Molecular cloning of two cysteine proteinases from paw-paw (*Carica papaya*)

Raymond A. McKEE, Susan ADAMS, Jayne A. MATTHEWS,* Christopher J. SMITH and Harry SMITH
Department of Botany, University of Leicester, University Road, Leicester LE1 7RH, U.K.

Two cDNA clones for plant cysteine proteinases have been isolated from a *Carica papaya* (paw-paw, papaya) leaf tissue cDNA library by using a mixture of 16 synthetic oligodeoxyribonucleotides as a hybridization probe. The inserted regions are 311 and 440 base-pairs in length and have the potential to encode a region corresponding to the C-terminal region of two proteins which are homologous with the known plant cysteine proteinases and the mammalian thiol cathepsins. One of the sequences shows a high (> 77%) homology with the plant cysteine proteinase papain, the other is closely related to papaya chymopapain. One sequence contains all, and the other most, of the 3' untranslated region of the mRNA. The inserts were used as specific probes in Northern Blot analyses giving an estimated size for the two mRNA species of 1.45 kilobases.

INTRODUCTION

Cysteine ('sulphydryl' or thiol) proteinases have been isolated from a wide variety of sources, both plant and animal. The enzymes from plants include papain, chymopapain and papaya peptidases A and B from *Carica papaya*; actinidin from *Actinidia chinensis* (kiwi-fruit, Chinese gooseberry); bromelain from *Ananas comosus* (pineapple); ficin from *Ficus carica* (common fig); and asclepains A and B from *Asclepia syriaca* (silk weed). Papain and actinidin have been completely amino-acid-sequenced, and their tertiary structures have been determined (Baker, 1980; Kamphuis *et al.*, 1984) and shown to be essentially identical. Papain and actinidin are about 40% homologous. Partial amino acid sequences of the other cysteine proteinases above have also been determined and found to exhibit varying degrees of homology. Fairly recently the amino acid sequences of the rat liver cathepsins B and H have been determined and shown to exhibit a high degree of homology with the plant protein primary structure, in particular that of papain (Takio *et al.*, 1983). The premise that all of the plant enzymes are descended from a common ancestor has been explored by Lynn (1983), and similar comparisons have been made between mammalian lysosomal cysteine proteinases (Kirschke *et al.*, 1984). In addition to the evolutionary question, disagreement still exists as to the exact number of enzymes, particularly the chymopapains, present in the proteolytic component of papaya latex (Buttle & Barrett, 1984; Brocklehurst *et al.*, 1985a,b; Barrett & Buttle, 1985).

In order to be able to answer these questions we have chosen to investigate the plant cysteine proteinases at the DNA level. We have isolated cDNA clones from a library constructed from papaya leaf-tissue mRNA, and shown these to be related to the papaya proteinases papain and chymopapain. In addition, by hybridization of these clones, we have been able to detect specific mRNA isolated from papaya leaf and fruit tissue, and estimated them to be about 1.45 kb in length. These mRNA species have the capacity to code for proteins of up to M_r 35000.

Furthermore, we have detected RNA fractions in total RNA isolated from papaya leaf and fruit tissue which are approx. 3.35 kb in length. These are taken to represent unprocessed mRNA precursors. Through the use of these clones it will be possible to characterize fully the cDNA and genomic sequences coding for the whole range of papaya cysteine proteinases. The extension of this work to other plant and animal cysteine proteinases will provide a means of determining the evolutionary relationships between this family of enzymes.

MATERIALS AND METHODS

AMV reverse transcriptase was obtained from Life Sciences; Klenow polymerase and phage-T₄ polynucleotide kinase were obtained from BRL; streptomycin, ampicillin and ribonuclease H were from Sigma; agarose (LE type) was from Seakem. DNA ligase was from NBL, and vectors pUC 8 and bacteriophages M13 mp18 and mp19 were from BRL. Nitrocellulose filters were obtained from Millipore or Schleicher and Schuell.

A mixed synthetic oligonucleotide with the sequence

corresponding to the protein sequence Trp-Gly-Glu-Asn-Gly (amino acids 181–185 of papain) was synthesized and purified by Celltech. *Carica papaya* seeds were obtained from paw-paw fruit bought in a local shop or from Chiltern Seeds, Bortree Stile, Ulverston, Cumbria, U.K.

