Volume 158, number 1 FEBS 0584 July 1983

Amino acid sequence of the chymotryptic protease II from the larvae of the hornet, *Vespa crabro*

Klaus-Dieter Jany and Harald Haug

Institut für Organische Chemie, Biochemie und Isotopenforschung der Universität Stuttgart, Pfaffenwaldring 55, D-7000 Stuttgart 80 FRG

Received 31 May 1983

The covalent structure of the chymotrypsin II from the larvae of the hornet, *Vespa crabro*, has been determined. The sequence has been deduced from 3 sets of overlapping peptides generated by trypsin after modification of the lysine or arginine residues and by chymotrypsin. The enzyme is a serine endopeptidase and contains 218 residues in a single polypeptide chain cross-linked by 3 disulfide bonds. Alignment of the sequence of this insect protease with those of chymotrypsin, elastase and trypsin shows about 35% identity with each and a homologous relationship is evident.

Serine endopeptidase *Chymotrypsin* *Invertebrate protease* *Evolution* *Amino acid sequence*

1. INTRODUCTION

Comparative studies of a functionally analogous protein from widely different phylogenetic sources can offer insights into the structure–activity relations and evolution. In this respect, the proteolytic enzymes of bacteria and vertebrates have been intensively investigated [1]. Much less is known about the structure of the invertebrate proteases [2,3], especially from insects, although 9 of the 10 phyla of the animal kingdom are represented by invertebrates.

We have isolated and characterized a chymotrypsin-like protease (henceforth VCP II) from *Vespa crabro* larvae [4]. Here, we report the complete covalent structure of this insect protease. This investigation was done to compare the sequence with those of the trypsin-related enzyme [1] which would extend the knowledge of the evolution of the serine endopeptidases beyond the vertebrate level.

2. MATERIALS AND METHODS

The procedures employed for the isolation of VCP II from the larval midgut, the inactivation of the enzyme with [^{3}H]diisopropyl fluorophosphate (DFP) or carbobenzyoxy-[^{14}C]phenylalanine methylchloroketone (Z-PheCH_2Cl), the *S*-carboxymethylation with iodo-[^{14}C]acetic acid and blocking of the lysine residues with citraconic anhydride were reported in [4]. Modification of the arginine residue with 1,2-cyclohexandione (CHD) (20 mg Cm-VCP II in 0.25 M sodium borate (pH 9.0); 75 μmol CHD) was performed according to [5].

Tryptic digestion of the citraconylated CM-*Z*-[^{14}C]PheCH_2Cl-VCP II or CHD-modified protease was carried out at 30:1 (w/w) substrate:enzyme in 0.1 M *N*-ethylmorpholine acetate (pH 8.5) at 35°C for 3 h. Tryptic subfragments were generated in 0.1 M NH_4HCO_3 (pH 7.8) (peptide:enzyme, 50:1, w/w) at 25°C for 1–6 h depending on the solubility of the fragments.

Chymotryptic fragmentation of [^{14}C]CM-[^{3}H]-DFP-VCP II was performed in 50 mM NH_4HCO_3 (pH 7.8) at 25°C at 100:1 protein:enzyme (w/w) for 40 min. Subdigestions were carried out at 50:1 (w/w) peptide:enzyme for 3 h. Subdigestions of the tryptic fragments (TA, TL) and chymotryptic

Published by Elsevier Science Publishers B.V.

peptides (C) with *Staphylococcus aureus* V8 protease was done in 50 mM NH_4HCO_3 (pH 7.8) at 35°C for 4–16 h.

Initial separation of the fragments was carried out on a Sephadex G-50f column (200 × 3 cm) equilibrated in 30% or 5% acetic acid. Subfragments were generally chromatographed on Sephadex G-50f or G-25f using 50 mM NH_4HCO_3 (pH 7.8) as eluant. Final purification of small peptides was achieved by chromatography on a chromo Beads P column (30 × 0.6 cm) or by thin layer chromatography as in [6]. The dansyl-Edman tech-

