

Genomic Organization of the Zinc-Endopeptidase Astacin^{1,2}

Gebhard Geier,* Elard Jacob,† Walter Stöcker,* and Robert Zwillig*³

**Zoologisches Institut der Universität Heidelberg, Im Neuenheimer Feld 230, D-69120 Heidelberg, Germany; and*

†*Department of Toxicology, BASF Aktiengesellschaft, D-67056 Ludwigshafen, Germany*

Received April 23, 1996, and in revised form August 28, 1996

The crayfish digestive protease astacin is the first described member of the astacin family of zinc-endopeptidases, for which it is regarded as a prototype. We have isolated and characterized the genomic sequence of astacin which spans a region of 2616 bp. The coding sequence is distributed over five exons and is interrupted by four introns. It was observed that structurally and functionally essential units of the protein, like the three α -helices, the five β -strands, the Zn-binding motif, and the Met turn are never disrupted by introns. The start site of transcription was determined by primer extension analysis, confirming the existence of a pre-pro-protein of 49 amino acids which so far had not been detectable at the protein level. In addition, when compared to the amino acid sequence of mature astacin, a carboxy-terminal extension of two additional amino acids was also found. The exon-intron pattern of the astacin gene was compared to those of three other astacin family members with known genomic sequences, i.e., tolloid of *Drosophila*, the fish hatching enzyme LCE, and the human BMP1 gene. In each of the four proteins one intron was found to be inserted in the codon for a similar Gly residue which is highly conserved in this position within the astacin family. © 1997 Academic Press

Key Words: astacin; zinc-metalloproteases; pre-pro-proteins in invertebrates; exon-intron pattern.

Astacin is the prototype for the astacin family of zinc peptidases which has been studied intensively in recent years. The enzyme can be isolated from the stomach-like cardia of the freshwater crayfish *Astacus astacus*

(1–3) and was shown to be synthesized in the hepatopancreas (midgut gland) of the animal (4). This digestive endopeptidase displays a distinct unique cleavage specificity, liberating peptides with short aliphatic amino acid side chains in the N-terminal position (5). Active astacin is a single-chain polypeptide, stabilized by two intramolecular disulfide bonds, and composed of 200 amino acid residues, which account for a molecular mass of 22,614 Da (6). One zinc ion which is essential for catalysis (7) is complexed to each astacin molecule.

X ray crystal structure analysis of astacin revealed a spherical molecule, divided by a long deep active site cleft into a N-terminal and a C-terminal domain (8). The catalytic zinc ion is located at the bottom of this cleft and forms a complex with the astacin family consensus sequence HExxHxxGFxHE. The N-terminal domain mainly consists of two long α -helices and five β -strands. The C-terminal domain, apart from two helical stretches, is largely organized in several turns and irregular structures. As a characteristic feature it also includes the so-called Met turn (9, 10), which provides a tyrosine Zn ligand as part of the short conserved motif SxMHY.

In recent years it has become evident that astacin is actually a member of a highly diverged protein family (11) with homologous sequences found in both vertebrates and invertebrate species as well as the prokaryotes (flavastacin; 12). These proteins exhibit a number of different physiological functions. Human bone morphogenetic protein 1 (BMP1)⁴ has been shown to be identical with the type I procollagen C-proteinase (13) and is likely to be involved in the *de novo* formation of bone and cartilage (14). A second class of astacin family members are the membrane-bound meprins which occur in the small intestine and the kidney and are believed to be involved in the processing of biologically active peptides (15, 16). Astacin-like enzymes are also

¹ The nucleotide sequence has been submitted to the EMBL Nucleotide Sequence Database under the Accession No. X95684.

² This work was supported by the Forschungsschwerpunkt-Programm "Molekulare Evolution" of the Land Baden-Württemberg (R.Z.) and a grant from the Deutsche Forschungsgemeinschaft (Sto185/3-1) to W.S. and R.Z.

³ To whom correspondence should be addressed.

⁴ Abbreviations used: BMP1, bone morphogenetic protein 1; PCR, polymerase chain reaction; EGTA, ethylene glycol bis(β -aminoethyl ether)-*N,N'*-tetraacetic acid; TPI, triosephosphate isomerase.

engaged in the breakdown of the egg envelope during hatching of fish (17) and bird embryos (18).

Other astacin-containing proteins are morphogenetically active in early embryogenesis in the frog (19) and the sea urchin (20). In *Drosophila*, tollid, another family member, plays a role in patterning the dorsoventral axis of the embryo (21, 22).

In all of these homologous proteins, except for the hatching enzymes and astacin itself, an astacin-like module with a constant chain length of about 200 residues is attached to a much larger protein. Presumably these C-terminal extensions serve as regulatory sites or as membrane anchors (8). On the basis of topological evidence, the astacin protein family can be grouped together with other zinc-protease families (adamalysins, serralysins, and matrix metalloproteases) into a common superfamily designated as "metzincins" (9, 10).

