THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 277, No. 25, Issue of June 21, pp. 22639–22647, 2002

Printed in U.S.A.

Unique Molecular Architecture of Silk Fibroin in the Waxmoth, *Galleria mellonella**

Received for publication, February 18, 2002, and in revised form, March 8, 2002
Published, JBC Papers in Press, March 8, 2002, DOI 10.1074/jbc.M201622200

Michal Žurovec and František Sehnal‡

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

Proteins of silk fibers are characterized by reiterations of amino acid repeats. Physical properties of the fiber are determined by the amino acid composition, the complexity of repetitive units, and arrangement of these units into higher order arrays. Except for very short motifs of 6–10 residues, the length of repetitive units and the number of these units concatenated in higher order assemblies vary in all spider and lepidopteran silks analyzed so far. This paper describes an exceptional silk protein represented by the 500-kDa heavy chain fibroin (H-fibroin) of the waxmoth, *Galleria mellonella*. Its non-repetitive N-terminal (175 residues) and C-terminal (60 residues) parts, the overall gene organization, and the nucleotide sequence around the TATA box show that it is homologous to the H-fibroins of other Lepidoptera. However, over 95% of the protein consists of highly ordered repetitive structures that are unmatched in other species. The repetitive region includes 11 assemblies $AB_1AB_1AB_1AB_2(AB_2)AB_2$ of remarkably conserved polypeptide repeats A (63 amino acid residues), B_1 (43 residues), and B_2 (18 residues). The repeats contain a high proportion of Gly (31.6%), Ala (23.8%), Ser (18.1%), and of residues with long hydrophobic side chains (16% for Leu, Ile, and Val combined). The presence of the GLGGLG and SSAASAA(AA) motifs suggests formation of pleated β-sheets and their stacking into crystallites. Conspicuous conservation of the apolar sequence VIVI followed by DD or ED is interpreted as indicating the importance of hydrophobicity and electrostatic charge in H-fibroin cross-linking. The environment of *G. mellonella* larvae within bee cultures requires continuous production of silk that must be both strong and elastic. The spectacular arrangement of the repetitive H-fibroin region apparently evolved to meet these requirements.

Silk is a proteinaceous polymer secreted by specialized exocrine glands in several groups of arthropods. Silk quality reached highest level of functional specialization in spiders and in lepidopteran larvae (caterpillars). The remarkable mechanical properties of the silks (1) inspired numerous studies on their composition and structural arrangement. Early amino acid analyses and x-ray diffraction studies revealed that the silks of caterpillars contain pleated β-sheets (2). The sheets formed by short iterated repeats of small amino acid residues are stacked into crystallites that reinforce the silk fiber (3). Recent data (1, 4) on the DNA sequence demonstrated the presence of short amino acid repeats in spider silks and proved that the composition of repeats determines physical properties of the silk fiber, such as strength and elasticity.

Caterpillars produce silk from a pair of labial glands, each of which consist of silk-secreting posterior and middle regions, and an outlet (5). The posterior region produces fibrous silk core, whereas the middle region provides a sticky coating of the fiber and adds several low molecular components with presumably protective functions to the silk (6, 7). The silk core is typically composed of 3 types of proteins as follows: heavy chain fibroin (H-fibroin),[1] light chain fibroin (L-fibroin), and P25 glycoprotein (8–10). For the silkworm *Bombyx mori* it was shown that the H-fibroin (~390 kDa) and L-fibroin (~30 kDa) molecules are linked by a disulfide bond, and six such heterodimers are assembled with a single P25 molecule into a complex called elementary fibroin unit (11). The core of tussah silk, which is produced by *Antheraea* silkmoths, is made up of H-fibroin homodimers (10, 12).

H-fibroin makes up the bulk of the lepidopteran silk fiber, and its structure determines the physical properties of the silk. Five structural silk categories were recognized on the basis of x-ray diffraction studies and amino acid analyses (13). The silk of *B. mori* was classified as group 1, which is characterized by dense molecule packing (intersheet distance 9.3 Å) corresponding to high glycine content (14), whereas the tussah silk of *Antheraea pernyi* was classified as group 3 with 10.6 Å intersheet packing that is consistent with high alanine content (15). These data were recently supported by full elucidation of the primary H-fibroin structures. The H-fibroin of *B. mori* was shown to consist of 12 large "crystalline" domains separated by short spacers; each crystalline domain includes a number of glycine-rich repeats dominated by the GAGAGS motif (16). By contrast, the H-fibroin of *A. pernyi* includes 80 tandemly arranged repeats, each containing a crystalline region of 5–15 alanine residues flanked by an "amorphous" motif (12).

The silk of the waxmoth, *Galleria mellonella*, a species distant in evolution from both *B. mori* and *A. pernyi*, was assigned to structural group 3 as the tussah silk (13), but partial sequence of the waxmoth H-fibroin did not disclose any regular polyalanine repeats (17). Also, the silk of *G. mellonella* has different physical properties than tussah silk. It has a tensile strength 7.5×10^8 Nm^{-2} and 101% extendibility (18), whereas

* This work was supported by Grants 204/96/1100 and 204/00/0019 from the Grant Agency of the Czech Republic and by Grant ME 204 from the Czech Ministry of Education, Youth, and Sports.The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EBI Data Bank with accession number(s) AF095239 and AF095240.

‡ To whom correspondence should be addressed: Entomological Institute, Academy of Sciences, Branišovská 31, 370 05 České Budějovice, Czech Republic. Tel.: 420-38-5300350; Fax: 420-38-5300354; E-mail: sehnal@entu.cas.cz.

[1] The abbreviations used are: H-fibroin, heavy-chain fibroin; L-fibroin, light-chain fibroin; nt, nucleotide.

the corresponding values for the tussah silk are 5.8×10^8 Nm^{-2} and 35%, respectively (19). Our goal was to clarify the structure of *G. mellonella* H-fibroin as a step toward understanding the molecular architecture that lends to this silk its unusual physical properties.

In a previous study (17), we isolated from the posterior silk gland region of *G. mellonella* larvae a cDNA clone designated PG-2. Except for a non-repetitive 0.3-kb 3′ terminus, the clone consisted of repetitive motifs. The corresponding mRNA was not detectable in any other body tissue. Its large size, high abundance in the posterior silk gland section, and the sequence of its putative translation product indicated that PG-2 was derived from the 3′ end of the *H-fibroin* gene. By using PG-2 as a probe, we have now identified genomic clones that encode considerable portions of both 3′ and 5′ ends of the gene. Sequence analysis revealed that the structure of *G. mellonella* H-fibroin is exceptionally homogenous and resembles certain spider silks more than the H-fibroins of *B. mori* and *A. pernyi*.

EXPERIMENTAL PROCEDURES

Insects and Tissue Preparation—Our stock of the waxmoth, *G. mellonella* L. (Lepidoptera, Pyralidae) was maintained at 30 °C on a semi-artificial diet (20). The posterior parts of silk glands were dissected from anesthetized last instar larvae whose age was measured from the preceding ecdysis. Dissected tissues were rapidly frozen in liquid nitrogen or on dry ice and stored at −80 °C.

Isolation of Genomic Clones and Hybridization Conditions—A genomic library was prepared from 15- to 20-kb *Sau*3A fragments of *G. mellonella* DNA (21). Fragments were ligated to *Xho*I half-site arms of the LambdaGEM™-12 vector (Promega) that was grown in the LE392 strain of *Escherichia coli*. About 4×10^4 plaques were screened with 1.2-kb cDNA probe PG-2 representing the 3′ end of the *H-fibroin* gene (17). The positive clones were re-screened with the extreme 3′ end of PG-2 (0.3 kb), which encoded non-repetitive C-terminal amino acid sequence. Hybridization with both the full and the truncated PG-2 probes was done under high stringency conditions at 65 °C in 5× SSPE, 5× Denhardt's solution, 0.5% SDS, and washing at room temperature twice in 2× SSPE, 0.1% SDS, and twice in 0.2× SSPE, 0.1% SDS. The clones, which failed to react with the short version of PG-2, were screened with the *Pst*I-*Bcl*I fragment of *B. mori H-fibroin* gene (positions −139 to +69) that contained the 5′ upstream region and the first exon of this gene (22). This hybridization was done under low stringency conditions at 55 °C in 5× SSPE, 5× Denhardt's solution, and 0.5% SDS, with two subsequent washings in 2× SSPE, 0.1% SDS at room temperature, and two washings in 1× SSC, 0.1% SDS at 55 °C. All probes were ^{32}P-labeled with the aid of the Random Primed DNA Labeling Kit (United States Biochemical Corp.).

