Volume 247, number 2, 419–424 FEB 07077 April 1989

Stem bromelain: amino acid sequence and implications for weak binding of cystatin

Anka Ritonja*, Andrew D. Rowan, 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 Laboratory, Worts Causeway, Cambridge CB1 4RN, England

Received 10 March 1989

The amino acid sequence of stem bromelain, the major cysteine proteinase from pineapple stem is described. It shows that the enzyme is a member of the papain superfamily of cysteine proteinases, but is not very closely related to any other known member of this group. The sequence shows mutation or deletion of several residues that have been conserved in cysteine proteinases examined previously, including Asn-175 (papain). We suggest that some of these changes have the effect of altering the active-site geometry of stem bromelain, and that this accounts for the resistance of the enzyme to inhibition by cystatins and E-64 [L-3-carboxy-2,3-*trans*-epoxypropionylleucylamido(4-guanidino)butane].

Bromelain; Amino acid sequence; Cysteine proteinase; Enzyme inhibition; Cystatin; Compound E-64; (Pineapple stem)

1. INTRODUCTION

Stem bromelain is the major cysteine proteinase amongst a number of such enzymes present in the stem of the pineapple plant, *Ananas comosus* [1]. The enzyme is distinct both immunologically and in specificity from fruit bromelain, found in the fruit of the pineapple (Rowan, A.D. and Buttle, D.J., unpublished), and for the purposes of the present paper, we shall use the term bromelain exclusively to mean stem bromelain. Bromelain is unusual amongst the cysteine proteinases that are homologous with papain in its weak inhibition by cystatin, a protein inhibitor from chicken egg-white, and by other related inhibitors (see section 3). Bromelain is also inactivated only very slowly by E-64 [1].

Correspondence address: A. Ritonja, Department of Biochemistry, J. Stefan Institute, Jamova 39, 61000 Ljubljana, Yugoslavia

Abbreviation: E-64, L-3-carboxy-2,3-*trans*-epoxypropionylleucylamido(4-guanidino)butane

Although crude bromelain (a mixture of cysteine proteinases) is of considerable commercial importance [2], bromelain as such has not been fully characterized. It is considered to be a glycoprotein consisting of a single polypeptide, and a single carbohydrate side chain. Sequence data for somewhat over half the residues have been published ([3,4], review [5]), and a carbohydrate component has been studied [6]. The purpose of the present paper is to report the complete amino acid sequence of bromelain, and to discuss this in relation to the resistance of the enzyme to the inhibitors mentioned above.

2. EXPERIMENTAL

The chemicals used for Edman degradation and amino acid analysis were of Sequenal grade from Applied Biosystems. Iodo[^{3}H]acetic acid was from Amersham, endoproteinase Lys-C from Boehringer Mannheim and protease V8 (*S. aureus*) from Miles Scientific (UK). All other chemicals were of analytical grade.

Bromelain was purified from commercially obtained pineapple stem extract as in [1]. The protein was reduced with 2-mercaptoethanol in 6 M guanidine hydrochloride, and carboxy-

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

methylated with iodo[^{3}H]acetic acid. CNBr cleavage was performed with a 300-fold molar excess of reagent over methionyl residues in 70% (v/v) formic acid [7].

Peptides were purified by gel chromatography and HPLC on an Applied Biosystems RP-18 column eluted with aqueous acetonitrile containing trifluoroacetic acid. Samples were sequenced with an Applied Biosystems model 470 A gas-phase sequenator [8]. Phenylthiohydantoin derivatives were identified on line with the 120 A HPLC [9].

Second-order rate constants for inactivation by iodoacetate and iodoacetamide were determined at pH 6.8, exactly as described for E-64 [1], except that the concentration of dithiothreitol in the incubation (but not in the 5 min pre-incubation at 40°C) was decreased to 0.2 mM. No allowance was made for reaction of the alkylating reagents with dithiothreitol.