After the extensive characterization of astacin at the protein level, we report here the genomic organization of this prototypical zinc-endopeptidase. The sequence information obtained by this work and the characterization of the transcription unit of the gene clearly confirm that a N-terminal pre-pro-peptide for astacin exists, even though an inactive precursor or zymogen has thus far not been isolated. The new data described here will help to assign structurally and functionally distinct domains of the protein to the corresponding exons and will yield insight into the molecular evolution of this protein family at the DNA level.

MATERIALS AND METHODS

Animals and isolation of genomic DNA. Crayfish, *A. astacus*, were obtained from a commercial crayfish farm (Keller, Augsburg, Germany) and kept as described elsewhere (4). For the preparation of genomic DNA, muscle tissue from animals in the intermolt stage was used. Frozen tissue (10 g) was ground by using a pestle in a mortar containing liquid nitrogen. The resulting powder was suspended in 50 ml of homogenization buffer (15 mM Tris, pH 7.5, 0.15 mM spermine, 0.5 mM spermidine, 60 mM KCl, 0.3 M sucrose, 2 mM EDTA, and 0.5 mM EGTA).

From this material nuclei were isolated using a standard protocol (23). Nuclei were suspended in lysis buffer (50 mM Tris, pH 8.0, 100 mM EDTA, 200 mM NaCl, 0.8% SDS) and incubated for 20 min at 50°C. An additional incubation at 37°C with proteinase K (1 mg/ml) was carried out with slow agitation for 6 h. Isolation of high-molecular-weight DNA was done by formamide treatment and dialysis according to Kupiec *et al.* (24) with an additional phenol extraction at the end of the procedure.

Molecular cloning. Unless otherwise stated, materials and methods for cloning and DNA manipulation were adopted from standard molecular biology protocols (25). Parts of the genomic sequence of the astacin gene were amplified from genomic DNA of *A. astacus* by PCR. Two sets of gene-specific oligonucleotide primers (see Fig. 3) were derived from cDNA fragments isolated by E. Jacob (manuscript in preparation). Primers EJ354 (5'-TATCTCTGGTCAGGTGTCATA-3') and EJ357 (5'-CGGGTGTGCTCATGGTAGAAG-3') resulted in a 488-bp product by using 100 ng of chromosomal DNA and the following temperature cycle: 94°C, 1 min; 58°C, 2 min; 72°C, 5 min repeated 35 times followed by an extension at 72°C for 10 min. Using the same

protocol, primers EJ356 (5'-GCAAGCCAATGGTTGTGTTTACCA-3') and EJ229 (5'-TGCATCAGTTTGAGCATGTGAGC-3') resulted in an amplification product of 1190 bp. The PCR products were analyzed in 2% agarose gels, excised from the gels, and purified with a NUCLEOTRAP gel extraction kit (Machery & Nagel, Germany). The purified fragments were then cloned into a T-vector generated from a *EcoRV*-digested pBluescript KS⁺ plasmid (Stratagene) by the protocol of Marchuk *et al.* (26). By this procedure the 488- and 1190-bp PCR products were cloned as pASP1 and pASP3, respectively.

For Southern blot analysis of the astacin gene, 10 µg of high-molecular-weight genomic DNA were completely digested using different restriction enzymes, separated in 0.8% agarose gels, and transferred to Hybond^N membranes (Amersham). Hybridization was performed in 6× SSC, 5× Denhardt's solution, 0.5% SDS, 100 µg/ml of salmon sperm DNA, and 50% formamide at 42°C for 20 h. The 1190-bp PCR product was used as a probe, radioactively labeled to a specific activity of 6 × 10⁹ cpm/µg by random priming. After hybridization, membranes were washed with 2× SSC, 0.1% SDS for 5 min twice at 37°C. Final washing was performed in 0.1× SSC, 0.1% SDS at 60°C.

For the creation of subgenomic libraries, 40 µg of a total *EcoRI*, *HindIII*, and *Clal* digest of genomic DNA was separated in a preparative 0.8% agarose gel. Genomic *EcoRI* fragments in the 2- to 3.5-kb range, *HindIII* fragments in the 5.5- to 6.0-kb range, and *Clal* fragments between 2.5- and 3.5-kb were excised and purified as described above and ligated to alkaline phosphatase-treated pBluescript KS⁺ vector that had been digested with the same restriction enzymes. In order to obtain a high transformation efficiency, the libraries were introduced into *Escherichia coli* DH5α (BRL) by electroporation (32). Preliminary characterization of mini-prep DNA (Promega Biotec, Magic mini-prep) from selected positive recombinants verified that the size-selected genomic fragments were ligated to pBluescript. Identification of clones containing astacin sequences was done by colony hybridization on nitrocellulose filters. Oligonucleotide primers EJ356 and EJ229 labeled on the 5'-position with [γ -³²P]ATP were used as probes. Hybridization conditions were as described for Southern blotting. By this procedure a 2.8-kb *EcoRI* fragment, a 3.2-kb *Clal* fragment, and a 5.8-kb *HindIII* fragment were identified and cloned as pAS25, pAS34, and pAS42, respectively.

