSPI 1* partly also in the posterior section) in the last larval instar prior to cocoon spinning. BmSPI 1 and GmSPI 1 are typical single-domain Kunitz-type inhibitors of about 6 kDa (Fig. 9B); their mutual homology (55%) is similar to the homology found among various Kunitz-type inhibitors that occur in insect tissues. BmSPI 2 (4.7 kDa) and GmSPI 2 (4.0 kDa) are single-domain Kazal-type inhibitors that are smaller than typical inhibitors of

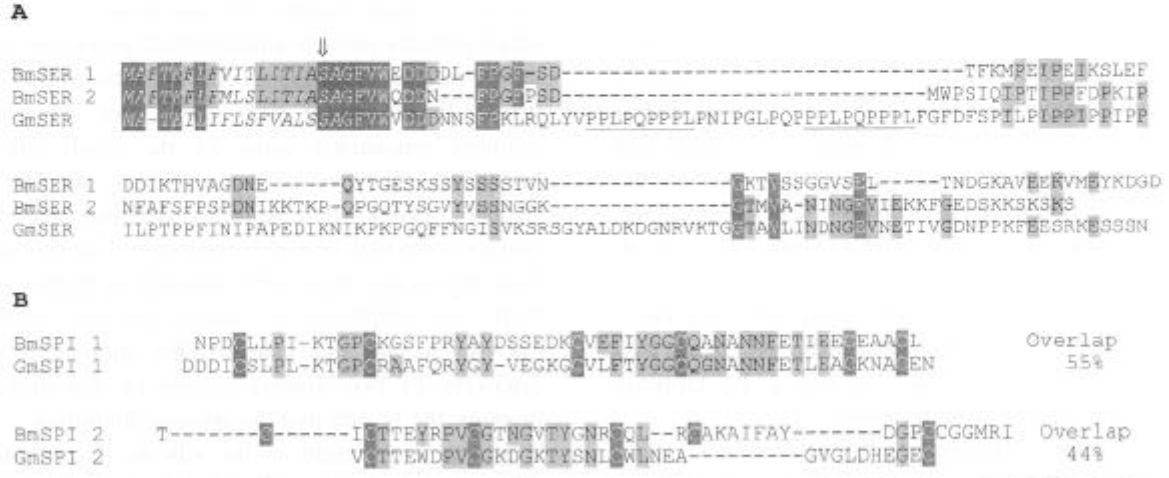


Fig. 9. Amino acid sequence of translated seroins (A) and secreted protease inhibitors (B). Arrow in A indicates the site of signal peptide cleavage. Underlines indicate proline-rich repeats in *G. mellonella* seroin. Dashes in SPI 2 mark positions of additional amino acid residues that are present in typical protease inhibitors.

this type. BmSPI 2 consists of 44 amino acid residues, of which 10 form the chain between CysV and CysVI, whereas typical Kazal inhibitors have 17 residues between these cysteines. GmSPI 2 is exceptionally small (36 amino acid residues), lacks CysI and CysV and therefore also the disulfide bond these cysteines would make, and the central α-helix segment in presumed peptide conformation is shorter than in typical Kazal inhibitors (Nirmala *et al.*, 2001a). The sequence homology between BmSPI 2 and GmSPI 2 is only 44% and the homology to other Kazal inhibitors is still lower.

ACKNOWLEDGMENTS

This review was prepared in frame of grant 204/00/0019 of the Grant Agency of the Czech Republic, and grant ME 204 of the Ministry of Education, Youth, and Sports of the Czech Republic.

REFERENCES

Adams, D.S., Eickbush, T.H., Herrera, R.J., and Lizardi, P.M. (1986) A highly reiterated family of transcribed oligo (A)-teminated interspersed DNA elements in the genome of *Bombyx mori*. *J. Mol. Biol.* **187**, 465-478.

Akai, H. (1965) Studies on the ultrastructure of the silk gland in the silkworm, *Bombyx mori* L. *Bull. Sericult. Exp. Sta.* **19**, 375-484.

Akai, H. (1984) The ultrastructure and functions of the silk gland cells of *Bombyx mori*. In *Insect Ultrastructure* (R. C. King and H. Akai, eds.), vol. 2. pp. 323-364. Plenum, New York.

Akai, H. (1997) Anti-bacteria function of natural silk materials. *Int J Wild Silkmoths Silk* **3**, 79-81.

