Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

[32] Families of Cysteine Peptidases

By NEIL D. RAWLINGS and ALAN J. BARRETT

Introduction

About 20 families of peptidases dependent on a cysteine residue at the active center can be recognized (Table I). As far as is known, the activity of all cysteine peptidases depends on a catalytic dyad of cysteine and histidine. The order of the cysteine and histidine residues (Cys/His or His/Cys) in the linear sequence differs between families (Table I), and this is among the lines of evidence suggesting that cysteine peptidases have had many separate evolutionary origins. As yet, tertiary structures are available for members of only two families of cysteine peptidases, so relationships between the families are not clear.

The order in which the families of cysteine peptidases are described in the present chapter is shown in Table I. The families C1, C2, and C10 can be loosely described as "papainlike," and form clan CA. Nearly half of the known families of cysteine peptidases (C3, C4, C18, C5, C6, C7, C8, C9, C16, and C21) are represented only in viruses, and the description of these forms the middle section of this chapter. The viral cysteine peptidases include families with both His/Cys and Cys/His orders of catalytic residues, and the distribution of the families among types of viruses is given in [2] in this volume. The final section of the present chapter deals with an additional eight families, three of which (C11, C15, and C20) are from bacteria, whereas the others are from eukaryotic organisms.

Papain Family (C1)

The best known family of cysteine peptidases is that of papain, often seen to be so typical of cysteine endopeptidases that a newly discovered enzyme of this type is automatically described as "papainlike" whether or not it is a homolog! The papain family contains peptidases with a wide variety of activities, including endopeptidases with broad specificity (such as papain), endopeptidases with very narrow specificity (such as glycyl endopeptidase), aminopeptidases, a dipeptidyl-peptidase, and peptidases with both endopeptidase and exopeptidase activities (such as cathepsins B and H) (Table II). There are also family members that show no catalytic activity.

Enzymes of the papain family are found in a wide variety of life forms:

TABLE I
CLANS AND FAMILIES OF CYSTEINE PEPTIDASES[a]

Clan	Family	Representative enzyme	Identified catalytic residues
CA	C1	Papain	Cys/His
CA	C2	Calpain	Cys
CA	C10	Streptopain	Cys/His
CB	C3	Polio virus picornain 3C	His/Cys
CB	C4	Tobacco etch virus NIa endopeptidase	His/Cys
—	C18	Hepatitis C virus endopeptidase 2	His/Cys
—	C5	Adenovirus endopeptidase	His/Cys
CC	C6	Tobacco etch virus HC-proteinase	Cys/His
CC	C7	Chestnut blight virus p29 endopeptidase	Cys/His
—	C8	Chestnut blight virus p48 endopeptidase	Cys/His
—	C9	Sindbis virus nsP2 endopeptidase	Cys/His
—	C16	Mouse hepatitis virus endopeptidase	Cys/His
—	C21	Turnip yellow mosaic virus endopeptidase	Cys/His
—	C11	Clostripain	Cys
—	C12	Deubiquitinating peptidase Yuh1	Cys/His
—	C19	Deubiquitinating peptidase Ubp1	Cys/His
—	C13	Hemoglobinase	—
—	C14	Interleukin 1β converting enzyme	Cys
—	C15	Pyroglutamyl-peptidase I	Cys
—	C17	Microsomal ER60 endopeptidase	—
—	C20	Type IV prepilin leader peptidase	—

[a] The order in which the families are listed here is that in which they are to be found in the text.

baculovirus,[1] eubacteria (*Porphyromonas* and *Lactococcus*), yeast,[2] and probably all protozoa, plants, and animals. In the present volume, articles [33]–[38] deal with enzymes of the papain family.

Papain homologs are generally either lysosomal (vacuolar) or secreted proteins. In plants, they are found in the vacuoles, but are also extracellular in the latex of papaya or fig, and in arthropods such as lobsters and mites they are among the digestive enzymes.[3,4] Exceptionally, bleomycin hydrolase is a cytosolic enzyme in fungi and mammals.[5]

[1] N. D. Rawlings, L. H. Pearl, and D. J. Buttle, *Biol. Chem. Hoppe-Seyler* **373,** 1211 (1992).
[2] C. Enenkel and D. H. Wolf, *J. Biol. Chem.* **268,** 7036 (1993).
[3] M. V. Laycock, R. M. MacKay, M. Di Fruscio, and J. W. Gallant, *FEBS Lett.* **292,** 115 (1991).
[4] K. Y. Chua, G. A. Stewart, W. R. Thomas, R. J. Simpson, R. J. Dilworth, T. M. Plozza, and K. J. Turner, *J. Exp. Med.* **167,** 175 (1988).
[5] S. M. Sebti, J. C. Deleon, and J. S. Lazo, *Biochemistry* **26,** 4213 (1987).

The catalytic residues of papain are Cys-25 and His-159, and these are conserved in all members of the family that are peptidases. Other residues important for catalysis include Gln-19, which helps form the "oxyanion hole," and Asn-175, which orientates the imidazolium ring of His-159 (see [33]). Bromelain is reported to lack Asn-175.[6] There is strong conservation of sequence in the vicinity of the essential Cys and His residues and Asn-175 (Fig. 1).

Members of the papain family in which Cys-25 has been replaced include the soya bean oil-body-associated protein (P34_SOYBN) (Cys25→Gly). In the *Plasmodium* surface protective protein (SERA_PLAFG) and a protein from *Schistosoma japonicum* (EMBL: X70969), Cys-25 has been converted to Ser.

Papain, like most other members of the family, shows a preference for a bulky hydrophobic residue occupying the S2 subsite, whereas cathepsin B prefers arginine (at least in small substrates). This can be explained by different amino acid side chains in the S2 binding pocket. In papain, Ser-205 lies at the bottom of the pocket, and this is replaced by Glu in cathepsin B, which is well suited to make a salt bridge with an arginine side chain.[7,8] Other homologs that contain Glu in this position include enzymes from *Brassica* (CYS4_BRANA), tomato (CYSL_LYCES), barley (CYS1_HORVU and CYS2_HORVU), *Plasmodium* (CYSP_PLAFA), the baculovirus from *Autographa* (CATV_NPVAC), lobster (CYS3_HOMAM), and rice (ORYA_ORYSA). Cysteine endopeptidases from *Entamoeba* that also prefer arginine in P2 have Ser-205 replaced by Asp (see [35]), as do stem bromelain and lobster digestive proteinase 1 (CYS1_HOMAM).

Being secreted or lysosomal enzymes, peptidases of the papain family are synthesized with signal peptides, and there are also propeptides at the N terminus. Proteolytic cleavage of the propeptides is necessary for activation of the proenzymes. The majority of the propeptides are homologous to that of papain. Although this group of propeptides does not contain any residues that are completely conserved, Glu-64, Arg-68, Phe-72, Asn-75, Asn-83, Phe-96, Asp-98, and Glu-103 (numbering according to prepropapain) are present in all but a few sequences. The first five of these amino acids are part of the "ERFNIN motif" that has been used to identify propeptides related to papain, and may have some structural significance.[9]

[6] A. Ritonja, A. D. Rowan, D. J. Buttle, N. D. Rawlings, V. Turk, and A. J. Barrett, *FEBS Lett.* **247,** 419 (1989).

[7] A. J. Barrett, M. J. H. Nicklin, and N. D. Rawlings, *Symp. Biol. Hung.* **25,** 203 (1984).

[8] D. Musil, D. Zucic, D. Turk, R. A. Engh, I. Mayr, R. Huber, T. Popovic, V. Turk, T. Towatari, N. Katunuma, and W. Bode, *EMBO J.* **10,** 2321 (1991).

[9] K. M. Karrer, S. L. Peiffer, and M. E. DiTomas, *Proc. Natl. Acad. Sci. U.S.A.* **90,** 3063 (1993).