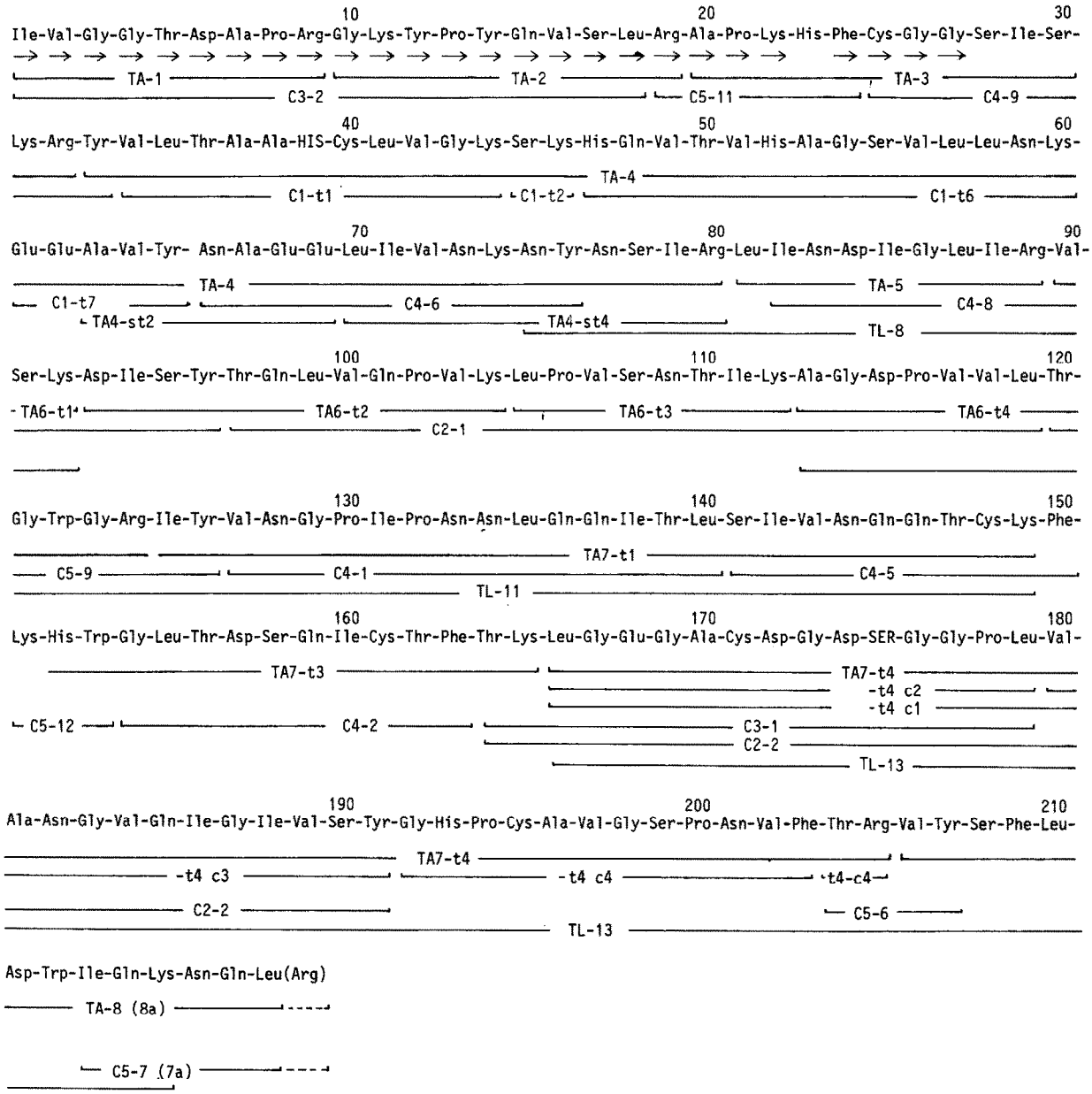

Fig.1. Amino acid sequence of the hornet chymotrypsin VCP II: TA, trypsin after blocking the lysine residues; TL, trypsin after modification of the arginine residues; C, chymotrypsin; small letters indicate subfragments; st, *Staphylococcus aureus* V8 (→) sequencer run of the intact CM-VCP II.

nique was used to sequence the peptides. The further procedures were as in [4,6,7].