Akai, H., Hakim, R.S., and Kristensen, N.P. (in press) Exocrine glands: saliva and silk. In *Lepidoptera: Moths and Butterflies 2* (N.P. Kristensen, ed.) Handbuch der Zoologie IV, vol. 36. Walther de Gruyter, Berlin.

Chinzei, Y. (1975a) Induction of histolysis by ecdysterone in vitro: Breakdown of anterior silk gland in silkworm, *Bombyx mori* (Lepidoptera: Bombycidae). *Appl. Entomol. Zool.* **10**, 136-138.

Chinzei, Y. (1975b) Biochemical evidence of DNA transport from the silk gland to the fat body of the silkworm, *Bombyx mori*. *J. Insect Physiol.* **21**, 163-171.

Chevillard, M., Couble, P., and Prudhomme, J.C. (1986a) Complete nucleotide sequence of the gene encoding the *Bombyx mori* silk protein P25 and predicted amino acid sequence of the protein. *Nucleic Acids Res.* **14**, 6341-6342.

Chevillard, M., Deleage, G., and Couble, P. (1986b) Amino acid sequence and putative conformational characteristics of the 25 kD silk protein of *Bombyx mori*. *Sericologia* **26**, 435-449.

Couble, P., Chevillard, M., Moine, A., Ravel-Chapuis, P., and Prudhomme, J.C. (1985) Structural organization of the P25 gene of *Bombyx mori* and comparative analysis of its 5' flanking DNA with that of the fibroin gene. *Nucleic Acids Res.* **13**, 1801-1814.

Couble, P. Michaille, J.J., Couble, M.L., and Prudhomme, J.C. (1987) Developmental switches of sericin mRNA splicing in individual cells of *Bombyx mori* silkgland. *Dev. Biol.* **124**, 431-440.

Couble, P., Moine, A., Garel, A., and Prudhomme, J.C. (1983) Developmental variations of a nonfibroin mRNA of *Bombyx mori* silkgland, encoding for a low-molecular-weight silk protein. *Dev. Biol.* **97**, 398-407.

Craig, C.L. (1997) Evolution of arthropod silks. *Annu. Rev. Entomol.* **42**, 231-267.

Datta A., Ghosh, A.K., and Kundu, S.C. (2001) Purification and characterization of fibroin from the tropical saturniid silkworm, *Antheraea myllita*. *Insect Biochem. Mol. Biol.* **31**, 1013-1018.

Denny, M.W. (1980) Silks - their properties and functions. In *The Mechanical Properties of Biological Materials*. Symposia of Society for Experimental Biology vol. 34, pp. 247-272. Cambridge University Press, Cambridge.

Dover, G. (1982) Molecular drive: A cohesive mode of species evolution. *Nature* **299**, 111-117.

Doira, H. (1983) Linkage map of *Bombyx mori* - Status quo in 1983. *Sericologia* **23**, 245-269.

Gamo, T. (1982) Genetics variants of the *Bombyx mori* silkworm encoding sericin proteins of different length. *Biochem. Genet.* **20**, 165-177.

Gamo, T., Inokuchi, T., and Laufer, H. (1977) Polypeptides of fibroin and sericin secreted from the different sections of the silk gland in *Bombyx mori*. *Insect Biochem.* **7**, 285-295.

Garel, A., Deleage, G., and Prudhomme, J.C. (1997) Structure and organization of the *Bombyx mori* Sericin 1 gene and of the Sericin 1 deduced from the sequence of the Ser 1B cDNA. *Insect. Biochem. Mol. Biol.* **27**, 469-477.

Garel, J.P., Mandel, P., Chavancy, G., and Daillie, J. (1970) Functional adaptation of tRNAs to fibroin biosynthesis in the silkgland of *Bombyx mori* L. *FEBS Letters* **7**, 327-329.

Goldsmith, M.R., and Kafatos, F.C. (1984) Developmentally regulated genes in silkmoths. *Annu. Rev. Genet.* **18**, 443-487.

Hamada, Y., Yamashita, O., and Suzuki, Y. (1987) Haemolymph control of sericin gene expression studied by organ transplantion. *Cell Differ.* **20**, 65-76.

Hui, C.C., and Suzuki, Y. (1995) Regulation of the silk protein genes and the homeobox genes in silk gland development. In *Molecular Model Systems in the Lepidoptera* (M.R. Goldsmith and A.S. Wilkins, eds.), pp. 249-271. Cambridge University Press, Cambridge.

Hui, C.C., Suzuki, Y., Kikuchi Y., and Mizuno, S. (1990) Homeodomain binding sites in the 5' flanking region of the *Bombyx mori* silk fibroin light-chain gene. *J. Mol. Biol.* **213**, 395-398.