TABLE II

PEPTIDASES OF PAPAIN, CALPAIN, AND STREPTOPAIN FAMILIES (CLAN CA)[a]

Peptidase	EC	Database code
Family C1: Papain		
Actinidain	3.4.22.14	ACTN_ACTCH
Aleurain (barley)	-	ALEU_HORVU
Allergen (*Dermatophagoides*)	-	MMAL_DERPT
Allergen (*Euroglyphus*)	-	EUM1_EURMA
Bleomycin hydrolase	-	BLMH_RABIT, BLH1_YEAST
Calotropin (*Calotropis*)	-	CAL1_CALGI
Caricain	3.4.22.30	PAP3_CARPA
Cathepsin B	3.4.22.1	CATB_*, CYSP_SCHMA, (M75822), (X73074), (X70968),[b]
Cathepsin H	3.4.22.16	CATH_*
Cathepsin L	3.4.22.15	CATL_*
Cathepsin S	3.4.22.27	CATS_*
Chymopapain	3.4.22.6	PAP2_CARPA
Cysteine aminopeptidase (*Lactococcus*)	-	(M86245)
Cysteine endopeptidases 2 and 3 (barley)	-	CYS1_HORVU, CYS2_HORVU
Cysteine endopeptidase (*Brassica napus*)	-	CYS4_BRANA
Cysteine endopeptidase (*Caenorhabditis*)	-	CYS1_CAEEL
Cysteine endopeptidases 1 and 2 (*Dictyostelium*)	-	CYS1_DICDI, CYS2_DICDI
Cysteine endopeptidase (*Entamoeba*)	-	(M27307), (M64712), (M64721), (M94163)
Cysteine endopeptidases 1 and 2 (*Haemonchus*)	-	CYS1_HAECO, CYS2_HAECO, (M80385), (M80386)
Cysteine endopeptidase (*Hemerocallis*)	-	(X74406)
Cysteine endopeptidases 1, 2 and 3 (*Homarus*)	-	CYS1_HOMAM, CYS2_HOMAM, CYS3_HOMAM
Cysteine endopeptidase (*Leishmania*)	-	LCPA_LEIME, (M97695), (Z14061)
Cysteine endopeptidase (mung bean)	-	CYSP_VIGMU
Cysteine endopeptidase (*Ostertagia*)	-	(M88505)
Cysteine endopeptidase (pea)	-	CYSP_PEA, (X66061)
Cysteine endopeptidase (*Plasmodium*)	-	CSP_PLACM, CYSP_PLAFA, (L08500), (L26362)
Cysteine protease tpr (*Porphyromonas*)	-	TPR_PORGI
Cysteine endopeptidase (*Tetrahymena*)	-	(L03212)
Cysteine endopeptidase (*Theileria*)	-	CYSP_THEPA, CYSP_THEAN
Cysteine endopeptidase (tobacco)	-	(Z13959), (Z13964)
Cysteine endopeptidase (tomato)	-	CYSL_LYCES, (Z14028)
Cysteine endopeptidase (*Trypanosoma*)	-	CYSP_TRYBR, (L25130), (M90067)
Dipeptidyl peptidase I	3.4.14.1	CATC_RAT
Endopeptidase (baculovirus of *Autographa*)	-	CATV_NPVAC
Endopeptidase EP-C1 (*Phaseolus vulgaris*)	-	CYSP_PHAVU
Glycyl endopeptidase	3.4.22.25	PAP4_CARPA
Oryzain (includes forms α, ß and γ) (rice)	-	ORYA_ORYSA, ORYB_ORYSA, ORYC_ORYSA
Papain	3.4.22.2	PAPA_CARPA
Stem bromelain	3.4.22.32	BROM_ANACO
Thaumatopain (*Thaumatococcus*)	-	THPA_THADA

TABLE II (*continued*)

Peptidase	EC	Database code
Family C2: Calpain		
Calpain (*Schistosoma*)	3.4.22.17	(M67499)
Calpain I	3.4.22.17	CAP1_CHICK, CAP1_HUMAN, CAP1_RABIT
Calpain II	3.4.22.17	CAP2_HUMAN, CAP2_RABIT
Calpain P94	3.4.22.17	CAP3_HUMAN, CAP3_RAT
Calcium-binding protein PMP41	3.4.22.17	CAP4_MOUSE
Sol gene product (*Drosophila*)	-	(M64084)
Family C10: Streptopain		
Cysteine endopeptidase (*Porphyromonas*)	-	(M83096)
Streptopain	3.4.22.10	STCP_STRPY

[a] EC is the enzyme nomenclature number (Nomenclature Committee of the International Union of Biochemistry and Molecular Biology, "Enzyme Nomenclature 1992." Academic Press, Orlando, 1992, and Supplement); where there is no EC number, none has been assigned. Literature references to the individual proteins are generally to be found in the database entries for which the codes are given; these codes are from the Swiss-Prot database, except those in parentheses, which are from the EMBL database.

[b] F. J. Cejudo, G. Murphy, C. Chinoy, and D. C. Baulcombe, *Plant J.* **2,** 937 (1992).

The papain propeptide is also homologous to mouse proteins (CT2A_MOUSE, CT2B_MOUSE) of unknown function from activated T cells.[10] These proteins are more closely related to cathepsin L and cathepsin S propeptides than to those from plants or protozoa, and thus seem to be derived from an ancestral proenzyme.

Not all of the propeptides in family C1 are related to that of papain. Among the exceptions are the propeptides of cathepsin B, dipeptidyl-peptidase I, and bleomycin hydrolase, which are also unrelated to each other. The propeptides of peptidases from the nematodes *Haemonchus* and *Ostertagia* are homologous, but unrelated to other members of family C1, and the nonenzymatic surface antigen from *Plasmodium* (SERA_PLAFA) has an N-terminal extension unrelated to any cysteine peptidase propeptide, but similar to other *Plasmodium* proteins. The enzymes with the propapain-like propeptides are seen to be more closely related to each other in overall structure than to other members of the family, and thus form a natural subfamily.

In the chymotrypsin family of serine peptidases (see [2]), propeptides show even greater variation than in the papain family, and the mechanism

[10] F. Denizot, J.-F. Brunet, P. Roustan, K. Harper, M. Suzan, M.-F. Luciani, M.-G. Mattei, and P. Golstein, *Eur. J. Immunol.* **19,** 631 (1989).

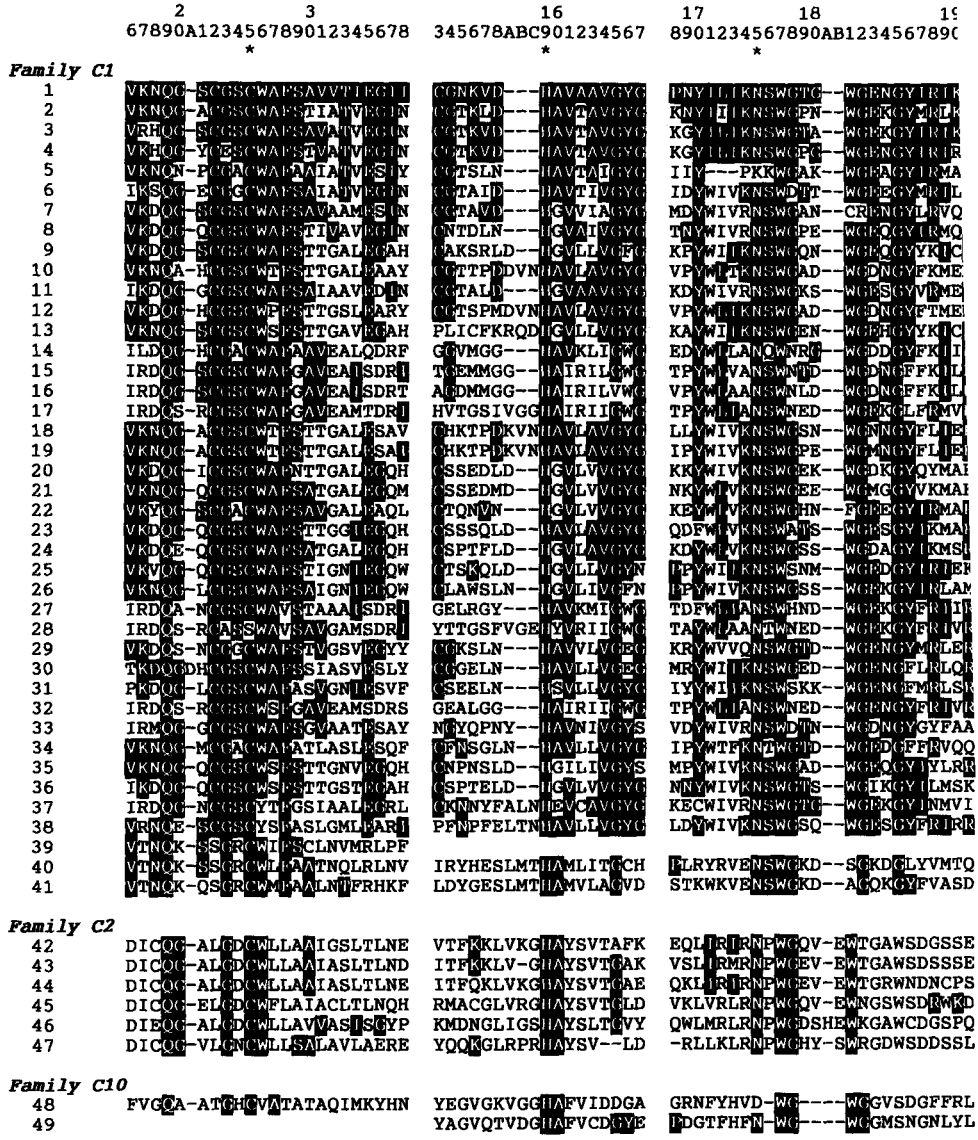

FIG. 1. Conservation of sequence around Cys-25, His-159, and Asn-175 in the papain family, and comparisons for the families of calpain and streptopain. Residues are numbered according to papain, and residues identical to those in papain are shown in white on black. Asterisks mark the catalytic residues. Key to sequences: 1, papain; 2, chymopapain; 3, caricain; 4, glycyl endopeptidase; 5, stem bromelain; 6, actinidain; 7, tomato cysteine endopeptidase; 8, mung bean cysteine endopeptidase; 9, pea cysteine endopeptidase; 10, aleurain; 11, oryzain α; 12, oryzain γ; 13, tobacco cysteine endopeptidase; 14, wheat cathepsin B

of exon shuffling has been invoked to explain this.[11] In the papain family, however, there is considerably less conservation of exon/intron junction positions or phasing,[12,13] so that exon shuffling appears to be an unlikely mechanism. Analysis of the GC content of the cathepsin H gene led Ishidoh *et al.*[12] to propose that the modern gene arose by fusion of several genes, of which that for the propeptide was one.

