Complete amino acid sequence of ananain and a comparison with stem bromelain and other plant cysteine proteases

Karen L. LEE[1], Karen L. ALBEE, Richard J. BERNASCONI and Tim EDMUNDS
Genzyme Corporation, 1 Mountain Road, Framingham, MA 01701-9322, U.S.A.

The amino acid sequences of ananain (EC 3.4.22.31) and stem bromelain (3.4.22.32), two cysteine proteases from pineapple stem, are similar yet ananain and stem bromelain possess distinct specificities towards synthetic peptide substrates and different reactivities towards the cysteine protease inhibitors E-64 and chicken egg white cystatin. We present here the complete amino acid sequence of ananain and compare it with the reported sequences of pineapple stem bromelain, papain and chymopapain from papaya and actinidin from kiwifruit. Ananain is comprised of 216 residues with a theoretical mass of 23464 Da. This primary structure includes a sequence insert between residues 170 and 174 not present in stem bromelain or papain and a hydrophobic series of amino acids adjacent to His-157. It is possible that these sequence differences contribute to the different substrate and inhibitor specificities exhibited by ananain and stem bromelain.

INTRODUCTION

Early work by Rowan et al. [1,2] identified ananain (EC 3.4.22.31) as a minor cysteine protease of pineapple (*Ananas comosus*) stem possessing distinct substrate and inhibitor binding properties compared with stem bromelain (EC 3.4.22.32), the major cysteine protease of pineapple stem. Specifically, it was determined that ananain, like other members of the papain family, is strongly inhibited by chicken egg-white cystatin and the active-site-directed inhibitor *trans*-epoxysuccinyl-L-leucylamido-(4-guanidino)butane (E-64), whereas stem bromelain is only weakly inhibited. Ananain preferentially hydrolyses the synthetic peptide substrate Bz-Phe-Val-Arg-*p*-nitroanilide (pNA) but not Bz-Arg-Arg-pNA, unlike stem bromelain which preferentially hydrolyses Bz-Arg-Arg-pNA and not Bz-Phe-Val-Arg-pNA [1–3]. Other reports indicated that ananain, unlike stem bromelain, does not contain N-linked glycosylation, but that a high degree of sequence similarity, particularly near the N-terminus, exists between the two proteases [4–7].

It has been suggested that the differences in inhibitor binding and substrate specificity exhibited by ananain and stem bromelain may be due to differences in protein geometry [8]. Kamphius et al. [9] compared the stem bromelain amino acid sequence with the crystal structures of actinidin from kiwifruit (*Actinidia chinensis*) and papain from papaya (*Carica papaya*) and determined that papain, actinidin and stem bromelain, all members of the papain family of cysteine proteases, share the same protein-folding pattern and catalytic mechanism. Early work characterizing the substrate- and inhibitor-binding site in the papain crystal structure identified a series of single residues important in substrate binding and specificity [10,11]. Although the cysteine proteases in the papain family share a common conformation, substrate specificity probably depends on specific amino acid residues in the substrate-binding site. In order better to understand these differences in specificity, we have determined the primary sequence of ananain. For this study, ananain was purified using the method of Napper et al. [12], and the complete amino acid sequence of ananain was determined and compared with the published amino acid sequences of papain [13] and chymopapain [14] from papaya, actinidin [15] from kiwifruit and stem bromelain [8] from pineapple.

EXPERIMENTAL

Materials

Stem bromelain, trypsin and endoproteinase Lys-C (Lys-C) were obtained from Boehringer-Mannheim. Ananain was purified by the method of Napper et al. [12]. Trifluoroacetic acid (TFA) and BNPS-skatole [2-(2′-nitrophenylsulphonyl)-3-methyl-3-bromoindolinene] were obtained from Pierce. HPLC solvents and water for HPLC analysis were from Burdick and Jackson. HPLC columns were from PE/Applied Biosystems. All other chemicals were obtained from Sigma Chemical Co.

CNBr cleavage of inactivated ananain and stem bromelain

Alkylated protein (1 mg in 70 % formic acid) was digested with CNBr using a 100-fold molar excess of CNBr to methionine residues. CNBr peptides were reduced with dithiothreitol and alkylated with iodoacetamide. Reduced and alkylated CNBr peptides were separated by reverse-phase HPLC on a Brownlee Aquapore RP300 column (2.1 mm × 100 mm) using a 30 min linear gradient of 0.1 % TFA to 100 % acetonitrile/0.08 % TFA at 0.25 ml/min.