Growth of papaya plants

The outer gelatinous seed coat was removed and the seeds were soaked in many changes of distilled water at 22 °C for 2–3 days. They were then washed under running water for a further 2–3 days, sown in potting compost, and allowed to germinate at 25 °C in the dark. The seedlings were potted up and transferred to a controlled-environment growth chamber. The growing condi-

Abbreviations used: kb, kilobase; AMV, avian myeloblastosis virus; BSA, bovine serum albumin; LB, Luria broth; X-Gal, 5-bromo-4-chloroindol-3-yl β -D-galactopyranoside; bp, base-pair.

* Present address: Department of Clinical Chemistry, Wolfson Research Laboratories, Queen Elizabeth Hospital, Edgbaston, Birmingham 15, U.K.

tions were as follows: 25 °C day and night; 12 h day/12 h night; 100% relative humidity; irradiance 500 $\mu\text{mol} \cdot \text{s}^{-1} \cdot \text{m}^{-2}$ (400–700 nm).

cDNA synthesis

Whole RNA was prepared from young fully expanded leaves of *C. papaya* by the method of Noyes *et al.* (1979). Poly(A)-containing RNA was isolated on a column of oligo(dT)-cellulose by the method of Bantle *et al.* (1976).

First-strand cDNA was synthesized from poly(A) mRNA (13–26 $\mu\text{g} \cdot \text{ml}^{-1}$; max. vol. 30 μl) in a solution containing 50 mM-KCl, 50 mM-Tris/HCl, pH 8.3, 10 mM-MgCl₂, 10 mM-dithiothreitol, 4 mM-spermidine, BSA (100 $\mu\text{g} \cdot \text{ml}^{-1}$), 1.3 mM each of the four deoxyribonucleotides, oligo(dT)_{12–18} (100 $\mu\text{g} \cdot \text{ml}^{-1}$) and AMV reverse transcriptase (500 units $\cdot \text{ml}^{-1}$). This reaction mixture was incubated at 42 °C for 1–2 h. Recovery was by the method of Okayama & Berg (1982).

Second-strand cDNA was synthesized by using the method of Gubler & Hoffman (1983); up to 400 ng of single-stranded cDNA was processed in a reaction volume of 80 μl containing 20 mM-Tris/HCl, pH 7.5, 5 mM-MgCl₂, 10 mM-(NH₄)₂SO₄, 100 mM-KCl, BSA (50 $\mu\text{g} \cdot \text{ml}^{-1}$), 40 μM -dNTPs, *Escherichia coli* ribonuclease H (8.5 units $\cdot \mu\text{l}^{-1}$) and DNA polymerase I (230 units $\cdot \text{ml}^{-1}$). The reaction mixture was incubated at 12 °C for 60 min and at 22 °C for 60 min. EDTA at 20 mM was added to stop the reaction. The double-stranded cDNA products were extracted twice with phenol and ethanol-precipitated twice from 2 M-ammonium acetate.

C-tailing

Double-stranded cDNA (85 ng) was C-tailed by the method of Okayama & Berg (1982), incorporating dideoxy-CTP at 0.5 μM in the reaction mix. The average length of C-tails was 22. The C-tailed double-stranded cDNA was ligated into pUC 8 plasmid DNA (Messing & Vieira, 1982) that had been digested with endonuclease *Pst*I, and tailed with an average of 23 guanosine residues. (Pritchard & Holland, 1985).

Transformation

The ligated DNA was used to transform *E. coli* JM83 *recA*, by the method of Mandel & Higa (1970). Transformants were recovered and incubated at 37 °C for 70 min in LB medium (Maniatis *et al.*, 1982). Cells were recovered by centrifugation and resuspended in LB medium.

