



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.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.

Introduction: The Clans and Families of Cysteine Peptidases

Peptidases in which the nucleophile that attacks the scissile peptide bond is the sulfhydryl group of a cysteine residue are known as cysteine-type peptidases. The catalytic mechanism is similar to that of the serine-type peptidases in that a nucleophile and a proton donor/general base are required, and the proton donor in all cysteine peptidases in which it has been identified is a histidine residue, as in the majority of serine-type peptidases. In some families only the dyad of Cys and His seems to be essential for catalysis, whereas in others there is evidence that a third residue is required to orientate the imidazolium ring of the His (a role analogous to that of the essential aspartate in some serine peptidases). The catalytic mechanisms of cysteine peptidases are described fully in Chapter 405.

The clans and families of cysteine peptidases are listed in [Table 404.1](#).

Clan CA

The first clearly recognized cysteine peptidase was papain (Chapter 418), and this forms the foundation of *clan CA*. The crystal structure ([Figure 404.1](#)) shows two structural domains separated by the active site cleft. The N-terminal domain consists largely of a bundle of α -helices whereas the C-terminal domain contains a β -barrel. A long helix runs through the middle of the molecule, and the catalytic cysteine is at the start of this.

Clan CA contains over 20 families (see [Table 404.1](#)). Almost all of these have been brought into the clan on the basis of structures that are known from crystallographic data to be similar. The other families are assigned to clan CA because they contain similar sequence motifs around the catalytic residues ([Figures 404.2–404.7](#)). Besides Cys158 and His292 of the catalytic dyad, two other residues are important for catalysis in papain and its relatives (numbering as in [MEROPS database](#)). These are a Gln152 that helps in the formation of the ‘oxyanion hole’, an electrophilic center that stabilizes the tetrahedral intermediate, and Asn308 (sometimes Asp in families C12, C28 and C39), which is thought to orientate the imidazolium

ring of the catalytic His (Chapter 405). The order of these residues in the sequence is Gln, Cys, His, Asn/Asp, and in mature papain, the numbering is Gln19, Cys25, His159 and Asn175. In family C19, asparagine serves the same function as Gln152. No equivalent of Gln152 has been identified in the foot-and-mouth disease virus L-endopeptidase. The catalytic cysteine is usually followed by an aromatic, hydrophobic amino acid, but glycine occupies this position in some peptidases of family C2 and all those of family C12.

[Figure 404.8](#) shows secondary structure diagrams of example peptidases from the clan. All the examples shown are α/β proteins, and all have the nucleophilic Cys at the start of a helix and the catalytic His at the start of a β -sheet.

Inhibitors are of some use in identifying peptidases of clan CA. Many members of the clan are irreversibly inhibited by compound E-64, but reversible inhibition by the arginine-like side chain of E-64 is not significant and should be ignored. Proteins of the cystatin family also inhibit many peptidases in clan CA, but it should be noted that some cystatins have a second reactive site that can inhibit legumain, a clan CD peptidase [1], and some cystatins can inhibit metallopeptidases [2].

Family C1. The [evolutionary tree](#) (see the *MEROPS* database) shows that the family is divided into two subfamilies, C1A (or papain subfamily) and C1B (or bleomycin hydrolase subfamily). It may well be that a peptidase of family C1 was present in the universal ancestor of all organisms, and that this evolved into a subfamily C1A-type protein in early archaea and a subfamily C1B-type protein in early bacteria. Further divergence would have occurred in the separate groups of organisms, and horizontal transfers can explain the presence of a C1A homolog in *Clostridium* and of bleomycin hydrolase in eukaryotes.

Subfamily C1A is the larger of these subunits, and consists of secreted and lysosomal enzymes. Subfamily C1A contains mainly endopeptidases from DNA viruses, protozoa, plants and animals, and exopeptidases from gram-positive bacteria, fungi and animals. Members

TABLE 404.1 The clans and families of cysteine peptidases

Clan CA	Catalytic residues: Gln, Cys, His, Asn (or Asp)	Fold: Helices + β -barrel
Family (or subfamily)	C1A	papain (<i>Carica papaya</i>)
	C1B	bleomycin hydrolase (<i>Saccharomyces cerevisiae</i>)
	C2A	calpain 2 (<i>Homo sapiens</i>)
	C2B	Tpr peptidase (<i>Porphyromonas gingivalis</i>)
	C10	streptopain (<i>Streptococcus pyogenes</i>)
	C12	ubiquitin C-terminal hydrolase UCH-L1 (<i>Homo sapiens</i>)
	C16A	murine hepatitis coronavirus papain-like endopeptidase 1
	C16B	murine hepatitis coronavirus papain-like endopeptidase 2
	C19	ubiquitin-specific protease 14 (<i>Homo sapiens</i>)
	C28	foot-and-mouth disease virus L-peptidase
	C31	porcine respiratory and reproductive syndrome arterivirus-type cysteine peptidase alpha (lactate-dehydrogenase-elevating virus)
	C32	equine arteritis virus-type cysteine peptidase (porcine reproductive and respiratory syndrome virus)
	C33	equine arterivirus Nsp2 cysteine peptidase
	C39	bacteriocin-processing peptidase (<i>Pediococcus acidilactici</i>)
	C47	staphopain (<i>Staphylococcus aureus</i>)
	C51	D-alanyl-glycyl endopeptidase (<i>Staphylococcus aureus</i>)
	C54	Aut2 peptidase (<i>Saccharomyces cerevisiae</i>)
	C58A	YopT peptidase (<i>Yersinia pestis</i>)
	C58B	HopN1 peptidase (<i>Pseudomonas syringae</i>)
	C64	Cezanne deubiquitylating peptidase (<i>Homo sapiens</i>)
	C65	otubain-1 (<i>Homo sapiens</i>)
	C66	IdeS peptidase (<i>Streptococcus pyogenes</i>)
	C67	CylD protein (<i>Homo sapiens</i>)
	C70	AvrRpt2 peptidase (<i>Pseudomonas syringae</i>)
	C71	pseudomurein endoisopeptidase Pei (Methanobacterium phage psiM2)
	C76	UL36 deubiquitylating peptidase (human herpesvirus 1)
	C78	UfSP1 peptidase (<i>Mus musculus</i>)
	C83	gamma-glutamylcysteine dipeptidyltranspeptidase (Nostoc sp. PCC 7120)
	C85	DUBA deubiquitylating enzyme (<i>Homo sapiens</i>)
	C86	ataxin-3 (<i>Homo sapiens</i>)
	C87	nairovirus deubiquitylating peptidase (Crimean-Congo hemorrhagic fever virus)
	C88	OTU1 peptidase (<i>Saccharomyces cerevisiae</i>)

(Continued)

TABLE 404.1 (Continued)

Clan CD	Catalytic residues: <u>His</u> , <u>Cys</u>	Fold: β , α , β -sandwich
Family	C11	clostripain (<i>Clostridium histolyticum</i>)
	C13	legumain (<i>Canavalia ensiformis</i>)
	C14A	caspase-1 (<i>Rattus norvegicus</i>)
	C14B	metacaspase Yca1 (<i>Saccharomyces cerevisiae</i>)
	C25	gingipain R (<i>Porphyromonas gingivalis</i>)
	C50	separase (<i>Saccharomyces cerevisiae</i>)
	C80	RTX self-cleaving toxin (<i>Vibrio cholerae</i>)
	C84	prtH peptidase (<i>Bacteroides forsythus</i>)
Clan CE	Catalytic residues: <u>His</u> , <u>Asp</u> (or <u>Asn</u>), <u>Gln</u> , <u>Cys</u>	Fold: β -barrel + helices
Family	C5	adenain (human adenovirus type 2)
	C48	Ulp1 endopeptidase (<i>Saccharomyces cerevisiae</i>)
	C55	YopJ protease (<i>Yersinia pseudotuberculosis</i>)
	C57	vaccinia virus I7 processing peptidase
	C63	African swine fever virus processing peptidase (African swine fever virus)
	C79	ElaD peptidase (<i>Escherichia coli</i>)
Clan CF	Catalytic residues: <u>Glu</u> , <u>His</u> , <u>Cys</u>	Fold: α , β , α -sandwich
Family	C15	pyroglutamyl peptidase I (<i>Bacillus amyloliquefaciens</i>)
Clan CL		
Family	C60A	sortase A (<i>Staphylococcus aureus</i>)
	C60B	sortase B (<i>Staphylococcus aureus</i>)
	C82	L,D-transpeptidase (<i>Enterococcus faecium</i>)
Clan CM		
Family	C18	peptidase 2 (hepatitis C virus)
Clan CN		
Family	C9	nsP2 peptidase (Sindbis virus)
Clan CO		
Family	C40	dipeptidyl-peptidase VI (<i>Bacillus sphaericus</i>)

(Continued)

TABLE 404.1 (Continued)

Subclan PA(C)	Catalytic residues: <u>His, Cys</u>	Fold: two β -barrels
Family (or subfamily)	C3A	poliovirus type 1 picornain 3C
	C3B	poliovirus type 1 picornain 2A
	C3C	foot-and-mouth disease virus picornain 3C
	C3D	cowpea mosaic comovirus picornain 3C
	C3E	hepatitis A virus picornain 3C
	C3F	human parechovirus 1 picornain 3C
	C3G	rice tungro spherical virus endopeptidase
	C4	nuclear-inclusion-a endopeptidase (plum pox potyvirus)
	C24	rabbit hemorrhagic disease virus 3C endopeptidase
	C30	murine hepatitis coronavirus picornain 3C-like endopeptidase
	C37	Southampton virus processing peptidase
	C38	parsnip yellow fleck virus putative processing peptidase
	C74	pestivirus NS2 peptidase (bovine viral diarrhea virus 1)
Clan PD(C)	Catalytic residues: <u>Cys, Thr, His</u>	Fold: all β
Family	C46	hedgehog protein (<i>Drosophila melanogaster</i>)
Subclan PC(C)	Catalytic residues: <u>Cys, His</u>	Fold: α , β , α -sandwich
Family	C26	γ -glutamyl hydrolase (<i>Rattus norvegicus</i>)
Family	C56	Pfpl endopeptidase (<i>Pyrococcus furiosus</i>)
Not assigned to clan	Catalytic residues: not applicable	Fold: not applicable
Family	C6	potyvirus helper component endopeptidase (potato virus Y)
	C7	chestnut blight fungus virus p29 endopeptidase (<i>Cryphonectria hypovirus</i>)
	C8	chestnut blight fungus virus p48 endopeptidase (<i>Cryphonectria hypovirus</i>)
	C21	tymovirus endopeptidase (turnip yellow mosaic virus)
	C23	carlavirus endopeptidase (apple stem pitting virus)
	C27	rubella virus endopeptidase
	C36	beet necrotic yellow vein furovirus papain-like endopeptidase
	C42	beet yellows virus papain-like endopeptidase
	C53	pestivirus Npro endopeptidase (classical swine fever virus)
C75	AgrB peptidase (<i>Staphylococcus aureus</i>)	

For each family the identity of the MEROPS type example is given.

