ELSEVIER

Molecular and Biochemical Parasitology 70 (1995) 95–106

MOLECULAR AND BIOCHEMICAL PARASITOLOGY

Brugia spp. and *Litomosoides carinii*: Identification of a covalently cross-linked microfilarial sheath matrix protein (shp2)

Jörg Hirzmann [a], Achim Schnaufer [a], Martin Hintz [b], Franz Conraths [c], Stephan Stirm [b], Horst Zahner [c], Gerd Hobom [a,*]

[a] *Institut für Mikrobiologie und Molekularbiologie, Justus-Liebig-Universität Giessen, Frankfurter Straße 107, 35392 Giessen, Germany*
[b] *Institut für Biochemie, Justus-Liebig-Universität Giessen, 35392 Giessen, Germany*
[c] *Institut für Parasitologie, Justus-Liebig-Universität Giessen, 35392 Giessen, Germany*

Received 17 October 1994; accepted 31 December 1994

Abstract

A microfilarial sheath protein gene (*shp2*) coding for the major constituent of the insoluble, cross-linked sheath remnant (SR) from *Brugia malayi*, *Brugia pahangi* and *Litomosoides carinii* has been cloned and sequenced, based on peptide partial amino-acid sequences. All three closely related single-copy *shp2* genes in the two genera carry a single intron in identical position; shp2 mRNAs are post-transcriptionally modified by both *cis*-splicing and *trans*-splicing. In accordance with their extracellular destinations the encoded proteins include signal peptide sequences; molecular masses of approx. 23 kDa are hence predicted for the mature secreted polypeptides. In their structures sheath matrix proteins shp2 may be regarded as extreme cases of a modular constitution, since these proteins largely consist of two different segments of multiple sequence repetitions, PAA and QYPQAP (or QYPQ), separated by elements of unique sequence. Extreme insolubility and cross-linking are likely to originate from these repetitive sequences within shp2, and to constitute the basic properties of a microfilarial matrix largely consisting of an shp2 network.

Keywords: *Brugia malayi*; *Brugia pahangi*; *Litomosoides carinii*; Microfilarial sheath; *trans*-splicing; Cross-linking

1. Introduction

In several filarial genera, such as *Brugia* and *Litomosoides*, the first-stage larvae are enclosed by a bag-like structure, the so-called sheath. It originates in part from the primary eggshell by modification in surface structure and composition during intrauterine development of the embryo [1,2]. While trematode eggshell proteins and their modes of expression have been studied extensively [3,4], little is known on the molecular development and composition of nema-

Abbreviations: aa, amino acid(s); PCR, polymerase chain reaction; shp, sheath protein; SL, *trans*-spliced leader segment; SR, sheath remnant

Note: Nucleotide sequence data reported in this paper have been submitted to the EMBL, GenBank™ and DDJB data bases under the accession numbers Z35443, Z35444 and Z35445.

* Corresponding author. Tel. (49-641) 702-9651; Fax (49-641) 702-9659.

SSDI 0166-6851(95)00011-9

tode eggshells, and in particular of such modified structures like the microfilarial sheath.

Efforts to characterize microfilarial sheaths from *Brugia spp.* biochemically have been hampered by limitation in material. Therefore, the closely related rodent filaria *Litomosoides carinii* was used to establish the respective purification procedures based on more abundant resources [5–7]. From *L. carinii* sheath material we obtained partial amino-acid sequences of at least 6 major polypeptide constituents, five of which (shp1, 1a, 3, 3a and 4) could be solubilized after disintegration of intact sheaths by reduction of disulfide bonds [8]. The gene coding for *L. carinii* shp1 (formerly called gp22) has been isolated based on peptide sequence information [9], and shp1 proved to be homologous to Mf22 of *B. pahangi* [10]. Since *Brugia spp.* sheaths do contain at least one further protein, which is closely related to a sheath polypeptide of *L. carinii*, negative staining shp3 [11], a substantial amount of homology between the proteins of these microfilarial sheaths may be expected in general.

At least one major sheath component, however, remained insoluble under the conditions applied [7,8], and was found retained in the sheath remnant (SR) after treatment. Starting with proteolytic peptides obtained from SR a second sheath protein: shp2 could be identified as a major component of the insoluble SR material. In this paper we describe the isolation and characterization of three such *shp2* genes, from *L. carinii, B. malayi* and *B. pahangi.*

2. Materials and methods

2.1. Parasites, infections and parasite isolation

Litomosoides carinii was maintained in cotton rats and *Mastomys coucha.* Animals were infected by allowing infected mites (*Bdellonyssus bacoti*) to suckle [12]. *Brugia malayi* (sub-periodic strain) and *B. pahangi* infections were raised in *M. coucha* by subcutaneous injection of 85 infective larvae each, isolated from *Aedes aegypti* [13]. Adult *L. carinii* were recovered from the pleural cavity 100 days post infection (p.i.). Adult *Brugia* worms were obtained from the lungs and hearts of the animals 120 days p.i. Microfilariae of *L. carinii* were isolated from the blood of highly microfilaraemic animals by Percoll gradient centrifugation [14].

2.2. Preparation of DNA and RNA

Live worms were washed in saline and quickly frozen in liquid nitrogen. DNA was extracted by a standard protocol including RNaseA treatment, proteinase K digestion and phenol/chloroform extraction [15]. Total RNA was isolated by acid guanidinium thiocyanate/phenol/chloroform extraction [16], and poly(A)$^+$ RNA was prepared with $(dT)_{25}$-paramagnetic beads (Dynal) according to producer's instructions.

2.3. RT-PCR amplification, cloning and sequencing

A *Bam*HI-$(dT)_{17}$ or *Sal*I/*Xho*I-$(dT)_{20}$ oligonucleotide was used to prime the first-strand cDNA synthesis from 1 μg poly(A)$^+$ RNA by MMLV-Superscript reverse transcriptase (Gibco BRL) in a 20 μl reaction volume. Excess oligo(dT) primer was removed by glass milk purification (own preparation) followed by ethanol precipitation. An aliquot of this product was used directly for PCR. For amplification of mRNA 3'-ends a degenerate mixture of oligonucleotides as deduced from (partial) peptide sequences [8] was employed as a 5'-specific primer. For amplification of mRNA 5'-ends the first-strand cDNA product was extended by homopolymer tailing with terminal deoxynucleotidyl transferase (Gibco BRL) in the presence of 0.2 mM dGTP. Subsequently a non-specific *Eco*RI-*Bam*HI-$(dC)_{12}$ anchor oligonucleotide [17] together with a degenerate mixture of sequence-derived anti-sense primers was used. Alternatively, an anchor-primer corresponding with the 5' *trans*-splice leader segment of *B. malayi* [18] was chosen, consisting of its first 18 (out of 22) nt [9]. PCR amplification was performed using *Taq* polymerase (Promega), and amplification products were subcloned either in T-vectors prepared as described by Marchuk et al. [19], or in conventional plasmid sequencing vectors after fragment-terminal restriction. Sequencing was done by the dideoxy method using Sequenase (US Biochemical) according to the manufacturer's protocol. A λgt11 cDNA

library from adult *B. pahangi* was a kind gift from M. Selkirk (London).