Tryptic digestion of ananain CNBr peptides

Ananain CNBr peptides (200 μg) were digested with trypsin using a 1:50 enzyme/protein ratio (w/w). The digest was quenched with 0.1 % TFA, and peptides were separated by reverse-phase HPLC using a Brownlee Aquapore RP300 column (1 mm × 220 mm) and a 60 min linear gradient of 0.1 % TFA to 60 % acetonitrile/0.08 % TFA at 0.1 ml/min.

Lys-C digestion of CNBr peptides

Two 100 μg aliquots of ananain CNBr peptides were digested with Lys-C using a 1:100 enzyme/protein ratio (w/w). One

Abbreviations used: pNA, *p*-nitroanilide; E-64, *trans*-epoxysuccinyl-L-leucylamido-(4-guanidino)butane; LC/MS, liquid chromatography/mass spectroscopy; Lys-C, endoproteinase Lys-C; TFA, trifluoroacetic acid; BNPS-skatole, 2-(2′-nitrophenylsulphonyl)-3-methyl-3-bromoindolinene.

[1] To whom correspondence should be addressed.

Figure 1 Proposed amino acid sequence of ananain

Tryp, trypsin; BNPS, BNPS-skatole.

digest was quenched with 0.1 % TFA, the other reduced with dithiothreitol then alkylated with iodoacetamide. Lys-C peptides were separated on a Brownlee Aquapore RP300 column (2.1 mm × 220 mm) and a 70 min linear gradient of 0.1 % TFA to 60 % acetonitrile/0.08 % TFA at 0.1 ml/min. Fractions from the CNBr, tryptic and Lys-C digestions were collected and sequenced.

BNPS-skatole digestion of inactivated ananain

Ananain was cleaved with BNPS-skatole using the method of Crimmins et al. [16], then loaded on to a Hewlett–Packard sequencing cartridge and reduced with dithiothreitol. Reduced peptides were eluted with 15 %, 30 % and 60 % acetonitrile. Eluates were loaded on to individual cartridges and sequenced.

N-Terminal sequence analysis

N-Terminal sequence analysis of alkylated ananain and individual peptides was performed using an Applied Biosystems model 477 or Hewlett–Packard model G1005A protein sequencer equipped with on-line phenylthiohydantoin analysis.

Liquid chromatography/mass spectroscopy (LC/MS)

Alkylated ananain and Lys-C peptides of ananain were analysed by LC/MS using a Finnigan TSQ700 triple quadrupole mass spectrometer with a Finnigan electrospray interface. Chromatography was performed on a Brownlee Aquapore RP300 column (1 mm × 220 mm) using a 64 min linear gradient of 0.1 % TFA to 54 % acetonitrile/0.08 % TFA at 0.05 ml/min.

Sequence comparison

The protein sequence of ananain was compared with the published protein sequences of stem bromelain [8], chymopapain [14], papain [13] and actinidin [15] using the Genetics Computer Group Sequence Analysis Software Package (Version 7.3-AXP). Hydropathy plots for ananain, stem bromelain, chymopapain, papain and actinidin were generated using the Kyte and Doolittle analysis of the MacVector Sequence Analysis Software (Version 4.5; Kodak).

RESULTS AND DISCUSSION

The stem bromelain amino acid sequence [8] and ananain amino acid composition [12] were used as references in the development of a cleavage strategy for ananain. Ananain was inactivated before enzymic and chemical cleavage by alkylation of the active-site cysteine with iodoacetamide. When alkylated ananain was incubated with protein substrates, significant digestion of the protein substrate was observed (results not shown), indicating that alkylated ananain was only partially inactivated. Cleavage of alkylated ananain with CNBr was chosen as a means of producing large peptides for sequence analysis and as a source of inactivated ananain for digestion by trypsin and Lys-C.

The proposed ananain amino acid sequence was determined using the combined data from the tryptic, Lys-C, CNBr and BNPS-skatole digests. Overlapping sequence information was obtained for most of the ananain peptide chain, except for residues 89–92 and 115–121 (Figure 1). Methionine residues were found at positions 120, 121 and 207. The position of these residues is similar to, but not identical with, the location of