Inoue, S., Tanaka, K., Arisaka, F., Kimura, S., Ohtomo, K., and Mizuno, S. (2000) Silk fibroin of *Bombyx mori* is secreted, assembling a high molecular mass elementary unit consisting of H-chain, L-chain, and P25, with a 6:6:1 molar ratio. *J. Biol. Chem.* **275**, 40517-40528.

Jindra, M., and Sehnal, F. (1989) Larval growth, food consumption, and utilization of dietary protein and energy in *Galleria mellonella*. *J. Insect Physiol.* **35**, 719-724.

Kikuchi, Y., Mori, K., Suzuki, S., Yamaguchi, K., Mizuno, S. (1992) Structure of the *Bombyx mori* fibroin light-chain-encoding gene: upstream sequence elements common to the light and heavy chain. *Gene* **110**, 151-158.

Kimura, K., Oyama, F., Ueda, H., Mizuno, S., and Shimura, K. (1985) Molecular cloning of the fibroin light chain complementary DNA and its use in the study of the expression of the light chain gene in the posterior silk gland

of *Bombyx mori*. *Experientia* **41**, 1167-1171.

Kodrík, D. (1992) Small protein components of the cocoons in *Galleria mellonella* (Lepidoptera, Pyralidae) and *Bombyx mori* (Lepidoptera, Bombycidae). *Acta Ent. Bohemoslov.* **89**, 269-273.

Komatsu, K. (1975) Studies on dissolution behaviors and structural characteristics of silk sericin. *Bull. Seric. Exp. Stn.* **26**, 135-256.

Kristensen, N.P. (1999) *Lepidoptera, Moths and Butterflies 1* Handbuch der Zoologie IV, vol. 35, 491 pp. Walther de Gruyter, Berlin.

Kusuda, J., Tazima, Y., Onimaru, K., Ninaki, O., and Suzuki, Y. (1986) The sequence around the 5' end of the fibroin gene from the wild silkworm, *Bombyx mandarina*, and comparison with that of domesticated species, *B. mori*. *Mol. Gen. Genet.* **203**, 359-364

Lazaris, A., Arcidiacono S., Huang, Y., Zhou J.-F., Duguay, F., Chretien, N., Welsh, E.A., Soares, J.W., and Karatzas, C.N. (2002) Spider silk fiber spun from soluble recombinant silk produced in mammalian cells. *Science* **295**, 472-476.

Lizardi, P.M., and Brown, D.D. (1975) The length of the fibroin gene in the *Bombyx mori* genome. *Cell* **4**, 207-15.

Lucas, F., and Rudall, K.M. (1968) Extracellular fibrous proteins: The silks. In *Comprehensive Biochemistry* (M. Florkin and E.H. Stota, eds.), vol. **26B**, pp. 475-558. Elsevier, Amsterdam.

Marsh, R.E., Corey, R.B., and Pauling, L. (1955a) An investigation of the structure of silk fibroin. *Biochem. Biophys. Acta* **16**, 1-34.

Marsh, R.E., Corey, R.B., and Pauling, L. (1955b) Structure of Tussah silk fibroin. *Acta Cryst.* **8**, 710-715.

Michaille, J.J., Couble, P., Prudhomme, J.C., and Garel, A. (1986) A single gene produces multiple sericin messenger RNAs in the silkgland of *Bombyx mori*. Biochimie **68**, 1165-1173.

Michaille, J.J., Garel, A., and Prudhomme, J.C. (1990a) Expression of Ser1 and Ser2 genes in the middle silk gland of *Bombyx mori* during the fifth instar. *Sericologia* **30**, 49-60.

Michaille, J.J., Garel, A., and Prudhomme, J.C. (1990b) Cloning and characterisation of the highly polymorphic Ser2 gene of *Bombyx mori*. *Gene* **86**, 177-184.

Mita, K., Ichimura, S., Zama, M., and James, T.C. (1988) Specific codon usage pattern and its implications on the secondary structure of silk fibroin mRNA. *J. Mol. Biol.* **203**, 917-925.

Mita, K., Ichimura, S., and James T.C. (1994) Highly repetitive structure and its organization of the silk fibroin gene. *J. Mol. Evol.* **38**, 583-592.

Nirmala, X., Kodrík, D., Žurovec, M., and Sehnal, F. (2001a) Insect silk contains both a Kunitz-type and a unique Kazal-type proteinase inhibitor. *Eur. J. Biochem.* **268**, 2064-2073.