2.4. Genomic libraries, Southern blot analysis

For preparation of genomic libraries high molecular mass DNA was partially digested by *Mbo*I, and DNA fragments were size-fractionated via gel electrophoresis, ligated to λEMBL3-vector arms (*Bam*HI digested), and packaged using Gigapack II Gold packaging extracts (Stratagene). For Southern analysis, filarial DNA was cleaved with various restriction enzymes; fragments were separated on 1.0% agarose and blotted by capillary transfer to Qiabrane Plus nylon membrane (Qiagen). Screening of libraries and Southern blot analysis were performed in parallel with selected cDNA coding region gene probes, labelled either with [α-^{32}P]dCTP (Amersham) or dig-dUTP (Boehringer-Mannheim) by random priming [20]. Hybridization of probes was detected by autoradiography or chemiluminescence, respectively.

2.5. Northern blot analysis

Approximately equal amounts of isolated filarial RNAs were fractionated in parallel on a 1.5% agarose-formaldehyde gel, transferred to Hybond N nylon membrane (Amersham) and attached to it by cross-linking under UV light. To demonstrate equal loading, the RNA blot was stained with methylene blue [21]. Hybridization was performed in a formamide-free system [22] using a cDNA fragment labelled with [α-^{32}P]dCTP by random priming.

3. Results

3.1. Isolation of shp2 cDNA clones

For cloning of cDNAs encoding the *L. carinii shp2* gene, we performed PCR amplification on reverse transcribed poly(A)$^+$ RNA using an oligo(dT) primer and degenerate sense-strand oligonucleotide primers corresponding to the N-terminal sequences of two isolated peptides. These had been obtained from proteolytic digests of total sheath material, but could be assigned to originate from partial proteolysis of the SR [8]; primers Ch6, 5′-tatctcgagTGKTAYCCDCCDATGTA-3′, and Ch7, 5′-tatctcgagCAGGGDCAAGCDCCDGC-3′. Following linear amplification at 40°C (annealing temperature) for three cycles with only the degenerate primer present, the anchor *Bam*HI-(dT)$_{17}$ primer was added, too, and PCR was continued for another 37 cycles at 58°C. Amplification resulted in two prominent products using primer Ch6, and in 5 fragments with primer Ch7. Sequencing revealed that in either amplification only one of the fragments showed the primer sequence to be extended correctly into distal codons corresponding to the known polypeptide sequences. Both cDNA clones isolated are largely overlapping, with both chymotryptic peptides belonging to the same open reading frame. In addition, most of the other SR peptides analyzed by Hintz et al. [8] could also be assigned to the same C-terminal region of the shp2 protein sequence as deduced from the partial cDNA clones isolated.

For isolation of a complete *shp2* cDNA clone a specific antisense oligonucleotide corresponding to codons 134–138 (5′-tccccgcggTTGCACGGGAGGCC-3′), together with a poly(dC) oligonucleotide primer served in a second PCR amplification, following dG-tailing of first-strand cDNA. A single fragment was isolated and subcloned. All full-length clones obtained started (behind a dC primer stretch) with the 22-nt *trans*-splice leader sequence (SL1) proving creation of the mature *shp2* transcript by trans-splicing. The N-terminal *shp2* nucleotide sequence codes for another two of the chymotryptic sheath fragments as described by Hintz et al. [8]. The full-size *L. carinii shp2* cDNA clone spans 1033 bp including 22 nt of the leader, as well as 48 and 294 nt of untranslated 5′ and 3′ sequences, respectively (Fig. 3A).

Amplification of the homologous sequences from *Brugia spp.* was achieved by a similar approach. The combination of primer Ch6 with an oligo(dT) primer proved appropriate also for isolation of the *B. malayi shp2* cDNA 3′ segment from RNA, and the *B. pahangi shp2* segment was analogously obtained using as template a cDNA library. An aliquot of the bacteriophage library was incubated for 10 min at 70°C and used directly for PCR amplification. Since we expected the *Brugia spp. shp2* mRNAs to be *trans*-spliced, an oligonucleotide primer representing

the first 18 nt of the 22-nt *trans*-splice leader sequence (5′-tatctcgaGGTTTAATTACCCAAGTT-3′) was chosen for amplification of the 5′ end of both *Brugia* cDNAs, together with a specific antisense oligonucleotide derived from a common region in the two *Brugia* 3′ segments (5′-atagtcgACGTAAGAAGTCTATTTCTGC-3′). Internal unique restriction sites served for combining the three 5′ cDNA segments with their corresponding 3′ cDNA fragments to yield full-size cDNA clones.

The sequences determined for the amplified cDNAs were confirmed over their entire lengths by sequencing the corresponding genomic clones (see below). The isolated cDNAs of *B. malayi* and *B. pahangi* including the *trans*-spliced leader segments comprise 947 and 920 bp, respectively; 59 and 153 (or 332/333, see below) nt of the *shp2* mRNAs remain untranslated in the 5′ and 3′ regions, in both cases. The *Brugia* sequences are almost identical to each other (95% homology). The similarity between *B. malayi* and *L. carinii* cDNAs accounts to 65% over their entire lengths, but reaches as much as 90% in the non-repetitive segments of the coding regions (see Fig. 4).

3.2. Northern blot analysis

In order to confirm the size of the mature *shp2* transcript and to determine in which stage of the parasite it is expressed, total RNA from *L. carinii* was isolated from adult females, adult males and microfilariae. Approximately equal amounts were separated by gel electrophoresis (Fig. 1A) and blotted. A 3′ fragment (*Pfl*MI-*Nsi*I) of the cDNA coding sequence used as hybridization probe recognized an abundant (and another weak) transcript in females, but never reacted with microfilarial RNA (Fig. 1B). In male RNA either no signal or at most a very weak signal was observed, in a minority of experiments, and was regarded to be unspecific. The size of the prominent *shp2* transcript in female RNA is around 1150 nt in length, and thus increased by 100 nt over the size as deduced from the cDNA sequence, attributable to the poly(A) tail of the mRNA. The additional, weak signal in the female RNA lane at 3 kb cannot be explained in relation to the *shp2* gene and, therefore, may result from an unrelated cross-reacting sequence.

