Vol. 110, No. 1, 1983 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
January 14, 1983 Pages 1-7

AMINO ACID SEQUENCE OF AN INSECT CHYMOTRYPSIN FROM THE LARVAE OF THE HORNET, Vespa orientalis

K.-D. Jany, K. Bekelar, G. Pfleiderer and J. Ishay*

Institut für Organische Chemie, Biochemie und Isotopenforschung der Universität Stuttgart, Pfaffenwaldring 55, D-7000 Stuttgart (F.R.G.)

* Tel-Aviv University, Sackler School of Medicine, Department of Pharmacology and Physiology, Tel-Aviv (Israel)

Received October 22, 1982

SUMMARY The complete amino acid sequence of the endopeptidase II from the larvae of the hornet, Vespa orientalis, has been determined. The enzyme is a single polypeptide chain of 216 residues. The protease is a serine endopeptidase. When aligned for optimal homology to the trypsin related proteases, the insect endopeptidase displays 37% identity with bovine chymotrypsin. The structure of the hornet protease is the first reported for a serine endopeptidase from an insect.

Mammalian and bacterial serine endopeptidases have been extensively studied in view of the evolution of these enzymes (1). Although serine endopeptidases have been isolated from many different invertebrate species (2,3) there is only scare information on their covalent structure (4,5). This is probably due to the difficulties of obtaining pure protein material in sufficient yields, especially from insects. In the course of our investigation of the digestive enzymes from the hornets, Vespa orientalis and V. crabro we have isolated two chymotrypsin-like endopeptidases (6,7), and the amino acid sequences around the active site serine and histidine residues were determined of the protease II from V. orientalis, which is designated as VOP II (8). The sequences of the active site peptides show homologies to the corresponding ones of the trypsin related endopeptidases. The present communication reports the complete amino acid sequence of the protease VOP II, which can give a more detailed evaluation of the evolutionary relationship of this invertebrate protease to the serine endopeptidases of the vertebrates.

MATERIALS AND METHODS

The protease VOP II was isolated from the midgut of the larvae as described previously (6). Tryptic digestion of the citraconylated, (14-C)-carboxymethy-

0006-291X/83/010001-07$01.50/0

lated and (3-H)-diisopropyl fluorophosphate inhibited VOP II as well as the chymotryptic fragmentation of the carboxymethylated and (14-C)-Z-PheCH_2Cl inactivated VOP II were performed as in (8). Digestion of VOP II with the Staphylococcus aureus protease V 8 (30:1, w/w) was carried out in 0.1 M NH_4HCO_3, pH 7.8 at 25 °C for 4 h. Subfragmentation with these enzymes was performed at a peptide to enzyme ratio of 50:1 (w/w) in 50 mM NH_4HCO_3, pH 8 at 25 °C for 2-6 h. Thermolytic subfragments were generated at a peptide to enzyme ratio of 100:1 (w/w) in 0.1 M N-ethylmorpholin acetate, pH 8 at 55 °C for 1-4 h. Modification of the arginine residues with 1,2 cyclohexandione was performed in 0.25 M sodium borate buffer, pH 9 (9).

Initial separation of the peptides was carried out by gel filtration on Sephadex G-50f (200 x 3 cm) using 30% or 5% acetic acid as eluant. Subfragments were separated on columns of Sephadex G-50f and G-25f equilibrated with 50 mM NH_4HCO_3, pH 7.8. Final purification of the small peptides was achieved by chromatography on a Chromo Beads P column (Technicon) using a linear gradient of 50 mM pyridine acetate, pH 2.9 to 2 M pyridine acetate, pH 4.9 or by thin layer chromatography on cellulose plates CEL 300 (Macherey-Nagel) in n-butanol-pyridine-acetic acid- water (68-40-14-25, v/v).

Amino acid analysis was carried out according to Spackman (10) on a Biotronik analyzer.

Carboxyl-Terminal residues were determined by carboxypeptidase A, B or Y digestion. Sequences were determined by the dansyl-Edman method (11), and Glu/Gln or Asp/Asn were determined by aminopeptidase M digestion and by thin layer chromatography of the PTH-derivatives.

