Mutagenesis and Kinetic Studies of a Plant Cysteine Proteinase with an Unusual Arrangement of Acidic Amino Acids in and around the Active Site[†,‡]

Carol E. Carter,[§] Howard Marriage,[||,⊥] and Peter W. Goodenough*[,§]

Wolfson Protein Engineering Facility, School of Plant Sciences, Plant Science Laboratories, University of Reading, Whiteknights, P.O. Box 221, Reading RG6 6AS, U.K., and Genzyme Ltd., 50 Gibson Drive, Kings Hill, West Malling, Kent ME19 6HG, U.K.

Received November 29, 1999; Revised Manuscript Received May 24, 2000

ABSTRACT: We report the cloning, overexpression, kinetic analysis, and modeling of the tertiary structure of an unusual plant cysteine proteinase. Ananain (EC 3.4.22.31), from *Ananas comosus (*pineapple) is distinguished from all other cysteine proteinases in the papain superfamily by having a unique combination of acidic amino acids. As well as lacking the acidic residue immediately preceding the active site histidine (position 158 in papain), it also lacks the extensive surface network of acidic residues that were postulated to compensate for the loss of charge at position 158 in mammalian cathepsins. Ananain has the fewest acidic residues, so far reported, of any plant cysteine proteinase, but two of the carboxyl residues (E50 and E35) postulated to have an enabling role in catalysis, the so-called "electrostatic switch", remain conserved. Comparisons of the kinetics of recombinant wild-type ananain with E50A and E35A mutants proves that these charged groups are not essential for catalysis. Hence this research does not confirm the presence of an electrostatic switch in this cysteine proteinase, and the role of acidic residues in the enhancement of catalytic competence in these enzymes is discussed in light of this new evidence.

The cysteine proteinases of plant origin, particularly the enzymes of the tropical plants *Carica papaya* (papain, chymopapain, caricain, papaya proteinase IV), *Actinidia chinensis* (actinidin), and *Ananas comosus* (fruit bromelain, stem bromelain, ananain, comosain), are of considerable commercial importance, due to their strong proteolytic activity against a broad range of protein substrates. Many cysteine proteinases are endopeptidases which have catalytic activity due to the nucleophilicity of an active site cysteine residue (*1*). An ion pair is formed between the active site cysteine thiolate ion and a histidine imidazolium group (C25 and H159 in papain) on opposite sides of a cleft between two structurally conserved globular domains.

It has been suggested that for the most studied examples, papain and caricain, the formation of the C25/H159 ion pair alone does not lead to catalytic competence without additional ionizations of surrounding carboxyl groups, possibly controlling ion pair geometry (*2*−*6*). This phenomenon has become known as the "electrostatic switch" (2). The arrangement of charged amino acids around the active site has been shown to alter the ionization (pK_a) of C25 and H159 by between 4 and 5 pH units. It was therefore originally thought that the ionization of D158, the nearest charged group to the active site, was the major factor leading to catalytic competence of the C25/H159 ion pair. Evidence against this theory includes the discovery that some animal cathepsins (homologues of plant cysteine proteinases) do not have a charged residue at position 158, this electrostatic role apparently being taken up by increased numbers of carboxyls in the vicinity of the ion pair. Mutagenesis experiments with papain and caricain have shown that the complete removal of charge at position 158 results in only a 90% loss in activity, indicating that a negative charge at position 158 is not essential for some catalytic competence (*6, 7*). Given these two pieces of evidence Pinitglang et al. (*2*) identified a pivotal electrostatic role for the conserved carboxyl at position 50 (i.e., E50), as there is a strong interaction energy with the catalytic ion pair in both papain and caricain. However, we have shown that removing the charge at position 50 in caricain does not significantly reduce the catalytic activity and, in fact, enhances the activity at low pH (*8*). In this report we evaluate conserved negative charges around the active site as potential candidates for the electrostatic switch by kinetic analysis of both wild-type and mutant molecules of ananain (EC 3.4.22.31) from *A. comosus* (pineapple). This cysteine proteinase has a remarkable arrangement of charged residues around the active site. Although not unique in lacking a charge at position 158, it is, however, unique in lacking the extensive network of carboxyls postulated to act as a compensating mechanism when the charge at position 158 is absent. In fact, ananain has fewer carboxyls than any other plant cysteine proteinase.

The crude extract of pineapple stem, which has been widely used in the food and brewing industries for many

[†] We acknowledge funding from the U.K. Biotechnology and Biological Sciences Research Council (grant award identification number T07406).

[‡] The following clones have been registered with EMBL: proananain (ACAJ2477) and putative comosain clones from *Ananas comosus* (AJ009829 and AJ009830).

* Corresponding author. Tel: (44) (0) 1189 316332. Fax: (44) (0) 1189753676. E-mail: P.W.Goodenough@Reading.ac.uk.

[§] University of Reading.

[||] Genzyme Ltd.

[⊥] Present address: Cyclacel Ltd., 5 Whitehall Crescent, Dundee DD1 4AR, U.K.

10.1021/bi992714w CCC: $19.00
Published on Web 08/15/2000

years (*9*), contains at least three immunologically distinct papain-like cysteine proteinases—stem bromelain, ananain, and comosain (*10, 11*) in the proportions 90:8:1, respectively (*10, 12*). It has been suggested that, although ananain comprises only a few percent of the proteinase in the crude stem extract, it is responsible for most of the proteolytic activity (*10*). It has also been shown that ananain is particularly efficient as a tissue debridement pharmaceutical (*11, 13, 14*).

This paper reports the testing of the hypothesis that an electrostatic switch is necessary for proteinase activity by biochemical analysis of recombinant wild-type and mutant proananain expressed using an *Escherichia coli* T7 RNA polymerase inducible system. Although this enzyme from *A. comosus* has recently been purified and sequenced at an amino acid level (*15*), the cDNA had not been isolated previously; hence degenerate primers were used to amplify a fragment of ananain cDNA. This was then used to isolate a full-length cDNA for pre-proananain from a pineapple stem cDNA library. Recombinant proananain was refolded and activated, producing approximately 10 mg of refolded ananain/L of *E. coli* culture. Cysteine proteinases are expressed as inactive zymogens, the N-terminal propeptide being essential for the production of functional enzyme, facilitating correct folding of the polypeptide, as well as preventing uncontrolled proteolysis (*16*−*18*). The recombinant proenzyme can be activated in vitro in low pH conditions, which mimic the acidic environment of the plant vacuole, the site of proteolytic processing in nature.

The ability to produce recombinant cysteine proteinases in *E. coli* has enabled major advances in the understanding of their structures and catalytic mechanisms, providing large amounts of pure protein for X-ray crystallography and enabling site-directed mutagenesis to identify residues important for structure or catalysis (*3, 4, 6, 8, 19*−*23*).

We also report a homology model for proananain, produced using the Homology module of Insight II (MSI), and correlate this structural information with the results of a pH activity (k_{cat}/K_m vs pH) profile of the recombinant enzyme, showing a broad pH optimum from pH 5.75 to pH 7.75. Significantly and perhaps unexpectedly, the activity profile exhibits clear evidence of multiple ionizations in the acid limb, despite there being significantly fewer solvent-accessible carboxyls in ananain than in any other plant cysteine proteinase. Of the carboxyls identified by Pinitglang et al. (*2*) as having high interaction energies with the catalytic ion pair, only E50 and E35 remain conserved in ananain. Site-directed mutagenesis and kinetic analyses of recombinant E50A and E35A proteinases give conclusive evidence that these charged groups are not essential for catalysis, therefore drawing into question the involvement of an electrostatic switch mechanism in this case.