3.3. Genomic organization of genes encoding shp2

For Southern hybridization analysis various restriction enzyme digests of *L. carinii* DNA were separated by gel electrophoresis, blotted, and probed with a cDNA fragment from the C-terminal shp2 coding region (primer Ch6 to *Nsi*I restriction site). *Nsi*I, *Hin*dIII and *Eco*RI patterns showed single responsive bands of 0.9, 15 and 5.4 kb, respectively. Genomic DNA from *B. malayi* and *B. pahangi* cleaved with *Bcl*I, *Eco*RI or *Nsi*I and probed with an equivalent *B. pahangi shp2* 3′-cDNA fragment, yielded single bands identical in both species of 0.8, 10 and 0.9 kb, respectively (Fig. 2). From these results we conclude that *shp2* is likely to be encoded as a unique gene in both genera.

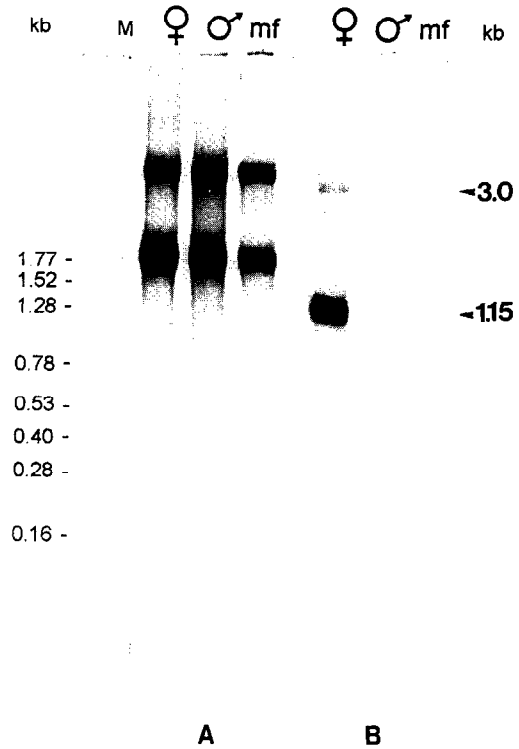

Fig. 1. Northern blot analysis. Total RNAs from *L. carinii* female and male adults as well as blood microfilariae (mf) were separated on a 1.5% formaldehyde-agarose gel (20 μg RNA per lane). The blot was stained with methylene blue (left-hand four lanes; panel A) to confirm the integrity and to compare the quantity of the transferred RNAs, and was hybridized (left-hand four lanes; panel B) with a radiolabelled *shp2* cDNA fragment. The sizes of RNA molecules were estimated by comparison with a co-electrophoresed RNA ladder (lane M, leftmost lane).

Genomic libraries of *L. carinii, B. malayi* and *B. pahangi* adult worms were constructed in bacteriophage vector λEMBL-3 from size-fractionated, partial *Mbo*I digests; the resulting libraries consisted of $(1-2) \times 10^6$ primary phages each. Screening of 6×10^4 plaques from amplified libraries with labelled cDNA probes (see Materials and methods) was sufficient to obtain several positive clones in all three cases. Two clones from each species were further plaque-purified and finally analysed in parallel by restriction enzyme hydrolysis and Southern blotting. The two corresponding clones always yielded equivalent results. After subcloning of the isolated genomic DNA segments, the entire sequences of all three *shp2* genes including proximal and distal flanking regions were obtained by dideoxy-sequencing of both DNA strands in overlapping fashion (Fig. 3).

In each of the three species the *shp2* gene is comprised of 2 exons. The intron segment interrupts the coding sequence of all three species at exactly the same basepair position, i.e., between the first and second G of glycine codon 26. The intron sequences of the two genera are similar in size, 305 bp in *B. malayi* (299 bp in *B. pahangi*) and 254 bp in *L. carinii*, but are much less homologous to each other than the contiguous exon sequences.

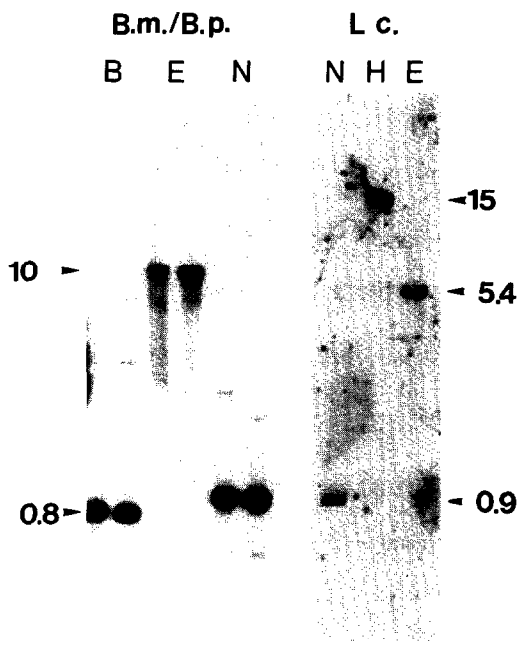

Fig. 2. Southern blot analysis of *B. malayi, B. pahangi* and *L. carinii* genomic DNAs hybridized with a cDNA 3′ fragment of the respective *shp2* gene. Each track contains 5 μg of DNA cleaved by restriction endonucleases: *Bcl*I (B), *Eco*RI (E), *Hin*dIII (H) or *Nsi*I (N).

3.4. Modular structure of shp2 polypeptides

The protein structures of shp2, as deduced from genomic DNA and cDNA sequences, are in part confirmed independently for *L. carinii* by amino-acid sequences derived from several of the sheath peptides, and further in their correspondence to sheath or SR fragmentation patterns by proteases, as observed for their N-terminal and C-terminal regions. No shp2 fragments are found in SDS/β-ME soluble sheath protein fractions after their proteolysis, but shp2 peptides are dominant in total sheath digests and constitute almost all of the SR peptides, demonstrating both the relative amount and the insolubility of protein shp2 [8].

Beginning with a proximal methionine codon in standard initation context [23], the open reading frames of the three shp2 genes encode polypeptides of 237 aa in *B. malayi*, 228 aa in *B. pahangi* and 222 aa in *L. carinii*, respectively (Fig. 4). All three shp2 proteins as deduced from DNA result in unusual amino-acid compositions, which however for *L. carinii* is in close agreement with the amino-acid analysis of total SR material, and is even reflected in the entire sheath composition, with regard to an indicative P/Q/A/Y ratio as observed earlier [6,7,24]. From these data we conclude that the insoluble SR matrix is likely to consist mainly of shp2 subunits. The minor deviations remaining between amino-acid composition of SR and deduced constitution for shp2 might be attributed to some other major sheath protein (i.e., shp1) being partially retained in a network of cross-linked shp2 units.