It has been shown by Fox *et al.*[14] that the propeptide of cathepsin B is a very potent inhibitor of the enzyme, and this presumably is part of the mechanism for the catalytic inactivity of the proenzyme. Comparable inhibitory activity of propeptides has been seen previously in a number of peptidases of other catalytic types, including members of the families of α-lytic endopeptidase (S2), subtilisin (S8), carboxypeptidase C (S10), pepsin (A1), and carboxypeptidase A (M14).

Many members of family C1 contain a proline residue at position 2 in the mature enzyme, which may serve to prevent unwanted N-terminal proteolysis. In addition to the N-terminal peptides, peptidases from tomato, pea, *Leishmania,* and *Trypanosoma* also have long C-terminal extensions.

[11] L. Patthy, *Semin. Thromb. Hemostasis* **16,** 245 (1990).
[12] K. Ishidoh, E. Kominami, N. Katunuma, and K. Suzuki, *FEBS Lett.* **253,** 103 (1989).
[13] M. Ferrara, F. Wojcik, H. Rhaissi, S. Mordier, M.-P. Roux, and D. Béchet, *FEBS Lett.* **273,** 195 (1990).
[14] T. Fox, E. De Miguel, J. S. Mort, and A. C. Storer, *Biochemistry* **31,** 12571 (1992).

(F. J. Cejudo, G. Murphy, C. Chinoy, and D. C. Baulcombe, *Plant J.* **2,** 937, (1992)]; 15, human cathepsin B; 16, mouse cathepsin B; 17, *Schistosoma japonicum* cathepsin B; 18, rat cathepsin H; 19, human cathepsin H; 20, chicken cathepsin L; 21, human cathepsin L; 22, human cathepsin S; 23, American lobster digestive cysteine endopeptidase 1; 24, American lobster digestive cysteine endopeptidase 3; 25, *Trypanosoma brucei* cysteine endopeptidase; 26, *Leishmania mexicana* cysteine endopeptidase; 27, *Haemonchus contortus* cysteine endopeptidase 1; 28, *Schistosoma japonicum* "serine endopeptidase"; 29, *Theileria parva* cysteine endopeptidase; 30, *Theileria annulata* cysteine endopeptidase; 31, *Plasmodium falciparum* cysteine endopeptidase; 32, *Schistosoma mansoni* cathepsin B; 33, house dust mite digestive cysteine endopeptidase; 34, baculovirus of alfalfa looper moth cysteine endopeptidase; 35, *Dictyostelium* cysteine endopeptidase 1; 36, *Dictyostelium* cysteine endopeptidase 2; 37, *Entamoeba histolytica* cysteine endopeptidase; 38, human dipeptidyl-peptidase I; 39, rabbit bleomycin hydrolase; 40, yeast bleomycin hydrolase; 41, *Lactococcus lactis* aminopeptidase; 42, chicken calpain; 43, human calpain I; 44, human calpain II; 45, *Schistosoma mansoni* calpain; 46, human calpain P94; 47, *sol* gene product, *Drosophila melanogaster;* 48, streptopain; 49, *Porphyromonas gingivalis* putative cysteine endopeptidase. Database codes for the above sequences may be found in Table II.

Three-dimensional structures have been eludicated for papain,[15] actinidain,[16] calotropin,[17] and cathepsin B,[8] and show bilobed molecules in which the catalytic site is located in a cleft between the lobes.

In an attempt to discover something of the evolution of the papain family, a dendrogram has been constructed using the KITSCH program of the PHYLIP suite.[18] This was calibrated by taking the divergence between cathepsin H and aleurain (the biochemically similar barley vacuolar enzyme) as 1000 million years ago. On this basis, the divergence between the *Lactococcus* aminopeptidase and the yeast bleomycin hydrolase homolog is found to be 2500 million years ago, which predates the origin of mitochondria, suggesting that papainlike cysteine peptidases were present in the organisms ancestral to eubacteria and eukaryotes, and that evolution of the papain family has not involved a horizontal transfer of genes.

The synthetic inhibitors of cysteine peptidases of the papain family have been reviewed by Rich,[19] Shaw,[20] and in articles [46]–[48] of the present volume. The most potent naturally occurring inhibitors are those of the cystatin family (see [49]).

Calpain Family (C2)

The calpain family includes the calcium-dependent cytosolic endopeptidase calpain, which is known from birds and mammals, and the product of the *sol* gene in *Drosophila*.[21] Calpain is a complex of two peptide chains. There are at least two variants of the enzyme in mammals, differing in calcium requirement (millimolar or micromolar concentrations) and with different heavy chains, but the same light chain. The heavy chain is a mosaic of four domains, in which domain 2 contains the catalytic cysteine and domain 4 binds calcium and regulates activity. The sequence of domain 2 shows some similarities to that of papain (Fig. 1), but the relationship is not statistically provable. The putative catalytic histidine residue in calpain has not been identified biochemically. Despite these uncertainties,

[15] I. G. Kamphuis, K. H. Kalk, M. B. A. Swarte, and J. Drenth, *J. Mol. Biol.* **179,** 233 (1984).
[16] E. N. Baker, *J. Mol. Biol.* **141,** 441 (1980).
[17] U. Heinemann, G. P. Pal, R. Hilgenfeld, and W. Saenger, *J. Mol. Biol.* **161,** 591 (1982).
[18] J. Felsenstein, *Evolution* **39,** 783 (1985).
[19] D. H. Rich, *in* "Proteinase Inhibitors" (A. J. Barrett and G. Salvesen, eds.), p. 153. Elsevier, Amsterdam, 1986.
[20] E. Shaw, *Adv. Enzymol.* **63,** 271 (1990).
[21] S. J. Delaney, D. C. Hayward, F. Barleben, K. F. Fischbach, and G. L. G. Miklos, *Proc. Natl. Acad. Sci. U.S.A.* **88,** 7214 (1991).

it seems very likely that the calpain and papain families are related, and belong in a single clan (CA). Domain 4 contains four calcium-binding EF-hand structures, and is homologous to sorcin, another EF-hand structure protein, although not to calmodulin. Domain 4 also is homologous to the calcium-binding domain of the calpain light chain.

The product of the *Drosophila sol* gene is also chimeric, but with a 1000-residue N-terminal domain containing six possible zinc fingers, and a C-terminal 300-residue domain of unknown function. Only the putative peptidase domain of the *sol*-encoded protein shows evidence of relationship to calpain.

Streptopain Family (C10)

Streptopain is the cysteine proteinase from a group A strain of *Streptococcus*,[22] probably *Streptococcus pyogenes*. Although there is no statistically significant relationship between the sequences of streptopain and papain, there are some similarities in structure and in properties.[23] In the primary structure, the catalytic residues of streptopain (Cys-46, the only cysteine in the protein, and His-194) occur in the same order as those in papain, and with some identical residues nearby. Asn-175 appears to be replaced by Asp (Fig. 1). The specificity of streptopain is also similar to that of papain, with preference for a hydrophobic residue at P2.[24] Streptopain is inactivated by E64 much more slowly than is papain.[25]

Streptopain is synthesized as a precursor with an 85 amino acid propeptide. The activation cleavage is at a Lys↓Gln bond.[26]

The *prtT* gene of *Porphyromonas gingivalis* is reported to encode an endopeptidase of novel sequence.[27] However, we have noticed that the strand complementary to this gene contains a sequence that is clearly homologous to that of streptopain. If some sequencing errors are assumed (as is indicated by results of analysis of the data by the NIP program of the Staden package[28]), a sequence of 69 amino acids over 50% identical to the C-terminal part of streptopain is revealed. This seems more than a coincidence, and merits further investigation.

[22] S. D. Elliott and T.-Y. Liu, this series, Vol. 19, p. 252.
[23] J. Y. Tai, A. A. Kortt, T.-Y. Liu, and S. D. Elliott, *J. Biol. Chem.* **251**, 1955 (1976).
[24] A. A. Kortt and T.-Y. Liu, *Biochemistry* **12**, 328 (1973).
[25] A. J. Barrett, A. A. Kembhavi, M. A. Brown, H. Kirschke, C. G. Knight, M. Tamai, and K. Hanada, *Biochem. J.* **201**, 189 (1982).
[26] K. Yonaha, S. D. Elliott, and T.-Y. Liu, *J. Protein Chem.* **1**, 317 (1982).
[27] J. I. Otogoto and H. K. Kuramitsu, *Infect. Immun.* **61**, 117 (1993).
[28] R. Staden, this series, Vol. 183, p. 163.

