Biol. Chem., Vol. 380, pp. 499 – 501, April 1999 · Copyright © by Walter de Gruyter · Berlin · New York

Short Communication

Trypsin from *Pacifastacus leniusculus* Hepatopancreas: Purification and cDNA Cloning of the Synthesized Zymogen

Patricia Hernández-Cortés[1,*], Lage Cerenius[1], Fernando García-Carreño[2] and Kenneth Söderhäll[1]

[1] Department of Physiological Mycology, University of Uppsala. Villavägen 6, S-75236 Uppsala, Sweden
[2] Northwest Biological Research Center (CIBNOR) P.O. Box 128 La Paz, BCS 23000, Mexico

* Corresponding author

Trypsin was purified from crayfish, *Pacifastacus leniusculus*, hepatopancreas, and the gene that encoded this enzyme was cloned from a hepatopancreas cDNA library. Crayfish trypsin is synthesized as a zymogen according to the sequence of the putative precursor peptide. The authenticity of the trypsinogen is supported by the deduced amino acid sequence and confirmed by the N-terminal amino acid sequence of the mature protein. The enzyme has features characteristic of a trypsin, such as a specific binding pocket. Sequence comparison shows that crayfish trypsin is similar to those of other species, with the exception that six cysteine residues present in vertebrates are missing. Some structural characteristics, such as the length of the signal peptide and a calcium binding site, are similar to bacterial trypsin.

Key words: Crustacean / Protease evolution / Serine proteinases.

Trypsin is a serine proteinase that hydrolyzes peptide bonds. Proteinases are subclassified according to the chemical nature of the amino acid residues that are responsible for their catalysis. There are four classes of proteinases: serine, cysteine, aspartic, and metallo proteinases. The catalytic triad of serine proteinases is His-57, Asp-102, and Ser-195. The name is based on the nature of the most prominent amino acid residue at the active site, in this case serine, and the numbering system corresponds to the position of the amino acid sequence of bovine chymotrypsin (Rypniewski *et al.*, 1994). All serine proteinases, such as trypsin, chymotrypsin, and elastase, are structurally similar and split internal peptide bonds of a protein. The variants of eukaryotic serine proteinases are thought to have arisen by gene duplication and mutations (Neurath, 1984). However, their cleavage specificities are different. Vertebrate proteinases acquired a higher degree of specialization by restricting their action to hydrolyzing peptide bonds at specific sites of a protein (Neurath, 1986). Zwilling *et al.* (1975) argued that the trypsin of *Astacus leptodactylus*, a freshwater crayfish, was a missing link between prokaryotic and vertebrate serine proteases. This statement was made because this trypsin has an acidic isoelectric point, is resistant to autodigestion, is irreversibly denatured at pH below 3, and is not a zymogen (Titani *et al.*, 1983).

Crustaceans synthesize their digestive proteinases in the hepatopancreas, also called the midgut gland. This organ combines the function of liver and pancreas of vertebrates (Gibson and Baker, 1979). To obtain information about the genes that encode proteinases in the freshwater crayfish *Pacifastacus leniusculus* and demonstrate their authenticity, we isolated, cloned, and sequenced a cDNA that encodes a trypsinogen present in *P. leniusculus* he-

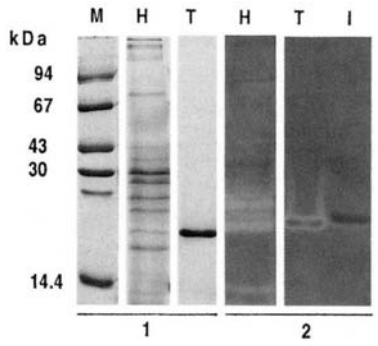

Fig. 1 SDS-PAGE and Zymogram of the Trypsin from Crayfish, *Pacifastacus leniusculus*.
Samples were electrophoresed through a discontinous gel (4% stacking gel and 12.5% separating gel) at 4 °C. Protein was visualized by Coomassie Brillant Blue and zymograms were done using casein as substrate (Garcia-Carreño *et al.*, 1993). 1, protein profile; 2, zymogram. Lanes: M is protein molecular marker, H is the hepatopancreas extract, T is the purified trypsin, and I is the inhibited trypsin with tosyl-Lys-chloromethyl ketone. The crayfish trypsin was purified from a hepatopancreas extract with a *p*-aminobenzamidine Sepharose column to which serine proteinases had been bound; trypsin was eluted with a buffer containing the competitive inhibitor butylamine (Eberhardt, 1992). A concentrated sample from the previous purification step was loaded onto a Mono-Q column and proteins were eluted using a linear gradient of 0.3 to 0.8 M NaCl. The purified trypsin was blotted onto polyvinylidene difluoride membrane with the methodology previously described (Johansson *et al.*, 1994) and the NH_2-terminal amino acid determined.

patopancreas and we also purified a trypsin encoded by the trypsinogen cDNA.