Nirmala, X., Mita, K., Vanisree, V., Žurovec, M., and Sehnal, F. (2001b) Identification of four small molecular mass proteins in the silk of *Bombyx mori*. *Insect Mol. Biol.* **10**, 437-445.

Nony, P., Prudhomme, J.C., and Couble, P. (1995) Regulation of the P25 gene transcription in the silk gland of *Bombyx*. *Biol. Cell* **84**, 43-52.

Nunome, J. (1937) The silk gland development in *Bombyx mori*. *Bull. Appl. Zool.* **9**, 68-92.

Obara, T., and Suzuki, Y. (1988) Temporal and spatial control of silk gene transcription analyzed by nucelar run-on assay. *Dev. Biol.* **127**, 384-391.

Ohta, S., Suzuki, Y., Hara, W., Takiya, S., and Suzuki, T. (1988) Fibroin gene transcription in the embryonic stages of the silkworm, *Bombyx mori*. *Develop. Growth Differ.* **30**, 293-299.

Okamoto, H., Ishikawa, E., and Suzuki, Y. (1982) Structural analysis of sericin genes. Homologies with fibroin gene in the 5' flanking nucleotide sequences. *J. Biol. Chem.* **257**, 15192-15199.

Pauling, L., and Corey, R.B. (1953) Two rippled-sheet configurations of polypeptide chains, and a note about the pleated sheets. *Proc. Natl. Acad. Sci. USA*, **39**, 253-256.

Peigler, R.S. (1993) Wild silks of the world. *Am. Entomol.* **39**, 151-161.

Prudhomme, J.C., Couble, P., Garel, J.P., and Daillie, J. (1985) Silk synthesis. In *Comprehensive Insect Physiology, Biochemistry and Pharmacology* (G.A. Kerkut and L.I. Gilbert, eds.), vol. **10**, pp. 571-594. Pergamon Press, Oxford.

Sasaki, T., and Noda, H. (1973) Studies on silk fibroin of *Bombyx mori* directly extracted from the silk gland. I. Molecular weight determination in guanidine hydrochloride or urea solutions. *Biochim. Biophys. Acta* **310**, 91-103.

Sehnal, F., and Akai, H. (1990) Insects silk glands: their types, development and function, and effects of environmental factors and morphogenetic hormones on them. *Int. J. Insect Morphol. Embryol.* **19**, 79-132.

Sehnal, F., Janda, V., and Němec, V. (1983) Composition, synthetic and cytolytic activities of *Galleria mellonella* silk glands during the last larval instar under the action of juvenile hormone. *J. Insect Physiol.* **29**, 237-248.

Sezutsu, H., and Yukuhiro, K. (2000) Dynamic rearrangement within the *Antheraea pernyi* silk fibroin gene is associated with four types of repetitive units. *J. Mol. Evol.* **51**, 329-338.

Sinohara, H. (1979) Glycopeptides isolated from sericin of the silkworm, *Bombyx mori*. *Comp. Biochem. Physiol.* **63B**, 87-91.

Shimura, K., Kikuchi, A., Ohtomo, K., Katagata, Y., and Hyodo, A. (1976) Studies on silk fibroin of *Bombyx mori*. I. Fractionation of fibroin prepared from the posterior silk gland. *J. Biochem.* **80**, 693-702.

Smith, G.P. (1976) Evolution of repeated DNA sequences by unequal crossover. *Science* **191**, 528-535.

Suzuki, Y. (1977) Differentiation of silk gland. A model system for the study of differential gene action. In *Results and Problems in Cell Diferentiation* (W. Beermann, ed.), pp. 1-44. Springer-Verlag, Berlin/Heidelberg.

Suzuki, Y., and Brown, D.D. (1972) Isolation and identification of the messenger RNA for silk fibroin from *Bombyx mori*. *J. Mol. Biol.* **63**, 409-429.

Tamura, T., Inoue, H., and Suzuki, Y. (1987) The fibroin genes of *Antheraea yamamai* and *Bombyx mori* are different in their core regions but reveal a striking sequences similarity in their 5' ends and 5' flanking regions. *Mol. Gen. Genet.* **206**, 189-195.

Tamura, T., and Kubota, T. (1989) A determination of molecular weight of fibroin polypeptides in the saturniid silkworms, *Antheraea yamamai, Antheraea pernyi*, and *Philosamia cynthia ricini* by SAD PAGE. In *Wild Silkmoths 88* (H. Akai and M. Kiguchi, eds.), pp. 67-72. Natl. Inst. Sericult. Insect Sci., Tsukuba.