TABLE III
VIRAL CYSTEINE PEPTIDASES OF PICORNAIN 3C, TOBACCO ETCH VIRUS NIa ENDOPEPTIDASE, ADENOVIRUS ENDOPEPTIDASE, AND HEPATITIS C VIRUS ENDOPEPTIDASE 2 FAMILIES[a]

Peptidase	EC	Database code
Family C3: Picornain		
Picornain 2A	3.4.22.29	POLG_POL1M, POLG_COXA2, POLG_SVDVH, POLG_BOVEV, POLG_HRV14
Picornain 3C	3.4.22.28	POLH_POL1M, POLG_COXA2, POLG_SVDVH, POLG_BOVEV, POLG_HPAV1, POLG_HRV14, POLG_ECHO9, POLG_TMEVD
Aphthovirus endopeptidase	-	POLG_FMDVD
Cardiovirus endopeptidase	-	POLG_EMCV
Comovirus endopeptidase	-	VGNB_CPMV, (D00657)
Nepovirus endopeptidase	-	POL1_GCMV, POL1_TBRVS, (D00915)
Family C4: Tobacco etch virus NIa endopeptidase		
NIa endopeptidase	-	POLG_PPVD, POLG_PPVYN, POLG_TEV, POLG_TVMV, POLG_WMV2, POLG_OMV, (D11118), (X68509)
Family C18: Hepatitis C virus endopeptidase 2		
Endopeptidase 2 (hepatitis C virus)	-	POLG_HCV1
Family C5: Adenovirus endopeptidase		
Adenovirus endopeptidase	-	VPRT_ADEB3, VPRT_ADEB7, VPRT_ADE02, VPRT_ADE03, VPRT_ADE04, VPRT_ADE05, VPRT_ADE12, VPRT_ADE40, VPRT_ADE41, VPRT_ADEM1, (L13161), (M72715)

[a] See Table II for general explanations.

Polio Virus Picornain 3C Family (C3)

Picornains are a family of polyprotein-processing endopeptidases from single-stranded RNA viruses. Examples are known from picornaviruses such as polio virus and coxsackie virus, as well as from plant viruses such as cowpea mosaic virus (a comovirus) and grapevine fanleaf virus (a nepovirus) (Table III). Each picornavirus has two picornains (known as 2A and 3C), whereas only one is known for the plant viruses. Cardioviruses and aphthoviruses have only one picornain, the second proteinase being

unrelated.[29] Picornains have been most thoroughly characterized from polio virus (see [40]).

An early cleavage in the processing of the polio virus polyprotein is that of a Tyr↓Gly bond by which picornain 2A releases the capsid protein precursor. With 161 amino acid residues, picornain 2A is one of the smallest cysteine peptidases. Picornain 3C performs the other nine cleavages of the polyprotein, mostly at Gln↓Gly bonds.[30] The Gln↓Gly specificity is also seen in picornain 3C from encephalomyocarditis virus and Mengo virus. The specificities of picornain 3C from other picornaviruses are less strict, but the P1 residue is generally glutamine.

The three dimensional structure of a mutant form of picornain 3C from a hepatitis A virus has recently been reported.[31] In the numbering of Fig. 2, His-40 and Cys-147 (replaced by Ala in the mutant) are located appropriately to form a catalytic dyad, confirming mutational studies.[32,33] However, a catalytic triad including Glu-71 (or Asp-85 in the coxsackie virus peptidase[33]) that had been proposed earlier was not seen. The tertiary fold of picornain is similar to those of chymotrypsin and α-lytic endopeptidase, and clearly unrelated to that of papain, confirming molecular modelling studies.[34] His-40 and Cys-147 in picornain 3C are functionally equivalent to His-57 and Ser-195 in chymotrypsin, and mutants of picornain 3C with the active site Cys replaced by Ser have been shown to retain some activity.[32,33,35] This is the only instance so far discovered among the many families of peptidases in which an evolutionary relationship crosses the boundary of catalytic type, and is presumed to have arisen from a single base change that converted a catalytic Ser to a Cys residue. In the Sindbis virus, the core protein is an endopeptidase also structurally related to chymotrypsin, but with the catalytic Ser retained (Chapter [2], this volume).

Tobacco Etch Virus NIa Endopeptidase Family (C4)

Tobacco etch virus is one of the potyviruses. These are plant viruses in which the single-stranded RNA encodes a large polyprotein that is

[29] A. C. Palmenberg, G. D. Parks, D. J. Hall, R. H. Ingraham, T. W. Seng, and P. V. Pallai, *Virology* **190,** 754 (1992).
[30] K. M. Kean, N. Teterina, and M. Girard, *J. Gen. Virol.* **71,** 2553 (1990).
[31] M. Allaire, M. M. Chernaia, B. A. Malcolm, and M. N. G. James, *Nature* **369,** 72 (1994).
[32] K. Miyashita, M. Kusumi, R. Utsumi, S. Katayama, M. Noda, T. Komano, and N. Satoh, *Protein Eng.* **6,** 189 (1993).
[33] J. T. Dessens and G. P. Lomonossoff, *Virology* **184,** 738 (1991).
[34] J. F. Bazan and R. J. Fletterick, *Proc. Natl. Acad. Sci. U.S.A.* **85,** 7872 (1988).
[35] M. A. Lawson and B. L. Semler, *Proc. Natl. Acad. Sci. U.S.A.* **88,** 9919 (1991).

FIG. 2. Amino acid sequences around His-40, Glu/Asp-71, and Cys-147 in the families C3 and C4 of viral cysteine endopeptidases (clan CB), with sequences from the chymotrypsin clan (SA) for comparison. Residues are numbered according to polio virus (Mahoney strain) picornain 3C, and residues identical to those in polio virus picornain 3C are shown in white on black. Key to sequences: 1, polio virus picornain 3C; 2, coxsackie virus picornain 3C; 3, cattle enterovirus picornain 3C; 4, human rhinovirus picornain 3C; 5, echo 9 virus picornain 3C; 6, pig vesicular virus picornain 3C; 7, foot-and-mouth disease virus picornain 3C; 8, encephalomyocarditis virus picornain 3C; 9, cowpea mosaic virus picornain 3C; 10, tomato black ring virus picornain 3C; 11, polio virus picornain 2A; 12, coxsackie virus picornain 2A; 13, cattle enterovirus picornain 2A; 14, pig vesicular virus picornain 2A; 15, human rhinovirus picornain 2A; 16, tobacco etch virus NIa endopeptidase; 17, tobacco vein mottling virus NIa endopeptidase; 18, bovine chymotrypsin; 19, *Lysobacter* α-lytic endopeptidase.

processed by at least three peptidases, all of which are virally encoded. Two of these enzymes are cysteine endopeptidases—NIa endopeptidase (48 kDa) and HC-proteinase—whereas the third is a serine endopeptidase (family S30, see [2]). The HC-proteinase falls into family C6 (see below).

The catalytic cysteine and histidine residues of the NIa endopeptidase have been identified by site-directed mutagenesis, and are found to occur in the order His/Cys. A catalytic triad formed by His-234, Asp-269, and Cys-339 has been proposed.[36] Slight similarities in the sequences around the catalytic residues (Fig. 2), together with a similar specificity for

[36] W. G. Dougherty, T. D. Parks, S. M. Cary, J. F. Bazan, and R. J. Fletterick, *Virology* **172,** 302 (1989).

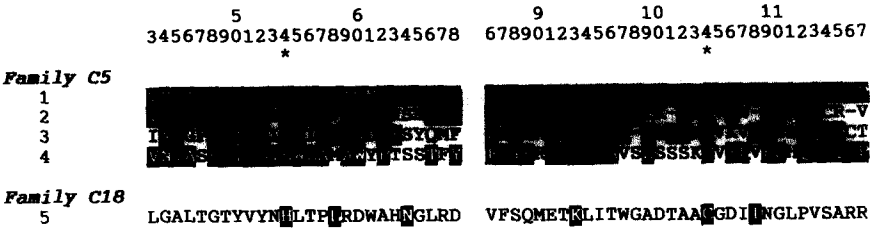

FIG. 3. Conservation of sequences around the catalytic residues in the families of adenovirus endopeptidase (C5) and hepatitis C virus endopeptidase 2 (C18). Residues are numbered according to human adenovirus type 40 endopeptidase. Residues identical to human adenovirus type 40 endopeptidase are shown in white on black. Key to sequences: 1, human adenovirus type 40 endopeptidase; 2, human adenovirus type 4 endopeptidase; 3, cattle adenovirus type 7 endopeptidase; 4, mastadenovirus mus1 endopeptidase; 5, hepatitis C virus endopeptidase 2.

Gln↓Gly cleavages, suggest that families C3 and C4 are members of a clan, CB.