EXPERIMENTAL PROCEDURES

E. coli Strains and Plasmid Vectors. *E. coli* strain DH5α F$^-$[ϕ80d*lac*ZΔM15 Δ(*lac*ZYA-*arg*F)U169 *deo*R *rec*A1 *end*A1 *hsd*R17(r_K^-,m_K^+) *pho*A *sup*E44λ^- *thi*-1 *gyr*A96 *rel*A1] was used as a general host for subcloning and plasmid propagation. *E. coli strain* BL21 (DE3)pLysS (*E. coli* B F$^-$ [*dcm ompT hsd*S(r_B^-, m_B^-) *gal*λ(DE3)][pLysS Cam^r]), purchased from Stratagene Ltd. (Cambridge, U.K.), was used as the host strain for proananain expression.

Bluescript SK plasmid (Stratagene Ltd., Cambridge, U.K.) was used as a general cloning vector. The pET3a vector (Stratagene Ltd., Cambridge, U.K.) was used to express proananain under the inducible control of a T7 promoter.

Preparation of Total RNA. The crown stems of ripe pineapple fruit, *A. comosus* (L.) Merr (smooth cayenne) were removed from the fruit and rooted in peat-based compost, regenerating rooted stem tissue. Total RNA was extracted from the youngest stem material using the method of Jordan et al. (*24*).

Reverse Transcription PCR Amplification of a Partial Ananain cDNA. First strand cDNA was produced via use of the Promega reverse transcription system (Promega Corp.); standard conditions were used with 5 μg of total RNA and 0.5 μg of RACEdT primer (5′ GCTAGGATCCGTCGACATCGAT[dT]$_{16}$ 3′).

Amplification. *(1) 3′ RACE PCR*. First strand cDNA was used as the template for initial PCR reactions; 10 μL of the product from the reverse transcription reactions was diluted 1 in 5 in water and used as template in 100 μL volume PCR reactions, using Biotaq polymerase (Bioline U.K. Ltd.). Primers were RACE (5′ GCTAGGATCCGTCGACATCGAT 3′) and PROT1 (5′ TTAGTNCCNCARWSNATHGAYTGG 3′). The reactions were cycled for 1 min at 94 °C, 1 min at 47 °C, and 1 min at 72 °C for 30 cycles.

(2) Nested Ananain-Specific PCR. The products of the first amplification (diluted 1 in 20 in water) were used as template in a nested amplification reaction, with ananain-specific AN2 (5′ GCTARTTNCKNACDATCCARAAKTT 3′) and PROT1. The same reaction conditions were used as previously, except that a lower annealing temperature of 45 °C was necessary.

Cloning of the Ananain PCR Product. PCR products were cloned into Bluescript SK plasmid vector via the Di/Trinucleotide Sticky End cloning procedure (*25*) between restriction enzyme sites *Eco*RI and *Hin*dIII.

Construction and Screening of the Pineapple Stem cDNA Library. Poly(A$^+$) mRNA was purified by using the Qiagen Oligotex mRNA system (Qiagen Ltd.). A cDNA library was constructed in the Uni-ZAP XR λ vector (Stratagene Ltd.) from 5 μg of poly(A$^+$) RNA, using Stratagene's ZAP-cDNA synthesis and ZAP-cDNA Gigapack Gold Cloning kits, according to their protocols. A clone containing the entire pre-proananain cDNA sequence (pBS-PPA) was isolated after three rounds of screening with ^{32}P-labeled AN2/PROT1 PCR product.

Homology Model Using the Homology Module of Insight II (MSI). A model of the structure of proananain was constructed on the basis of the published crystallographic structures of actinidin (2act.pdb; *26*), chymopapain (1yal.pdb; *27*), papain (9pap.pdb; *28*), procaricain (1pci.pdb; *22*), and procathepsin L (1cjl.pdb; *29*).

Expression of Proananain Recombinant Protein in E. coli BL21 Cells. The coding sequence of proananain was amplified from pBS-PPA via Pfu DNA polymerase (Strategene Ltd.), using the primers EX1 (5′ CGCGCTTCATATGGACGAACCCAGTGATCCCATG 3′) and EX2 (5′ CGCGCGGATCCCTAAATGACTTCGACACTGGG 3′). These primers enabled the deletion of the 26 aa N-terminal preregion, the addition of an in-frame methionine codon at the new N-terminus, and incorporation of sites for *Nde*I (EX1) and *Bam*HI (EX2) (underlined sequences). The reactions were cycled for 45 s at 95 °C, 45 s at 45 °C, and 2.5 min at 72 °C

for 15 cycles. The PCR product was cloned between *Nde*I and *Bam*HI sites in the pET3a expression vector (*30*). This construct is referred to as pET-PA.

E. coli BL21(DE3)pLysS cells were transformed with the pET-PA plasmid DNA. Protein production was induced by 0.4 mM IPTG[1] as described (*19*).

SDS−PAGE Electrophoresis and Western Blot Analysis. Cell pellets were harvested prior to induction and at various times postinduction before being resuspended in SDS gel-loading buffer (31). Protein samples were heated at 95 °C for 5 min, debris was removed by centrifugation, and supernatants were analyzed by SDS−PAGE in 15% gels (*31*).

Protein bands containing ananain were detected, after blotting onto nitrocellulose membranes, by alkaline phosphatase activity, following sequential binding of an affinity-purified anti-ananain polyclonal antibody, raised in rabbit (courtesy of Tim Edmunds and Jean Gamble, Genzyme Corp., Framingham, MA.), and an alkaline phosphatase conjugated anti-rabbit secondary antibody (Sigma).

Refolding and Activation of the Recombinant Protein. Cells were harvested from 3 L of *E. coli* culture 3−4 h postinduction, and inclusion bodies were prepared. Four hundred milligrams of inclusion body pellet was solubilized, and recombinant protein was refolded in 5 L of refolding buffer, as described (*4*). Refolded protein was concentrated using the Millipore minitan filtration system with a PK5000 NMWL high-flux polysulfone membrane (5 L to 100 mL) and then aquacide II powder [Calbiochem Novabiochem (U.K.) Ltd.] (100 to 10 mL).

Enzyme Purification and Kinetic Analysis. The enzyme was purified by cation-exchange chromatography on a Mono S HR5/5 column with a gradient of sodium acetate (pH 5.0) from 50 mM to 1.0 M in the presence of 10 mM cysteine, pure ananain being eluted with about 0.2 M salt.