The nucleotide sequences of the plasmid inserts were determined by double-stranded sequencing according to the dideoxy chain termination method, using Sequenase version 2.0 (UBS). Either universal M13 primers or the specific primers mentioned above were used. In some cases, plasmid inserts partially deleted by exonuclease III were utilized.

Primer extension analysis. For primer extension analysis, Poly(A)⁺ RNA was isolated from the crayfish hepatopancreas by standard methods (25). The oligonucleotide primer PE 3 (5'-GCGCAC-TGCATGTGGTAGGTA-3'), complementary to nucleotides 326 to 346 in Fig. 3) was 5'-end-labeled with polynucleotide kinase and [γ -³²P]-ATP and purified on a 20% urea-acrylamide gel. Poly(A)⁺ RNA (15 µg) was primed and reverse-transcribed with 12 units of reverse transcriptase for 45 min at 37°C as described (30). The primer extension product was analyzed on a 6% sequencing gel, and the size of the product was determined by comparison with sequencing reactions that were simultaneously loaded.

RESULTS AND DISCUSSION

Isolation of Genomic Fragments of the Astacin Gene

A first approach to obtain genomic sequences of astacin was made by applying PCR techniques. Two fragments of 488 and 1190 bp were amplified from genomic DNA of *A. astacus* and cloned as described under Materials and Methods. Since both of these cloned genomic

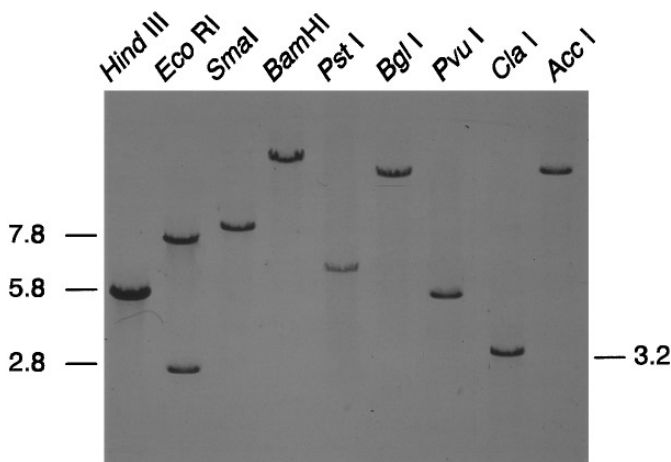


FIG. 1. Southern blot analysis of *Astacus astacus* genomic DNA. 10 μ g of genomic DNA was digested with the restriction enzymes indicated in the top line, separated in a 0.8% agarose gel, and transferred to a nylon membrane. The filter was hybridized with a 1190-bp genomic PCR product of the astacin gene, labeled internally with 32 P. Filters were washed at high stringency and exposed to an X ray film for 2 days.

PCR products are significantly longer than the corresponding cDNA derived amplification products (326 and 296 bp, respectively; data not shown) the presence of intron sequences was supposed. As described below, this assumption was confirmed by sequencing.

Southern blot analysis of genomic *A. astacus* DNA, applying the 1190-bp PCR product of pASP3 as a probe, revealed a simple pattern of hybridization (Fig. 1). Except for *EcoRI*, all restriction enzymes used for the total digest of genomic DNA resulted in a single hybridization signal, suggesting that the astacin gene is present in a single copy within the genome of *A. astacus*. However, for active astacin at least three isoforms have been observed after purification (2). Several reasons may be envisaged to account for this observation:

There may be several genes with sequences that are too dissimilar in their nucleotide sequences to be detectable by the high stringency conditions used here. It is also conceivable that there are diverging sequences beyond the sequence stretch covered by the probe. Post-translational trimming of one or both C-terminal residues which are deduced from the nucleotide sequence of the gene (see below), but absent in the predominant isoform at the protein level, may also result in different isoforms. In addition, splicing variants of the astacin gene might also exist.

In a second approach to isolate genomic sequences of the astacin gene, restriction fragments of an appropriate size range (as determined by southern blot analysis, Fig. 1) were used for a shotgun cloning. By this procedure subgenomic libraries were established as described under Materials and Methods. A screening for

astacin sequences by colony hybridization resulted in the isolation of a 2.8-kb *EcoRI* fragment, a 3.2-kb *ClaI* fragment, and a 5.8-kb *HindIII* fragment.