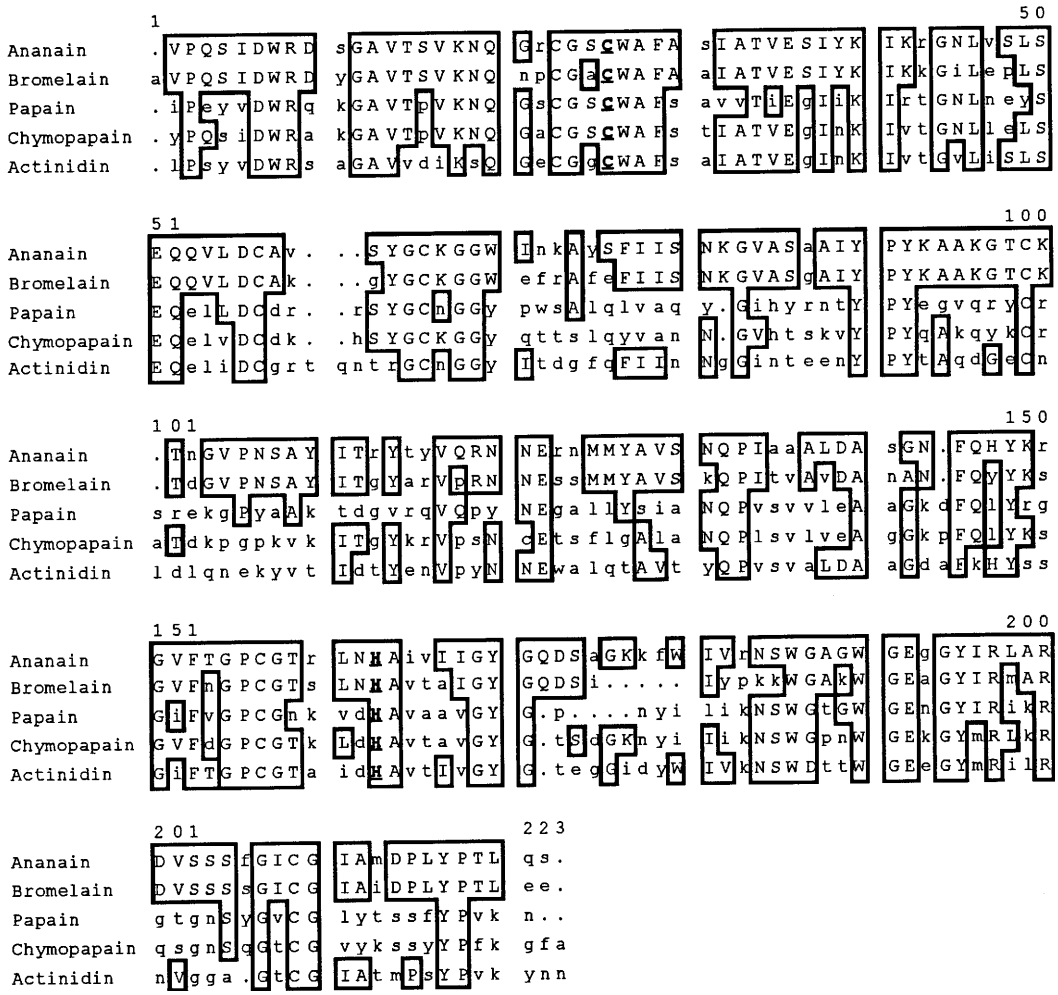

Figure 2 Amino acid sequence alignment of ananain with stem bromelain, papain, chymopapain and actinidin

Active-site residues Cys-25 and His-157 are in bold print and underlined.

methionine residues in stem bromelain (120, 121 and 192; ananain numbering scheme). Direct sequence data through Met-120 and Met-121 could not be obtained using the methods described here; however, LC/MS analysis of the Lys-C peptides of ananain confirm the presence of adjacent methionine residues (Table 1). Peptides derived from cleavage of intact alkylated ananain with the tryptophan-specific reagent BNPS-skatole [16] were used to obtain sequence information through Met-207 and for residues 89–92. Cys-25, identified as carboxamidomethylcysteine in CNBr peptide CB1 and Lys-C peptide K2, was confirmed as the free cysteine of ananain, identical with the position of the free active-site cysteine in stem bromelain, papain, chymopapain and actinidin (Figure 2). Similarly, the active-site His-157 residue [1] is positioned identically in ananain, stem bromelain, papain, chymopapain and actinidin (Figure 2).

To confirm the proposed ananain sequence shown in Figure 1, two LC/MS experiments were performed. In one experiment, masses of reduced Lys-C peptides of ananain were obtained and compared with theoretical peptide masses based on the proposed sequence. The data in Table 1 indicate that the measured masses correlate well with the theoretical values, with errors ranging from 0.003 % to 0.13 %. Peptide K13 was not found by LC/MS in the reduced Lys-C digests, but was found by LC/MS in non-reduced Lys-C peptide maps as an incomplete cleavage product (results not shown). In the second experiment, electrospray MS analysis of intact alkylated ananain gave an average mass of 23478 Da, which is within 0.07 % of the theoretical mass of 23464 Da. Results from both of the LC/MS experiments confirm that the sequence presented here is accurate.

