

Amino-acid sequence and glycan structures of cysteine proteases with proline specificity from ginger rhizome *Zingiber officinale*

Kyung H. Choi and Richard A. Laursen

Department of Chemistry, Boston University, MA, USA

The ginger proteases (GP-I and GP-II), isolated from the ginger rhizome *Zingiber officinale*, have an unusual substrate specificity preference for cleaving peptides with a proline residue at the P₂ position. The complete amino-acid sequence of GP-II, a glycoprotein containing 221 amino acids, and about 98% that of GP-I have been determined. Both proteases, which are 82% similar, have cysteine residues at positions 27 and histidines at position 161, corresponding to the essential cysteine–histidine diads found in the papain family of cysteine proteases, and six corresponding cysteine residues that form the three invariant disulfide linkages seen in this family of proteins. The sequence homology with other members (papain, bromelain, actinidin, protease omega, etc.) of this family is ≈ 50%. GP-II has two predicted glycosylation sites at Asn99 and Asn156. Analysis by electrospray and collision-induced dissociation MS showed that both sites were occupied by the glycans (Man)₃(Xyl)₁(Fuc)₁(GlcNAc)₂ and (Man)₃(Xyl)₁(Fuc)₁(GlcNAc)₃, in a ratio of ≈ 7 : 1. Both glycans are xylose containing biantennary complex types that share the common core structural unit, Man1→6(Man1→3)(Xyl1→2)Man1→4GlcNAc1→4(Fuc1→3)GlcNAc for the major form, with an additional N-acetylglucosamine residue being linked, in the minor form, to one of the terminal mannose units of the core structure.

Keywords: amino-acid sequence; cysteine protease; ginger; glycoprotein; glycosylation site; proline peptidase; proteinase; mass spectrometry.

Some 20 families of peptidases dependent on a cysteine residue at the active center are known [1]. The most prominent of these is the papain family, which includes peptidases with a wide variety of activities, including endopeptidases with broad specificity (e.g. papain) and narrow specificity (glycyl endopeptidase), amino peptidases, and a dipeptidyl peptidase. Enzymes of the papain family are found in a wide variety of life forms: baculovirus, eubacteria, yeast, probably all protozoa, plants and animals. Papain homologs are generally either lysosomal (vacuolar) or secreted proteins. In plants, they are found in the vacuoles, but also extracellularly, as in the latex of papaya or fig. Plant cysteine proteinases include papain from papaya (*Carica papaya*), bromelain from pineapple stem and actinidin from the Chinese gooseberry or kiwi fruit (*Actinidia chinensis*) [1].

Ginger proteases from ginger rhizome, *Zingiber officinale*, were first reported by Ichikawa *et al.* [2] in 1973. Two proteases, designated GP-I and GP-II were isolated by

chromatography on DEAE-cellulose and tentatively classified as cysteine proteases based on their inhibition by thiol reagents and their relative molecular masses, estimated at ≈ 22 500 by gel filtration. More recently, these proteases were separated as mercuribenzoate derivatives into three fractions by IEF [3]. It has been reported [4] that these proteases have both proteinase and collagenase activities; ginger root extracts have also been tested as a meat tenderizer [5,6].

Recently, using a peptide specificity mapping technique developed in our laboratory, Duan [7,8] observed the apparent preference of GP-I and GP-II for cleavage at prolyl peptide bonds. Because cleavage at proline is quite unusual and has a number of interesting applications, we decided to characterize these proteases in detail. In this paper we report on the primary structure analysis of GP-I and GP-II, showing that they are members of the papain family of cysteine proteases and that GP-II is glycosylated at two sites. This work provides the basis for the X-ray crystal structure of GP-II [9]. We have also found that GP-I and GP-II prefer to cleave peptide bonds at which proline is in the P₂ position, rather than P₁, as initially thought. A detailed analysis of the substrate specificity of GP-II will be described elsewhere.

MATERIALS AND METHODS

Isolation of ginger proteases

Ginger proteases were isolated by a modification of the procedure of Ichikawa *et al.* [2]. Ginger roots (800 g, purchased from local markets) were cut into small pieces and ground in a food blender at 3 °C (cold room) in 1000 mL 20 mM phosphate buffer (pH 7.0), containing 10 mM cysteine. The resulting mixture was stirred for 1 h and filtered through cheesecloth, and the filtrate saved. The residue was re-extracted with

Correspondence to R. A. Laursen, Department of Chemistry, 590 Commonwealth Ave., Boston University, 590 Commonwealth Ave, Boston, MA 02215, USA. Fax: + 1 617 353 6466, Tel.: + 1 617 353 2491, E-mail: laursen@bu.edu

Abbreviations: GP-I and GP-II, ginger proteases I and II; Cbz-Lys-pNP, N-carbobenzoyloxycarbonyllysine *p*-nitrophenyl ester; Lys-C, endoproteinase Lys-C; Glu-C, endoproteinase Glu-C; DITC-, *p*-phenylene diisothiocyanate-modified; MALDI-TOF, matrix assisted laser desorption ionization time-of-flight; CID, collision induced dissociation.

Note: the ginger protease amino-acid sequences have been registered at the National Biomedical Research Foundation Protein Information Resource under the accession numbers A59040 for ginger protease-I and A59041 for ginger protease-II.

(Received 26 August 1999, revised 10 January 2000, accepted 17 January 2000)

500 mL of buffer and the mixture filtered. The combined filtrates (1600 mL) were adjusted to 2% (w/v) NaCl and stirred for 2 h, after which the suspension was centrifuged at 15 000 *g* at 4 °C for 30 min. The supernatant was filtered through Celite to remove suspended material and brought to 20% saturation with ammonium sulfate. The solution was stirred for 1 h and then centrifuged at 13 000 *g* for 30 min, following which the supernatant was brought to 60% saturation with ammonium sulfate, adjusted to pH 7.2 with NaOH and stirred for 3 h at 4 °C. The resulting suspension was centrifuged, the supernatant discarded, and the precipitate suspended in $\approx$ 150 mL 20 mM sodium phosphate buffer (pH 7.0) containing 5 mM sodium tetrathionate. The solution was then dialyzed (using dialysis tubing with a cutoff of 12 000 Da) against three 2-L changes of 20 mM phosphate buffer (pH 7.0), 5 mM sodium tetrathionate, 1 mM EDTA, over 16 h.

The dialysate was centrifuged at 10 000 *g* for 30 min at 4 °C to remove any insoluble material and was immediately applied to a column (2.5 $\times$ 30 cm) of DEAE cellulose which had been equilibrated with 20 mM sodium phosphate buffer (pH 7.0). The column was washed with $\approx$ 500 mL equilibration buffer to remove unretained proteins and eluted with a linear gradient of 500 mL 0.25 M NaCl (in equilibration buffer) into 500 mL equilibration buffer at a flow rate of 20 mL·h⁻¹; 8-mL fractions were collected. The column was monitored at 280 nm to detect protein, and fractions were assayed for proteolytic activity. Fractions of interest were pooled. For sequence studies, peak center fractions were purified further by HPLC on a Vydac C₁₈ column (8.0 $\times$ 250 mm, 5- μ m particle size) using a linear gradient of 0.05% trifluoroacetic acid/water and 0.05% trifluoroacetic acid/acetonitrile. Protease-containing peaks were collected and lyophilized.

Protease activity assay

Protease activity was measured by a modification of the method of Boland and Hardman [10], by following the release of *p*-nitrophenolate ion from *N*- α -carbobenzyl-L-lysine *p*-nitrophenyl ester (Cbz-Lys-pNP) at 348 nm. The assay mixture was 2.5 mL 0.1 M phosphate buffer (pH 6.0) containing 1 mM EDTA and 1 mM dithiothreitol, to which was added (with mixing) 50 μ L Cbz-Lys-pNP in dimethylformamide (10 mM stock solution). The absorbance at 348 nm was recorded for 1 min to determine the rate of spontaneous hydrolysis, and then 10 μ L enzyme solution was added to the mixture. The absorbance at 348 nm was recorded for 2 min. For each measurement, the nonenzymatic hydrolysis rate was subtracted from the apparent rate. Relative activities of various fractions were recorded for convenience in arbitrary absorbance units (ΔA_{348} , 1 cm·min⁻¹). The molar extinction coefficient of 5400 M⁻¹·cm⁻¹ was used for *p*-nitrophenol at pH 6.0. Assays were performed at room temperature (23 °C).

