

## Plant collagenase: Unique collagenolytic activity of cysteine proteases from ginger<sup>☆</sup>

Misook Kim<sup>a,\*</sup>, Susan E. Hamilton<sup>a</sup>, Luke W. Guddat<sup>a</sup>, Christopher M. Overall<sup>b</sup>

<sup>a</sup> School of Molecular and Microbial Sciences, The University of Queensland, St Lucia, Australia

<sup>b</sup> Department of Biochemistry and Molecular Biology, The University of British Columbia, Vancouver, Canada

Received 13 May 2007; received in revised form 31 July 2007; accepted 1 August 2007

Available online 24 August 2007

### Abstract

Two cysteine proteases, GP2 and GP3, have been isolated from ginger rhizomes (*Zingiber officinale*). GP2 is virtually identical to a previously identified ginger protease GPII [K.H. Choi, and R.A. Laursen, Amino-acid sequence and glycan structures of cysteine proteases with proline specificity from ginger rhizome *Zingiber officinale*, *Eur. J. Biochem.* 267 (2000) 1516–1526.], and cleaves native type I collagen at multiple discrete sites, which are in the interior of the triple helical region of this molecule. In reaction with proline-containing peptides GP2 shows preference for Pro in the P2 position, and at least 10-fold higher efficiency of hydrolysis than papain. Comparison of models of GP2 and GP3 with the crystal structure of papain shows that the three enzymes have different S2 pocket structures. The S2 pocket in GP2 and GP3 is half the size of that of papain. GP2 is the only reported plant cysteine protease with a demonstrated ability to hydrolyse native collagen. The results support a role for ginger proteases as an alternative to papain, in commercial applications such as meat tenderization, where collagen is the target substrate. Crown Copyright © 2007 Published by Elsevier B.V. All rights reserved.

**Keywords:** Plant; Ginger; Collagenase; Hydrolysis; Proline specificity; Meat tenderization

A restricted range of proteolytic enzymes hydrolyse native collagen. These include bacterial enzymes from *Clostridium histolyticum* [2–5], matrix metalloproteinases from mammalian tissues [6], selected cysteine cathepsins [7,8], invertebrate serine proteases from fiddler crab [9] and a cysteine protease from arctic shrimp [10]. All of these enzymes hydrolyse native type I triple helical collagen at discrete, highly specific sites. All three chains of the collagen triple helix are usually hydrolysed at these specific sites.

The collagenolytic activity of these enzymes has commercial application. Bacterial collagenase hydrolyses bovine trachea cartilage allowing the preparation of intact mammalian cells in culture [11]. Potential therapeutic applications include the enzymatic debridement of necrotic tissue associated with burns [12], cleaning blood cells for improved screening in medical diagnostics [13]. Recently collagen peptides produced by collagenolytic enzymes have been used for the treatment of diseases such as collagen-induced arthritis [14] as well. The enzyme most commonly used in meat tenderization by the hydrolysis of connective tissue proteins is papain [15]. The hydrolysis of connective tissue proteins by papain occurs at elevated temperatures during cooking, with the substrate being denatured collagen [16]. Other connective tissue proteins are indiscriminately hydrolysed by papain at these temperatures, leading to undesirable characteristics such as “mushiness” [17]. Application of alternative collagenase to this process may therefore provide an improved product. Recent reports using ginger extracts for meat tenderization indicate that the extracts may contain collagenolytic activity, though this has not yet been biochemically characterised [18,19]. The results also suggested that ginger extracts showed a greater proteolytic activity towards

**Abbreviations:** GP, ginger protease; CHC, *Clostridium histolyticum* collagenase; DTT, dithiothreitol; Dpa, *N*-3-(2,4-dinitrophenyl)-L-2,3-diaminopropionyl; E64, *L*-trans-epoxysuccinyl-leucylamido (4-guanidino)-butane; HyP, hydroxyproline; Nva, *L*-norvaline; SDS-PAGE, polyacrylamide gel electrophoresis in the presence of SDS; STT, sodium tetrathionate; Mca, 7-methoxycoumarin-4-acetyl; MMP, matrix metalloproteinase

<sup>☆</sup> The nucleotide sequences reported in this work have been deposited in the NCBI (GenBank®). The accession numbers are: gp2a, AY700585; gp2b, AY700586; gp3a, AY700587; gp3b, AY700588.

\* Corresponding author. CSIRO, Livestock Industries, 306 Carmody Rd, St Lucia, QLD4067, Australia. Tel.: +61 7 3214 2544; fax: +61 7 3214 2900.

E-mail address: [misook.kim@csiro.au](mailto:misook.kim@csiro.au) (M. Kim).

[illegible]

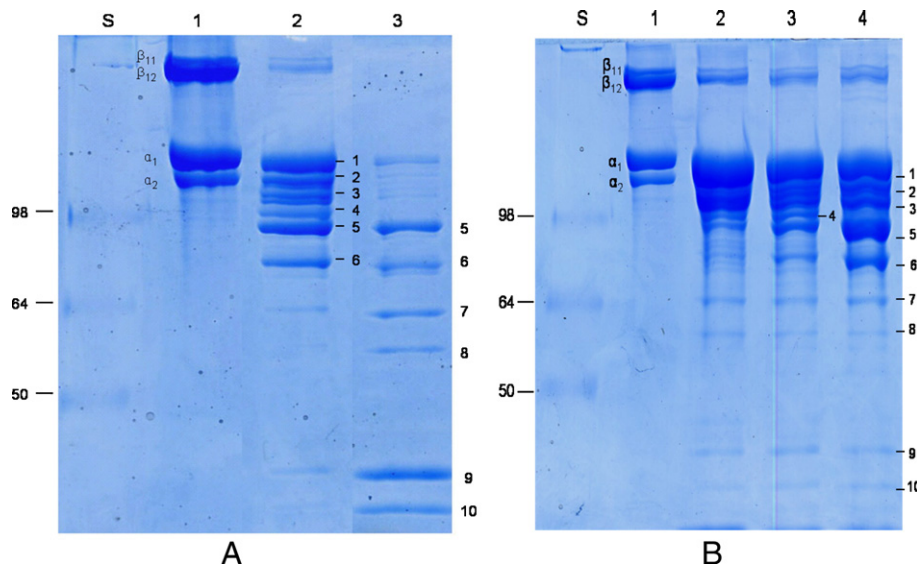


Fig. 1. SDS-PAGE showing the digestion of bovine type I collagen by GP2. (A) Native calf skin type I collagen was incubated with GP2 (1.2 U and 5.7 U) for 64 h at  $\sim 22^\circ\text{C}$ . The reactions were stopped by the addition of  $200\ \mu\text{M}$  E-64. Lane S: molecular weight standards; lane 1: collagen incubated without enzyme; lanes 2–3: collagen incubated with 1.2 U and 5.7 U GP2 respectively. Estimated apparent molecular weights of the fragments are 219, 200, 124, 116, 119, 119, 115–110, 106–103, 98, 92, 81, 64, 58, 40 and 36 kDa for band numbers  $\beta_{11}$ ,  $\beta_{12}$ ,  $\alpha_1$ ,  $\alpha_2$ , 1 to 10 in order, respectively. (B) Time course for digestion; native calf skin type I collagen was incubated with GP2 (3.2 U) at ambient temperature ( $\sim 22^\circ\text{C}$ ). The reactions were stopped by the addition of  $200\ \mu\text{M}$  E-64 after 1, 2, 5 days and the samples were analysed by SDS-PAGE as described in Materials and methods. Lane S: molecular weight standards; lane 1: collagen incubated without enzyme; lanes 2–4: collagen incubated for 24 h, 48 h, and 5 days respectively. Note that the estimates are taken from the front of the broad, high molecular weight bands. Estimated apparent molecular weights of the fragments are 219, 200, 124, 116, 119, 111, 103, 98, 89, 78, 64, 58, 40, and 36 kDa for band numbers  $\beta_{11}$ ,  $\beta_{12}$ ,  $\alpha_1$ ,  $\alpha_2$ , 1 to 10 in order, respectively. Numbering is chosen to facilitate the comparison with the ones in panel A.

### 1.6. Quenched fluorescent substrate assay

Assays were performed at  $25^\circ\text{C}$  in a  $200\ \mu\text{l}$  reaction mixture containing  $0.1\ \text{M}$  sodium acetate, pH 5.5 with  $5\ \mu\text{l}$  of activated enzymes (with ca.  $70\ \text{U/mg}$  of caseinolytic activity) and  $5\ \mu\text{M}$  peptide substrates containing intra-molecularly quenched 7-methoxycoumarin-4-acetyl (Mca), of which peptide sequences are Mca-PLGL-DpaAR-NH<sub>2</sub>, Mca-PLANva-DpaAR-NH<sub>2</sub>, and Mca-QPVGINTS-DpaRR-NH<sub>2</sub>, respectively. Substrate 1 is a general MMP substrate [24], substrate 2 a MMP3/stromelysin1 substrate [25], and substrate 3 resembles the N-terminus of MCP-3 (monocyte chemo-attractant protein) [26]. Stock substrate solutions were prepared at a concentration of  $0.1\ \text{mM}$  in dimethyl sulfoxide (DMSO). Hydrolysis reactions were initiated by the addition of the substrates diluted in water from DMSO stock, and enzymatic activities were followed by monitoring the release (during 1 h) of the fluorogenic group, Mca, at an excitation wavelength of  $320\ \text{nm}$  and an emission wavelength of  $405\ \text{nm}$ . Assays were performed in a microplate reader (Fluorescence Microplate Reader, Molecular Device, California) using SOFTmaxPRO software. MMP2 was used as control under the

same conditions in  $0.1\ \text{M}$  Tris pH 7.5, containing  $10\ \text{mM}$  CaCl<sub>2</sub>,  $0.1\ \text{M}$  NaCl and  $0.05\%$  Brij35 to establish the assay. Initial velocities were obtained from plots of fluorescence versus time, using data points from the linear portion of the hydrolysis curve. In this assay, one unit of enzyme activity was defined as the amount of enzyme causing a change of fluorescence intensity per second, and normalised based on the caseinolytic activity (GP2,  $70.5\ \text{U/mg}$ ; GP3,  $68.2\ \text{U/mg}$ ; papain,  $80.6\ \text{U/mg}$ ). Cleaved products after 24-h incubation at room temperature were purified by C<sub>18</sub> mini-column (ZipTipC<sub>18</sub>, Millipore), and  $\sim 10\ \text{pmol}$  of the unseparated peptide mixture was analysed by mass spectrometry (Fig. S3, Supplementary material) in a Voyager System 4311 (Applied Biosystems).