To isolate trypsin, *P. leniusculus* hepatopancreas were dissected, homogenized, and centrifuged to eliminate lipids and tissue debris. The aqueous supernatant obtained, called the hepatopancreas extract, was applied to an affinity column followed by ion-exchange chromatography. Protein and zymogram patterns of the hepatopancreas extract and the purified trypsin are shown in Figure 1. Five bands exhibited proteolytic activity. The 22 kDa band was identified as trypsin. This proteinase was inhibited by TLCK, a specific trypsin inhibitor, and also hydrolyzed TAME, a trypsin substrate (Worthington, 1993). The NH_2 terminus of this enzyme had the sequence IVGGT, which is homologous to the NH_2 terminus of mature trypsin of other organisms.

To obtain the trypsin cDNA sequence, the hepatopancreas cDNA library was screened with a 459 bp cDNA clone that encodes a trypsin-like proteinase of the hemocyte of *P. leniusculus* (Huang *et al.*, unpublished). A clone of 900 bp was isolated, amplified, and subcloned into the *Eco*RV site of pBluescript KS(+) (Stratagene). The DNA sequence was determined in both directions using the dideoxy-chain termination reaction and the sequence was analyzed with MacVector (Eastman Chemical Co.) software. A hydrophobic region of 15 amino acids was found and a putative precursor peptide of 13 residues, including a Lys cleavage site, was present in the cDNA sequence

```
1   GCCAGCCATGAAGACCCTCGTGTTCTGCCTGCTCCTCGTCGGGGCTCTCGCCGCCCCCTC
1     P  A  M  K  T  L  V  F  C  L  L  L  V  G  A  L  A  A  P  S
61  CCGCAGGCTCAGATTTCCACCCAACAATTACAAGATCGTTGGTGGCACTGATGCCAGTCT
21    R  R  L  R  F  P  P  N  N  Y  K  I  V  G  G  T  D  A  S  L
121 CGGTGAGTTCCCATACCAGCTCAGCTTCCAGGAAAAATTCCTAGGGTTTTCTTTCCATTT
41    G  E  F  P  Y  Q  L  S  F  Q  E  K  F  L  G  F  S  F  H  F
181 CTGCGGAGCTTCCATCTACAACGAGAACTATGCCATCACTGCTGGCCACTGTGTCTACGG
61    C  G  A  S  I  Y  N  E  N  Y  A  I  T  A  G  H  C  V  Y  G
241 CGATGACTACGACAATCCAAGTGGCCTCAATATTGTGGCTGGAGAACTCGACATGAGCGT
81    D  D  Y  D  N  P  S  G  L  N  I  V  A  G  E  L  D  M  S  V
301 AAACGAAGGATCGGAGCAGACCATTGCAGTGTCAAAAATCATCCTTCATGAGAACTTGGA
101   N  E  G  S  E  Q  T  I  A  V  S  K  I  I  L  H  E  N  L  D
361 CTACGATCTTCTAGATAATGACATCTCTCTTCTTAAGCTGGCTACACCCCTCACTTTCAA
121   Y  D  L  L  D  N  D  I  S  L  L  K  L  A  T  P  L  T  F  N
421 CAATAACGTCGCTCCCATTGCTCTTCCAGCACAAGGTCATACAGCCACTGGCAACGTCAT
141   N  N  V  A  P  I  A  L  P  A  Q  G  H  T  A  T  G  N  V  I
481 TGTCACAGGTTGGGGCACCACAAGCGAAGGAAGGAACACTCCTGATGTGCTGCAGAAGGT
161   V  T  G  W  G  T  T  S  E  G  R  N  T  P  D  V  L  Q  K  V
541 GACTATCCCTCTCGTGTCGGACGCAGAGTGCCGCTCTGATTATGGTGCTGATGAAATTTT
181   T  I  P  L  V  S  D  A  E  C  R  S  D  Y  G  A  D  E  I  F
601 TGACTCCATGATCTGCGCCGGAGTTCCTGAGGGAGGGAAGGACTCGTGCCAGGGTGACTC
201   D  S  M  I  C  A  G  V  P  E  G  G  K  D  S  C  Q  G  D  S
661 TGGTGGACCTCTGGCCGCCAGCGACACCGGGTCTACATACCTGGCCGGCATTGTGTCCTG
221   G  G  P  L  A  A  S  D  T  G  S  T  Y  L  A  G  I  V  S  W
721 GGGCTACGGCTGCGCTCGTGCGGGCTACCCCGGCGTCTACACTGAGGTCTCTTACCATGT
241   G  Y  G  C  A  R  A  G  Y  P  G  V  Y  T  E  V  S  Y  H  V
781 CGACTGGATCAAAGCAAATGCAGTCtaaacagtgataatagtaacactcagattatgtga
261   D  W  I  K  A  N  A  V  *
841 acatcaataaacaaatttaatttgtttactaatttttaaataaattcaaatataaaaaaa
```