Sequencing—Selected genomic clones were characterized with restrictases and the restriction fragments subcloned in pBluescript SK(+/−) (Stratagene). Part of the non-repetitive region was sequenced with primer walking. Single- or double-stranded templates were sequenced either manually or with capillary DNA sequencer ABI 310 (PerkinElmer Life Sciences). The dideoxy chain termination reaction with [α-^{35}S]dATP, T3 or T7 primers, and Sequenase version 2.0 DNA Sequencing Kit (Amersham Biosciences), was employed in the manual sequencing. Automatic sequencing was done with Thermosequenase Cycle Sequencing Kit (Amersham Biosciences) or ABI Dye Terminator Cycle Sequencing Kit (PerkinElmer Life Sciences). Most of the analyzed gene regions were sequenced repeatedly from both DNA strands as depicted in Fig. 1. Established sequences were analyzed with the DNAstar software (Lasergene) and with the aid of EMBL network services.

Southern and Northern Analysis—For the Southern analysis, high molecular weight genomic DNA was prepared from the newly ecdysed last instar larvae, and 5-μg DNA aliquots were digested with restrictases specified under "Results," electrophoresed on 0.8% agarose gel, and blotted on nylon membrane. Aliquots of 5 μg of total RNA, which was prepared from the posterior and middle silk gland regions and from the body carcass devoid of the silk glands, were taken for the Northern blotting. By using high stringency conditions described above, Southern blots were probed with the ^{32}P-radiolabeled 1.2-kb PG-2 probe and Northern blots with the *Xba*I-*Bam*HI fragment of the λX1 genomic clone. Relative positions of *Hind*III-digested λ phage DNA fragments were used as size markers.

Reverse Transcription-PCR—The reverse transcription step of reverse transcription-PCR was performed with 5 μg of total RNA, 5 units of avian leukemia virus reverse transcriptase, and 50 ng of primer 5′ ATCCAGATGAACCACCT 3′ (positions +2869 to +2853 of the gene, *i.e.* 4136–4110 in the GenBank™ accession number AF095239) in a 20-μl reaction mix. Subsequently, 1 μg of the reaction mixture was taken for cDNA amplification using a primer corresponding to the gene sequence between positions +9 and +28 (Fig. 3*A*), and a reverse primer corresponding to positions +2851 to +2832 (4108–4309 in the GenBank™ accession number AF095239). PCR included initial denaturation as follows: 1 min at 94 °C, 35 cycles of 20 s at 94 °C, 20 s at 55 °C, and 20 s at 72 °C, and final extension 10 min at 72 °C.

Primer Extension—The procedure was described previously (6). A 35-nt synthetic oligonucleotide complementary to the region −30 to +4 of the gene (Fig. 3*A*) was employed for primer extension analysis using total RNA from the posterior section of silk glands of well staged last instar larvae. The primer was 5′-labeled with [γ-^{32}P]ATP with the aid of T4 polynucleotide kinase and hybridized to 10 μg of total RNA for 16 h at 32 °C. The following extension reaction was carried out with murine leukemia virus reverse transcriptase (Amersham Biosciences) for 2 h at 42 °C.

RESULTS

Isolation of the Genomic Clones—Screening of our *G. mellonella* genomic library with the PG-2 cDNA clone led to isolation of nearly 60 positive recombinant phages. Most of them were likely to contain inserts from the central gene region with repeated sequences that are difficult to clone and analyze. We therefore decided to select clones extending to the 5′ or the 3′ regions of the *H-fibroin* gene. Eight clones of the 3′ region were obtained using the unique 0.3-kb 3′ terminus of PG-2 as probe. Clone λXF, which gave the strongest hybridization signal, was chosen for further work. The remaining 52 clones hybridizing with the whole PG2 were screened with a probe derived from the 5′ end of *B. mori H. fibroin*. Based on experience with other silk genes (21), we expected some similarity between the *H. fibroin* genes of *B. mori* and *G. mellonella*. Indeed, under low stringency conditions, we detected weak hybridization of the *B. mori* probe to two genomic clones of *G. mellonella*. The DNA of these clones was isolated, cut with *Pst*I, and re-hybridized with the *H-fibroin* probe of *B. mori*. Both clones proved to contain a single 0.7-kb fragment similar in sequence to the first exon of *B. mori H-fibroin* gene. One of the two clones was taken for further research under the name λX1. Extensive mapping and subcloning revealed the presence of *Sty*I, *Acc*I, and *Sac*I restriction sites that were instrumental in elucidating large parts of the repetitive gene region. The 50 clones, which hybridized only with the full PG-2 probe, apparently contained only the internal sequences of the *H-fibroin* gene and were not analyzed.

H-fibroin Gene Structure and Expression—Genomic Southern blotting (data not shown) allowed us to construct a restriction map of the whole *H-fibroin* gene and to assess the size of the entire repetitive region. Localization of the restriction sites *Xba*I and *Bam*HI in the clone λX1 and of the *Sal*I and *Sac*I sites in the clone λXF was essential for the overall gene analysis. Genomic DNA fragments obtained by double digestions with *Xba*I + *Sal*I, *Xba*I + *Sac*I, *Bam*HI + *Sal*I, and *Bam*HI + *Sac*I, respectively, were probed with PG-2 and some also with the *Xba*I-*Bam*HI fragment of the clone λX1 and/or the *Sal*I-*Sac*I fragment of the clone λXF (data not shown). Analysis of the data allowed one possible arrangement of the fragments as shown in Fig. 1. The overlapping digestion fragments of the λX1 and λXF inserts were subcloned into pBluescript SK(+) according to the strategy summarized in Fig. 1 and sequenced. The results (GenBank™ accession numbers AF095239 and AF095240) confirmed that the clones represented the 5′ and 3′ ends, respectively, of the *H-fibroin* gene. By analogy with the *H-fibroin* gene structure in other lepidopteran species (12, 16), we assume that the internal *Xba*I-*Xba*I fragment (7 kb) and adjacent regions represent central part of the second exon (Fig.

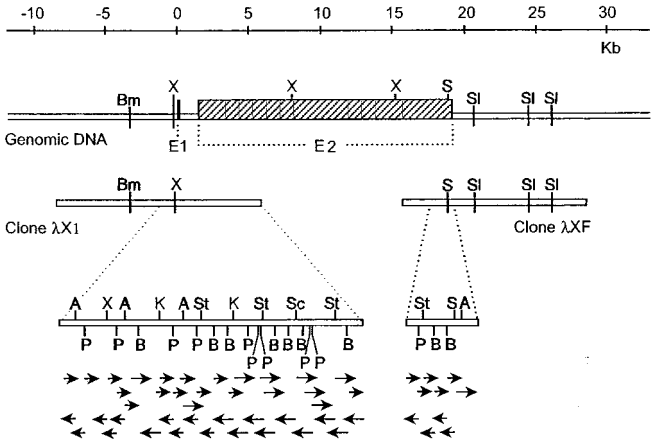

FIG. 1. **Restriction map of the *H-fibroin* gene in *G. mellonella* and the sequencing strategy.** The general gene organization (exons E1 and E2 are represented by *hatched boxes*) and several restriction sites are presented in the composite genomic map below the size bar. Additional restriction sites, which were used for subcloning, are shown in the aligned genomic clones λX1 and λXF. The sequenced regions are indicated by *arrows* (both the vector primers and the specific synthesized primers were used for sequencing). Restriction enzymes included *Acc*I (*A*), *Bgl*II (*B*), *Bam*HI (*Bm*), *Kpn*I (*K*), *Pst*I (*P*), *Sac*I (*S*), *Sac*II (*Sc*), *Sal*I (*Sl*), *Sty* (*St*), and *Xba*I (*X*).