Proananain was activated in 50 mM sodium acetate (pH 4.0), in the presence of 20 mM cysteine and 4 mM EDTA, at 55 °C for 30−60 min. Activity was determined by following the release of *p*-nitroaniline from the substrate Pyr-Glu-Phe-Leu-pNA, at 25 °C over a pH range from 3.0 to 9.5, which is measurable at 410 nm ($E_{410} = 8.8 \times 10^{-3}$ M^{-1} cm^{-1}). Assay buffers and conditions were as described in ref *8*. The value of $(k_{cat}/K_m)_{obs}$ was obtained from the relationship $v_i = e_0 s(k_{cat}/K_m)_{obs}$, where e_0 = enzyme concentration, as determined by E64 titration (*6*, *7*, *32*). This substrate has poor solubility in aqueous buffers and is used at a concentration where $s \ll K_m$. Detail of this method is given in ref *33*, and this relationship has been used for all cysteine proteinases so far investigated. Data were analyzed using the reaction scheme described in ref *6* for four hydrogenations and three active species, and pK_a data were modeled using the nonlinear curve-fitting program Kaleidograph (Synergy Software).

Site-Directed Mutagenesis. Mutagenic primers hybridizing to complementary strands of the pET-PA clone included single nucleotide point mutations leading to either an E35A or E50A mutation. The 5′ ends also incorporated point mutations such that a new unique *Bst*EII site was formed on recircularization of the mutated PCR product. The primer sequences were as follows: Glu50Ala (forward), 5′ ACcTAGTATCTTTATCGGcGCAGCAAGTTCT 3′; Glu35Ala (reverse), 5′ TaCCTCTTTTGATCTTGTAGATTGATgCTAC 3′; Glu50wt (forward), 5′ ACcTAGTATCTTTATCGGAGCAGCAAGTTCT 3′; Glu35wt (reverse), 5′ TaCCTCTTTTGATCTTGTAGATTGATTCTAC 3′. Lower case letters indicate point mutations, and underlined letters indicate parts of the *Bst*EII restriction site which is formed on correct recircularization of the PCR product.

Primers Glu50Ala with Glu35wt and Glu50wt with Glu35Ala were used to amplify the entire pET-PA sequence. The 5.5 kb PCR products were purified by using agarose gel electrophoresis and the Wizard PCR preps DNA purification system (Promega Life Science). The 5′ ends were phosphorylated and ligated as described in ref *6*. *E. coli* BL21(DE3)pLysS cells were transformed with the recircularized plasmid DNA, as described previously.

RESULTS

Cloning of Pre-proananain Complete Coding cDNA. A 537 bp proananain cDNA fragment was amplified from total pineapple stem RNA by RT-PCR, using degenerate oligonucleotide primers designed from the ananain amino acid sequence (*15*). A nested PCR approach was necessary; in the first reaction 3′ RACE PCR (*34)* nonspecifically amplified cDNAs for numerous members of the cysteine proteinase family, while subsequently a nested PCR reaction amplified ananain specifically from this enriched population.

A clone containing the entire coding sequence of pre-proananain (submitted to EMBL, ACAJ2477) was isolated from a pineapple stem cDNA library using the AN2/PROT1 PCR product as a homologous probe. The translated sequence includes a 26 aa hydrophobic signal peptide, a 96 aa N-terminal propeptide, a 216 aa mature enzyme sequence, and a 7 aa C-terminal propeptide. The mature enzyme sequence is identical to that of P80884 (*15)* except for two substitutions—serine for alanine at position 169 and leucine for isoleucine at position 202 (numbers refer to the mature enzyme sequence).

Construction of a Modeled Structure Using Homology. The amino acid sequence of pre-proananain is 44% homologous to that of pre-prochymopapain (P14080), 43% to pre-procaricain (P001056), 41% to pre-proactinidin (P00785), 39% to pre-propapain (P00784), and 34% to human pre-procathepsin L (P07711). Multiple sequence alignments are shown in Figure 1, and the structures of the five cysteine proteinases mentioned above were used to construct a homology model of proananain using the Insight II package.

The structure of the proananain sequence was aligned onto the structures of five cysteine proteinases according to sequence homology: (P0−96) propeptide aligned to procaricain (except for insertion of G at P50 in proananain, loop generated), (1−24) chymopapain, (25−29) caricain, (30−58) actinidin, (59−71) chymopapain, (72−74) caricain, (75−76) actinidin, (77−78) cathepsin L, (79−100) caricain, (101) chymopapain, (102−103) actinidin, (104−109) chymopapain, (110−113) caricain, (114−118) actinidin, (119−126) papain, (127−149) actinidin, (150−154) chymopapain, (155−165) actinidin, (166−168) generated loop, (169−171) caricain, (172−195) cathepsin L, (196−200) caricain, (201−

[1] Abbreviations: Pyr-Glu-Phe-Leu-pNA, L-pyroglutamyl-phenylalanyl-L-leucine *p*-nitroanilide; E64, 1-[(L-*trans*-epoxysuccinyl)-L-leucylamino]-4-guanidinobutane; IPTG, isopropyl β-D-thiogalactopyranoside; DMSO, dimethyl sulfoxide.

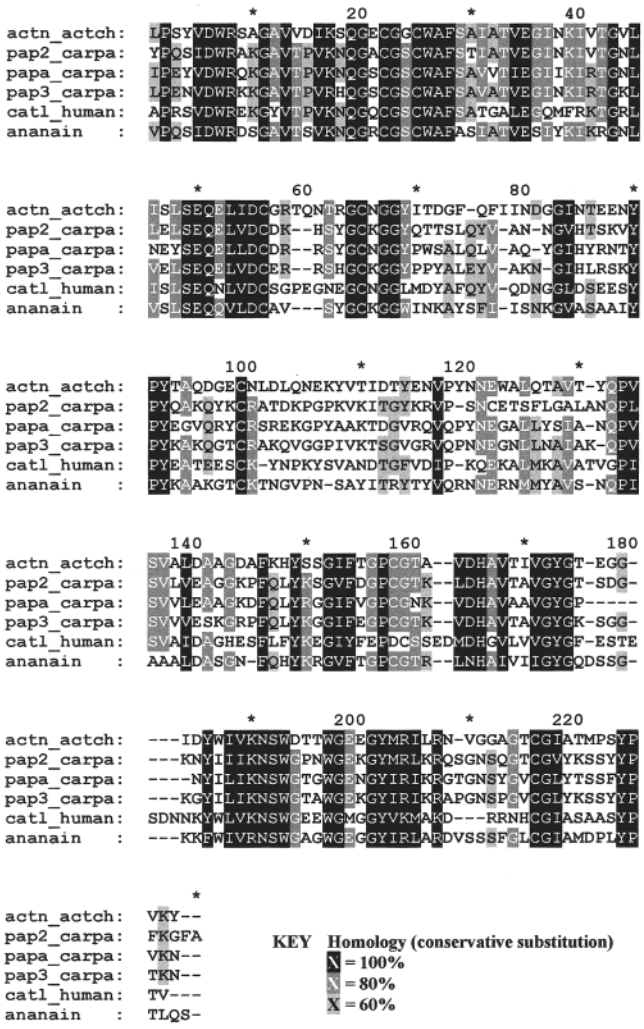

FIGURE 1: Multiple sequence alignment of the mature enzyme sequences of actinidin (actn_actch.swissprot; accession number P00785), chymopapain (pap2_carpa.swissprot; P14080), papain (papa_carpa.swissprot; P00784), caricain (pap3_carpa.swissprot; P001056), human cathepsin L (catl_human.swissprot; P07711), and ananain (ACAJ2477; AJ002477).

212) actinidin, (213−214) cathepsin L, (214−216) generated loop. (P denotes the propeptide sequence.)