### 1.7. Modelling of GP2 and GP3 structures

The sequences for GP2 and GP3 were based on the deduced sequences of cloned cDNAs (GenBank gp2b; AY700586, gp3a; AY700587). The coordinates of the structure of GPII (PDB:1CQD [27]) were used as the starting template.

Table 2  
N-terminal amino acid sequencing of fragments generated by the action of GP2 on native bovine type I collagen

| Band no. | Sequence obtained                                                              | Best matching sequence within $\alpha_1$ or $\alpha_2$ chains <sup>a</sup> |
|----------|--------------------------------------------------------------------------------|----------------------------------------------------------------------------|
| 5        | G E E G K<br><i>G E A R P<sup>b</sup></i><br><i>G E E G A<sup>b</sup></i>      | $\alpha_1$ or $\alpha_2$ <sup>286</sup> GEEGK                              |
| 6        | G E E G K R<br><i>G E L G P<sup>2b</sup></i><br><i>G A L T P A<sup>b</sup></i> | $\alpha_2$ <sup>286</sup> GEEGKR                                           |
| 7        | G E R G F <sup>HY</sup> P                                                      | $\alpha_1$ <sup>499</sup> GERGF <sup>HY</sup> P                            |
| 8        | G E R G P <sup>HY</sup> P L A I                                                | $\alpha_2$ <sup>499</sup> GERGP <sup>HY</sup> P                            |
| 9        | G N A G P <sup>HY</sup> P                                                      | $\alpha_1$ <sup>712</sup> GNAGP <sup>HY</sup> P                            |
| 10       | G I S G P <sup>HY</sup> P G P <sup>HY</sup> P G P                              | $\alpha_2$ <sup>712</sup> GISGPP <sup>C</sup> GPP <sup>C</sup> GP          |

<sup>a</sup>The numbering of the residues starts from the first triplet (G–X–Y) of the triple helix of bovine type I collagen. <sup>b</sup>Residual underlying sequences are italicized.

<sup>c</sup>Sequence database do not distinguish between Pro and HyP.

Table 3  
Comparison of the experimental and calculated sizes of the fragments derived from the cleavage of bovine type I collagen by the action of GP2

| Band no. | N-terminal amino acid sequence             | Possible origin of fragment <sup>a</sup> | Calculated size of C-terminal fragment <sup>b</sup> (kDa) | Observed size on SDS-PAGE (kDa) |
|----------|--------------------------------------------|------------------------------------------|-----------------------------------------------------------|---------------------------------|
| 5        | <sup>286</sup> GEEGK                       | α1 or α2                                 | α1 91.3, α2 83.1                                          | 92                              |
| 7        | <sup>499</sup> GERGF <sup>Hyp</sup>        | α1                                       | α1 64.5                                                   | 64                              |
| 9        | <sup>712</sup> GNAGP <sup>Hyp</sup>        | α1                                       | α1 38.1                                                   | 40                              |
| 6        | <sup>286</sup> GEEGKR                      | α2                                       | α2 83.1                                                   | 81                              |
| 8        | <sup>499</sup> GERGP <sup>Hyp</sup>        | α2                                       | α2 58.8                                                   | 58                              |
| 10       | <sup>712</sup> GISGP <sup>Hyp</sup> PGPPGP | α2                                       | α2 34.4                                                   | 36                              |

<sup>a</sup> Alignments of the sequences of the triple helical region of type I collagen are in Appendices 1 and 2, Supplementary material.  
<sup>b</sup> Calculated based on a single cleavage.

The “mutate and replace” option within the program O [28] was used to make the necessary changes to GPII to make starting models of GP2 (Fig. 5) and GP3. Once the correct amino acids were in place, the choice of rotamer was based on avoiding steric clashes with nearby residues. A suitable rotamer could always be found that resulted in reasonable van der Waals and/or hydrogen bonding interactions with residues in close proximity. None of the mutations for GP2 were within the active site.

2. Results

2.1. Cleavage of native type I collagen

Native calf skin type I collagen was incubated for 64 h at 22 °C with various concentrations of enzyme (Fig. 1A). The β bands of collagen were rapidly and completely degraded as a result of hydrolysis in the non-helical, cross-linked ends of the molecule. At the lower enzyme concentration, 6 major high molecular weight fragments (labelled bands 1–6) were produced, with the apparent molecular weights indicated in the figure legend, with bands 2 and 3 appearing as doublets. At the higher enzyme concentration, a series of 6 fragments (bands 5 to 10) were produced with apparent molecular weights indicated. When the time of digestion was varied from 1 to 5 days at a single enzyme concentration (3.2 U/ml) the final pattern was essentially identical to that shown in lane 2, Fig. 1A (Fig. 1B). The data are consistent with an accumulation of fragments corresponding to bands 5 and 6 prior to the appearance of the four smaller bands.

In order to identify the cleavage sites within collagen triple helix, the N-terminal amino acid sequences of the six major bands (5–10, Fig. 1) were determined by Edman degradation. Sequencing of bands 7 to 10 resulted in single sequences whereas sequence information obtained for bands 5 and 6 was consistent with the presence of one major sequence and one or more minor sequences (Table 2). The experimentally obtained sequences were compared with the known sequence for the

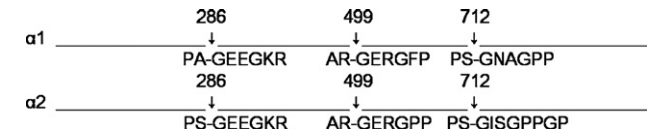


Fig. 2. GP2 cleavage sites in bovine collagen type I.

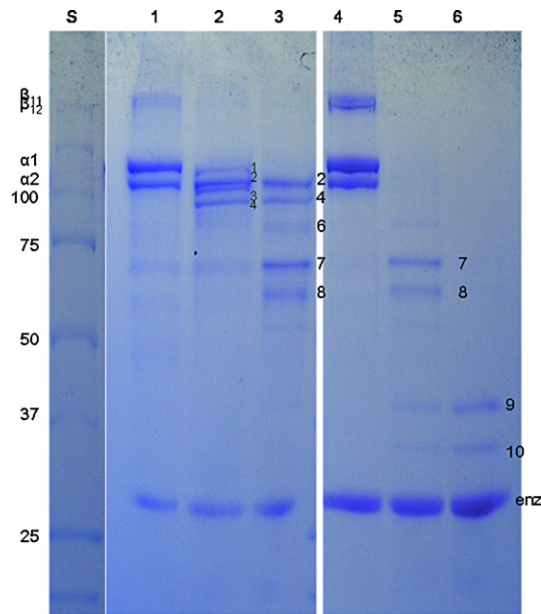


Fig. 3. SDS-PAGE showing cleavage of rat tail type I collagen by GP2. Native rat tail type I collagen was incubated with GP2 (2.5 U and 6.5 U) for up to 48 h at 22 °C. The reactions were stopped by the addition of E-64. Lane S: molecular weight standards (kDa); lanes 1–3: 0 h, 24 h, 48 h digests with 2.5 U GP2, respectively; lanes 4–6: 0 h, 24 h, 48 h digests with 6.5 U GP2, respectively. Estimated apparent molecular weights of the fragments are 217, 200, 126, 112, 119, 111, 105, 97, 68, 60, 39 and 34 kDa for band numbers β11, β12, α1, α2, 1, 2, 3, 4, 7, 8, 9, and 10, respectively. Numbering is chosen to facilitate the comparison with bovine collagen cleavage shown in Fig. 1.

α1 or α2 chains of bovine type I collagen (see Appendices, Supplementary material). For bands 6 to 10, single cleavage sites were located unequivocally in one of the two chains. For band 5, the sequence GEEGK was identical to a sequence in the α1 and α2 chains commencing at Gly286. It was therefore not possible

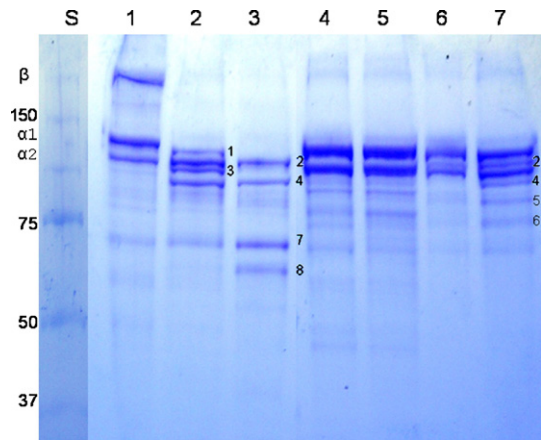


Fig. 4. Comparison of the cleavage of rat collagen by GP2, GP3 and papain. Native rat tail type I collagen was incubated with GP2, GP3, or papain (2.5 U) at ~22 °C and pH 5.5 for 2 days. The reactions were stopped by the addition of E64 and analysed by SDS-PAGE (4–20% gradient). Lane S: molecular weight standards (kDa); lane 1: rat collagen incubated without enzyme; collagen incubated for 1 day and 2 days with GP2 (lanes 2–3), papain (lanes 4 to 5), and GP3 (lanes 6–7), respectively. The numbering of the bands was chosen to facilitate comparison with bovine (Fig. 1) and rat (Fig. 3) collagen cleavage.