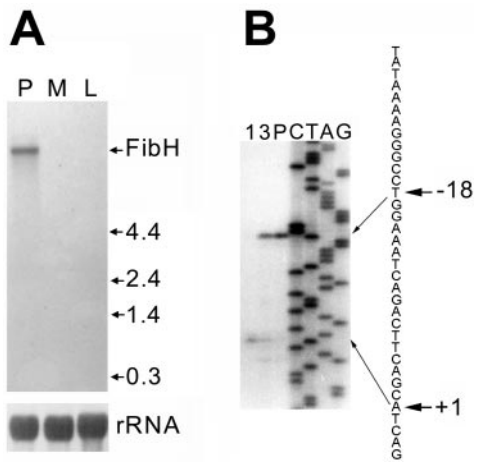

FIG. 2. **Transcription of the *H-fibroin* gene.** *A*, Northern analysis demonstrating tissue specificity of the expression. 5 μg of total RNA aliquots from the posterior (*P*) and middle (*M*) silk gland regions and from the body carcass without silk glands (*L*) were probed with the *Xba*I-*Bam*HI fragment (5′ end of the gene) of clone λX1. Ribosomal RNA stained with methylene blue was used to check RNA integrity. *B*, primer extension mapping of the transcription start. 5′-Radiolabeled (5×10^4 cpm) primer corresponding to the gene region −30 to +4 (Fig. 3*A*) and 10 μg of total RNA were used to synthesize the cDNA with the aid of murine leukemia virus reverse transcriptase. The template RNA was prepared from the posterior section of silk glands taken from the last instar larvae 1 day (*lane 1*) and 3 days (*lane 3*) after ecdysis and at the prepupal stage (*lane P*). The products were analyzed in a sequencing gel (8% polyacrylamide, 7 M urea) along with the C, T, A, G ladders of the λX1 genomic clone that was sequenced with the same primer. The results reveal two transcription starts (*arrows*) whose use depends on the age of the larva.

1) that is made up of similar repeats as found in clones λX1 and λXF (see below).

Northern blots probed with the labeled *Xba*I-*Bam*HI fragment of clone λX1 (encoding H-fibroin N terminus) disclosed the presence of a single, very large posterior silk gland-specific mRNA (Fig. 2*A*). A similar transcript was previously detected with the PG2 probe that encodes H-fibroin C terminus (17).

The Gene Sequence—The sequenced part of the λX1 clone included 1285 nt of the 5′ non-coding region, 42 nt putatively encoding 14 amino acids (beginning with an initiation codon in position +29), 1310 nt evidently non-coding and therefore regarded as an intron, and 4362 nt with a continuous open reading frame. Sequences between positions −426 and −346 in the upstream region, and between +1123 and +1203 in the intron, exhibited about 90 and 70% identity, respectively, with the 3′ end of the repetitive element *Gm1* (21).

Transcription start of the *H-fibroin* gene was determined by primer extension analysis using posterior silk gland RNA (Fig. 2*B*). The analysis disclosed two different transcription starts that appeared to be used singly or simultaneously, depending on the developmental stage of the last instar larva. The transcription start used by the newly ecdysed last instar larvae was assigned +1, and the whole gene was numbered relative to this position. This site was also identified in the *H-fibroin* gene of other Lepidoptera (Fig. 3*A*). In the course of the last larval instar of *G. mellonella*, the transcription start was shifted to position −18, and the +1 position ceased to be used. Two sequences possibly served as TATA boxes, each of them being associated with one transcription start: a TATAAAA sequence at −30 to −24 and a TAATATA sequence at −49 to −43 (Fig. 3*A*). An initiator sequence TCAGT (23) at +2 to +6 followed immediately after the transcription start at +1, and a similar sequence TCAGA was localized at −12 to −8, *i.e.* just a few nucleotides downstream from the transcription start at −18.

The exact delimitation of the intron was determined by sequencing across the exon/intron junctions in a cDNA clone obtained by reverse transcription-PCR. The amplified cDNA fragment began at +9 and ended at +2851, thus including almost the entire exon 1 and a part of the large exon 2. The sequencing proved that the first exon consisted of the 28-nt leader sequence and the 42-nt coding sequence (Fig. 3*A*). The second exon began with a unique sequence of 532 nt and continued as regular repeats that are described below. The repeats were iterated until the 3′ end of the sequenced part of clone λX1 and the same type of repeats appeared at the 5′ end of clone λXF (Fig. 4). The sequenced portion of the latter clone included 1331 nt of the coding region, of which about 70% consisted of the repeats, and 183 nt of adjacent non-coding sequence (GenBank™ accession number AF095240). The point of transcription termination, *i.e.* the 3′ terminus of the second exon, was established by comparing the sequence of λXF with that of the PG-2 cDNA clone (17). It was confirmed that the 3′-untranslated region consisted of 94 nt and included polyadenylation signal AATAAA (24) localized 23 nt upstream from the transcription terminus. The last 89 nt identified in the λXF clone represented non-transcribed spacer DNA flanking the 3′ end of the *H-fibroin* gene.

Predicted Amino Acid Sequence—The identified λX1 sequence encodes a putative peptide of 1468 amino acid residues (Fig. 5*A*). The expanse of the first 188 residues exhibits no obvious sequence iterations, but the remainder consists of regular repeats. Compliance with the general rules (25) indicates that a signal peptide is cleaved off between residues 18 and 19. The following tract of 170 residues is remarkable by the high content of charged amino acids that are often grouped by 2. The 4.5 p*K* value reflects prevalence of the negatively charged residues. About one-third of residues in the non-repetitive N terminus is hydrophobic and one-third is polar. Notable is a degenerate doubling of two short sequences.

The putative translation product of the sequenced part of clone λXF includes 443 amino acid residues, of which 373 are arranged in repeats, and 70 constitute a non-repetitive C terminus (Fig. 5*A*). It is characterized by high content of arginine that shifts the p*K* value to 11.6, and by the presence of two similar stretches of 7–8 amino acids.