Sulfhydryl group analysis

Approximately 3.5 mg of protein was dissolved in 3 mL 2% SDS, 0.08 M sodium phosphate buffer (pH 8) and 0.5 mg·mL⁻¹ EDTA, and was followed by 0.1 mL of 5,5'-dithio-bis(2-nitrobenzoic acid), prepared by dissolving 40 mg of the reagent in 10 mL phosphate buffer pH 8. After 15 min, the absorbance at 412 nm was read and the sulfhydryl content calculated using the molar absorptivity value of 12 000·M⁻¹·cm⁻¹. Protein concentration was calculated from the absorbance at 280 nm using $\epsilon_{280} = 2.0$ mL·mg⁻¹.

Disulfide bond analysis

The number of cysteine residues involved in disulfide bonds was estimated by measuring total thiol content (see above) after reduction of disulfide bonds with sodium borohydride [11] and subtracting the thiol content of the native protein.

Carboxymethylation of cysteine

One milligram (38 nmol) of the lyophilized protein was dissolved in 0.5 mL 0.5 M Tris buffer (pH 8.1), 1 mM EDTA, 6 M guanidinium chloride, and the solution was deaerated with nitrogen for 10 min. Dithiothreitol (1.4 μ mol; fivefold excess over the number of cysteines) was added and the sealed solution was kept at room temperature overnight. Then 3.4 μ mol of iodoacetic acid was added and the solution was incubated in the dark for 1.5 h. The reaction was stopped by addition of 100 μ L of 2-mercaptoethanol, after which the solution was dialyzed against 3 $\times$ 1 L of water for 5 h and then lyophilized.

Protein and peptide fragmentation

Endoproteinase Lys-C. S-Carboxymethylated protein (0.4 mg) was dissolved in 150 μ L 0.1 M ammonium bicarbonate buffer (pH 8.5), and 8 μ g (2% w/w) endoproteinase Lys-C (Wako) was added. The solution was incubated at 37 °C for 4 h and subjected directly to HPLC separation.

Endoproteinase Glu-C. S-Carboxymethylated protein (0.45 mg) was dissolved in 150 μ L 0.1 M ammonium bicarbonate buffer (pH 8.5), and 9 μ g (2% w/w) endoproteinase Glu-C (Boehringer) was added. The solution was incubated at 37 °C for 4 h and fractionated by HPLC.

Cyanogen bromide cleavage. S-Carboxymethylated protein (0.3 mg) was cleaved in 1 mL cyanogen bromide (50 mg·mL⁻¹) in 70% formic acid in the dark for 16 h. The solution was then diluted to $\approx$ 10 mL and lyophilized.

Sequential, multiprotease digestion for analysis of glycopeptides. S-Carboxymethylated GP-II (0.35 mg) was dissolved in 150 μ L 0.1 M ammonium bicarbonate buffer (pH 8.5), and 2% (w/w) endoproteinase Lys-C solution was added. The solution was incubated at 37 °C for 2 h and then analyzed by matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) MS. Trypsin (2% w/w) was added to the digest and the solution was incubated at 37 °C. Aliquots of 1 μ L were removed for MS analysis every 30 min and at the appropriate stage the digest was fractionated by HPLC to isolate suspected glycopeptides. The lyophilized tryptic fractions were dissolved in 200 μ L 0.1 M ammonium bicarbonate buffer (pH 8.5) and 2 μ g endoproteinase Glu-C in 10 μ L of the same buffer was added. Digestion was monitored by MS of 1- μ L aliquots of the reaction mixture at 2, 4, 24 and 48 h, after which the digests (above) were heated to 100 °C for 15 min to denature endoproteinase Glu-C. After cooling, 2 μ g thermolysin in 0.1 M ammonium bicarbonate buffer (pH 8.5) was added to the sample, which was kept at 37 °C and analyzed periodically by MALDI-TOF MS.

Purification of peptides

All peptides were purified by reversed-phase HPLC on a Vydac C₁₈ column (4.6 mm $\times$ 250 mm, 5- μ m particle size) using linear gradients of 0.05% trifluoroacetic acid/water and 0.05%

trifluoroacetic acid/acetonitrile on a Millipore/Waters HPLC system with detection at 214 nm.

Protein and peptide sequencing

Protein and peptide samples were sequenced by solid-phase Edman degradation on a Millipore model 6625 ProSequencer. Peptides were immobilized, according to standard Millipore protocols, by covalent attachment of carboxyl groups to Millipore Sequelon-AA arylamine membrane disks or of amino groups to Sequelon-DITC (*p*-phenylene diisothiocyanate membrane).

Deglycosylation of glycoprotein and peptides

Endoglycosidase F/N-glycosidase F. Endoglycosidase F/N-glycosidase F was purchased from Boehringer-Mannheim. Buffers in the range pH 7–9 were used to increase N-glycosidase F activity and reduce endoglycosidase F activity, to obtain more homogeneous carbohydrate fragmentation patterns. S-Carboxymethylated glycoprotein (10–20 nmol) was dissolved in $\approx 50 \mu\text{L}$ 0.2% SDS and heated for 10 min at 90 °C. After the solution had cooled to room temperature, 100 μL 90 mM sodium phosphate buffer (pH 7.5) containing 10 mM EDTA, 2% Triton X-100, 0.2% SDS and 1% 2-mercaptoethanol was added, followed by 1 U N-glycosidase F in 20 μL 20 mM potassium phosphate (pH 7.2), 5 mM EDTA and 0.05% sodium azide. The solution was incubated at 37 °C for 12–72 h.

N-glycosidase A digestion. S-Carboxymethylated glycoprotein (10–20 nmol), or a proteolytic digest thereof, was dissolved in 10 μL water, and 0.5 U N-glycosidase A (Boehringer-Mannheim) in 50 mM citrate/phosphate buffer (pH 5.0) containing 50% (v/v) glycerol, was added. After mixing, the solution was kept at 37 °C for 24–72 h. For tryptic peptide mixtures, the pH of the digestion solution was adjusted to 5.0 with acetic acid before addition of N-glycosidase A.

Mass determination of peptides and sugars

MALDI-TOF mass spectra were taken on a Voyager-DE Biospectrometry Workstation (PerSeptive Biosystems, Framingham, MA, USA). Approximately 0.5–1 μL of peptide mixture was spotted on a Voyager-DE sample plate, after which an equal volume of saturated sinapinic acid in 0.1% trifluoroacetic acid acetonitrile (1:1) and water was added. For the smaller peptides ($M_r < 2000$) α -cyano-4-hydroxycinnamic acid was used as matrix. Peptide mixtures containing large and small peptides were analyzed in separate experiments using both matrixes. To enhance the signal-to-noise ratio, 64–128 single shots from the nitrogen laser (337 nm) were averaged for each mass spectrum.

Determination of carbohydrate composition and structure

Carbohydrate structural analyses were provided by V. Reinhold at the Boston University Medical Center Mass Spectrometric Facility. For compositional analysis, a sample (100 μg) of GP-II was hydrolyzed in anhydrous methanolic HCl at 80 °C for 16 h, after which the amino sugars were re-acetylated. Monosaccharides were then converted to trimethylsilyl ethers and the derivatized products were analyzed by gas-liquid chromatography.

Structural analysis of the intact oligosaccharides was performed, after permethylation of N-glycosidase A digested glycopeptides, by electrospray and collision induced

dissociation (CID) MS on a Finnigan MAT TSQ-700 mass spectrometer. For CID studies, parent ions were selectively transmitted by the first mass analyzer (RF electric quadrupole field) and directed into the collision cell with a kinetic energy set at -37eV . Argon was used as the collision gas.

RESULTS

Isolation of ginger proteases

The ginger proteases were purified essentially as described by Ichikawa *et al.* [2], except that sodium tetrathionate was introduced during the purification to react with and protect the active site thiol group. Protein precipitating between 20% and 60% ammonium sulfate saturation was chromatographed on DEAE-cellulose, which resolved two fractions having protease activity (Fig. 1). These were designated GP-I and GP-II. Reversed-phase HPLC analysis of the material isolated by DEAE-cellulose chromatography showed that GP-II was essentially homogeneous, eluting as a single peak, but that the GP-I fraction contained other components. Under the specific conditions used, GP-I eluted about 0.6 min earlier than GP-II, but the two proteins could be resolved (data not shown). Both fractions were purified further by preparative HPLC before being used for sequence studies.

