Autocatalytic processing of pro-papaya proteinase IV is prevented by crowding of the active-site cleft

Kenneth C. Baker1, Mark A. J. Taylor2, Nicola J. Cummings, María-Antonieta Tuñón3, Kathryn A. Worboys and Ian F. Connerton

Department of Food Macromolecular Science, Institute of Food Research, Reading Laboratory, Earley Gate, Whiteknights Road, Reading RG6 2EF, UK

1Present address: Institute of Animal Health, Compton, Newbury RG20 7NN, UK
2Present address: Centre for Genetic Engineering and Biotechnology, Cubanacan 6162, Havana, Cuba
3To whom correspondence should be addressed

The DNA coding for pro-papaya proteinase IV (PPIV) has been cloned and expressed in Escherichia coli. Heterologous expression of the protein, followed by refolding in vitro, yields an enzymatically active pro-enzyme which fails to autodigest to form the mature protein. Mutagenesis of the active site of papain to simulate that of PPIV yields a pro-enzyme which also fails to autoactivate. Complementary mutagenesis of the pro-region/mature boundary of PPIV, to introduce its own substrate recognition sequence, has, however, produced a pro-enzyme that will autocatalytically cleave. This is the first report of enzymatic activity in a recombinant pro-cysteine proteinase, and the first time that such a protein has been shown to fail to autocatalytically cleave because of its stringent substrate specificity.

Keywords: cysteine proteinase/enzyme activation/papaya proteinase IV/pro-enzyme activity

Introduction

Cysteine proteinases are a class of proteinases ubiquitous in plant and animal cells. Commercially, they are used as meat tenderizers and as a means of chill-proofing beer. The exemplar of the family is papain (EC 3.4.22.2), extracted from Carica papaya, the pawpaw. Although the structure and catalytic action of papain have been studied extensively, and papain was the second enzyme to be crystallized and have its structure determined by X-ray crystallography, it comprises only ~10% of the proteolytic capacity of crude papaya latex. The other constituents are chymopapain (EC 3.4.22.6), caricain (EC 3.4.22.30, also known as papaya proteinase A, III or omega) and papaya proteinase IV (EC 3.4.22.25, PPIV, glycy1 endopeptidase, chymopapain M; Thomas et al., 1994), which account equally for the remaining proteinase activity of the latex. All except PPIV are general endopeptidases, PPIV alone being a highly specific endopeptidase for the peptide linkage Gly–X (Ritona et al., 1989).

Three of the cysteine proteinases of C. papaya have now been cloned, sequenced and expressed in heterologous hosts (Vernet et al., 1989; Cohen et al., 1990; Taylor et al., 1992; Revell et al., 1993). Comparison of the sequences shows a high degree of homology. The structure of PPIV has been determined recently (O’Hara et al., 1995), and indicates that the substitution of the highly conserved residues Gly23 and Gly65, by Glu and Arg respectively in PPIV, leads to considerable occlusion of the active site, and thus to the observed high specificity first reported by Ritonja et al. (1989). Substitution of the corresponding residues in cathepsin B has been reported to yield an inactive protein, although single-residue substitutions did demonstrate a more pronounced reduction in activity towards Arg–X sites than Gly–X sites compared with the wild-type enzyme (Fox et al., 1995).

All cysteine proteinases studied to date require, for their correct folding, a large pro-region which can comprise up to 30% of the total molecular weight of the pro-enzyme. This pro-region probably serves a dual function, as both a folding template and an intrinsic inhibitor, preventing ectopic activation of the newly synthesized protein. Evidence for both functions in other proteinases is increasing. Zhu et al. (1989) have demonstrated that the pro-region of the serine proteinase subtilisin acts as an essential cofactor in the folding pathway of the enzyme, and will even act in trans to guide the folding of denatured subtilisin when added exogenously in vitro. In a similar fashion, the pro-region of the asparatic acid proteinase, proteinase A, from Saccharomyces cerevisiae, is required for the formation of stable enzyme in vivo, even when the two regions are translated independently in a proteinase-deficient host (van den Hazel et al., 1993). We have also shown that the pro-regions of papain and caricain are required for the refolding in vitro of recombinant enzyme synthesized in Escherichia coli (Taylor et al., 1992; Revell et al., 1993). In addition, the proteinase pro-regions have been shown to be potent inhibitors of the mature enzymes (Fusek et al., 1991; Baker et al., 1992; Fox et al., 1992; Taylor et al., 1995).