Fig. 2 The Nucleotide Sequence of *P. leniusculus* Trypsinogen cDNA and Its Deduced Amino Acid Sequence.
A 900 bp clone was isolated from the crayfish hepatopancreas library and sequenced. The putative open reading frame is composed of 798 nucleotides (8–705), corresponding to 266 amino acids. A hydrophobic region of 15 amino acids, the signal peptide (boxed), was present in the cDNA sequence. The predicted zymogen peptide of 13 amino acids begins at nucleotide 56 (Pro, horizontal arrow) and the mature trypsin begins at nucleotide 96 (Ile, vertical arrow). The corresponding NH_2-terminal amino acid sequence of the purified trypsin is double underlined. The stop codon, TAA, is indicated by an asterisk. EMBL Nucleotide Sequence Database, accession number AJ22658.

(Figure 2). The putative precursor peptide of *P. leniusculus* trypsin supports the idea that digestive proteinases are synthesized as zymogens, as are other crayfish proteinases (Aspan and Söderhäll, 1990). The molecular mass of the deduced amino acid sequence of the mature trypsin was 25 kDa, which is close to the mass determined for the purified trypsin.

Multiple sequence alignments of trypsinogens are shown in Figure 3. The sequence following the first Met is similar in shrimp and crayfish trypsin. However, we did not

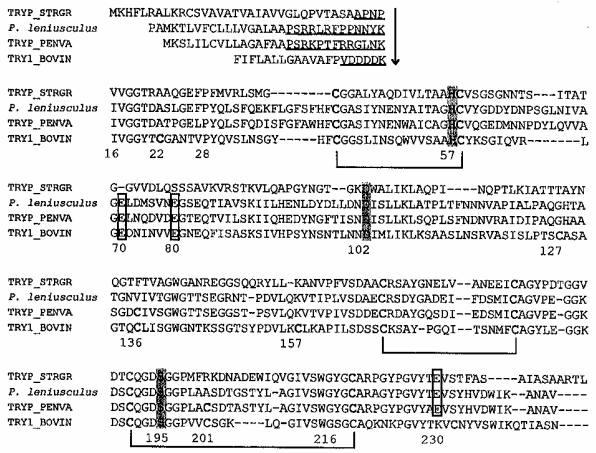

Fig. 3 Amino Acid Sequence Alignment of Trypsinogens.
The amino acids that comprise the zymogen peptide are underlined. The putative activation cleavage site is indicated with an arrow. The catalytic triad (His, Asp, Ser) is denoted by shadows. The cysteine residues are in bold and solid lines denote the conserved intermolecular disulfide bonds. Boxes indicate calcium-binding sites. Sequences are shown for bacteria, *Streptomyces griseus* (TRYP_STRGR), crayfish, *Pacifastacus leniusculus*, shrimp, *Penaeus vannamei* (TRYP_PENVA), and bovine *Bos taurus* (TRYP_BOVIN) trypsinogens (SWISSPROT data-library accession codes). The numbering system is according to bovine chymotrypsin, beginning at the N-terminal residue of the mature protein (Rypniewski *et al.*, 1994).

determine whether the Met in the deduced protein sequence represents the start codon, but the similarity between the signal peptide identified in shrimp trypsinogen (Klein *et al.*, 1996) and the crayfish supports this conclusion. Trypsin zymogens of shrimp and crayfish appear to be longer than those of other invertebrates (Peterson *et al.*, 1994), vertebrates or bacteria. The reason for this is, however, still unknown.

The amino acids that determine serine catalysis (His-57, Asp-102, and Ser-195, numbers corresponding to the sequence of bovine chymotrypsin) and trypsin specificity (Asp-189, Gly-216, and Gly-226) are conserved in *P. leniusculus* trypsinogen. Three cysteine residues that form disulfide links in vertebrates trypsin are also conserved in crayfish trypsinogen, namely Cys 42:58 (His loop), Cys 168:82 (Met loop), and Cys 191:220 (Ser loop). In vertebrates, three additional disulfide bridges, 22:157, 127:232 and, 136:201, are present. Titani *et al.* (1983) hypothesized that these additional cysteine bridges in vertebrate

trypsins were acquired after decapod crustaceans and mammals separated during the course of evolution. The disulfide linkage Cys 136:201 is characteristic of vertebrate serine proteinases and is also present in shrimp trypsin (Klein *et al.*, 1996), and in abalone, a molluscan, chymotrypsin (Groppe and Morse 1993). The digestive trypsinogen of *P. leniusculus* and *Astacus fluviatilis* trypsin (Titani *et al.*, 1983) does not possess any of those six residues. This supports the suggestion from Rypniewski *et al.* (1994) that those disulfide bridges are not essential, although it may affect enzymatic activity.

