

Amino Acid Sequence of Sp23, a Structural Protein of the Spermatophore of the Mealworm Beetle, *Tenebrio molitor**

(Received for publication, April 15, 1992)

Guido C. Paesen, Margaret B. Schwartz, Marnix Peferoen†, Frantisek Weyda§, and George M. Happ¶

From the Department of Zoology, University of Vermont, Burlington, Vermont 05405

In this paper we present the amino acid sequence of Sp23, a structural protein of the spermatophore of the mealworm beetle (*Tenebrio molitor*). This is the first report of the primary structure of a spermatophorin. The protein is rich in proline (24%), relatively rich in tyrosine (9%) and glutamine (10%), and does not contain sulfur-containing amino acids. In the carboxyl-terminal half of the protein a peptide motif is repeated which is similar to a repetitive motif in a group of dipteran chorion proteins.

Spermatophorins are structural proteins of spermatophores, sacs that package sperm and seminal fluid in many insect species. In the yellow mealworm beetle, *Tenebrio molitor*, the spermatophore is an elongate closed cylinder with a solid core at its anterior end (1). The semen lies in the lumen between the outer wall and the core. The wall, core, and seminal plasma are products of two pairs of accessory reproductive glands of the male. From the tubular accessory glands (TAGs)¹ come more soluble proteins, of which some mix with the sperm. Four classes of TAG proteins, distinguished by their molecular weight ranges and isoelectric points, are abundant in the adult male (2, 3). The best known are the highly repetitive D proteins (4).

The larger bean-shaped glands (BAGs) contribute an array of proteins and lipoproteins that form the wall and core of the spermatophore. There are eight distinct cell types in the BAG, and each type secretes a different product. The secretions from the different cells collect in an aggregate plug within the lumen at the center of each BAG (5). As the plug flows from the BAGs into the ejaculatory duct, the blocks of secretion from the various cell types are molded into the thin sheets of the multilayered tubular wall and the solid core of

the primary spermatophore (5-7). With the injection of the sperm, the spermatophore is forced down the long male tract and ejected into the female. Next, the spermatophore spontaneously elongates as semen flows forward until finally the inflated anterior tip ruptures to liberate semen.

Three intriguing features of the spermatophore must somehow be explained by the properties of its component proteins: ordered assembly in extracellular space, stabilization once the layers are formed, and the plasticity of the wall which allows elongation in the female.

In a previous paper, Shinbo *et al.* (8) reported the purification of the spermatophorin Sp23 by immunoaffinity chromatography using the monoclonal antibody designated PL21.1. Immunohistochemistry with PL21.1 suggested that Sp23 is produced exclusively by cell type 4 of the BAG. The antigen also showed up in a discrete layer of the spermatophore wall. On denaturing SDS gels, it has an apparent molecular mass of 23 kDa, but after separation of cellular proteins and purified fractions on native gels followed by Western blotting, the PL21.1 antigen has an apparent molecular mass of 370 kDa. Compositional analysis showed that this protein is rich in proline (25.2%) and in glutamine/glutamic acid (15.4%).

The present paper is, to our knowledge, the first report of the amino acid sequence of a spermatophorin. The sequence was inferred from the cDNA clones isolated from libraries in λ expression vectors, which were constructed from messenger RNA of the adult bean-shaped glands.

EXPERIMENTAL PROCEDURES

Animals—Mealworms (*T. molitor*) were purchased from a commercial supplier, they were kept at room temperature and reared on Purina Chick Labchow. Males and females were separated in the pupal stage.

Immunoelectron Microscopy—Formaldehyde (4%; Polysciences) in 0.1 M PIPES (pH 7.2-7.3) was used for fixation (2 h, room temperature) of spermatophores. After a short dehydration in 50% and 75% ethanol, the samples were incubated in an ethanol/LR White resin mixture (1/2) and thereafter embedded in undiluted LR White resin (LADD) (30). Sections were immunostained using the monoclonal antiserum PL21.1 and goat anti-mouse antibodies labeled with colloidal gold particles (Sigma). Antibodies were dissolved in a 10 mM phosphate buffer solution (pH 7.3) containing 1% bovine serum albumin. 10% goat serum (Sigma) was used as blocking solution. Control sections underwent the same procedure, except for the incubation with primary antiserum, which was omitted. The stained sections were examined on Philips EM 201 and 300 transmission electron microscopes.

Protein Purification—BAGs were dissected from adult mealworms and homogenized in an 8 M urea solution, which was deionized by elution over a mixed bed resin (AG 501-X8, Bio-Rad) column. The homogenate was left at room temperature for 16 h, and then centrifuged (12,000 \times g; 3 \times 10 min). The supernatant was submitted to HPLC, using a Vydac C₄ reverse phase column and a 0.1% trifluoroacetic acid (Sigma) running buffer with a 10-80% acetonitrile gradient. Fractions containing Sp23 protein were further purified by a

* This work was supported by National Institutes of Health Grant AI-15662 and United States Department of Agriculture Grant CSRS 87-CRCR-1-2406. 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™/EMBL Data Bank with accession number(s) M92928.

† Present address: Plant Genetics Systems NV, Plateaustr. 22, B-9000 Ghent, Belgium.

§ Present address: Dept. of Developmental Morphology, Institute of Entomology, 37005 Ceske Budejovice, Czechoslovakia.

¶ To whom all correspondence should be addressed.