3. RESULTS AND DISCUSSION

3.1. *Amino acid sequence of bromelain*

Native bromelain eluted as a single symmetrical peak from a Sephacryl S-200 column (apparent M_r about 24 000), and contained two sequences (fig.1). The minor sequence (about 25%) started with Ala (residue 1 in fig.1), and the major (about 75%) with Val (residue 2). The reduced and carboxymethylated protein was used for chemical cleavage with CNBr, which proved to be complete at methionyl bonds and partial (about 30%) at tryptophanyl bonds. The complete sequence was determined from the CNBr-generated peptides, subfragments obtained by endoproteinase Lys-C cleavage of peptide CN1, and peptides from a proteinase V8 digest of the carboxymethylated protein (see fig.1). There was about 50% cleavage of the Asp–Pro (204–205) bond during the treatment with CNBr, which was expected in view of the acid-labile character of this bond. Non-specific cleavage with protease V8 on the carboxyl side of Gly-177 was attributed to the long incubation necessitated by insolubility of the reduced and alkylated protein.

The results show that bromelain exists as a single polypeptide chain with 211 or 212 residues. The M_r or the longer form of bromelain is calculated to be 22 828 (fig.1). Residue 117 was the only position in which no normal amino acid phenylthiohydantoin derivative was detected; hydrolysis of this derivative yielded aspartic acid, suggesting that it represented a glycosylated asparaginyl residue. This is consistent with the fact that residues 117–119 are Asn-Glu-Ser, the only potential glycosylation site in the molecule. The structure of an oligosaccharide from a bromelain preparation has been reported [6]; this would have an M_r of about 1000, so that the total M_r of bromelain may be about 23 800.

The alignment of the amino acid sequence of bromelain with those of other cysteine proteinases (fig.2) shows unequivocally that the enzyme is a member of the papain superfamily, although it is not very closely related to any other known member. Percentages (in parentheses) of identical residues when the sequences are aligned as in fig.2 decrease in the order: papaya proteinase III (43), actinidin (42), human cathepsin L (38), papain (38), human cathepsin H (33) and human cathepsin B (22). Bromelain resembles human cathepsin H [11] in existing partially as a form having two rather than one residue prior to the conserved Pro-2 residue at the N-terminus.

Bromelain is unusual amongst plant cysteine proteinases in its weak inhibition by the protein inhibitor cystatin from chicken egg-white, and by other inhibitors of the cystatin superfamily including L-kininogen [16,17]. Rowan et al. [1] showed that the K_i of pure bromelain with chicken cystatin is not less than 3.6×10^{-5} M, at least seven orders of magnitude weaker than that of papain ($< 5 \times 10^{-12}$ M: [18]). Bromelain is also distinguished from most other cysteine proteinases by its very slow rate of inactivation by the active-site directed inhibitor, E-64, the second-order rate constant of 512 $M^{-1} \cdot s^{-1}$ [1] contrasting with that of 638 000 for papain [19].

We have considered whether the amino acid sequence of bromelain gives any indication of the reasons for its resistance to these inhibitors. The nature of the interaction between cysteine proteinases and cystatins is not known in detail, because it has not yet been possible to obtain crystals of enzyme-inhibitor complexes suitable for X-ray crystallography, but a very plausible proposal has been put forward by Bode et al. [20] on the basis of computer 'docking' of a shortened form of chicken cystatin in the active-site cleft of papain. In this model, one end of the cystatin molecule forms a wedge that fits accurately into the active-site cleft of papain. The wedge is constructed of three segments of the cystatin polypeptide chain: a segment close to the N-terminus involving the conserved Gly-9, a segment containing the conserved Gln-Xaa-Val-Gly sequence (residues 53–57), and a segment from the C-terminus in-

```
1       10        20        30        40        50        60        70
AVPQSIDWRDYGAVTSVKNQNPCGACWAFAAIATVESIYKIKKGILEPLSEQQVLDCAKGYGCKGGWEFR

{Intact protein}
AVPQSIDWRDYGAVTSVKNQNPCGACWAFAAIATVESIYKIKK-->

{CN1}
AVPQSIDWRDYGAVTSVKNQNPCGACWAFAAIATVESIYKIKKGIL-->

{CNA}                                              {CNC}              {CND}
AVPQSIDX       {CNB}       AFAAIATVESIYKIKKGILEPLSEQQVLDCAKGYGCKGGX
        RDYGAVTSVKNQNPCGACX                                          EFR

{K1}                                        {K3}            {K4}  {K5}
AVPQSIDWRDYGAVTSVK         {K2}            KGILEPLSEQQVLDCAK     GGWEFR
                  NQNPCGACWAFAAIATVESIYK                    GYGCK

{E1}                                  {E3}           {E4} {E5}       {E6}
AVPQSIX            {E2}               SIYKIKKGILEPLSE     CAKGYGCKGGWE
       WRDYGAVTSVKNQNPCGACWAFAAIATVE                QQVLX            FR


        80        90        100       110     * 120       130       140
AFEFIISNKGVASGAIYPYKAAKGTCKTDGVPNSAYITGYARVPRNNESSMMYAVSKQPITVAVDANANF

{CND}                                                  {CN2}
AFEFIISNKGVASGAIYPYKAAKGTCKTDGV-->                     MYAVSKQPITVAVDANANF

{K5}                    {K7}
AFEFIISNK     {K6}      GTCK          {K8}
         GVASGAIYPYK        TDGVPNSAYITGYARVPRNXESSX

{E6}                           {E8}                                  {E10}
AFE          {E7}              GVPNSAYITGYARVPRNXE       {E9}        ANANF
   FIISNKGVASGAIYPYKAAKGTCKTX                      SSMMYAVSKQPITVAVX


        150       160       170       180       190       200       210
QYYKSGVFNGPCGTSLNHAVTAIGYGQDSIIYPKKWGAKWGEAGYIRMARDVSSSSGICGIAIDPLYPTLEE

{CN2}                                {CNE} {CNF}                    {DP}
QYYKSGVFNGPCGTSLNHAVTAIGYGQDSI-->        GEAGYIRX       {CN3}     PLYPTLEE
                                  GAKX          ARDVSSSSGICGIAIDPLYPTLEE

{E10}                          {E11}   {E12}
QYYKSGVFNGPCGTSLNHAVTAIGYGQX           AKWGE              {E13}
                              SIIYPKKWG      AGYIRMARDVSSSSGICGIAIDPLYPTLE
```