Table 4

Relative enzyme activities<sup>a</sup> and cleavage sites (↓) within peptide substrates by GP2, GP3, and papain

| Peptides | GP2                                          |                   | GP3                |                   | Papain    |                   |
|----------|----------------------------------------------|-------------------|--------------------|-------------------|-----------|-------------------|
|          | Cleavage                                     | Relative activity | Cleavage           | Relative activity | Cleavage  | Relative activity |
| 1        | PL↓GLAR <sup>b</sup><br>PLG↓LAR <sup>b</sup> | 1.41              | PL↓GLAR<br>PLG↓LAR | 1.23              | PLG↓LAR   | 0.10              |
| 2        | PL↓ANva<br>PLA↓Nva                           | 2.80              | PL↓ANva<br>PLA↓Nva | 4.62              | PLA↓Nva   | 0.11              |
| 3        | QPV↓GINTS<br>QPVG↓INTS <sup>a</sup>          | 1.29              | QPVG↓INTS          | 4.94              | QPVG↓INTS | 0.29              |

<sup>a</sup> Relative caseinolytic activity (see Materials and methods).<sup>b</sup> Known cleavage site of MMP.

to assign this cleavage to one of the two chains. In order to confirm the cleavage sites, the experimentally observed sizes of the major fragments from Fig. 1 were compared with the calculated sizes of the expected C-terminal fragments (Table 3) using the cleavage sites shown in Table 2, and with those of the complete sequences of bovine  $\alpha 1$  and  $\alpha 2$  chains.

Thus the following conclusions can be drawn from the comparison of expected and observed fragments sizes: (i) there is a good agreement between calculated and observed sizes, supporting the evidence that the 5 identified cleavage sites yielded bands 6 to 10. (ii) The size of fragment 5 is consistent with the location of the cleavage in the  $\alpha 1$  chain (91.3, calculated versus 92 kDa, observed sizes). (iii) The results also indicate that the cleavages occur in pairs at positions 286, 499 and 712 of both  $\alpha 1$  and  $\alpha 2$  chains (Fig. 2). None of the sequenced fragments corresponded to N-terminal fragments of the collagen molecule. Cleavages at position 286 would have produced an N-terminal

fragment of  $\sim 28$  kDa, similar in size to that of the enzyme band and too small to be resolved on 7.5% SDS-PAGE gels. Cleavages at residues 499 and 712 would have produced N-terminal fragments of  $\sim 60$  and  $\sim 84$  kDa ( $\alpha 1$  chain) and  $\sim 58$  and  $\sim 80$  kDa ( $\alpha 2$  chain), respectively. No sequences corresponding to the N-terminal region of collagen were found in any of the fragments in these size ranges. It is concluded that cleavages at these positions must have occurred subsequent to cleavage at residue 286 or other uncharacterised cleavages closer to the N-termini, yielding N-terminal fragments too small to be seen on the gels. It is possible that some of the higher molecular weight fragments corresponding to bands 1 to 4, produced in the first stages of the digestion, correspond to such cleavage products. N-terminal sequences of these high molecular weight fragments 1 to 4 were not determined.

Cleavage of native rat tail collagen by GP2 was also examined (Fig. 3), and the results were very similar to those

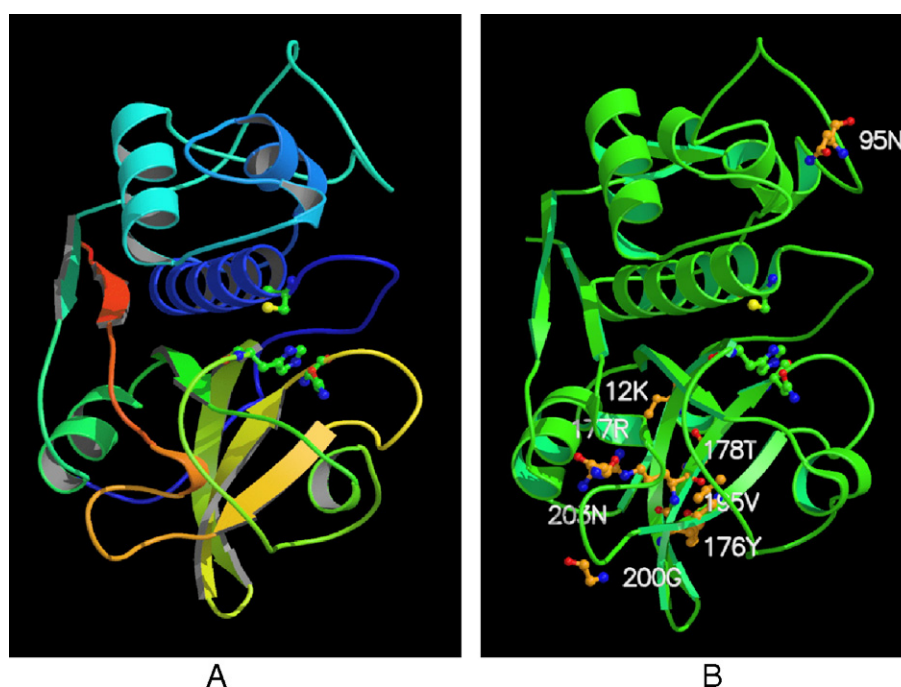


Fig. 5. Modelled structure of GP2 showing the location of the amino acid residues that differ from the published GPII. (A) The 3D structure of GPII [27]. The coordinates were obtained from the protein data bank, code 1CQD. (B) The picture was generated by superimposed structure of GP2 to 3D structure of GPII. Different residues relative to GPII are shown in ball-and-stick representation in orange. Cys27, His161 and Asn181 of active site are shown in CDK colour.

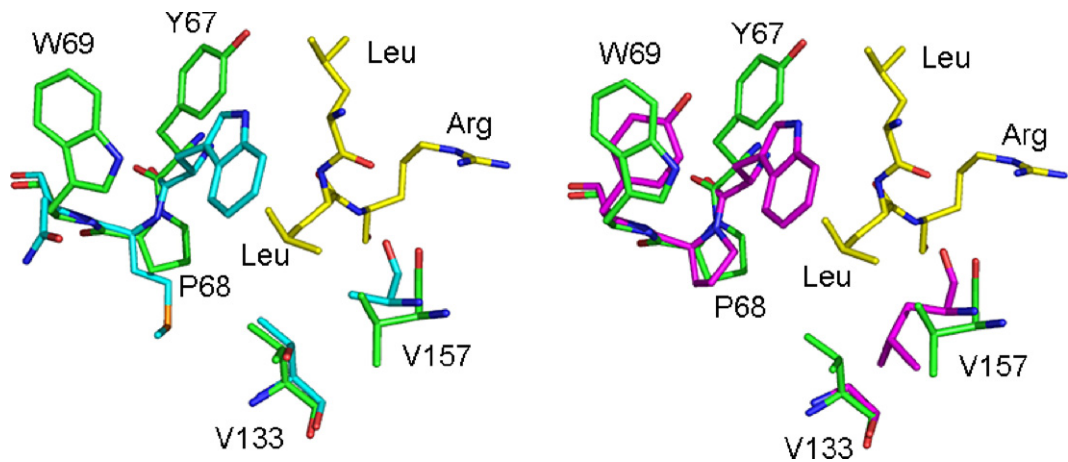


Fig. 6. Comparison of the S2 binding pocket residues for papain and GP2 (left), and papain and GP3 (right) The coordinates for papain were taken from the protein data bank entry 1POP [33]. In this structure papain (green carbon atoms) is complexed with the tripeptide leupeptin (shown in yellow carbon atoms). For all molecules oxygen, nitrogen and sulphur atoms are drawn as red blue and tan sticks respectively. The numbering and naming are based on the papain structure.

obtained for bovine collagen. Compare, for example, lane 3 in Fig. 3 and lane2 in Fig. 4, the 48-h digests using 2.5 U of enzyme. Bands 9 and 10 appear only in the calf skin digest using 6.5 U of enzyme (lane 6, Fig. 3). No comparative digest at the higher enzyme concentration was undertaken for rat tail collagen. Fragments with sizes corresponding to cleavages at residues 499 and 709 in bovine collagen are evident (bands 7–10), as are the major high molecular weight bands 1 to 4. The data are consistent with preferential hydrolysis of two peptide bonds at ~ 50 and ~ 170 residues close to the N-terminus of both

chains based on the size of fragments of bovine collagen digest (Fig. 1) and were comparable with bovine collagen hydrolysis. The activities of ginger proteases (GP2 and GP3) and papain towards native rat tail type I collagen were compared (Fig. 4), using equal concentrations of all three enzymes, based on caseinolytic activity measurement (2.5 U, as for Fig. 3, lane 3). The results for GP2 are as in Fig. 3: essentially complete digestion by GP2 after 2 days to yield four major fragments with sizes corresponding to those of fragments 2, 4, 7 and 8 in Fig. 3. GP3 showed very limited digestion of collagen after 2 days to

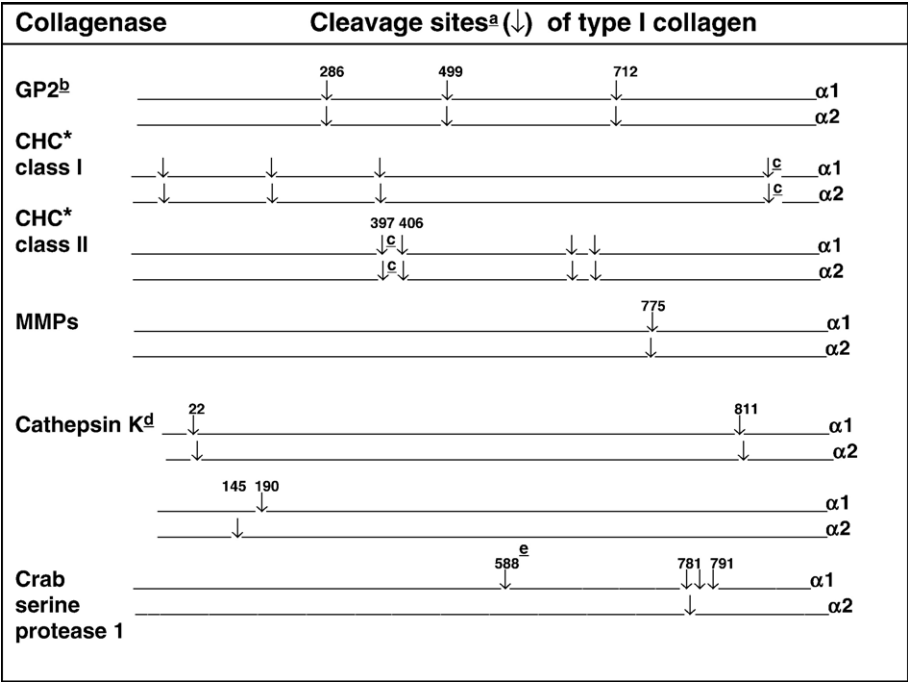


Fig. 7. Comparison of known sites of cleavage of type I collagen by the various collagen-digesting enzymes. \*Clostridium histolyticum collagenase. (a) Cleavage sites identified by N-terminal sequencing are indicated by commencing residue numbers in the triplets of collagen, and cleavage sites are approximately located by the calculation of the sizes of the cleaved fragments. (b) Ginger protease 2 (GP2) of this work. (c) The initial cleavage site [3]. (d) Cleavage sites of bovine (top) [8] and rat (bottom) type I collagen [7]. (e) Secondary cleavage site [32].