Growth and screening of recombinants

Recombinants were grown on LB agar plates containing streptomycin (100 $\mu\text{g} \cdot \text{ml}^{-1}$), ampicillin (50 $\mu\text{g} \cdot \text{ml}^{-1}$) and X-Gal (40 $\mu\text{g} \cdot \text{ml}^{-1}$) at 37 °C overnight. β -Galactosidase-negative (white) colonies were transferred to sterile nitrocellulose filter papers, grown overnight at 37 °C and processed as described by Grunstein & Hogness (1975). The mixed oligonucleotide probes were labelled by using phage-T4 polynucleotide kinase and [γ -³²P]ATP (sp. radioactivity approx. 3000 Ci/mmol). The labelled probe mixture was purified by centrifugation through small columns of Sephadex G-50 equilibrated in 10 mM-Tris/HCl, pH 7.5, containing 1 mM-EDTA. Filters were prehybridized and hybridized at 33 °C and washed at 40 °C by the method of Singer-Sam *et al.* (1983).

Analysis of putative recombinants

Colonies that hybridized with the oligonucleotide probe were purified by re-streaking on the same medium (LB, streptomycin, ampicillin, X-Gal), transferred to nitrocellulose filters and rehybridized. DNA was prepared initially from 1.5 ml small-scale plasmid preparations by the alkaline-SDS method of Birnboim & Doly (1979), and subsequently as large-scale plasmid preparations (500 ml). Recombinant DNA was then characterized by restriction-enzyme analysis. The products of restriction-endonuclease digestion were either immediately ligated into bacteriophage M13 mp18 and M13 mp19 or ligated into the bacteriophage after separation by electrophoresis on 0.7–1.0% agarose gel and subsequent recovery of the restriction fragments on to DE81 paper by using the method of Dretzen *et al.* (1981).

DNA sequencing of the resultant M13 clones was carried out by the dideoxy-chain-termination method of Sanger *et al.* (1977) and Biggin *et al.* (1983). The two sequences were completely determined in both orientations. The sequence data was compiled by hand, and amino-acid-sequence comparisons with known protein sequences were made by direct comparison by eye.

Northern hybridization analysis

Samples of poly(A) RNA (1 μg) and total RNA (10 μg) were allowed to react at 55 °C for 15 min in a solution containing 50% (v/v) formamide, 2.15 M-formaldehyde, 20 mM-Mops, pH 7.0, 5 mM-sodium acetate and 0.5 mM-EDTA, pH 8.0. Samples were subjected to electrophoresis in a 1%-agarose gel by the method of Maniatis *et al.* (1982). The RNA was transferred to nitrocellulose filters and hybridized with a single-stranded [α -³²P]dCTP-labelled probe cut from the cDNA inserts and containing the cDNA sequence. Hybridization was carried out overnight at 42 °C in a solution containing 40% (v/v) formamide, 10% (w/v) dextran sulphate, 20 mM-sodium phosphate, pH 7.5, denatured salmon sperm DNA (20 $\mu\text{g} \cdot \text{ml}^{-1}$), 0.8 $\times$ Denhardtts solution (0.2% Ficoll/0.2% BSA/0.2% polyvinylpyrrolidone) and 6 $\times$ SSC (0.09 M-trisodium citrate/0.9 M-NaCl). The blot was then washed sequentially at 50 °C in (a) 1 $\times$ SSC/0.1% SDS for 15 min, and (b) two changes of 0.1 $\times$ SSC/0.1% SDS for 15 min, and blotted dry and dried at 37 °C.

RESULTS AND DISCUSSION

Isolation and sequences of cDNA clones

The cDNA libraries were constructed from poly(A) RNA isolated from young fully expanded leaves of *Carica papaya* plants grown under standard conditions. The libraries contained approx. 10⁴ colonies/ μg of double-stranded cDNA insert, of average size 800 bp. By using the mixed synthetic oligonucleotide probe for the papain protein sequence Trp-Gly-Glu-Asn-Gly, which corresponds to residues 181–185, four positively hybridizing clones were identified from a total of 557 recombinants screened.