¹ The abbreviations used are: TAG, tubular accessory gland; BAG, bean-shaped accessory gland; SDS, sodium dodecyl sulfate; HPLC, high performance liquid chromatography; NBT, nitro blue tetrazolium; X-phosphate, 5-bromo-4-chloro-3-indolyl phosphate; PCR, polymerase chain reaction; PIPES, 1,4-piperazinediethanesulfonic acid.

second HPLC run (Vydac C₁₈ reverse phase column, 0.1% *n*-heptafluorobutyric acid (Pierce Chemical Co.), 0–100% acetonitrile). For detection of Sp23-containing fractions and estimation of their purity, aliquots of the eluate fractions were resolved by SDS-polyacrylamide gel electrophoresis and immunoblotted using the monoclonal antiserum PL21.1 (8).

NH₂-terminal sequence analyses were performed in the protein analytical facility of the Medical Biochemistry Department at the University of Vermont. Sequence analysis was performed on an Applied Biosystems 475A protein sequencing system employing a gas phase Edman degradation protocol with on-line HPLC identification of phenylthiohydantoin derivatives of amino acids.

cDNA Library Construction—Two libraries were used, one constructed in λ gt11, the other in λ Zap II (Stratagene). The construction protocol was almost the same for both libraries. BAGs were excised from 5–8-day adult mealworms, rinsed with distilled water, and collected in liquid nitrogen. Total RNA was prepared by a guanidinium thiocyanate/phenol/chloroform extraction (31). mRNA was purified from total RNA by means of oligo(dT)-cellulose chromatography (32). cDNA was prepared according to Gubler and Hofman (33), using Amersham and Pharmacia LKB cDNA synthesis kits (for λ gt11 and λ Zap II, respectively). After insertion of the cDNA and packaging of the vectors (34), the libraries were amplified in Y1090 (λ gt11) or XL-I-Blue (λ Zap) cells.

cDNA Library Screening and Subcloning—The λ gt11 library was screened according to Mierendorf *et al.* (35), using a polyclonal mouse antiserum against electrophoretically purified Sp23. The library was expressed in Y1090 cells on LB plates. Three immunopositive clones were isolated and grown in liquid LB medium. The cultures were lysed, and the bacteriophage was recovered by precipitation with polyethylene glycol 6000 (36). DNA was extracted from the phage by phenol/chloroform extraction. cDNA inserts were excised with *EcoRI* and separated from the phage DNA by electrophoresis through a 1% agarose gel onto a piece of NA-45 DEAE-cellulose membrane (Schleicher & Schuell). The longest insert (583 bp) was ligated into the pBluescript II KS(+) plasmid (Stratagene) at the *EcoRI* site. The ligation product, which was termed gt11-Sp23, was amplified in XL-I-Blue cells and submitted to sequencing.

Sequencing (Fig. 1)—Sequencing of the gt11-Sp23 clone was carried out by the Sanger deoxy-mediated chain termination reaction (37), using Sequenase (UCB). Double-stranded as well as single-stranded templates were prepared. Single-stranded DNA was rescued and purified according to Short *et al.* (38). Double-stranded phagemid DNA was purified on PlasmidQuik columns (Stratagene) and alkali-denatured (39). Compressions were resolved by substituting dGTP for dGTP. Digestion of the plasmid with the restriction endonucleases *NarI* and *ClaI* (both Stratagene) followed by religation, resulted in the removal of the 5' half of the cDNA and allowed sequencing from the middle of the cDNA toward the 3' end. Using *NarI* and *NotI* (Stratagene), the 3' half was removed and sequencing from the middle to the 5' end was made possible. Sequence data were analyzed using IBI Pustell and Genetics Computer Group (GCG; Ref. 40) sequence analysis software.

Polymerase Chain Reaction (PCR)—Immunoscreening of the λ gt11 library did not yield full-length cDNAs. In order to obtain the 5' end of the Sp23 message, we submitted purified DNA from the λ Zap II library to PCR. Protein-free λ Zap DNA was obtained after five extractions with equal volumes of phenol/chloroform (1:1) and one extraction with an equal volume of chloroform, followed by ethanol precipitation. In the PCR reaction, about 5 μ g of this DNA was combined with a vector-specific primer and a Sp23-specific primer (5'-GTTTTCCAGTCAAGACG-3' and 5'-CGCCTGGTAAATGACCG-3', respectively). The Sp23-specific primer corresponds with bases 326–308 of the cDNA sequence from the λ gt11 library. Primers were purchased from Oligos Etc. Inc.; polymerase buffer, dNTPs, and AmpliTaq DNA polymerase were Perkin Elmer Cetus products. Template DNA was melted at 94 °C (50 s), primers were allowed to anneal for 2 min at 57 °C, and polymerization took place at 72 °C for 1.5 min. The cycle was repeated 37 times. After amplification, the vector part of the PCR products (125 base pairs) was removed by digestion with the restriction nuclease *NotI* (Stratagene); the cDNAs were provided with a *NotI* site at both ends during construction of the library). Thereafter, the PCR products were analyzed by agarose gel electrophoresis and DNAs of the appropriate length (more than 310 base pairs) were recovered from gel slices by means of SpinBind DNA extraction units (FMC BioProducts). The DNA was made blunt-ended by incubation with T7 polymerase (BRL, Ref. 13) and was inserted into pBluescript II KS(+) (Stratagene) at the *SmaI* site,

prior to sequencing (as described above). In total, the products of five separate PCR reactions (termed zap-5'-Sp23₁₋₅) were subjected to sequencing (Fig. 1B).