Table 5  
Cleavage sites in collagen for collagen-digesting enzymes

| Collagenase      | $\alpha 1$ chain cleavage site |    |    |     |     |                | $\alpha 2$ chain cleavage site |    |    |     |     |     | Source of collagen (Reference) |
|------------------|--------------------------------|----|----|-----|-----|----------------|--------------------------------|----|----|-----|-----|-----|--------------------------------|
|                  | ↓                              |    |    |     |     |                | ↓                              |    |    |     |     |     |                                |
|                  | Sequence of cleavage site      |    |    |     |     |                |                                |    |    |     |     |     |                                |
|                  | P3                             | P2 | P1 | P1' | P2' | P3'            | P3                             | P2 | P1 | P1' | P2' | P3' |                                |
| GP2              | G                              | P  | A  | G   | E   | E              | G                              | P  | S  | G   | E   | E   | Bovine [this paper]            |
|                  | G                              | A  | R  | G   | E   | R              | G                              | A  | R  | G   | E   | R   |                                |
|                  | G                              | P  | S  | G   | N   | A              | G                              | P  | S  | G   | I   | S   |                                |
| CHC <sup>a</sup> | G                              | A  | R  | G   | Q   | A              | G                              | P  | R  | G   | E   | A*  | Rat [3]                        |
| Class II         | G                              | F  | P  | G   | P   | K              | G                              | F  | P  | G   | P   | K   | Human [31]                     |
| MMPs             | P                              | Q  | G  | I   | A   | G              | P                              | Q  | G  | L   | L   | G   | Bovine [8]                     |
| Cathepsin K      | G                              | L  | D  | G   | A   | K              | G                              | L  | K  | G   | P   | Q   | Guinea pig                     |
|                  | G                              | P  | S  | G   | A   | S*             | P                              | P  | G  | A   | R   | G   | Guinea pig                     |
|                  | G                              | P  | Q  | G   | F   | Q              | G                              | P  | Q  | G   | F   | Q   | Guinea pig                     |
| Crab collagenase | G                              | P  | R  | G   | L   | P*             | G                              | P  | R  | G   | P   | P   | Rat [7]                        |
|                  | G                              | P  | Q  | G   | V   | R              |                                |    |    |     |     |     | Bovine [7]                     |
|                  | K                              | P  | G  | K   | S   | G <sup>b</sup> | G                              | A  | R  | G   | S   | D   | Bovine [32]                    |
|                  | A                              | G  | Q  | R   | G   | V              | G                              | F  | L  | G   | L   | P   |                                |
|                  | G                              | Q  | R  | G   | V   | V <sup>c</sup> |                                |    |    |     |     |     |                                |
|                  | G                              | E  | R  | G   | F   | P*             |                                |    |    |     |     |     |                                |
|                  | R                              | G  | L  | T   | G   | P <sup>d</sup> |                                |    |    |     |     |     |                                |

<sup>a</sup>*Clostridium histolyticum* collagenase. <sup>b</sup>Type II collagen. All other data refer to type I collagen. <sup>c</sup>Major primary cleavage. <sup>d</sup>Secondary cleavage. Asterisks (\*) represent the sequences which appear twice in collagen.

fragments similar in size to bands 2 and 4 and small amounts of fragments corresponding to bands 5 and 6. Incubation with papain shows no significant digestion over the same period.

## 2.2. Activity towards peptide substrates

The ability of GP2, GP3 and papain to hydrolyse three fluorescent proline-containing substrates which correspond to known cleavage sites for mammalian matrix metalloproteinase (MMP) was examined. The activities of each enzyme were normalised based on the caseinolytic activity in the standard assay for the comparison. Papain showed lower relative activity towards all substrates, typically by 10- to 40-fold than GP2 and GP3, respectively (Table 4). Cleaved products were analysed by mass spectrometry at the completion of the reaction to determine the site(s) of cleavage relative to proline (Fig. S3, Supplementary material). Cleavage by both GP2 and GP3 occurred with Pro at P2 and/or P3, but GP2 showed more preference for Pro at P2 than GP3. For papain, cleavage occurred with Pro in P3 only (Table 4).

## 2.3. Structural analysis of ginger proteases

As there is a 96% identity between the GPII [1] and GP2 amino acid sequences, the published structure of ginger protease GPII [27] was used to construct molecular models for GP2 and GP3. GP2 does not appear to exhibit any significant structural differences in the active site when compared with GPII although we cannot rule out small distortions in the overall structure that might impinge upon the active site (Fig. 5). Likewise the backbone structure of GP3 would appear to be essentially indistinguishable from GP2 (sequence alignment is found in Table S4, Supplementary material). The primary specificity of papain family cysteine

proteases is determined by interactions at the S2 site. GPII [27] has been reported to be specific for proline-containing peptides in contrast with the broad specificity of papain [29]. In papain, the amino acid residues that form the S2 pocket are Tyr67, Pro68, Trp69, Val133 and Val157. However, GP2 has Trp, Met, Asn, Thr and Ala and GP3 has Trp, Pro, Tyr, Gly and Leu in the equivalent positions. The enhanced proline specificity of GPII has been attributed to the presence of Trp69 which limits the size of the S2 specificity pocket to about half that of papain [29]. The most significant differences between GP2 and GP3 in the S2 specificity pocket are Met68 → Pro and Thr133 → Gly. The two mutations reshape the pocket making it larger towards the bottom and more compact (and more similar to papain) at the top for GP3, as viewed in Fig. 6. The inclusion of the extra hydroxyl group as part of the Thr side chain is unlikely to add a hydrogen bonding partner for the substrate, since this group can be rotated such that the hydroxyl faces away from the active site. Thus it is likely that the overall hydrophobic nature of the pocket is conserved in all of the enzymes. It is therefore more likely that the differences in specificity are due to orienting effects of the substrate as the shape of the S2 pocket changes between the three enzymes.

## 3. Discussion

The reported efficacy of ginger proteases in meat tenderization and the previously reported specificity of ginger proteases for proline-containing peptide substrates [27,30] led to an investigation of the collagenolytic activity of this group of enzymes. Digestion of type I collagen from calf skin and rat tail by ginger protease GP2 is shown to occur at 22 °C, well below the denaturing temperature of the intact triple helical structure. Multiple discrete cleavage sites were identified within the triple helix of native bovine collagen (Fig. 2). The observed cleavage

in the non-helical cross-linked regions of the molecule, resulting in degradation of  $\beta$  bands on SDS-PAGE gel, is a feature shared by many proteases, and of lesser interest in the present study. This is the first report of a plant protease which hydrolyses native triple helical collagen. The GP2 cleavage pattern is different from those of other known collagenases (Fig. 7).

In order to further compare the hydrolysis of collagen by GP2 and other collagenases, the amino acid sequences surrounding the cleavage site have been compared. From an examination of the six-residue GP2 cleavage sites (P3–P3') (Table 5), the following conclusions were drawn: (i) Gly always occurs in P1' and P3 positions; (ii) 4 cleavage sites have Pro, and 2 have Ala in the P2 position; (iii) none of the sequences surrounding the identified cleavage sites of GP2 corresponded to those of other collagenases; (iv) there is a preference (4/6 cleavages) for Glu in P2'. A consensus sequence G(P/A)X↓GEX thus exists for four of the six cleavage sites. Within the entire type I collagen sequence there are four additional sites with this motif: at residues 55, 175, 649 and 832. Since fragments of 100 kDa or higher were not sequenced in this study, cleavages at positions 55 and 832 would not have been identified. Fragments of approximately the correct size were however observed (~120 kDa and ~110 kDa, respectively). An N-terminal sequence corresponding to the C-terminal fragment from cleavage at residue 175 of the  $\alpha 2$  chains (175GELGP) is discernable in the underlying sequence data for band 6, although the size of the fragment which would have resulted from a single cleavage (96 kDa) is higher than the fragment size seen (81 kDa), suggesting that the cleavage may have occurred subsequent to cleavage at the C-terminal end of the molecule. The expected fragment size for a fragment corresponding to residues 175–832 would be 82 kDa. Neither the N- or C-terminal fragments of a single cleavage at Gly649 were observed. The remaining GP2 cleavage site, at Gly712 (GPS↓GNA in  $\alpha 1$  and GPS?GIS in  $\alpha 2$  chain), is distinguished by the fact that while the sequence GPSG occurs many times throughout the  $\alpha 1$  and  $\alpha 2$  chains. This is the only site at which the sequence occurs in both chains at the same position.

The majority of the other collagenase cleavage sites are unique sequences in type I collagen. Exceptions are two of the cleavage sites of cathepsin K, one each of crab collagenase and *C. histolyticum* collagenase (CHC) class II, which are found twice in corresponding collagen sequences, but only one of the two is cleaved (asterisks in Table 5). The six-residue sequence alone is apparently not sufficient to define the specificity of cleavage in these cases. An alternative explanation is that the stability of the triple helix is lower in the vicinity of the cleavage sites, but there are conflicting views regarding the effect of Pro/HyP content on triple helix stability (whether the stability will increase or decrease when this content increases). Examination of the 12–18 residues around the GP2 cleavage sites did not reveal notable differences from the rest of the sequences in the collagen in this regard. Pro/HyP comprises 20–23% of the ca. 1000 residue triple helical region of collagen. In the 12 residues around the cleavage sites of the  $\alpha 1$  and  $\alpha 2$  chains by GP2, two cleavage sites contain 17–25% Pro/HyP, and the site closest to the carboxyl terminus, 42%.

The significance of this region and its relevance to the stability/flexibility of collagen are not clear.