Tanaka, K., Mori, K., and Mizuno, S. (1993) Immunological identification of the major disulfide-linked light component of silk fibroin. *J. Biochem.* **114**, 1-4.

Tanaka, K., Kajiyama, N., Ishikura, K., Waga, S., Kikuchi, A., Ohtomo, K., Takagi, T., and Mizuno, S. (1999) Determina-

tion of the site of disulfide linkage between heavy and light chains of silk fibroin produced by *Bombyx mori*. *Biochim. Biophys. Acta* **1432**, 92-103.

Tanaka, K., and Mizuno, S. (2001) Homologues of fibroin L-chain and P25 of *Bombyx mori* are present in *Dendrolimus spectabilis* and *Papilio xuthus* but not detectable in *Antheraea yamamai*. *Insect Biochem. Mol. Biol.* **31**, 665-677.

Takei, F., Kikuchi, Y., Kikuchi, A., Mizuno, S., and Shimura, K. (1987) Further evidence for importance of the subunit combination of silk fibroin in its efficient secretion from the posterior silk gland cells. *J. Cell Biol.* **105**, 175-180.

Tripoulas, N.A., and Samols, D. (1986) Developmental and hormonal regulation of sericin mRNA in the silkworm, *Bombyx mori*. *Dev. Biol.* **116**, 328-336.

Valluzi, R., and Gido, S.P. (1997) The crystal structure of *Bombyx mori* silk fibroin at the air-water interface. *Biopolymers* **42**, 705-717.

Wilson, E.T., Condliffe, D.P., and Sprague, K.U. (1988) Transcriptional properties of BmX, a moderately repetitive silkworm gene that is an RNA polymerase III template. *Mol. Cell. Biol.* **8**, 624-631.

Warwicker, J.O. (1960) Comparative studies of fibroins. II. The crystal structures of various fibroins. *J. Mol. Biol.* **2**, 350-362.

Yang, C., Krishnan, M., Žurovec, M., Sehnal, F., and Scheller, K. (1995) Correlation betwen ecdysteroid titre and silk gene expression in the last instar larvae of *Galleria mellonella* (Lepidoptera: Pyralidae). *Eur. J. Entomol.* **92**, 229-234.

Yang, C., Teng, X., Žurovec, M., Scheller, K., and Sehnal, F. (1998) Characterization of the P25 silk gene and associated insertion elements in *Galleria mellonella*. *Gene* **209**, 157-165.

Yamaguchi, K., Kikuchi, Y., Takagi, T., Kikuchi, A., Oyama, F., Shimura, K., and Mizuno, S. (1989) Primary structure of the silk fibroin light chain determined by cDNA sequencing and peptide analysis. *J. Mol. Biol.* **210**, 127-139.

Yukuhiro, K., Kanda, T., and Tamura, T. (1997) Preferential codon usage and two types of repetitive motifs in the fibroin gene of the Chinese oak silkworm, *Antheraea pernyi*. *Insect. Mol. Biol.* **6**, 89-95.

Zhou, C.Z., Confalonieri, F., Medina, N., Zivanovic, Y., Esnault, C., Yang, T., Jacquet, M., Janin, J., Duguet, M., and Perasso, R. (2000) Fine organization of *Bombyx mori* fibroin heavy chain gene. *Nucleic Acid Res.* **28**, 2413-2419.

Zhou, C.Z., Confalonieri, F., Jacquet, M., Perasso, R., Li, Z.G., and Janin, J. (2001) Silk fibroin: Structural implications of a remarkable amino acid sequence. *Proteins Struct. Func. Genet.* **448**, 119-122.

Žurovec, M., Kodrik, D., Yang, C., Sehnal, F., and Scheller, K. (1998a) The P25 component of *Galleria* silk. *Mol. Gen. Genet.* **257**, 264-270.

Žurovec, M., Sehnal, F., Scheller, K., and Kumaran, A.K. (1992) Silk gland specific cDNA from *Galleria mellonella* L. *Insect Biochem. Mol. Biol.* **22**, 55-67.

Žurovec, M., Vašková, M., Kodrík, D., Sehnal, F., and Kumaran, A.K. (1995) Light-chain fibroin of *Galleria mellonella* L. *Mol. Gen. Genet. 247*, 1-6.

Žurovec, M., Yang, C., Kodrík, D., and Sehnal, F. (1998b) Identification of a novel type of silk protein and regulation of its expression. *J. Biol. Chem.* **273**, 15423-15428.