Nucleotide Sequence and Structural Analysis of the Astacin Gene

The cloned genomic PCR products and the genomic fragments isolated from the subgenomic libraries were sequenced on both strands. Using overlapping sequences, a continuous genomic region of 7.6 kb was established (Fig. 2). As deduced from the amino acid sequence (6), additional cDNA sequence data, and primer extension analysis (see below) the complete astacin gene was localized to a 2.6-kb *EcoRV*–*PstI* subfragment. The nucleotide sequence of this fragment and the deduced amino acid sequence of 251 residues is shown in Fig. 3. It differs from the cDNA sequence by only three nucleotides, of which only one leads to a different amino acid residue. The amino acid residue in position –41 was found to be valine instead of alanine in the cDNA sequence (E. Jacob, manuscript in preparation).

The coding region of the astacin gene is interrupted by four introns, designated (in the 5'–3' direction) as intron 1 to 4 (Figs. 2 and 3). Two long intronic sequences of 268 bp (intron 1) and 894 bp (intron 2) are followed by two small introns of 106 bp (intron 3) and 58 bp (intron 4). These introns display donor and acceptor sites fitting the invertebrate splice junction consensus (27) and contain stop codons in all three reading frames. Except for intron 4 the astacin introns also contain an internal sequence corresponding to a putative lariat branch point upstream to the 3'-splice site in the region between 21 and 35 nucleotides upstream from the 3'-splice point (28).

The amino acid sequence deduced from the genomic DNA shows an amino-terminal extension of 49 residues when compared to active astacin (Fig. 3). The ATG translation initiation codon (Met –49) starts at nucleotide position 336 (according to the nucleotide position numbering in Fig. 3). Only five codons upstream from it, a TAA stop codon is located in frame.

The existence of a signal- and a pro-sequence has long been postulated for astacin, since all members of the astacin family are either secreted or are membrane bound proteins. Besides a transient signalpeptide, they also contain a pro-part which is removed during maturation. In the case of astacin no inactive zymogen (proenzyme) has been observed, presumably because during biosynthesis the enzyme is not accumulated intracellularly, but is rather immediately released and stored in the stomach as an active mature proteinase (4). The crystal structure of astacin indicates that the removal of a prosequence may be the major mechanism of activation. An amino-terminal extension will prevent