Hydropathy analysis, presented in Figure 3, indicated that the ananain sequence between residues 159 and 161, Ile-Val-Ile, was more hydrophobic than the same residues in stem bromelain and chymopapain [8,14] and similar in hydrophobicity to the same region in papain [13] and actinidin [15]. It is possible that sequence differences that alter the overall hydrophobicity of those residues adjacent to His-157 could affect the specificity and inhibitor-binding properties of ananain, particularly compared with stem bromelain. When peptides CB3 and K15 were sequenced, it was determined that ananain possesses a sequence insert from position 170 to 174 (ananain numbering scheme) not present in stem bromelain [8] or papain [13], but similar to inserts found in chymopapain [14] and actinidin [15]. Whether the presence of the sequence insert affects substrate specificity or inhibitor binding is not known. When Kamphius et al. [9] compared the crystal structure of actinidin with that of papain it was noted that insertions in the amino acid sequence did not

Table 1 LC/MS analysis of Lys-C peptides of CNBr-cleaved ananain

Theoretical mass values are based on the proposed ananain amino acid sequence shown in Figure 1. Peptides labelled with an asterisk (*) are short peptides which were not recovered using the LC/MS methods described in the Experimental section. HS (homoserine) and HSL (homoserine lactone) are products of CNBr cleavage at methionine. Peptides labelled ND were not determined.

Lys-C Peptide	Residues	Theoretical mass (Da)	Observed mass (Da)	Deviation (%)
K1	1 to 17	1844.95	1844.90	−0.003
K2	18 to 39	2505.17	2505.90	+0.003
K3	40 to 41	260.20	*	*
K4	42 to 63	2483.20	2483.90	+0.03
K5	64 to 69	674.36	673.50	+0.13
K6	70 to 78	1042.56	1042.00	−0.05
K7	79 to 89	1139.61	1139.20	−0.03
K8	90 to 92	289.19	*	*
K9	93 to 96	465.21	*	*
K10	97 to 120 (HS)	2831.38	2832.20	+0.006
K10	97 to 120 (HSL)	2813.37	2814.00	+0.02
K11	121	150.06	*	*
K12	122 to 143	2365.16	2365.50	+0.014
K11-12	121 to 143	2496.2	2496.3	+0.004
K13	144 to 171	2946.36	ND	ND
K14	172	147.11	*	*
K15	173 to 207 (HS)	3845.90	3847.70	+0.05
K15	173 to 207 (HSL)	3827.89	3829.20	+0.034
K16	208 to 216	1033.52	1033.10	−0.04

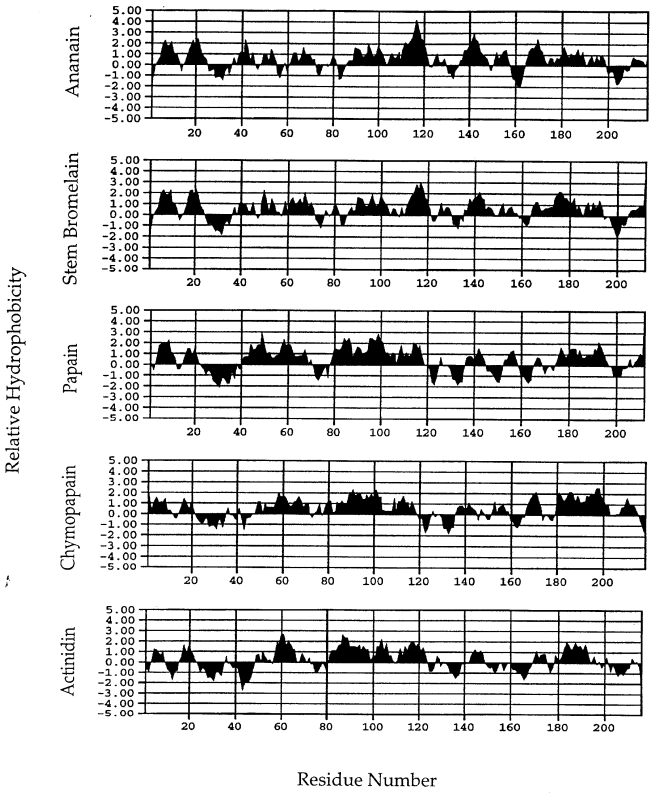

Figure 3 Hydropathy plots of ananain, stem bromelain, chymopapain, papain and actinidin

Hydropathy plots were performed on the aligned sequences shown in Figure 2.

