Volume 258, number 1, 109–112 FEB 07856 November 1989

Papaya proteinase IV amino acid sequence

Anka Ritonja, David J. Buttle+, Neil D. Rawlings+, Vito Turk and Alan J. Barrett+

Department of Biochemistry, J. Stefan Institute, Jamova 39, 61000 Ljubljana, Yugoslavia and *+Biochemistry Department, Strangeways Research Laboratory, Worts Causeway, Cambridge CB1 4RN, England*

Received 4 October 1989

The amino acid sequence of papaya proteinase IV (PPIV), a major proteinase from the latex of *Carica papaya* [(1989) Biochem. J. 261, 469–476] is described. The enzyme has a high degree of sequence identity with papaya proteinase III, chymopapain and papain (81, 70 and 67%, respectively), and is clearly a member of the papain superfamily of cysteine proteinases. Nevertheless, the sequence shows substitution of certain residues conserved in all other known members of the superfamily. It is suggested that some of these substitutions may account for the unusual specificity of PPIV.

Proteinase IV; Amino acid sequence; Cysteine proteinase; Papain; (Papaya)

1. INTRODUCTION

The application of a new affinity ligand to the fractionation of cysteine proteinases from papaya latex led to the discovery of an enzyme named papaya proteinase IV [1]. This has a substrate specificity very different from the other known papaya cysteine proteinases, and is not inhibited by chicken cystatin [1]. We now report the complete amino acid sequence of the enzyme, showing that it is a member of the papain superfamily of proteins, but with some unusual amino acid substitutions.

2. EXPERIMENTAL

The chemicals used for Edman degradation were of Sequenal grade from Applied Biosystems. Iodo[^{3}H]acetic acid was from Amersham, and endoproteinase Lys-C from Boehringer Mannheim. β-Trypsin was prepared according to [2]. All other chemicals were of analytical grade.

PPIV was purified from papaya latex as described [1], except that elution from the affinity column was with 2,2′-dipyridyl disulphide (30 mM) in place of hydroxyethyl disulphide. Amino acid analysis was by post-column derivatization with *o*-phthalaldehyde (Fluka).

Peptides were purified by gel chromatography and HPLC on Applied Biosystems and/or Bio-Rad C_{18} columns eluted with aqueous acetonitrile containing trifluoroacetic acid. Samples were sequenced with an Applied Biosystems gas phase sequenator model 470A [3]. Phenylthiohydantoin derivatives were identified on-line with the 120A HPLC [4]. Reversible blocking of amino groups with maleic anhydride and unblocking were as described [5].

Correspondence address: A.J. Barrett, Department of Biochemistry, Strangeways Research Laboratory, Worts Causeway, Cambridge CB1 4RN, England

Abbreviations: $NHPhNO_2$, *p*-nitroanilide; $OPhNO_2$, *p*-nitrophenyl ester; PPIV, papaya proteinase IV

3. RESULTS AND DISCUSSION

Native, active enzyme gave a single sequence up to 17 residues (fig.1). The protein was reduced with 2-mercaptoethanol in 6 M guanidine-HCl, and carboxymethylated with iodo[^{3}H]acetic acid for determination of the complete sequence as summarized in fig.1. The first set of peptides was generated by arginine-specific cleavage with β-trypsin after maleylation. Nine peptides (R1–R9) were isolated and sequenced. Two of the peptides, R2 (residues 9–65) and R7 (residues 140–192) could not be sequenced completely because of their length, and were therefore de-maleylated and cleaved with trypsin at the lysine residues. R2 yielded the subfragments T1–T3, and R7 gave T4–T8. Peptide R9 was concluded to be the C-terminal peptide, since its C-terminal residue is asparagine, which would not have been a site of cleavage by trypsin.

To determine the order of the tryptic peptides, a second digest of the whole protein was made with endoproteinase Lys-C. Only the peptides containing arginine (K1–K5) were sequenced.

The results show that PPIV exists as a single polypeptide chain with 216 residues, and M_r 23313 (assuming 3 disulphide bonds (fig.1). No site of potential glycosylation was detected.