It was of interest to determine whether collagen cleavage specificity of ginger proteases was consistent with specificity toward synthetic substrates. The results of assays using synthetic peptides containing proline were consistent with the observed preference of GP2 for proline in the P2 position. In contrast, papain showed a preference for Pro in P3, and exhibited much lower efficiency of hydrolysis than GP2, consistent with its inability to hydrolyse native collagen (Fig. 4). Structural modelling of the residues surrounding the S2 site in the three enzymes (Fig. 6) shows that the S2 pocket in the ginger enzymes is only half the size of the papain pocket, which may contribute to the difference in specificity.

In summary, the combined evidence shows that ginger protease GP2 is able to hydrolyse native type I collagen at the same discrete sites in all 3 chains of type I collagen. A second ginger protease, GP3, is less effective in this regard, and papain exhibits no detectable activity under the same conditions. A preference for Pro in the P2 site is seen in the hydrolysis by GP2 of some proline-containing peptide substrates, consistent with the specificity demonstrated in the cleavage of native collagen at positions Gly286 and Gly712 and with the previous report on the specificity of the ginger protease GP2 [27]. Modelling of the structure of GP2 based on the known structure of GP2, and comparison with papain, indicates that the active sites of GP2 and GP2 are essentially identical and more constrained than that of papain. GP2 is thus the only documented example of a plant protease with collagenolytic activity. Whether the collagenolytic activity is related to a biological function in the rhizome, for example in the defence against plant pests such as nematodes, or whether the natural substrate is an unrelated protein remains to be established.

## Acknowledgments

We thank Mr. Chris Wood (UQ, Australia) for the protein sequencing and Dr. George Buttler (UBC, Canada) for the supply of MMP2 and assistance in peptide cleavage assays. Special thanks go to Mr. Len Palmer (Buderim Ginger Co, Australia) for maintaining the field and supplying ginger for 3 years for this project. This work was supported by Strategic Partnerships with Industry-Research and Training (SPIRT) grant of Australian Research Council and Graduate School Research Travel Awards (GSRTA) of University of Queensland.

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at [doi:10.1016/j.bbagen.2007.08.003](https://doi.org/10.1016/j.bbagen.2007.08.003).

## References

- [1] K.H. Choi, R.A. Laursen, Amino-acid sequence and glycan structures of cysteine proteases with proline specificity from ginger rhizome *Zingiber officinale*, Eur. J. Biochem. 267 (2000) 1516–1526.

- [2] I. Mandl, H. Zipper, L.T. Ferguson, *Clostridium histolyticum* collagenase: its purification and properties, *Arch. Biochem. Biophys.* 74 (1958) 465–475.
- [3] M.F. French, A. Bhowan, H.E. Van Wart, Identification of *Clostridium histolyticum* collagenase hyper reactive sites in type I, II, and III collagens: lack of correlation with local triple helical stability, *J. Protein Chem.* 11 (1992) 83–97.
- [4] O. Matsushita, C.M. Jung, J. Minami, S. Katayama, N. Nishi, A. Okabe, A study of the collagen-binding domain of a 116-kDa *Clostridium histolyticum* collagenase, *J. Biol. Chem.* 273 (1998) 3643–3648.
- [5] J.J. Wilson, O. Matsushita, A. Okabe, J. Sakon, A bacterial collagen-binding domain with novel calcium-binding motif controls domain orientation, *EMBO J.* 22 (2003) 1743–1752.
- [6] J.L. Lauer-Fields, D. Jaska, G.B. Fields, Matrix metalloproteinases and collagen catabolism, *Biopolymers* 66 (2002) 19–32.
- [7] W. Kafienah, D. Bromme, D.J. Buttle, L.J. Croucher, A.P. Hollander, Human cathepsin K cleaves native type I and II collagens at the N-terminal end of the triple helix, *Biochem. J.* 331 (1998) 727–732.
- [8] P. Garnero, O. Borel, I. Byrjalsen, M. Ferreras, F.H. Drake, M.S. McQueney, N.T. Foged, P.D. Delmas, J.M. Delaisse, The collagenolytic activity of cathepsin K is unique among mammalian proteinases, *J. Biol. Chem.* 273 (1998) 32347–32352.
- [9] C.A. Tsu, J.J. Perona, V. Schellenberger, C.W. Turck, C.S. Craik, The substrate specificity of *Uca pugnator* collagenolytic serine protease 1 correlates with the bovine type I collagen cleavage sites, *J. Biol. Chem.* 269 (1994) 19565–19572.
- [10] H. Aoki, M.N. Ahsan, K. Matsuo, T. Hagiwara, S. Watabe, Purification and characterization of collagenolytic proteases from the hepatopancreas of northern shrimp (*Pandalus eous*), *J. Agric. Food Chem.* 51 (2003) 777–783.
- [11] S. Bjornsson, D. Heinegard, Isolation and culture techniques of foetal calf chondrocytes, *Biochem. J.* 198 (1981) 141–148.
- [12] D.P. Orgill, P.Y. Liu, L.S. Ritterbush, E.M. Skrabut, J.A. Samuels, S.L. Shames, Debridement of porcine burns with a highly purified, ananain-based cysteine protease preparation, *J. Burn Care Rehabil.* 17 (1996) 311–322.
- [13] T. Mazda, K. Makino, R. Yabe, K. Nakata, K. Fujisawa, H. Ohshima, Use of standardized protease enzymes for antibody screening of blood donor samples with the microplate system AutoAnalyzer, *Transfus. Med.* 5 (1995) 43–50.
- [14] S. Khare, C. Krc, M. Griffiths, H. Luthra, C. David, Oral administration of an immunodominant human collagen peptide modulates collagen-induced arthritis, *J. Immunol.* 155 (1995) 3653–3659.
- [15] R.A. Lawrie, The eating quality of meat, *Meat Science*, Pergamon Press, Oxford, 1991, p. 212.
- [16] D.S. Miyada, A.L. Tappel, The hydrolysis of beef proteins by various proteolytic enzymes, *Food Res.* 21 (1956) 217–225.
- [17] T.L. Rolan, G.W. Davis, S.C. Seideman, T.L. Wheeler, M.F. Miller, Effects of blade tenderization and proteolytic enzymes on restructured steaks from beef bullock chucks, *J. Food Sci.* 53 (1988) 1062–1064.
- [18] Y.B. Lee, D.J. Sehnert, C.R. Ashmore, Tenderization of meat with ginger rhizome protease, *J. Food Sci.* 51 (1986) 1558–1559.
- [19] A. Hashimoto, Y. Takeuti, Y. Kawahara, K. Yasumoto, Proteinase and collagenase activities in ginger rhizome, *J. Jpn. Soc. Nutr. Food Sci.* 44 (1991) 127–132.
- [20] E.H. Thompson, I.D. Wolf, C.E. Allen, Ginger rhizome: a new source of proteolytic enzyme, *J. Food Sci.* 38 (1973) 652–655.
- [21] B.M. Naveena, S.K. Mendiratta, Tenderisation of spent hen meat using ginger extract, *Br. Poult. Sci.* 42 (2001) 344–349.
- [22] U.K. Laemmli, Cleavage of structural proteins during the assembly of the head of bacteriophage T4, *Nature* 227 (1970) 680–685.
- [23] R.T. Aimes, J.P. Quigley, Matrix metalloproteinase-2 is an interstitial collagenase. Inhibitor-free enzyme catalyzes the cleavage of collagen fibrils and soluble native type I collagen generating the specific 3/4- and 1/4-length fragments, *J. Biol. Chem.* 270 (1995) 5872–5876.
- [24] C.G. Knight, F. Willenbrock, G. Murphy, A novel coumarin-labelled peptide for sensitive continuous assays of the matrix metalloproteinases, *FEBS Lett.* 296 (1992) 263–266.
- [25] G. Murphy, Q. Nguyen, M.I. Cockett, S.J. Atkinson, J.A. Allan, C.G. Knight, F. Willenbrock, A.J. Docherty, Assessment of the role of the fibronectin-like domain of gelatinase A by analysis of a deletion mutant, *J. Biol. Chem.* 269 (1994) 6632–6636.
- [26] G. Oudenakker, G. Froyen, P. Fiten, P. Proost, J. Vandamme, Human monocyte chemotactic protein-3 (MCP-3): molecular cloning of the cDNA and comparison with other chemokines, *Biochem. Biophys. Res. Commun.* 191 (1993) 535–542.
- [27] K.H. Choi, R.A. Laursen, K.N. Allen, The 2.1 Å structure of a cysteine protease with proline specificity from ginger rhizome, *Zingiber officinale*, *Biochemistry* 38 (1999) 11624–11633.
- [28] T.A. Jones, J.Y. Zou, S.W. Cowan, M. Kjeldgaard, Improved methods for building protein models in electron density maps and the location of errors in these models, *Acta Crystallogr., A* 47 (Pt 2) (1991) 110–119.
- [29] J.L. Harris, B.J. Backes, F. Leonetti, S. Mahrus, J.A. Ellman, C.S. Craik, Rapid and general profiling of protease specificity by using combinatorial fluorogenic substrate libraries, *Proc. Natl. Acad. Sci. U. S. A.* 97 (2000) 7754–7759.
- [30] K.H. Choi, PhD Dissertation in Chemistry, 1998 Boston University.
- [31] H. Nagase, A.J. Barrett, J.F. Woessner Jr., Nomenclature and glossary of the matrix metalloproteinases, *Matrix, Suppl.* 1 (1992) 421–424.
- [32] K. Kishi, K. Yamazaki, I. Yasuda, N. Yahagi, M. Ichinose, Y. Tsuchiya, S.B. Athauda, H. Inoue, K. Takahashi, Characterization of a membrane-bound arginine-specific serine protease from rat intestinal mucosa, *J. Biochem. (Tokyo)* 130 (2001) 425–430.
- [33] E. Schroder, C. Phillips, E. Garman, K. Harlos, C. Crawford, X-ray crystallographic structure of a papain–leupeptin complex, *FEBS Lett.* 315 (1993) 38–42.