Number of sulfhydryls and disulfides in GP-I and GP-II

Titration of native GP-I and GP-II with 5,5'-dithio-bis(2-nitrobenzoic acid) gave respective values of 0.72 and 0.76 mole thiol group per mole protein, indicating the presence of a single

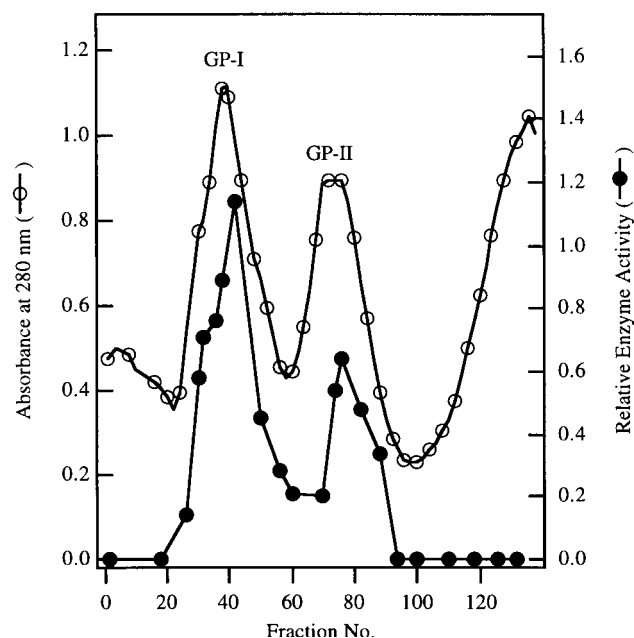


Fig. 1. Separation of protease activity into GP-I and GP-II on DEAE-cellulose. The fraction of protein precipitating at 20–60% saturated ammonium sulfate was chromatographed on a DEAE-cellulose column with a gradient of sodium chloride. Not shown is a large unretained peak ($\approx 300 \text{ mL}$) that eluted with the equilibration buffer wash prior to application of the gradient (cf. [2]). Enzyme activity was measured by monitoring the hydrolysis of CBz-Lys-pNP at 348 nm. The GP-I and GP-II fractions were pooled separately and the proteins purified further by HPLC for sequence analysis.

sulfhydryl group in each of the proteins, (this is typical of cysteine proteases which have a single active site sulfhydryl group [1]). Titration of GP-II after reduction of disulfides with sodium borohydride showed 6.6 sulfhydryls per mole protein, consistent with the premise that native GP-II contains one sulfhydryl group and three disulfide bonds, as do most cysteine proteases of the papain family [1]. Papain, used as a control, gave a value of 6.9 sulfhydryls per mole protein under the same conditions.

Amino-acid sequences of GP-I and GP-II

The amino-acid sequence of GP-II was deduced by sequencing of overlapping endoproteinase Lys-C (Lys-C), endoproteinase Glu-C (Glu-C) and cyanogen bromide peptides as illustrated in Fig. 2. In addition, sequencing of the intact protein by Edman degradation confirmed the first 35 residues of GP-II. Peptide mixtures were separated by reversed-phase HPLC. All of the expected Lys-C peptides were isolated, with the exception of the polar C-terminal tripeptide, which probably eluted in the void volume. In all, 11 Lys-C peptides were isolated and at least partially sequenced by solid-phase Edman degradation. The Lys-C fragment pairs, K7 and K8, and K10 and K11, showed the same sequences through 20 and 21 residues, respectively. However, their measured masses were about 1100–1200 Da

higher than the masses calculated from the sequences eventually deduced (see below) suggesting the presence of post-translational modifications, probably the addition of oligosaccharides.

Seventeen Glu-C peptides were isolated and sequenced by solid-phase Edman degradation after immobilization on arylamine membranes. Most of these peptides were sequenced in their entirety even though the phenylthiohydantoin signals for aspartic acid and glutamic acid were often weak due to their side chains being coupled to the sequencing membrane. The measured molecular masses of these peptides matched the calculated masses, except for peptides E11' and E12, which contained suspected glycosyl groups. Most of the Glu-C cleavages occurred as expected at glutamic acid, although there were also a number of aspartic acid cleavages also.

GP-II contains two methionine residues, so it was expected that cleavage with cyanogen bromide would yield three peptides. Initially, however, only two peptides, CNBr-1 and CNBr-2 could be isolated by HPLC. In an attempt to detect suspected CNBr-3, the intact protein was immobilized on a DITC-sequencing membrane and treated first with acetic anhydride to block the N-terminal residue (in CNBr-1), and then with CNBr to cleave at the methionine residues. Edman sequencing gave two signals and, by subtracting out the sequence of CNBr-2, a partial sequence for CNBr-3 could be

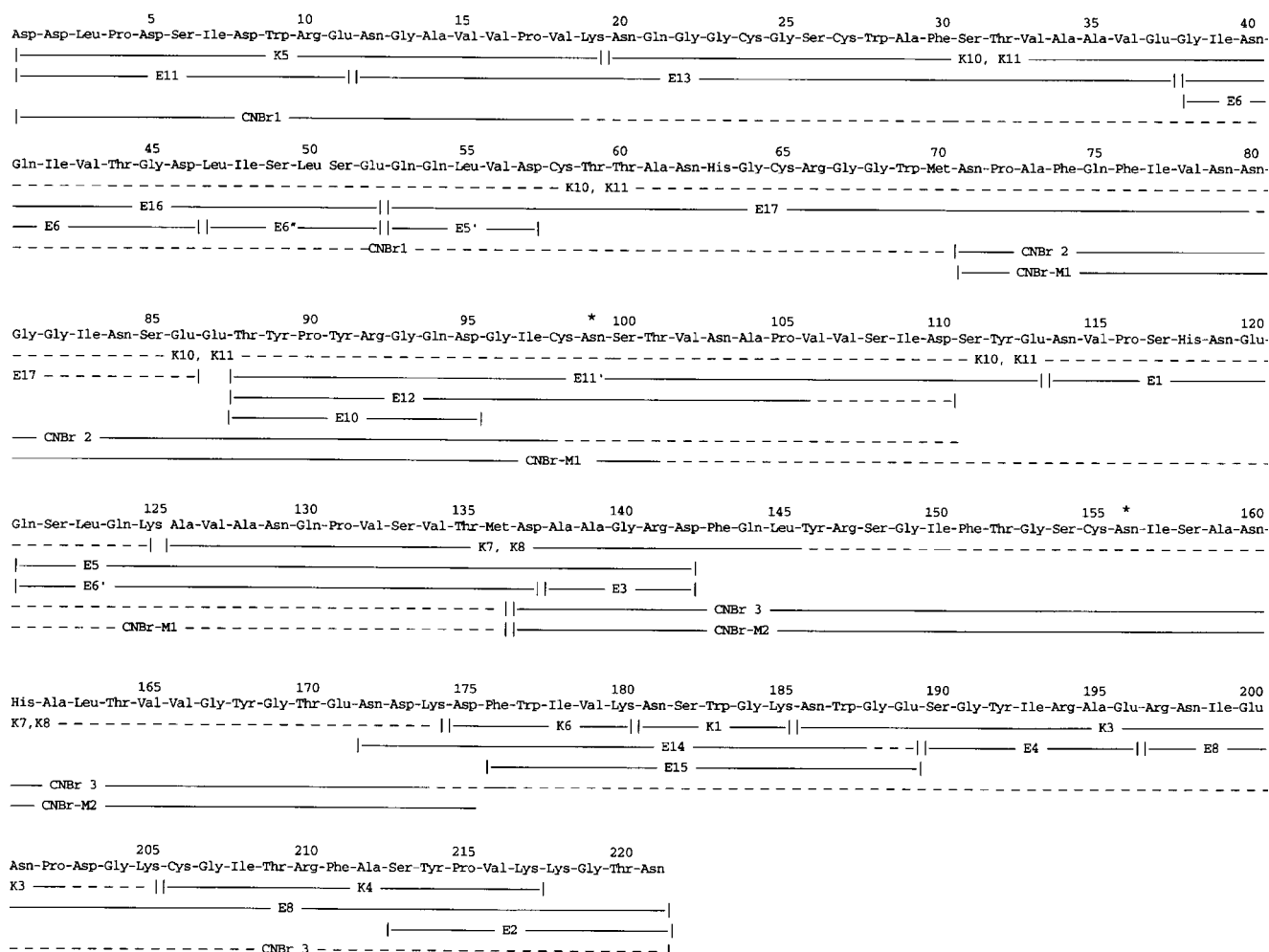


Fig. 2. Summary of data used to deduce the sequence of GP-II. Cleavage reagents: K, endoproteinase Lys-C; E, endoproteinase Glu-C; CNBr, cyanogen bromide. M refers to peptide sequences obtained by immobilizing GP-II on a DITC-membrane, then cleaving with CNBr and sequencing. Solid lines indicate regions sequenced by Edman degradation; dashed lines, nonsequenced portions. Glycosylation sites are marked by asterisks.