The modeled proananain C_α structure is a very close match to the structure of procaricain, having an rms deviation of 0.61 Å for 95.7% of the residues. The model of the mature ananain structure is most similar to the structure of actinidin with an rms deviation of 0.64 Å for 91.7% of the residues and is also very similar to that of the other cysteine proteinases used for the model (chymopapain 0.55 Å for 88.4% of the residues, papain 0.65 Å for 92.1% of the residues, and caricain 0.71 Å for 93.5% of the residues).

The pro region corresponds to the caricain and cathepsin L subfamily of cysteine proteinases; cathepsin B is an example of a second subfamily with a shorter nonglobular proregion of 62 residues (*22*, *35*). The homology model of ananain is adequately represented by the ribbon representation shown in Figure 2 of ref *22* and so is not reproduced here. Residues 12−112 and 208−212 (papain numbering, *36*) are highly conserved regions of mainly the α-helix. This is particularly true of α-helix A (residues 24−42 in ananain), which, in all cases, runs through the center of the molecule at the interface of the two domains and contains the active site C at position 25, and helix B (residues 50−57) containing the conserved E50. In the R-domain (residues 1−11 and 113−207, papain numbering), the mainly *β*-sheet secondary structure (*36*, *37*) is even more highly conserved than the α-helices in the L-domain. In this region there is a four amino acid insertion relative to papain (167−170), but similar insertions are found in all other plant cysteine proteinases. This high degree of homology in the family makes it highly likely that the Homology model will be a good representation of any experimentally determined structure.

Table 1: Comparison of Charged Side Chains in Papain, Caricain, and Ananain[a]

papain	caricain	ananain	papain	caricain	ananain
		(a) Acidic			
E3	E3		E99		
D6	D6	D6	D108		
	k9	D9	E118	E118	E117
E35	E35	E35	E135	E135	D134
E47	E47		D140		
E50	E50	E50		E150	
E52	E52		D158	D158	
D55	D55	D55		k168	D167
D57	E57		E183	E187	E186
	E73				D195
E89	k89	k89			D208
		(b) Basic			
R8	R8		K106	K106	
	K9	d9			R108
K10	K10		R111		
K17	R17	K17		R112	
	H18				R114
		R21			R118
K39	K39	K39		K127	
R41	R41	K41		K137	
		R42	K139	R139	
	K44				H141
R58	R58		R145	K145	K143
R59	R59				R144
	H61		K156	K156	R154
	K64	K63	H159	H159	H157
		K69		K168	d167
	K77	K78		K172	K171
H81	H81				K172
R83	R83		K174	K178	R177
	K85			K188	
e89	K89	K89	R188	R192	R191
	K91		K190	K194	
R93		K92	R191	R195	R194
R96	R96			K208	
R98	K98	K98	K211	K215	
K100					

[a] Noncharged residues in equivalent positions are not shown. Opposite charges are in lower case.

Comparison of Charged Residues in Mature Enzymes. It is important in the study of cysteine proteinases that the number and position of charged carboxyls are evaluated to determine their possible mechanistic roles. As mentioned earlier, the mechanism of activation is postulated to require ionization of carboxyls in addition to formation of the C25 thiolate ion. Table 1 summarizes a comparison of charged residues in the most studied of the papain-like cysteine proteinases. Ananain has the lowest number of negatively charged residues, 11, of any plant cysteine proteinase sequenced to date (data not shown). P34 thiol proteinase of soybean (accession number P22895) has the most negative charges at 35, and chymopapain (P001056) and caricain (P001056) the second least at 14. The low number of

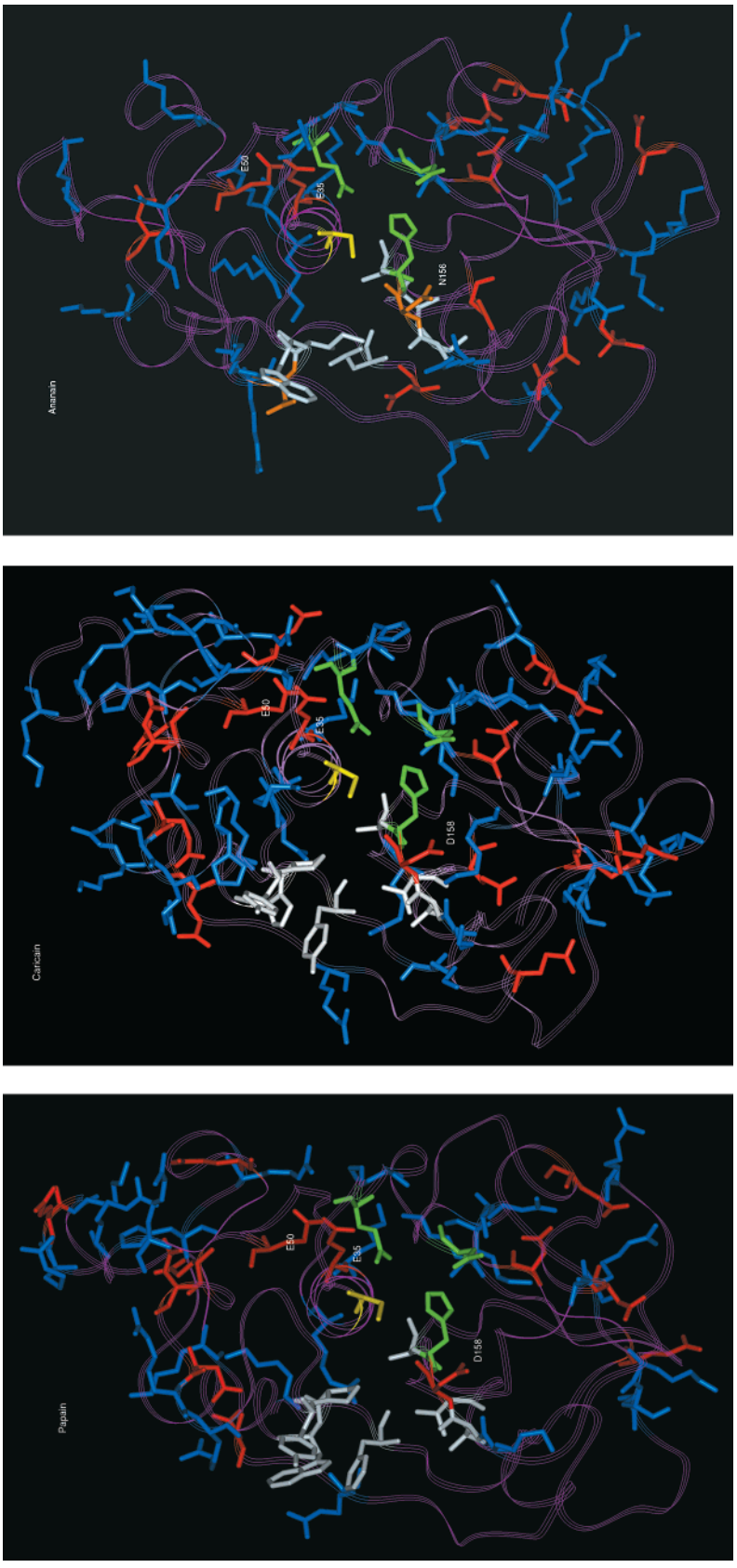

FIGURE 2: Tertiary structures of (a, left) papain and (b, middle) caricain compared to (c, right) the modeled structure of ananain. The first two structures are displayed using the coordinates from the PDB (9pap.pdb, papain; 1pci.pdb, caricain), and the third structure was constructed using the program Homology (Insight II, MSI). Residues colored red are carboxyls, residues colored blue are amines, and residues colored green are the amino acids involved in the chemical reaction, H158, N175, and Q19. Yellow denotes the active site cysteine (C25). Those residues colored white are the hydrophobic residues which delineate the P2 pocket. D158 is colored red in papain and caricain but gold, as it is N, in ananain.