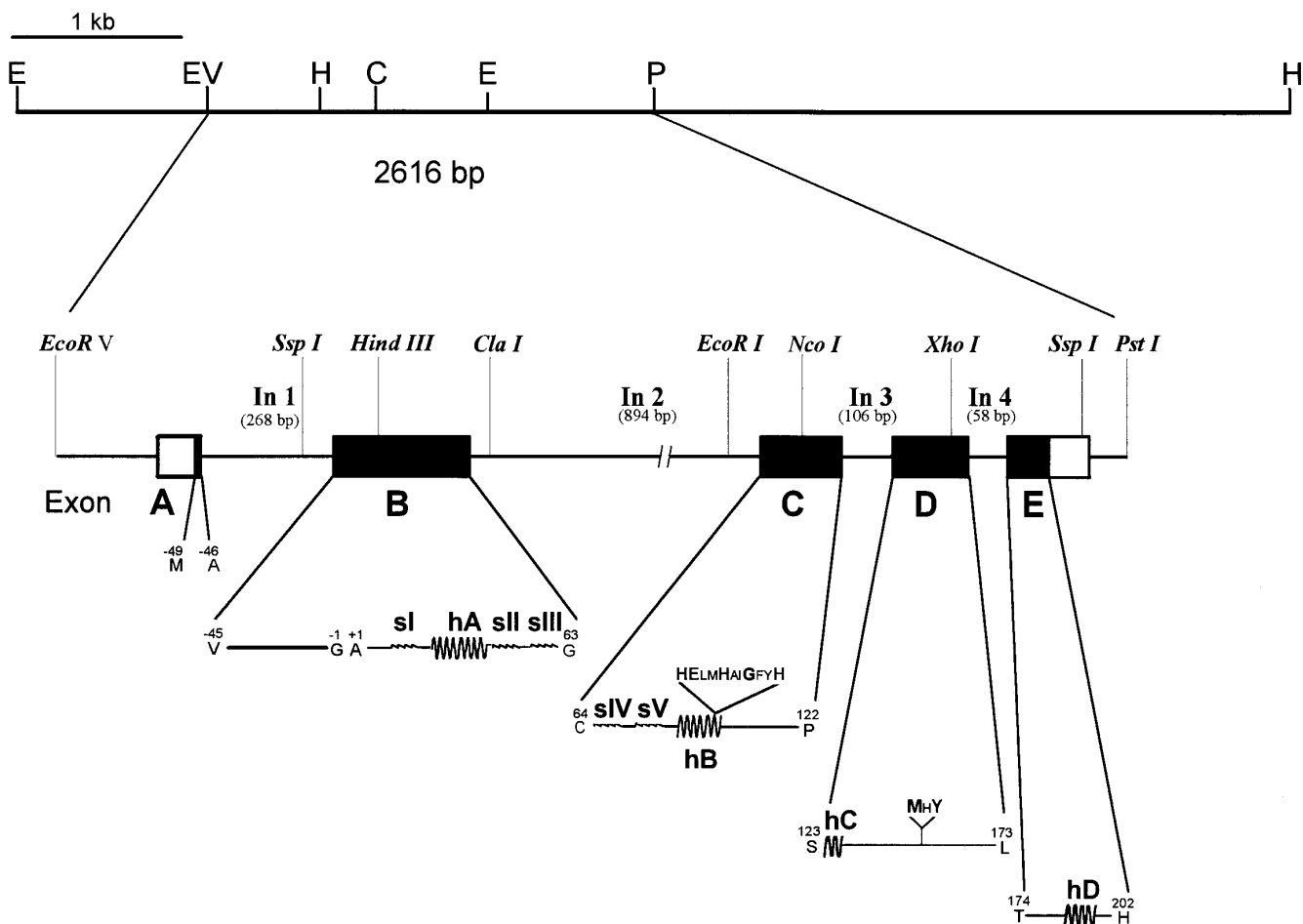


FIG. 2. Structure of the astacin gene. The restriction map of the 7.6-kb genomic fragment carrying the gene is presented on the top line. (abbreviations: E, *EcoRI*; EV, *EcoRV*; H, *Hind III*; C, *Cla I*; P, *Pst I*). The distribution of secondary structure elements of the enzyme over the five exons is outlined in the lower part of the figure. α -helices hA to hD and β -strands sI to sV are indicated according to Stöcker *et al.* (8).

the formation of a critical salt bridge between the amino-terminal residue of mature astacin (Ala 1) and Glu 103, which is next to the zinc ligand His 102 at the active site (8).

The pre-pro-peptide deduced from the genomic sequence of astacin contains in its N-terminal part a putative signalpeptide of 15 residues as predicted by the program PSIGNAL (PC/Gene-package, IntelliGenetics). The potential cleavage site, which conforms to the “(-3, -1)-rule” of von Heijne (29), is located between Ala -35 and Ser -34. For the removal of the remaining prosequence two modes of proteolytic cleavage are conceivable. There could be an autoactivation by astacin itself, cleaving between Gly -1 and Ala +1. The sequence around the mature N-terminus would meet the requirements for an astacin substrate cleavage site (5). Another possible way for activation might consist in a tryptic cleavage after Arg -3, followed by subsequent trimming of the N-terminus by an aminopeptidase.

Surprisingly, the C-terminus of astacin as deduced from the 3' region of the gene carries two additional residues when compared to the mature enzyme. The mature C-terminus (Leu 200) is followed by Arg 201 and His 202 (Fig. 3), which have also been found in the cDNA sequence. Possibly the two additional amino acids are removed by the activity of a carboxypeptidase B, which is also present in *A. astacus* (2).

At a distance of 83 bp downstream from the TAG stop codon (nucleotide position 2415) a putative polyadenylation signal is located, which occurs 11 bp upstream from the predicted end of the cDNA (indicated in Fig. 3).

The 5' region of the astacin gene was further characterized by primer extension of mRNA isolated from hepatopancreas to determine the start site of transcription. An oligonucleotide primer (PE3) complementary to nucleotides 326–346 was radiolabeled and extended, and the product was analyzed by denaturing gel elec-

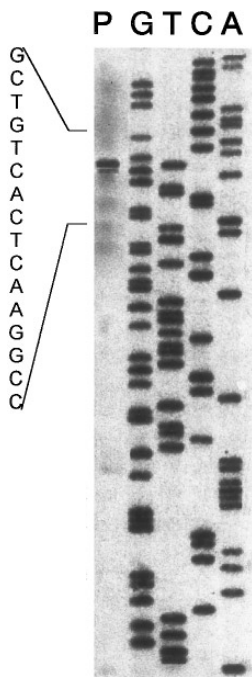


FIG. 4. Mapping of the transcriptional start site of the astacin gene by primer extension analysis. Lane P shows the product obtained by primer extension of oligonucleotide PE 3 (see Fig. 3). Lanes G, T, C, and A are sequencing reactions used to determine the exact size of the extended product.

trophoresis (Fig. 4). A major product was found as well as a minor product corresponding to a fragment one nucleotide shorter in length, which was detected only upon a very long exposure of the gel. The major product corresponded to a 101-base extended fragment as deduced from the dideoxy sequencing reactions, primed with the same oligonucleotide, and loaded onto the gel at the same time. This result localizes the transcription start site at the A in position 246, which is 90 bp upstream from the translation initiation codon. The analysis of the 5'-flanking region revealed a TATA box (putative RNA polymerase II promoter sequence) in positions 217 to 223. This is centered 27 bp upstream from the identified transcription initiation site, which is in good accordance with the preferred position of such a promoter element (30). The sequence CA at the transcription start site corresponds to the sequence found in the majority of eukaryotic promoters analyzed (30).