		10		20		30		40
GP-I	D V L P D S I D W R	E K G A V V P V K N	Q G G C G S C W A F	D A I A A V E G I N	Q I V T G D L I S L			
GP-II	D D L P D S I D W R	E N G A V V P V K N	Q G G C G S C W A F	S T V A A V E G I N	Q I V T G D L I S L			
Brom	- A V P Q S I D W R	D Y G A V T S V K N	Q N P C G A C W A F	A A I A T V E G I Y	K I K K G I L E P L			
Actd	- - L P S Y V D W R	S A G A V V D I K S	Q G E C G G C W A F	S A I A T V E G I N	K I V T G V L I S L			
Pap	- - I P E Y V D W R	Q K G A V T P V K N	Q G S C G S C W A F	S A V V T I E G I I	K I R T G N L N E Y			
PPO	- - L P E N V D W R	K K G A V T P V R H	Q G S C G S C W A F	S A V A T V E G I N	K I R T G K L V E L			
GEP	- - L P E S V D W R	A K G A V T P V K H	Q G Y C E S C W A F	S T V A T V E G I N	K I K T G N L V E L			
	50	60	70	80	90			
GP-I	S E Q Q L V D C S T	- - R N H G C E G G	W P Y R A F Q Y I I	N N G G I N S E E H	Y P Y T G T N G T C			
GP-II	S E Q Q L V D C T T	- - A N H G C R G G	W M N P A F Q F I V	N N G G I N S E E T	Y P Y R G Q D G I C			
Brom	S E Q Q V L D C A K	- - G Y G C K G G	W E F R A F E F I I	S N K G V A S G A I	Y P Y K A A K G T C			
Actd	S E Q E L I D C G R	T Q N T R G C D G G	Y I T D G F Q F I I	N N G G I N T D E N	Y P Y T A Q D G D C			
Pap	S E Q E L L D C D R	- - R S Y G C N G G	<u>Y P W</u> S A L Q L V -	A Q Y G I H Y R N T	Y P Y E G V Q R Y C			
PPO	S E Q E L V D C E R	- - R S H G C K G G	Y P P Y A L E Y V -	A K N G I H L R S K	Y P Y K A K Q G T C			
GEP	S E Q E L V D C D L	- - Q S Y G C N R G	Y Q S T S L Q Y V -	A Q N G I H L R A K	Y P Y I A K Q Q T C			
	100	110	120	130	140			
GP-I	D T K E N A H V V S	I D - S Y R N V P S	N D E K S L Q K A V	- A N Q P V S V T M	D A A G R D F Q L Y			
GP-II	N* S T V N A P V V S	I D - S Y E N V P S	H N E Q S L Q K A V	- A N Q P V S V T M	D A A G R D F Q L Y			
Brom	K T D G V P N S A Y	I T - G Y A R V P -	R N N E S S M M Y A	V S K Q P I T V A V	D A N A - N F Q Y Y			
Actd	D V A L Q N E K Y V	T I D T Y E N V P -	Y N N E W A L Q T A	V T Y Q P V S V A L	D A A G D A F K Q Y			
Pap	R S R E K G P Y A A	K T D G V R Q V Q P	Y N E G A L L Y S I	- A N Q P V S V V L	E A A G K D F Q L Y			
PPO	R A K Q V G G P I V	K T S G V G R V Q P	N N E G N L L N A I	- A K Q P V S V V V	E S K G R P F Q L Y			
GEP	R A N Q V G G P K V	K T N G V G R V Q S	N N E G S L L N A I	- A H Q P V S V V V	E S A G R D F Q N Y			
	150	160	170	180	190			
GP-I	R N G I F T G S C N	I S A N H Y R T V G	G R E T E N D K D Y	W T V K N S W G K N	W G E S G Y I R V E			
GP-II	R S G I F T G S C N*	I S A N H A L T V V	G Y G T E N D K D F	W I V K N S W G K N	W G E S G Y I R A E			
Brom	K S G V F N G P C G	T S L N H A V Y A I	G Y G Q D S I I - Y	P P K - - K W G A K	W G E S G Y I R M A			
Actd	S S G I F T G P C G	T A I D H A V T I V	G Y G T E G G I D Y	W I V K N S W D T T	W G E E G Y M R I L			
Pap	R G G I F V G P C G	N K V D H A V A A V	G Y G P N - - - Y	I L I K N S W G T G	W G E N G Y I R I K			
PPO	K G G I F E G P C G	T K V D H A V T A V	G Y G K S G G K - Y	I L I K N S W G T A	W G E K G Y I R I K			
GEP	K G G I F E G S C G	T K V D H A V T A V	G Y G K S G G K G Y	I L I K N S W G P G	W G E N G Y I R I R			
	200	210						
GP-I	R N I A E S S G K C	G I A I S P S Y P I	K E X X X					
GP-II	R N I E N P D G K C	G I T R F A S Y P V	K K G T N					
Brom	R D V S S S S G I C	G I A I D P L Y P T	L E E					
Actd	R N V G G A - G T C	G I A T M P S Y P V	K Y N N					
Pap	R G T G N S Y G V C	G L Y T S S F Y P V	K N					
PPO	R A P G N S P G V C	G L Y K S S Y Y P T	K N					
GEP	R A S G N S P G V C	G V Y R S S Y Y P I	K N					

Fig. 3. Alignment of the sequences of GP-I and GP-II with those of some related cysteine proteases of the papain family. Numbering of the sequences corresponds to the papain sequence. Known N-glycosylation sites (in GP-II and bromelain) are indicated by asterisks. Double underlined residues in papain are the active site cysteine and histidine residues found in all cysteine proteases and single underlined residues are residues whose side chains line the S_2 binding pocket in papain. Brom, Stem bromelain; actd, actinidin; pap, papain; PPO, protease omega; GEP, glycyl endopeptidase.

determined. Subsequently it was found that CNBr-3 had precipitated from acidic solution before the CNBr digest was separated by HPLC, which explains its absence. When the precipitated peptide was dissolved in N-methylmorpholine buffer, it could be attached to an arylamine sequencing membrane and sequenced through 36 residues. The measured mass of CNBr-1 exactly matched the calculated mass, but the measured masses of CNBr-2 and CNBr-3 were both higher (> 1200 Da) than the masses calculated from the sequences, again reflecting sites of glycosylation in each of these peptides.

The sequence of GP-I was determined by a similar strategy, except that CNBr peptides were not used and some alignments were made based on the strong (82%) homology between GP-I and GP-II. Thus the sequence of GP-I was not determined with the same rigor as for GP-II. Certain ambiguities remain: for example, the nature of the C-terminus, as no peptide containing the C-terminal tripeptide Gly-Thr-Asn seen in GP-II was isolated. Sequencing was complicated by the difficulty of obtaining pure samples of GP-I. The GP-I appears to have been contaminated by a third, homologous protein, the existence of which has been suggested by IEF studies [3]. Another

ambiguity is at the N terminus, where sequencing of intact GP-I (through 35 cycles) and of a tryptic peptide revealed aspartic acid at position 2, as in GP-II, but sequencing of N-terminal Lys-C and Glu-C peptides showed valine at this position. The proposed sequence of GP-I is shown in Fig. 3 aligned with GP-II and related cysteine proteases.