RESULTS

The covalent structure of the hornet protease VOP II is given in Fig. 1, together with the peptides necessary to establish the sequence. The structure

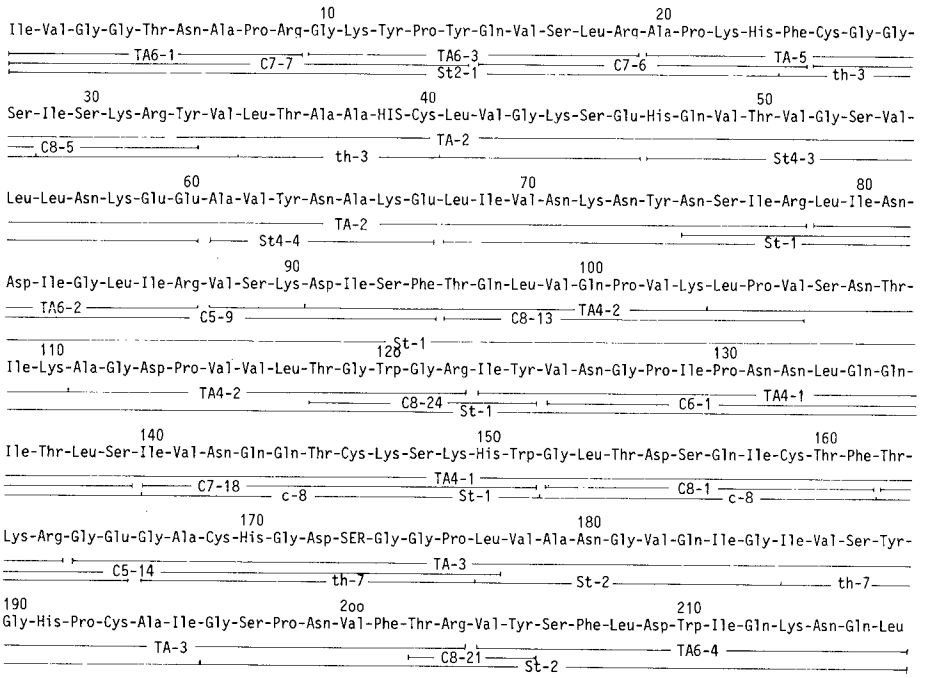

Fig. 1 Amino acid sequence of the hornet chymotrypsin VOP II
TA - trypsin after citraconylation, C - chymotrypsin, St - Staphylococcus aureus protease V 8, th - thermolytic subfragment

was constructed from three sets of overlapping peptides, generated by trypsin after modification of the lysine residues (TA), chymotrypsin (C) and staphylococcal protease V 8 (St). The large fragments (the staphylococcal peptides St-1, St2-1) after modification of the arginine residues) were subdigested with these enzymes.

The bulk of the sequence work was carried out on the arginine fragments. The peptides TA-2 to TA-5 were obtained by gel filtration and TA6-1 to TA6-4 were purified on a Chromo Beads P column. The sequences of TA-2 and TA-3 have been already determined (8). The staphylococcal peptides St4-3, St4-4 and St1-t9 (Asn-75 - Lys-90) now gave the necessary overlaps within TA-2, and St1-t9/C5-9 linked TA-2 to TA4-2. Tryptic subfragmentation of TA4-2 yielded four peptides and the sequence of TA4-2 could be completed by the chymotryptic fragments C5-9, C8-13 and C8-24. C8-24 overlapped TA4-2 and TA4-1. Fragment TA4-1 is completely insoluble, and only the first 22 residues could be determined. Digestion of TA4-1 with trypsin generated again an insoluble fragment, which was found to be the aminoterminal part of TA4-1 and the soluble peptides His-150 - Lys-164, Ser-Lys and free arginine. The alignment of these peptides could be made by the chymotryptic fragments and thermolytic peptides of St-1. C5-14 provided the overlap of TA4-1 and TA-3.

Since VOP II contains no free thiol groups (6), the 6 cysteine residues must be incorporated in three disulfide bridges. This was confirmed by the staphylococcal peptides which were obtained from VOP II with intact disulfide bonds. The thermolytic subfragment St2-1th-3 indicated, that Cys-25 and Cys-40 are involved in a disulfide bond. From the chymotryptic digest of the large St-1-fragment the peptide St-1c-8 was isolated. Thus the second disulfide bond is formed by Cys-146 and Cys-159. The third disulfide bond is Cys-169 - Cys-193 as found from the peptide St-2th-7.

DISCUSSION

The protease VOP II is a single polypeptide chain of 216 amino acid residues with three disulfide bridges linking residues 25 to 40, 146 to 159 and 169-193. The amino acid composition of VOP II (Tab. 1) is in close agreement with

Table 1 Amino acid composition of VOP II

	Amino acid analysis*	Sequence
Asp	21.2	21
Thr	10.5	12
Ser	15.2	16
Glu	16.6	17
Pro	9.8	11
Gly	21.6	22
Ala	9.8	10
Cys/2**	6.2	6
Met	0	0
Val	19.9	22
Ile	16.7	19
Leu	16.3	16
Tyr	6.8	8
Phe	4.9	5
Lys	13.6	14
His	5.9	6
Arg	7.8	8
Trp***	3.1	3