The repeats found in clones λX1 and λXF are identical,

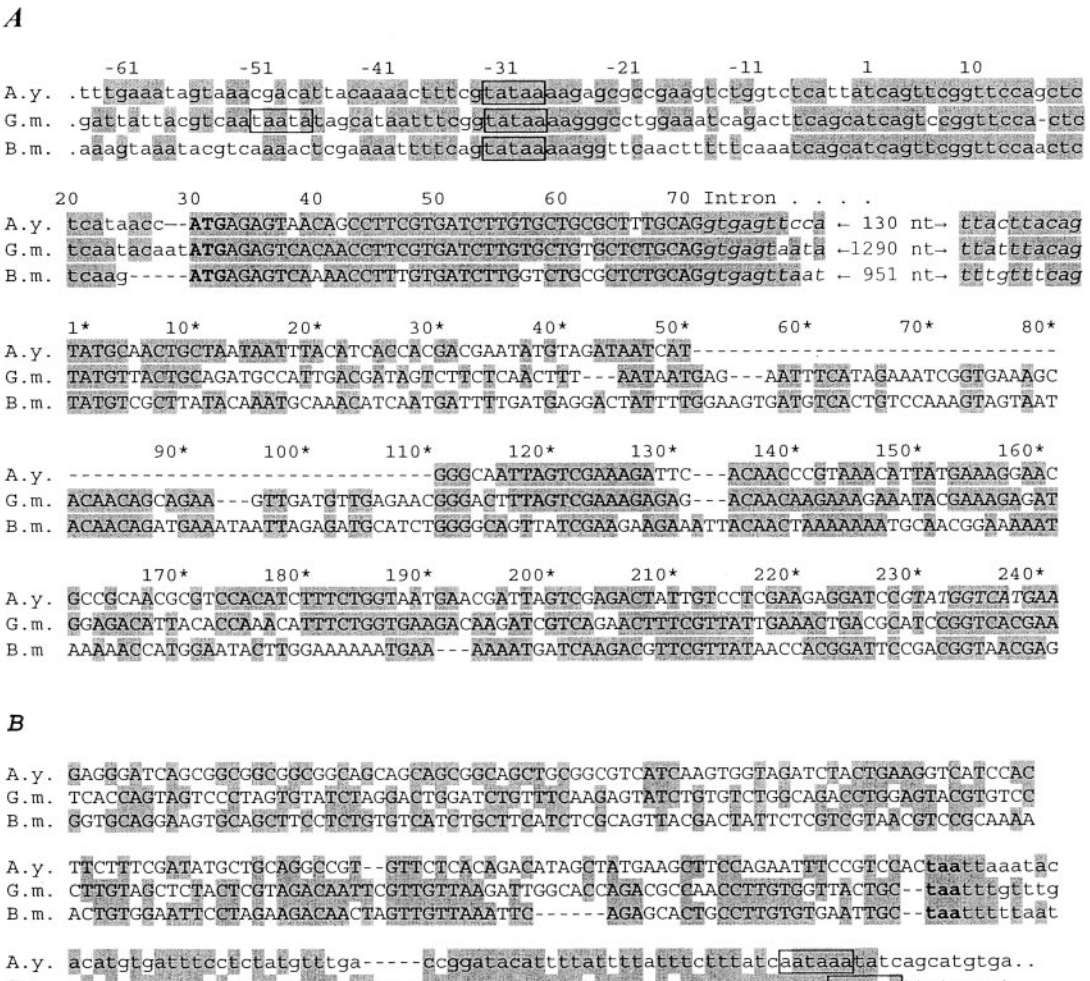

FIG. 3. **Terminal parts of the *H-fibroin* gene of *A. yamamai* (*A.y.*, GenBank™ entry X05578; last 13 residues shown in italics actually represents the closely related gene of *A. pernyi*, AF083333), *G. mellonella* (*G.m.*, AF095239 and AF095240), and *B. mori* (*B.m.*, V00094 and AF226688).** Sequences are aligned to maximize homologies indicated with *gray background.* The non-coding parts are printed in *lowercase letters*, and the coding parts in *uppercase letters. A*, the *H-fibroin* 5′ ends with the upstream flanking regions. The TATA motifs are *boxed*, and the initiation codons are printed in *bold.* Sequences of the upstream region and the first exon are numbered from the commonly used transcription start; 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.*

indicating that they are iterated throughout the central gene region between the two clones. From the genomic Southern mapping summarized in Fig. 1, we estimate that the entire coding region is ~18,000 nt long and encodes ~6000 amino acids. The deduced molecular mass of *G. mellonella* H-fibroin is thus close to 500 kDa, and over 95% of the molecule consists of repeats. They are built up of predominantly neutral amino acids; only about 2.7% residues carry a strong negative charge and 0.6% a positive charge. The composition of repeats is decisive for the amino acid representation in the whole fibroin. Relative representation of amino acid residues in the deduced translation products of λX1 and λXF is consistent with the results of direct amino acid analysis of fibroin extracted from the silk glands (26) or the cocoons (27, 28).

Repeated Motifs in the H-fibroin Protein—Beginning at the predicted amino acid residue 189, the sequence of *G. mellonella* H-fibroin continues as a regular repetitive arrangement of three types of complex repeats (Fig. 5*A*). The repeat A consists of 63 residues, and the repeats B_1 and B_2 consist of 43 and 18 amino acid residues, respectively. The repeats are combined in higher order repetitions AB_1 and AB_2, and these are arranged into assemblies ($AB_1AB_1AB_1AB_2(AB_2)AB_2$). We estimate that fibroin encompasses about 11 such ensemble repeats. Both the absence of any variability in the length of repeats and the high conservation of their sequence are striking. Hydrophobic residues dominate, but their stretches are broken into short motifs by intercalated polar residues. Charged amino acids are rare and occur exclusively in the repeat A. Only a few amino acid residues are occasionally replaced with others of similar properties, *e.g.* Leu alternates with Ile and Val, and Asp with Glu. Replacements between Gly, Ala, and Ser are also relatively common. Most variations are confined to certain positions and occur in several copies of the respective repeat. The total variation never exceeds 20%, and certain regions are identical in all copies.

The short motif GLGGLG, in which L may be replaced with I, V, and exceptionally by S, is shared by all repeats (Fig. 5*B*). The A repeats usually contain a GVGGLG modification in positions 38–43 (numbered from the repeat beginning) and a PLGGIG modification in positions 49–54. Each B_1 repeat contains 3 similar motifs in positions 4–9, 15–20, and 33–38, respectively. The B_2 repeat includes just one motif that is