## Supplementary material - Kim

### Contents

1. cDNA cloning of ginger proteases
  - 1.1 Plant material and isolation of total RNA
  - 1.2 cDNA cloning strategy (Tables S1 - S3)
2. Purification of ginger proteases, GP2 and GP3. (Figures S1 – S2)
3. Mass spectrometry analysis of cleaved products of synthetic substrates by ginger proteases and papain (Figure S3)
4. Appendices: Alignments of GP2, GP3  $\alpha$ 1 and  $\alpha$ 2 sequences of bovine, human, and rat collagen triple helical regions

### 1. cDNA cloning of ginger proteases

#### 1.1 Plant material and isolation of total RNA

Ginger rhizomes were harvested from the field, stored at ambient temperature and processed within 1 day. The tissue was disrupted by slicing the rhizomes and grinding them to a fine powder under liquid nitrogen. Total RNA was extracted from 0.1 g of tissue using 1.5 ml of phenol and guanidine isothiocyanate (TRIzol Reagent, Gibco-BRL) according to the manufacturer's instructions. Total RNA was affinity purified using RNeasy Mini Spin Column (QIAGEN), eluted in 100  $\mu$ l RNase free (DEPC treated) water and stored at -70 °C.

#### 1.2 cDNA cloning strategy

Cloning of cDNA sequences corresponding to the published amino acid sequences of ginger protease I and II (GPI and GPII) [1] was a four step procedure: (i) Three sets of degenerate primers derived from the conserved amino acid sequences common to GPI and GPII, or selected GPII- specific sequences, were used to amplify internal cDNA fragments; (ii) 5' RACE PCR was used to amplify the upstream end of the cDNA using a non-degenerate primer based on the sequence of the initial amplified product; (iii) 3' RACE was performed in an analogous manner to 5' RACE to amplify the downstream end of the cDNA; and (iv) non-degenerate primers based on terminal sequences of the 3' and 5' RACE clones were used to amplify the full length sequence.

#### *Amplification of ginger protease gene fragments using degenerate primers*

Degenerate primers were designed that corresponded to the conserved amino acid sequences CGSCWAF (primers 1 and 2) and WGESGYI (primers 3 and 4) of GP I and GP II [1]. First-strand cDNA was synthesized from total RNA using 3' anchored oligo(dT) in the Superscript II RT kit (Life Technology). A ~500bp PCR product amplified by the degenerate primers was directly cloned into T-

tailed pBluscript SK(+) (Stratagene) using E.coli DH5 $\alpha$  cells as a host. Purified plasmid cDNA clones were sequenced by the ABI Prism Big Dye Terminator Cycle Sequencing method. Because a PCR fragment corresponding to the anticipated GPII sequence was not initially identified, a set of degenerate primers specific to the GPII sequences WMNPAFQ (primer 5) and NIENPDG (primer 6) was designed to amplify the sequence (Table S1).

#### *5' and 3' RACE*

Gene-specific primers (7 and 8) corresponding to the cloned gp2 and gp3 fragments were used to prime first-strand cDNA, using the SMART RACE cDNA Amplification Kit (CLONTECH). This kit provides a 5' template extension to facilitate second-strand synthesis. For 3' RACE, first strand cDNA was synthesized using the 3' anchored oligo(dT) primer. Non-degenerate sequence specific sense primers (9 and 10) were used together with the oligo(dT) for PCR amplification of the 3' cDNA fragments (Table S2).

#### *Full length cDNA cloning*

The hypothetical full length sequences were assembled from central and terminal sequences generated as described above. To confirm the accuracy of these sequences, full-length gp2 and gp3 cDNA clones were then amplified by PCR from total rhizome RNA using the 5' and 3' sequence specific primers (Table S3). All the clones were completely sequenced in both orientations. DNA sequencing was performed by ABI Prism Big Dye Terminator Cycle Sequencing method in AGRF (Australian Genomic Research Facility).

## **2. Purification of ginger proteases by ion exchange chromatography (Figures S1 and S2)**

Mature ginger proteases were purified by DEAE-Sephadex column (see Methods and materials in the main text). As shown in Figure S1, the three proteolytic activities coincided well with the first peak of protein, suggesting that the enzyme in this peak was essentially homogenous. The second peak of activity corresponded to a shoulder in the protein profile, indicating that this enzyme was not fully separated from impurities. The purity of representative fractions from chromatography, and supernatant before chromatography, was assessed by SDS-PAGE (Figure S2).

## **3. Mass spectromy analysis of cleaved products of synthetic substrate 1 by ginger proteases and papain (Figure S3)**

# Appendix 1. Alignment of $\alpha 1$ (I) sequences of bovine, human, and rat collagen triple helical regions<sup>1</sup>

```

bovine_      GPMGPGSRGLPGPPGAPGPQGFQGGPEGEPEGASGPMGPRGPPGPPGKNGDDGEAGKP  60
human_      GPMGPGSRGLPGPPGAPGPQGFQGGPEGEPEGASGPMGPRGPPGPPGKNGDDGEAGKP
rat_        GPMGPGSRGLPGPPGAPGPQGFQGGPEGEPEGASGPMGPPGPPGPPGKNGDDGEAGKP
            *****
bovine_      GRPGERGPPGPQGARGLPGTAGLPGMKGHRGFSGLDGAKGDAGPAGPKGEPGSPGENGAP 120
human_      GRPGERGPPGPQGARGLPGTAGLPGMKGHRGFSGLDGAKGDAGPAGPKGEPGSPGENGAP
rat_        GRPGERGPPGPQGARGLPGTAGLPGMKGHRGFSGLDGAKGDTGPAGPKGEPGSPGENGTP
            *****
bovine_      GQMGRGLPGERGRPGAPGPAGARGNDGATGAAGPPGPTGPAGPPGFPGAVGAKGEGGPQ 180
human_      GQMGRGLPGERGRPGAPGPAGARGNDGATGAAGPPGPTGPAGPPGFPGAVGAKGEAGPQ
rat_        GQMGRGLPGERGRPGPPGTAGARGNDGAVGAAGPPGPTGPTGPPGFPGAAGAKGEAGPQ
            *****
bovine_      GPRGSEGPQGVRGEPGPPGPAGAAGPAGNPGADGQPGAKGANGAPGIAGAPGFPARGPS 240
human_      GPRGSEGPQGVRGEPGPPGPAGAAGPAGNPGADGQPGAKGANGAPGIAGAPGFPARGPS
rat_        GARGSEGPQGVRGEPGPPGPAGAAGPAGNPGADGQPGAKGANGAPGIAGAPGFPARGPS
            *
bovine_      GPQGSPGPPGPKNGSGEPGAPGSKGDTGAKGEPGPTGIQGGPPGAGEEGKRGARGEPGPA 300
human_      GPQGGPGPPGPKNGSGEPGAPGSKGDTGAKGEPGPVGQGGPPGAGEEGKRGARGEPGPT
rat_        GPQGSPGAPGPKGTSGEPEGAPGNKGDTGAKGEPGPAGVQGGPPGAGEEGKRGARGEPGPS
            *****
bovine_      GLPGPPGERGGPGSRGFPAGDVAGPKGPAGERGAPGPAGPKGSPGEAGRPGEAGLPGAK 360
human_      GLPGPPGERGGPGSRGFPAGDVAGPKGPAGERGSPGPAGPKGSPGEAGRPGEAGLPGAK
rat_        GLPGPPGERGGPGSRGFPAGDVAGPKGPSGERGSPGPAGPKGSPGEAGRPGEAGLPGAK
            *****
bovine_      GLTGSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGFPGPKGAAGEPGKAGER 420
human_      GLTGSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGFPGPKGAAGEPGKAGER
rat_        GLTGSPGSPGPDGKTGPPGPAGQDGRPGPAGPPGARGQAGVMGFPGPKGTAGEPGKAGER

```

<sup>1</sup> Bovine, human, and rat collagen sequences are from Swiss-Prot/TrEMBL entry Nos.; Q3MHM2, P02452, and P02454, respectively, and [2].



|         |                                                              |      |
|---------|--------------------------------------------------------------|------|
| human_  | GAEGSPGRDGSPPGAKGDRGETGPAGPPGAPGAPGAPGVPAGKSGDRGETGPAGPAGPV  |      |
| rat_    | GAEGSPGRDGAPGAKGDRGETGPAGPPGAPGAPGAPGVPAGKNGDRGETGPAGPAGPI   |      |
|         | *****.*****.*****.                                           |      |
| bovine_ | GPVGARGPAGPQGPRGDKGETGEQDGRGIKGHRGFSGLQGPPGPPGSPGEQGPSGASGPA | 960  |
| human_  | GPAGARGPAGPQGPRGDKGETGEQDGRGIKGHRGFSGLQGPPGPPGSPGEQGPSGASGPA |      |
| rat_    | GPAGARGPAGPQGPRGDKGETGEQDGRGIKGHRGFSGLQGPPGSPGSPGEQGPSGASGPA |      |
|         | **.******.*****.                                             |      |
| bovine_ | GPRGPPGSAGSPGKDGLNGLPGPIGPPGPRGRTGDAGPAGPPGPPGPPGPPGPP       | 1214 |
| human_  | GPRGPPGSAGAPGKDGLNGLPGPIGPPGPRGRTGDAGVGPVPPGPPGPPGPPGPP      |      |
| rat_    | GPRGPPGSAGSPGKDGLNGLPGPIGPPGPRGRTGDSGPAGPPGPPGPPGPPGPP       |      |
|         | *****.******.*.*.******                                      |      |