The coding sequence of the astacin gene is distributed over a total of five exons, designated as exon A to E in Fig. 2. Sequences coding for secondary structure elements of the protein like the three α -helices hA, hB, hD, the short 3^{10} -helix hC, and the five β -strands sI to sV (8) are never interrupted by introns (Fig. 2). Likewise, also the main Zn-binding motif and the Met turn consensus remain uninterrupted.

Exon A only codes for the four N-terminal amino

acid residues of the pre-pro-sequence from the start methionine (position -49) to Ala -46. It also includes a 90-bp nontranslated leader sequence, which is also part of the transcription unit. Exon B, which is the longest of the five astacin exons, encodes the region between Val -45 and Gly 63. It contains the major part of the pre-pro-sequence (Val -45 to Gly -1), the N-terminus of mature astacin (Ala 1), and includes β -strands sI, sII, sIII, and α -helix hA. Exon C comprises the region between Cys 64 and Pro 122 which includes β -strands sIV and sV as well as α -helix hB. It also includes the main zinc-binding motif, which is represented by the consensus HExxHxxGFxH (8, 11). Exon D carries the coding sequence between Ser 123 and Leu 173. It therefore contains the short 3^{10} -helix hC and the conserved SxMHY consensus for the characteristic Met turn (9), which provides the zinc ligand Tyr 149. Exon E codes for the C-terminal part of astacin from Thr 174 up to the extended C-terminus His 202 and includes the C-terminal α -helix D. The stop codon is followed by a 99-bp untranslated trailer containing a putative polyadenylation signal.

Until now, tolloid from *Drosophila* (22), the fish hatching enzymes HCE and LCE from *Oryzias latipes* (17, 31), and the human BMP1 gene (32) are the only astacin family members that have been characterized at the genomic level. Fig. 5 outlines an alignment of the intron positions in the astacin-like modules found in these three genes (36 to 39% amino acid sequence identity to astacin) and the astacin gene characterized in this work.

From an evolutionary point of view it is quite interesting that the comparison of a crustacean, an insect, and two vertebrate members of the astacin family seems to reveal conserved positions for some of the introns. Intron 2 of the astacin gene, intron e of the LCE gene (31), one of the introns in the BMP1 gene (32), and the only intron found in the astacin-like portion of tolloid are all inserted into the codon for a similar Gly residue (Gly63 in the astacin sequence), which is highly conserved in this position within the astacin family. At the protein level this intron occurs between β -strands sIII and sIV. However, while the astacin intron separates the codon triplet for this residue after the first nucleotide (phase I intron), the three other introns lie between the second and third nucleotide (phase II introns). The genes for astacin, BMP1, and LCE also share a second conserved intron position which corresponds to astacin intron 3. According to the standard amino acid sequence alignment of the astacin family (8, 11) these three phase I introns occur in homologous positions.

There are two major theories to explain the origin and evolutionary significance of introns. One suggests that precursor genes consisted entirely of coding sequences and introns were inserted later in the course