The DNA sequences obtained were translated in all six possible reading frames, and the derived protein sequences screened against the known amino acid sequences of plant cysteine proteinases and mammalian cathepsins (Goto *et al.*, 1976, 1980; Mitchell *et al.*, 1970; Carne & Moore, 1978; Takio *et al.*, 1983; Ohno *et al.*, 1984; Gay & Walker, 1985; San Segundo *et al.*, 1985;

Thr ACC	Lys AAG	Val GTA	Asp GAT	His CAT	Ala GCT	Val GTA	Thr ACA	Ala GCA	Val GTT	Gly GGA	Tyr TAT	Gly GGA	Lys AAA	Ser AGT	Gly GGA		
10			20			30			40								
Gly GGA	Lys AAA	Gly GGT	Tyr TAT	Ile ATA	Leu CTC	Ile ATA	Lys AAG	Asn AAT	Ser TCA	Trp TGG	Gly GGT	Pro CCA	Gly GGA	Trp TGG	Gly GGT		
50			60			70			80			90					
<u>Glu GAG</u>			<u>Asn AAT</u>			<u>Gly GGG</u>			<u>Tyr TAT</u>			<u>Ile ATA</u>			<u>Arg AGA</u>		
100			110			120			130			140					
Val GTG	Cys TGT	Gly GGA	Val GTT	Tyr TAT	Arg AGA	Ser AGC	Ser TCA	Tyr TAC	Tyr TAT	Pro CCT	Ile ATT	Lys AAA	Asn AAT	TGA			
150			160			170			180								
GATAATGGGC			TTTTATACTA			TCACTTACAT			ACATATATAT			ATATGTGTGT			GTGTGTGTCA		
190			200			210			220			230			240		
TATCTGTGTG			TTTACTATAA			TAAAATCGTT			GGATTAATAA			TTAAAAAAA			AA		
260			270			280			290			300			310		

Fig. 1. DNA sequence, and derived protein sequence, of papaya cDNA insert in plasmid pLBPC18

The sequence hybridizing the oligonucleotide probe is underlined, the 'active-site' histidine residue is marked (*), and the putative poly(A) addition signal is boxed. Restriction-endonuclease sites useful in the sequencing of the insert are shown.

K. Lynn, D. Watson & M. Yaguchi, personal communication).

Two clones, pLBPC13 and pLBPC18, were completely DNA-sequenced (Figs. 1 and 2) (pLBPC = plasmid Leicester Botany Papaya clone). The other isolated clones were partially DNA sequenced and shown to be identical with pLBPC13. All of the clones are related to the cysteine proteinases papain, actinidin, bromelain, cathepsin B and H, a bovine cathepsin cDNA (Gay & Walker, 1985) and a Ca^{2+} -dependent proteinase (Fig. 3).

The amino acid coding sequence of the pLBPC18 insert is of particular interest, as the overall homology, when comparing this sequence with that of papain, is 77%. The only gross mismatching occurs at a four-residue insert in the cDNA sequence corresponding to position 167 of papain. Of the 13 amino acids which do not show an exact match with the papain sequence, 77% could be shown to match with a single base change at the DNA level. The amino acid sequence of clone pLBPC13 is clearly less related to papain, although it has a number of the amino acid residues which appear to be conserved throughout the cysteine proteinases. This clone is, however, related more closely to the amino acid sequence of chymopapain A (K. Lynn, D. Watson & M. Yaguchi, personal communication), and it may be one form of chymopapain B. Neither of the two clones extend to the active-site

cysteine residue found in papain at amino acid residue position 25 (Drenth *et al.*, 1971).

The derived amino acid sequence of the 'papain-like' clone, pLBPC18, contains the sequence Trp-Gly-Glu-Asn-Gly encoded by the DNA sequence TGG GGT GAG AAT GGG. This sequence matches the mixed oligonucleotide probe in all 14 consecutive positions. The derived amino acid sequence of the 'chymopapain-like' clone, pLBPC13, contains the sequence Trp-Gly-Glu-Asn-Gly encoded by the DNA sequence TGG GGA GAG AAC GGA. This sequence also matches the mixed oligonucleotide probe in all 14 consecutive positions. Interestingly the clone pLBPC13 has, immediately before the above amino acid sequence, the additional sequence Trp-Gly-Glu-Asn encoded by the DNA sequence TGG GGA GAA AAC. This sequence matches the probe in 12 consecutive positions. The two perfectly hybridizing duplexes, under our conditions, would have a 'melting' temperature of 44 °C for pLBPC13 and of 46 °C for pLBPC18.