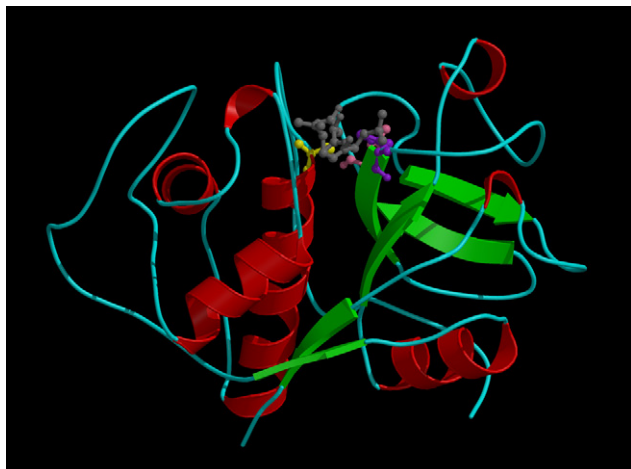


FIGURE 404.1 Richardson diagram of the papain/E-64 complex. The image was prepared from the Protein Data Bank entry (1PE6) as described in the Introduction (p. li). Catalytic residues are shown in ball-and-stick representation: Cys25, His159 and Asn175 (numbering as in PDB entry). E-64 is shown in gray in ball-and-stick representation.



FIGURE 404.3 Richardson diagram of streptopain. The image was prepared from the Protein Data Bank entry (1DKI) as described in the Introduction (p. li). Chain A of the tetramer is shown. The catalytic residues are shown in ball-and-stick representation: Cys192 (engineered to be Ser), His340 and Asn356 (numbering as in PDB entry).

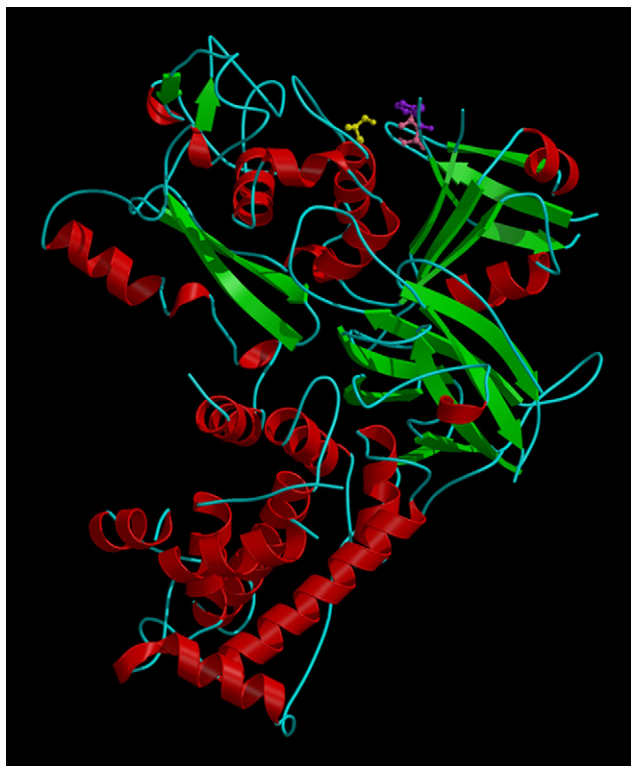


FIGURE 404.2 Richardson diagram of rat calpain 2. The image was prepared from the Protein Data Bank entry (1DF0) as described in the Introduction (p. li). Only the catalytic chain is shown. Catalytic residues are shown in ball-and-stick representation: Cys105 (engineered to be Ser), His262 and Asn286 (numbering as in PDB entry).



FIGURE 404.4 Richardson diagram of human ubiquitin C-terminal hydrolase UCH-L3. The image was prepared from the Protein Data Bank entry (1UCH) as described in the Introduction (p. li). Catalytic residues are shown in ball-and-stick representation: Cys95, His169 and Asp184 (numbering as in PDB entry).

of this subfamily are absent from some fungal genomes (for example, *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*) and the microsporidian *Encephalitozoon cuniculi*. Uncharacterized homologs are known from several archaean species, including *Archaeoglobus fulgidus* and *Methanosarcina acetivorans*,

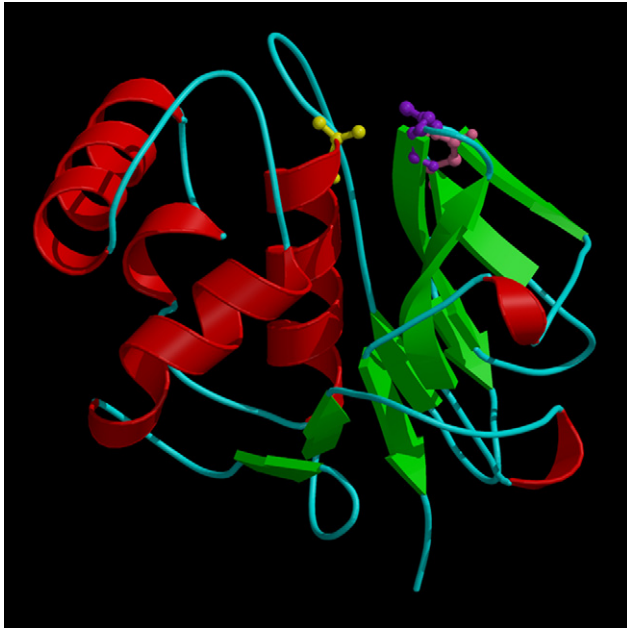


FIGURE 404.5 Richardson diagram of foot-and-mouth disease virus L-peptidase. The image was prepared from the Protein Data Bank entry (1QMY) as described in the Introduction (p. li). One molecule of the trimer is shown. Catalytic residues are shown in ball-and-stick representation: Cys51 (engineered to be Ala), His148 and Asp163 (numbering as in PDB entry).

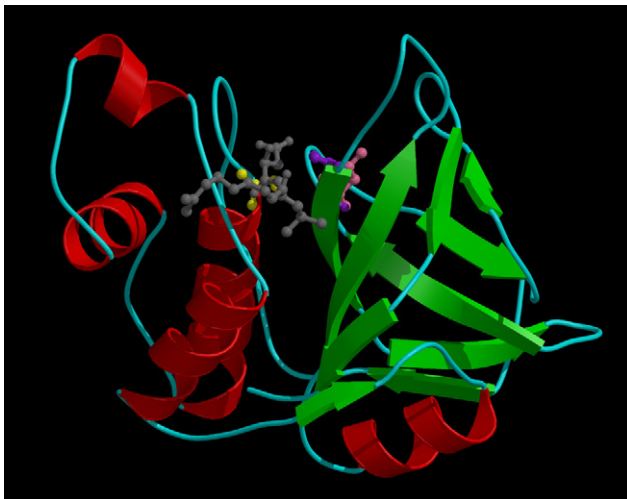


FIGURE 404.6 Richardson diagram of the staphopain/E-64 complex. The image was prepared from the Protein Data Bank entry (1CV8) as described in the Introduction (p. li). Catalytic residues are shown in ball-and-stick representation: Cys24, His120 and Asn141 (numbering as in PDB entry). E-64 is shown in gray in ball-and-stick representation.

but the majority of completed archaean genomes lack C1A homologs. Homologs are patchily distributed in bacteria, with most genomes lacking a homolog. Most bacterial homologs are from members of the phyla *Bacteroidetes* and *Chloroflexi*. Even within a genus, some

species will possess a homolog and some will not; for example, half of the *Mycoplasma* species with completed genomes have homologs. The subfamily includes the plant enzymes papain, chymopapain and actinidain, and the animal lysosomal enzymes cathepsins B, H and L. [Figure 404.1](#) shows the three-dimensional structure of papain.

There is a clear division within the subfamily between the cathepsin B-like enzymes and the papain-like enzymes. Among the cathepsin B-like enzymes are dipeptidyl-peptidase I, and the endopeptidases from *Giardia*. Included among the papain-like enzymes are cathepsins O, H, L, K and S. Cathepsin O is very divergent, whereas the others are more closely related. The close relationship between mammalian cathepsin H and plant aleurain is apparent.

The distinction between exopeptidases and endopeptidases is blurred for some members of subfamily C1A. Dipeptidyl-peptidase I (Chapter 447) acts principally as an exopeptidase, removing N-terminal dipeptides, but may have limited endopeptidase activity. Cathepsins B and H (Chapters 406 and 408) both possess endopeptidase activity, but arguably are more important for their exopeptidase activities. Cathepsin B acts as a peptidyl-dipeptidase, releasing C-terminal dipeptides, and this activity is attributable to the existence of an extended loop that forms a cap to the active-site cleft, and carries a pair of His residues that are thought to bind the C-terminal carboxylate of the substrate ([Figure 404.9](#)). Cathepsin X is a rather specific carboxypeptidase (Chapter 415).

The enzymes that enter the secretory pathway, destined either for secretion or for the lysosomal system, are synthesized as precursors, with N-terminal propeptides as well as the signal peptides. Most members of family C1 have propeptides homologous to that of papain, but the propeptide of cathepsin B is much shorter and very different in sequence. However, the propeptides of both cathepsin B and papain are thought to act in the same way in the proenzymes, blocking the active site by being bound in the reverse orientation to a substrate, as can be seen from the structure of procathepsin B ([Figure 404.10](#)). Papain-like propeptides may be identified from the presence of the ERFNIN motif, in which some of the following residues (numbered according to the papain propeptide) are conserved: Glu64, Arg68, Phe72, Asn75, Asn83, Phe96, Asp98 and Glu103. The papain propeptide is homologous to proteins CTLA-2 α and CTLA-2 β of activated T cells, which have been shown to be cysteine endopeptidase inhibitors [3]. It is notable that residue Pro2 is often conserved in the mature peptidases of family C1 (see the alignment in *MEROPS*), and this may prevent attack by aminopeptidases, since the Xaa-Pro bond is resistant to many such enzymes.