Ala	A	25	Gln	Q	7	Leu	L	6	Ser	S	17
Arg	R	6	Glu	E	9	Lys	K	15	Thr	T	9
Asn	N	10	Gly	G	22	Met	M	3	Trp	W	5
Asp	D	8	His	H	1	Phe	F	6	Tyr	Y	14
Cys	C	7	Ile	I	17	Pro	P	11	Val	V	14

Total: 212 residues; M_r: 22,828.

Fig.1. Amino acid sequence and composition of bromelain, and strategy of the sequence determination. The amino acid sequence of *S*-[^{3}H]carboxymethylated bromelain was determined by automated Edman degradation of the whole molecule, the eight CNBr peptides (CN1–CN3 and CNA–CNE), the eight endoproteinase Lys-C-generated peptides (K1–K8) and 13 protease V8-generated peptides (E1–E13). The residues not directly identified in the individual peptides are marked (×). Cysteine residues were identified by the radioactivity of the *S*-[^{3}H]carboxymethylated derivative. Asn-117 is the point of attachment of carbohydrate (indicated by *).

cluding Pro-103 and Trp-104. There was already biochemical evidence that the first two of these segments are involved in the binding to papain [21,22], and a further study has implicated Trp-104 [23].

When the amino acid sequence of bromelain is

Fig.2. Alignment of the bromelain (sbrom) sequence with those of papain (pap) [10], human cathepsin H (catHh) [11], human cathepsin L (catLh) [11], human cathepsin B (catBh) [12], papaya proteinase III (ppIII) [13], actinidin (actdn) [14], and human calpain II (calph) [15]. Residue numbers refer to the numbering scheme of papain [10]. For several of the sequences, alternative residues have been reported at a few positions, as mentioned in the references cited. An insertion of 17 residues that occurs in cathepsin B between residues 114 and 115 is deleted from the figure to save space. Segments of the calpain sequence that are not related to the other sequences are simply replaced by dots, for clarity.

compared with those of papain, human cathepsin H and human cathepsin L (all tightly bound by cystatin [17]) as in fig.2, a number of differences that must affect the geometry of the active-site cleft are apparent (fig.3). Perhaps the most striking is the deletion of Asn-175 and two adjacent residues, and the mutation of the highly conserved Ser-176 to Lys. Asn-175 has been found in all other papain homologues sequenced previously, with the exception of the calpains, which incidentally are unaffected by most cystatins [17]. In papain, Asn-175 is hydrogen-bonded to His-159, and orients it for its