Identification of the glycosylation sites in GP-II

The strategy used to locate the glycosylation sites in GP-II was similar to one described by Carr and Roberts [12]. Mass analysis of some of the proteolytic digests described above showed that they contained some high molecular weight peptides (presumably glycopeptides) that did not fit the determined sequence (Figs 2 and 3) and that they lacked certain expected peptides, specifically those that contained consensus glycosylation sites. To confirm the presence of carbohydrate moieties, the endoproteinase Lys-C digest was treated with N-glycosidase A, an enzyme known [13] to cleave the β -aspartylglycosylamine linkage of all known types of asparagine-linked sugars and to convert the asparagine to aspartic acid. After this treatment it was found that the high

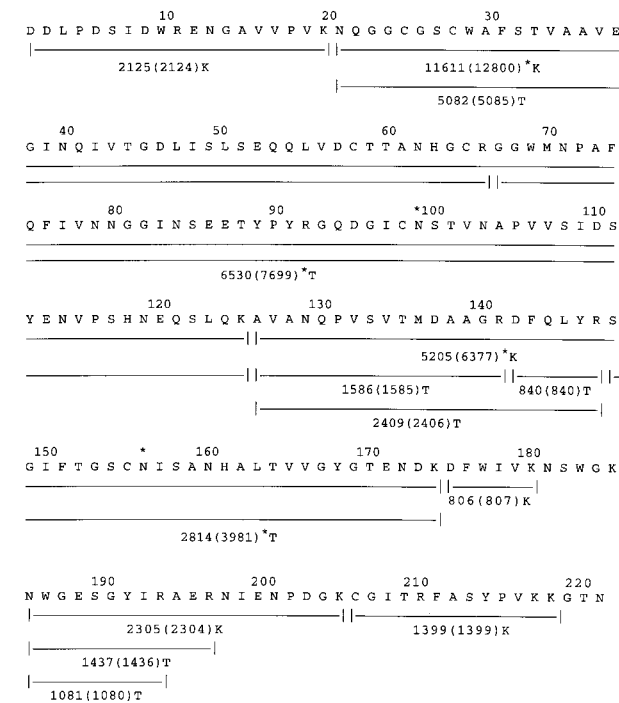


Fig. 4. Amino-acid sequence of GP-II indicating the masses of peptides detected by MALDI-TOF-MS of mixtures of endoprotease Lys-C (K) and of tryptic (T) peptides. Calculated masses ($M + H^+$) are shown, with observed masses in parentheses; putative glycopeptides are indicated by asterisks and are recognized by having masses ≈ 1170 Da higher than those calculated from the sequence. Cysteine residues are carboxy-methylated.

mass peptides had disappeared and peptides corresponding to the missing ones appeared.

MALDI-TOF MS of an endoproteinase Lys-C digest of GP-II produced six signals, four of which could be assigned to the known sequence (Fig. 4). Two expected signals with ($M + H^+$) = 5205 Da (residues 126–174) and 11611 Da (residues 20–125) were not found, but two with

($M + H^+$) = 6377 Da and 12800 Da were. In each case the difference between the observed and expected masses was ≈ 1180 Da, which suggested the presence of a glycan.

To narrow down the glycosylation sites, the endoproteinase Lys-C peptide mixture was further digested with trypsin. The resulting mixture showed 10 signals which could be assigned to positions in the sequence without ambiguity (Fig. 4). Previously unassigned positions 126–147 and 21–66 were accounted for by three tryptic peptides, the signals at ($M + H^+$) = 1586.8, and 841.4 Da being assigned to positions 126–141 and 142–147, respectively, and that at 5083 Da to position 20–66. These peptides, in addition to the endoproteinase Lys-C peptides account for 76% of the primary sequence of GP-II. At this point residues 67–125 and 148–174, which include the two potential glycosylation sites, remained unaccounted for. Two unidentified signals at ($M + H^+$) = 7704 Da and 3981 Da most likely corresponded to the glycosylated forms of residues of these peptides. Portions of the tryptic peptide mixture were then treated in parallel with N-glycosidase A and endoglycosidase F/N-glycosidase F. MS analysis of the N-glycosidase A-treated digest showed the loss of the 7704 Da and 3981 Da signals and appearance of two new signals at 6531 Da and 2813 Da, which correspond to residues 67–125 and 148–174, respectively (Table 1). No deglycosylation was noted in a parallel experiment with endoglycosidase F/N-glycosidase F. This experiment confirmed the presence of glycan moieties in the two tryptic peptides and also gave information about the nature of the glycans. Both N-glycosidase F and N-glycosidase A cleave the bond between N-linked glucosamine in the glycan and asparagine in glycoproteins and convert this residue to aspartic acid [13]. N-Glycosidase A has broader specificity than N-glycosidase F in that the former can release oligosaccharides containing core fucosylation at C-3 of the N-glucosamine residue attached to asparagine, whereas the latter cannot [14]. Endoglycosidase F hydrolyzes the glycosidic bond of the penultimate N-acetylglucosamine leaving the ultimate amino sugar (glucosamine) attached to asparagine, provided that this residue is not fucosylated. The unreactivity of the glycopeptides to endoglycosidase F/N-glycosidase F therefore implies that both carbohydrate chains contain core

Table 1. Identification of tryptic glycopeptides in GP-II. Carboxymethylated GP-II was digested first with Lys-C protease and then with trypsin. A portion of this digest was then treated with N-glycosidase A. Samples of the treated and untreated digests were analyzed by MALDI-TOF-MS in order to detect peptides which lost mass due to deglycosylation. ND, not determined.

Sequence location	C-terminal residue	[M + H] ⁺		Observed (+ Glycosidase)	Glycopeptide
		Calculated	Observed (– Glycosidase)		
1–19	K	2125	2125	2125	No
20–66	R	5085	5083	5083	No
67–125	K	6530	7704	6531	Yes
126–141	R	1586	1586	1585	No
142–147	R	841	840	840	No
148–174	K	2814	3981	2813	Yes
175–180	K	807	807	807	No
181–185	K	590	ND	ND	
185–195	R	1081	1081	1081	No
185–198	R	1437	1436	1436	No
206–217	K	1399	1399	1399	No
218–221	N	400	ND	ND	

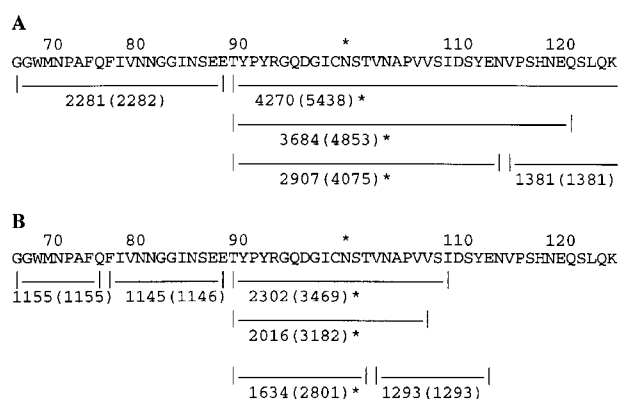


Fig. 5. Amino-acid sequence of the 7704-Da tryptic glycopeptide indicating those fragments identified by MALDI-TOF-MS after digestion (A) by endoproteinase Glu-C and (B) further digestion by thermolysin. Calculated masses ($M + H$)⁺ are shown, with observed masses in parentheses; putative glycopeptides are indicated by asterisks. The action of these enzymes reduced the 7704-Da peptide to the 2801-Da glycopeptide fragment, TYPYRGQDGICN*ST, which contains the glycosylation consensus site.

fucosylation at C-3 of the N-linked glucosamine residue attached to asparagine. From the foregoing experiments, 96% of the protein sequence could be accounted for. Only two small peptides whose masses were undetectable in the matrix noise of the mass spectra were not detected. Cleavage was observed at all the expected sites except for Arg10 and Arg92, which resisted proteolysis for unknown reasons. The digestion data are summarized in Table 1 and Fig. 4. In both cases the difference in mass between the glycosylated and nonglycosylated peptides is ≈ 1170 , which is close to the mass of the glycan found in the homologous cysteine protease, bromelain, which contains three hexose, two pentose and two *N*-acetylhexosamine residues [15,16].

As there is more than one asparagine residue in each of the tryptic glycopeptides, further degradation was required to locate the precise N-glycosylation site. The two tryptic glycopeptides were purified by reversed-phase HPLC and subjected to further proteolytic digestion, the larger [$M + H$]⁺ = 7704 Da by endoproteinase Glu-C and thermolysin, and the smaller [$M + H$]⁺ = 3981 by thermolysin alone. The resulting mixtures were analyzed by MALDI-TOF MS. As outlined in Fig. 5A, endoproteinase Glu-C cleaved the 7704-Da peptide into five fragments, two (corresponding to residues 67–87 and 114–125) of which could be assigned without ambiguity. Three other mass signals could also be identified if the carbohydrate mass of about 1168 was subtracted from their masses. The endoproteinase Glu-C peptide mixture was then

digested further with thermolysin (Fig. 6B). One nonglycosylated peptide of ($M + H$)⁺ = 1293 Da corresponds to residues 102–113, which eliminates Asn103 as a glycosylation site. Three apparent glycopeptides were observed with ($M + H$)⁺ = 3469, 3182 and 2801 Da, the latter of which corresponds to the sequence TYPYRGQDGICN*ST, with N* (Asn99), in the consensus sequence -Asn-Ser-Thr-, being the probable glycosylation site.