Most members of subfamily C1A are monomeric, but dipeptidyl-peptidase I is a homotetramer in which each

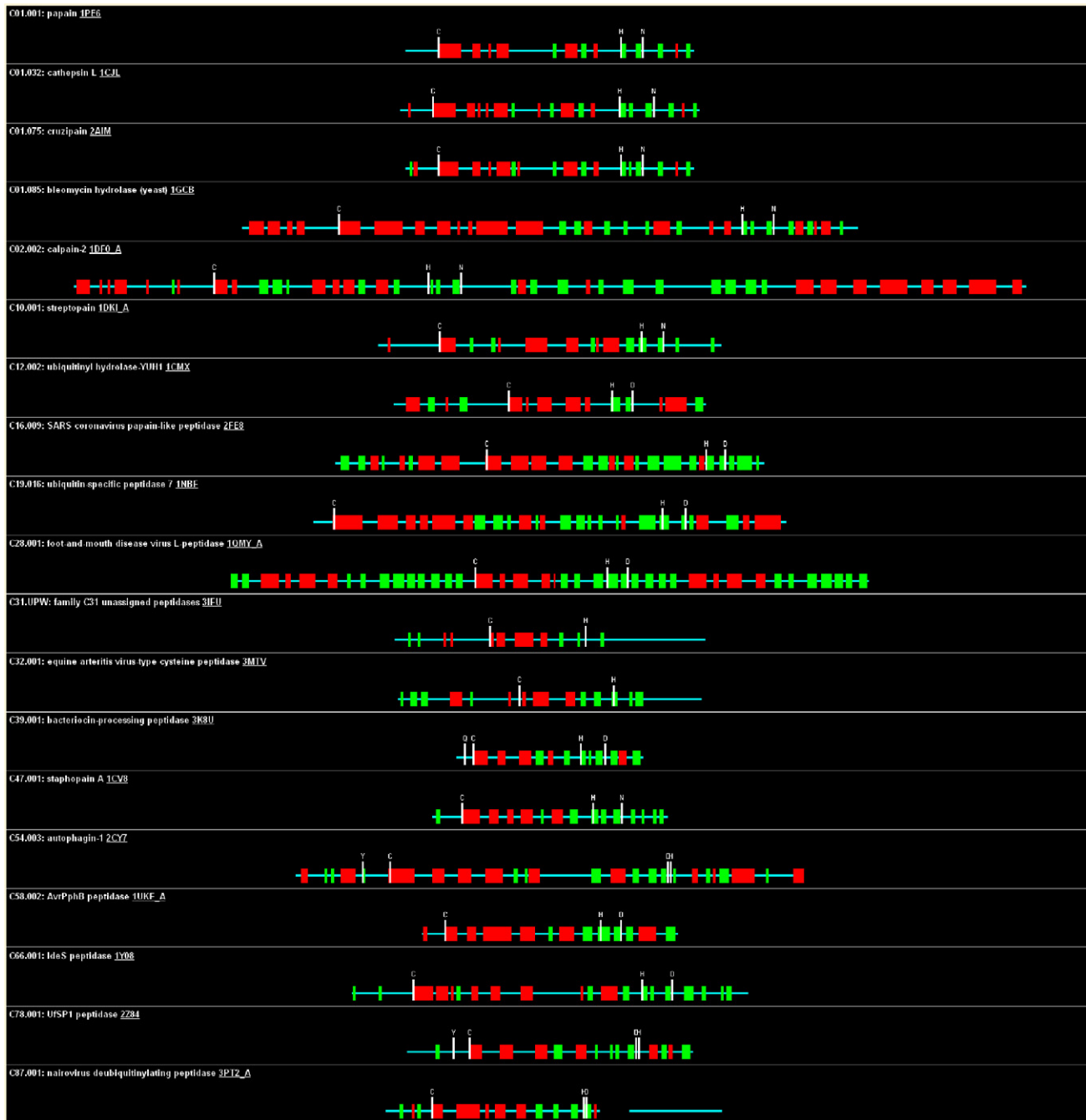


FIGURE 404.8 Comparison of secondary structure elements for selected members of clan CA.

Brassica (UniProt: [P25251](#)), tomato (UniProt: [P20721](#)), barley (UniProt: [P25249](#)), *Plasmodium* (UniProt: [P25805](#)), the baculovirus of *Autographa* (UniProt: [P25783](#)), lobster (UniProt: [P25784](#)) and oryzain α (UniProt: [P25776](#)). Ser338 is replaced by Asp in endopeptidases from *Entamoeba* that show a preference for Arg in S2, and this replacement also occurs in stem bromelain and a lobster digestive endopeptidase (UniProt: [P13277](#)). Glycyl endopeptidase (Chapter 422) from

papaya has a very restricted substrate binding site that can only accept glycine in P1.

The smaller *subfamily CIB* (numbered 1–5 in [Figure 404.4](#)) contains several soluble, intracellular peptidases. These are aminopeptidases that commonly show selectivity for release of N-terminal Arg residues, and include aminopeptidase C (PepC) from bacteria (Chapter 451), and bleomycin hydrolase (Chapter 449) from eukaryotes. Note that no bleomycin hydrolase homolog is

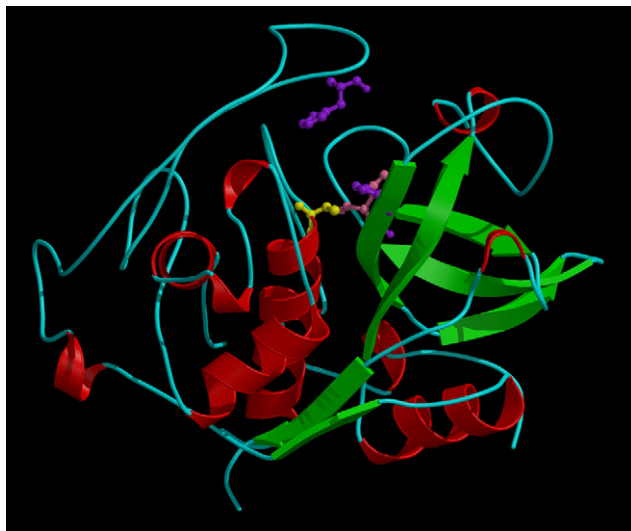


FIGURE 404.9 Richardson diagram of the human cathepsin B/CA030 complex. The image was prepared from the Protein Data Bank entry (1CSB) as described in the Introduction (p. li). Catalytic residues are shown in ball-and-stick representation: Cys29, His199 and Asn219 (numbering as in PDB entry). One of the pair of histidines (His110) on the occluding loop is also shown in ball-and-stick representation.

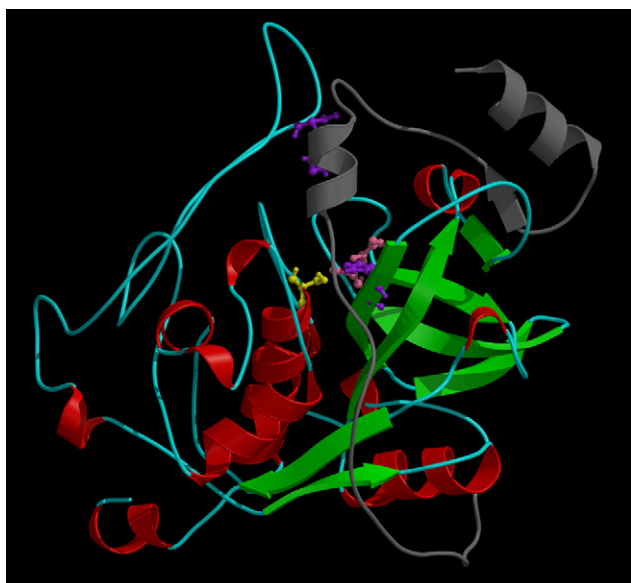


FIGURE 404.10 Richardson diagram of rat procathepsin B. The image was prepared from the Protein Data Bank entry (1MIR) as described in the Introduction (p. li). Catalytic residues are shown in ball-and-stick representation: Cys29 (engineered to be serine), His199 and Asn219 (numbering as in PDB entry). One of the pair of histidines (His110) on the occluding loop is also shown in ball-and-stick representation. The propeptide is shown in gray.

present in the genomes of *Caenorhabditis elegans*, *Arabidopsis thaliana*, *Encephalitozoon cuniculi* or *Schizosaccharomyces pombe*.

The aminopeptidase C-like enzymes are oligomeric. The yeast bleomycin hydrolase is probably representative,

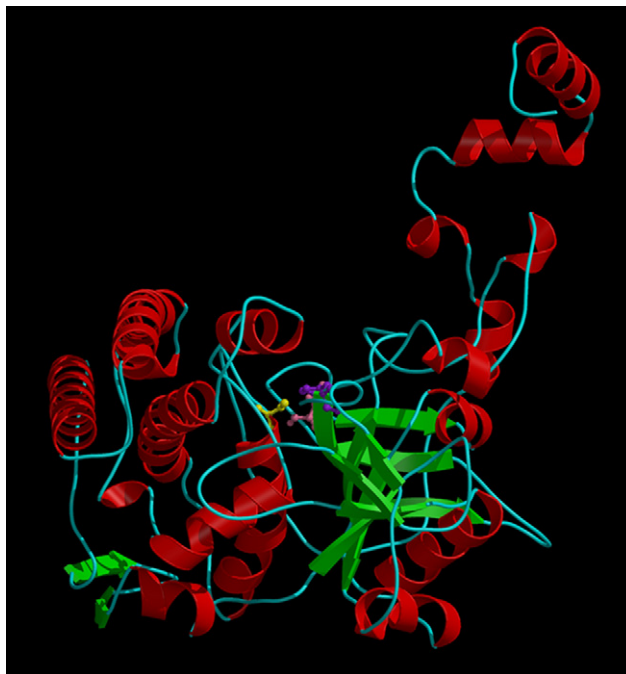


FIGURE 404.11 Richardson diagram of a subunit of yeast bleomycin hydrolase. The image was prepared from the Protein Data Bank entry (1GCB) as described in the Introduction (p. li). Catalytic residues are shown in ball-and-stick representation: Cys73 (engineered to be serine), His369 and Asn392 (numbering as in PDB entry).

and is a homohexamer, with the active sites arranged on the inner face of the central channel, in an arrangement reminiscent of that in the proteasome. This arrangement evidently allows only small molecules to interact with the catalytic site.

Unlike papain and cathepsin B, the aminopeptidases of family C1B do not contain disulfide bonds and are synthesized without propeptides. As is discussed in Chapter 449, the mature bleomycin hydrolase subunit consists of three domains, the peptidase domain, an oligomerization (or 'hook') domain, and a helical domain. Half of the hook domain is N-terminal to the catalytic domain, but the other half is an insert (relative to papain) preceding the catalytic His. The helical domain corresponds to two inserts in the catalytic domain with respect to papain. Figure 404.11 shows the tertiary structure of the yeast bleomycin hydrolase subunit.