Received 17 January 1997/1 May 1997; accepted 4 June 1997

affect overall protein conformation, but were more likely to affect overall protein solubility and charge. Data from the papain crystal structure have been used to identify individual residues important in substrate binding and specificity. These residues include but are not limited to Cys-25, Gly-66, Trp-69, Asp-158, His-159 and Trp-177 [10,11]. The equivalent residues in ananain are Cys-25, Gly-65, Asn-68, Asn-156, His-157 and Trp-180 and for stem bromelain they are Cys-25, Gly-66, Phe-69, Asn-157, His-158 and Trp-181. Although these residues are similar among papain, ananain and stem bromelain, one notable difference is observed. Trp-69, a component of the S_2 subsite in papain [10], is replaced by an asparagine residue in ananain and a phenylalanine in stem bromelain. The difference in hydrophobicity between asparagine and phenylalanine could contribute to the unique substrate specificities exhibited by ananain and stem bromelain towards synthetic peptide substrates.

A comparison of the proposed ananain sequence with the published sequences of stem bromelain [8], chymopapain [14], papain [13] and actinidin [15] is presented in Figure 2 and shows a high degree of identity between all five plant cysteine proteases. Ananain possesses 77% identity in amino acid sequence with stem bromelain, 52% identity with chymopapain, 50% identity with actinidin and 44% identity with papain. The identity shared between ananain and stem bromelain (77%) is greater than the identity shared between chymopapain and papain from papaya (60%).

The ananain amino acid sequence presented here possesses some unique features, which include a hydrophobic series of amino acids near His-157 and a sequence insert between residues 170 and 174 not found in stem bromelain. It is possible that these amino acid sequence differences may contribute to the unique inhibitor and substrate-binding properties of ananain compared with stem bromelain.

We acknowledge Dr. Parthasarathy Manavalan for assisting with the amino acid sequence homology analysis and for review of the manuscript. We also thank Dr. Marianne Borowski, Marcella Holdridge and Marina Leonard for purifying the ananain used in these studies, and Dr. John McPherson, Dr. Edward Cole and Dr. Robert Mattaliano for critical review of the manuscript.

REFERENCES

1 Rowan, A. D., Buttle, D. J. and Barrett, A. J. (1988) Arch. Biochem. Biophys. **267**, 262–270
2 Rowan, A. D., Buttle, D. J. and Barrett, A. J. (1990) Biochem. J. **266**, 869–875
3 Rowan, A. D., Brzin, J., Buttle, D. J. and Barrett, A. J. (1990) FEBS Lett. **269**, 328–330
4 Lee, K., Bernasconi, R., O'Day, C., Borowski, M. and Edmunds, T. (1992) Sixth Symposium of the Protein Society, San Diego, CA, p. 78
5 Albee, K., Lee, K., Bernasconi, R. and Edmunds, T. (1995) Protein Sci. **4**, 143
6 Murachi, T., Yasui, M. and Yasuda, Y. (1964) Biochemistry **3**, 48–55
7 Ota, S., Moore, S. and Stein, W. H. (1964) Biochemistry **3**, 180–185
8 Ritonja, A., Rowan, A. D., Buttle, D. J., Rawlings, N. D., Turk, V. and Barrett, A. J. (1989) FEBS Lett. **247**, 419–424
9 Kamphius, I. G., Drenth, J. and Baker, E. N. (1985) J. Mol. Biol. **182**, 317–329
10 Lowe, G. (1976) Tetrahedron **32**, 291–302
11 Drenth, J., Kalk, K. H. and Swen, H. M. (1976) Biochemistry **15**, 3730–3738
12 Napper, A. D., Bennett, S. P., Borowski, M., Holdridge, M. B., Leonard, M. J., Rogers, E. E., Duan, Y., Laursen, R. A., Reinhold, B. and Shames, S. L. (1994) Biochem. J. **301**, 727–735
13 Husain, S. S. and Lowe, G. (1969) Biochem. J. **114**, 279–288
14 Watson, D. C., Yaguchi, M. and Lynn, K. R. (1990) Biochem. J. **266**, 75–81
15 Carne, A. and Moore, C. H. (1978) Biochem. J. **173**, 73–83
16 Crimmins, D. L., McCourt, D. W., Thoma, R. S., Scott, M. G., Macke, K. and Schwartz, B. D. (1990) Anal. Biochem. **187**, 27–38