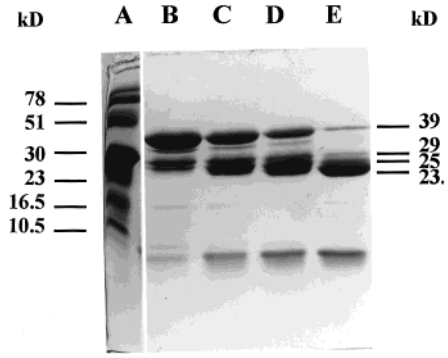

FIGURE 3: Activation of recombinant proananain. Purified refolded proananain was dialyzed into buffer containing 50 mM sodium acetate, pH 4.0, 20 mM cysteine, and 4 mM EDTA and then incubated at 55 °C. Samples were removed at time intervals during activation, and E64 was added to 50 μM to stop further activation. Samples were analyzed as described above. Lanes: A, color markers wide range (Sigma); B−E, protein samples during activation by incubation at 55 °C for (B) 0 min, (C) 10 min, (D) 20 min, and (E) 30 min. The activity of the crude enzyme against Pyr-Glu-Phe-Leu-pNA [mM s^{-1} (μg of protein)$^{-1}$ × 10^{-6}] at each time point was (B) 0, (C) 0.37, (D) 0.52, and (E) 1.6.

negative charges in ananain is all the more unusual because the residue next to the catalytic H, in the position analogous to that of D158 in papain, is an uncharged N. It has previously been observed that in the cathepsins, where there is an uncharged residue in this position, there are many (around 30) negatively charged carboxyls, and it has been suggested that the increased electrostatic field compensates for the loss of the aspartic acid in the active site (*4*). This is clearly not the case in ananain, and there is no correlation between the absence of D158 and the number of negative charges, when the sequences of other plant cysteine proteinases are examined. This suggests that one or more of the carboxyls conserved in ananain must be the electrostatic switch, if this is indeed a conserved function in cysteine proteinases and is essential for catalysis. Glutamate ions at positions 35 and 50 are the conserved negatively charged residues which have the greatest interaction energy with the active site ion pair and are good candidates for such a role in ananain (see Discussion).

Comparison of Active Sites. A ribbon representation of the homology model of the active site of ananain is compared to the equivalent structures of papain and caricain in Figure 2. The side chains of charged residues are highlighted, as are side chains important for catalysis, i.e., C25, H159(157), N175(178), and Q19, and side chains of hydrophobic residues in the P2 substrate binding pocket (numbers as in papain and in parentheses where different in ananain).

Expression of Proananain in E. coli. Crude protein extracts from pET3a-PA cultures,which had been induced with IPTG, were analyzed by SDS−PAGE. A prominent band was produced which corresponded to the expected size for proananain (39 kDa). The sizes of homologous recombinant cysteine proteinase precursors propapain (*19*) and procaricain (*21*) are 39 and 41 kDa, respectively. The protein was further confirmed to be recombinant proananain by reactivity with an affinity purified anti-ananain polyclonal antibody (data not shown).

Recombinant proananain was refolded from inclusion body extracts with a final yield of approximately 10 mg/L of starting culture. The proenzyme was activated in vitro as described earlier, but it was found that the conditions for processing are stringent, particularly the requirement for pH 4.0; no activation occurred at pH 5.0. Figure 3 shows a time course of proenzyme processing at pH 4.0 and 55 °C. Prior to activation the 39 kDa proenzyme predominates, and intermediate species of 29 and 25 kDa can be identified and represent the proenzyme at different stages of processing. The fully processed 23.5 kDa enzyme, active against Pyr-Glu-Phe-Leu-pNA, predominates after 30 min. The molecular mass of pineapple-stem-derived active ananain is 23.42 kDa (*12*).

Enzyme Efficiency over the pH Range 3.0−9.5. The efficiency $(k_{cat}/K_m)_{obs}$ of the enzyme against the substrate Pyr-Glu-Phe-Leu-pNA was determined over a pH range from 3.0 to 9.5. Titration against the inhibitor E64 was used to accurately calculate the concentration of active sites/μg of protein in each reaction. The enzyme efficiency $(k_{cat}/K_m)_{obs}$ was calculated at each pH, and the results are shown in Figure 4a. Plots show a skewed bell-shaped curve of enzyme efficiency against pH with maximal activity in the neutral region from pH 5.75 to pH 7.75. The data were fitted to the 4 pK_a model described in ref *6*, and the calculated pK_as are 2.80, 4.89, 5.43, and 8.37.

Site-Directed Mutagenesis. Mutated E50A or E35A proananain was able to be expressed and refolded as for the wild-type proenzyme. In both cases the size and expression level of proenzyme were similar to the wild-type expression. The mutant proenzymes had different requirements for activation, described below, and activity against the neutral substrate Pyr-Glu-Phe-Leu-pNA was determined over a pH range from 3.0 to 9.0.

The E50A mutant was easily activated, and the $(k_{cat}/K_m)_{obs}$ vs pH profile (Figure 4b) was remarkably similar to that of the wild-type enzyme. The only significant difference was in the acid limb between pH 3.0 and pH 4.0, where the wild-type enzyme showed a significant "shoulder", while in E50A this shoulder is much less pronounced. Also, the activity does appear to be slightly enhanced for E50A, around pH 5.0, as for the E50A mutant of caricain (*8*), although in ananain this is barely significant. The calculated pK_as were 2.68, 4.95, 4.55, and 8.09.

It was observed that the E35A mutation led to increased instability of the refolded recombinant proenzyme, as substantial precipitation of the refolded protein occurred in acidic buffer, which was not observed for either wild type or E50A. Activation of E35A under the same conditions as wild type, i.e., pH 4.0 and 55 °C for 30 min, did not yield significant activity due to protein denaturation. Further optimization of the activation process led to the development of a less destructive procedure, where only a 5 min incubation at 55 °C at pH 4.0 was necessary to produce a small amount of active enzyme. Autocatalytic activation of the remaining proenzyme molecules occurred due to the action of the initial small pool of active proteinase in a 20 min preincubation at 40 °C immediately prior to the assay. The $(k_{cat}/K_m)_{obs}$ with the neutral substrate Pyr-Glu-Phe-Leu-pNA was determined over a pH range from 3.0 to 9.0.