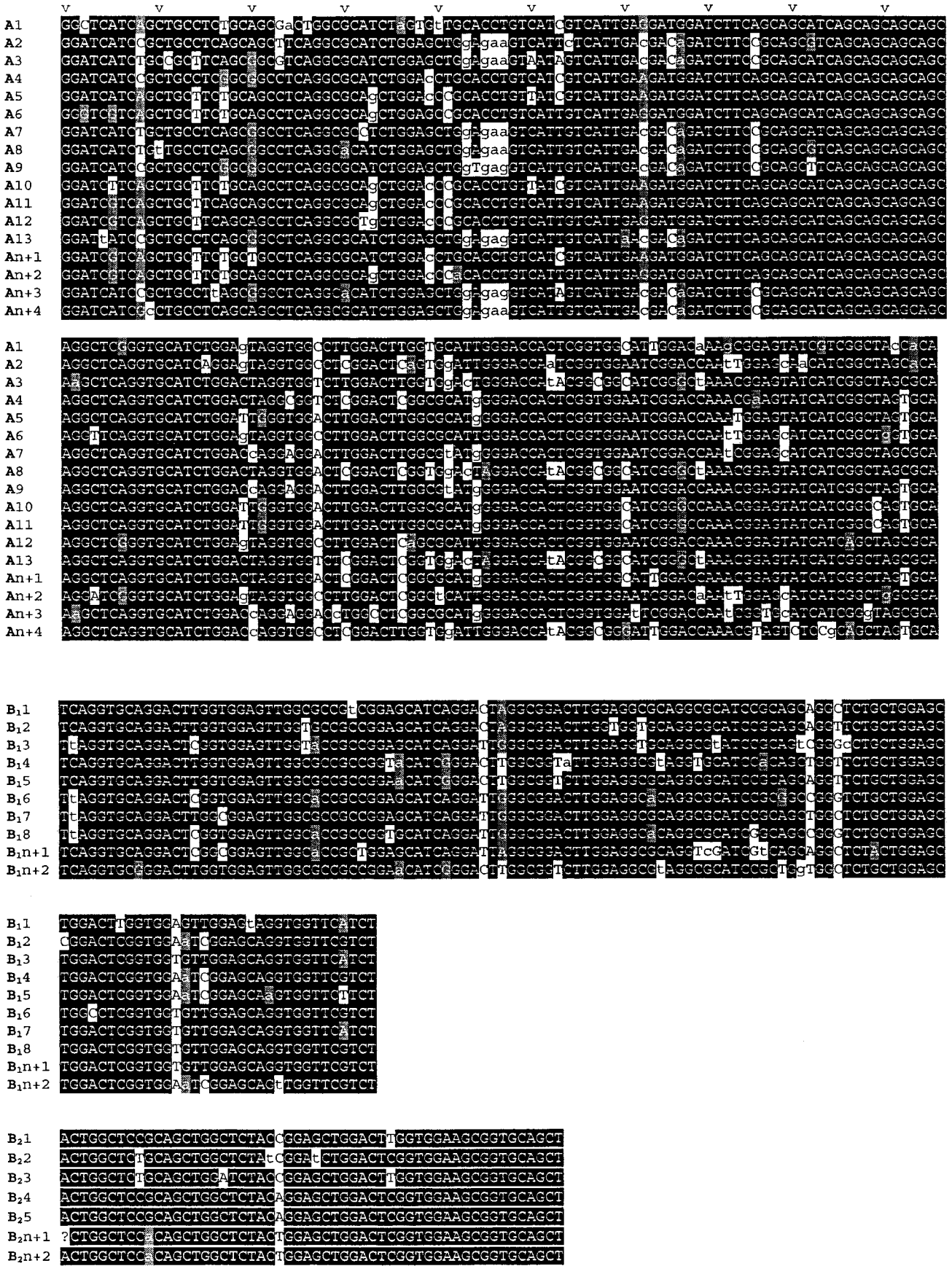

FIG. 4. **Sequences of the *H-fibroin* repeats.** The repeats A (189 nt), B_1 (129 nt), and B_2 (54 nt) are numbered according to their position in the gene (numbering begins with 1) or in the λXF clone (numbering starts with $n + 1$). The sequenced parts of clones λX1 (5742 nt) and λXF (1514 nt) contained $AB_1AB_1AB_1AB_2AB_2AB_1AB_1AB_1AB_2AB_2AB_2AB_1AB_1A$ and $B_2AB_2AB_1AB_1A$, respectively, repeat iterations (*cf.* Fig. 5). Nucleotides deviating from the usual sequence of the repeats are *highlighted,* and deviations changing codon assignment are printed in *lowercase letters.*

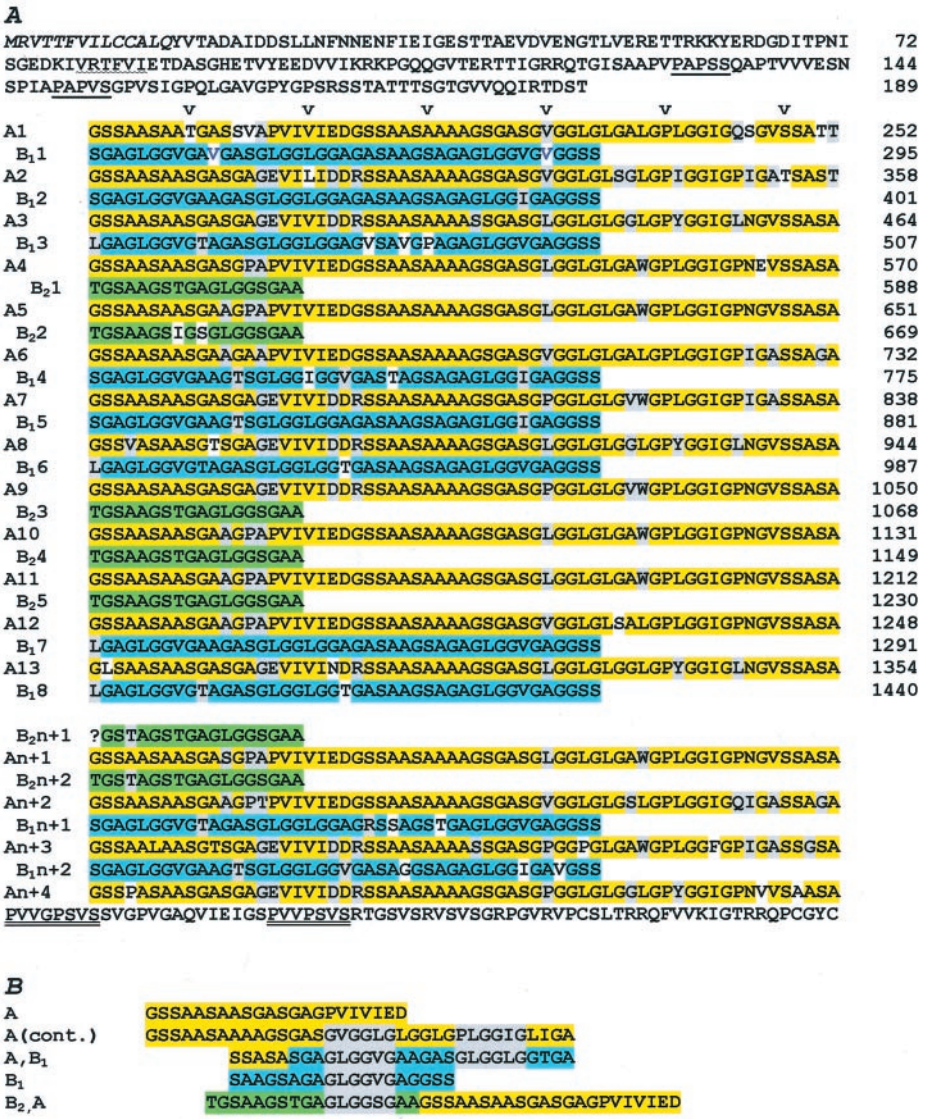

FIG. 5. **H-fibroin structure deduced from the gene sequence.** *A*, amino acid sequences of the non-repetitive N terminus (189 residues numbered from the translation start, with the presumed signal peptide printed in *italics*), the repetitive central parts, and non-repetitive C terminus. The N-terminal and C-terminal regions contain doubled short motifs (*underlined*, each pair in different way). The repetitive central region is arranged in the repeats A (*yellow*), B_1 (*blue*), and B_2 (*green*), which are numbered as the corresponding polynucleotide blocks in Fig. 4. Variations in the amino acid residues that occur in at least two copies of the respective repeat are shown on the *gray*, and singular replacements on the *white background*. *B*, consensus sequences of the repeats A (*yellow*), B_1 (*blue*), and B_2 (*green*) aligned to accentuate homologies. The hydrophobic motif GLGLG and its variants are shown on *gray background*.

flanked by similar amino acid residues as the last such motif in B_1, indicating that B_2 and the second portion of B_1 may have common origin. The hydrophobic motifs always alternate with short sequences containing polar residues Ser or Thr. Two similar and relatively long motifs, SSAASAAAA and SSAASAAS, are present in the A repeat and as highly altered and truncated modifications also in the B_1 and B_2 repeats (Fig. 5*B*). The hydrophobic group VIVI, which is followed by two acidic residues (in a single case, DD was changed to ND) and preceded either by PAP or AGE tripeptide, is a notable feature of the A repeat (Fig. 5*A*).

DISCUSSION

H-fibroin gene of Lepidoptera—The *H-fibroin* gene of *G. mellonella* is composed of one very short and one very long exon (Fig. 1), as are the homologous genes in *B. mori* (16) and *A. pernyi* (12). Because *G. mellonella* is phylogenetically rather remote from the two other species, the described arrangement of the fibroin gene is probably standard for most Lepidoptera. Sequence homologies at the 5′ and 3′ ends indicate the existence of a common ancestral *H-fibroin* gene.

The DNA sequence of several hundreds nucleotides upstream from the transcription start was identified in the *H-fibroin* gene of *B. mori* (29), *Antheraea yamamai* (30), *A. pernyi* (12), and now *G. mellonella*. In all these cases, the upstream region is AT-rich, with frequent concatenations of up to 7 A or T nucleotides. Interspecies homology is obvious downstream from about −50 position, including the presence of a TATA box at −0 to −24 (Fig. 3*A*). *G. mellonella* acquired, probably by coincidental mutations, another TATA box at −49 to −43. The existence of two transcription starts (standard at +1 and additional at −18) indicates strongly that both TATA boxes are functional. We cannot explain, however, why the start is at +1 in young larvae and at −18 in old larvae (Fig. 2*B*).