In a second set of PCR reactions we used 5'-CTAAAGGGAA-CAAAAGCTGG-3' as vector-specific primer, and 5'-TCAATCGAA-GAGGGTGGCG-3' as the Sp23-specific primer, the latter corresponding with bases 197–215 of the λ gt11 cDNA. Reaction conditions were as above. Three separate PCR products were inserted into pBluescript II KS(+). These PCR products constitute the 3' end of the gene (Fig. 1C) and were termed zap-3'-Sp23₁₋₃. They contain a 93-base overlap with the products of the first set of PCR reactions, and a 466-base overlap with the λ gt11 clone. They were submitted to double-stranded sequencing to make sure that the 5' end sequence obtained from the λ Zap library and the 3' end sequence from the λ gt11 library were derived from identical mRNAs.

Southern Blotting—*Tenebrio* pupae (total animals) were homogenized in a Dounce tissue grinder containing preheated (50 °C) lysis buffer (50 mM Tris-HCl, pH 8.0, 20 mM EDTA, 1 mM CaCl₂, 2% SDS). 50 μ g/ml proteinase K was added, and the homogenate was incubated at 50 °C for 2 h. Proteins were precipitated by adding 1/3 volume of a saturated NaCl solution. After spinning (2000 \times g, 4 °C, 15 min) the supernatant was incubated with RNase A (20 μ g/ml) at 37 °C for 15 min. The DNA was extracted with phenol (once) and chloroform (once) and precipitated with ethanol. It was redissolved in distilled water and digested with restriction endonucleases (BRL, see Fig. 6). After resolution on a 2% agarose gel, the DNA was transferred to a nylon membrane (Hybond N+, Amersham Corp.) and hybridized with a digoxigenin (Boehringer Mannheim) labeled probe. The probe was constructed with purified insert from the gt11-Sp23 clone by random primer labeling. The anti-digoxigenin alkaline phosphatase conjugate was used for detection of bound probe. Enzyme activity was visualized by adding the lumigen LumiPhos (Boehringer Mannheim) and exposing the blot to x-ray film (X-OMAT AR, Kodak), or by staining with nitro blue tetrazolium salt (NBT) and 5-bromo-4-chloro-3-indolyl phosphate (X-phosphate).

Northern Blotting—mRNA was isolated from BAGs from pupal and adult beetles, using the Micro-FastTrack isolation system (Invitrogen). RNA samples (5 μ g) were subjected to electrophoresis on formaldehyde-containing agarose (0.8%) gels, according to Sambrook *et al.* (41). The RNA was transferred to a nylon membrane (Hybond N+, Amersham), using 20 \times SCC as transfer buffer, and hybridized with the digoxigenin-labeled probe. LumiPhos was used for probe detection.

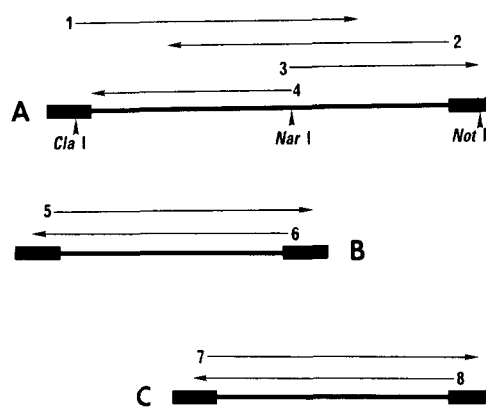


FIG. 1. Strategy for sequencing of gt11-Sp23 (A), zap-5'-Sp23 (B), and zap-3'-Sp23 (C). Arrows 1, 3, and 5 indicate single-stranded sequencing; arrows 2, 4, and 6–8 represent double-stranded sequencing. The fatter part of the bars stand for the pBluescript-polylinker regions, which flank the inserts. The zap-5'-Sp23 and zap-3'-Sp23 sequences are shown in alignment with the gt11-Sp23 sequence. A, at first, the 583-base pair insert was sequenced in both directions, using the KS and SK primer sites of pBluescript (arrows 1 and 2). After removal of the sequence between the *ClaI* and *NarI* restriction sites, the clone was religated and sequenced again, using the T3 primer site (arrow 3). Removal of nucleotides in between the *NarI* and *NotI* sites allowed sequencing from the T7 primer site on, as shown by arrow 4. B and C, the KS and T7 primer sites of the pBluescript polylinker region were used for sequencing the inserted PCR products.

RESULTS

Immunohistochemistry—Immunoelectron microscopy with PL21.1 revealed antigens in the loose layer of the wall and in its continuation in the core (Fig. 2). The term “loose” refers to the arrangement of the fibers, which appears to be less compact than in the other, adjacent fibrous layer (termed compact layer). The reaction in the loose layer was very strong. Some mild staining occurred in the granular peripheral layer and in the central granular mass of the core. These mild reactions were probably due to nonspecific binding of the antiserum.

HPLC Purification of Sp23—After two cycles of HPLC purification, a fraction was recovered which contained a single Sp23 protein band on Coomassie-stained SDS gels and a single PL21.1-immunoreactive spot on Western blots. We attempted to determine the sequence of the first 12 amino acids at the NH₂-terminal end. However, it was not clear whether the first residue is glycine or glutamic acid, and residues 8 and 12 could not be identified at all. This leaves us with the following sequence: (Gly or Glu)-Glu-Glu-Pro-Ala-Ala-Glu-Xaa-Ser-Gln-Gln-Xaa-Pro.