The calculated $(k_{cat}/K_m)_{obs}$ vs pH profile of E35A, summarized in Table 2, is quite different from that of the wild-type enzyme. The enzyme efficiency is maximal at pH 5.0−5.5 and then falls rapidly with increasing pH. At pH 5.0 the $(k_{cat}/K_m)_{obs}$ of E35A is actually enhanced by over 75%

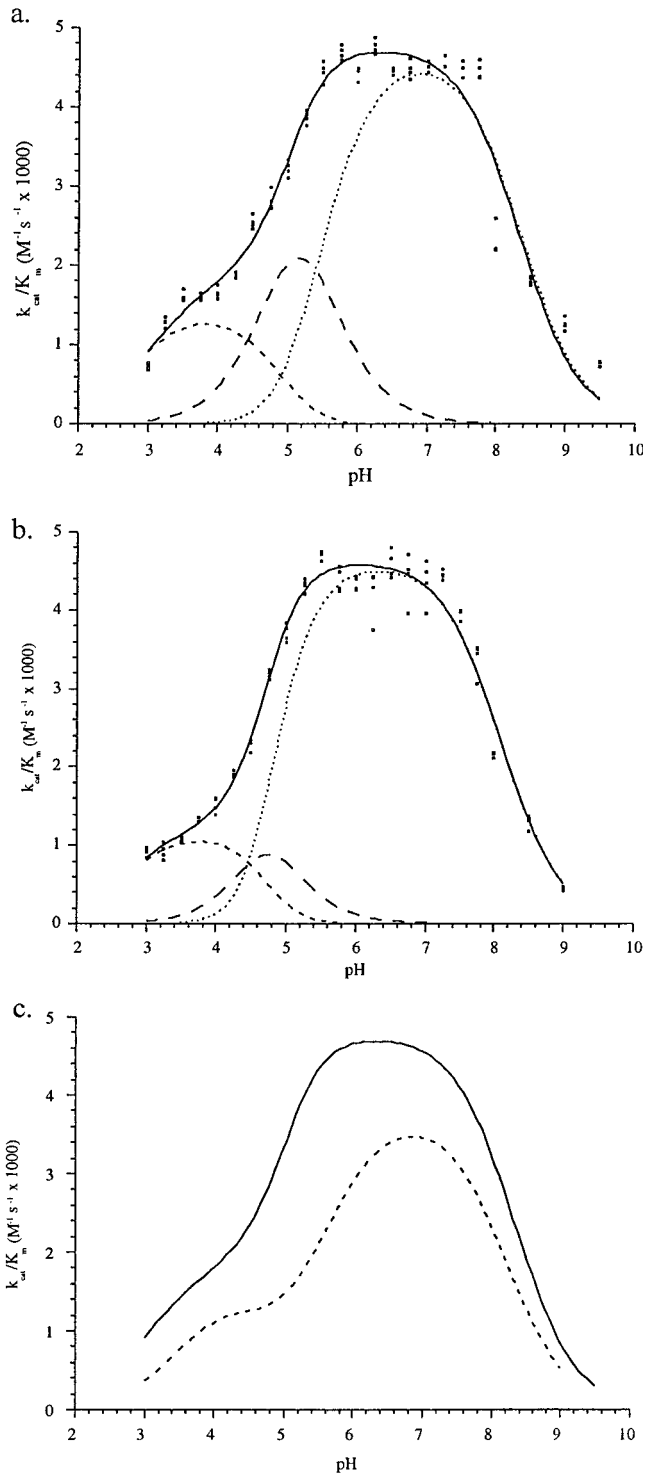

FIGURE 4: Efficiency of activity (k_{cat}/K_m), using the substrate Pyr-Glu-Phe-Leu-pNA, determined over a wide pH range for recombinant wild-type and mutant ananain species. These data was fitted to the equation describing four hydrogenations and three active species (shown as eq 6 in ref *6*). Panels: a, ananain; b, E50A ananain [in both cases solid lines represent $(k_{cat}/K_m)_{obs}$, while broken lines represent contributions of individual components of the curve]; c, wild-type ananain (solid line) and caricain (broken line). The latter data were taken from ref *6* with permission.

compared to the value of the wild-type enzyme; hence, clearly the charge at position 35 is not essential for catalytic activity. The major effect of this mutation seems to be a decrease in structural stability; hence, the analysis was not as extensive as for either E50A or wild type, as the time needed to obtain an extensive data set led to significant loss of enzyme by denaturation. This instability can be explained as, in the wild type, E35, E50, K17, and R177 form electrostatic interactions at the domain−domain interface. E35 may be particularly important in this respect, while it seems that E50 is less important. The data were fitted to the 4 pK_a model described in ref *6*, and the calculated pK_as were 3.88, 6.46, 4.41, and 8.26.

DISCUSSION

Comparison of Mature Enzyme Sequences and Tertiary Structures of Cysteine Proteinases. A comparison of the amino acid sequence of mature ananain with the published sequences of actinidin, chymopapain, papain, caricain, and cathepsin L is presented in Figure 1. The amino acids associated with important catalytic functions are conserved; in the early part of the discussion, sequence numbers refer to papain numbering, and those in parentheses refer to ananain if different from papain. Thus C25 and H159(157) form the catalytic ion pair, N175(178) orientates the catalytic H residue toward C25 via hydrogen bonding, and this bonding is shielded from solvent by the aromatic ring of W177(180). Another important conserved active site residue is Q19, which interacts with the substrate, stabilizing the oxyanion transition state, during catalysis.

Hydrophobic residues which form the hydrophobic binding pocket, accommodating large hydrophobic groups at the P2 site, are partially conserved, Y67, P68, W69, V133, V157, A160, and F207 in papain are replaced, in ananain, by W66, I67, N68, I129, L155, A158, and L210, all hydrophobic, nonpolar residues except for the substitution of the large hydrophobic side chain of W69 with the polar N68 side chain. Variation in this residue is thought to be important in defining substrate specificity (*15*).

The numbers and positions of charged residues in ananain compared to papain and caricain are summarized in Table 1. Of ananain's 11 negatively charged residues, six are conserved exactly, D6, E35, E50, D55, E118(117), and D183(186) (187 in caricain), while one, E135 in papain and caricain, is replaced by similarly charged D134 in ananain. D158 in papain and caricain is replaced by N(157) in ananain, removing this negative charge close to the active site but maintaining important H-bonding patterns. It has been found that ananain has, additionally, four negatively charged residues substituting uncharged or positively charged side chains that are found in papain and caricain. Of these, three, D9, 167, and 195, appear to be located within variable loop regions exposed to the solvent, according to the homology model of ananain's structure (Figure 2). The fourth, D208, is located at the back of the hydrophobic S2 pocket in a position analogous to that of S205 in papain. The side chain at this position markedly affects substrate specificity. In actinidin S205 is replaced by M, and the bulky side chain reduces the size of the S2 pocket, thus reducing the binding of substrates with large aromatic P2 groups. In cathepsin B where this position is occupied by a negatively charged acidic residue, as in ananain, substrates with aromatic groups still bind well due to the predominantly hydrophobic environment, but substrates with basic groups at the P2 position also bind well.

Ananain has 23 positively charged amino acids compared to 33 for caricain but 22 in papain (Figure 2).