On the basis of structural elements and homology, the shp2 polypeptide chains may be divided into six different domains (see Fig. 4). The amino terminus extends into a standard secretory signal sequence (aa 1–19; first region), very similar in all three species. A positively charged amino acid (Arg) at position 6 is followed by a stretch of hydrophobic amino acids, of which the last one is a helix-breaking proline. The putative cleavage site is not completely certain, but

as defined by the algorithm of von Heijne [25] is most likely located between Cys^{19} and Tyr^{20}. The mature polypeptides (in the absence of further processing reactions) thus have predicted molecular masses of 23.6 kDa in *B. malayi*, 22.7 kDa in *B. pahangi* and 22.2 kDa in *L. carinii*; experimental data are missing because of insolubility.

The second region (*B. malayi / B. pahangi* aa 20–53, *L. carinii* aa 20–57) is largely identical between the two genera, without a distinct structural motif. Interestingly, three conserved Lys are found in that region, with Lys^{35} being interspersed between two Gly residues.

The third region of shp2 (*B. malayi* aa 54–100, *B. pahangi* aa 54–99, *L. carinii* aa 58–104) consists almost exclusively of Ala and Pro residues, in lose adherence to a PAA repeat motif. With its very hydrophobic and repetitive structure region III is likely to contribute to the insolubility of the entire protein. An almost regular spacing of proline residues interspersed with only the methyl-group side-chains of Ala may also give rise to specific secondary structures which, in intermolecular contact, might further increase insolubility of shp2 aggregates.

A short region IV (*B. malayi* aa 101–118, *B. pahangi* aa 100–117, *L. carinii* aa 111–130) located

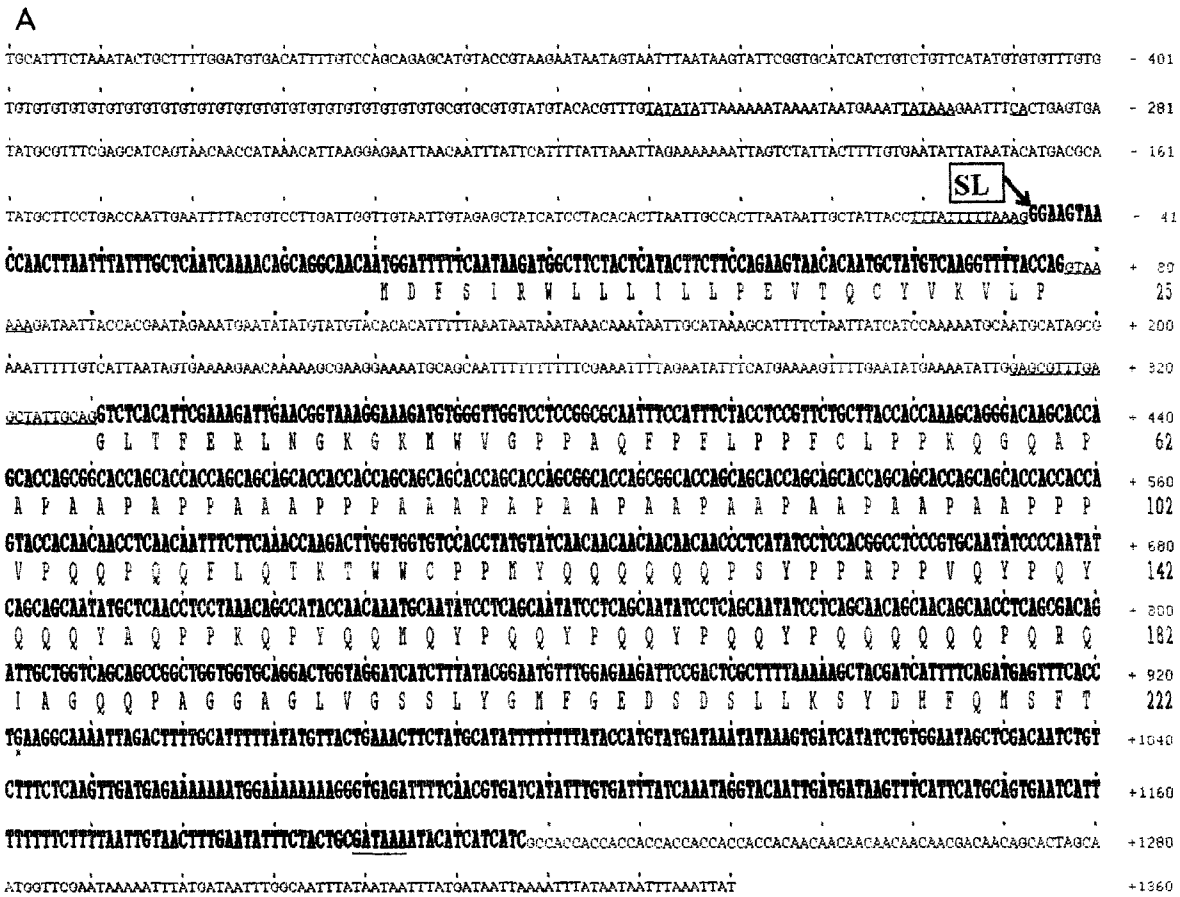

Fig. 3. (A) Nucleotide sequence and coding regions of the *L. carinii shp2* gene. Nucleotides of the two exon segments are shown in large, bold type. The amino-acid sequence derived from their open reading frames is given below the exon sequences. Nucleotide and amino-acid numbering systems are in reference to the translation start codon and its first nucleotide (A). Putative TATA boxes, the *trans*-splice acceptor sequence, as well as both cis-splice signals and the putative poly(A) signal are underlined. The position at which the addition of the *trans*-splice leader sequence (SL) is observed in the amplified cDNA is indicated by an arrow pointing to the first nucleotide of exon 1.