The NIa endopeptidase releases itself from the viral polyprotein by cleavages at Gln↓Gly bonds at its N and C termini.[37] It then processes the viral polyprotein by cleavage at five similar sites in the C-terminal half of the polyprotein.[38]

The NIa endopeptidase is a bifunctional molecule; the C-terminal domain is the proteinase whereas the N-terminal domain functions as the VPg protein, which is attached covalently to the viral mRNA.[39]

Hepatitis C Virus Endopeptidase 2 Family (C18)

Hepatitis C virus is a flavivirus encoding a single polyprotein. As in other flaviviruses, the NS3 protein is a serine peptidase that is probably structurally related to chymotrypsin ([2], family S29). A second proteinase from this virus has been described, cleaving the Leu↓Ala bond between the NS2 and NS3 proteins.[40] The limits of this second peptidase within the polyprotein have not been completely established, but deletion studies have shown that portions of the NS2 and NS3 proteins are essential for activity. Site-directed mutagenesis has identified His-952 and Cys-993 as essential for catalytic activity, and both of these reside in the NS2 protein (Fig. 3).

[37] K. Rorrer, T. D. Parks, B. Scheffler, M. Bevan, and W. G. Dougherty, *J. Gen. Virol.* **73,** 775 (1992).

[38] C.-S. Oh and J. C. Carrington, *Virology* **173,** 692 (1989).

[39] J. F. Murphy, R. E. Rhoads, A. G. Hunt, and J. G. Shaw, *Virology* **178,** 285 (1990).

[40] A. Grakoui, D. W. McCourt, C. Wychowski, S. M. Feinstone, and C. M. Rice, *Proc. Natl. Acad. Sci. U.S.A.* **90,** 10583 (1993).

Adenovirus Endopeptidase Family (C5)

Adenoviruses are double-stranded DNA, nonenveloped viruses that cause tumors in mammals. Unlike potyviruses, the adenoviruses do not encode a polyprotein, but have different genes for different proteins, as would be expected in a DNA virus. However, several adenovirus proteins are synthesized as precursors, which have to be processed before the virion is assembled.[41] Temperature-sensitive mutants fail to process the proteins, implying that the virus encodes its own proteinase (see [41]).

Although the endopeptidase has now been sequenced, there has been confusion about its catalytic type. The current thinking is that this is a cysteine peptidase, but for many years the enzyme was considered to be a serine peptidase, on the basis of reported inhibition by standard inhibitors of such enzymes.[42] Inhibition by cysteine proteinase inhibitors has been described, and the catalytic residues are thought to be His-54 and Cys-104 (Fig. 3) (see [41]).

The endopeptidase is synthesized as an active enzyme, not needing proteolytic processing.[43] This is unusual for an endopeptidase, but is perhaps explained by the very strict specificity of the enzyme.

Tobacco Etch Virus HC-Proteinase Family (C6)

A second potyvirus proteinase, helper component proteinase (HC-Pro), is known, and is responsible for only one cleavage, that of a Gly↓Gly bond that releases the precursor of HC-Pro. Further processing of the precursor is mediated by the third potyvirus proteinase (see [2]; family S30).[44] HC-Pro is probably a two-domain protein, the C-terminal domain carrying the endopeptidase activity and the N-terminal part being the helper component required for virus transmission from plant to plant by aphids.[38]

Site-directed mutagenesis of HC-Pro from tobacco etch virus has indicated that Cys-303 and His-376 are the essential catalytic residues, and these are completely conserved in all members of the family (Fig. 4; Table IV).[38]

Chestnut Blight Virus p29 and p48 Endopeptidase Families (C7 and C8)

The chestnut blight fungus *Cryphonectria parasitica,* which causes canker on chestnut trees, exhibits reduced pathogenicity if a viruslike,

[41] H.-G. Kräusslich and E. Wimmer, *Annu. Rev. Biochem.* **57,** 701 (1988).
[42] A. R. Bhatti and J. M. Weber, *Virology* **96,** 478 (1979).
[43] A. Webster and G. Kemp, *J. Gen. Virol.* **74,** 1415 (1993).
[44] Y. Stram, A. Chetsrony, H. Karchi, M. Karchi, O. Edelbaum, E. Vardi, O. Livneh, and I. Sela, *Virus Genes* **7,** 151 (1993).

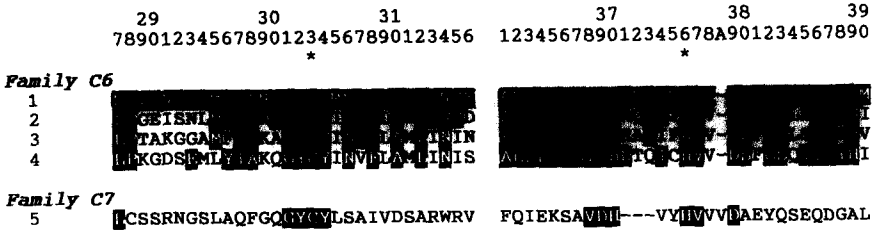

FIG. 4. Conservation of sequences around the catalytic residues in clan CC. Residues are numbered according to tobacco etch virus HC-proteinase. Residues identical to those in tobacco etch virus HC-proteinase are shown in white on black, and asterisks indicate the catalytic residues identified in both families. Key to sequences: 1, tobacco etch virus HC-proteinase; 2, tobacco vein mottling virus HC-proteinase; 3, plum pox virus HC-proteinase; 4, potato virus Y HC-proteinase; 5, chestnut blight virus p29 endopeptidase.

double-stranded RNA element is present in the fungal hyphae. During anastomosis (joining) of hyphae, this hypovirulence particle is transmissible to other fungal strains lacking the particle. The hypovirulence particle encodes two polyproteins, both of which are proteolytically processed. The processing of the smaller polyprotein yields two proteins of 29 and 48 kDa, known as p29 and p48, respectively. The p29 component has been shown to be a proteinase, performing the single cleavage in the polyprotein at a Gly↓Gly bond.[45] Site-directed mutagenesis has identified the possible catalytic residues in the p29 protein as Cys-162 and His-215. The postulated active site residues in the endopeptidases of family C6 (above) occur within similar motifs (Fig. 4), and a common ancestor has been postulated.[46] This suggests that families C6 and C7 comprise a clan (CC).

The second, larger polyprotein is also proteolytically processed, at a Gly↓Ala bond,[47] and the proteinase responsible has been identified as the p48 protein. The catalytic residues have been identified by site-directed mutagenesis as Cys-341 and His-388 (Fig. 5). The p29 and p48 proteins show no sequence similarity, and are considered here to be representatives of separate families.

Sindbis Virus nsP2 Endopeptidase Family (C9)

Togaviruses such as the Sindbis virus produce two polyproteins, and these are processed by a combination of virally encoded peptidases and cellular peptidases. The p130 polyprotein is processed by a serine endo-

[45] G. H. Choi, R. Shapira, and D. L. Nuss, *Proc. Natl. Acad. Sci. U.S.A.* **88,** 1167 (1991).
[46] G. H. Choi, D. M. Pawlyk, and D. L. Nuss, *Virology* **183,** 747 (1991).
[47] R. Shapira and D. L. Nuss, *J. Biol. Chem.* **266,** 19419 (1991).

TABLE IV
OTHER CYSTEINE PEPTIDASES[a]

Peptidase	EC	Database code
Family C6: Tobacco etch virus HC-proteinase		
HC-Proteinase	-	POLG_PPVD, POLG_PVYN, POLG_TEV, POLG_TVMV
Family C7: Chestnut blight virus p29 endopeptidase		
p29 Endopeptidase	-	(M57938)
Family C8: Chestnut blight virus p48 endopeptidase		
p48 Endopeptidase	-	(M57938)
Family C9: Sindbis virus nsP2 endopeptidase		
Togavirus cysteine endopeptidase	-	POLN_SINDV, POLN_RRVN, POLN_SFV,POLN_ONNVG, V180_CGMVS, (J02246), (L01443),(X63135)
Family C16: Mouse hepatitis virus endopeptidase		
Mouse hepatitis virus endopeptidase	-	RRPA_CVMJH
Avian infectious bronchitis virus endopeptidase	-	VGF1_IBVB
Family C21: Turnip yellow mosaic virus endopeptidase		
Turnip yellow mosaic virus endopeptidase	-	POLR_TYMV
Family C11: Clostripain		
α-Clostripain	3.4.22.8	CLOS_CLOHI
Family C12: De-ubiquitinating peptidase Yuh1		
Ubiquitin carboxyl-terminal hydrolase	-	UBL1_*, UBL3_HUMAN, {8676}
Family C19: De-ubiquinating peptidase Ubp1		
Deubiquinating enzyme (DOA4 protein)	-	SSV7_YEAST, (U02518)
Ubiquitin-specific processing peptidase 1	-	UBP1_YEAST
Ubiquitin-specific processing peptidase 2	-	UBP2_YEAST
Ubiquitin-specific processing peptidase 3	-	UBP3_YEAST
tre oncogene protein (human)	-	(X63547)
unp protein (mouse)	-	(L00681)
Family C13: Hemoglobinase		
Hemoglobinase (*Schistosoma*)	-	HGLB_SCHMA, (X70967)
Legumain (jack bean)	-	[b]
Family C14: Interleukin-1β converting enzyme		
Interleukin-1β converting enzyme	-	I1BC_*
Family C15: Pyroglutamyl-peptidase I		
Pyroglutamyl-peptidase I	3.4.19.3	PCP_*, (X75919)