* Hydrolysis in 6 N HCl at 108 °C for 24, 72, 96 and 144 h.
** Determined as cysteic acid
*** Hydrolysis in 3 M mercaptoethane sulfonic acid

the amino acid sequence. The molecular weight calculated from the sequence is 23411. This value is in excellent agreement with the molecular weight of 23000 estimated by sodium dodecylsulfate disc electrophoresis (6). However, this molecular weight is approximately twice that determined by gel filtration and sedimentation equilibrium analysis (6). At present, no satisfactory explanation is available for such a large discrepancy. Probably VOP II retards on the gel matrix and a too low molecular weight was determined. These retardation effects are often observed on endopeptidases (12). Also a wrong molecular weight is determined by the sedimentation equilibrium analysis, if the partial specific volume of VOP II differs significantly from that generally estimated for the majority of proteins.

VOP II is a serine protease and exhibits a substrate specificity as bovine chymotrypsin (6). When the sequence of VOP II is aligned to those with bovine chymotrypsin, trypsin and porcine elastase (Fig.2) a homologous relationship

```
                20                  30                  40                  50                  60
VOP II  I V G G T N A P R G K Y P Y Q V S L R - - - - - A P K H F C G G S I S K R - Y V L T A A H C L V G
CA      I V N G E E A V P G S W P W Q V S L Q D K T G F - - - H F C G G S L I N E N W V V T A A H C G V T
E       V V G G T E A Q R N S W P S Q I S L Q Y R S G S S W A H T C G G T L I R Q N W V M T A A H C V D R
T       I V G G Y T C G A N T V P Y Q V S L - - N S G Y - - - H F C G G S L I N S Q W V V S A A H C Y K S

                        70                  80                  90                  100
VOP II  K S E H Q V T V - - G S V - L L N K E E A V Y - - N A E E L I V N K N Y N S I R L I - - N D I G L
CA      T S D - - - V V V A G E F D Q G S S S E K I Q K L K I A K V F K N S K Y N S L T I N - - N D I T L
E       E L T F R - - V V V G E H N L N Q N N G T E Q Y V G V Q K I V V H P Y W N T D D V A A G Y D I A L
T       G I Q V R L - - - - G Q D N I N V V E G N Q Q F I S A S K S I V H P S Y N S N T - - A N N D I M L

                110                 120                 130                 140                 150
VOP II  I R V S K D I S F T Q L V Q P V K L P - - - V S N T I K A G D P V V L T G W G - - R I Y V N G P I
CA      L K L S T A A S F S Q T V S A V C L P S - - A S D D F A A G T T C V T T G W G L T R - Y T N A N T
E       L R L A Q S V T L N S Y V Q L G V L P R - - A G T I L A N N S P C Y I T G W G L T R - - T N G Q L
T       I K L K S A A S L N S R V A S I S L P T S C A S - - - - A G T Q C L I S G W G N T K - S S G T S Y

                        160                 170                 180                 190
VOP II  P N N L Q Q I T L S I V N Q Q T C K - S K H W G - - L T D S Q I C T F - T K R G E G A C H G D S G
CA      P D R L Q Q A S L P L L S N T N C K K - - Y W G T K I K D A M I C A G - A S G V - S S C M G D S G
E       A Q T L Q Q A Y L P T V D Y A I C S S S S Y W G S T V K N S M V C A G - G N G V R S G C Q G D S G
T       P D V L K C L K A P I L S N S S C K - S A Y P G - Q I T S N M F C A G Y L E G G K D S C Q G D S G

                200                 210                 220                 230                 240
VOP II  G P L V A - - N G - - V Q I G I V S Y G H - - P C A I G S - P N V F T R V Y S F L D W I Q K N Q L
CA      G P L V C K K N G A W T Q I G I V S W G S S - T C S - T S T P G V Y A R V T A L V N W I Q - Q T L A A N
E       G P L H C L V N G Q Y A V H G V T S F V S R L G C N V T R K P T V F T R V S A Y I S W I N - N V I A S N
T       G P V V C - - S G K - - L Q G I V S W G S - - G C A Q K N K P G V Y T K V C N Y V S W I K - Q T I A S N
```

Fig. 2 Comparison of the sequence of the hornet chymotrypsin (VOP II) with those of bovine chymotrypsin A (CA), trypsin (T) and porcine elastase (E). The sequences of the vertebrate proteases are taken from deHaen et al.(1). Residues are numbered according to the chymotrypsinogen numbering system.