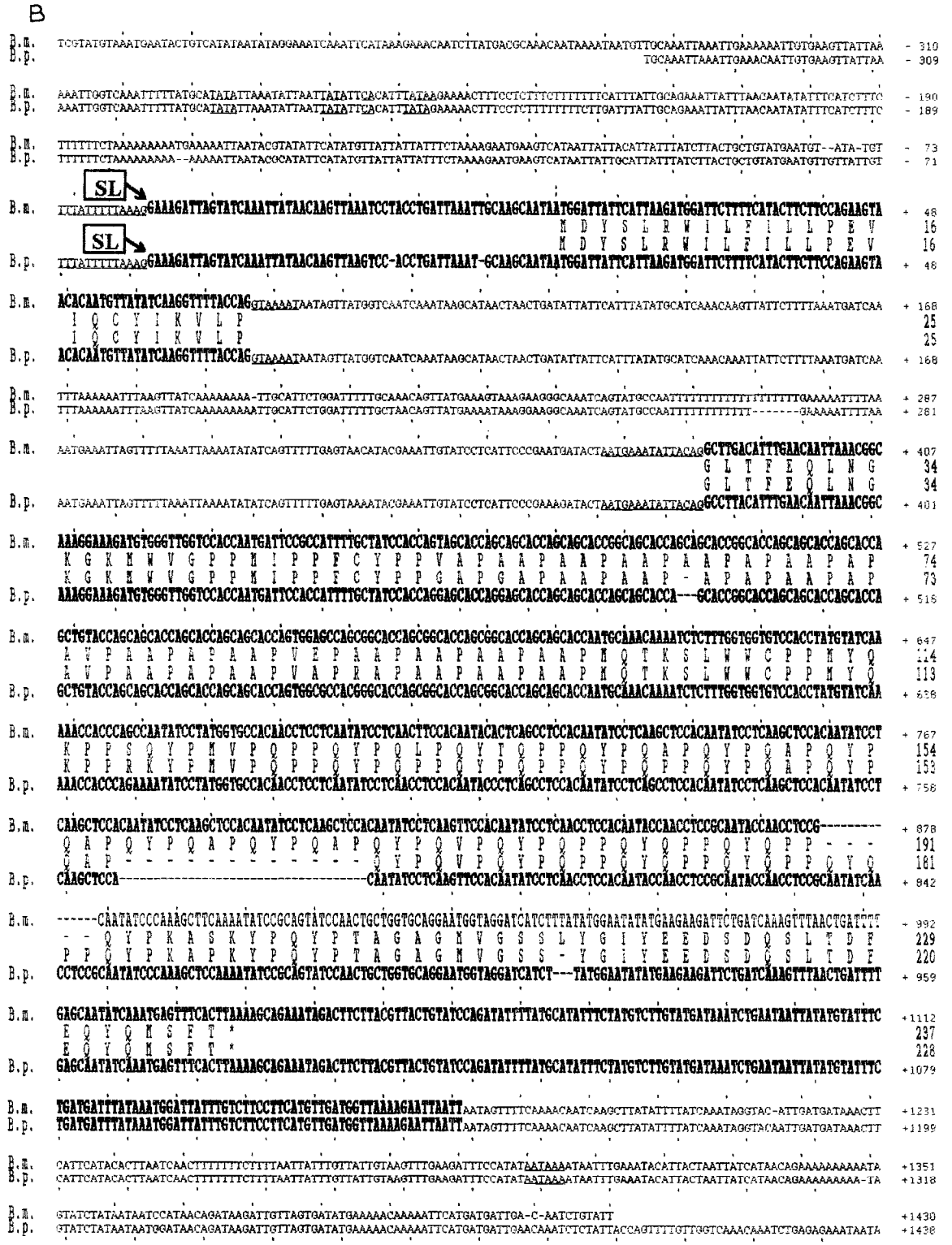

Fig. 3. (B) Nucleotide sequences and coding regions of the *B. malayi* and *B. pahangi shp2* genes. Gaps have been inserted to improve the alignment between the two sequences. Regions of the genes are organized and pointed out as described above, including underlignment of a putative polyadenylation signal at position 1300 (1268), while the 3′ end of the longest clones isolated is located at position 1171 (1138).

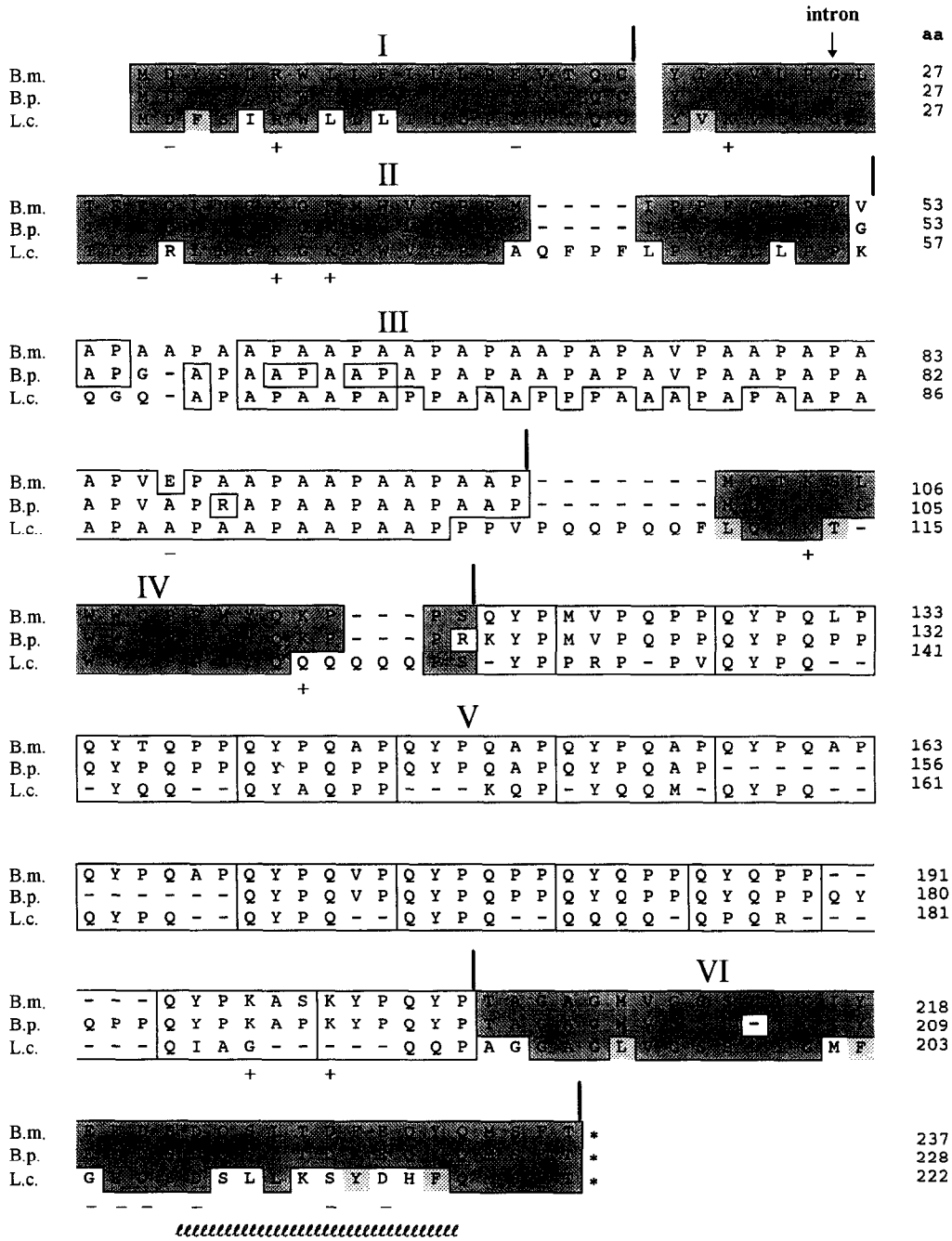

Fig. 4. Amino-acid sequence comparison between the shp2 proteins from *B. malayi*, *B. pahangi* and *L. carinii*. Gaps have been inserted to improve regional homology. Six different modular regions (I–VI), as described in the text, are indicated and separated by vertical bars. In regions I, II, IV and VI identical positions are boxed and shaded. The repeat units in region V are marked by boxes. The putative signal peptidase cleavage site (gap) and the known intron position identical for all three species (arrow) are indicated. For *B. malayi* charged residues (+, −) and a predicted C-terminal α-helix (*ℓℓℓℓℓ*; PHD method [40]) are shown below the sequence comparison.