Intact pro-enzymes have not been isolated from plant tissues, and it is thought that the enzyme is activated by proteolytic removal of the pro-region during, or just after, translocation into the lacteal vessels in the plant. Details of the proteolytic steps involved in this processing are, however, more obscure. Both papain and caricain from recombinant sources are able to process themselves in an autocatalytic manner in vitro at pH 4.0 (Taylor et al., 1992; Revell et al., 1993). This is some three pH units away from the pH optimum of the enzyme (Vernet et al., 1991). Recently, Mach et al. (1994) have shown that human pro-cathepsin B, produced from a recombinant source, is able to process itself in an autocatalytic step which is concentration independent and therefore unimolecular. However, examination of the protein sequence of pro-PPIV reveals that there are no glycine residues at the junction of the pro-sequence and the mature region, and that it is therefore unlikely that processing is caused by proteolytic cleavage by PPIV itself, either in an intra- or intermolecular process. Here we demonstrate that pro-PPIV is unable to process itself in an autocatalytic manner. Simulation of the active site of PPIV by the site-directed mutagenesis of papain, and mutation of the junction of the pro- and mature regions of pro-PPIV, demonstrates that this is caused solely by the interaction of active-site crowding by residues 23 and 65, and the presence of the
charged and relatively bulky aspartic acid at the C-terminus of the pro-region.

Materials and methods

Materials

(1,3-trans-carboxiran-2-carbonyl)-L-Leu-agmatin (E64), t-butyloxy carbonyl-1,6-alanyl-1,6-alanyl-1-glycine-p-nitroanilide (Boc-Ala-Ala-Gly-pNA) and t-pyroglutamyl-1-phenylalanyl-L-leucine-p-nitroanilide (pyr-Phe-Leu-pNA) were obtained from Bachem (Bubendorf, Switzerland). All other reagents were of the highest quality commercially available.

Cloning of PPIV

Carica papaya leaf RNA was isolated using the method of Sokolowski et al. (1990), and cDNA was produced using the Stratagene ZAP cDNA kit (Stratagene, Cambridge, UK). cDNA was cloned into the unidirectional vector λ UniZapXR to produce a library with an efficiency of 5 × 10^8 p.f.u./µg and amplified to 2 × 10^10 p.f.u./µg. The PPIV sequence was obtained by PCR from this library using an extended SK primer (GGCGCCCGCTCTAGAATCTGATGATC) and a postcoding sequence derived from a partial PPIV clone obtained by McKe et al. (1986). The clone was then sequenced using fluorescently labelled dideoxy terminators and an ABI 373A-Stretch automated sequencer (ABI, Warrington, UK) using a combination of Taq cycle sequencing and T7 sequencing.

Production of mutant sequences

The first round of mutagenesis on papain was used to introduce an enterokinase site at the junction of the mature and pro-regions. Mutations were produced using inverse PCR (Hemsley et al., 1989; Taylor et al., 1995). Mutagenic oligonucleotides were synthesized using a DuPont Coder 300 synthesizer and chemicals from Cruachem (Glasgow, UK). Primers were designed as follows: forward primer, CAA-GATCCCGGAATATGTCATTGGAAGAC; reverse primer, AGTGACAGTGGTAATTCGTGAAGAAGCATA. PCR amplification was performed using 20 ng of template papain (Taylor et al., 1992) cloned into pBlueScriptKS (Stratagene), 500 ng of each of two primers and 25 nmol dNTP (Boehringer Mannheim UK, Lewes, UK) in a total volume of 100 µl PCR buffer II/3 mM MgCl2 (Perkin Elmer, Beaconsfield, UK). Conditions of thermal cycling were 94°C for 1.5 min, then 25 cycles of 1 min at 94°C, 1 min at 55°C and 12 min at 72°C. The resultant PCR product was treated following the protocol of Taylor et al. (1992a) and used to transform E.coli SURE cells (Stratagene) using the method of Hanahan (1983).

Following production of the mutant, the DNA was sequenced using an ABI 373A-Stretch sequencer, employing Taq cycle sequencing to confirm fidelity. Only the resulting PCR product was treated following the protocol of Taylor et al. (1992). Briefly, inclusion body material was solubilized in 100 mM Tris/acetate, pH 8.6 buffer, containing 1 mM EDTA, 10 mM dithiothreitol and 6 M guanidinium hydrochloride. This material was then diluted 10-fold into 0.5 M Tris/acetate, pH 8.6 buffer, with similar concentrations of EDTA and guanidinium hydrochloride and 0.1 M oxidized glutathione, to form mixed disulfides. This material was refolded by 50-fold dilution into 100 mM Tris/acetate, pH 8.6 buffer, containing 0.4 M L-arginine and 3 mM cysteine under anoxic conditions. Refolded protein was concentrated ~500 times using a 10 kDa cut-off NMWCO holofilte filter (Bios-Flo Ltd, Glasgow, UK) and an Amicon stirred cell (Amicon, Stonehouse, UK) with an Omega OM10 membrane (Flowgen, Sittingbourne, UK).