There are a number of proteins homologous to papain that lack peptidase activity because catalytic residues have been replaced. These include an oil-body-associated protein from soybean (UniProt: P22895) in which the catalytic Cys is replaced by Gly, a surface protective protein from *Plasmodium* (UniProt: P08676) and a protein from *Schistosoma japonicum* (GenBank: X70969) in which the catalytic Cys has been replaced by Ser. In mammals, non-peptidase homologs include

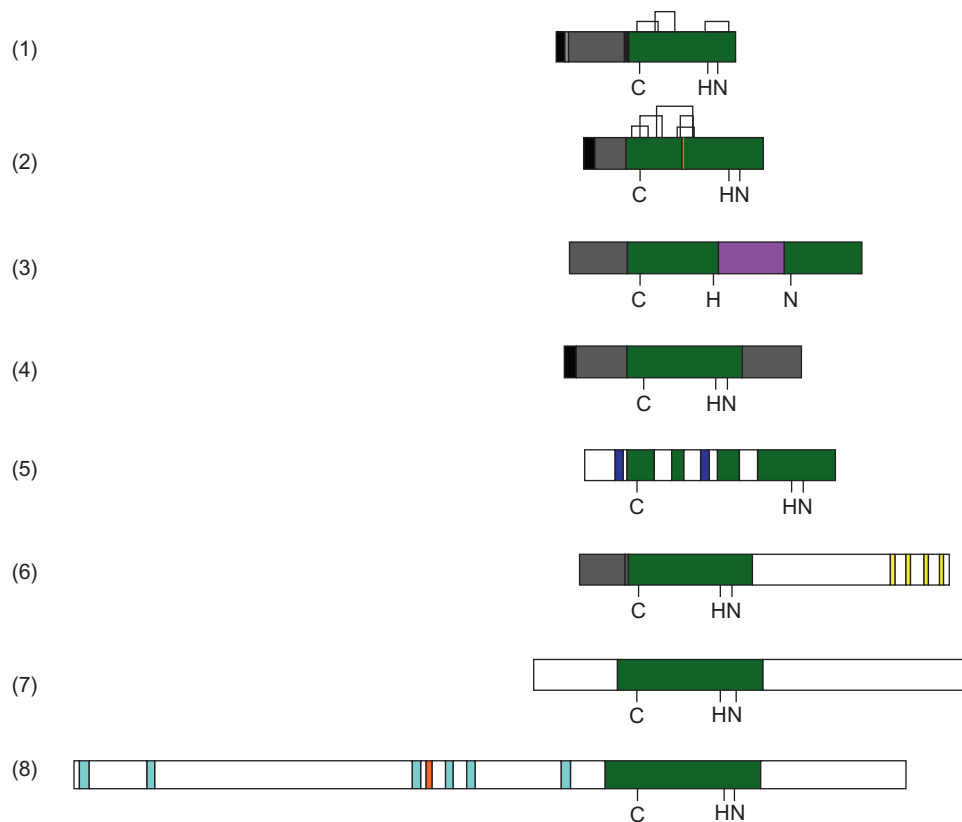


FIGURE 404.12 Domain structures for selected endopeptidases from clan CA. Endopeptidase sequences are shown as rectangles. The length of the rectangle is a representation of the sequence length. Disulfide bridges are shown as open boxes above the sequence rectangle. The positions of the catalytic triad residues are shown below the sequence rectangles, and all structures are aligned at the catalytic cysteine. Structures are arranged so that the simplest is at the top, and the most complex at the bottom. A dashed vertical line marks the position of His111 on the occluded loop of cathepsin B (2). Shaded boxes within the sequence rectangle represent different structural domains. Key to domains: black, signal peptide; gray, N- or C-terminal propeptide; green, catalytic domain; purple, serine-rich domain; dark blue, internal repeats; yellow, calcium-binding EF-hand structure domain; light blue, zinc finger-like domains; red stripes, opa repeat. Key to sequences: (1) papain, (2) *Homo sapiens* procathepsin B, (3) *Dictyostelium discoideum* cysteine proendopeptidase 7, (4) *Trypanosoma brucei* cruzipain, (5) *Saccharomyces cerevisiae* bleomycin hydrolase, (6) *H. sapiens* calpain-1, (7) *Aspergillus nidulans* palB protein, (8) *Drosophila melanogaster* sol protein.

testin, which is secreted by Sertoli cells of the testis but has no known function, and tubulointerstitial nephritis antigen, an extracellular matrix basement protein that is a target for antibody-mediated interstitial nephritis [5]. Both mammalian proteins have the catalytic Cys replaced by Ser. Several pseudogenes are known in mammals, including a human cathepsin L-like pseudogene (GenBank: L07772).

Family C2 contains the calpains. The molecule of a 'typical' calpain is a heterodimer of a heavy chain and a light chain (Chapters 454, 456). The heavy chain is a mosaic protein containing the peptidase domain (domain II) (Figure 404.12) and also a C-terminal domain (domain IV) with four calcium-binding EF-hand structures. The calcium-binding domain is homologous to calcium-binding domains in other proteins. The functions of heavy chain domains I and III are unknown. Crystal structures of calcium-free human calpain have been described [6,7],

and they show that in the absence of calcium the peptidase unit adopts a conformation that disrupts the catalytic site. In most mammalian cells there are two calpains with different calcium requirements for activity: calpain-1 (Chapter 454) and calpain-2 (Chapter 455). Calpain 3 (Chapter 456) has a 48 residue insert in the catalytic domain between the catalytic Cys and His. A number of atypical calpains are known, in which the domains N-terminal and C-terminal to the protease domain are different from those in the typical calpains, and a small subunit may not be required for activity. A *Drosophila* homolog of calpain possesses an insert in the C-terminal domain, and there is evidence that differential splicing can give rise to a form lacking the calcium-binding domain [8]. Another calpain homolog exists in *Drosophila* and is involved in the development of the small optic lobe. This putative endopeptidase is the product of the *sol* gene, and the protein is predicted to be multidomain, including six

zinc fingers as well as a peptidase domain, but no calcium-binding domain [9] (Figure 404.12). Again, an alternatively spliced form exists, this time lacking the peptidase domain.

The calpain family is not restricted to animals: there are sequences from fungi, including *Aspergillus* and *Saccharomyces*, the protozoa *Trypanosoma* and *Leishmania*, the plant *Arabidopsis*, and the cyanobacterium *Nostoc* (but not *Synechocystis*). Homologs are also known from members of a wide variety of bacterial phyla, but homologs are absent from archaea. The only bacterial homolog that has been characterized is the Trp peptidase from *Porphyromonas gingivalis*, which is so dissimilar in sequence to other members of the family that it had been considered a member of a separate family, and is now the only representative in subfamily C2B. The absence of homologs from archaea and many bacteria with completely sequenced genomes, implies that horizontal gene transfer has occurred, probably from a protist or fungus to a bacterium (see the [evolutionary tree](#) in the *MEROPS* database). *Aspergillus* palB protein and a hypothetical protein from *Caenorhabditis* are unusual among peptidases in clan CA in that the catalytic Cys is followed by Ser rather than a hydrophobic residue (see the alignment in *MEROPS*). Both the *Aspergillus* and *Caenorhabditis* proteins are multidomain, but show only limited sequence similarity to domains I and III of human calpain. Neither protein possesses the calcium-binding domain IV. A non-peptidase homolog is known from *Trypanosoma brucei* in which the catalytic Cys has been replaced by Ser.

Family C10 contains streptopain (Chapter 483) and a few similar enzymes, all of which are from gram-positive bacteria. Streptopain is inactivated by E-64 much more slowly than papain, but has a similar specificity, with a preference for a hydrophobic residue at P2 [10,11]. In some members of the family, Asn308 (numbering as in Figure 404.7) is replaced by Asp. The prtT peptidase from *Porphyromonas* is a mosaic protein with a C-terminal hemagglutinin domain unrelated to those in gingipains R and K in family C25 but 98% identical to the C-terminal domain of hemin-regulated protein from *Porphyromonas* (UniProt: P72198). Surprisingly, streptopain homologs are found in some *Bacteroides* species but not others.

There are several families in clan CA in which the peptidases release ubiquitin from conjugated proteins. Ubiquitin is a 76 amino acid protein that is attached to other proteins as a signal for intracellular translocation or degradation (usually mediated by the proteasome). Ubiquitin is attached through the carboxyl group of its C-terminal glycine residue either to the N-terminus of another polypeptide or to the α -amino group of a lysine residue, which forms an isopeptide bond (Chapters 460–462). Ubiquitin may also be attached to another

molecule of ubiquitin, and proteins targeted for degradation may be polyubiquitinated. In order for ubiquitin to be recycled there is a need for peptidases to liberate the free ubiquitin by hydrolysis of the C-terminal glycylyl bond. Peptidases that release ubiquitin from conjugates are known as ‘ubiquitin C-terminal hydrolases’ or ‘deubiquitinating enzymes’. Because release of ubiquitin may involve cleavage of an isopeptide bond, ubiquitin C-terminal hydrolases are not classified as peptidases in *Enzyme Nomenclature* but as thiolester hydrolases. There are at least eight families of deubiquitinating enzymes in clan CA, and there are enzymes in other, unrelated, families, including the metallopeptidase AMSH (family M67) which is a component of the 26S proteasome (Chapter 817). There are over 100 different deubiquitinating enzymes in humans, which implies a sophisticated network controlling translocation and degradation of proteins which is only just being understood.

Family C12 includes peptidases that can hydrolyze the ubiquitin-conjugate glycylyl bond whether it is an α -peptide bond or an isopeptide bond, with various specificities. The tertiary structure of the human ubiquitin C-terminal hydrolase UCH-L3 has been solved [12], and is shown in Figure 404.4. There are many structural similarities to papain. The molecule is bilobed, one lobe consisting mainly of helices, the other containing a β -barrel surrounded by helices. The catalytic Cys is at the start of one of the helices, and the catalytic His is the start of a β strand. A difference from papain is that the first strand of the β -barrel precedes the helix bearing the catalytic cysteine in the sequence, whereas in papain the helix precedes all the strands of the β -barrel. The spacing between Gln152 and Cys158 (numbering as in Figure 404.7) is identical to that in families C1, C2 and C10. Peptidases in family C12 are synthesized without propeptides, and are intracellular; they are known only from animals, plants and fungi.