Homology between *H-fibroin* genes of different species is very clear through the first exon, including the exon/intron boundary (Fig. 3*A*). A few deletions and insertions modified the length of the leader sequence, but the following part of the first exon, which encodes the signal peptide, is remarkably preserved. The length of the intron and its internal sequence are dissimilar, but similarities between the compared species are found in the sequence of about 250 nt beginning around the intron/exon boundary and continuing in the second exon. The following major part of the exon is composed of highly species-specific repeated motifs. Similarity between the species occurs again at the 3′ end of the gene (Fig. 3*B*). For example, the non-translated tail sequence is of similar length and includes a GGATACAT motif localized 27–31 nt prior to the polyadenylation signal.

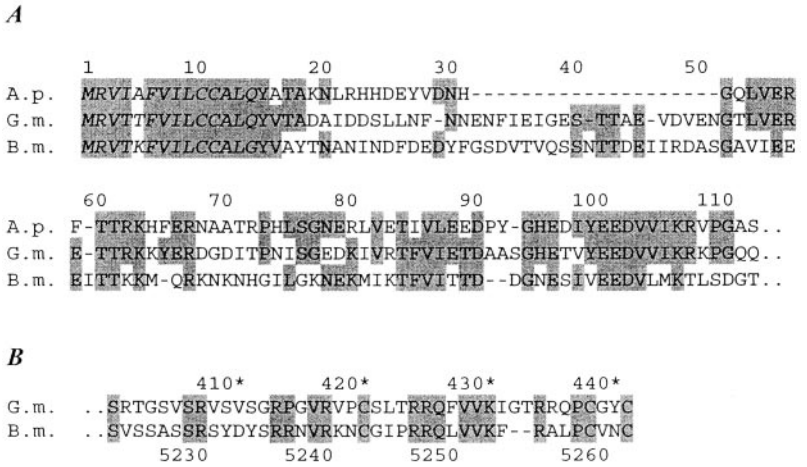

FIG. 6. **Comparison of the non-repetitive terminal sequences of H-fibroin proteins.** *A*, N-terminal parts exhibiting similarities between the three compared species: *A. pernyi* (*A.p.*), *G. mellonella* (*G.m.*), and *B. mori* (*B.m.*). Signal sequence, encoded by the first exon, is typed in *italics*. *B*, extreme C termini of *G. mellonella* (*G.m.*) numbered as in Fig. 4, and *B. mori* (*B.m.*) numbered according to Ref. 16.

The H-fibroin Proteins—The electrophoretic mobility of *G. mellonella* H-fibroin in polyacrylamide gel indicated a size of 240 kDa (25), which is considerably less than the 500 kDa deduced from the gene size (Fig. 1). However, size estimation in the gel is inaccurate, especially in the case of large hydrophobic molecules such as H-fibroin. Partial degradation of the analyzed H-fibroin during solubilization in 8 N urea cannot be excluded. We trust size assessment based on the gene analysis, and we emphasize that the H-fibroin of *G. mellonella* is larger than in other lepidopterans so far examined. A size of 391 kDa was computed for the H-fibroin of *B. mori* (16), and 216 kDa can be deduced for *A. pernyi* from the molecular data of Ref. 12. The last figure is in good agreement with the 220-kDa assessment of the extracted H-fibroin (31). The size of H-fibroins in *Antheraea myllita* was estimated as ~197 kDa (32), and H-fibroins of two other saturniid species, *A. yamamai* and *Philosamia cynthia ricini,* behaved as proteins of 250 and 230 kDa, respectively (31).

Interspecific diversity of the H-fibroin size is due to differences in the large central protein region that is composed of repeats, whereas the terminal non-repetitive sequences are of similar length in all species. In *G. mellonella*, the non-repetitive N terminus encompasses 173 amino acid residues, of which a stretch of more than 100 is clearly homologous with the H-fibroins of *B. mori* and *A. pernyi* (Fig. 6*A*), as well as *Bombyx mandarina* (33) and *A. yamamai* (30). The major part of the signal sequence, which is encoded by the first exon, is nearly identical in all these species. The non-repetitive sequence encoded by the second exon has diversified but more than 20 residues have maintained similar spacing. They are arranged in conserved groups, with occasional insertions or deletions between them. A major deletion of 20 amino acids (in comparison with the H-fibroin of *B. mori*) follows after the residue 31 in the H-fibroins of *A. pernyi* and *A. yamamai.* Despite great phylogenetic distance, the conservation of amino acid positions between *G. mellonella* (superfamily Pyraloidea) and *Antheraea* is similar or even higher than between *Antheraea* and *Bombyx* which both belong to the superfamily Bombycoidea. The non-repetitive H-fibroin N terminus of all species contains a high proportion 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. This and other conserved attributes of the N terminus suggest that it may play a role in silk fiber assembly.

The C terminus of the H-fibroin is characterized by the presence of 3 cysteines that are located in the proximity of eight basic residues (Fig. 6*B*). For *B. mori* it has been shown that the cysteine most distant from the protein end forms a disulfide linkage to the L-fibroin, whereas the more proximal and the terminal cysteines form an intramolecular disulfide bond (34). We have previously demonstrated homology between the L-fibroins of *B. mori* and *G. mellonella* and suggested that they have similar functions (35). The finding of conserved positions of the 3 cysteines in the H-fibroin (Fig. 6*B*) strongly indicates that the H-fibroin-L-fibroin complex of this species is assembled in a way similar to *B. mori.*

The C terminus of *Antheraea* H-fibroin is short and includes only 3 basic residues. It does contain 3 cysteines, but the alignment to the H-fibroin C termini of *B. mori* and *G. mellonella* is ambiguous. The diversification of *Antheraea* H-fibroin C terminus from the consensus found in *B. mori* and *G. mellonella* is associated with the loss of L-fibroin (10). It was proposed that *Antheraea* H-fibroin forms homodimers that are held together by a disulfide bridge in which one of the cysteines is engaged (30).

Organization of the Repetitive H-fibroin Core—Amino acid composition of the repetitive region of *G. mellonella* H-fibroin is dominated by residues Gly, Ala, Ser, and those with bulky apolar side chains (Table I). Amino acid sequence and high conservation of the A, B_1, and B_2 repeats, and the regularity of their assemblies $AB_1AB_1AB_1AB_2(AB_2)AB_2$ (Fig. 5*A*), are very spectacular features of *G. mellonella* H-fibroin. The basic repeats A, B_1, and B_2 are longer and more complex than the elementary repetitive units in other spider and lepidopteran silks. Their arrangement in the AB_1 and AB_2 subdomains and concatenation in about 11 assemblies are extremely regular. The H-fibroins of *B. mori* and *A. pernyi* are also composed of about 12 large ensemble repeats, but their sizes vary due to irregular numbers of the basic repetitive units they contain. For example, the numbers of residues range from 147 to 596 in the ensemble repeats of *B. mori* H-fibroin, and only the length (42–44 residues) of the amorphous spacers that separate the assemblies is conserved (36). The H-fibroin of *A. pernyi* is made up of 80 units that are each composed of a track of 4–14 alanines followed by one of 4 different types of non-crystalline motifs (12). A periodic occurrence of the shortest non-crystalline motif (11 residues) after each 5th to 8th unit breaks the repetitive region into the 12 repeat assemblies of unequal length.

The Maintenance of Uniform Repeats—All three types of repeats in *G. mellonella* H-fibroin are remarkably conserved throughout the sequenced part of the gene. Their uniformity at the DNA level (Fig. 4) is strengthened by the preferential use of certain isocodons. The high content of codons rich in G and C in the first two positions is compensated for by high frequency of isocodons ending with U and A. Similar compensation occurs in various types of DNA sequences (37). In the *H-fibroin* genes, however, the rate of recurrence of U or A in the third position is not random, and the frequency of isocodons is species-specific

TABLE I

Total number and percent of selected amino acids in the repeated region of known H-fibroins

Only part of the region (1651 residues) was sequenced in *G. mellonella*, but complete sequences are available for *B. mori* (5263 residues) and *A. pernyi* (2639 residues).