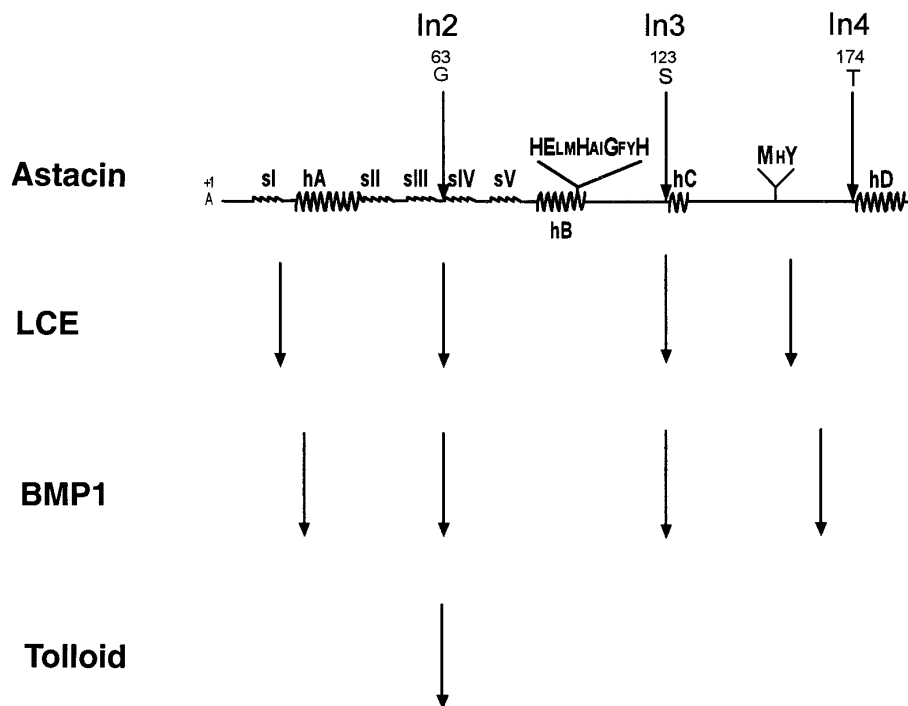


FIG. 5. Comparison of intron positions in the astacin-like domain of the tolloid gene of *Drosophila* (22), the gene for the fish hatching enzyme LCE (31), the human BMP1 gene (32), and the astacin gene (this work). Only the section representing the mature astacin protease is shown. The intron positions are marked by arrows. For orientation, the position of astacin introns 2, 3, and 4 relative to the amino acid sequence are indicated.

of evolution (33). The alternative theory argues that introns were present in the earliest cells (34, 38) and that gene evolution is dominated by recombination within the introns ("exon shuffling"), by the loss of introns, and by sliding of the introns.

A conservation of certain intron positions within the genes of some protein families has also been observed in the actin family (35), the serine protease family (36), in evolutionary distinct genes for glyceraldehyde-3-phosphate dehydrogenases (37), and triosephosphate isomerase (TPI) (38). For the different TPI genes, for example, it was shown that intron positions are not random in respect to the three-dimensional structure of the protein (38), implicating that exons are related to units of protein structure. In general this is also true for the astacin gene where elements of secondary structure are not interrupted by introns. Intron 2, which may be seen as an exception, is only inserted into the outermost residue (Gly63) of β -strand sIV.

It is remarkable that the position of the astacin intron 2 seems to be conserved in invertebrates (insect and crustacean) and vertebrates (fish and human), although both are separated by an evolutionary distance of about 500 million years. This conservation, however, is entirely true only in terms of the same amino acid position since the exact position of intron 2 differs from those of the corresponding introns in the other three

genes by one nucleotide (see above). Such cases of quasi-conservation have also been observed in other protein families (37, 38) and may be interpreted as examples of intron sliding caused by a passive series of mutations that must leave an intron present, but not necessarily in the same nucleotide position (38).

It is also noteworthy that the exon-intron patterns of the genes with a crustacean, fish, and human origin show a closer similarity (two conserved intron positions) than the two invertebrate genes (one conserved intron position). The fact that the tolloid gene of *Drosophila* contains only one single intron in the astacin-like module may reflect a tendency toward a loss of introns, which seems to be reasonable for organisms with small genomes and short generation times.

The complete loss of introns within a gene of an astacin family member has also been observed for the fish hatching enzyme HCE. Both HCE and LCE, which are closely related, were shown to be expressed in the same fish species (17, 31). In this case the intron-less HCE gene is thought to be the consequence of a mRNA-mediated process of retrotransposition.

Exon-intron patterns are also known for several members of the matrix metalloproteinases (matrixins) (39), another group of the metzincins (9). Interestingly, the presently known astacin and matrixin genes share a single intron position, which corresponds to the re-

gion between β -strands III and IV at the protein level. Considering the three-dimensional structure of the astacins and the other metzincins, this particular conserved intron position is located at the border between the N-terminal framework of the protein structure and the catalytic apparatus, which starts with β -strand IV that is essential for substrate binding. This specific feature might account for the fact that intron 2 has been conserved beyond the boundaries of different protease families which belong to the same superfamily, the metzincins.

ACKNOWLEDGMENTS

The authors thank Monika Dörhöfer and Susanne Petersen-Werner for excellent technical assistance.

