THE AMINO-TERMINAL SEQUENCE OF AN INVERTEBRATE TRYPSIN (CRAYFISH *ASTACUS LEPTODACTYLUS*): HOMOLOGY WITH OTHER SERINE PROTEASES

Robert ZWILLING, Hans NEURATH, Lowell H. ERICSSON and David L. ENFIELD
Zoologisches Institut II der Universität Heidelberg, D-69 Heidelberg, Germany, and Department of Biochemistry, University of Washington, Seattle, WA 98195, USA

Received 22 September 1975
Revised version received 23 October 1975

1. Introduction

The serine proteases represent one of the best characterized families of proteins which have evolved from a common ancestor [1]. They include such functionally diverse and distinct enzymes as the digestive proteases of the pancreas, trypsin, chymotrypsin and elastase, and the hepatic proteases of the blood coagulation system, thrombin, factor X_a and factor IX_a [2].

Plasmin [1] and kallikrein [3] also belong to this family. Among these enzymes, the trypsins, defined by common features of their active site, seem to have been preserved during the whole span of evolution ranging from bacteria to mammals. The complete covalent structures of three vertebrate trypsins, i.e., bovine [4], porcine [5] and dogfish [6] have been reported and structural details of several other vertebrate trypsins are known in part [4]. In contrast, little information is available on trypsins of invertebrate animals despite the fact that nine of ten phyla of the animal kingdom are represented by invertebrates. A proteolytic enzyme which exhibits tryptic specificity toward peptides and synthetic substrates [7,8] was found in the invertebrate crayfish *Astacus leptodactylus* and was later identified as a serine protease [9].

In contrast to bovine trypsin, the crayfish trypsin is a rather acidic protein which is irreversibly inactivated at low pH [8]. The molecular weight, calculated from the amino acid composition is 24 000 [10] and in contrast to the vertebrate trypsins crayfish trypsin contains 6 rather than 12 half-cystine residues. The enzyme occurs in multiple forms, varying in number from species to species, as determined by immunological methods [11]. Since bovine and crayfish trypsin have evolved independently for at least 500 million years, a structural analysis of the crayfish enzyme would extend significantly our knowledge of the evolution of trypsins beyond the vertebrate level and provide a link to the homologous bacterial trypsins among which that of *Streptomyces griseus* is the most thoroughly studied case [12]. The present communication describes the amino-terminal sequence of crayfish trypsin and its homologous relationship to those of other biological species.

2. Materials and methods

Crayfish trypsin (*A. leptodactylus*) was prepared as previously described [8]. The preparations contained two similar forms which differed slightly in electrophoretic mobility but showed no other differences in either composition or amino-terminal sequence (vide infra). Sequence analysis was performed with the automatic Sequencer (Beckman, Model 890) after reduction and pyridyl ethylation of disulfide bonds as described by Hermodson et al. [5].

3. Results and discussion

The sequence of the first 20 amino-terminal residues of crayfish trypsin is given in the top line of table 1 and compared to those of related serine proteases.

Table 1

Amino-terminal sequence of crayfish trypsin and other serine proteases[a]

	16				20					25					30					35
Trypsin, Crayfish	Ile	Val	Gly	Gly	Thr	Asp	Ala	Thr	Leu	Gly	Glu	Phe	Pro	Tyr	Gln	Leu	Ser	Phe	Gln	Asn
Trypsin, Bovine	Ile	Val	Gly	Gly	Tyr	Thr	Cys	Gly	Ala	Asn	Thr	Val	Pro	Tyr	Gln	Val	Ser	Leu	Asn	–
Trypsin B, Lungfish	Ile	Val	Gly	Gly	Tyr	Glu	Cys	Pro	Lys	His	Ser	Val	Pro	Trp	Gln	Val	Ser	Leu	–	–
Trypsin, Dogfish	Ile	Val	Gly	Gly	Tyr	Glu	Cys	Pro	Lys	His	Ala	Ala	Pro	Trp	Thr	Val	Ser	Leu	Asn	–
Trypsin, *Str. griseus*	Val	Val	Gly	Gly	Thr	Arg	Ala	Ala	Gln	Gly	Glu	Phe	Pro	Phe	Met	Val	Arg	Leu	Ser	Met
Chymotrypsin A, Bovine	Ile	Val	Asn	Gly	Glu	Glu	Ala	Val	Pro	Gly	Ser	Trp	Pro	Trp	Gln	Val	Ser	Leu	Gln	Asp
Elastase, Porcine	Val	Val	Gly	Gly	Thr	Glu	Ala	Gln	Arg	Asn	Ser	Trp	Pro	Ser	Gln	Ile	Ser	Leu	Gln	Tyr
Thrombin B, Porcine	Ile	Val	Glu	Gly	Gln	Asp	Ala	Glu	Val	Gly	Leu	Ser	Pro	Trp	Gln	Val	Met	Leu	Phe	Arg
Factor X_a, Bovine	Ile	Val	Gly	Gly	Arg	Asp	Cys	Ala	Glu	Gly	Glu	Cys	Pro	Trp	Gln	Ala	Leu	Leu	Val	Asn

[a] The numbering system is that of bovine chymotrypsinogen A. For references, see the text. All positions which are occupied by identical amino acid residues in crayfish trypsin and in one or more other serine protease are enclosed by boxes.