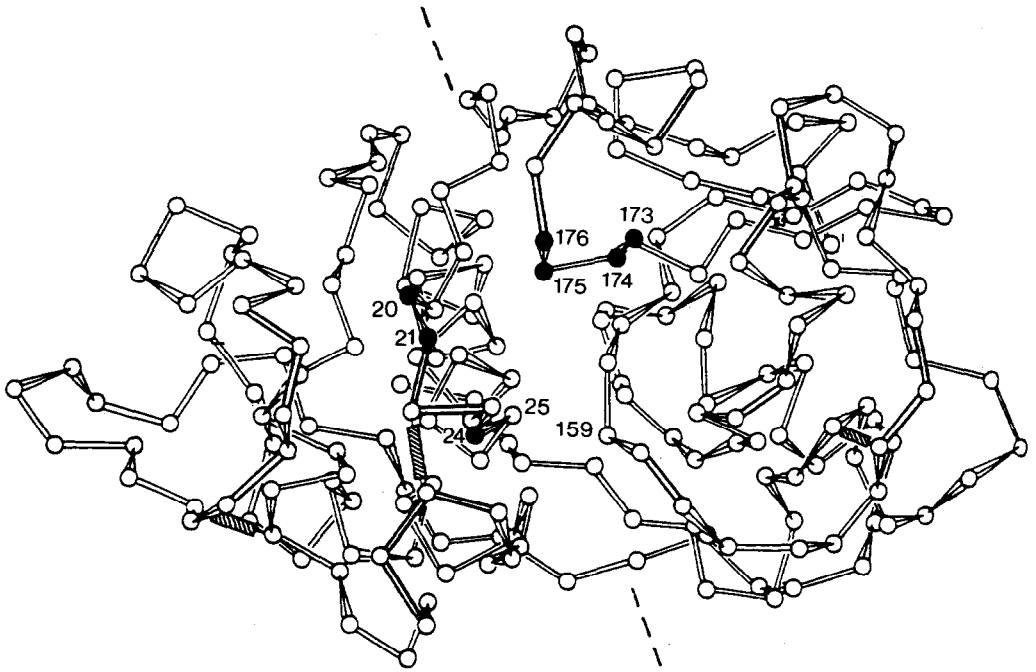

Fig.3. Perspective drawing of the structure of papain (α-carbon atoms only, adapted from [24]), emphasising residues that we suggest may contribute to the conformation of the catalytic site, and are mutated or deleted in bromelain (closed circles). The broken lines mark the approximate line of the active-site cleft. Residues 25 and 159 are the cysteine and histidine residues that form the catalytic thiolate-imidazolium pair.

catalytic function [5]. The substitution of Ser-21 by Pro, and of the highly conserved Ser-24 by Ala may well also affect the orientation of the catalytic Cys-25. In view of these factors, we suggest that the putative catalytic thiolate-imidazolium pair at the active site of bromelain is likely to have a different conformation from that in the cysteine proteinases that are tightly inhibited by cystatin.

There are other replacements that modify the structure of the active-site cleft and the specificity subsites, but the residues concerned are generally not highly conserved, and moreover, we have not been able to identify any correlation between the substrate specificities of the cysteine proteinases and their susceptibility to inhibition by cystatin or E-64.

3.2. *Reactivity of the catalytic site of bromelain*

In view of the indications for an atypical disposition of the catalytic groups in bromelain, we examined the reactivity of the active site directly, using iodoacetamide and iodoacetate. Negatively charged reactants are sensitive probes of the reactivity of the active sites of cysteine proteinases, and the thiolate-imidazolium ion pair of papain reacts at an enhanced rate with iodoacetate because of a favourable interaction between the positive imidazolium ion of the ion pair and the negative carboxylate of the alkylating reagent [25]. We determined the second-order rate constants for inactivation of papain by iodoacetate and iodoacetamide to be 1422 and 84 $M^{-1} \cdot s^{-1}$, respectively (a ratio of 17.0); these values are in good agreement with those in the literature [25]. The corresponding values for bromelain were 17.7 and 5.5 $M^{-1} \cdot s^{-1}$ (ratio 3.2), showing a much lower level of reactivity, and a smaller discrimination in favour of the negatively charged reagent.

In conclusion, it is clear that the geometry and reactivity of the catalytic site of bromelain are different from those of cysteine proteinases that bind cystatins tightly and react rapidly with E-64. Since the earliest studies of chicken cystatin, it has been known that its binding to cysteine proteinases does not require an active catalytic site (review [17]), but it is known that some large substituents greatly

weaken the interaction, showing that this is a structurally sensitive area for binding of this and related inhibitors. It has been pointed out that Cys-25 resembles a high saddle-point that subdivides the active-site cleft into two adjacent depressions [20]; in this case, one can easily imagine that a changed structure at this point might seriously affect the fit of the cystatin 'wedge' into the active-site cleft.