3. RESULTS

The complete amino acid sequence of the hornet protease VCP II is shown in fig.1 together with the key peptides used for its elucidation. The structure was primarily deduced from analyses of the tryptic and chymotryptic peptides prepared from the intact, modified and inactivated enzyme. The protease is a single polypeptide chain consisting of 218 amino cid residues. The amino acid composition of VCP II based on M_r 24 000 is in a close agreement with the residues found in the structure. The homology which exists between the sequences of VCP II and the protease II from *Vespa orientalis* [6] helped us to align the fragments. Gel filtration of the tryptic digest of CM-Z-[^{14}C]PheCH$_2$Cl-VCP II on Sephadex G-50f yielded the pure peptides TA-7 (125–205), TA-4 (33–80), TA-6 (90–124) and two fractions of the small peptides TA-1, -2, -3, -5, -8 and -8a which were separated by chromatography on Chromo Beads P. TA-6 was subfragmented with trypsin and the resulting peptides TA6-t1 to -t4 were isolated and sequenced. Its final ordering was achieved by the peptides C4-8 and C2-1 obtained from the chymotryptic digest of the intact protease. TA-4 contains the active-site histidine residue and the labelled residue (His-39) was clearly identified by testing the butylacetate phases for radioactivity during the analysis of the intact peptide or its chymotryptic subfragments. Subfragmentation of the insoluble

```
                        10                  20                  40
VCP II  I V G G T D 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 K S K H
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 K S E H
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 T S D -
            20                  40                  60

                    60                  80
VCP II  Q V T V H A 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 I R V S K D I
VOP II  Q V T V - - G S V - L L N K E E A V Y - - N A K E L I V N K N Y N S I R L I N D I G L I R V S K D I
CA      - - 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 L K L S T A A
                        80                  100

            100                 120                 140
VCP II  S Y 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 P N N L Q Q I T L S
VOP II  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 P N N L Q Q I T L S
CA      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 P D R L Q Q A S L P
                120                 140                 160

                        160                 180
VCP II  I V N Q Q T C K - F K H W G - - S T D S Q I C T F T K L G E G A C D G D S G G P L V A - - N G - - V
VOP II  I V N Q Q T C K - S K H W G - - S T D S Q I C T F T K R G E G A C H G D S G G P L V A - - N G - - V
CA      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 G P L V C K K N G A W T
                        180                 200

                    200                 218
VCP II  Q I G I V S Y G H P - C A V G S - P N V F T R V Y S F L D W I Q K N Q L(R)
VOP II  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      L V 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
                220                 240
```

Fig.2. Comparison of the sequence of hornet chymotrypsin VCP II with those of VOP II and bovine chymotrypsin (CA). Differences between VCP II and VOP II are underlined: (top) VCP II numbering; (bottom) chymotrypsinogen numbering. The sequence of bovine chymotrypsin is taken from [1].

TA-7 with trypsin yielded again in insoluble peptide (TA7-t1) and as soluble fragments -t3, -t4 and the dipeptide Phe-Lys. TA7-t4 (166–205) was further digested with chymotrypsin and the subfragments -t4c1 to -tac5 were purified from the digest. The peptides -t4c2/c3 were generated by an incomplete cleavage of the Leu–Val bond 179–180. Two ^{3}H-radioactively labelled peptides (C2-2, C3-1) were obtained from the chymotryptic digest of the [^{14}C]CM-[^{3}H]DFP inactivated protease. These peptides linked TA7-t3 and -t4 and corresponded to the C-terminal part of TA7-t4c1/c2. The position of the labelled residue was indentified by counting the butylacetate phases and the nature of this residue (Ser-175) was also established by sequencing the peptides TA7-t4c1/c2. Most of the arginine fragments as well as their subfragments could be alinged by the chymotryptic peptides C. The chymotryptic subfragments of TA-4 were ordered by the tryptic peptides of C-1 and by the fragments obtained after digestion of VCP II with modified arginine residues. The C-terminus of VCP II was found to be heterogeneous. All peptides isolated from this region terminate either in –Leu or –Leu–Arg. This heterogeneity had led to the misinterpretation in the M_r-detemination of VCP II by carboxypeptidase digestions [4].

4. DISCUSSION

In fig.2 the sequence of VCP II is aligned with those of the protease II from *Vespa orientalis* (VOP II) and bovine chymotrypsin. Not surprisingly, the structures of the insect proteases obtained from such closely related species are nearly identical. Only 8 loci are changed whereby 5 replacements are not conservative (res. 46, 68, 150, 166, 172). The substitution of Arg→Leu-166 enables the purification of the large arginine fragments (TA-7, -4, -6) of VCP II by simple gel filtration whereas the corresponding peptides of VOP II are nearly equal in size. The replacement of Asp→His-172 probably reflects the moleclar basis of the large differences in the catalytic constants for the hydrolysis of synthetic substrates; e.g., glut-Phe-pNan, $k_{cat} = 8.6 \times 10^{-2}\ s^{-1}$ (VCP II), $92.9 \times 10^{-2}\ s^{-1}$ (VOP II) [8]. The homology between the sequences of both insect proteases justifies the conclusion that in VCP II the cysteines 25–40, 148–161 and 171–195 are linked by disulfide bonds. However, the deletion 52, 53 is surprising in the structure of VOP II. Normally, no deletions are found in the aligned sequences of the same proteases in closely related species. In this respect, both the insect enzymes differ more than the mammalian chymotrypsins [1].