Family C19 contains the second group of ubiquitin C-terminal hydrolases. These are also intracellular peptidases, but are able to release ubiquitin from much larger peptides that have been polyubiquitinated (Chapters 463–473). Because ubiquitination is also essential for the assembly of multimeric proteins, there is a need to remove the ubiquitin molecules from polyubiquitinated subunits, which is presumably one of the functions of peptidases of family C19. Peptidases in family C19 have more complicated structures than those in C12, and many are multidomain proteins. There is also a much greater variety, with *Saccharomyces* possessing at least 19 homologs, and even greater numbers in mammals. The tertiary structure of ubiquitin-specific peptidase 7 has been solved and a catalytic triad consisting of Cys, His and Asp was identified (Figure 404.13; also see later Figure 404.35). The Asp is equivalent to the Asn152 of papain. The side

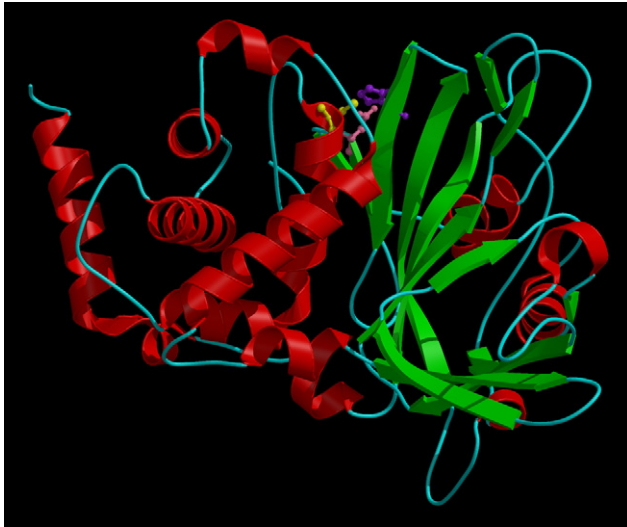


FIGURE 404.13 Richardson diagram of human ubiquitin-specific peptidase 7. The image was prepared from the Protein Data Bank entry (INBF) as described in the Introduction (p. li). One molecule of the homodimer is shown. Catalytic residues are shown in ball-and-stick representation: Cys223 in yellow, His464 in purple and Asp481 in pink (numbering as in PDB entry).

chain of an asparagine stabilizes the oxyanion hole and is equivalent of Gln19 in papain. The fold has been likened to that of an open hand, consisting of three subdomains, with the active site Cys on the ‘thumb’ subdomain, the active site His on the ‘palm’ and the active site cleft between the ‘thumb’ and the ‘palm.’ The ‘fingers’ interact with ubiquitin [13].

Members of family C19 are present in all the eukaryotic genomes so far completely sequenced, including *Encephalitozoon cuniculi*. Family C19 sequences are known from several protozoan species that do not contain family C12. C19 homologs are known from bacteria such as the pathogen *Burkholderia pseudomallei*, but the sequences are shorter (c. 280 residues) and lack inserts between the catalytic Cys and His. The bacterial homologs are derived from the most divergent branch in the evolutionary tree. The TssM protein from *Burkholderia mallei* has been shown to be a deubiquitinating enzyme [14].

Other families of deubiquitinating enzymes were initially postulated by Makarova *et al* [15], which have now been experimentally verified. The human Cezanne protein, from *family C64*, was shown to release ubiquitin from linear or branched synthetic ubiquitin chains and from ubiquitinated proteins [16], and from the solved structure of tumor necrosis factor α -induced protein 3, the fold was shown to be similar to that of papain [17]. *Family C65* includes the otubains (see Chapter 477). The tertiary structure of otubain-2 has been determined and shows that the fold is similar to that of papain [18]. Otubain-1 has been shown to cleave Lys48-linked

polyubiquitin [19]. The CylD protein from *family C67* (see Chapter 476) has been shown to have de-ubiquitinating activity that is directed towards non-Lys48-linked polyubiquitin chains [20]. The tertiary structure has been solved [21]. *Family C85* includes the DUBA deubiquitinating peptidase, which has been shown to cleave the lysine-63-linked polyubiquitin chains on tumor necrosis factor receptor-associated factor 3 [22]. However, no tertiary structure has been solved for any member of the family, and the active site residues are predictions only [15]. Ataxin-3 (Chapter 479) is yet another deubiquitinating enzyme and a member of *family C86*. The nuclear magnetic resonance solution structure of the domain containing the active site of ataxin-3, known as the Josephin domain, has been determined [23], and shows the characteristics of a cysteine peptidase in clan CA. *Family C88* includes OTU1 peptidase from *Saccharomyces cerevisiae* (Chapter 478) which preferentially cleaves longer polyubiquitin chains with Lys48 linkages. The tertiary structure has been solved for the catalytic domain, and the structure is similar to that of other clan CA members [24].

Surprisingly, enzymes that release ubiquitin from proteins have been discovered in viruses. The UL36 protein from herpes simplex virus releases ubiquitin from proteins ubiquitinated via Lys48 and is the founder of *family C76* (Chapter 480). The tertiary structure has been solved, and shows a papain-like fold [25]. The Crimean-Congo hemorrhagic fever virus (a *Nairovirus*) also possesses a deubiquitinating enzyme which can downregulate antiviral responses of the host by release of ubiquitin and interferon-stimulated gene product 15 from proteins of the antiviral and inflammatory pathways [26]. The *Nairovirus* enzyme is a member of *family C87*, and the order of active site residues, identified by mutagenesis [26], led us to include the family in clan CA.

Other proteins act like ubiquitin and can be tagged to proteins, these include SUMO and NEDD-8. Ubiquitin-fold modifier-1 (Ufm1) is another ubiquitin-like protein, and like ubiquitin attaches to proteins via the C-terminal glycine. The physiological relevance for tagging proteins with Ufm1 is unknown, but tagged proteins are not degraded. Peptidases in *family C78*, which includes UfSP1 peptidase (Chapter 481), release Ufm1 from conjugates and process the Ufm1 precursor by removing a C-terminal dipeptide [27]. Once again, resolution of the tertiary structure has confirmed the relationship to papain [28] and the family C78 is included in clan CA. The aspartic acid that is the third member of the catalytic triad unusually is N-terminal of the histidine, whereas in most papain-like proteins follows the histidine. The residue stabilizing the oxyanion hole is a tyrosine (whereas it is glutamine in papain). In both these respects, members of family C78 resemble members of the family C54.



FIGURE 404.14 Richardson diagram of human autophagin-2. The image was prepared from the Protein Data Bank entry (2P82) as described in the Introduction (p. li). Catalytic residues are shown in ball-and-stick representation: Tyr57 in green, Cys77 in yellow, Asp279 in pink and His281 in purple (numbering as in PDB entry).

Family C54 includes the yeast peptidase Aut2 that is responsible for the removal of a C-terminal arginine residue from the Apg8 (or Aut7) protein. The Apg8FG protein, as the processed substrate is known, is then able to form a conjugate with an unidentified protein via the newly exposed C-terminal Gly and the conjugate binds tightly to membranes permitting the formation of the double-membrane vesicles required for the transport of material from the cytoplasm to an autophagic vesicle. Homologs are known from human, mouse, *Drosophila melanogaster*, *Caenorhabditis elegans* and *Arabidopsis thaliana*. Inhibition by *N*-ethylmaleimide and site-directed mutagenesis of Cys159 (see [Alignment C54](#) in *MEROPS*) has shown that yeast peptidase Aut2 is a cysteine peptidase (Chapter 482), but other catalytic residues have not been determined. The tertiary structure of human autophagin-1 has been solved [29,30], and shows a catalytic triad consisting of Cys74, Asp278 and His280 that is unusual amongst papain-like proteins, in which the third member of the catalytic triad follows the histidine in the sequence (see [Figure 404.14](#)). Families C54 and C78 are unique in having the third member (in both families an aspartic acid) preceding the histidine. Members of both families are also unique in having a tyrosine helping to stabilize the oxyanion hole.

Family C47 includes staphopain (see Chapters 484 and 485), a cysteine peptidase from *Staphylococcus epidermidis*. The tertiary structure has been determined and shows a papain-like fold [31] (see [Figure 404.6](#)). Staphopain is inhibited by E-64 and cystatin A.

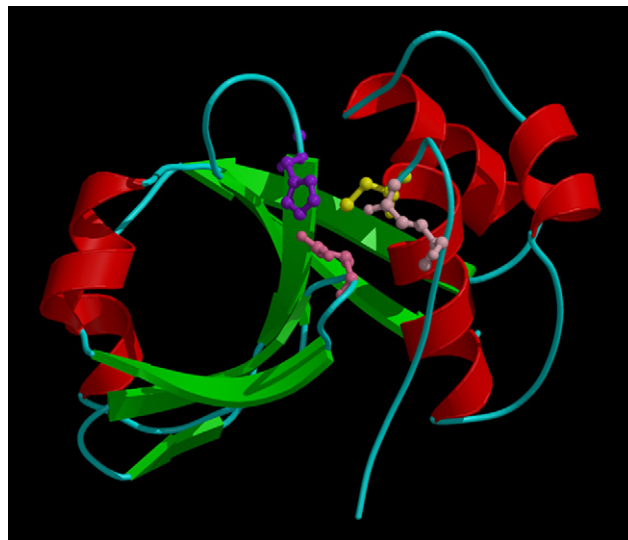


FIGURE 404.15 Richardson diagram of the bacteriocin-processing peptidase from *Streptococcus mutans*. The image was prepared from the Protein Data Bank entry (3K8U) as described in the Introduction (p. li). Catalytic residues are shown in ball-and-stick representation: Gln11 and Asp114 in pink, Cys17 in yellow and His96 in purple (numbering as in PDB entry).

Family C39 includes endopeptidases that are involved in the processing and export of bacteriocins. Bacteriocins are antibiotic proteins secreted by some species of bacteria that have the effect of inhibiting the growth of other bacterial species. The bacteriocin is synthesized as a precursor with an N-terminal leader peptide, and processing involves removal of the leader peptide at a Gly-Gly↓ bond, followed by translocation of the mature bacteriocin across the cytoplasmic membrane. The endopeptidase serves both functions, and is also known as an ATP-binding cassette transporter or ABC-transporter. The endopeptidases are integral membrane proteins containing an N-terminal peptidase domain followed by six or more transmembrane domains and a C-terminal ATP-binding domain. The structure of the peptidase domain has been solved, and shows a papain-like fold (see [Figure 404.15](#)) [32]. The non-peptidase-domains are homologous to a variety of other ABC-transport proteins. Peptidase specificity is not unlike that of the ubiquitin peptidases.