Sequence of the Sp23 Gene—The five products of the first PCR reaction (zap-5'-Sp23₁₋₅) were all identical to each other, meaning that the *Taq* polymerase did not introduce any errors

during amplification of the cDNA. There were no differences among the three zap-3'-Sp23 clones either. Fig. 3 shows the combined sequences of the gt11-Sp23 and zap-5'-Sp23 clones. Apart from the extra bases at the 5' end, the sequences obtained from the λ Zap library were identical to the gt11-Sp23 sequence. This suggests that the cDNAs were derived from identical mRNAs and that the two cDNAs can be combined to reconstitute the original gene. The sequence contains termination codons and a polyadenylation signal (AATAAA). The cDNA did not contain a poly(A) tail. Bases 2, 3, and 4 (ATG) possibly represent the start codon.

cDNA-deduced Amino Acid Sequence—The translation product of the largest open reading frame (Fig. 3) is rich in proline (23%), glutamine (10%), and tyrosine (9%). Methionine, cysteine, and tryptophan residues are absent.

The cDNA translation product contains the NH₂-terminal sequence of the purified protein (after the arrow in Fig. 3), and clarifies that the first (NH₂-terminal) residue is a glutamic acid.

In the cDNA translation, the NH₂-terminal sequence is preceded by a 7-amino acid piece (Met-Val-Ala-Ser-Ile-Ala-Gly). This very hydrophobic region most probably constitutes the signal peptide. A signal cleavage site at the position of the arrow in Fig. 3 is in accordance with the predictions of Von

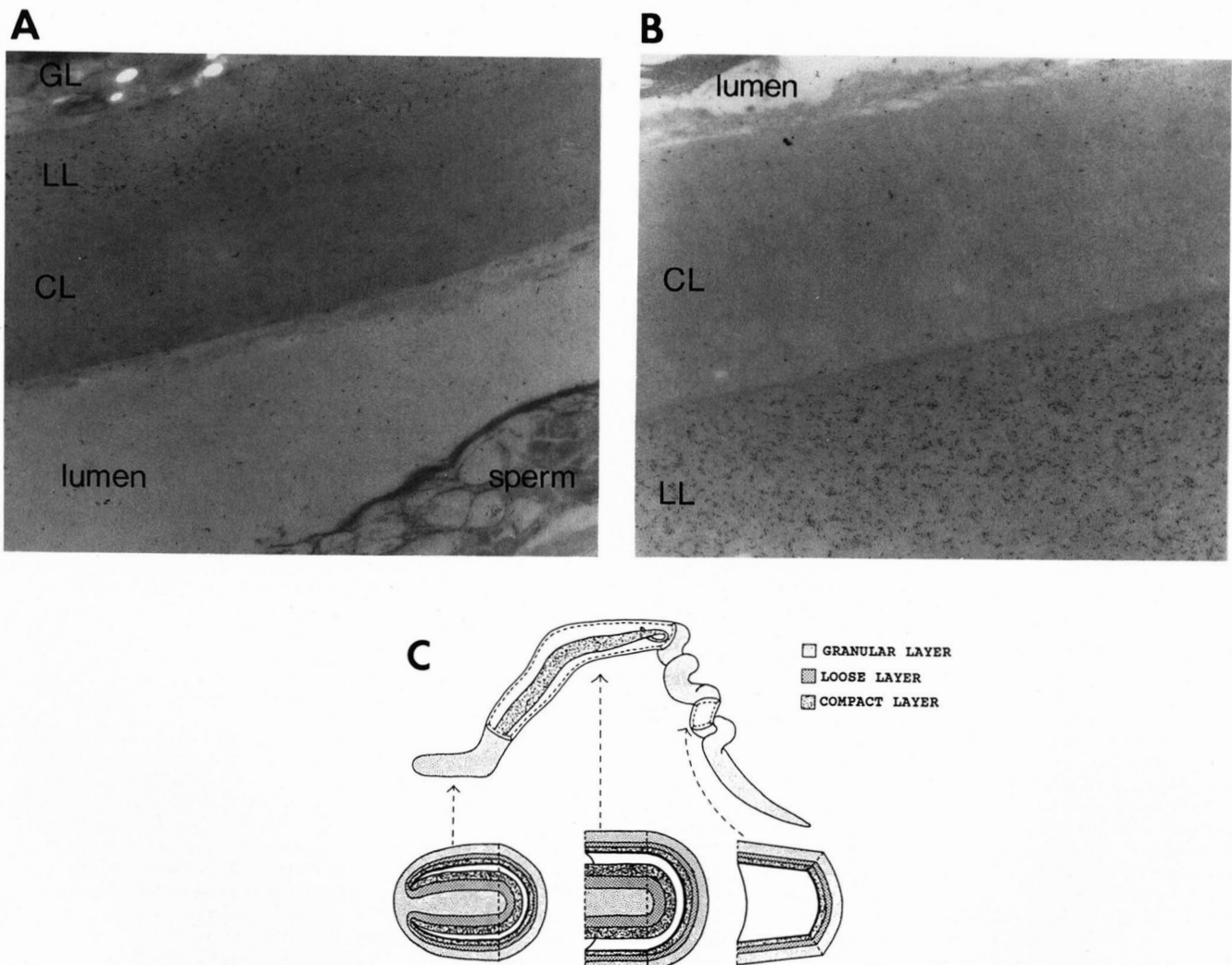


FIG. 2. Immunoelectron microscopy results. PL21.1-bound epitopes were found in one of the major layers of the wall (A) and in its continuation in the core (B). The layer is fibrous, with the fibrils loosely arranged (hence the name “loose layer”). It is situated in between the hydrophobic granular layer and the compact fibrous layer (C).