between two segments of repetitive sequences, is again highly conserved among all three species. It includes another constant Lys residue flanked by Thr and Ser, and a prominent hydrophobic sequence element, WWCPPMY.

A rather large region V (*B. malayi* aa 119–203, *B. pahangi* aa 118–195, *L. carinii* aa 131–188) very prominently displays a total of 14 sequence repetitions, variations of a single structural motif. In *B. malayi* and *B. pahangi* this motif consists of hexapeptide QYPQAP, while in *L. carinii* a shorter, but related element QYPQ takes its place (see Fig. 4). The high content of glutamine in both cases may favour the involvement of this region in ϵ-(γ-glytamyl)lysine isodipeptide cross-links, most likely together with the conserved lysine residues mentioned above for segments II and IV. Such cross-links have been detected in *L. carinii* sheaths [26] and are expected to contribute to the insolubility of the resulting shp2 network.

C-terminal segment VI is again largely conserved between *Brugia* and *Litomosoides* (80% homology). It is Gly + Ser-rich and negatively charged due to four to six acidic residues.

3.5. Structural elements within the shp2 genes

The genomic sequences of *B. malayi* and *B. pahangi shp2* genes are more than 95% identical over 1.8 kb (Fig. 3B), with most of the few deviations resulting only from a different number of repetitions in segment V.

Both observations, interspecies conservation and differences in the numbers of repetitions, suggest an intragenic duplication process to have created the repetitive sequences of *shp2* gene segment V. If duplications accumulated in a stepwise manner, those that occurred more recently might be expected to result in coherent groups of repeats resembling each other most closely. Consecutive repeats 5–9 of *B. malayi* for example are identical to each other and thus may have evolved recently. Similarly, *B. malayi* repeats 12 and 13, as well as *B. pahangi* repeats 10–12 (QYQPP) are again identical to each other, but differ from repeats 5–9 in three nucleotide substitutions plus the deletion of one entire codon. Equally, in *B. pahangi*, consecutive repetitions 2–5 (QYPQPP) are identical to each other, but deviate by single G → C substitutions in nucleotide position 13. In *L. carinii*, the QYPQ repetitive element (repeats 9–11) is identical in DNA sequence to the first eleven nt of the *Brugia* QYPQAP repeats, with the terminal six basepairs being deleted entirely from the shorter *L. carinii* element.

The proximal *trans*-splice-acceptor site (TTTAT-TTTTAAAG | G) is present in the three species without any deviation, while the 5′ flanking sequences, when aligned via their ATG translation initiation codons, show little obvious similarity. The overall homology within 450 bp of upstream sequence is only 36% and is mainly restricted to the *trans*-splice acceptor sequence and to two candidate promotor TATA boxes, underlined in Fig. 3A,B. No sequence regions matching standard *cis*-acting control elements are observed in the putative promoter regions or become evident because of homology. In *L. carinii*, a tandem repeat of 25 GT dinucleotides at position −403 appears to set a limit to the *shp2* promoter region, just 15 bp upstream of the aforementioned candidate TATA-box.

Based on a comparison with *L. carinii shp2* cDNA, we like to predict for the *Brugia* species that the isolated cDNA clones may have been incomplete at their 3′ ends. The longest *B. malayi* cDNA clone obtained terminates at position 1171, 152 bp downstream of the TAA termination codon. This position is not correlated with an upstream consensus poly(A) signal, however. Instead, a first AATAAA poly(A) signal sequence can be identified at position 1300 (see Fig. 3B), which exactly agrees with the 3′ end of several *L. carinii* full-size cDNA clones and the (irregular) *L. carinii* signal sequence GATAAA at this position, 19 bp upstream of the poly(A) tail segments. Although not common in vertebrates [27], a number of deviations from AATAAA have been found in nematodes including the *L. carinii* or *B. pahangi* shp1 gene: AGTAAA [9,10], as well as several *C. elegans* genes (cited in Ref. 28).

4. Discussion

On the basis of N-terminal amino-acid sequences of chymotryptic fragments obtained from total *L. carinii* microfilarial sheaths, we have isolated and sequenced the gene coding for a second sheath pro-

tein (shp2) in *L. carinii, B. malayi* and *B. pahangi.* Calculations based on amino-acid analyses of both, total sheaths and SR (M.H., unpublished) confirm that shp2 is one of the main constituents of the sheath and a dominant component within the insoluble sheath fraction. In addition to its concentration the role of shp2 as a structural matrix protein appears to be supported by its two prominent repeat regions of different character (PAA and QYPQAP) which may serve in providing a hydrophobic nucleus and a substrate region for intermolecular interlinkage, respectively. A series of covalent interactions between shp2 subunits is expected to result in an insoluble matrix network underlying the sheath structure, which should also be contacting and supporting other components of the sheath. Also, immunhistological studies (F.J.C., unpublished) confirm that protein shp2 is located in the sheath matrix.

Covalent interlinkage between shp2 molecules is very unlikely to result from cystine bridges, as shp2 (or SR) is insensitive to mercaptoethanol, even at high temperatures, and since only two Cys residues are present in the mature shp2 sequence. While also dityrosine linkages have been excluded, analysis of non-reducible covalent cross-links in *L. carinii* microfilarial sheaths revealed that they contain ε-(γ-glutamyl)lysine bonds [26]. A most likely candidate to serve as a substrate for these cross-links appears to be shp2, due to the presence of Gln residues in large numbers and to several Lys in the expected structural context [29,30], i.e., with small side-chain residues in flanking positions. Also, the extreme insolubility of shp2 would be in agreement with an interlinked network of protein subunits resulting from such covalent cross-links. Recently, transglutaminase, an enzyme catalyzing the formation of such ε-(γ-glutamyl)lysine isopeptide bonds was identified in adult female worms of *B. malayi* [31], and it was found that inhibition of this enzyme arrests maturation of *B. malayi* microfilariae [32].