E-64 inactivates papain by first forming a reversible, Michaelis-type complex, and then reacting covalently and irreversibly with Cys-25. We have no direct evidence as to which of these two steps is impaired in bromelain, to account for its slow interaction with E-64, but the simplest possibility is that the specific conformation of the catalytic site is responsible for the weak inhibition by E-64 as well as the loose binding of cystatins.

Acknowledgement: We thank the Research Council of Slovenia (grant no. C1-0515) for financial support of part of this work.

REFERENCES

[1] Rowan, A.D., Buttle, D.J. and Barrett, A.J. (1988) Arch. Biochem. Biophys. 267, 262–270.
[2] Caygill, J.C. (1979) Enzyme Microb. Technol. 1, 233–242.
[3] Goto, K., Murachi, T. and Takahashi, N. (1976) FEBS Lett. 62, 93–95.
[4] Goto, K., Takahashi, N. and Murachi, T. (1980) Int. J. Peptide Protein Res. 15, 335–341.
[5] Kamphuis, I.G., Drenth, J. and Baker, E.N. (1985) J. Mol. Biol. 182, 317–329.
[6] Ishihara, H., Takahashi, N., Oguri, S. and Tejima, S. (1979) J. Biol. Chem. 254, 10715–10719.
[7] Gross, E. (1967) Methods Enzymol. 11, 238–255.
[8] Hewick, R.M., Hunkapiller, M.W., Hood, L.E. and Dreyer, W.J. (1981) J. Biol. Chem. 256, 7990–7997.
[9] Hunkapiller, M.W. and Hood, L.E. (1983) Methods Enzymol. 91, 486–493.
[10] Cohen, L.W., Coghlan, V.M. and Dihel, L.C. (1986) Gene 48, 219–227.
[11] Ritonja, A., Popovic, T., Kotnik, M., Machleidt, W. and Turk, V. (1988) FEBS Lett. 228, 341–345.
[12] Ritonja, A., Popovic, T., Turk, V., Wiedenmann, K. and Machleidt, W. (1985) FEBS Lett. 181, 169–172.
[13] Dubois, T., Kleinschmidt, T., Schnek, A.G., Looze, Y. and Braunitzer, G. (1988) Biol. Chem. Hoppe-Seyler 369, 741–754.
[14] Carne, A. and Moore, C.H. (1978) Biochem. J. 173, 73–83.
[15] Aoki, K., Imajoh, S., Ohno, S., Emori, Y., Koike, M., Kosari, G. and Suzuki, K. (1986) FEBS Lett. 205, 313–317.
[16] Gounaris, A.D., Brown, M.A. and Barrett, A.J. (1984) Biochem. J. 221, 445–452.
[17] Barrett, A.J., Rawlings, N.D., Davies, M.E., Machleidt, W., Salvesen, G. and Turk, V. (1986) in: Proteinase Inhibitors (Barrett, A.J. and Salvesen, G. eds) pp. 515–569, North-Holland, Amsterdam.
[18] Nicklin, M.J.H. and Barrett, A.J. (1984) Biochem. J. 223, 245–253.
[19] Barrett, A.J., Kembhavi, A.A., Brown, M.A., Kirschke, H., Knight, C.G., Tamai, M. and Hanada, K. (1982) Biochem. J. 201, 189–198.
[20] Bode, W., Engh, R., Musil, D., Thiele, U., Huber, R., Karshikov, A., Brzin, J., Kos, J. and Turk, V. (1988) EMBO J. 7. 2593–2599.
[21] Abrahamson, M., Ritonja, A., Brown, M.A., Grubb, A., Machleidt, W. and Barrett, A.J. (1987) J. Biol. Chem. 262, 9688–9694.
[22] Barrett, A.J. (1987) Trends Biochem. Sci. 12, 193–196.
[23] Lindahl, P., Alriksson, E., Jörnvall, H. and Björk, I. (1988) Biochemistry 27, 5074–5082.
[24] Drenth, J., Jansonius, J.N., Koekoek, R., Swen, H.M. and Wolthers, B.G. (1968) Nature 218, 929–932.
[25] Polgar, L. and Csoma, C. (1987) J. Biol. Chem. 262, 14448–14453.