FIG. 3. Nucleotide sequence and deduced amino acid sequence of Sp23. The sequence as obtained from the *λ*gt11 clone is underlined. The dotted line represents the sequence of the 5'-PCR products. The double-underlined region was confirmed by sequencing the 3'-PCR products. The thick line shows the polyadenylation signal, putative phosphorylation signals are boxed, and the putative signal peptidase cleavage site is indicated by the arrow. The numbering of the amino acids starts after the arrow.

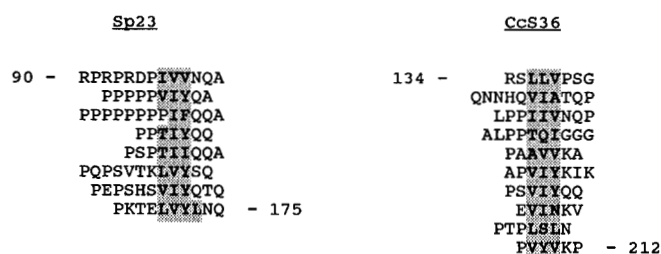
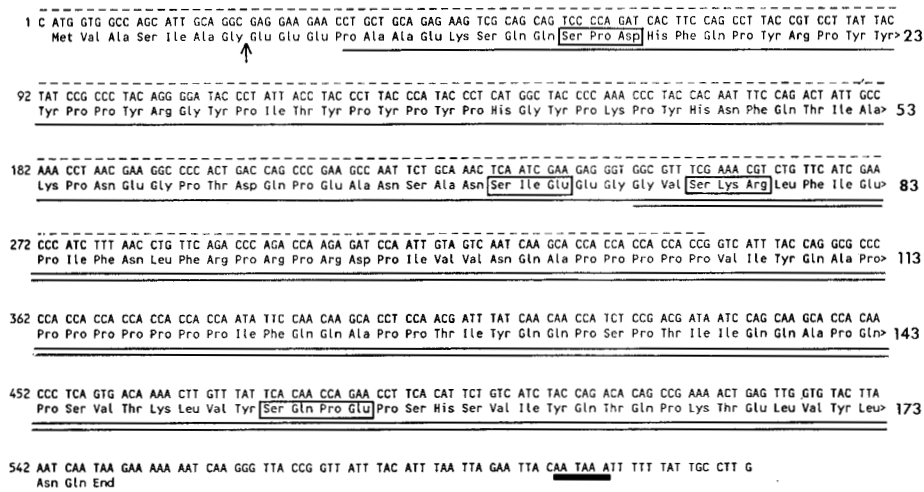


FIG. 4. The tandemly repeated peptide motif of the carboxyl-terminal half of the Sp23 translation product is compared with the repetitive peptide motif observed by Agelli *et al.* (23) in the central region of the *ceratitis capitata* eggshell protein S36. The shaded triplets most likely form hydrophobic β -sheet domains (10). The residues that compose these triplets are mainly the same in Sp23 and S36. The flanking regions of the eggshell protein consist of β -breakers (10, and are believed to give rise to β -turns (23). In the Sp23 protein, the amino acids preceding the shaded region are also strong β -structure breakers (10. According to Chou and Fasman (10), the glutamines directly following the hydrophobic triplets are indifferent to the local secondary structure and are predicted to extend the β -strands. However, Levitt (29) considers glutamine a β -breaker, suggesting that the β -sheet structures in Sp23, as in S36, consist of 3 residues only. β -Strands this short are not uncommon (29).

Heijne (9). The cDNA-deduced protein, without the putative signal sequence, has a calculated molecular mass of 19,939 Da and is slightly acidic (pI 6.1).

Different domains can be distinguished in the Sp23 protein. The NH₂-terminal end (amino acids 1–17 of the mature protein) is rich in glutamine and glutamic acid. It is followed by an easily distinguishable region rich in proline and tyrosine (residues 18–46). As if to accentuate its distinct character, the domain is flanked by two very similar peptides (His-Phe-Gln and His-Asn-Phe-Gln). The sequence from amino acid 47–89 has a higher diversity of amino acids but is still glutamic acid-rich and lacks tyrosine. In the rest of the protein a peptide motif is tandemly repeated (Fig. 4). The motif consists of 2–8 residues that are typical β -sheet breakers (Pro, Ser, Glu, and Lys), followed by 2–4 consecutive, hydrophobic residues that are considered to be the most common β -sheet formers (Val, Leu, Ile, Tyr, and Phe) (10). This hydrophobic cluster is mostly followed by 1 or 2 hydrophilic glutamines, which may or may not extend the β -sheet supposedly formed by the hydrophobic residues. According to computer predictions following the methods of Chou and Fasman (11) and Garnier *et al.* (12), the carboxyl-terminal half indeed consists of β -sheets

(in which the glutamines take part), interrupted by β -turns. The sequence analysis software also predicts a few short β -sheet structures in the NH₂-terminal half of the protein, and a short α -helix, immediately following the signal sequence.