The long evolutionary distance between vertebrate and invertebrate trypsins is reflected by the observation that chymotrypsin, elastase, and even the hepatic proteases, thrombin and blood coagulation factor X_a, are structurally no less related to crayfish trypsin than are bovine and dogfish trypsin. Interestingly, the highest degree of identity seems to exist between crayfish and bacterial trypsin rather than between bacterial and any of vertebrate trypsins. Crayfish and *S. griseus* trypsin have 9 of the first 13 amino acid residues in common and of the remaining 4 amino acid substitutions, 3 can be accounted for by single base changes. The highest degree of identity exists between crayfish trypsin and bovine factor X_a, 10 of the first 20 residues occurring in identical positions.

Crayfish trypsin starts with the conservative sequence Ile–Val–Gly–Gly – common to most serine proteases. However in the following region (position 20–27) there is no obvious similarity between crayfish and the vertebrate trypsins but coincidence with the remaining serine proteases is evident. This is true of Thr-20 in bacterial trypsin and porcine elastase, of Asp-21 which has been conserved only in thrombin and factor X_a and of position 22 which is alanine in crayfish trypsin, bacterial trypsin, chymotrypsin, elastase and thrombin but half-cystine in vertebrate trypsins. Positions 25–27 are identical only in crayfish and bacterial trypsins indicative again of the remoteness of the former from the vertebrate trypsins. On the other hand, Pro-28, Gln-30 and Ser-32 are common in crayfish trypsin and in most of the other serine proteases included in this comparison. Only position 33 is unique in crayfish trypsin which contains phenylalanine instead of leucine in all other serine proteases. The present findings are also of interest in relation to the evolution of zymogens [13]. All attempts to detect in freshly collected secretions of *A. leptodactylus* a zymogen form of trypsin have failed [8] and the enzyme was fully active even in freshly prepared homogenates of the tissue where it is synthesized. It has been recently suggested that zymogens have arisen by the addition of a peptide segment to preexisting enzymes [13] and the present observations are in accord with this hypothesis.

Acknowledgements

That portion of the work which was carried out in Heidelberg was supported by a grant from the Deutsche Forschungsgemeinschaft, Bonn - Bad Godesberg. The work in Seattle was supported by a grant from the National Institutes of Health (GM 15731). We thank Fr. Brigitte Fritsch, Heidelberg, for excellent technical assistance. D.L.E. is an Associate Investigator of the Howard Hughes Medical Institute.

References

[1] de Haen, C., Neurath, H. and Teller, D. C. (1975) J. Mol. Biol. 92, 225–229.
[2] Enfield, D. L., Ericsson, L. H., Fujikawa, K., Titani, K. Walsh, K. A. and Neurath, H. (1974) FEBS Lett. 47, 132–135.
[3] Fiedler, F., Ehret, W., Hirschauer, C., Kutzbach, C., Schmidt-Kastner, G. and Tschesche, H. (1975) Abstracts of papers presented at the Fifth Linderstrøm-Lang Conference, Vingsted Conference Center, Denmark, 47.
[4] Dayhoff, M. O. (1972) Atlas of Protein Sequence and Structure, Vol. 5, National Biomedical Research Foundation, Washington, D. C.
[5] Hermodson, M. A., Ericsson, L. H., Neurath, H. and Walsh, K. A. (1973) Biochemistry 12, 3146–3153.
[6] Titani, K., Ericsson, L. H., Neurath, H. and Walsh, K. A. (1975) Biochemistry 14, 1358–1366.
[7] Pfleiderer, G., Zwilling, R. and Sonneborn, H. H. (1967) Z. Physiol. Chem. 348, 1319–1331.
[8] Zwilling, R., Pfleiderer, G., Sonneborn, H. H., Kraft, V. and Stucky, I. (1970) Comp. Biochem. Physiol. 28, 1275–1287.
[9] Tomásek, V., Sorm, F., Zwilling, R. and Pfleiderer, G. (1970) FEBS Lett. 6, 229–231.
[10] Zwilling, R. and Tomásek, V. (1970) Nature 228, 5266, 57–58.
[11] Linke, R., Zwilling, R., Herbold, D. and Pfleiderer, G. (1969) Z. Physiol. Chem. 350, 877–885.
[12] Olafson, R. W., Jurasek, L., Carpenter, M. R. and Smillie, L. B. (1975) Biochemistry 14, 1168–1177.
[13] Neurath, H. (1975) in : Cold Spring Harbor Conference on Cell Proliferation, Vol. 2, in the press.