Table 2: Comparison of Enzyme Efficiency $(k_{cat}/K_m)_{obs}$ for Wild-Type and Mutant Ananain against the Neutral Substrate Pyr-Glu-Phe-Leu-pNA at Specified pH Values[a]

	enzyme efficiency $(k_{cat}/K_m)_{obs}$ ($M^{-1}\,s^{-1} \times 1000$)								
	pH 3.0	pH 3.5	pH 4.0	pH 5.0	pH 5.5	pH 6.0	pH 7.0	pH 8.0	pH 9.0
wild type	0.74	1.61	1.65	3.23	4.45	4.44	4.53	2.30	1.26
	±0.04	*±0.06*	*±0.08*	*±0.10*	*±0.12*	*±0.08*	*±0.06*	*±0.20*	*±0.08*
E50A	0.92	1.08	1.52	3.71	4.70	4.34	4.36	2.16	0.45
	±0.06	*±0.04*	*±0.10*	*±0.12*	*±0.06*	*±0.08*	*±0.28*	*±0.02*	*±0.02*
E35A[b]	0.25	1.03	1.96	5.67	5.91	4.47	2.07	1.44	0.08

[a] For wild type and E50A, values represent the mean of four replicate assays, and values in italic type indicate standard deviation. [b] Due to the instability of the E35A mutant enzyme, only limited data were collected.

In caricain there are five examples of buried charge pairs, R8-D6, R192-E187, R195-E118, K137-D158, and H81-E52 (*37*). The second and third examples of these charge pairs are conserved in ananain, and the calculated interaction energy of the carboxylate with the ion pair is small (0.1 $kcal^{-1}$ or below). Removal of these charges, either singly or as an ion pair, and replacement with an uncharged residue would almost certainly result in an unstable and unfolded enzyme. Of the other three pairs in caricain only D6 remains as a charged residue in ananain. In addition, D55, D134, and K178 are partially buried in caricain and have adequate compensation by noncharge interactions or partial solvent accessibility. The interaction energies of D55 and D134 with the ion pair are about 50% of those of E50.

The buried charges of the domain−domain interface in papain, E35, E50, K17, and K174 are conserved, except for the substitution of K174 with R177 in ananain. Electrostatic interactions between these charged groups as well as a hydrogen-bonding network with several water molecules are likely to be important in maintaining the correct conformation of the active site cleft between the two domains and are thought to allow flexing between the domains on substrate binding (*39*).

Activity of Recombinant Ananain in the pH Range 3.0−9.5. The pH dependence of k_{cat}/K_m of recombinant ananain (Figure 4a) shows a typical papain-like cysteine proteinase pH profile with three ionizations identified as occurring at acidic pHs (pK_as between 4 and 7). The pH dependence of k_{cat}/K_m appears to reflect ionizations which are directly involved in catalysis or are important for maintenance of the active site's structural conformation. When the k_{cat}/K_m vs pH data for hydrolysis of the neutral substrate Pyr-Phe-Leu-pNA by ananain were fitted to a 4 pK_a model (*6*), pK_as of 2.80, 4.89, 5.43, and 8.37 were calculated. The pK_as of 2.80 and 8.37 are attributed to the ionizations of the active site C thiolate S^- and H imidazolium ImH^+ ions, respectively. We have assumed the other acidic pK_as are analogous to the additional acidic ionizations observed in papain and caricain, postulated as being important in enhancing the catalytic competence.

For papain, reacting with the charged substrate CBZ-Phe-Arg-MCA, there are two detectable acidic pK_as of about 4.0 and a basic pK_a at 8.5 (*7*, *32*), while in caricain (against Pyr-Glu-Phe-Leu-pNA), where the ionizations are more widely spaced than in papain and therefore easier to define, pK_as of 3.47, 5.23, 5.57, and 8.25 are detected (*6*). Figure 4c shows the pH profiles of k_{cat}/K_m for ananain and caricain plotted on the same axes for comparison. It is immediately obvious that ananain is more active against this substrate over the entire pH range. It is also significant that the activity seems particularly enhanced in the acidic region between pH 4.0 and pH 7.0 and that the profile is widened, the pK_as of C and H deprotonation being 3.47 and 8.25 in caricain and 2.80 and 8.37 in ananain, suggesting that the S^- ion is stabilized at a lower pH in ananain. As well as pK_a1 of ananain (2.80) being shifted to a more acidic pK_a than in caricain, the two additional acidic pK_as are also shifted to the acidic side. It must, therefore, be a possibility that this cysteine proteinase does not require an electrostatic switch to ensure catalytic competence of the ion pair. However, one explanation for the increased efficiency brought about by transition through the pK_as of 4.89 and 5.43 would be deprotonation of other sites on the molecule. Experiments with active site probes would be necessary to establish whether the ion pair was fully ionized at pH values below those where catalytic activity could be measured.

Electrostatic Enhancement of Catalysis in Ananain. Ananain is the first plant cysteine proteinase to be cloned which naturally lacks a negative charge analogous to D158 in papain and where the pH dependence of k_{cat}/K_m has been studied in detail. If several ionizations of acidic residues distant from the active site stabilize the ion pair, enabling catalytic competence as suggested by Pinitglang et al. (*2*), the fact that ananain has the fewest number of acidic residues of any plant cysteine proteinase known should have enabled rapid identification of these important charged groups by site-directed mutagenesis.

The homology model shows very high similarity to the experimentally determined structures of caricain and papain, and there is no reason to suppose that interaction energies calaculated for acidic residues with the ion pair are significantly different between papain, caricain, and ananain. E35 and E50 are the major candidates for catalytically essential ionizable groups, able to take the role of an electrostatic switch in ananain. The role of E50 as an electrostatic switch has already been disproved in caricain by site-directed mutagenesis (*8*), but this might well not have been the case in ananain. However, this research conclusively indicates E50 is not an electrostatic switch. The results of the analysis of E35A in ananain reveal that although removal of the charge at position 35 decreases the stability of the enzyme, high activity levels against Pyr-Glu-Phe-Leu-pNA are still observed, and there is a significant enhancement of activity around pH 5.0. Therefore, our results support the theories that the activity of cysteine proteinases is affected by a network of hydrogenations in addition to the active site ion pair, as we have observed several pK_as influencing the activity of ananain in the acidic limb of pH activity profiles. Indeed, removal of the specific negative charges at positions 50 and 35 clearly affects the pH activity profiles, in both

cases causing decreased activity below pH 4.0 and increased activity around pH 5.0, particularly in the case of E35A. These effects may be due, particularly in the case of E35A, to decreased stability of the enzyme, or changes in the active site or substrate binding site conformation, rather than direct effects on the catalytic activity of the C/H ion pair. This rules out the hypothesis that the interaction between these residues and the ion pair is an essential electrostatic switch necessary for catalytic competence in these enzymes.

CONCLUSIONS

Ananain appears to have three kinetically important ionizations affecting catalytic efficiency in the acid limb of the activity vs pH profile, These have been shown not to be controlled by deprotonations of E50 or E35.

Can other possible candidates exist? D55 and E134 have around half the interaction energy with the ion pair that is found with E50 and E35 and are either at a greater distance (D50) or similar distance (E134) to the ion pair. Logically it would seem that if E50 and E35 are not essential for catalytic competence, then it is unlikely that D55 and E135 are candidates either, unless the calculated interaction energies are grossly in error.

On the other hand, it is perhaps necessary to consider if ananain really does possess an electrostatic switch. Due to its unusual arrangement of charged residues, ananain may have a different mechanism to the closely related enzymes. If that was the case, then residues found in papain and caricain but not ananain may well be important in controlling the electrostatic enhancement. Recombinant material of ananain could be made available to those laboratories which have the active site probes.

Alternatively, there are three other charged residues found in ananain that are not found in papain or caricain, and these may yet prove to be involved in enhancement of activity in ananain, acting as an electrostatic switch.