The alignment of the amino acid sequence of PPIV with those of other plant cysteine proteinases (fig.2) shows unequivocally that the enzyme is a member of the papain superfamily, containing many residues identical with those in papaya proteinase III (81%), chymopapain (70%) and papain (67%). No insertion or deletion of residues with respect to the other 3 papaya sequences was apparent, apart from that associated

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/89/$3.50

Fig.1. Amino acid sequence and composition of PPIV, and strategy of the sequence determination. The amino acid sequence of the *S*-[^{3}H]carboxymethylated protein was determined by automated Edman degradation of the native enzyme, the 9 tryptic peptides from the maleylated protein (R1–R9), 8 sub-fragments derived from R2 and R7 by further digestion with trypsin after de-maleylation (T1–T8) and 5 peptides produced by digestion with endoproteinase Lys-C (K1–K5). Cysteine residues were identified by the radioactivity of the *S*-[^{3}H]carboxymethylated derivative.

with the deletion of a tetrapeptide at residues 169–170 in papain.

It can now be seen that the N-terminal 17 residues of a protein called papaya peptidase B [7] are identical with those of papaya proteinase IV. Papaya peptidase B was found in the fraction from papaya latex that did not bind to an agarose-mercurial column; its properties resembled those of PPIV in that it did not hydrolyse Bz-Arg-NHPhNO_2, but cleaved a -Gly-OPhNO_2 substrate [13]. Also, a partial cDNA sequence from a *C. papaya* leaf tissue library (pLBPc18) [8] codes for the C-terminal 62 residues of papaya proteinase IV (fig.2).

All of the features thought to be essential to the catalytic mechanism of papain are also present in PPIV, including Asn^{175} (missing in stem bromelain [14]) to hydrogen-bond and orient the imidazole ring of His^{159} [15].

Previous work [1] has shown that the substrate specificity of PPIV is very different from that of pa-