is evident to the mammalian enzymes. The degree of sequence identity is ~ 37%. The catalytic essential residues His-57 - Asp-102 - Ser-195*, forming the charge relay system of bovine chymotrypsin are located in the corresponding regions of VOP II and as in all other serine proteases, the sequences around these residues are highly conserved. Also, the sequence Ser-Trp-Gly-216 and isoleucine-99 in the secondary substrate binding site of bovine chymotrypsin (13)correspond to Ser-Tyr-Gly-216 and leucine-99 in VOP II. In addition, the three disulfide bonds of VOP II occur in homologous positions in the mammalian enzymes. They correspond Cys-42 - Cys-58, Cys-168 - Cys-182 and Cys-191 - Cys-220. The degree of sequence identity and the structural features indicate that the hornet protease II is closely related to the serine proteases of the trypsin-type. It is likely that this insect protease, which can be also designated as hornet chymotrypsin, and the vertebrate enzymes have evolved from a common ancestral serine protease.

* The chymotrypsinogen numbering is used throughout in marking comparisons with the amino acid sequences of other proteases.

Despite these similarities, VOP II shows some differences to the vertebrate proteases. In bovine chymotrypsin and trypsin, serine-189 and aspartic acid, respectively, are placed on the bottom of the primary substrate binding site. In VOP II and also in the crab collagenase (14), this position occupied by a glycine. The glycine-226 of bovine chymotrypsin and trypsin is replaced by an asparagine in VOP II. In elastase this residue is a threonine, which hinders, at least in part, the binding of "large" hydrophobic amino acid side chains. But in VOP II the specificity for such amino acids is not affected by the asparagine-226.

The amino-terminal Ile-16 and aspartic acid-194 form an ion pair in bovine chymotrypsin, and both residues are also found in the corresponding regions of VOP II. But in VOP II, this salt bridge seems to be not essential for the active conformation as the amino group of Ile-16 can be modified without any loss of enzymic activity (unpublished results).

Three disulfide bonds are present in VOP II which are conserved in all other serine endopeptidases. However, the vertebrate proteases, with exception of the group-specific protease GSP (15), have always more than three disulfide bridges. Three disulfide bonds seem to be characteristic for invertebrate proteases (4,14). DeHaën et al.(1) have pointed out, that the number and location of the disulfide bridges can be used for the estimation of the evolutionary relationship of proteins. Based on this criterion, the point of divergence of the progenitor of VOP II should be placed between the bacterial trypsin and the vertebrate proteases. However, since the crayfish trypsin has also three disulfide bridges (4) in homologous positions as VOP II, the question arises whether the chymotryptic cleavage specificity has developed several times during the evolution.

ACKNOWLEDGMENTS

This work has been supported by the Deutsche Forschungsgemeinschaft. We wish to thank Mrs. E. Galaboff and Mrs. E. Nitsche for their skillful technical assistance.

REFERENCES

1. DeHaën, G., Neurath, H. and Teller, D.C. (1975) J.Mol.Biol. 92, 225-259
2. Law, J.H., Dunn, P.E. and Kramer, K.J. (1977) Advances in Enzymol.45, 389-425
3. Pfleiderer, G. and Zwilling, R. (1972) Naturwissenschaften 59, 396-405
4. Zwilling, R. and Neurath, H. (1981) Methods Enzymol. 80C, 633-664
5. Lecroisey, A., deWolf, A. and Keil, B. (1980) Biochem.Biophys. Res.Commun. 94, 1261-1265
6. Jany, K.-D., Pfleiderer, G. and Molitoris, H.-P. (1974) Eur.J.Biochem. 42, 419-428
7. Jany, K.-D., Haug, H., Pfleiderer, G. and Ishay, J. (1978) Biochemistry 17, 4675-4682
8. Jany, K.-D., Bekelar, K. and Ishay, J. (1981) Biochim.Biophys.Acta 668, 197-200
9. Patthy, L. and Smith, E.L. (1975) J.Biol.Chem. 250, 557-564
10. Spackman, D.H. (1967) Methods Enzymol. 11, 3-27
11 Chen, R. (1976) Hoppe-Seyler's Z. Physiol.Chem. 357, 873-886
12 Voordouw, G. Gaucher, G.M. and Roche, R.S. (1974) Biochem. Biophys.Res. Commun. 58, 8-12
13. Segal, D.M., Powers, J.C., Cohen, G.H., Davies, D.R. and Wilcox, P.E. (1971) Biochemistry 10, 3728-3738
14. Grant, G.A., Henderson, K.O., Eisen, A.Z. and Bradshaw, R.A. (1980) Biochemistry 19, 4653-4659
15 Woodbury, R.G., Katunuma, N., Kobayashi, K., Titani, K. and Neurath, H. (1978) Biochemistry 17, 811-819