Activation of pro-enzymes

Recombinant pro-cysteine proteinases were routinely activated by incubation at 60°C, pH 4.0, for various time periods in the presence of 20 mM cysteine. Enterokinase (enteropeptidase; Boehringer Mannheim) digestion of pro-papain(Ent)G23E/G65R was performed in 50 mM Tris buffer, pH 8.0, with or without 1 M urea and 10 mM methylamine HCl, and with or without 0.1% Triton X-100 at 37°C with an enzyme:substrate ratio of 1:100, as suggested by the supplier. Native papain and PPIV were purified from papaya latex (M.A.J. Taylor, unpublished results). Enterokinase authenticity was tested using the synthetic substrate H-Gly-Asp-Asp-Asp-Asp-Lys-pNA.

Kinetic determinations

Enzymes were assayed in 100 mM sodium phosphate buffer, pH 7.0, containing 1 mM EDTA, 0.1% brj 35 and 10 mM cysteine at 37°C. Enzymes were individually standardized by titration with E64 using Boc-Ala-Ala-Gly-pNA substrate (Barrett and Kirschke, 1981). K_m and V_max were determined using the same substrate between 0.02 and 1.00 mM. Released
p-nitroaniline was detected at OD\textsubscript{410} in a Perkin Elmer Lambda 15 spectrophotometer thermostatted to 37°C. Kinetic parameters were determined using the Enzfitter program (Leatherbarrow, 1987). The dimethylsulfoxide solvent concentration was kept constant at 2% (v/v).

### N-terminal sequencing

N-terminal sequences of peptides were determined using an ABI 407A sequencer following PAGE and electrophoretic transfer onto Pro-Blot (ABI).

### Results

**Production of pro-PPIV and pro-papain(Ent)G23E/G65R**

Yields of protein were similar to those reported by us previously (Taylor et al., 1992; Revell et al., 1993), routinely yielding 700 mg of inclusion body material. This material was refolded with 3–5% efficiency, as determined by active-site titration with the substrate Boc-Ala–Ala–Gly–pNA following the method of Barrett and Kirschke (1981).

**Activation of pro-enzymes.**

In an effort to ensure that inactive pro-enzymes could be proteolytically deprived of their pro-regions, the first mutation made was to insert an enterokinase cleavage site (DDDDD) into pro-papain by substitution of the C-terminal five residues of the pro-region to form pro-papain(Ent). Wild-type recombinant pro-papain and pro-caricain are not able to autoprocess significantly in vitro at 37°C and pH 7.0 under reducing conditions. They require extended incubation at low pH and elevated temperatures under reducing conditions (Vernet et al., 1989; Taylor et al., 1992; Revell et al., 1993). The effect of the enterokinase cleavage site mutation was to increase the rate of autoprocessing at physiological pH so that the inclusion of enterokinase into the reaction mixture did not noticeably increase the rate of conversion (results not shown). The pro-papain(Ent) sequence was then utilized as a template to produce pro-papain(Ent)G23E/G65R. In contrast to pro-papain(Ent), pro-papain(Ent)G23E/G65R did not autoprocess (as determined by PAGE) under conditions appropriate for wild-type pro-papain or by prolonged incubation with enterokinase, despite the inclusion of 1 M urea and 1 mM methylamine or 0.1% Triton X-100.

It was interesting to note, however, that both pro-papain (Ent)G23E/G65R and pro-PPIV were active against the small peptide substrate Boc–Ala–Ala–Gly–pNA (see Table I).

In contrast, pro-PPIV(D-1G) was rapidly autoactivated under standard conditions. The greatest activity was achieved after 1–2 min, after which activity declined (data not shown), presumably because of autodigestion. An electrophoretic analysis of the processed material revealed three major digestion products (Figure 1). The N-terminal sequence of the largest of these, with a molecular weight of 26 kDa, was SLPE; this was consistent with cleavage at residue −22 (see Figure 2 for PPIV numbering). Band 2 had a molecular weight of 23 kDa and an N-terminal sequence of LP, as would be expected for mature PPIV. The smallest band, band 3, initiated at YSQD, which is consistent with a pro-region remnant initiating at residue −100 and continuing to the end of the pro-region with a molecular weight of 12 kDa. All of the products result from the digestion of Gly–X recognition sites (Figure 2).