Amino acid residue	*G. mellonella*		*B. mori*		*A. pernyi*	
Gly	522	31.6%	2415	45.9%	720	27.3%
Ala	393	23.8%	1592	30.2%	1137	43.1%
Ser	298	18.1%	635	12.1%	297	11.3%
Ile, Leu, Val	264	16.0%	117	2.2%	42	1.6%
Pro	52	3.2%	14	0.3%	5	0.2%
Asp, Glu	45	2.7%	55	1.0%	135	5.1%
Thr	35	2.1%	47	0.9%	9	0.3%
Tyr	4	0.2%	277	5.3%	139	5.3%

TABLE II

Codon usage (in %) for the most common amino acids in repeated region of the H-fibroins

Residue	Codon	*G. mellonella*	*B. mori*	*A. pernyi*
Ala	gca	59.7	18.1	58.2
	gcc	11.9	10.4	6.7
	gcg	3.8	3.0	20.0
	gcu	24.4	71.3	14.9
Gly	gga	50.1	39.4	26.6
	ggc	21.6	4.1	27.5
	ggg	2.1	1.0	4.3
	ggu	26.0	55.5	41.5
Ser	agc	6.0	13.4	1.7
	agu	3.0	1.9	11.4
	uca	40.2	68.0	57.5
	ucc	9.0	0.6	5.1
	ucg	13.1	1.7	9.8
	ucu	28.2	14.3	14.5

(Table II). The incidence of "silent" mutations is similar to the rate of base replacements that change codon specification. For example, within the 4950 nt of the analyzed *G. mellonella* repeats (Fig. 4), we found 5.47% silent and 3% assignment-altering mutations.

The codon choice pattern in *B. mori H-fibroin* is correlated with skewed cellular contents of isoaccepting tRNAs (38, 39). Mita *et al.* (40) proposed that the restricted codon usage is dictated by the most stable conformations of chromatin or the encoded mRNA, whereas the composition of the tRNA pool is a secondary adaptation. We propose that the codon bias is also related to the stability of reiterated motifs. The tandems of short motifs, such as the minisatellites, are prone to replication slippage, unequal sister chromatid exchange, and unequal allelic recombination that generate variations in the length of the tandems (41). Allelic divergence in spider silk genes (42) and 15% length variation of *B. mori* H-fibroin (43) demonstrate the importance of this mechanism in the silk genes with short iterated motifs. By contrast, larger DNA blocks without intrinsic repeats are less likely to be internally misaligned during crossing over. We assume that a lack of reiterated short motifs within the A, B_1, and B_2 DNA blocks is a prerequisite for the maintenance of their length. The unique parts of some other silk genes, for example introns and coding "spacers" in a spider silk gene (44) and amorphous linkers joining the repetitive domains of *B. mori* H-fibroin (36), are also characterized by length conservation.

The crossover and misreplication between the iterated DNA segments drive their concerted evolution and lead to segment homogenization (45, 46). Homogenization of DNA repeats has been reported for microsatellites, repetitive regions of mitochondrial DNA, tandem gene arrays, and repetitive sequences within large genes. The process of sequence homogenization on one hand suppresses and on the other hand promotes fast spreading of base substitutions among the repeats. Mutations changing codon assignment are initially stochastic events, but the function of the encoded protein imposes constraints on their maintenance and propagation (47). Mutations improving the function are favored by natural selection that fosters spreading of the altered nucleotide to other repeats.

Molecular Conformation of G. mellonella H-fibroin—Classical x-ray diffraction studies of lepidopteran silk revealed that fibroin molecules are cross-linked by hydrogen bridges formed between the amide and the carbonyl groups of adjacent polypeptide strands (2). Amino acid repeats such as GAGAS or AAAAA are thereby aligned into pleated β-sheets, and interactions between the side chains of the amino acids of adjacent sheets cause their stacking into crystallites reinforcing the silk fiber. Intersheet distances in the crystallites are determined by the side chains of participating amino acids. The results of measurements of these distances (14, 15) are consistent with the domination of glycine in the H-fibroin of *B. mori* and the presence of polyalanine tracks in *A. pernyi* H-fibroin. The x-ray spectrum of *G. mellonella* silk indicated the presence of crystallites with a similar intersheet packing as in the silk of *A. pernyi* (3). Indeed, the sequences SSAASAAAA and SSAASAAS in the A repeat of *G. mellonella* H-fibroin (Fig. 4) resemble the polyalanine tracks of the *A. pernyi* H-fibroin. However, the repeats of *G. mellonella* H-fibroin also contain a high proportion of Gly alternating with Ala, Ser, and a few other residues. Some of these alternations are regular, and their conservation in reiterated repeats indicates that they can also form crystallites, albeit of different periodicity. New physical measurements are needed to resolve the structure of crystallites in *G. mellonella* H-fibroin.

In the H-fibroins of *B. mori* (16) and *A. yamamai* (12), the reiterated crystalline motifs are periodically interspaced with distinct "boundary" sequences. They contain residues with bulky side chains, *e.g.* Tyr, Trp, Glu, and Arg, that interrupt the β-sheet regions and allow random orientation of the H-

fibroin molecule conferring flexibility to the silk fiber (19). In the H-fibroin of *G. mellonella*, residues Glu, Asp, Trp, and Asn occupy specific sites in close proximity to short crystalline regions. This arrangement suggests that this H-fibroin may contain numerous small crystallites separated by short peptide strands, as has been suggested for a spider dragline silk (48). The residues with apolar bulky side chains, such as Leu, Ile, and Val, may have a dual role in *G. mellonella* H-fibroin. In some positions they appear to be incidentally replaced by Pro or Tyr, indicating that in these positions they are involved in disturbing the formation of β-sheets and prevent their stacking. The high conservation of Leu, Ile, and Val in other positions, usually in alternation with Gly, indicates a more specific role, including formation of some kind of crystallites.

Some highly elastic spider silks are composed of repeats allowing formation of β-spirals rather than sheets (4). Peptide chain conformations other than the β-sheet crystallites may also be important for the physical properties of *G. mellonella* silk. Similarly, high conservation of the apolar sequence VIVI followed by DD or ED (Fig. 5*A*) in all analyzed A repeats (in a single case DD was replaced by ND) of the H-fibroin suggests that hydrophobic interactions and electrostatic bonds may be involved in H-fibroin cross-linking. More data are needed to reconcile molecular conformation of the H-fibroin fibers.

Functional Considerations—In the presence of long repeats of distinct internal structures, the H-fibroin of *G. mellonella* resembles certain spider silks (1, 49). Spiders typically produce several types of silk whose propensities match the functional requirements. The dragline and web frame silks are very strong, whereas the orb web silks excel in elasticity. The caterpillars possess only one type of silk gland, and their silk is generally used for the construction of cocoons in which they pupate. Cocoon silk should be strong and persistent, and elasticity is irrelevant. These properties are met in the silk of *B. mori* and *A. pernyi* that have tensile strengths 7.4×10^8 and 5.8×10^8 Nm^{-2}, and extensibility 24 and 35%, respectively (19).

In a number of lepidopteran families, the use of silk is not limited to cocoon spinning. The larvae of *G. mellonella*, which develop in bee colonies, produce large quantities of silk from the second larval instar on, to construct silk tubes protecting them against the detection and killing by bees. The tubes are continuously renewed, but each section is used for several days. It must be strong and extensible to allow the larva inside to grow and to turn around without leaving the tube. Older parts of the tube are eaten and digested by the larva. Upon reaching its full size in the seventh larval instar, the larva abandons the tube and seeks a suitable place for cocoon spinning. The larva of 200 mg can then suspend on the silk thread and descend while spinning.