VCP II is a serine endopeptidase and exhibits the same cleavage specifity as bovine chymotrypsin [4]. Therefore the sequences of both enzymes were aligned. Due to the different chain length of the insect and mammalian enzymes (218 and 230 residues, respectively), the alignment requires 12 deletions and 5 insertions in the sequence of VCP II (fig.2). After this alignment both enzymes differ in 151 amino acid residues which corresponds to 37% identity. In addition, for other serine endopeptidases the following identities are calculated: porcine elastase/VCP II 38%, bovine trypsin/VCP II 31%, mast cell protease II [9] 31% and crayfish trypsin [10] 37%. The homology is not only extended to the catalytic essential residues His(57)-Asp(102)-Ser-(195) (chymotrypsinogen numbering is used), but also the 3 disulfide bonds correspond to those of the His-loop (42–58), Met-loop (168–182) and Ser-loop (191–220). The degree of sequence identity clearly indicates the relationship of the insect protease to the trypsin-related enzymes. It is likely that VCP II, now designated as hornet chymotrypsin, and the vertebrate enzymes have evolved from a common ancestral serine protease. Since amino acid replacements occur more frequently than gaps or changes in the number or location of disulfide bonds in the aligned sequences of homologous proteins [11], these facts are used to estimate the points of divergence of the serine endopeptidases [1]. Based on these criteria, the progenitor of hornet chymotrypsin may have arisen before those of the vertebrate trypsins and the chymotrypsin family (chymotrypsin, elastase, mast cell protease II). The characteristic Ile–Val–Gly–Gly– amino-terminus is conserved in the vertebrate and invertebrate trypsins as well as in hornet chymotrypsin. But this region is changed in the vertebrate chymotrypsin-family [1,9]. The substitutions may have occurred after the lines of development leading to hornet chymotrypsin and to the vertebrate enzyme have separted from one another. The same may apply to the additional disulfide bridges of the vertebrate

enzymes. But regarding the crayfish trypsin which has 3 disulfide bonds in the same positions [10] as hornet chymotryposin, the point of divergence of the trypsin- and chymotrypsin-family is rather unclear. But the possibility exists that the chymotryptic cleavage specificity has evolved several times as still a mutation at aspartic acid (189) at the primary substrate binding site of trypsin can lead to a chymotrypsin-like enzyme.

ACKNOWLEDGEMENTS

This work was supported by the Deutsche Forschungsgemeinschaft. We thank Mrs E. Galaboff and Mrs E. Nitsche for their skilful tecnical assistance. We are also indebted to Professor Ponstingl, Deutsches Krebsforschungszentrum, Heidelberg, for using his sequencer.

REFERENCES

[1] De Haën, 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) Adv. Enzymol. 45, 389–425.

[3] Zwilling, R. and Neurath, H. (1981) Methods Enzymol. 80C, 633–664.

[4] Jany, K.-D., Haug, H., Pfleiderer, G. and Ishay, J. (1978) Biochemistry 17, 4675–4682.

[5] Toi, K., Bynum, E., Norris, E. and Itano, A. (1967) J. Biol. Chem. 242, 1036–1043.

[6] Jany, K.-D., Bekelaer, K., Pfleiderer, G. and Ishay, J. (1983) Biochem. Biophys. Res. Commun. 110, 1–7.

[7] Jany, K.-D., Bekelaer, K. and Ishay, J. (1981) Biochim. Biophys. Acta 668, 197–200.

[8] Jany, K.-D. (1980) Habilitationsschrift, Universität Stuttgart.

[9] Woodbury, R.G., Katunuma, N., Kobayashi, K., Titani, K. and Neurath, H. (1978) Biochemistry 17, 811–819.

[10] Titani, K., Sasagawa, T., Woodbury, R.G., Ericsson, L.H., Dörsam, H., Kramer, M., Neurath, H. and Zwilling, R. (1983) Biochemistry 22, 1459–1465.

[11] Dayhoff, M.O. (1972) Atlas of Protein Sequence and Structure, vol. 5, Nat. Biomed. Res. Found., Washington DC.