The remainder of both sequences is made up of the 3' untranslated regions of the mRNAs. In clone pLBPC18, the sequence AATAAA (Fig. 2) is found at position 278, and corresponds to the signal for poly(A) addition found in the RNA sequence UUAUUU (Proudfoot & Brownlee, 1976). This sequence is followed 19 bases

Gly GGT	Pro CCC	Leu CTG	Ala GCG	Val GTG	Ala GCC	Ile ATC	Asn AAT	Ala GCA	Ala GCA	Tyr TAC	Met ATG	Gln CAG	Thr ACA	Tyr TAT	Ile ATT		
10			20			30			40			50					
Gly GGA	Gly GGA	Val GTG	Ser TCA	Cys TGC	Pro CCG	Tyr TAT	Ile ATA	Cys TGC	Ser TCA	Arg AGG	Arg CGC	Leu CTC	Asn AAC	His CAT	Gly GGT		
60		70		80		90		100		110		120		130			
Val GTG	Leu TTG	Leu TTG	Val GTG	Gly GGA	Tyr TAT	Gly GGT	Ser TCA	Ala GCT	Gly GGC	Tyr TAC	Ala GCG	Pro CCC	Ile ATC	Arg CGA	Leu CTG		
140			150			160			170			180			190		
Lys AAG	Glu GAG	Lys AAG	Pro CCA	Tyr TAC	Trp TGG	Val GTC	Ile ATA	Lys AAG	Asn AAT	Ser TCA	Trp TGG	Gly GGA	Glu GAA	Asn AAC			
200			210			220			230			240			250		
Trp TGG	Gly GGA	Glu GAG	Asn AAC	Gly GGA	Tyr TAT	Tyr TAC	Lys AAG	Ile ATC	Cys TGC	Arg AGG	Gly GGT	Arg CGC	Asn AAT	Ile ATT	Cys TGT		
260			270			280			290			300			310		
320			330			340			350			360			370		
380			390			400			410			420			430		
440			450			460			470			480			490		
500			510			520			530			540			550		
560			570			580			590			600			610		
620			630			640			650			660			670		
680			690			700			710			720			730		
740			750			760			770			780			790		
800			810			820			830			840			850		
860			870			880			890			900			910		
920			930			940			950			960			970		
980			990			1000			1010			1020			1030		
1040			1050			1060			1070			1080			1090		
1100			1110			1120			1130			1140			1150		
1160			1170			1180			1190			1200			1210		
1220			1230			1240			1250			1260			1270		
1280			1290			1300			1310			1320			1330		
1340			1350			1360			1370			1380			1390		
1400			1410			1420			1430			1440			1450		
1460			1470			1480			1490			1500			1510		
1520			1530			1540			1550			1560			1570		
1580			1590			1600			1610			1620			1630		
1640			1650			1660			1670			1680			1690		
1700			1710			1720			1730			1740			1750		
1760			1770			1780			1790			1800			1810		
1820			1830			1840			1850			1860			1870		
1880			1890			1900			1910			1920			1930		
1940			1950			1960			1970			1980			1990		
2000			2010			2020			2030			2040			2050		
2060			2070			2080			2090			2100			2110		
2120			2130			2140			2150			2160			2170		
2180			2190			2200			2210			2220			2230		
2240			2250			2260			2270			2280			2290		
2300			2310			2320			2330			2340			2350		
2360			2370			2380			2390			2400			2410		
2420			2430			2440			2450			2460			2470		
2480			2490			2500			2510			2520			2530		
2540			2550			2560			2570			2580			2590		
2600			2610			2620			2630			2640			2650		
2660			2670			2680			2690			2700			2710		
2720			2730			2740			2750			2760			2770		
2780			2790			2800			2810			2820			2830		
2840			2850			2860			2870			2880			2890		
2900			2910			2920			2930			2940			2950		
2960			2970			2980			2990			3000			3010		
3020			3030			3040			3050			3060			3070		
3080			3090			3100			3110			3120			3130		
3140			3150			3160			3170			3180			3190		
3200			3210			3220			3230			3240			3250		
3260			3270			3280			3290			3300			3310		
3320			3330			3340			3350			3360			3370		
3380			3390			3400			3410			3420			3430		
3440			3450			3460			3470			3480			3490		
3500			3510			3520			3530			3540			3550		
3560			3570			3580			3590			3600			3610		
3620			3630			3640			3650			3660			3670		
3680			3690			3700			3710			3720			3730		
3740			3750			3760			3770			3780			3790		
3800			3810			3820			3830			3840			3850		
3860			3870			3880			3890			3900			3910		
3920			3930			3940			3950			3960			3970		
3980			3990			4000			4010			4020			4030		