The complex structure of bacterial cell walls means that there are a large number of peptidases involved in cell wall processing and lysis. Newly synthesized bacteriophage virions escape from the host cell by lysing it, and the bacteriophage genomes encode a cell wall lytic enzyme that acts either as a peptidase to degrade the cell wall crosslinks, or as an amidase to disrupt the acetylmuramyl bonds. *Staphylococcus aureus* has a multi-domain enzyme known as autolysin the C-terminal domain of which is related to lysostaphin (peptidase family M37). The

autolysin gene has been acquired by a number of staphylococcal phages, and in bacteriophage φ 11, autolysin has been shown to cleave the cell wall cross-linking peptide at D-Ala↓Gly. The peptidase activity is expressed by the N-terminal domain [33]. An alignment shows that residues Gln31, Cys32, His95 and Asn113 are conserved, suggesting that the N-terminal domain is a cysteine peptidase with a papain-like fold, which we include in *family C51*.

Pseudomonas syringae is a plant pathogen which secretes disease-causing factors into the host cell via its type III secretion pathway [34]. The AvrRpt2 protein (Chapter 552) is one of these virulence factors which activates itself and probably cleaves the RIN4 protein of the host plant cell. By comparison with staphopain, AvrRpt2 protein has been predicted to be a cysteine peptidase with a Cys, His and Asp, papain-like, catalytic triad, and mutation of any of these residues prevented autoactivation or RIN4 digestion [35]. The AvrRpt2 protein is included in *family C70*.

Family C58 includes the YopT effector protein from the plague organism *Yersinia pestis*. The Yop effectors are proteins secreted into host cells that disrupt the immune response; two of these effectors are now known to be peptidases, YopT and YopJ (see Chapter 488). The YopT effector has been shown to be a cysteine peptidase, and is known to remove a C-terminal prenylated cysteine (or a short peptide containing the C-terminal cysteine) from GTPases. The effect is to release the GTPase from the membrane. The tertiary structure has been solved for a homolog from the plant pathogen *Pseudomonas syringae*, which is the avirulence protein AvrPphB (see Figure 404.16). The only proteolytic activity that could be demonstrated was that for autocatalytic activation of the precursor protein. A papain-like catalytic triad (Cys/His/Asp) has been identified in both the YopT effector and the AvrPphB protein by site-directed mutagenesis. However, there does not appear to be a conserved residue in the family that would act like Gln19 in papain [36].

Another related peptidase from *Pseudomonas syringae*, HopN1 is a virulence factor and is injected into host plant cells by a type III secretion system where it suppresses cell death [37]. The HopN1 peptidase is only distantly related and consequently, family C58 is divided into two subfamilies with YopT and AvrPphB members of subfamily C58A and HopN1 in C58B.

The IdeS peptidase (Chapter 489) from *Streptococcus pyogenes* is included in *family C66*. This peptidase cleaves the γ -chains of human IgG. From the crystal structure [38], a papain-like fold is apparent and a catalytic triad consisting of Cys94, His262 and Asp284 (see Figure 404.17).

Family C83 includes phytochelatin synthases (Chapter 490), which are γ -glutamylcysteine dipeptidyltranspeptidases, from cyanobacteria and plants. These enzymes

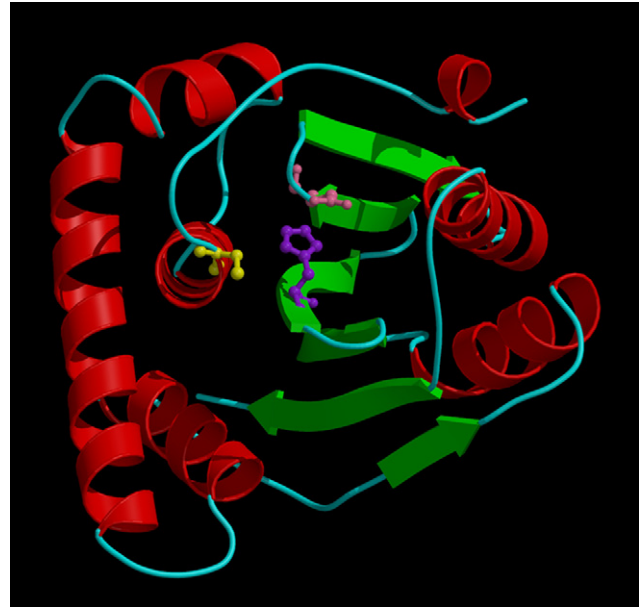


FIGURE 404.16 Richardson diagram of the AvrPphB peptidase from *Pseudomonas syringae*. The image was prepared from the Protein Data Bank entry (1UKF) as described in the Introduction (p. li). Catalytic residues are shown in ball-and-stick representation: Cys98 in yellow and His212 in purple (numbering as in PDB entry). The molecule is rotated 90° in the Y direction as compared with papain.



FIGURE 404.17 Richardson diagram of the IdeS peptidase from *Streptococcus pyogenes*. The image was prepared from the Protein Data Bank entry (1Y08) as described in the Introduction (p. li). Catalytic residues are shown in ball-and-stick representation: Cys94 in yellow, His262 in purple and Asp284 in pink (numbering as in PDB entry).

remove the C-terminal Gly from glutathione (γ -Glu-Cys-Gly). This is in contrast to the action of the γ -glutamyltransferases (Chapter 820) that remove the N-terminal residue. Removal of the C-terminal Gly is an important step towards formation of phytochelatin from linear

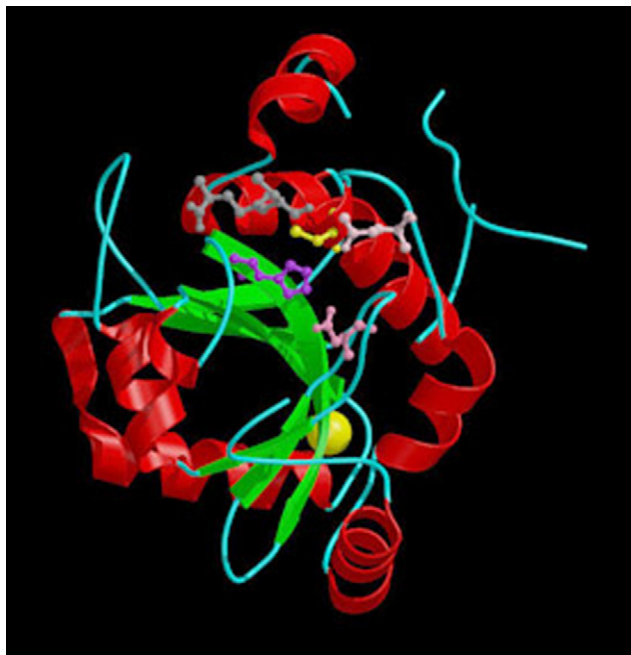


FIGURE 404.18 Richardson diagram of the gamma-glutamylcysteine dipeptidyltranspeptidase from *Nostoc* sp. PCC 7120. The image was prepared from the Protein Data Bank entry (2BU3) as described in the Introduction (p. li). One monomer of the dimer is shown. Catalytic residues are shown in ball-and-stick representation: Gln64 in dark pink, Cys70 in yellow, His183 in purple and Asp201 in light pink. Glutathione is shown in grey ball-and-stick representation (numbering as in PDB entry).

polymers of the γ -Glu-Cys, and phytochelatin is important because ions of heavy metal ions such as mercury and cadmium are rendered harmless by forming tight complexes with phytochelatin. The tertiary structure of γ -glutamylcysteine dipeptidyltranspeptidase from the cyanobacterium *Nostoc* sp. PCC 7120 has been solved, and shows a papain-like fold and an active site also similar to that of papain, containing Gln, Cys, His and Asp [39] (see Figure 404.18).

The ‘papain-like’ endopeptidases of RNA viruses form a large group of cysteine peptidases that contain the catalytic dyad residues in the order Cys, His. All of the families had been included in clan CA, even though for many no tertiary structure for any member had been solved. When the crystal structure of the nsP2 (non-structural protein 2) from Sindbis virus (a member of family C9) was determined, the fold was quite unlike that of papain, and the family was assigned to its own clan (CN, see below). Consequently, many families of peptidases from RNA viruses that had been included in clan CA have been removed, because until a tertiary structure has been solved it will not be possible from sequence alone to determine if any of these families belongs to clan CA or clan CN. The families of RNA viruses retained in

clan CA, and for which tertiary structures have been solved, are C28, C31, C32 and C87. Each family contains peptidases from only one family of viruses. Thus, endopeptidases from aphthoviruses are in family C28; arteriviruses in C31 and C32, and nairoviruses in C87. The cleavages mediated by the clan CA viral endopeptidases usually show Gly in P1 (contrasting with Gln in P1 for many cleavages by clan PA(C) endopeptidases).

In most of these viruses there is a single polyprotein containing a single peptidase, but there may be up to three peptidases in a single polyprotein. In the foot-and-mouth disease virus, potyviruses and togaviruses, the other cysteine peptidases are not related to papain (having the catalytic dyad in the order His, Cys). In the coronaviruses and arteriviruses, all the endopeptidases have a Cys, His catalytic dyad, suggesting papain-like tertiary folds, and are probably derived from one or two sequence duplications. Such duplications appear to have been as ancient as the speciation events that gave rise to the different viruses, because there is no significant sequence similarity, and hence the peptidases are in different families.

The crystal structure of the L-peptidase from the foot-and-mouth disease virus (in **family C28**) shows a clear relationship to papain. Although aphthoviruses such as the foot-and-mouth disease virus possess picornain 3C (Chapter 537) as a general polyprotein-processing enzyme, they lack a picornain 2A, and the functions of that endopeptidase are performed by the L-peptidase or an autolytic activity (see Chapter 491). The L-peptidase is sited at the N-terminus of the polyprotein, from which it releases itself. The L-peptidase also cleaves the eukaryotic initiation factor eIF-4G (formerly known as p220 or eIF-4 γ), thus preventing cap-dependent synthesis of host-cell proteins.

Family C16 contains the mouse hepatitis virus p28 endopeptidase, which releases itself from the N-terminus of the polyprotein encoded by gene A. The catalytic dyad has been identified by mutagenesis [40]. The polyprotein encoded by gene A includes a picornain-like endopeptidase and a putative second papain-like endopeptidase, formerly assigned to a separate family, C29, but now considered to be a member of family C16. The compound E-64c inhibits polyprotein processing in the mouse hepatitis virus [41], presumably by inhibiting the action of one or both of the two papain-like endopeptidases. The tertiary structure of the papain-like endopeptidase from SARS virus has been solved (see Figure 404.19).