Sp23 is predominantly hydrophilic, except for the small hydrophobic regions in the carboxyl half of the molecule and for one bigger hydrophobic region in the middle (corresponding with the sequence Phe-Ile-Glu-Pro-Ile-Phe-Asn-Leu). The sequence analysis software did reveal four putative serine phosphorylation sites (13, 14) but no glycosylation signals (Fig. 3).

Northern and Southern Blotting—The Northern blot in Fig. 5 reveals the presence of only minute amounts of Sp23 mRNA in pupal animals, compared with the quantities found in the adult stage. This is completely in accordance with the Western blots obtained with the PL21.1 antiserum (8), on which the Sp23 protein is detectable from day 8 of the pupal stage on. Northern hybridization also shows some faint bands of higher molecular weight, especially in the adult stage. They seem to be the result of nonspecific binding to sequences that are somewhat related to the Sp23 mRNA, and they are only visible because of the high amounts of mRNA that were applied to the gel (in order to detect the pupal signal).

As a rule, male Tenebrionidae carry a Y-chromosome (15). The Southern blot in Fig. 6A demonstrates that the Sp23-gene is not sex-specific; it appears in both male and female beetles. The two enzymes used to digest the genomic DNA prior to blotting have recognition sites in the (combined) cDNA at positions 40 and 564 (after the first stop codon). The size of the genomic DNA reacting most strongly with the probe is approximately 520 base pairs, suggesting that the gene has the same size as the cDNA, and therefore does not contain introns between the enzyme restriction sites. Digestions with *EcoRI*, *HinI*, *PvuII*, and *EcoRI* + *HindIII* result in a single band (Fig. 6B). Double digestion with *EcoRI* and *BamHI* gave two bands in the pooled DNA of five animals. This doublet is due to a difference in the location of a *BamHI* site from one animal to the other as was demonstrated on a Southern blot of DNA from individual animals (results not shown). Additional bands do appear when the exposure of the blot to x-ray film is extended (Fig. 6A), or when a lower stringency is used during hybridization (Fig. 6C). Under these conditions genomic DNA from the flour beetle *Tribolium brevicornis* also reacts with the probe (Fig. 6C).

DISCUSSION

Properties of the Sp23 Protein—The estimated molecular weight of Sp23 proteins on SDS-gels (23,000) is higher than

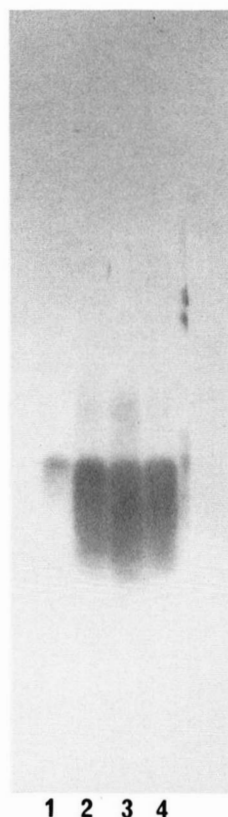


FIG. 5. Northern hybridization, using digoxigenin-labeled *gt11*-Sp23 insert as a probe. *Tenebrio* mRNA (5 μ g) was applied to each of the lanes of an agarose (0.8%)/formaldehyde gel. Lanes 1-4 contain mRNA from, respectively 8-day-old pupae, 1-day-old adults, 3-day-old adults, and 5-day-old adults.

is predicted from the translated mRNA sequence (19,900). Apart from the fact that its mobility on SDS gels is not always a reliable measure for a protein's molecular weight (16), especially when it concerns structural proteins (17, 18), the discrepancy in this case may well be due to posttranslational modification. Phosphorylation is possible, and some of the proline residues may undergo hydroxylation (the amino acid composition determination did not distinguish prolines from hydroxyprolines). Glycosylation, which is common in cuticular proteins (19, 20), probably also affects spermatophorins. Indeed, Sp23 can be stained with periodic acid-Schiff's reagent, an indication for covalently bound carbohydrate (8). However, the amount of carbohydrate added to Sp23 is probably small, since the inner and outer loose layers of the spermatophore, to which Sp23 contributes, could not be stained with periodic acid-Schiff reagent (1).

While the cDNA does not contain codons for cysteine, Shinbo *et al.* (8) reported the presence of half-cystine in immunoaffinity-purified Sp23. The amino-acid composition data, however, were from a single hydrolysis time and are therefore not completely accurate. The possible presence of contaminants in the protein sample, which are not detectable by SDS-polyacrylamide gel electrophoresis, could also account for the discrepancy. The existence of Sp23 variants with a somewhat different amino acid composition, is a less likely explanation. Southern analysis of genomic *Tenebrio* DNA, digested with different enzymes, mostly results in the appearance of a single band (Fig. 6). When genomic DNA is digested with *Hinf*I, the only clear band that is detected corresponds with an (approximately) 800-base pairs-long DNA piece. This piece can contain no more than one Sp23 gene. If there would be a closely related sequence, a second band would surely be visible. Moreover, Western blots of SDS gels did not reveal