TABLE IV (*continued*)

Peptidase	EC	Database code
Family C17: Microsomal ER60 endopeptidase		
Microsomal ER60 protein	-	ER60_*
Family C20: Type IV-prepilin leader peptidase		
Prepilin leader peptidase	-	TCPJ_VIBCH, COMC_BACSU, XCPA_PSEAE, PULO_KLEPN, HOPO_ECOLI, (L11715)

[a] See Table II for general explanations.
[b] Y. Abe, K. Shirane, H. Yokosawa, H. Matsushita, M. Mitta, I. Kato, and S. Ishii, *J. Biol. Chem.* **268,** 3525 (1993).

peptidase (see [2], family S3). The p270 polyprotein contains nonstructural proteins, among which nsP2 is the cysteine endopeptidase that processes the polyprotein. The active site residues have been identified by site-directed mutagenesis as Cys-481 and His-558 in the Sindbis virus (Fig. 5).[48]

Protein nsP2 is a bifunctional, mosaic molecule, with the cysteine endopeptidase restricted to the C-terminal domain. The N-terminal domain is probably involved in RNA-binding during replication and is homologous to some plant virus proteins, such as the 180 kDa protein of cucumber green mottle mosaic virus, which is an unrelated tobamovirus. There is no sequence relationship to the endopeptidase domain.

Mouse Hepatitis Virus Endopeptidase Family (C16)

Mouse hepatitis is caused by one of the coronaviruses, which are single-stranded RNA viruses that encode several polyproteins. The polyprotein encoded by gene A of the virus is known to be autolytically processed to release the 28-kDa N-terminal p28 protein.[49] Site-directed mutagenesis has identified Cys-1137 and His-1288 as the catalytic dyad in this peptidase (Fig. 5).[50]

[48] E. G. Strauss, R. J. De Groot, R. Levinson, and J. H. Strauss, *Virology* **191,** 932 (1992).
[49] S. C. Baker, C. K. Shieh, L. H. Soe, M. F. Chang, D. M. Vannier, and M. M. C. Lai, *J. Virol.* **63,** 3693 (1989).
[50] S. C. Baker, K. Yokomori, S. Dong, R. Carlisle, A. E. Gorbalenya, E. V. Koonin, and M. M. C. Lai, *J. Virol.* **67,** 6056 (1993).

FIG. 5. Comparison of sequences in the vicinity of the catalytic residues of cysteine peptidases from families C8, C9, C11, C14, C15, C16, C17, C20, and C21 with those in papain. Residues identical to those in papain are shown in white on black. Key to sequences: 1, papain; 2, chestnut blight virus p48 endopeptidase; 3, Sindbis virus nsP2 endopeptidase; 4, clostripain; 5, human interleukin 1β converting enzyme; 6, *Bacillus subtilis* pyroglutamyl-peptidase I; 7, mouse hepatitis virus endopeptidase; 8, rat ER60 endopeptidase repeat 1; 9, rat ER60 endopeptidase repeat 2; 10, *Pseudomonas aeruginosa* type IV prepilin leader peptidase; 11, turnip yellow mosaic virus endopeptidase. Asterisks indicate identified catalytic residues (predictions only for ER60 endopeptidase, and His in pyroglutamyl-peptidase I).

Turnip Yellow Mosaic Virus Endopeptidase Family (C21)

Turnip yellow mosaic virus is a single-stranded RNA virus in which two polyproteins are encoded. The larger polyprotein (206 kDa) includes an endopeptidase that cleaves the 206-kDa polyprotein at a single bond to release an N-terminal 150-kDa protein, which contains a helicase, and a C-terminal 70-kDa protein, which includes a polymerase. The endopeptidase has been delimited to residues 731–885 (within the 150-kDa protein), and site-specific mutagenesis has identified Cys-783 and His-869 as the catalytic dyad (Fig. 5).[51] Homologous sequences occur in the ononis yellow

[51] K. L. Bransom and T. W. Dreher, *Virology* **198**, 148 (1994).

mosaic virus (POLR_OYMV), the eggplant mosaic virus (POLR_EPMV), the kennedya yellow mosaic virus (EMBL: D00637), and the erysimum latent virus.[52]

Clostripain Family (C11)

Clostripain, from the anaerobic bacterium *Clostridium histolyticum*, is a cysteine endopeptidase with strict specificity for the cleavage of arginyl bonds, only rarely cleaving lysyl bonds.[53] The two-chain enzyme is synthesized as a precursor protein that is processed to yield (from the N terminus) a 50-residue propeptide, the light chain of the mature enzyme, a nonapeptide, and the heavy chain. The C-terminal residues of both the light chain and nonapeptide linker are arginine, suggesting that their processing is autolytic, but the activation cleavage at the N terminus is of a Lys↓Asn bond, and is probably mediated by another enzyme.[54]

Clostripain differs from the enzymes of the papain family both in its calcium dependence and its inhibition characteristics. The enzyme is more rapidly inactivated by iodoacetamide than by iodoacetate, and E64 gives only reversible inhibition.[55] The active site cysteine has been identified as Cys-181,[56] and occurs in a sequence unrelated to those in family C1 (Fig. 5). The dependence of catalytic activity on a histidine residue has been demonstrated by use of diethyl pyrocarbonate,[55] but the histidine remains unidentified.

Deubiquitinating peptidase Yuh1 Family (C12)

Ubiquitin is a protein of 76 amino acids that exists in eukaryotic cells and commonly occurs conjugated to ubiquitin or other proteins through the carboxyl group of its C-terminal glycine residue. The link may be to the N terminus of another polypeptide or to the ε-amino group of a lysine, in which case an isopeptide bond is formed.[57] Ubiquitination can act as a signal to mark proteins for rapid degradation, or may have a chaperone function in the assembly of oligomeric proteins and ribosomes. Whatever the function of the attachment of ubiquitin, the ubiquitin molecule is

[52] P. Srifah, P. Keese, G. Weller, and A. Gibbs, *J. Gen. Virol.* **73,** 1437 (1992).
[53] B. Keil, "Specificity of Proteolysis." Springer-Verlag, Berlin, 1992.
[54] H. Dargatz, T. Diefenthal, V. Witte, G. Reipen, and D. Von Wettstein, *Mol. Gen. Genet.* **240,** 140 (1993).
[55] A. A. Kembhavi, D. J. Buttle, P. Rauber, and A. J. Barrett, *FEBS Lett.* **283,** 277 (1991).
[56] A.-M. Gilles, A. De Wolf, and B. Keil, *Eur. J. Biochem.* **130,** 473 (1983).
[57] K. D. Wilkinson, K. Lee, S. Deshpande, P. Duerksen-Hughes, J. M. Boss, and J. Pohl, *Science* **246,** 670 (1989).

eventually released by the action of a peptidase hydrolyzing the glycyl bond at the C terminus, and is recycled.

There are a number of distinct deubiquitinating peptidases,[58] and they belong to at least two families. Nevertheless, they have important properties in common. They are activated by thiol compounds[59] and are inhibited by thiol-blocking reagents, but not by inhibitors of other classes of peptidases. They are also inhibited by ubiquitin aldehyde.[60] In these respects, the deubiquitinating enzymes have properties expected of cysteine peptidases. Also, they are all specific for the C terminus of ubiquitin, but seem to show almost no selectivity for residues on the prime side of the scissile bond (only proline being unacceptable).[61] A generic assay for the deubiquitinating enzymes is made with ubiquitin ethyl ester as substrate.[62]

The deubiquitinating peptidases generally fall into one of two molecular mass ranges, 20–30 or 100–200 kDa.[58] The small enzymes are those of the Yuh1 family (C12), and the large ones belong to the Ubp1 family (C19).

Family C12 comprises the product of the yeast *YUH1* gene and its mammalian counterparts. The Yuh1 protein is known to cleave only short ubiquitin conjugates, being inactive against ubiquitinated β-galactosidase.[63] One mammalian homolog, known as ubiquitin conjugate hydrolase (Uch) isozyme L1 or PGP 9.5, is among the most abundant proteins of the brain (1–5% of total soluble protein), and has been described as "neurone-specific", although it does occur in other tissues at lower concentrations. The other, Uch-L3, is the predominant form of ubiquitin C-terminal hydrolase in bovine thymus.[57]

The alignment of sequences[57] shows only one cysteine residue conserved among the members of this family, Cys-100 (Fig. 6). There are two conserved histidine residues, His-107 and His-181, the second of which has the spacing from the Cys that is more consistent with participation in a catalytic dyad.