## Appendix 2. Alignment of $\alpha 2$ (I) sequences of bovine, human, and rat collagen triple helical regions<sup>2</sup>

|        |                                                                      |     |
|--------|----------------------------------------------------------------------|-----|
| Bovine | GPMGLMGRGPPGASGAPGPQGFQGGPPGEPGEGQTGPAGARGPPGPPGKAGEDGHPGKP          | 60  |
| Human  | GPMGLMGRGPPGAAGAPGPQGFQGPAGEPGEGQTGPAGARGPAGPPGKAGEDGHPGKP           |     |
| Rat    | GPMGLMGRGPPGAVGAPGPQGFQGPAGEPGEGQTGPAGSRGAPPPGKAGEDGHPGKP            |     |
|        | ***** . ***** . ***** . ***** . *****                                |     |
| <br>   |                                                                      |     |
| Bovine | GRPGERGVVGPQGARGFPGTGLPGFKGIRGHNGLDGLKGQPGAPGVKGEPGAPGENGTP          | 120 |
| Human  | GRPGERGVVGPQGARGFPGTGLPGFKGIRGHNGLDGLKGQPGAPGVKGEPGAPGENGTP          |     |
| Rat    | GRPGERGVVGPQGARGFPGTGLPGFKGIRGHNGLDGLKGQPGAQGVKGEPGAPGENGTP          |     |
|        | *****                                                                |     |
| <br>   |                                                                      |     |
| Bovine | GQTGARGLPGERGRVGAPGPAGARGSDGSVGPVGPAGPIGSAGPPGFPAGPGKELGPV           | 180 |
| Human  | GQTGARGLPGERGRVGAPGPAGARGSDGSVGPVGPAGPNGSAGPPGFPAGPGKELGAV           |     |
| Rat    | GQAGARGLPGERGRVGAPGPAGARGSDGSVGPVGPAGPIGSAGPPGFPAGPGKELGPV           |     |
|        | ** . ***** ***** . *                                                 |     |
| <br>   |                                                                      |     |
| Bovine | GNPGPAGPAGPRGEVGLPGLSGPVGPPGNPGANGLPGAKGAAGLPGVAGAPGLPGRGIP          | 240 |
| Human  | GNAGPTGAPAGPRGEVGLPGLSGPVGPPGNPGANGLTGAKGAAGLPGVAGAPGLPGRGIP         |     |
| Rat    | GNPGPAGPAGPRGEAGLPGLSGPVGPPGNPGANGLTGAKGATGLPGVAGAPGLPGRGIP          |     |
|        | ** . ** . ***** . ***** . ***** . *****                              |     |
| <br>   |                                                                      |     |
| Bovine | GPVGAAGATGARGLVGEPGPAGSKGESGNKGEPGAVGQPPGPS <u>GEEGKR</u> GSTGEIGPA  | 300 |
| Human  | GPVGAAGATGARGLVGEPGPAGSKGESGNKGEPGSAGPQPPGPS <u>GEEGKR</u> GPNGEAGSA |     |
| Rat    | GPVGAAGATGPRGLVGEPGPAGSKGETGNKGEPGSAGAQQPPGPS <u>GEEGKR</u> SPGEPGSA |     |
|        | ***** . ***** . ***** . * ***** . ** *                               |     |
| <br>   |                                                                      |     |
| Bovine | GPPGPPGLRGSPGSRGLPGADGRAGVMGPPGSRGASGPAGVRGPNGDSGRPGEPGLMGPR         | 360 |
| Human  | GPPGPPGLRGSPGSRGLPGADGRAGVMGPPGSRGASGPAGVRGPNGDAGRPGEPGLMGPR         |     |
| Rat    | GPAGPPGLRGSPGSRGLPGADGRAGVMGPPGNRGSTGPAGVRGPNGDAGRPGEPGLMGPR         |     |
|        | ** . ***** . * . * . ***** . *****                                   |     |
| <br>   |                                                                      |     |
| Bovine | GFPGPSGNIGPAGKEGPVGLPGIDGRPGPIGPAGARGEPGNIGFPGPKGPSGDPGKAGEK         | 420 |
| Human  | GLPGSPGNIGPAGKEGPVGLPGIDGRPGPIGPAGARGEPGNIGFPGPKGPTGDPGKNGDK         |     |

<sup>2</sup> Bovine, human, and rat collagen sequences are from Swiss-Prot/TrEMBL entry Nos.; P02465, P08123, and P02466, respectively, and [2]

Rat GLPGSPGNVGPAGKEGPVGLPGIDGRPGPIGPAGPRGEAGNIGFPGPKGPSGDPGKPGEK  
 \*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*.\*

Bovine GHAGLAGARGAPGPDGNNGAQGPPGLQGVQGGKGEQGPAGPPGFQGLPGPAGTAGEAGKP 480  
 Human GHAGLAGARGAPGPDGNNGAQGPPGQGVQGGKGEQGPDPGPPGFQGLPGSPGAGEVGKP  
 Rat GHPGLAGARGAPGPDGNNGAQGPPGQGVQGGKGEQGPAGPPGFQGLPGPSGTAGEVGKP  
 \*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*.\*.\*.\*

Bovine GERGIPGEFGLPGPAGARGERGPPGESGAAGPTGPIGSRGPSGPPGPDGNKGEPGVVGAP 540  
 Human GERGLHGEFGLPGPAGPRGERGPPGESGAAGPTGPIGSRGPSGPPGPDGNKGEPGVVGAV  
 Rat GERGLPGEFGLPGPAGPRGERGPPGESGAAGPSGPIGIRGPSGAPGPDGNKGEGAVGAP  
 \*\*\*.\*.\*\*\*\*\*.\*.\*\*\*\*\*.\*.\*\*\*\*\*.\*.\*\*\*\*\*.\*.\*

Bovine GTAGPSGSGLPGERGAAGIPGGKGEKGETGLRGDIGSPGRDGARGAPGAIGAPGPAGAN 600  
 Human GTAGPSGSGLPGERGAAGIPGGKGEKGEPLRGEIGNPGRDGARGAHGAVGAPGPAGAT  
 Rat GSAGASGPGGLPGERGAAGIPGGKGEKGETGLRGEIGNPGRDGARGAPGAIGAPGPAGAS  
 \*.\*\*.\*.\*.\*\*\*\*\*.\*\*\*.\*.\*.\*\*\*\*\*.\*.\*.\*\*\*\*\*.\*

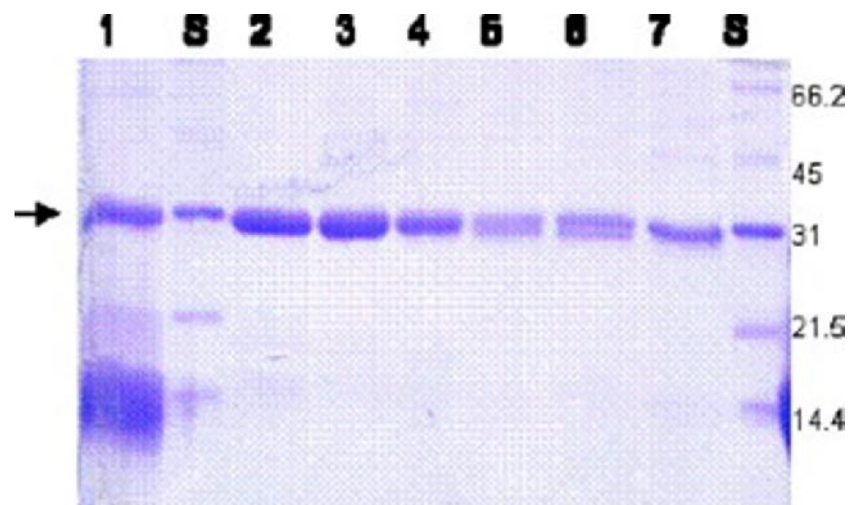
Bovine GDRGEAGPAGPAGPAGPRGSPGERGEVGPAGPNGFAGPAGAAGQPGAKGERGTKGPKGEN 660  
 Human GDRGEAGAAGPAGPAGPRGSPGERGEVGPAGPNGFAGPAGAAGQPGAKGERGGKGPGEN  
 Rat GDRGEAGAAGPSGPAGPRGSPGERGEVGPAGPNGFAGPAGSAGQPGAKGEKGTKGPKGEN  
 \*\*\*\*\*.\*.\*.\*\*\*\*\*.\*\*\*\*\*.\*.\*\*\*\*\*.\*

Bovine GVPVGTGPVGAAGPSGPNPPGAGSRGDGGPPGATGFPGAAGRTGPPGPSGISGPPGPP 720  
 Human GVVGTGPVGAAGPAGPNPPGAGSRGDGGPPGMTGFPGAAGRTGPPGPSGISGPPGPP  
 Rat GIVGTGPVGAAGPSGPNPPGAGSRGDGGPPGMTGFPGAAGRTGPPGPSGITGPPGPP  
 \* \*\*\*\*\*.\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*

Bovine GPAGKEGLRGPRGDQGPVGRSGETGASGPPGFVGEKGPSGEPGTAGPPGTGPQGLLAP 780  
 Human GPAGKEGLRGPRGDQGPVGRTEVGAVGPPGFAGEKGPSGEAGTAGPPGTGPQGLLAP  
 Rat GAAGKEGIRGPRGDQGPVGRTEIGASGPPGFAGEKGPSGEPGTTGPPGTAGPQGLLAP  
 \*.\*\*\*\*\*.\*.\*\*\*\*\*.\*.\* \*\*\*\*\*.\*.\*\*\*\*\*.\*.\*.\*\*\*\*\*.\*

Bovine GFLGLPGSRGERGLPGVAGSVGEPGPLGIAGPPGARGPPGNVGNPGVNGAPGEAGRDGNP 840  
 Human GILGLPGSRGERGLPGVAGAVGEPGPLGIAGPPGARGPPGAVGSPGVNGAPGEAGRDGNP  
 Rat GILGLPGSRGERGQPGIAGALGEPGPLGIAGPPGARGPPGAVGSPGVNGAPGEAGRDGNP  
 \*.\*\*\*\*\*.\*.\*.\*.\*.\*\*\*\*\*.\*.\*.\*\*\*\*\*.\*

[illegible]