4040			4050			4060			4070			4080			4090		
4100			4110			4120			4130			4140			4150		
4160			4170			4180			4190			4200			4210		
4220			4230			4240			4250			4260			4270		
4280			4290			4300			4310			4320			4330		
4340			4350			4360			4370			4380			4390		
4400			4410			4420			4430			4440			4450		
4460			4470			4480			4490			4500			4510		
4520			4530			4540			4550			4560			4570		
4580			4590			4600			4610			4620			4630		
4640			4650			4660			4670			4680			4690		
4700			4710			4720			4730			4740			4750		
4760			4770			4780			4790			4800			4810		
4820			4830			4840			4850			4860			4870		
4880			4890			4900			4910			4920			4930		
4940			4950			4960			4970			4980			4990		
5000			5010			5020			5030			5040			5050		
5060			5070			5080			5090			5100			5110		
5120			5130			5140			5150			5160			5170		
5180			5190			5200			5210			5220			5230		
5240			5250			5260			5270			5280			5290		
5300			5310			5320			5330			5340			5350		
5360			5370			5380			5390			5400			5410		
5420			5430			5440			5450			5460			5470		
5480			5490			5500			5510			5520			5530		
5540			5550			5560			5570			5580			5590		
5600			5610			5620			5630			5640			5650		
5660			5670			5680			5690			5700			5710		
5720			5730			5740			5750			5760			5770		
5780			5790			5800			5810			5820			5830		
5840			5850			5860			5870			5880			5890		
5900			5910			5920			5930			5940			5950		
5960			5970			5980			5990			6000			6010		
6020			6030			6040			6050			6060			6070		
6080			6090			6100			6110			6120			6130		
6140			6150			6160			6170			6180			6190		
6200			6210			6220			6230			6240			6250		
6260			6270			6280			6290			6300			6310		
6320			6330			6340			6350			6360			6370		
6380			6390			6400			6410			6420			6430		
6440			6450			6460			6470			6480			6490		
6500			6510			6520			6530			6540			6550		
6560			6570			6580			6590			6600			6610		
6620			6630			6640			6650			6660			6670		
6680			6690			6700			6710			6720			6730		
6740			6750			6760			6770			6780			6790		
6800			6810			6820			6830			6840			6850		
6860			6870			6880			6890			6900			6910		
6920			6930			6940			6950			6960			6970		
6980			6990			7000			7010			7020			7030		
7040			7050			7060			7070			7080			7090		
7100			7110			7120			7130			7140			7150		
7160			7170			7180			7190			7200			7210		
7220			7230			7240			7250			7260			7270		
7280			7290			7300			7310			7320			7330		
7340			7350			7360			7370			7380			7390		
7400			7410			7420			7430			7440			7450		
7460			7470			7480			7490			7500			7510		
7520			7530			7540			7550			7560			7570		
7580			7590			7600			7610			7620			7630		
7640			7650			7660			7670			7680			7690		
7700			7710			7720			7730			7740			7750		
7760			7770			7780			7790			7800			7810		
7820			7830			7840			7850			7860			7870		
7880			7890			7900			7910			7920			7930		
7940			7950			7960			7970			7980			7990		
8000			8010			8020			8030			8040			8050		
8060			8070			8080			8090			8100			8110		
8120			8130			8140			8150			8160			8170		
8180			8190			8200			8210			8220			8230		
8240			8250			8260			8270			8280			8290		
8300			8310			8320			8330			8340			8350		
8360			8370			8380			8390			8400			8410		
8420			8430			8440			8450			8460			8470		
8480			8490			8500			8510			8520			8530		
8540			8550			8560			8570			8580			8590		
8600			8610			8620			8630			8640			8650		
8660			8670			8680			8690			8700			8710		
8720			8730			8740			8750			8760			8770		
8780			8790			8800			8810			8820			8830		
8840			8850			8860			8870			8880			8890		
8900			8910			8920			8930			8940			8950		
8960			8970			8980			8990			9000			9010		
9020			9030			9040			9050			9060			9070		
9080			9090			9100			9110			9120			9130		
9140			9150			9160			9170			9180			9190		
9200			9210														