Families C31, C32 and C33 include polyprotein-processing endopeptidases from arteriviruses. In the Lelystad (porcine reproductive and respiratory syndrome) virus, the three papain-like endopeptidases are consecutive in the pol polyprotein, with PCP α at the N-terminus, followed by PCP β and then Nsp2, which are so different in

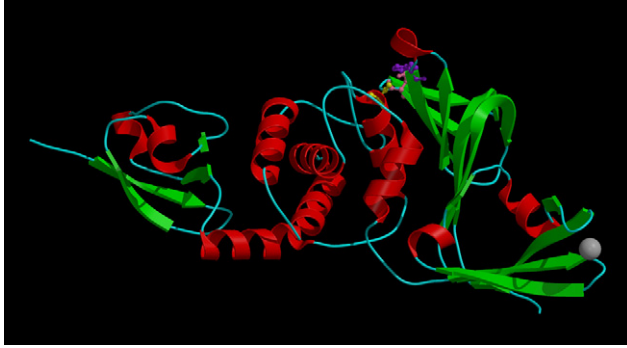


FIGURE 404.19 Richardson diagram of the SARS coronavirus papain-like peptidase. The image was prepared from the Protein Data Bank entry (2FE8) as described in the Introduction (p. li). One molecule of the homotrimer is shown. Catalytic residues are shown in ball-and-stick representation: Cys1651 in yellow, His1812 in purple and Asp1826 in pink (numbering as in PDB entry). The structural zinc ion is shown as a light gray sphere in CPK representation.



FIGURE 404.20 Richardson diagram of hepatitis A virus picornain 3C. The image was prepared from the Protein Data Bank entry (1HAV) as described in the Introduction (p. li). Catalytic residues are shown in ball-and-stick representation: His44 and Cys172 (engineered to be serine) (numbering as in PDB entry).

sequence that they are assigned to the separate families C31, C32 and C33, respectively. PCP α and PCP β together constitute the Nsp1 protein. The catalytic dyads have been determined by mutagenesis for all three endopeptidases [42,43]. Besides the three papain-like cysteine endopeptidases, the arteriviruses possess a fourth, serine-type, endopeptidase (see Chapter 691) in the pol polyprotein. Not all arteriviruses require all four endopeptidases, because in the equine arteritis virus, PCP α is inactive, the catalytic Cys being replaced by Lys. The PCP α endopeptidase releases itself from the N-terminus of the Nsp1 protein [42]. The PCP α endopeptidase cleaves between the Nsp1 and Nsp2 proteins in the pol polyprotein [42]. The Nsp2 endopeptidase cleaves between the Nsp2 and Nsp3 proteins, and, unusually, the catalytic Cys is followed by Gly. Cys-Gly is generally regarded as diagnostic for the subclass PA(C) viral endopeptidases, but it has also been seen in some calpain homologs [43] and in family C12.

In the archaean *Methanothermobacter*, the cell wall is analogous to that of bacteria, but the crosslinking peptide contains amino acids only in the normal, *laevo*, orientation. The crosslink is made between the ϵ -amino group of L-Lys and the α -carboxyl group of L-Ala. It is this bond that is broken by the pseudomurein endoisopeptidase (*family C71*) from *Methanothermobacter* phage ψ IM2. The enzyme can be inhibited by EDTA, suggesting that metal ions are important for activity. However, a cysteine-type catalytic triad (Cys, His, Asp), similar to that of papain, has also been proposed [44], and family C71 is included in clan CA.

Clan PA(C)

A second clan of cysteine peptidases contains processing endopeptidases of RNA viruses. Formerly termed clan CB, this has the catalytic dyad in the order His, Cys in

the sequence. A comparison of sequence motifs led Bazan and Fletterick [45] to suggest that the polyprotein processing cysteine endopeptidases of several single, positive-stranded RNA viruses were structurally related to chymotrypsin, and this was confirmed when the crystal structures of picornains 3C from human hepatitis A virus (Chapter 541) (Figure 404.20) and human rhinovirus type 14 (Chapter 537) were found to share the two β -barrel fold of chymotrypsin. There now seems no doubt that these peptidases have evolved from a common ancestor with chymotrypsin. The catalytic His of picornain 3C (in family C3) is equivalent in position to His57 of chymotrypsin, and the Cys replaces Ser195 (see the representation of two-dimensional structures for clan PA, see Figure 559.2). In rhinovirus picornain 2A a His, Asp, Cys catalytic triad is seen [46], but the Asp is replaced by Glu in some other members of family C3. The viral endopeptidases and chymotrypsin clearly belong in the same clan, and because the clan contains both cysteine-type and serine-type peptidases it has been named *clan PA*. The clan is divided into subclass PA(S), for the serine peptidases, and *subclass PA(C)* for the cysteine peptidases. Families C4, C24, C30, C37 and C38 also contain endopeptidases from RNA viruses that are predicted to have the chymotrypsin fold, and are placed in subclass PA(C) together with family C3.

The viral peptidases in subclass PA(C) contrast in several respects with those in clan CA. In subclass PA(C) the active site Cys is followed by Gly (like the Ser in chymotrypsin) whereas in clan CA the catalytic Cys is generally followed by a large hydrophobic residue. As regards

function, the endopeptidases of subclan PA(C) tend to be general processing enzymes, cleaving several bonds in the polyprotein, whereas those in clan CA generally perform only a single cleavage. There is also a difference in P1 substrate specificity: in subclan PA(C) cleavage is commonly at a Gln↓Gly bond, whereas in clan CA cleavage usually occurs at a Gly↓Gly bond. The subclan PA(C) endopeptidase is usually positioned between the helicase and the RNA polymerase in the polyprotein, whereas the clan CA enzyme usually precedes both the helicase and the RNA polymerase. Taken together with other indications that the positive-stranded RNA viruses are evolutionarily related, the similar locations of the subclan PA(C) endopeptidases in the genomes is entirely consistent with a common evolutionary origin of all of the peptidases in the subclan. The high rate of mutation of RNA viruses, lacking proofreading in the replication of the genome, could rapidly have obscured sequence relationships between the sequences.

Family C3 contains the processing endopeptidases of picornaviruses, aphthoviruses, nepoviruses and comoviruses. The first group includes animal viruses such as those that cause polio and encephalomyocarditis. Picornaviruses encode one polyprotein that contains coat and core proteins, an RNA polymerase and two endopeptidases that are homologous to each other and responsible for excising the individual proteins. The larger of the endopeptidases, picornain 3C, is responsible for most of the cleavages, mainly at Gln↓Gly bonds, whereas the smaller endopeptidase, picornain 2A, releases itself by cleavage of a Tyr↓Gly bond. Picornain 2A has the further function of cleaving the eukaryote initiation factor 4G of the host cell [47]. Picornaviruses are unusual in possessing two homologous endopeptidases; in aphthoviruses, for example, cleavage of the eukaryote initiation factor 4G is performed by an endopeptidase unrelated to the picornains, the L-peptidase of family C28, clan CA (Chapter 491). Although the picornains C3 and 2A are homologous, the sequences are so different that they are assigned to the separate subfamilies C3A and C3B, respectively.

The polyprotein processing enzyme from aphthoviruses, which include foot-and-mouth disease, is so different in sequence from the picornains that it too is assigned to a separate subfamily, C3C. The processing endopeptidase for the hepatitis A virus is a member of a fourth subfamily, C3E, and a fifth subfamily, C3F, includes the protease from parechovirus 1.

Not all members of family C3 are animal pathogens, and subfamily C3D includes processing peptidases for the polyprotein of cowpea mosaic virus and other plant pathogens. A seventh subfamily, C3G, includes a polyprotein processing endopeptidase from rice tungro spherical virus. Catalytic activity has been demonstrated and the cleavage sites determined by site directed mutagenesis

to be Gln2526↓Asp and Gln2852↓Ala [48]; the catalytic residues are predicted to be His2680, Asp2722 and Cys2811 (see Alignment C3G in *MEROPS*). A homolog is also known from maize chlorotic dwarf waikavirus.

Family C4 contains one of the three polyprotein processing endopeptidases from potyviruses, which are plant pathogens. The NIa endopeptidase has been shown by site-directed mutagenesis to have at least a catalytic dyad that is in the order His, Cys [49]. Another similarity with family C3 is the preference for cleavage of Gln↓Gly bonds. The endopeptidase is a multifunctional molecule, acting also as the Vpg protein, which is attached covalently to the viral RNA [50].

Family C24 includes the processing endopeptidase from caliciviruses. Catalytic residues have been identified in the rabbit hemorrhagic disease virus, and a His, Cys dyad is the most likely order in the sequence, but a histidine C-terminal to the cysteine is also important, probably for substrate binding [51]. The endopeptidase cleaves Glu↓Gly bonds in the polyprotein [52].

Family C30 includes the processing endopeptidase from coronaviruses. Again, there is a His, Cys catalytic dyad, and cleavage is preferentially at Gln↓Gly bonds [53]. A representation of the three-dimensional structure of the processing peptidase from porcine transmissible gastroenteritis virus main protease is shown in Figure 404.21.

Family C37 contains the processing endopeptidases of Southampton and Norwalk caliciviruses. The catalytic cysteine has been identified in Southampton virus, and there is a preference for cleavage at Gln↓Gly bonds [54].

Family C38 comprises a putative endopeptidase from parsnip yellow fleck virus identified by comparison with the picornavirus polyprotein [55].

Family C74 includes the NS2 protein, one of the three polyprotein processing peptidases from pestiviruses. The catalytic triad has been established by mutagenesis as

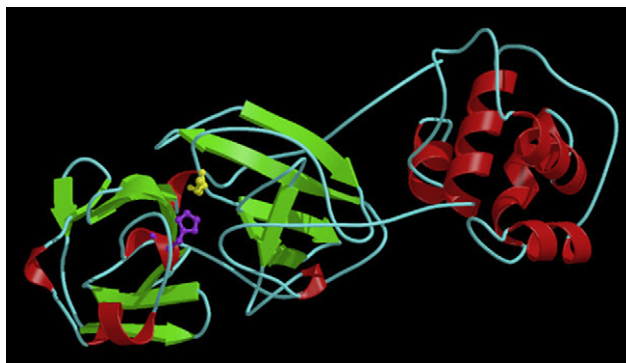


FIGURE 404.21 Richardson diagram of porcine transmissible gastroenteritis virus peptidase. The image was prepared from the Protein Data Bank entry (1LVO) as described in the Introduction (p. li). One monomer of the homohexamer is shown. Catalytic residues are shown in ball-and-stick representation: His41 and Cys144 (numbering as in PDB entry).

His, Glu, Cys, consistent with other families in subclass PA(C) [56].