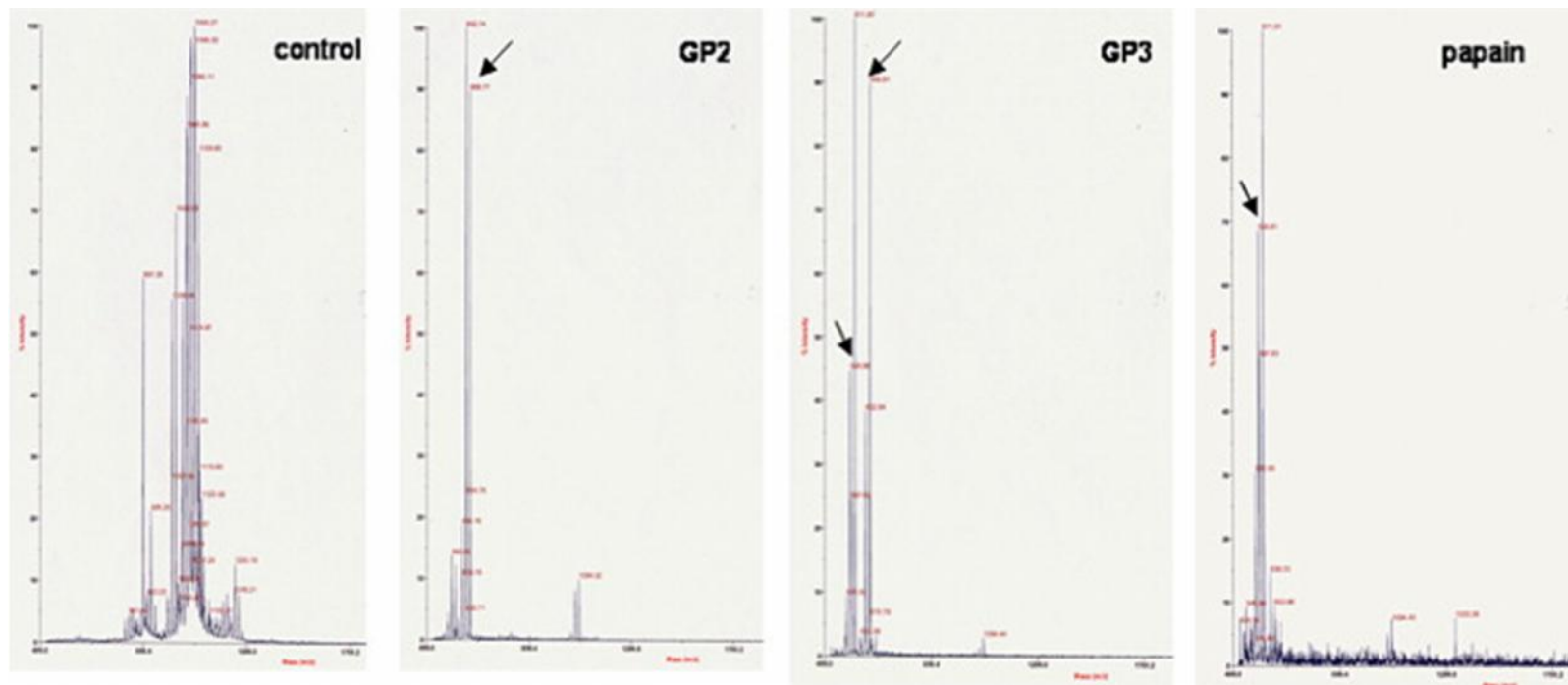
Supplementary Fig. S2

Misook Kim , Susan E. Hamilton , Luke W. Guddat , Christopher M. Overall

**Plant collagenase: Unique collagenolytic activity of cysteine proteases from ginger**

Biochimica et Biophysica Acta (BBA) - General Subjects Volume 1770, Issue 12 2007 1627 - 1635

<http://dx.doi.org/10.1016/j.bbagen.2007.08.003>



Supplementary Fig. S3

Misook Kim , Susan E. Hamilton , Luke W. Guddat , Christopher M. Overall

**Plant collagenase: Unique collagenolytic activity of cysteine proteases from ginger**

Biochimica et Biophysica Acta (BBA) - General Subjects Volume 1770, Issue 12 2007 1627 - 1635

<http://dx.doi.org/10.1016/j.bbagen.2007.08.003>

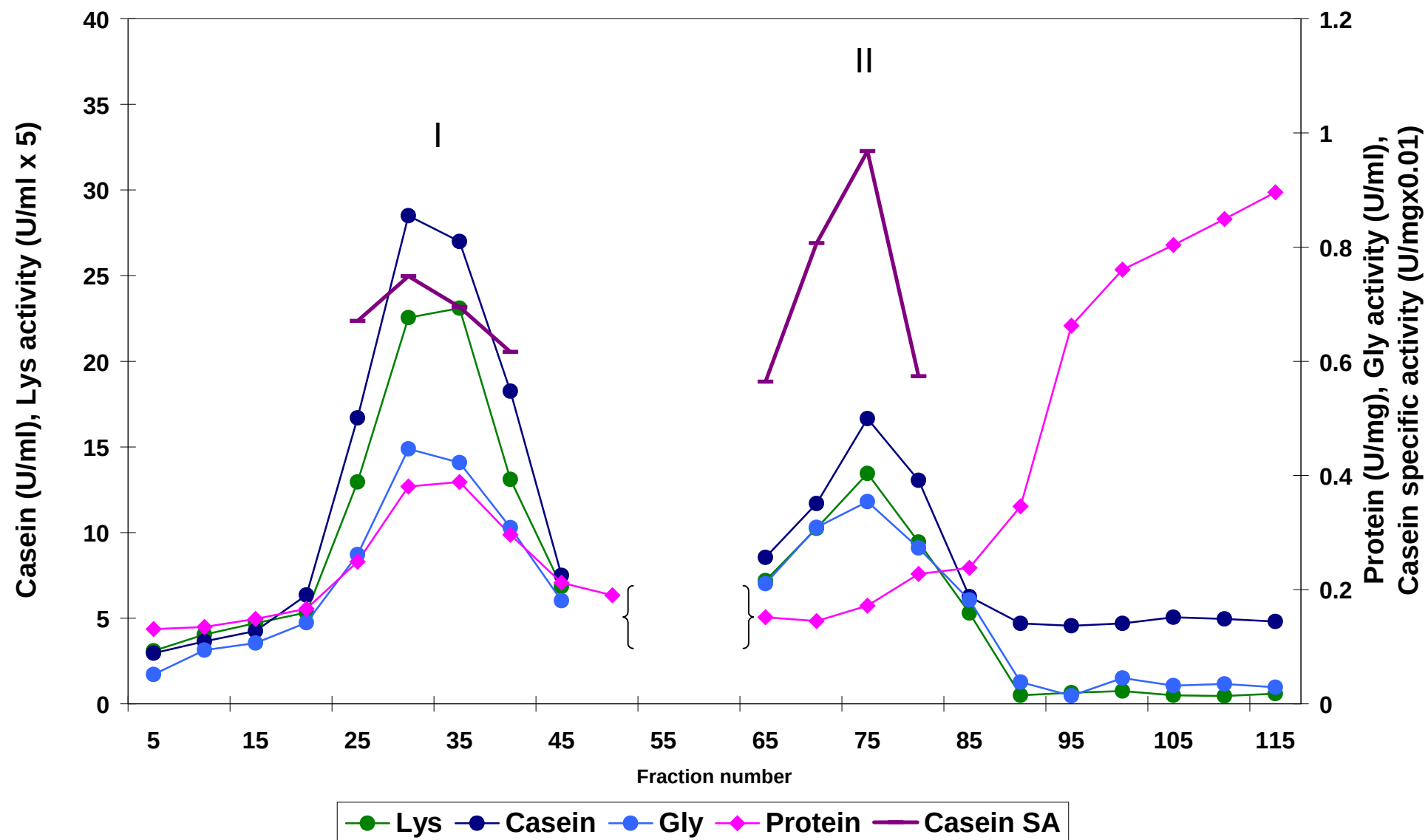


**Table S1. Degenerate primers**

**Table S2. 5' and 3' RACE primers**

**Table S3. Full length cDNA cloning primers**

**Figure S1. Elution of ginger proteases from DEAE-Sephadex column.**

(I) Peak I: fractions 25 – 39, (II) Peak II: fractions 61 – 75. Brackets ( { } ) indicate the fractions accidentally lost in the elution.

**Figure S2. SDS-PAGE Analysis of purification of ginger proteases**

Dialysed proteins before chromatography and fractions of chromatography (see Figure S1) were analysed by SDS-PAGE. Lane S: molecular mass standards (kDa); Lane 1: dialysed precipitate before chromatography; Lanes 2 - 4: fraction nos. 30, 33, and 39 from peak I, respectively; Lanes 5 - 7: fraction nos. 61, 69, and 75 from peak II, respectively.

**Figure S3. Mass spectra of the cleaved products of peptide substrate 1(Mca-PLGL-DpaAR-NH<sub>2</sub>, see Materials and methods in main text) by three enzymes.**

Arrows from right side and from left side in the picture indicate the cleaved products of PL↓GLAR and PLG↓LAR, respectively. Refer to Table 3 in main text.

| Primer         | Sequence                                             |
|----------------|------------------------------------------------------|
| primer 1 and 2 | 5`TGYGGNAGYTG YTG GGCNTT and 5`TGYGGNTCNTGYTG GGCNTT |
| primer 3 and 4 | 5`ATRTANCCRCTYTCNCCCCA and 5`ATRTANCCNGAYTCNCCCCA    |
| primer 5       | 5'TGG ATG AAY CCN GCN TTY CA                         |
| primer 6       | 5'CCR TCN GGR TTY TCD ATR TT                         |
|                | (R=A and G, Y=C and T, N=A, G, C, and T)             |

| RACE    | Primer    | Sequence                               |
|---------|-----------|----------------------------------------|
| 5` RACE | primer 7  | 5'AATCCCATTTTGCCCCCTGTACGG (for gp2)   |
| 5` RACE | primer 8  | 5'CTGTAAATGGCAACGACGACTACTGG (for gp3) |
| 3` RACE | primer 9  | 5'GCTTTGACCGTAGTAGGCTATGGC (for gp2)   |
| 3' RACE | primer 10 | 5'CCAAATCGTCACCGGTGATCTG (for gp3)     |

| Primer              |                    |    |  | Sequence                    |
|---------------------|--------------------|----|--|-----------------------------|
| 5'sequence specific | gp2 forward primer | 11 |  | 5'TCCAGTTTCTCTCTTCTTTCC     |
| 3'sequence specific | gp2 reverse primer | 12 |  | 5'GCGATCGAGTTGCTACTTATAGTAC |
| 5'sequence specific | gp3 forward primer | 13 |  | 5'GACCATCTCGCTTCTCTTCCCT    |
| 3'sequence specific | gp3 reverse primer | 14 |  | 5'GCCGTTGCTTATAGTCGTCTCATC  |