Fig. 2. DNA sequence, and derived protein sequence, of papaya cDNA insert in plasmid pLBPC13

The sequence hybridizing to the oligonucleotide probe is underlined, and the 'active-site' histidine residue is marked (*). Restriction-endonuclease sites useful in the sequencing of the insert are shown.

downstream by the start of the 3' poly(A) tail. In clone pLBPC13, no such precise poly(A) addition signal is observed, although the sequence CATAA at position 404 could be a variant of this sequence; however, the 3'-sequence of this clone lacks a poly(A) tail.

Determination of mRNA sizes

The size of the mRNAs was determined by 'Northern' hybridization analysis, as shown in Fig. 4. mRNA coding for each protein is estimated to be about 1.45 kb. There is no qualitative difference between mRNA from leaf and

fruit tissue, and the mRNA hybridizing the endonuclease-*Hind*III/*Bam*HI DNA fragments from clones pLBPC13 and pLBPC18 is of similar size. Poly(A) RNA gives one major hybridizing band, whereas total RNA gives a second band of approx. 3.35 kb. This second band observed in fruit and leaf tissue total RNA is taken to be the unprocessed mRNA precursor. The poly(A) RNA is sufficient to code for a protein of 30–35 kDa, depending on the size of the 5' untranslated region. Several of the clones that were partially DNA-sequenced were identical with pLBPC13; furthermore, although isolated from

Fig. 3. Homology between the papaya cDNA derived protein sequences, plant cysteine proteinases, cathepsins B and H, a 'cathepsin-like' protein and a Ca²⁺-dependent proteinase

For the papain, actinidin and bromelain sequences, see Ramshaw (1982); for cathepsins B and H, see Takio *et al.* (1983) and San Segundo *et al.* (1985). For the 'cathepsin-like' protein, see Gay & Walker (1985), and for the Ca²⁺-dependent proteinase, see Ohno *et al.* (1984). The active-site histidine residue of papain is marked (*).

Fig. 4. 'Northern' blot hybridization of poly(A) RNA and total RNA with papaya cDNAs

Poly(A) RNA from paw-paw fruit (b) and (f) and from leaf (c) and (g) tissue (1 µg) was probed with ³²P-labelled DNA probes corresponding to the complete inserts from (1) pLBPc13 and (2) pLBPc18 clones (*Hind*III, *Bam*HI fragments). Total RNA from fruit (a) and (e) and from leaf (d) and (h) tissue was hybridized with the same probes. The origin, and sizes in kb, are shown. Marker sizes (not shown) were determined from *Hind*III digestion products of bacteriophage-λ DNA and *Hae*III products of bacteriophage φX174 DNA replicative form.

different cDNA banks, they also appear to terminate at the same 5' sequences as does pLBPc13. The results of the 'Northern' blots indicate that this premature termination is not due to limited availability of full-length mRNA molecules, but may be a consequence of secondary structure in a full-length RNA molecule. The complete sequence of the cDNA and genomic DNA will provide information concerning the identity of the mRNA and the proteins. These preliminary data confirm our protein fractionation and characterization studies (R. A. McKee & H. Smith, unpublished work), which have indicated that the papaya latex may contain a large number of very closely related cysteine proteinases, differing only in a few amino acid residues, and reflecting the presence of a multigene family.

This work was funded by a Science and Engineering Research Council Co-operative Award to H.S. We are grateful to Ms. Fiona Page and Ms. Karen Smith for excellent technical assistance.