Thermolysin digestion of tryptic peptide of M_r 3981 showed four signals in the mass spectrum. That with ($M + H$)⁺ = 1296 Da corresponds to residues 163–174 (Fig. 6). The other peaks did not match directly, and small fragments (< 500 Da) could not be observed due to matrix interference. However, by subtracting 1168 Da from the masses of the unassigned peptides, signals with ($M + H$)⁺ = 2706, 2561, 2317 could be matched to position 148–162, 150–162, 148–158, respectively (Fig. 6). In this way the asparagine residue at position 156 in the sequence SGIFTGSCN*IS was deduced to be another N-glycosylation site, as predicted from the consensus sequence -Asn-Ile-Ser-.

These experiments demonstrate that both predicted glycosylation sites, Asn99 and Asn156, in GP-II are in fact occupied by glycans. Among the other cysteine proteases, only bromelain is known to have an asparagine-linked glycosylation site, namely, Asn119 (in GP-II numbering) in the sequence -Asn-Glu-Ser-[15] which corresponds to -Asn-Glu-Gln- (not a consensus glycosylation site) in GP-II.

Carbohydrate composition of the oligosaccharides

A sample of intact GP-II was hydrolyzed with methanolic HCl and the carbohydrates released were derivatized as described and analyzed by GLC. This analysis of the total carbohydrate in GP-II revealed the presence of glucosamine (GlcNAc), fucose (Fuc), mannose (Man) and xylose (Xyl) in the molar ratios of 2.0 : 1.0 : 2.7 : 1.0. A small amount of glucose detected was probably due to sample contamination.

Carbohydrate structure determination: ES MS and CID MS

A tryptic digest of GP-II was treated with N-glycosidase A and the intact oligosaccharide chains were isolated by reversed-phase HPLC and then permethylated. Electrospray (ES) MS revealed two carbohydrate forms: (Man)₃(Xyl)₁(Fuc)₁(GlcNAc)₂ and (Man)₃(Xyl)₁(Fuc)₁(GlcNAc)₃ (Fig. 7 and Table 2). Each methylated carbohydrate form appeared in two charged states: at m/z 1751.8⁺ and 887.1⁺ for (Man)₃(Xyl)₁(Fuc)₁(GlcNAc)₂ and at 1505.65⁺ and 764.6⁺ for (Man)₃(Xyl)₁(Fuc)₁(GlcNAc)₃. These two forms also appeared as the fragments (Man)₃(Xyl)₁(GlcNAc)₁ and (Man)₂(Xyl)₁(GlcNAc)₁ in which one GlcNAc and one fucose residue are missing. These fragments may have been due to a

Table 2. Composition of two glycans from GP-II determined by ES MS.

Composition	Charge state m/z (Da)		Glycoform abundance (%) ^a
	1 +	2 +	
Gn ₃ Man ₃ XylFuc ^b	1751.8	887.08	12.5
Gn ₂ Man ₃ XylFuc	1505.6	764.54	87.5
Gn ₂ Man ₃ Xyl	1331.6		
Gn ₁ Man ₃ Xyl	1086.5	554.83	

^a Glycoform abundances were calculated from the ion abundances at all charge states; ^b Gn, *N*-acetylglucosamine.

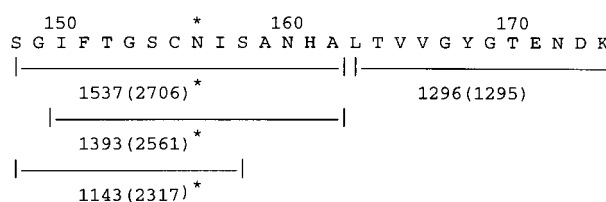
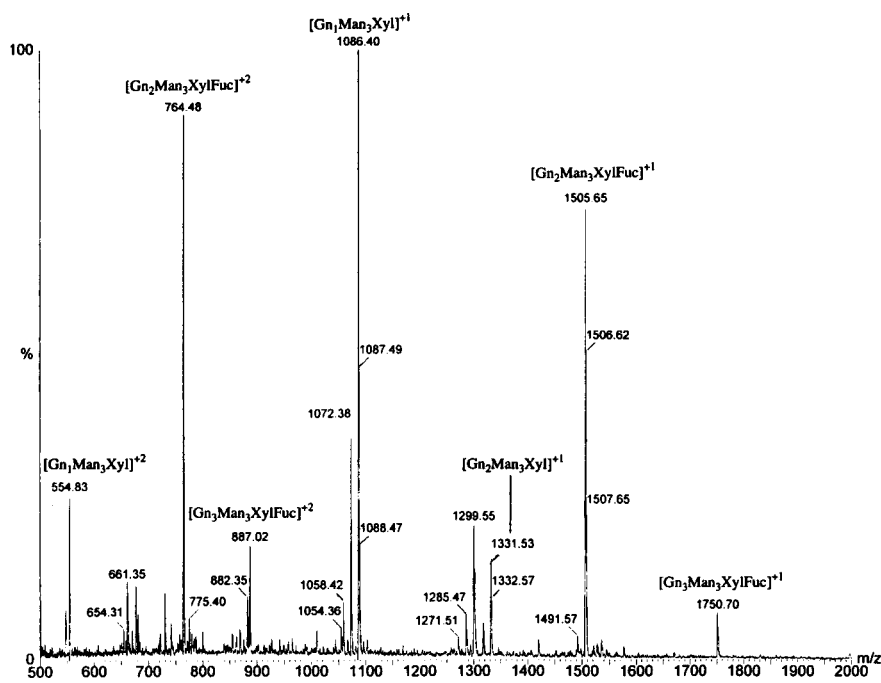


Fig. 6. Amino-acid sequence of the 3981-Da tryptic glycopeptide indicating fragments identified by MALDI-TOF MS of the thermolysin digest. Calculated masses ($M + H$)⁺ are shown, with observed masses in parentheses; putative glycopeptides are indicated by asterisks. Thermolysin reduced the tryptic peptide to the 2317-Da fragment, SGIFTGSCN*IS.

Fig. 7. Electrospray MS on the N-linked oligosaccharides of GP-II. The assigned fragments are shown above the peaks and are summarized in Table 2.



prolonged incubation time of N-glycosidase digestion at acidic pH. Of the two glycoforms $(\text{Man})_3(\text{Xyl})_1(\text{Fuc})_1(\text{GlcNAc})_2$ was the major component (87.5%) compared with $(\text{Man})_3(\text{Xyl})_1(\text{Fuc})_1(\text{GlcNAc})_3$ (12.5%). As it had been determined that glycopeptides representing both N-glycosylation sites contained approximately the same molar mass of carbohydrate, it seems likely that each glycosylation site contains both $(\text{Man})_3(\text{Xyl})_1(\text{Fuc})_1(\text{GlcNAc})_2$ and $(\text{Man})_3(\text{Xyl})_1(\text{Fuc})_1(\text{GlcNAc})_3$ in a ratio of about 7 : 1, rather than one site being occupied by the larger glycan and the other site by the smaller one.

Monomer sequence and branching information was obtained by CID MS of the methylated glycan. Monosaccharide residues are connected by labile glycosidic bonds which are extensively fragmented by low energy collision. For methylated and alkali metal-cationized oligosaccharides, glycosidic fragments are a consequence of $\text{C}_1\text{--O}$ cleavage with a transfer of a hydrogen to the oxygen by elimination across the proximal $\text{C}_1\text{--C}_2$ bond [17]. To ascertain linkage and branching information $(\text{Man})_3(\text{Xyl})_1(\text{GlcNAc})_1$ (m/z 1086.2⁺) was chosen for CID MS (Fig. 8). Major ions identified as glycosidic bond rupture products are shown in Table 3. The ion series at m/z -values of

Fig. 8. CID MS of a fragment for the N-linked oligosaccharide of GP-II. The parent ion, m/z 1086.2 of the major peak in Fig. 7 was fragmented. The assigned fragments are summarized in Table 3.