Deubiquitinating Peptidase Ubp1 Family (C19)

Work on the high molecular mass deubiquitinating peptidases of yeast led to the sequencing of Ubp1,[64] Ubp2, and Ubp3.[61] The Ubp2 protein can

[58] A. N. Mayer and K. D. Wilkinson, *Biochemistry* **28,** 166 (1989).
[59] C. M. Pickart and I. A. Rose, *J. Biol. Chem.* **261,** 10210 (1986).
[60] A. Hershko and I. A. Rose, *Proc. Natl. Acad. Sci. U.S.A.* **84,** 1829 (1987).
[61] R. T. Baker, J. W. Tobias, and A. Varshavsky, *J. Biol. Chem.* **267,** 23364 (1992).
[62] K. D. Wilkinson, M. J. Cox, A. N. Mayer, and T. Frey, *Biochemistry* **25,** 6644 (1986).
[63] H. I. Miller, W. J. Henzel, J. B. Ridgway, W. J. Kuang, V. Chisholm, and C. C. Liu, *Bio/Technology* **7,** 698 (1989).
[64] J. W. Tobias and A. Varshavsky, *J. Biol. Chem.* **266,** 12021 (1991).

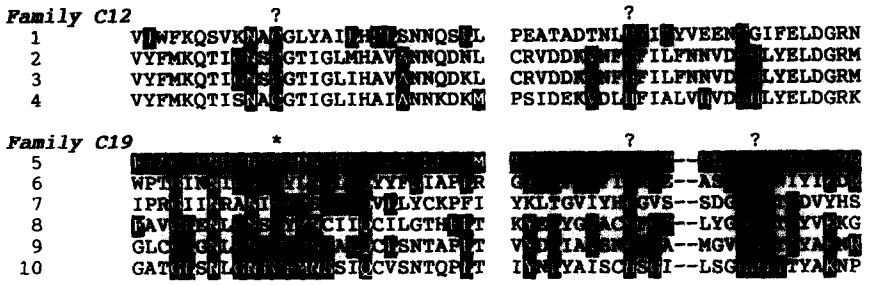

FIG. 6. Conservation of sequences around the potential catalytic cysteine and histidine residues of deubiquitinating peptidases. Residues identical to yeast deubiquitinating peptidase Ubp1 are shown in white on black, and the asterisk marks the cysteine residue shown to be involved in the activity of Doa4. Key to sequences: 1, yeast deubiquitinating peptidase Uch; 2, rat deubiquitinating peptidase Uch-L1; 3, human deubiquitinating peptidase Uch-L1; 4, human deubiquitinating peptidase Uch-L3; 5, yeast deubiquitinating peptidase Ubp1; 6, yeast deubiquitinating peptidase Ubp2; 7, yeast deubiquitinating peptidase Ubp3; 8, yeast deubiquitinating enzyme Doa4; 9, mouse protooncoprotein Unp; 10, human *tre* oncogene protein.

deubiquitinate any size of ubiquitinated protein, including polyubiquitin, whereas Ubp1 and Ubp3 do not act on polyubiquitin.[61] Subsequently, a fourth homolog (Doa4) was discovered in yeast. Doa4 appears to be involved in the later stages of ubiquitin recycling, perhaps working in conjunction with the 26S ubiquitin–conjugate-degrading enzyme.[65]

Site-directed mutagenesis has identified Cys-571 as catalytically important in Doa4,[65] and the closest similarities between the sequences of these proteins occur in the vicinity of this residue and a more C-terminal region containing two histidines, either of which may play a role in activity (see Fig. 6).

Mammalian proteins in the family include the products of the human *tre-2* oncogene, which is also a deubiquitinating enzyme,[65] and the mouse *unp* gene (Table IV). Also related is dog mucin (EMBL: L03387), which overlaps only the C-terminal half of the Doa4 protein and does not contain the active site Cys.

Hemoglobinase Family (C13)

Schistosoma mansoni, a blood fluke, is a human parasite causing schistosomiasis. The adult worms burrow through the skin and take up residence in the bloodstream, where they feed on hemoglobin. There are two cysteine endopeptidases in the parasite digestive tract, one a cathepsin

[65] F. R. Papa and M. Hochstrasser, *Nature (London)* **366**, 313 (1993).

B-like enzyme of family C1, and a second, termed hemoglobinase,[66] which shows no homology to papain. Hemoglobinase has proved difficult to isolate free from the cathepsin B, and consequently rather little is known about its catalytic activity, although it is trapped by α_2-macroglobulin and is inactivated by Z-Tyr-Ala-CHN_2 (C. L. Chappell, personal communication, 1991). An attempt at expressing hemoglobinase in *Escherichia coli* did not produce a product with detectable peptidase activity,[67] and it may be that the enzyme acts synergistically with the cathepsin B. A second clone has recently been sequenced from *Schistosoma japonicum* (Table IV).

It has been discovered (see [42]) that legumain, an atypical cysteine endopeptidase from legume seeds, is a homolog of hemoglobinase. Legumain is a very strict asparaginyl endopeptidase that acts on both small substrates and proteins. This enzyme has been assigned two separate functions in the legume seed: first, posttranslational splicing of seed proteins, including concanavalin A during maturation of the seed, and second, a role in the degradation of seed proteins during germination (see Ref. 68 and [42]). It was first suggested by Csoma and Polgár[69] that the bean asparaginyl endopeptidase might be a member of a family of cysteine peptidases different from that of papain, because it has the unusual characteristic of reacting more rapidly with iodoacetamide than with iodoacetate. No catalytic residues have been identified in the hemoglobinase family. A sequence tag from *Arabidopsis thaliana* is homologous to the N terminus of hemoglobinase (EMBL: Z17798).

Interleukin 1β Converting Enzyme Family (C14)

The precursor of the cytokine, interleukin 1β, is a 31- to 33-kDa protein synthesized by monocytes, and the active 17.5-kDa molecule is released by cleavage of an Asp↓Ala bond. The interleukin 1α precursor is not processed by this enzyme.[70]

The interleukin 1β converting enzyme (ICE) has been purified, cloned, and sequenced (see [43]). The enzyme shows strict specificity for the cleavage of aspartyl bonds.

[66] A. H. Davis, J. Nanduri, and D. C. Watson, *J. Biol. Chem.* **262,** 12851 (1987).
[67] B. Götz and M.-O. Klinkert, *Biochem. J.* **290,** 801 (1993).
[68] A. A. Kembhavi, D. J. Buttle, C. G. Knight, and A. J. Barrett, *Arch. Biochem. Biophys.* **303,** 208 (1993).
[69] C. Csoma and L. Polgár, *Biochem. J.* **222,** 769 (1984).
[70] N. A. Thornberry, H. G. Bull, J. R. Calaycay, K. T. Chapman, A. D. Howard, M. J. Kostura, D. K. Miller, S. M. Molineaux, J. R. Weidner, J. Aunins, K. O. Elliston, J. M. Ayala, F. J. Casano, J. Chin, G.J.-F. Ding, L. A. Egger, E. P. Gaffney, G. Limjuco, O. C. Palyha, S. M. Raju, A. M. Rolando, J. P. Salley, T.-T. Yamin, and M. J. Tocci, *Nature (London)* **356,** 768 (1992).

ICE is synthesized as a 45-kDa precursor. Four peptide bond cleavages occur in the formation of the active enzyme, which is a heterodimer of a heavy chain (22 kDa) and a light chain (10 kDa). All of the cleavages in the maturation of ICE are of aspartyl bonds, and it is very probable that they are mediated by preexisting molecules of the active enzyme.

Although ICE has little reactivity with most inhibitors that are effective against enzymes of the papain family, it is inhibited by aldehyde, diazomethane, and (acyloxy)methane compounds of appropriate structure (see [43]). ICE is also inhibited by a protein inhibitor of the serpin family that is encoded by the cowpox virus.[71]

A broader physiological role for ICE has been suggested by the discovery that the product of the *ced-3* gene involved in programmed cell death in *Caenorhabditis elegans* is a homolog of ICE,[72] and that either this protein or ICE can cause programmed cell death in transfected cells.[73]

The catalytic residue of ICE has been identified as Cys-285 (Fig. 5), and the only His residues conserved in the family are N-terminal to this.

Pyroglutamyl-peptidase I Family (C15)

Pyroglutamyl-peptidase I removes an N-terminal pyroglutamyl residue from a polypeptide. Sequences are known from eubacteria only, although a mammalian enzyme with similar activity exists.[74] The enzyme from *Bacillus amyloliquefaciens* (23-kDa monomer mass) is inhibited by diazomethane and chloromethane inhibitors,[75] and site-directed mutagenesis has identified Cys-144 as the catalytic residue.[76] There are two conserved His residues, His-168 and His-215, the former the more likely to be part of a catalytic dyad because the latter forms the C terminus (Fig. 5).