### Discussion

Unlike other plant cysteine proteinases, pro-PPIV cannot independently autoactivate in either an inter- or intramolecular process. Previous work has indicated that the restricted substrate specificity, so notable in this enzyme, is caused by encroachment into the active-site cleft of bulky residues at positions 23 and 65 (Thomas et al., 1994; O’Hara et al., 1995; our unpublished data). The mutation of papain, to engineer these same two residues at the same positions, abolished the capacity of this enzyme to autoactivate. This occurred despite the additional insertion of an enterokinase substrate site at the pro-region/mature region boundary. The insertion of this latter
site into otherwise wild-type papain increases the rate of autoactivation many-fold. This enterothiolase mutation may act by preventing the correct folding of the pro-region/mature region boundary and forming a localized disordered region of the molecule which would be more susceptible to proteolysis by the more open active-site cleft of papain. This disordered structure is not, however, accessible to the restricted cleft present in PPFV. To allow this molecule to autoactivate, it was necessary to mutate this region to insert a Gly-X site, the natural substrate of PPFV.

Whilst this mutant pro-PPFV(D1G) was able to autoprocess itself, electrophoretic analysis of the cleavage products revealed that three species were formed during this process. The smallest, band 3, is of a molecular weight and N-terminal sequence consistent with it being a partially degraded remnant of the pro-region. The fact that its staining intensity reduces with time (Figure 1) shows that it is susceptible to further proteolysis at Gly-X sites at the low pH used for autocatalysis. Band 2 is the mature product, as indicated by both the N-terminal sequence and the molecular weight. Band 1 consists of a partially clipped pro-enzyme, digested at position $\sim 22$ (PPFV numbering). These results are interesting considering the work of Vernet et al. (1995), who mutated the pro-region of papain and examined the activation process of this enzyme. Based on these experiments, Vernet et al. (1995) proposed two possible mechanisms of activation of pro-papain. One of these involved the initial digestion of the pro-region at position $\sim 36$ (papain numbering), followed by a secondary digestion event which removed the remnant of the pro-region. Primary digestion of pro-PPFV(D1G) could occur, however, at position $\sim 22$ in a process analogous to that suggested by Vernet et al. (1995). Although pro-PPFV lacked the ability to autoprocess, it did, however, show hydrolytic activity towards the small peptide substrate Boc-Ala-Ala-Gly-pNA, but not against the substrate pyr-Phe-Leu-pNA which would position the relatively bulky leucine residue in the S1 subsite of the enzyme (Schechter and Berger, 1967). Activity against the first, but not the second, of these substrates is also a feature of mature PPFV. Interestingly, this activity is of similar magnitude to that observed for the pro-papain(Ent)G23E/G65R mutant (data not shown). Proteolytic activity by pro-cathepsin L secreted by mouse fibroblasts has been reported previously (Mason et al., 1987), and a precursor form of the aspartyl proteinase cathepsin D, secreted by human breast cancer cells, has also been observed to be active (Capony et al., 1987). The reports of enzyme activity in these pro-proteases occur despite the fact that pro-regions of cysteine and aspartyl proteinases are potent inhibitors of mature enzymes when added in trans (Fuske et al., 1991; Taylor et al., 1994b, 1995).

It seems probable that this lack of inactivation of pro-PPFV and pro-papain(Ent)G23E/G65R by their pro-regions is because of steric occlusion of the active-site cleft by residues E23 and R65, thus preventing close contact of the inhibitory pro-region with the mature enzyme. This allows ingress of the small peptide substrates, but larger peptides or proteins are still excluded.

These observations are interesting in view of the evolutionary distinctions between the pro-regions of this family of enzymes. The papaya cysteine proteinases (papain, chymopapain, carciain and PPFV) have amino acid similarities of between 72 and 86% for their mature sequences. The similarity of the corresponding pro-regions is higher, from 86 to 90%. This is probably because of different selection pressures on the two regions. Significant mutations in the pro-region are likely to result in misfolded or ectopically active enzyme, either of which will be lost during evolution. However, significant mutations in the mature region might result in an enzyme with altered properties, which would still, nonetheless, be a substrate for further evolutionary change. The fact that the pro-regions of these molecules are able to stabilize mature enzymes bearing mutations is, however, graphically demonstrated by the stabilization of papain G23E/G65R by an essentially unchanged pro-region.

Acknowledgements

The DNA sequence for pro-PPFV has been submitted to the EMBL database under accession number X78056. The authors would like to acknowledge the financial support of the BBSRC.

References

Vernet, T., Bert, P.J., de Montigny, C., Musil, R., Tesser, D.C., Menard, R.,

Received December 8, 1995; revised January 31, 1996; accepted February 16, 1996