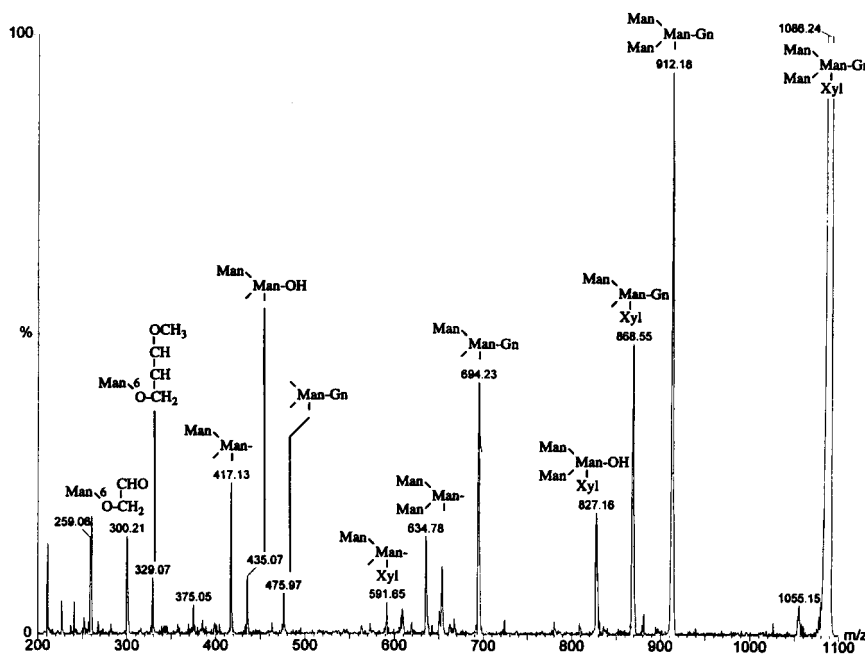


Table 3. Structural assignment of CID fragments.

<i>m/z</i>	Composition ^a
912.18	Man ₃ Gn
868.55	Man ₂ XylGn
827.16	Man ₃ Xyl-OH
694.11	Man ₂ Gn
634.72	Man ₃
591.65	Man ₂ Xyl
475.97	ManGn
435.07	Man ₂ -OH
417.13	Man ₂
328.94	Man-O-CH ₂ -CH ₂ -CH ₂ -OCH ₃
300.27	Man-O-CH ₂ -CHO
260.11	Xyl-O-CH ₂ -CH ₂ -OH

^a Gn, *N*-acetylglucosamine.

417, 435, 475, 591, 634, 694, 827.2, 868 and 912.2 represents a complete sequence for the (Man)₃(Xyl)₁(GlcNAc)₁ structure. Specifically, the fragment at *m/z* 476 clearly indicates that two terminal mannose residues are branched, rather than linear, as seen in stem bromelain [15] (Fig. 9). The *m/z* 476 peak represents the fragment, Man-GlcNAc, having a single methyl group on mannose, the other three hydroxyls having been occupied by two mannose units and a xylose; in a linear structure this fragment would have contained two methyl groups (cf. Fig. 9).

Three small fragments represent important cross-ring cleavage products that define linkage types. The first one, at *m/z* 260 atomic mass units defines the linkage of the terminal xylose residue to the 2-hydroxyl of the core mannose residue. The latter two, at *m/z* 300 and 329 atomic mass units are cross-ring fragments of a nonreducing terminal mannose residue specific for the 1→6 linkage [17]. The primary structure for (Man)₃(Xyl)₁(Fuc)₁(GlcNAc)₂ was thus deduced to be Man1→6(Man1→3)(Xyl1→2)Man1→4GlcNAc1→4(Fuc1→3)GlcNAc, which is a biantennary complex-type oligosaccharide typically, in plant

proteins [18], with a xylose residue linked to the mannosyl residue by a β1→2 linkage and a fucose residue attached by an α1→3 linkage to the *N*-acetylglucosamine that is linked to the polypeptide chain (Fig. 9). Unfortunately CID MS of the minor component oligosaccharide was not available but ES MS gave some information about the structure, namely that the additional *N*-acetylglucosamine is attached to a terminal mannosyl residue.

DISCUSSION

Comparison of the sequences of the ginger proteases with those of other cysteine proteases shows clearly that they are members of the papain family (Fig. 3) [1]. The ginger proteases contain seven cysteine residues, and they are located at the same positions as the seven cysteines found in other cysteine proteases [1], suggesting identical disulfide crosslinking patterns. On this basis, we conclude that Cys25 (Fig. 4, papain numbering) provides the free thiol group of the active site, and that the remaining six residues form the pairs Cys22–Cys63, Cys 56–Cys95 and Cys153–Cys200. These conclusions have been confirmed by analysis of an X-ray crystal structure of GP-II [9]. Other conserved, catalytically important residues are His159, which participates in general acid/base catalysis; Gln19, which helps form the 'oxanion hole' and Asn175, which orients the imidazolium ring of His159 in positions optimal for various steps in the catalytic mechanism [19,20]. GP-II shows 82% homology with GP-I, 58% with actinidin, 50% with glycyl endopeptidase, 48% with protease omega and 47% with papain. GP-II also contains two consensus glycosylation sites at Asn96 and Asn154 (Fig. 3), both of which are occupied; potential GP-I glycosylation sites are at residues Asn92 and Asn154 (papain numbering).

The nature of the residue at position P₂ of peptide substrates has been shown to be the most significant in terms of determining specificity for most of the cysteine proteases. For example, it has been established that papain prefers to cleave polypeptides with an aromatic amino acid (e.g. Phe or Tyr) in the P₂ position [21], although cleavage occurs more slowly at many other sites, too. The ginger proteases are unusual in showing a preference for proline in the P₂ position (unpublished observations). Examination of the crystal structure of an Acetyl-Ala-Ala-Phe-Ala-methylene derivative of papain allows identification of the residues in the S₂ subsite, whose side chains make the most intimate contacts with the P₂ side chain of the substrate (or inhibitor), namely Tyr67, Pro68, Trp69, Val133, and Val157, which together form a hydrophobic pocket [22,23]. In GP-II these are replaced by Trp67, Met68, Asn69, Thr133, and Ala157 (Fig. 3). As shown by crystallographic studies [9], the volume of the GP-II S₂ subsite is only about half that of papain, presumably preventing the binding of large hydrophobic residues.

Using IEF Ohtsuki *et al.* [3] detected comparable amounts of three forms of ginger protease having pI values of 4.5, 4.6 and 4.8. The pI for GP-II, calculated from its amino-acid sequence (http://expasy.hcu.ge.ch/cgi-bin/pi_tool), is 4.82 and that for GP-I is either 5.05 or 5.16, depending, respectively, on whether residue 2 is aspartic acid or valine. In any case, the lower calculated pI value for GP-II indicates that it is the more acidic protein, which is consistent with its eluting later from the DEAE-cellulose column. GP-I may be a mixture of two closely related, but more basic proteins, a complication that was not appreciated until late into the sequencing studies. For this reason, and because of the ambiguities discussed above, the GP-I sequence reported here must be regarded as tentative. The

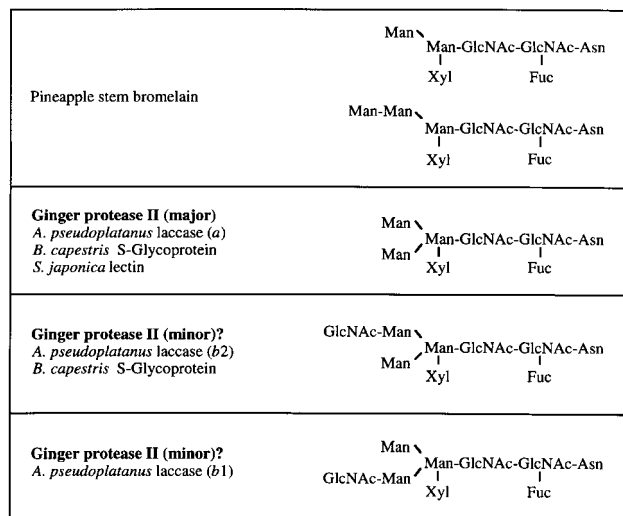


Fig. 9. Comparison of GP-II and other plant glycan structures.

appearance of multiple forms of the ginger proteases is not remarkable, however, as such heterogeneity has also been seen in other species, for example, papaya [24] and pineapple stem [25].

The N-linked oligosaccharides of plant glycoproteins typically have 5–13 glycosyl units and fall into two main classes: high-mannose-type glycans and complex-type glycans. Both classes of glycans have a core of two N-glucosamine residues (*N,N'*-diacetylchitobiose) linked at the reducing end to asparagine and at the nonreducing end to mannose. The high-mannose-type glycans have up to nine mannose residues in a characteristic Man–Man branching pattern, whereas in the complex glycans some mannose residues are replaced by additional sugars such as fucose, xylose, and galactose [18,26]. Characteristic of plant glycans are Xyl β -1 $\rightarrow$ 2-linked to the core mannose residue and fucose α -1 $\rightarrow$ 3-linked to the GlcNAc proximal to the protein. Animal glycoproteins contain fucose α -1 $\rightarrow$ 6-linked to the proximal GlcNAc residue and sialic acid, which is not found in plants. Therefore, the oligosaccharides of GP-II appear to be typical plant glycans of the complex type.