Microsomal ER60 Endopeptidase Family (C17)

ER60 endopeptidase is a protein from the mammalian rough endoplasmic reticulum that is believed to degrade other endoplasmic proteins,

[71] C. A. Ray, R. A. Black, S. R. Kronheim, T. A. Greenstreet, P. R. Sleath, G. S. Salvesen, and D. J. Pickup, *Cell (Cambridge, Mass.)* **69,** 597 (1992).

[72] J. Yuan, S. Shaham, S. Ledoux, H. M. Ellis, and H. R. Horvitz, *Cell (Cambridge, Mass.)* **75,** 641 (1993).

[73] M. Miura, H. Zhu, R. Rotello, E. A. Hartwieg and J. Yuan, *Cell (Cambridge, Mass.)* **75,** 653 (1993).

[74] J. K. McDonald and A. J. Barrett, "Mammalian Proteases: A Glossary and Bibliography. Volume 2: Exopeptidases." Academic Press, London, 1986.

[75] K. Fujiwara, E. Matsumoto, T. Kitagawa, and D. Tsuru, *Biochim. Biophys. Acta* **702,** 149 (1980).

[76] T. Yoshimoto, T. Shimoda, A. Kitazono, T. Kabashima, K. Ito, and D. Tsuru, *J. Biochem. (Tokyo)* **113,** 67 (1993).

such as protein disulfide-isomerase and calreticulin.[77] Degradation of these proteins is inhibited by leupeptin and E64. These compounds also inhibit calpain, as well as the lysosomal cysteine endopeptidases that pass through the endoplasmic reticulum in the course of their biosynthesis. However, the ER60 activity has a neutral pH optimum, and is unaffected by EGTA (which inhibits calpain) and by other inhibitors that might affect the cathepsins.

The sequence reported for ER60 endopeptidase is that of a member of a large family of proteins related to thioredoxin. Thioredoxin acts as a protein disulfide oxidoreductase, and catalyzes dithiol–disulfide exchanges. The three-dimensional structure of *E. coli* thioredoxin has been resolved,[78] and a catalytic mechanism has been proposed in which the proximity of two Cys residues in a sequence -Cys-Xaa-Xaa-Cys- permits the reversible formation of a disulfide bond. The ER60 endopeptidase contains two thioredoxin-like sequences separated by an unrelated segment of 240 residues. The Cys residues involved in the catalytic activity of thioredoxin are conserved in the ER60 endopeptidase. It is notable that a -Cys-Xaa-Xaa-Cys- sequence contains the catalytic Cys in papain, and is suggested to contain that of the type IV prepilin leader peptidase also (Fig. 5).

Type IV Prepilin Leader Peptidase Family (C20)

Some of the biosynthetic precursors of proteins secreted by bacteria have a special leader peptide that is removed by type IV prepilin leader peptidase. Pili are hairlike structures that occur on the surface of bacteria, and each is assembled from one or more protein subunits known as pilins. Type IV pili are found in the gram-negative pathogens, including *Pseudomonas aeruginosa* and *Neisseria gonorrhoeae,* and are thought to be responsible for attaching the organism to the surface of host epithelial cells. The prepilin subunits are synthesized with six- to eight-residue leader peptides that contain charged amino acids, unlike the leader peptides removed by leader peptidase 1. All mature type IV pilins have a methylated N-terminal Phe residue. N-Methylation is unusual in bacteria, but is a prerequisite for assembly of pilin subunits.[79]

The type IV prepilin leader peptidase is found in the inner membrane, and cleavage and methylation of the pilin precursors appear to occur on

[77] R. Urade, M. Nasu, T. Moriyama, K. Wada, and M. Kito, *J. Biol. Chem.* **267,** 15152 (1992).
[78] H. J. Dyson, G. P. Gippert, D. A. Case, A. Holmgren, and P. E. Wright, *Biochemistry* **29,** 4129 (1990).
[79] M. S. Strom and S. Lory, *J. Biol. Chem.* **266,** 1656 (1991).

the cytoplasmic face of the membrane.[80] In strains of *Pseudomonas* in which activity is deficient, there is accumulation not only of prepilins, but also of proteins that are normally secreted, such as exotoxin A, phospholipase C, alkaline phosphatase, and pseudolysin.[81] The accumulation of these proteins occurs because type IV prepilin leader peptidase is required to process proteins that mediate secretion of the affected proteins.[82]

The prepilin leader peptidase cleaves Gly↓Phe bonds, and specificity is for residues in the prime sites. Thus, the peptidase recognizes a -Gly↓Phe-Thr-Leu- (or -Ile-) -Glu consensus in which the Gly in P1 is obligatory.[79] The peptidase is bifunctional, however, for not only is the leader peptide removed, but the newly exposed N terminus is methylated.[83]

Type IV prepilin leader peptidase is a product of the *pilD* gene in *Pseudomonas*, and homologs are known from other bacteria (Table IV). The enzyme is sensitive to thiol-blocking reagents such as *N*-ethylmaleimide, iodoacetamide, and *p*-mercuribenzoate, and the inhibition by *p*-mercuribenzoate is reversible by dithiothreitol. Site-directed mutagenesis has implicated four Cys residues in both peptidase and methylase activities to differing extents,[80] although one of these (Cys-97) is naturally replaced by Ser in the *Klebsiella* homolog.[84] However, there is no conserved His residue in an alignment of the sequences of members of the family.[80]

Conclusions

In the past, the cysteine proteinases of the papain family have been given so much more attention than the others that it has been easy to assume that all such enzymes were essentially "papainlike", but it is now abundantly clear that that is not the case. In the near future, this point seems likely to be brought into sharp focus by the elucidation of three-dimensional structures for cysteine peptidases of other families. In reviewing the structures and mechanisms of the catalytic sites of peptidases generally, James[85] has observed that for serine peptidases (e.g., chymotrypsin and subtilisin), aspartic endopeptidases (e.g., pepsin), and metallopeptidases (e.g., carboxypeptidase A and thermolysin), the attack on the

[80] M. S. Strom, P. Bergman, and S. Lory, *J. Biol. Chem.* **268,** 15788 (1993).
[81] D. N. Nunn and S. Lory, *Proc. Natl. Acad. Sci. U.S.A.* **88,** 3281 (1991).
[82] M. S. Strom and S. Lory, *J. Bacteriol.* **174,** 7345 (1992).
[83] M. S. Strom, D. N. Nunn, and S. Lory, *Proc. Natl. Acad. Sci. U.S.A.* **90,** 2404 (1993).
[84] M. R. Kaufman, J. M. Seyer, and R. K. Taylor, *Genes Dev.* **5,** 1834 (1991).
[85] M. N. G. James, *in* "Proteolysis and Protein Turnover" (J. S. Bond and A. J. Barrett, eds.), p. 1. Portland Press, London, 1993.

peptide bond that leads to hydrolysis of the substrate is from the *re* face of the bond. In contrast, the active site of papain attacks its substrate from the *si* face. As James points out, it will be of the greatest interest to learn, as more structures become available, whether other families of cysteine peptidases share this unusual characteristic of papain, or resemble the majority of peptidases of other catalytic types. It may well be that some of the other families will have geometries more like that of chymotrypsin. As we have seen, the picornains have been suggested to be distantly related to the chymotrypsin family. Also, Salvesen[86] has pointed out that the reactivity of interleukin 1β converting enzyme with a serpin suggests that it too may have geometry more like that of chymotrypsin than that of papain.

[86] G. Salvesen, *in* "Proteolysis and Protein Turnover" (J. S. Bond and A. J. Barrett, eds.), p. 57. Portland Press, London, 1993.

[33] Catalytic Mechanism in Papain Family of Cysteine Peptidases

By ANDREW C. STORER and ROBERT MÉNARD

Introduction

Cysteine peptidases are a class of enzymes that have been widely studied over the years. The overall principles of substrate recognition, catalysis, and inhibition are now reasonably well documented. However, the molecular basis of these properties is still not clearly established. For example, although it has formed the subject of numerous reviews (see Refs. 1–5), the mechanism by which cysteine peptidases hydrolyze their substrates is still poorly defined at the atomic level. By far the bulk of the literature reports dealing with enzymes in this class describe results

[1] L. Polgar and P. Halasz, *Biochem. J.* **207,** 1 (1982).
[2] E. N. Baker and J. Drenth, *in* "Biological Macromolecules and Assemblies, Volume 3—Active Sites of Enzymes" (F. A. Jurnak and A. McPherson, eds.), p. 314. Wiley, New York, 1987.
[3] K. Brocklehurst, *in* "Enzyme Mechanisms" (M. I. Page and A. Williams, eds.), p. 140. Royal Society of Chemistry, London, 1987.
[4] K. Brocklehurst, F. Willenbrock, and E. Salih, *in* "Hydrolytic Enzymes" (A. Neuberger and K. Brocklehurst, eds.), p. 39. Elsevier, Amsterdam, 1987.
[5] L. Polgar, *in* "Mechanisms of Protease Action," p. 123. CRC Press, Boca Raton, Florida, 1990.