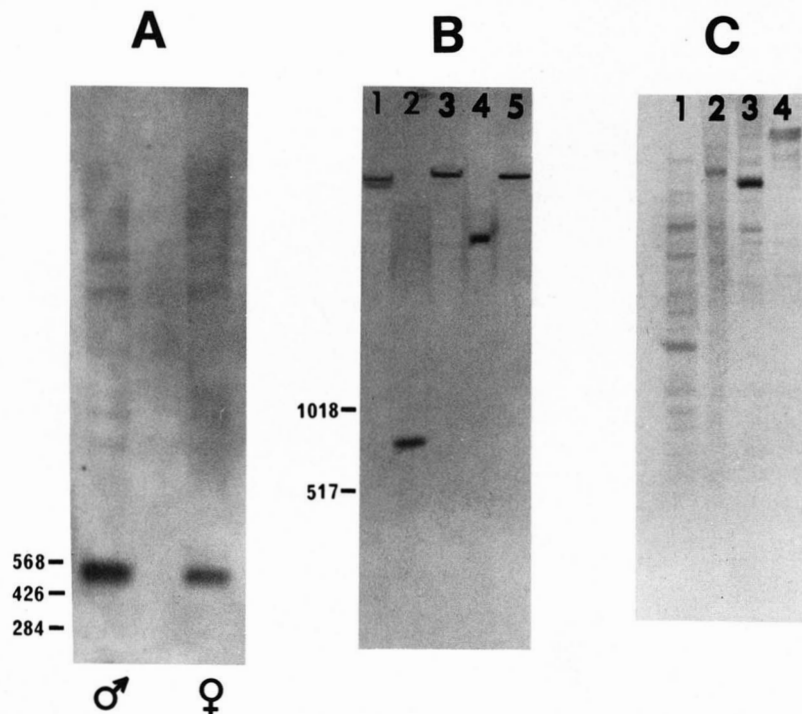


FIG. 6. Southern analysis of genomic DNA (~5 μ g/lane). A, genomic DNA from male (lane 1) and female (lane 2) *T. molitor* was double-digested with *Pst*I and *Bst*EII, which cut at position 40 and 564 of the Sp23-cDNA. LumiPhos was used for detection of the probe. Hybridization temperature was 65 $^{\circ}$ C. B, genomic DNA from male mealworm beetles was digested with *Eco*RI + *Bam*HI (lane 1), *Hinf*I (lane 2), *Eco*RI (lane 3), *Pvu*II (lane 4), and *Eco*RI combined with *Hind*III (lane 5). Hybridization temperature was 65 $^{\circ}$ C; NBT and X-phosphate were used for probe detection. C, genomic DNA from the flour beetle, *T. brevicornis*, was digested with *Hinf*I (lane 1) and *Eco*RI + *Bam*HI (lane 2). The remaining lanes contain *Tenebrio* DNA, digested with *Pvu*II (lane 3) and *Eco*RI + *Bam*HI (lane 4). Hybridization temperature was 58 $^{\circ}$ C instead of 65 $^{\circ}$ C; NBT and X-phosphate were used for probe detection.

PL21.1-immunopositive bands, corresponding to molecular masses other than 23 kDa (8), nor did the Northern hybridization suggest a lot of variation in the size of the Sp23 messenger RNA (Fig. 5).

Since proline residues are dispersed all over the Sp23 molecule, large regions of regularly organized conformation (α -helices or β -sheets) are practically impossible. According to the sequence analysis software, the secondary structure of Sp23 consists of a small, NH₂-terminal, α -helix, small β -sheet domains, a lot of (proline-induced) turns, and some undetermined regions.

Comparison to Sequences in the Data Base—Hitherto, no amino acid composition of spermatophorins, other than Sp23, has been published. Cheeseman *et al.* (21) did however report the amino acid composition of the SDS-insoluble spermatophore fraction of the grasshopper *Melanoplus sanguinipis* and of the insoluble secretion of the short hyaline gland, which is believed to be the source of a major spermatophorin. The resemblance with the amino acid composition of Sp23 did not go further than the rather high content of glutamine/glutamic acid.

Comparing the Sp23 sequence with other structural insect proteins in the data bases did not reveal large domains of homology, either. There is some structural resemblance, however, with a few chorion proteins in fruit flies, termed S36 and S38 (22, 23). The central domains of these structural eggshell proteins contain tandemly repeated peptide motifs similar to the one observed in the carboxyl-terminal half of Sp23 (Fig. 4).

Significant sequence homology with polypeptides from sources other than insects was not found in the data bases. However, some resemblance in amino acid composition was observed with a group of plant cell wall proteins, termed extensions, which are, like Sp23, rich in proline and relatively rich in tyrosine (24, 25).

Interaction of Sp23 with Other Spermatophore Components—BAGs do not produce one homogeneous mixture containing all the spermatophorins necessary for the construction of the spermatophore. Instead they produce several separate mixtures, each of which has a defined selection of spermatophorins and contributes to a specific layer of the spermatophore wall and core. This implies that there must be forces that keep the components of a given mixture together from the moment they are produced, until they become a fixed part of the spermatophore. Shinbo *et al.* (8) observed aggregation of the PL21.1 antigen and suggested that freshly synthesized Sp23 molecules undergo polymerization, presumably by hydrophobic bonds. The forces that, directly or indirectly, connect the Sp23-polymers with other molecules (of the same mixture) possibly involve hydrogen bonds. At least, this would explain why the use of 8 M urea for the extraction of Sp23 from BAG homogenates, resulted in a substantially increased yield.² Fixation of Sp23-proteins in the spermatophore-und-

er-structure suggests the formation of firm, covalent cross-links. These cross-links are possibly generated in the tyrosine-rich regions of Sp23 molecules, by the peroxidase-catalyzed formation of di- or trityrosines, a common mechanism for fixation of insect structural proteins (26). Endogenous peroxidase activity was already reported in the cuticle of *T. molitor* (27), and di- and trityrosines were identified in the insoluble fraction of the *Tenebrio* eggshell (28).