Detailed NMR studies have established the structures and conformations of the N-linked oligosaccharides of the cysteine proteinase bromelain [16,27] (Fig. 9). Although largely similar, ginger protease oligosaccharides differ in two respects from those of bromelain. First, the additional mannose is linked to the core mannosyl residue by an α 1 $\rightarrow$ 3 linkage. Second, the structure (of the minor form) is extended with an additional GlcNAc rather than a mannose. On the other hand, the major carbohydrate moiety of GP-II appears to be identical to those found in several species: a lectin isolated from the seeds of the Japanese pagoda tree (*Sophora japonica*) [28], laccase a from sycamore cells (*Acer pseudoplatanus*) [29] and the S-glycoprotein of *Brassica campestris* [30]. Although we have not yet established which of the terminal mannose residues the additional GlcNAc residue is attached to in GP-II, both possible forms have been found in *Acer pseudoplatanus* (laccase b1 and b2) and one in S-protein isolated from *Brassica campestris* [13]. As the glycan of GP-II seems to be a typical plant glycan with a covalent structure identical to that of other plant glycans of known conformation, we predict that the GP-II has the structure Man α 1 $\rightarrow$ 6(Man α 1 $\rightarrow$ 3)(Xyl β 1 $\rightarrow$ 2)-Man β 1 $\rightarrow$ 4GlcNAc β 1 $\rightarrow$ 4(Fuc α 1 $\rightarrow$ 3)GlcNAc, although the assignment of the anomeric configurations remains to be confirmed experimentally.

ACKNOWLEDGEMENTS

The authors thank V. Reinhold and K. Linsley at the Boston University Medical Center for mass spectral analyses of the glycan structures.

REFERENCES

- Rawlings, N.D. & Barrett, A.J. (1994) Families of cysteine peptidases. *Methods Enzymol.* **244**, 461–486.
- Ichikawa, Y., Sasa, H. & Michi, K. (1973) Purification of ginger protease. *J. Jap. Soc. Food Nutr.* **26**, 337–383.
- Ohtsuki, K., Taguchi, K., Sato, K. & Kawabata, M. (1995) Purification of ginger proteases by DEAE-Sepharose and isoelectric focusing. *Biochim. Biophys. Acta* **1243**, 181–184.
- Hashimoto, A., Takeuti, Y., Kawahara, Y. & Yasumoto, K. (1991) Proteinase and collagenase activities in ginger rhizome. *Japan. Soc. Nutr. Food Sci.* **44**, 127–132.
- Lee, Y.B., Shenert, D.J. & Ashmore, C.R. (1986) Tenderization of meat with ginger rhizome protease. *J. Food Sci.* **51**, 1558–1559.
- Thompson, E.H., Wolf, I.D. & Allen, C.E. (1973) Ginger rhizome: a new source of proteolytic enzyme. *J. Food Sci.* **38**, 652–655.
- Duan, Y. & Laursen, R.A. (1994) Protease substrate specificity mapping using membrane-bound peptides. *Anal. Biochem.* **216**, 431–438.
- Duan, Y. (1994) *Protease substrate specificity mapping using membrane-bound peptides and characterization of proteases from ginger root*. PhD Thesis, Boston University, USA.
- Choi, K.H., Laursen, R.A. & Allen, N.N. (1999) The 2.1 Å structure of a cysteine protease with proline specificity from ginger rhizome, *Zingiber officinale*. *Biochemistry* **38**, 11624–11633.
- Boland, M.J. & Hardman, M.J. (1972) Kinetic studies on the thiol protease from *Actinidia chinensis*. *FEBS Lett.* **27**, 282–284.
- Cavallini, D., Graziani, M.T. & Dupré, S. (1966) Determination of disulphide groups in proteins. *Nature* **212**, 294–295.
- Carr, S.A. & Roberts, G.D. (1986) Carbohydrate mapping by mass spectrometry: a novel method for identifying attachment sites of Asn-linked sugars in glycoproteins. *Anal. Biochem.* **157**, 396–406.
- O'Neill, R.A. (1996) Enzymatic release of oligosaccharides from glycoproteins for chromatographic and electrophoretic analysis. *J. Chromatogr. A* **720**, 201–215.
- Tretter, V., Altmann, F. & März, L. (1991) Peptide-N4-(*N*-acetyl-beta-glucosaminyl) asparagine amidase F cannot release glycans with fucose attached alpha 1–3 to the asparagine-linked *N*-acetylglucosamine. *Eur. J. Biochem.* **199**, 647–652.
- Ishihara, H., Takahashi, N., Oguri, S. & Tejima, S. (1979) Complete structure of the carbohydrate moiety of stem bromelain. *J. Biol. Chem.* **254**, 10715–10719.
- van Kuik, J.A., Hoffman, R.A., Mutsaers, J.H.G.M., van Halbeek, H., Kamerling, J.P. & Vliegthart, J.F.G. (1986) A 500-Mhz ¹H-NMR study on the N-linked carbohydrate chain of bromelain. *Glyconjugate J.* **3**, 27–34.
- Reinhold, V.N., Reinhold, B.B. & Costello, C.E. (1995) Carbohydrate molecular weight profiling, sequence, linkage, and branching data: ES-MS and CID. *Anal. Chem.* **67**, 1772–1784.
- Sturm, A. (1991) N-Glycosylation of plant proteins. *Eur. J. Biochem.* **199**, 169–179.
- Storer, A.C. & Ménard, R. (1991) Catalytic mechanism in papain family of cysteine peptidases. *Methods Enzymol.* **244**, 486–500.
- Drenth, J., Jansonius, J.N., Koekkoek, R., Swen, H.M. & Wolthers, B.G. (1968) Structure of papain. *Nature* **218**, 929–932.
- Berger, A. & Schechter, I. (1970) Mapping the active site of papain with the aid of peptide substrates and inhibitors. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **257**, 249–264.
- Drenth, J., Kalk, K.H. & Swen, H.M. (1976) Binding of chloromethyl ketone substrate analogues to crystalline papain. *Biochemistry* **15**, 3731–3738.
- Topham, C.M., Salih, E., Frazao, C., Kowlessur, D., Overington, J.P., Thomas, M., Brocklehurst, S.M., Patel, M., Thomas, E.W. & Brocklehurst, K. (1991) Structure–function relationships in the cysteine proteinases actinidin, papain and papaya proteinase Ω . *Biochem. J.* **280**, 79–92.
- Watson, D.C., Yaguchi, M. & Lynn, K.R. (1990) The amino acid sequence of chymopapain from carica papaya. *Biochem. J.* **266**, 75–81.
- Napper, A.D., Bennett, S.P., Borowski, M., Holdridge, M.B., Leonard, M.J.C., Rogers, E.E., Duan, Y., Laursen, R.A., Reinhold, B. & Shames, S.L. (1994) Purification and characterization of multiple forms of the pineapple-stem-derived cysteine proteinases ananain and comosain. *Biochem. J.* **301**, 727–735.
- Sturm, A. (1995) N-Glycosylation of plant proteins. In: *Glycoproteins* (Montreuil, J., Vliegthart, J.F.G. & Schacter, H., eds), pp. 521–541. Elsevier Science Publishers, Amsterdam.
- Bouwstra, J.B., Spoelstra, E.C., de Waard, P., Leeftang, B.R., Kamerling, J.P. & Vliegthart, J.F.G. (1990) Conformational studies

- on the N-linked carbohydrate chain of bromelain. *Eur. J. Biochem.* **190**, 113–122.
28. Fournet, B., Leroy, Y., Wieruszeski, J.-M., Montreuil, J. & Poretz, R.D. (1987) Primary structure of an N-glycosidic carbohydrate unit derived from *Sophora japonica* lectin. *Eur. J. Biochem.* **166**, 321–324.
29. Takahashi, N., Hotta, T., Ishihara, H., Mori, M., Tejima, S., Bligny, R., Akazawi, T., Endo, S. & Arata, Y. (1986) Xylose-containing common structural unit in N-Linked Oligosaccharides of laccase from sycamore cells. *Biochemistry* **25**, 388–395.
30. Takayama, S., Isogai, A., Tsukamoto, C., Ueda, Y., Hinata, K., Okazaki, K. & Suzuki, A. (1987) Sequences of S-glycoproteins, products of the *Brassica campestris* self-incompatibility locus. *Nature* **326**, 102–105.