REFERENCES

- Pfleiderer, G., Zwilling, R., and Sonneborn, H. H. (1967) *Hoppe Seyler's Z. Physiol. Chem.* **348**, 1319–1331.
- Zwilling, R., and Neurath, H. (1981) *Methods Enzymol.* **80**, 633–664.
- Stöcker, W., and Zwilling, R. (1995) *Methods Enzymol.* **248**, 305–325.
- Vogt, G., Stöcker, W., Storch, V., and Zwilling, R. (1989) *Histochemistry* **91**, 373–381.
- Krauh, E., Dörsam, H., Little, M., Zwilling, R., and Ponstingl, H. (1982) *Anal. Biochem.* **119**, 153–157.
- Titani, K., Torff, H.-J., Hormel, S., Kumar, S., Walsh, K. A., Rödl, J., Neurath, H., and Zwilling, R. (1987) *Biochemistry* **26**, 222–226.
- Stöcker, W., Wolz, R. L., Zwilling, R., Strydom, D. J., and Auld, D. S. (1988) *Biochemistry* **27**, 5026–5032.
- Stöcker, W., Gomis-Rüth, F.-X., Bode, W., and Zwilling, R. (1993) *Eur. J. Biochem.* **214**, 215–231.
- Bode, W., Gomis-Rüth, F.-X., and Stöcker, W. (1993) *FEBS Lett.* **331**, 134–140.
- Stöcker, W., Grams, F., Baumann, U., Reinemer, P., Gomis-Rüth, F.-X., McKay, D. B., and Bode, W. (1995) *Protein Science* **4**, 823–840.
- Bond, J. S., and Beynon, R. J. (1995) *Protein Sci.* **4**, 1247–1261.
- Tarentino, A. L., Quinones, G., Grimwood, B. G., Hauer, C. R., and Plummer, T. H. (1995) *Arch. Biochem. Biophys.* **319**, 281–285.
- Kessler, E., Takahara, K., Biniaminov, L., Brusel, M., and Greenspan, D. S. (1996) *Science* **271**, 360–362.
- Wozney, I. M., Rosen, V., Celeste, A. L., Mitsock, L. M., Whitters, M. J., Kriz, R. W., Hewick, R. M., and Wang, E. A. (1988) *Science* **242**, 1528–1534.
- Dumermuth, E., Sterchi, E. E., Jiang, W., Wolz, R. L., Bond, J. S., Flannery, A. V., and Beynon, R. J. (1991) *J. Biol. Chem.* **266**, 21381–21385.
- Dumermuth, E., Eldering, J. A., Grünberg, H., Jiang, W., and Sterchi, E. E. (1993) *FEBS Lett.* **335**, 367–375.
- Yasumasu, S., Jamada, K., Akasaka, K., Mitsunage, K., Iuchi, I., Shimada, H., and Yamagami, K. (1992) *Dev. Biol.* **153**, 250–258.
- Elaroussi, M. A., and DeLuca, H. F. (1994) *Biochim. Biophys. Acta* **1217**, 1–8.
- Sato, S. M., and Sargent, T. D. (1990) *Dev. Biol.* **137**, 135–141.
- Reynolds, S. D., Angerer, L. M., Palis, J., Nasir, A., and Angerer, R. C. (1991) *Development* **114**, 769–786.
- Shimell, M. J., Ferguson, E. L., Childs, S. R., and O'Conner, M. B. (1991) *Cell* **67**, 469–481.
- Finelli, A. L., Bossie, C. A., Xie, T., and Padgett, W. (1994) *Development* **120**, 861–870.
- Bingham, P. M. (1981) *Cell* **25**, 693.
- Kupiec, J. J., Giron, M. L., Vilette, D., Jeltsch, J. M., and Emanoil-Ravier, R. (1987) *Anal. Biochem.* **164**, 53–59.
- Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) "Molecular Cloning: A Laboratory Manual," 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Marchuk, D., Drumm, M., Saulino, A., and Collins, F. S. (1990) *Nucleic Acids Res.* **19**, 1154.
- Shapiro, M. B., and Senapathy, P. (1987) *Nucleic Acids Res.* **15**, 7155–7174.
- Keller, E. B., and Noon, W. A. (1985) *Nucleic Acids Res.* **13**, 4971–4981.
- Von Heijne, G. (1986) *Nucleic Acids Res.* **14**, 4683–4690.
- Bucher, P. (1990) *J. Mol. Biol.* **212**, 563–578.
- Yasumasu, S., Shimada, H., Inohaya, K., Ymazaki, K., Iuchi, I., Yasumasu, I., and Yamagami, K. (1996) *Eur. J. Biochem.* **237**, 752–758.
- Takahara, K., Lee, S., Wood, S., and Greenspan, D. S. (1995) *Genomics* **29**, 9–15.
- Cavalier-Smith, T. (1991) *Trends Genet.* **7**, 145–148.
- Gilbert, W. (1987) *Cold Spring Harbor Symp. Quant. Biol.* **52**, 901–905.
- Lewin, B. (1994) In "Genes V," p. 700, Oxford University Press, New York.
- Rogers, J. (1985) *Nature* **315**, 458–459.
- Kersanach, R., Brinkmann, H., Liaud, M.-F., Zhang, D.-X., Martin, W., and Cerff, R. (1994) *Nature* **367**, 387–389.
- Gilbert, W., and Glynias, M. (1993) *Gene* **135**, 137–144.
- Gaire, M., Magbanua, Z., McDonnell, S., McNeil, L., Lovett, D. H., and Matrisian, L. M. (1994) *J. Biol. Chem.* **269**, 2032–2040.