Acknowledgments—We thank Christine Yuncker Happ for HPLC purification. We also acknowledge the assistance of Xavier Villarreal in amino-terminal sequence analysis.

REFERENCES

- Gadzama, N. M., and Happ, G. M. (1974) *Tissue Cell* **6**, 95–108
- Happ, G. M., Yuncker, C., and Huffnir, S. A. (1977) *J. Exp. Zool.* **200**, 223–236
- Black, P. N., Landers, M. H., and Happ, G. M. (1982) *Dev. Biol.* **94**, 106–115
- Paesen, G. C., Weyda, F., and Happ, G. M. (1992) *Insect Biochem.* **22**, 387–398
- Dailey, P. J., and Happ, G. M. (1983) *J. Morphol.* **178**, 139–154
- Grimnes, K. A., and Happ, G. M. (1986) *Insect Biochem.* **16**, 635–643
- Grimnes, K. A., Bricker, C. S., and Happ, G. M. (1986) *J. Exp. Zool.* **240**, 275–286
- Shinbo, H., Yaginuma, T., and Happ, G. M. (1987) *J. Biol. Chem.* **262**, 4794–4799
- Von Heijne, G. (1983) *Eur. J. Biochem.* **133**, 17–21
- Prevelige, P., and Fasman, G. D. (1989) in *Prediction of Protein Structure and the Principles of Protein Conformation* (Fasman, G. D., ed) pp. 391–416, Plenum Press, New York
- Chou, P. Y., and Fasman, G. D. (1978) *Adv. Enzymol.* **47**, 45–147
- Garnier, J., Osguthorpe, D. J., and Robson, B. (1978) *J. Mol. Biol.* **120**, 97–120
- Pinna, L. A. (1990) *Biochim. Biophys. Acta* **1054**, 267–284
- Woodgett, G. R., Gould, K. L., and Hunter, T. (1986) *Eur. J. Biochem.* **161**, 177–184
- Smith, S. G. (1960) *Can. J. Genet. Cytol.* **2**, 66–88
- Tung, J.-S., and Knight, C. A. (1972) *Anal. Biochem.* **48**, 153–163
- Cox, D. L., and Willis, J. H. (1987) *Insect Biochem.* **17**, 457–468
- Pau, R. N., Weaver, R. J., and Edwards-Jones, K. (1986) *Arch. Insect Biochem. Physiol.* **3**, 59–73
- Willis, J. H. (1987) *Arch. Insect Biochem. Physiol.* **6**, 204–215
- Talbo, G., Hojrup, P., Rahbek-Nielsen, H., Andersen, S. O., and Roepstorff, P. (1991) *Eur. J. Biochem.* **195**, 495–504
- Cheeseman, M. T., Gillott, C., and Ahmed, I. (1990) *J. Exp. Zool.* **255**, 193–204
- Konsolaki, M., Komitopoulou, K., Tolia, P. P., King, L. D., Swimmer, C., and Kafatos, F. C. (1989) *Nucleic Acids Res.* **18**, 731–737
- Aggeli, A., Hamodrakas, S. J., Komitopoulou, K., and Konsolaki, M. (1991) *Int. J. Biol. Macromol.* **13**, 307–315
- Chen, J., and Varner, J. E. (1985) *EMBO J.* **4**, 2145–2151
- Gatehouse, L. N., Evans, I. M., Gatehouse, J. A., and Croy, R. R. D. (1990) *Plant Sci.* **71**, 223–231
- Sugumar, M. (1988) *Adv. Insect Physiol.* **21**, 179–231
- Lemoine, A., and Delachambre, J. (1986) *Insect Biochem.* **16**, 483–489
- Kawasaki, H., Sato, H., and Suzuki, M. (1975) *Insect Biochem.* **5**, 25–34
- Levitt, M. (1978) *Biochemistry* **17**, 4277–4285
- Kann, M.-L., and Fouquet, J.-P. (1989) *Histochemistry* **91**, 221–226
- Chomczynski, P., and Sacchi, N. (1987) *Anal. Biochem.* **162**, 156–159
- Gilham, P. T. (1964) *J. Am. Chem. Soc.* **86**, 4982–4985
- Gubler, U., and Hoffman, B. J. (1983) *Gene (Amst.)* **25**, 263–269
- Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982) in *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
- Mierendorf, R. C., Percy, C., and Young, R. A. (1987) *Methods Enzymol.* **152**, 458–469
- Miller, H. (1987) *Methods Enzymol.* **152**, 145–170
- Sanger, F., and Coulson, A. R. (1975) *J. Mol. Biol.* **94**, 441–448
- Short, J. M., Fernandez, J. M., Sorge, J. A., and Huse, W. D. (1988) *Nucleic Acids Res.* **16**, 7583–7599
- Mierendorf, R. C., and Pfeffer, D. (1987) *Methods Enzymol.* **152**, 556–562
- Devereux, J., Haeblerli, P., and Smithies, O. (1984) *Nucleic Acids Res.* **12**, 387–395
- Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) in *Molecular Cloning: A Laboratory Manual*, 2nd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

² G. C. Paesen, M. B. Schwartz, M. Peferoen, F. Weyda, and G. M. Happ, unpublished observation.