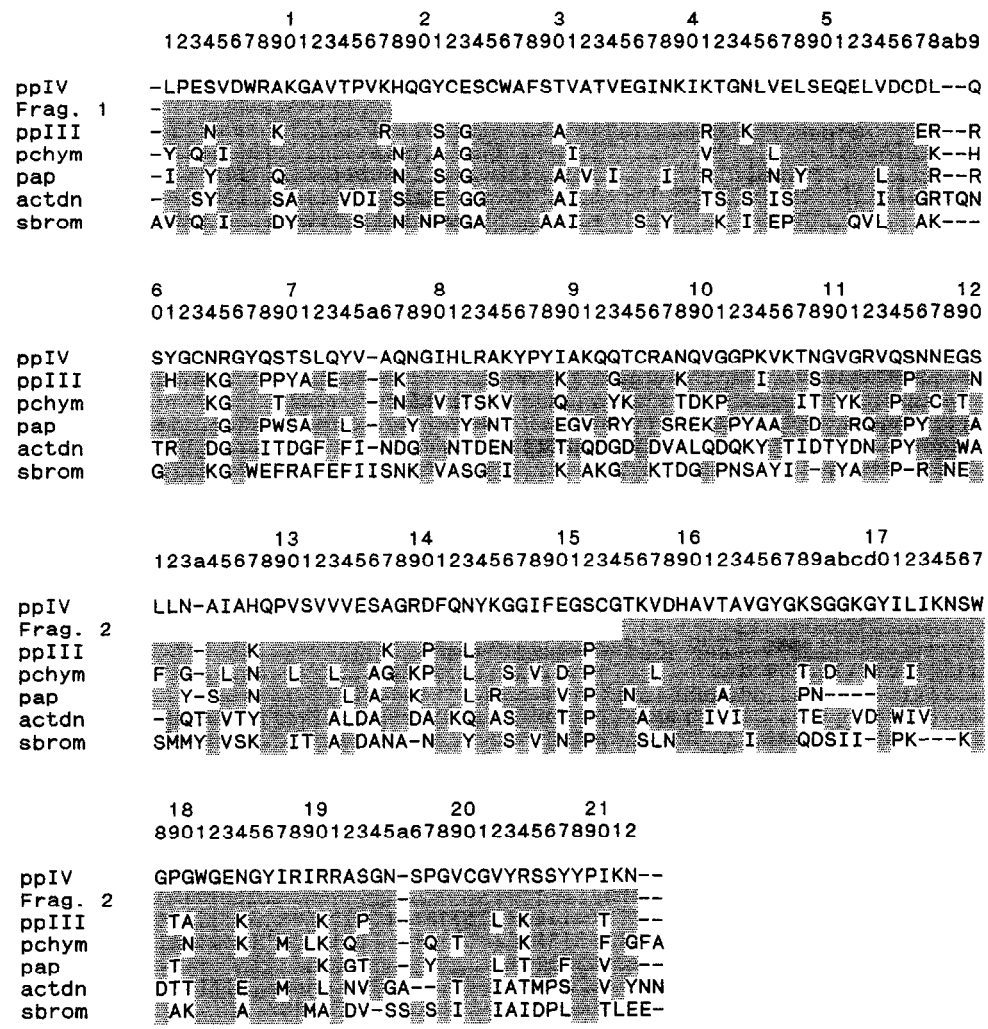

Fig.2. Alignment of the papaya proteinase IV (PPIV) sequence with those of other plant cysteine proteinases. Residues identical to those in PPIV are shaded in the other sequences, and the residue numbering is according to the scheme for papain [10]. For several of the sequences, alternative residues have been reported at a few positions. Key: (Frag. 1) an N-terminal sequence from papaya latex attributed to 'papaya peptidase B' [7]; (Frag. 2) sequence derived from the cDNA of plasmid pLBPc18 (Leicester Botany papaya clone 18) [8]; (ppIII) papaya proteinase III [6]; (pchym) chymopapain [9]; (pap) papain [10]; (actdn) actinidin [11,12]; and (sbrom) stem bromelain [14].

pain, and more restricted. Thus, the enzyme produced only limited cleavage of casein, and was inactive on Bz-Arg-$NPhNO_2$, although it hydrolysed Boc-Gly-$OPhNO_2$. These properties would be consistent with crowding of the specificity sites, and the substitution of Gly^{23} by Glu is particularly interesting in this connection. Gly^{23} is totally conserved in the 20 or so homologous sequences reported to date, and it has been pointed out that a side chain in this position would seriously crowd the active site cleft [15]. The replacement of the highly conserved Gly^{65} by Arg, in the wall of the active site cleft, may also lead to restriction of space for the substrate molecule.

Acknowledgement: We thank the Research Council of Slovenia (Grant C1-0515) for financial support of part of this work.

REFERENCES

[1] Buttle, D.J., Kembhavi, A.A., Sharp, S., Schute, R.E., Rich, D.H. and Barrett, A.J. (1989) Biochem. J. 261, 469–476.
[2] Strop, P. and Cechova, D. (1981) J. Chromatogr. 207, 55–62.
[3] Hewick, R.M., Hunkapiller, M.W., Hood, L.E. and Dreyer, W.J. (1981) J. Biol. Chem. 256, 7990–7997.
[4] Hunkapiller, M.W. and Hood, L.E. (1983) Methods Enzymol. 91, 486–493.
[5] Butler, P.J.G. and Hartley, B.S. (1972) Methods Enzymol. 25, 191–199.
[6] Dubois, T., Kleinschmidt, T., Schnek, A.G., Looze, Y. and Braunitzer, G. (1988) Biol. Chem. Hoppe-Seyler 369, 741–754.
[7] Lynn, K.R. and Yaguchi, M. (1979) Biochim. Biophys. Acta 581, 363–364.
[8] McKee, R.A., Adams, S., Matthews, J.A., Smith, C.J. and Smith, H. (1986) Biochem. J. 237, 105–110.
[9] Jacquet, A., Kleinschmidt, T., Schnek, A.G., Looze, Y. and Braunitzer, G. (1989) Biol. Chem. Hoppe-Seyler 370, 425–434.

[10] Cohen, L.W., Coghlan, V.M. and Dihel, L.C. (1986) Gene 48, 219–227.
[11] Carne, A. and Moore, C.H. (1978) Biochem. J. 173, 73–83.
[12] Baker, E.N. (1980) J. Mol. Biol. 141, 441–484.
[13] Lynn, K.R. (1979) Biochim. Biophys. Acta 569, 193–201.
[14] Ritonja, A., Rowan, A.D., Buttle, D.J., Rawlings, N.D., Turk, V. and Barrett, A.J. (1989) FEBS Lett. 247, 419–429.
[15] Baker, E.N. and Drenth, J. (1987) in: Biological Macromolecules and Assemblies, vol.3 (Jurnak, F.A. and McPherson, A. eds) pp.313–368, John Wiley and Sons, New York.