Insolubility of shp2 may also result from structural properties of the protein subunit itself and from non-covalent interactions. Within the modular structure of shp2, the P + A-rich segment III may serve this function in two ways. First, it is very hydrophobic in general, and second, the prolines within the repetitive sequence may force the polypeptide chain into a special secondary structure such as a narrow elongated helix, in particular since the methyl groups of the alanines would not pose a barrier against such a conformational deviation. This may perhaps favour intermolecular aggregation of two or more shp2 molecules which could further contribute to the insolubility of shp2.

Similar P/A repeats are found in *Drosophila melanogaster* defective chorion-I protein precursor (dec-1) and a component (SV23) of the vitelline membrane, in a chorion protein (S-36) of *Ceratitis capitata*, in arabino galactan proteins (AGPs) of various plants, in larval cuticular proteins (Pr-8, -37 and -38) of *Locusta migratoria*, and most interestingly also in the cuticlins CUT-1 and CUT-2 of *Caenorhabditis elegans* [33–35]. The latter finding suggests a relationship between shp2 and nematode cuticlins in general. It is also worth noting, that an interspersion of such P/A repeats with stretches comprising Tyr residues, as present in *C. elegans* CUT-2, but neither in CUT-1 nor in shp2, appears to correlate with the formation of dityrosine cross-links [33].

With insolubility being such a dominant feature of the shp2-derived matrix structure it may be appropriate to ask the question of how shp2 molecules can be synthesized so that they will stay soluble until they will reach their final position and become integrated into an shp2 matrix network. We propose that the shp2 segment V with its QYPQAP repetitive structure may be similar to the functional role of repetitive protein segments (e.g., PQGGYQQYN [36,37]) in the so-called prion proteins, with the insoluble material catalysing the transition of additional units from the soluble to the insoluble state. Two alternative conformations for a single protein may be stabilized by such extended, Gln-rich repeats because of a particularly large number of hydrogen bonds supported by the glutamine carboxamide side-chains together with Tyr hydroxyl groups, in the presence of Pro residues. Cross-linking reactions catalysed by transglutaminase may, therefore, not only establish such covalent bonds to other shp2 subunits, but at the same time initiate conformational changes extending over the entire Gln + Pro-rich repetitive segment and the shp2 subunit altogether, when it becomes integrated into a growing shp2 extracellular network.

The model proposed is somewhat reminiscent of

the assembly of a fertilization envelope in sea urchin eggs. Here, an egg surface transglutaminase is activated within the first 4 min after fertilization to catalyze, via cross-linking reactions, the conversion of soluble vitelline subunits within the surface layer into a cross-linked, highly insoluble fertilization envelope of low permeability, accompanied by a major conformational change of the protein [38].

Sheath matrix protein shp2 appears to be located in the lower homogeneous layer of the microfilarial sheath, since it is accessible by shp2-specific antisera only upon pretreatment with chymotrypsin (F.J.C., unpublished observations). As observed by in situ hybridization with an antisense RNA probe comprising the entire coding region of *shp2* cDNA, expression of the *shp2* mRNA inspite of the final destination of that protein is located not in the embryo, but in uterine epithelial cells (J.H., unpublished data).

While all microfilarial sheath protein mRNAs isolated so far: shp1 [9], shp2, shp3 and shp3a (unpublished results) similarly undergo *cis*-splicing and *trans*-splicing reactions during their maturation, they are expressed in different cells, oocyte and embryo [39] or uterine layer cells. Hence, their individual promoters may not constitute a closely related set of regulatory elements, but are more likely to belong to different, more complex control circuits. Such regulatory interactions must be able to organize the development of an integrated sheath structure, even if based on *shp* gene expression in a number of cells in different developmental stages. Although a large number of protein molecules is produced in a rather short time, only unique genes have been discovered, in every case of an *shp* gene so far. It is unknown at present whether any of the heterogeneous, repetitive elements observed in the 3′ segments of *shp* mRNAs may be involved in a regulatory control over mRNA stability and frequency of translation to assist in that result.

Acknowledgements

This article represents in part the Ph.D. thesis of J.H. We gratefully acknowledge the expert technical assistance of M. Hoffmann, B. Hofmann, N. Pohl and U. Ruppert. This work has been supported by the Deutsche Forschungsgemeinschaft via SFB 272.

References

[1] Schraermeyer, U., Peters, W. and Zahner, H. (1987) Formation by the uterus of a peripheral layer of the sheath in microfilariae of *Litomosoides carinii* and *Brugia malayi*. Parasitol. Res. 73, 557–564.

[2] Zaman, V. (1987) Ultrastructure of *Brugia malayi* egg shell and its comparison with microfilarial sheath. Parasitol. Res. 73, 281–283.

[3] Lo-Verde, P.T. and Chen, L. (1991) *Schistosoma* female reproductive development. Parasitol. Today 7, 303–308.

[4] Rice-Ficht, A.C., Dusek, K.A., Kochevar, G.J. and Waite, J.H (1992) Eggshell precursor proteins of *Fasciola hepatica*, I. Structure and expression of vitelline protein B. Mol. Biochem. Parasitol. 54, 129–142.

[5] Bardehle, G., Klonisch, T., Schott, H.H., Stirm, S. and Zahner, H. (1987) Isolation of pure sheaths of *Litomosoides carinii* microfilariae. Parasitol. Res. 74, 188–190.

[6] Klonisch, T., Bardehle, G., Linder, D., Boschek, B., Schott, H.-H., Zahner, H. and Stirm, S. (1991) The sheaths of *Brugia microfilariae*: isolation and composition. Parasitol. Res. 77, 448–451.

[7] Bardehle, G., Jepp-Libutzki, A., Linder, D., Moehnle, K., Schott, H.-H., Zahner, H., Zähringer, U. and Stirm, S. (1992) Chemical composition of *Litomosoides carinii* microfilarial sheaths. Acta Trop. 50, 237–247.

[8] Hintz, M., Hirzmann, J., Hobom, G., Linder, D., Lottspeich, F., Schott, H.-H., Conraths, F.-J., Zahner, H. and Stirm, S. (1994) *Litomosoides carinii* microfilarial sheaths: partial amino acid sequences of several major polypeptide constituents. Mol. Biochem. Parasitol. 67, 69–77.

[9] Christ, H., Hirzmann, J., Conraths, F., Zahner, H., Stirm, S. and Hobom, G. (1992) Trans-splicing of an early embryo mRNA in *Litomosoides carinii*, coding for the major microfilarial sheath protein gp22. Gene 121, 219–226.