The functional requirement for a combination of strength and elasticity probably drove the silk evolution in *G. mellonella*. The elasticity of *G. mellonella* silk is obvious from the observations of larvae moving in their silken tubes, but it has not been measured. A note on extensibility hysteresis upon cyclical application of the strain load indicates elasticity, but no details were given (18). The tensile strength of 7.5×10^8 Nm^{-2} and the 101% extensibility of *G. mellonella* silk are the highest values found in Lepidoptera (18). The reported increase of extensibility in water may be regarded as an adaptive feature. The environment of bee nests is humid, whereas cocoons, for which extensibility is rather disadvantageous, are usually formed outside the nests in a dry place. *G. mellonella* silk is "designed" to be strong and flexible in the silk tubes produced in the warm and moist environment of the bee nest, digestible at the alkaline pH in the gut of caterpillars, and rigid and persistent in the cocoons that must last for weeks or months. These features are embodied in the structure and extreme regularity of the H-fibroin repeats.

Acknowledgments—We thank Drs. Dalibor Kodrík and Changsong Yang for sharing unpublished data on the fibroin protein analysis and on the developmental changes in *H-fibroin* gene expression, respectively. The *Pst*I-*Bcl*I fragment of *B. mori H-fibroin* gene was kindly provided by Dr. P. Couble of the University of Lyon, France. We also thank Dr. J. S. Edwards of the University of Washington, Seattle, WA, for critical reading of the manuscript.

REFERENCES

1. Gosline, J. M., Guerette, C. S., Ortlepp, C. S., and Savage, K. N. (1999) *J. Exp. Biol.* **202,** 3295–3303
2. Pauling, L., and Corey, R. B. (1953) *Proc. Natl. Acad. Sci. U. S. A.* **39,** 253–256
3. Lucas F., and Rudall, K. M. (1968) in *Comprehensive Biochemistry* (Florkin, M., and Stota, E. H., eds) Vol. 26B, pp. 475–558, Elsevier Science Publishers, Amsterdam
4. Hayashi, C. Y., and Lewis, R. V. (1998) *J. Mol. Biol.* **275,** 773–784
5. Sehnal, F., and Akai, H. (1990) *Int. J. Insect Morphol. Embryol.* **19,** 79–132
6. Žurovec, M., Yang, C., and Sehnal, F. (1998) *J. Biol. Chem.* **273,** 15423–15428
7. Nirmala, X., Kodrík, D., Žurovec, M., and Sehnal, F. (2001) *Eur. J. Biochem.* **268,** 1–10
8. Tanaka, K., Mori, K., and Mizuno, S. (1993) *J. Biochem. (Tokyo)* **114,** 1–4
9. Žurovec, M., Kodrík, D., Yang, C., Sehnal, F., and Scheller, K. (1998) *Mol. Gen. Genet.* **257,** 264–270
10. Tanaka, K., and Mizuno, S. (2001) *Insect Biochem. Mol. Biol.* **31,** 665–677
11. Inoue, S., Tanaka, K., Arisaka, F., Kimura, S., Ohtomo, K., and Mizuno, S. (2000) *J. Biol. Chem.* **275,** 40517–40528
12. Sezutsu, H., and Yukuhiro, K. (2000) *J. Mol. Evol.* **51,** 329–338
13. Warwicker, J. O. (1960) *J. Mol. Biol.* **2,** 350–362
14. Marsh, R. E., Corey, R. B., and Pauling, L. (1955) *Biochim. Biophys. Acta* **16,** 1–34
15. Marsh, R. E., Corey, R. B., and Pauling, L. (1955) *Acta Crystallogr.* **8,** 710–715
16. Zhou, C.-Z., Confalonieri, F., Medina, N., Zivanovic, Y., Esnault, C., Yang, T., Jacquet, M., Janin, J., Duguet, M., and Perasso, R. (2000) *Nucleic Acids Res.* **28,** 2413–2419
17. Žurovec, M., Sehnal, F., Scheller, K., and Kumaran, A. K. (1992) *Insect Biochem. Mol. Biol.* **22,** 55–67
18. Hepburn, H. R., Chandler, H. D., and Davidoff, M. R. (1979) *Insect Biochem.* **9,** 69–77
19. Denny, M. W. (1980) *Symp. Soc. Exp. Biol.* **34,** 247–272
20. Sehnal, F. (1966) *Z. Wiss. Zool.* **174,** 53–82
21. Yang, C., Teng, X., Žurovec, M., Scheller, K., and Sehnal, F. (1998) *Gene (Amst.)* **209,** 157–165
22. Ohshima, Y., and Suzuki, Y. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74,** 5363–5367
23. Cherbas, L., and Cherbas, P. (1993) *Insect Biochem. Mol. Biol.* **23,** 81–90
24. Proudfoot, N. J. (1991) *Cell* **64,** 671–674
25. Nielsen, H., Engelbrecht, J., Brunak, S., and von Heijne, G. (1997) *Protein Eng.* **10,** 1–6
26. Michalik, J., Sehnal, F., and Lassota, Z. (1984) *Acta Biochim. Pol.* **31,** 139–148
27. Lucas, F., Shaw, J. T. B., and Smith, S. G. (1960) *J. Mol. Biol.* **2,** 339–349
28. Bonnot, G., and Delobel, B. (1970) *Ann. Zool. Ecol. Anim.* **2,** 595–605
29. Tsujimoto, Y., and Suzuki, Y. (1979) *Cell* **18,** 591–600
30. Tamura, T., Inoue, H., and Suzuki, Y. (1987) *Mol. Gen. Genet.* **206,** 189–195
31. Tamura, T., and Kubota, T. (1989) in *Wild Silkmoths 88* (Akai, H., and Kiguchi, M., eds) pp. 67–72, National Institute Sericultural Insect Science, Tsukuba, Japan
32. Datta, A., Ghosh, A. K., and Kundu, S. C. (2001) *Insect Biochem. Mol. Biol.* **31,** 1013–1018
33. Kusuda, J., Tazima, Y., Onimaru, K., Ninaki, O., and Suzuki, Y. (1986) *Mol. Gen. Genet.* **203,** 359–364
34. Tanaka, K., Kajiyama, N., Ishikura, K., Waga, S., Kukuchi, A., Ohtomo, K., Takagi, T., and Mizuno, S. (1999) *Biochim. Biophys. Acta* **1432,** 92–103
35. Žurovec, M., Vašková, M., Kodrík, D., Sehnal, F., and Kumaran, A. K. (1995) *Mol. Gen. Genet.* **247,** 1–6
36. Zhou, C.-Z., Confalonieri, F., Jacquet, M., Perasso, R., Li, Z.-G., and Janin, J. (2001) *Proteins Struct. Funct. Genet.* **448,** 119–122
37. Nakamura, T., Soyama, A., and Wada, A. (1991) *FEBS Lett.* **289,** 123–125
38. Garel, J. P., Mandel, P., Chavancy, G., and Daillie, J. (1970) *FEBS Lett.* **7,** 327–329
39. Suzuki, Y., and Brown, D. D. (1972) *J. Mol. Biol.* **63,** 409–429
40. Mita, K., Ichimura, S., Zama, M., and James, T. C. (1988) *J. Mol. Biol.* **203,** 917–925
41. Armour, J. A. L., Monckton, D. G., Neil, D. L., Tamaki, K., MacLeod, A., Allen, M., Crosier, M., and Jeffreys, A. J. (1993) *Genome Anal.* **7,** 43–57
42. Beckwitt, R., Arcidiacono, S., and Stote, R. (1998) *Insect Biochem. Mol. Biol.* **28,** 121–130
43. Manning, R. F., and Gage, L. P. (1980) *J. Biol. Chem.* **255,** 9451–9457
44. Hayashi, C. Y., and Lewis, R. V. (2000) *Science* **287,** 1477–1479
45. Smith, G. P. (1976) *Science* **191,** 528–535
46. Dover, G. (1982) *Nature* **299,** 111–117
47. Meeds, T., Lockhard, E., and Livingston, B. T. (2001) *J. Mol. Evol.* **53,** 180–190
48. Termonia, Y. (1994) *Macromolecules* **27,** 7378–7381
49. Gatesy, J., Hayashi, C., Motriuk, D., Woods, J., and Lewis, R. (2001) *Science* **291,** 2603–2605