Glu-70 and Glu-80 are known calcium binding sites of bovine trypsin (Rypniewski *et al.*, 1994), which are conserved in the *P. leniusculus* trypsinogen. The bacterial trypsin calcium binding site, Glu-230, is present in crayfish and shrimp trypsinogen but not in vertebrates or insects (Müller *et al.*, 1993).The position of *P. leniusculus* trypsinogen in the course of evolution of serine proteinases could be located between vertebrates and bacterial trypsins because it shares structural similarities with both eukaryotic and prokaryotic enzymes. Further biochemical characteristics will help to understand the relation between structure and function of trypsins.

Acknowledgements

This work has been financed by a grant from the Swedish Science Research Council and from the Swedish Forestry and Agriculture Research Council to Kenneth Söderhäll. A Consejo Nacional de Ciencia Tecnologia fellowship (871447) and grant (3589-N) supported Patricia Hernandez.

References

Aspan, A., and Söderhäll, K. (1990). Purification of prophenoloxidase from crayfish blood cell, and its activation by an endogenous serine proteinase. Insect. Biochem. *21*, 263–373.

Eberhardt, J. (1992). Isolation and characterization of five serine proteases with trypsin-, chymotrypsin and elastase-like characteristics from the gut of the lugworm *Arenicola marina* (L.) (Polychaeta). J. Comp. Physiol. *162B*, 159–167.

Garcia-Carreño, F., Dimes, L., and Haard, N. (1993). Substrate-gel electrophoresis for composition and molecular weight of proteinases or proteinaceous proteinase inhibitors. Anal. Biochem. *214*, 65–69.

Gibson, R., and Baker, P. (1979). The decapod hepatopancreas. Oceanogr. Mar. Biol. Ann. Rev. *17*, 197–204.

Groppe, J., and Morse, D. (1993). Molluscan chymotrypsin-like protease: structure, localization and substrate specificity. Arch. Biochem. Biophys. *305*, 159–169.

Johansson, M., Keyser, P., and Söderhäll, K. (1994). Purification and cDNA cloning of a four-domain Kazal proteinase inhibitor from crayfish blood cells. Eur. J. Biochem. *223*, 389–394.

Klein, B., Le Moullac, G., Sellos, D., and Van Wormhoudt, A. (1996). Molecular cloning and sequencing of trypsin cDNAs from *Penaeus vannamei* (Crustacea, Decapoda): Use in assessing gene expression during the moult cycle. Int. J. Cell Biol. *28*, 551–563.

Müller, H., Crampton, A., della Torre, A., Sinden, R., and Crisanti, A. (1993). Members of a trypsin gene family in *Anophels gambiae* are induced in the gut by blood meal. EMBO J. *12*, 2891–2900.

Neurath, H. (1984). Evolution of proteolytic enzymes. Science *224*, 350–357.

Neurath, H. (1986). The versatility of proteolytic enzymes. J. Cell. Biochem. *32*, 35–49.

Peterson, A.M., Barillas-Mury, C.V., and Wells, M.A. (1994). Sequence of three cDNAs encoding an alkaline midgut trypsin from *Manduca sexta*. Insect Biochem. Mol. Biol. *24*, 463–471.

Rypniewski, W., Perrakis, A., Vorgias, C., and Wilson, K. (1994). Evolutionary divergence and conservation of trypsin. Prot. Engin. *7*, 57–64.

Titani, K., Sasagawa, T., Woodbury, R.G., Ericcsson, L.H., Dörsam, H., Kraemer, M., Neurath, H., and Zwilling, R. (1983). Amino acid sequence of crayfish (*Astacus fluviatilis*) trypsin I_f. Biochemistry. *22*, 1459–1465.

Worthing, V. (1993). Trypsin. In: Worthington Enzyme Manual. (Freehold, NJ, USA: Worthington Biochemical), pp. 374–391.

Zwilling, R., Neurath, H., Ericsson, L., and Enfield, D.L. (1975). The amino terminal sequence of an invertebrate trypsin (crayfish *Astacus leptodactylus*). FEBS Lett. *60*, 247–249.

Received October 19, 1998; accepted January 4, 1999