This research is highly significant as it clearly demonstrates that control of enzyme efficiency in cysteine proteinases has not, as yet, been fully explained by mutagenesis experiments and use of sophisticated active site probes, and continued probing of this intriguing active site mechanism is justified.

REFERENCES

1. Brocklehurst, K., Willenbrock, F., and Salih, E. (1987) in *Hydrolytic enzymes* (Neuberger, A., and Brocklehurst, K., Eds.) pp 39−158, Elsevier, Amsterdam.
2. Pinitglang, S., Watts, A. B., Patel, J. D., Noble, M. A., Gul, S., Bokth, A., Naeem, A., Patel, H., Thomas, E. W., Sreedharan, S. K., Verma, C., and Brocklehurst, K. (1997) *Biochemistry 36*, 9968−9982.
3. Katerelos, N. A., Taylor, M., Scott, M., Goodenough, P. W., and Pickersgill, R. W. (1996) *FEBS Lett. 392*, 35−39.
4. Katerelos, N. A., and Goodenough, P. W. (1996) *Biochemistry 35*, 14763−14772.
5. Plou, F. J., Kowlessur, D., Malthouse, J. P. G., Mellor, G. W., Hartshorn, M. J., Pinitglang, S., Patel, H., Topham, C. M., Thomas, E. W., Verma, C., and Brocklehurst, K. (1996) *J. Mol. Biol. 257*, 1088−1111.
6. Taylor, M. A. J., Baker, K. C., Connerton, I., Cummings, N. J., Harris, G. W., Henderson, I. M., Pickersgill, R. W., Sumner, I. G., Warwicker, J., and Goodenough, P. W. (1994) *Protein Eng. 7*, 1267−1276.
7. Menard, R., Khouri, H. E., Plouffe, C., Laflamme, P., Dupras, R., Vernet, T., Tessier, D. C., Thomas, D. Y., and Storer, A. C. (1991) *Biochemistry 30*, 5531−5538.
8. Ikeuchi, Y., Katerelos, N. A., and Goodenough, P. W. (1998) *FEBS Lett. 437*, 91−96.
9. Caygill, J. C. (1979) *Enzyme Microbiol. Techmol. 1*, 233−242.
10. Rowan, A. D., Buttle, D. J., and Barrett, A. J. (1988) *Arch. Biochem. Biophys. 267*, 262−270.
11. Rowan., A. D., Buttle, D. J., and Barrett, A. J. (1990) *Biochem. J. 266*, 869−875.
12. Napper, A. D., Bennett, S. P., Borowski, M., Holdridge, M. B., Leonard, M. J., Rogers, E. E., Duan, Y., Laursen, R. A., Reinhold, B., and Shames, S. L. (1994) *Biochem. J. 301*, 727−735.
13. Rowan., A. D., and Buttle, D. J. (1992) U.S. Patent 5 106 621.
14. Orgill, D. P. (1996) *J. Burn Care Rehabil. 17*, 311−322.
15. Lee, K. L., Albee, K. L., Bernasconi, R. J., and Edmunds, T. (1997) *Biochem. J. 327*, 199−202.
16. . Cohen, L. W., Fluhaarty, C., and Dihel, L. C. (1990) *Gene 88*, 263−267.
17. Smith, S. M., and Gottesman, M. M. (1989) *J Biol. Chem. 264*, 20487−20495.
18. Vernet, T., Tessier, D. C., Richardson, C., Laliberte, F., Khouri, H. E., Bell, A. W., Storer, A. C., and Thomas, D. Y. (1990) *J. Biol. Chem. 265*, 16661−16666.
19. Taylor, M. A. J., Pratt, K. A., Revell, D. F., Baker, K. C., Sumner, I. G., and Goodenough, P. W. (1992) *Protein Eng. 5*, 455−460.
20. Taylor, M. A. J., Baker, K. C., Briggs, G. S., Connerton, I., Cummings, N. J., Pratt, K. A., Revell, D. F., Freedman, R. B., and Goodenough, P. W. (1995) *Protein Eng. 8*, 59−62.
21. Revell, D. F., Cummings, N. F., Baker, K. C., Collins, M. E., Taylor, M. A. J., Sumner, I. G., Pickersgill, R. W., Connerton, I. F., and Goodenough, P. W. (1993) *Gene 127*, 221−225.
22. Groves, M. R., Taylor, M. A. J., Scott, M., Cummings, N. J., Pickersgill, R. W., and Jenkins, J. A. (1996) *Structure 4*, 1193−1203.
23. Baker, K. C., Taylor, M. A. J., Cummings, N. J., Tunon, M., Worboys, K. A., and Connerton, I. F. (1996) *Protein Eng. 9*, 525−529.
24. Jordan, B. R., He, T., Chow, W. G., and Anderson, J. M. (1992) *Plant Cell Environ. 15*, 91−98.
25. Dietmaier, W., Fabry, S., and Schmitt, R. (1993) *Nucleic Acids Res. 21*, 3603−3604.
26. Baker, E. N., and Dodson, E. J. (1980) *Acta Crystallogr. A 36*, 559−572.
27. Maes, D., Bouckaert, J., Poortmans, F., Wyns, L., and Looze, Y. (1996) *Biochemistry 35*, 16292−16298.
28. Kamphuis, I. G., Kalk, K. H., Swarte, M. B. A., and Drenth, J. (1984) *J. Mol. Biol. 179*, 233−256.
29. Coulombe, R., Grochulski, P., Sivaraman, J., Menard, R., Mort, J. S., and Cygler, M. (1996) *EMBO J. 15*, 5492−5503.
30. Studier, F. W., Rosenberg, A. H., Dunn, J. J., and Dubendorff, J. W. (1990) *Methods Enzymol. 185*, 60−89.
31. Laemmli, U. K. (1970) *Nature 227*, 680−685.
32. Menard, R., Khouri, H. E., Plouffe, C., Dupras, R., Ripoll, D., Vernet, T., Tessier, D. C., Laliberte, F., Thomas, D. Y., and Storer, A. C. (1990) *Biochemistry 29*, 6706−6713.
33. Sumner, I. G., Vaughan, A., Eisenthal, R., PickersgilL, R. W., Owen, A. J., and Goodenough, P. W. (1992) *Biochim. Biophys. Acta 1164*, 243−251.
34. Frohman, M. A., Dush, M. K., and Martin, G. R. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 8998−9002.
35. Turk, D., Podobnik, M., Kuhelj, R., Dolinar, M., and Turk, V. (1996) *FEBS Lett. 384*, 211−214.
36. Baker, E. N., and Drenth, J. (1987) in *Biological macromolecules and assemblies* (Jurnak, F. A., and McPherson, A., Eds.) Vol. 3, pp 314−368, John Wiley and Sons, New York.
37. Pickersgill, R. W., Rizkallah, P., Harris, G., and Goodenough, P. W. (1991) *Acta Crystallogr. B 47*, 766−771.
38. Vernet, T., Tessier, D. C., Chatellier, J., Plouffe, C., Lee, T. S., Thomas, D. Y., Storer, A. C., and Menard, R. (1995) *J. Biol. Chem. 28*, 16645−16652.
39. Lowe, G., and Yuthong, Y. (1971) *Biochem. J. 124*, 107−115, 117−122.

BI992714W