REFERENCES

- Baker, E. N. (1980) *J. Mol. Biol.* **141**, 441–484
 Bantle, J. A., Maxwell, I. H. & Hahn, W. E. (1976) *Anal. Biochem.* **72**, 413–427
 Barrett, J. & Buttle, D. J. (1985) *Biochem. J.* **228**, 527
 Biggin, M. D., Gibson, T. J. & Hong, K. F. (1983) *Proc. Natl. Acad. Sci. U.S.A.* **80**, 3963–3965
 Birnboim, H. C. & Doly, J. (1979) *Nucleic Acids Res.* **7**, 1513–1523
 Brocklehurst, K., Salih, E., McKee, R. A. & Smith, H. (1985a) *Biochem. J.* **228**, 525–527
 Brocklehurst, K., McKee, R. A., Salih, E. & Smith, H. (1985b) *J. Protein Chem.* **4**, 103–127
 Buttle, D. J. & Barrett, A. J. (1984) *Biochem. J.* **223**, 81–88
 Carne, A. & Moore, C. H. (1978) *Biochem. J.* **173**, 73–83
 Drenth, J., Jansonius, J. N., Koekoek, R. & Walther, B. G. (1971) *Adv. Protein Chem.* **25**, 79–115
 Dretzen, G., Bellard, M., Sassone-Corsi, P. & Chambon, P. (1981) *Anal. Biochem.* **112**, 295–298
 Gay, N. J. & Walker, J. E. (1985) *Biochem. J.* **225**, 707–712
 Goto, K., Murachi, T. & Takahashi, N. (1976) *FEBS Lett.* **62**, 93–95
 Goto, K., Takahashi, N. & Murachi, T. (1980) *Int. J. Peptide Protein Res.* **15**, 335–341
 Grunstein, M. & Hogness, D. S. (1975) *Proc. Natl. Acad. Sci. U.S.A.* **72**, 3961–3965
 Gubler, U. & Hoffman, B. J. (1983) *Gene* **25**, 263–269
 Kamphuis, I. G., Kalk, K. H., Swarte, M. B. A. & Drenth, J. (1984) *J. Mol. Biol.* **179**, 233–256
 Kirschke, H., Locnikar, P. & Turk, V. (1984) *FEBS Lett.* **174**, 123–127
 Lynn, K. R. (1983) *Phytochemistry* **22**, 2485–2487
 Mandel, H. & Higa, A. (1970) *J. Mol. Biol.* **53**, 159–162
 Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, New York
 Messing, J. & Vieira, J. (1982) *Gene* **19**, 269–276
 Mitchell, R. E. J., Chaiken, I. M. & Smith, E. L. (1970) *J. Biol. Chem.* **245**, 3485–3492
 Noyes, B. E., Mavareck, M., Stein, R. & Agarwal, K. L. (1979) *Proc. Natl. Acad. Sci. U.S.A.* **76**, 1770–1774
 Ohno, S., Emori, V., Imajoh, S., Kawasaki, H., Kisaragi, M. & Suzuki, K. (1984) *Nature (London)* **312**, 566–570
 Okayama, H. & Berg, P. (1982) *Mol. Cell. Biol.* **2**, 161–170
 Pritchard, R. H. & Holland, I. B. (1985) *Basic Cloning Techniques: a Manual of Experimental Procedures*, Blackwell Scientific Publications, Oxford
 Proudfoot, N. J. & Brownlee, G. G. (1976) *Nature (London)* **263**, 211–214
 Ramshaw, J. A. M. (1982) in *Nucleic Acids and Proteins in Plants I* (Boulter, D. & Parthier, B., eds), pp. 229–290, Springer Verlag, Berlin and New York
 Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463–5467
 San Segundo, B., Chan, S. J. & Steiner, D. F. (1985) *Proc. Natl. Acad. Sci. U.S.A.* **82**, 2320–2324
 Singer-Sam, J., Simmer, R. L., Keith, D. H., Shively, L., Teplitz, M., Itakura, K., Gartler, S. M. & Riggs, A. R. (1983) *Proc. Natl. Acad. Sci. U.S.A.* **80**, 802–806
 Takio, K., Towatari, T., Katunuma, N., Teller, D. C. & Titani, K. (1983) *Proc. Natl. Acad. Sci. U.S.A.* **80**, 3666–3670