[10] Selkirk, M.E., Yazdanbakhsh, M., Freedman, D., Blaxter, M.L., Cookson, E., Jenkins, R.E. and Williams, S.A. (1991) A proline-rich structural protein of the surface sheath of larval *Brugia* filarial nematode parasites. J. Biol. Chem. 266, 11002–11008.

[11] Zahner, H., Hobom, G. and Stirm, S. (1995) The microfilarial sheath and its proteins. Parasitol. Today 11, 116–120.

[12] Lämmler, G., Saupe, E. und Herzog, H. (1968) Infektionsversuche mit der Baumwollrattenfilarie *Litomosoides carinii* bei *Mastomys natalensis* (Smith, 1834). Z. Parasitenk. 30, 281–290.

[13] Sänger, I., Lämmler, G. and Kimmig, P. (1981) Filarial infections of *Mastomys natalensis* and their relevance for chemotherapy. Acta Trop. 38, 277–288.

[14] Chandrashekar, R., Rao, U.R., Rajasekariah, G.R. and Subrahmanyam, D. (1984) Separation of viable microfilariae free of blood cells on Percoll gradients. J. Helminth. 58, 69–70.

[15] Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular Cloning. A Laboratory Manual, 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

[16] Chomczynski, P. and Sacchi, N. (1987) Single-step method

of RNA Isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal. Biochem. 162, 156–159.

[17] Hirzmann, J., Luo, D., Hahnen, J. and Hobom, G. (1993) Determination of messenger RNA 5′-ends by reverse transcription of the cap structure. Nucleic Acids Res. 21, 3597–3598.

[18] Takacs, A.M., Denker, J.A., Perrine, K.G., Maroney, P.A. and Nilsen, T.W. (1988) A 22-nucleotide spliced leader in the human parasitic nematode *Brugia malayi* is identical to the trans-spliced leader exon in *Caenorhabditis elegans*. Proc. Natl. Acad. Sci. USA 85, 7932–7936.

[19] Marchuk, D., Drumm, M., Saulino, A. and Collins, F.S. (1990) Construction of T-vectors, a rapid and general system for direct cloning of unmodified PCR products. Nucleic Acids Res. 19, 1154.

[20] Feinberg, A.P. and Vogelstein, B. (1983) A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. 132, 6–13 (Addendum (1984) Anal. Biochem. 137, 266–267).

[21] Wilkinson, M., Doskow, J. and Lindsey, S. (1990) RNA blots: staining procedures and optimization of conditions. Nucleic Acids Res. 19, 679.

[22] Yang, H., McLeese, J., Weisbart, M., Dionne, J.-L., Lemaire, I. and Aubin, R.A. (1993) Simplified high throughput protocol for Northern hybridization. Nucleic Acids Res. 21, 3337–3338.

[23] Cavener, D.R. and Ray, C. (1991) Eukaryotic start and stop translation sites. Nucleic Acids Res. 19, 3185–3192.

[24] Bardehle, G. Hintz, M., Linder, D., Schares, G., Schott, H.-H., Stirm, S. and Zahner, H. (1992) *Litomosoides carinii*: extraction of the microfilarial sheath components and antigenicity of the sheath fractions. Parasitol. Res. 78, 501–508.

[25] von Heijne, G. (1986) A new method for predicition of signal sequence cleavage sites. Nucleic Acids Res. 14, 4683–4690.

[26] Tarcsa, E., Eckerstorfer, M., Breitenbach, M., Hintz, M., Schott, H.-H., Zahner, H. and Stirm, S. (1992) ϵ-(γ-Glutamyl)lysine cross-links in *Litomosoides carinii* microfilarial sheaths. Parasitol. Res. 78, 623–624.

[27] Wickens, M. (1990) How the messenger got its tail. Trends Biochem. Sci. 15, 277–281.

[28] Spieth, J., Brooke, G., Kuersten, S., Lea, K. and Blumenthal, T. (1993) Operons in *Caenorhabditis elegans*: Polycistronic mRNA precursors are processed by trans-splicing of SL2 to downstream coding regions. Cell 73, 521–532.

[29] Molhuizen, H.D.F., Alkemade, H.A.C., Zeeuwen, P.L.J.M., de Jongh, G.J., Wieringa, B. and Schalwijk, J. (1993) SKALP/elastin: an elastase inhibitor from cultured human keratinocytes. J. Biol. Chem. 268, 12028–12032.

[30] Hohl, D., Mehrel, T., Lichti, U., Turner, M.L., Roop, D.R. and Steiner, P.M. (1991) Characterization of human loricrin. J. Biol. Chem. 266, 6626–6636.

[31] Mehta, K., Rao, U.R., Vickery, A.C. and Fesus, L. (1992) Identification of a novel transglutaminase from the filarial parasite *Brugia malayi* and its role in growth and development. Mol. Biochem. Parasitol. 53, 1–16.

[32] Mehta, K., Rao, U.R., Vickery, A.C. and Birckbichler, P.J. (1990) Significance of transglutaminase-catalyzed reaction in growth and development of filarial parasite *Brugia malayi*. Biochem. Biophys. Res. Commun. 173, 1051–1057.

[33] Lassandro, R., Sebastiano, M., Zei, F. and Bazuzicalupo, P. (1994) The role of dityrosine formation in the crosslinking of CUT-2, the product of a second cuticlin gene of *Caenorhabditis elegans*. Mol. Biochem. Parasitol. 65, 147–159.

[34] Sebastiano, M., Lassandro, F. and Bazzicalupo, P (1991) CUT-1, a *Caenorhabditis elegans* gene coding for a dauer-specific noncollagenous component of the cuticle. Dev. Biol. 146, 519–530.

[35] Sticher, L., Hofsteenge, J., Milani, A., Neuhaus, J.-M. and Meins, F. (1992) Vacuolar chitinases of tobacco: a new class of hydroxyproline-containing proteins. Science 257, 655–657.

[36] Tuite, M.F. (1994) Psi no more for yeast prions. Nature, 370, 327–328.

[37] Prusiner, S.B. (1991) Molecular biology of prion diseases. Science, 1515–1521.

[38] Battaglia, D.E. and Shapiro, B.M. (1988) Hierarchies of protein cross-linking in the extracellular matrix: involvement of an egg surface transglutaminase in early stages of fertilization envelope assembly. J. Cell Biol. 107, 2447–2454.

[39] Conraths, F.J., Schütze, B., Christ, H., Hobom, G. and Zahner, H. (1993) The gene coding for the major sheath protein of *Litomosoides carinii* microfilariae, gp22, is transcribed in oocytes and embryonic cells. J. Mol. Biochem. Parasitol. 60, 111–120.

[40] Rost, B., Sander, C. and Schneider, R. (1994) Redefining the goals of protein secondary structure prediction. J. Mol. Biol. 235, 13–26.