Clan CD

Clan CD contains endopeptidases in which only a catalytic dyad appears to exist, in the order His, Cys in the sequence. Chen *et al.* [57] recognized that sequence motifs around the catalytic residues in mammalian legumain were similar to those in the caspases (family C14). Moreover, they saw similar motifs in the cysteine peptidases of families C11 (clostripain) and C25 (gingipain R). They proposed the existence of a clan CD of cysteine peptidases with a protein fold similar to that already known for the caspases (family C14). The proposal soon received support from the crystal structure of gingipain R2. The structure of caspase 1 (see [Figure 404.22](#),



FIGURE 404.22 Richardson diagram of the human caspase-1/Ac-Tyr-Val-Ala-Asp complex. The image was prepared from the Protein Data Bank entry (1ICE) as described in the Introduction (p. li). Catalytic residues are shown in ball-and-stick representation; His237 and Cys285 (numbering as in PDB entry). Ac-Tyr-Val-Ala is shown in gray in ball-and-stick representation.

Chapter 503) shows a distinctive α/β fold quite unlike those of clans CA and PA. In the sequence motif described by Chen *et al.* [57], the catalytic histidine is preceded by a block of hydrophobic residues and a glycine, and the catalytic cysteine is preceded by a second block of hydrophobic residues, a glutamine and an alanine (see [Figures 404.23](#) and [404.24](#)). A similar motif has been recognized in separase (family C50), RTX self-cleaving toxin (family C80) and prtH peptidase (family C84), and all three families have been added to clan CD. Similar motifs are present in proteins not yet known to be peptidases [58] so the clan may expand still further.

The peptidases in clan CD have some biochemical similarities. They are all endopeptidases with restricted specificity dominated by selectivity for the P1 residue of the substrate. Thus, in family C14 cleavage occurs predominately after Asp, in C13 after Asn, in C11 and C50 after Arg, and in C25 after Arg or Lys depending on the peptidase. This P1-dominated type of specificity resembles that of the unrelated chymotrypsin-like serine peptidases in subclass PA(S), and is quite unlike clan CA. The endopeptidases in the clan are not irreversibly inhibited by E-64, in contrast to most enzymes in clan CA. Clostripain and legumain react more rapidly with iodoacetamide than with iodoacetate, whereas iodoacetate is the faster inhibitor for papain and other members of family C1 (Chapter 419).

CrmA, a serpin from cowpox virus, is an inhibitor of the caspases [59] that also inhibits the serine endopeptidase granzyme B (Chapter 598) [60]. Rather than reflecting any structural relationship, however, this is

	23	24	28	29
	678901234567890123456		456789012345678901234	
		#		#
C14A	SDSTFLVFMSE	GLQEGICGIT	DKPKVIIQA	RGEKQGVVLL
C14B	DSLFLHYSGH	GGQTELDLGD	GVRLTALFDS	CHSGTVLDLPY
C11	ADKYVLIMAN	GGGAREKSNP	QSDLLAFDA	LMGTAEVAYQ
C13	EDRIFIFYSD	GGPGVLGMPN	YKEMVIYIEA	ESGSIFEGIM
C25	GGISLVNYTG	GSETAWGTSH	NQLPFIQDVA	VNGDFLFSMP
C50	QNSNLFVYIG	GGGEQYVRSK	TKIAPSFLLG	SSAAMKYYGK
C80	DGKLRWQLVG	GRDHSETNNT	NKPDHISIVG	SLVSDDKQKG

FIGURE 404.23 Conservation of sequence around catalytic residues in clan CD.

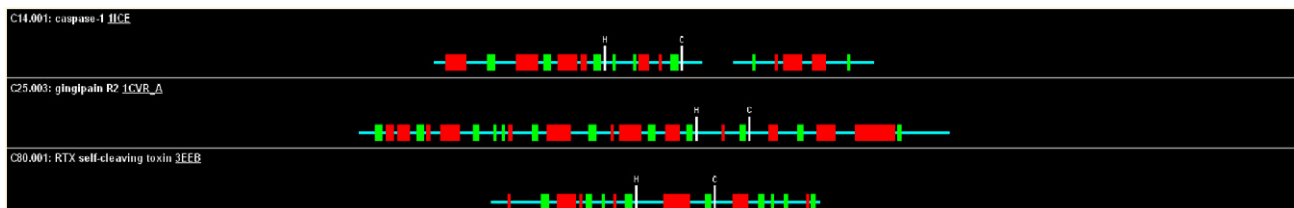


FIGURE 404.24 Comparison of secondary structure elements for selected members of clan CD.

undoubtedly the result of a similar specificity of the enzymes for aspartyl bonds.

Family C14 comprises a number of cytosolic endopeptidases that have strict specificity for hydrolysis of aspartyl bonds, the best known of which are caspase-1 (Chapter 505) and caspase-3 (Chapter 509). As can be seen in the alignment in *MEROPS*, there is a catalytic dyad in the order His, Cys in the sequence, as expected in clan CD.

Caspase-1 is synthesized as a single-chain precursor, and the activation by cleavage of four aspartyl bonds is presumably autocatalytic. The mature endopeptidase is a heterodimer of a 22 kDa heavy chain and a 10 kDa light chain [61]. Caspase-1 also mediates the processing of interleukin 1 β at aspartyl bonds [62], and was formerly known as interleukin 1 β -converting enzyme (ICE).

The discovery that the product of the *ced3* gene in *Caenorhabditis elegans*, which when defective increases the lifespan of the nematode [63] is a homolog of caspase-1 led to the discovery of a proteolytic cascade that leads to apoptosis in mammalian cells. Apoptosis is receptor mediated, and once the system is activated, proteins associate with one another by means of 'death domains' and 'death-effector domains'. Death-effector domains occur in a number of intracellular proteins and in caspases-8 and -10 [64]. FLIP, a newly reported inhibitor of caspase-8, occurs in two alternatively spliced forms, the longer of which, FLIP_L, contains two death-effector domains, and a caspase-like domain in which the catalytic Cys is replaced by Tyr, and the His by Arg (human) or Leu (mouse) [65] (see the alignment for **family C14** in *MEROPS*).

The **evolutionary tree** shows a pronounced divergence within the family, which is therefore divided into two subfamilies. Subfamily C14A includes caspases, whereas subfamily C14B includes metacaspases (see Chapter 517) and paracaspases (see Chapter 515). Paracaspase is known from a variety of animal species, including human, and provides an important link to a number of bacterial species, including homologs in the completed genomes of *Nostoc*, *Mesorhizobium loti* and *Xylella fastidiosa* [58]. The family is generally absent from archaea, but homologs have been detected in the proteomes of *Methanosarcina barkeri*, *Methanothermus fervidus* and *Thermococcus onnurineus*, which are presumably derived from horizontal gene transfers. There are also homologs in two insect ascoviruses, which are double-stranded DNA viruses.

Family C13 includes a number of endopeptidases that specifically cleave asparaginyl bonds. An asparaginyl endopeptidase was first identified in seeds of leguminous plants and was named 'legumain' [66]. The legumain of plant seeds has now been found in a wide variety of dicotyledonous plants and is proposed to be responsible for the post-translational processing of seed proteins prior to storage [67]. The processing of a few

of these proteins, such as concanavalin A, also involves protein splicing at asparaginyl bonds, and this transamidase reaction also can be mediated by legumain [68]. Homologous asparaginyl endopeptidases have been characterized from the blood flukes *Schistosoma mansoni* and *S. japonicum*; they were initially thought to be responsible for hemoglobin breakdown and were termed 'hemoglobinases' but this name is no longer appropriate. Mammalian legumain has been characterized and shown to be a lysosomal enzyme [69].

Legumain is not inhibited by compound E-64 and like clostripain it reacts more rapidly with iodoacetamide than with iodoacetate. The enzyme is inhibited by egg-white cystatin, which is also an inhibitor of peptidases in family C1 (clan CA), but it has been found that cystatin possesses two reactive sites, one for legumain and one for the papain-like enzymes, and is able to inhibit both simultaneously and independently [1] (Chapter 518).

Not a conventional endopeptidase, but also in the legumain family, is glycosylphosphatidylinositol:protein transamidase. This has been identified in *Saccharomyces* and humans, and is responsible for attaching glycosylphosphatidylinositol (GPI) anchors to the C-termini of newly synthesized proteins in the endoplasmic reticulum. A C-terminal peptide is removed before the preformed GPI anchor is attached, apparently in a single reaction. The transamidase differs considerably in sequence from the legumains, and the Cys residue that had been predicted to be catalytic in *Schistosoma* legumain [70] is replaced by Leu. The **evolutionary tree** demonstrates the marked difference between the forms of legumain and the GPI8 transamidase (Chapter 520).

Members of family C13 are known principally from animals, plants and fungi: the family is absent from the genome of the microsporidian *Encephalitozoon cuniculi* and not known from any other protozoa. However, despite generally being absent from prokaryotes, homologs are present in the gram-negative bacteria *Pseudomonas aeruginosa* and *Caulobacter crescentus*. The *Pseudomonas* homolog is the most divergent member of the family appears to be derived from a divergence that predates that of legumain and glycosylphosphatidylinositol:protein transamidase.

Family C11 contains only clostripain (Chapter 521). This is an endopeptidase discovered in *Clostridium histolyticum* that has a preference for hydrolysis of arginyl bonds. Clostripain is synthesized as a proprotein, and activation requires loss of a 23 residue propeptide and autolytic removal of an internal nonapeptide to produce a heterodimer consisting of a 131-residue light chain and a 336-residue heavy chain that contains the residues of the catalytic dyad. Clostripain differs from enzymes of the papain family in a number of respects: it is calcium dependent, it is more rapidly inactivated by

iodoacetamide than by iodoacetate, and E-64 gives only reversible inhibition attributable to the presence of an arginine-like sidechain in the compound. Forms of clostripain are known from *Clostridium perfringens* and *Thermotoga maritima*.

Family C25 is represented by gingipains R and K (Chapters 522 and 523), secreted endopeptidases from the pathogenic bacterium *Porphyromonas gingivalis*. Gingipain is a multidomain protein, containing not only a peptidase unit but also a C-terminal hemagglutinin domain (Figure 404.25). Gingipain R has specificity for arginyl bonds, whereas gingipain K cleaves after lysine residues. Both enzymes are thought to contribute to the disease processes in gingivitis.

Family C50 contains separase (Chapter 524), an endopeptidase that is involved in the separation of sister chromatids during the anaphase periods of mitosis or meiosis. In mitosis, the chromatids are bound together by the multi-subunit protein cohesin, and cleavage of the Scc1 subunit of cohesin by separase promotes chromatid separation. Cleavage occurs at the Arg180↓Arg and

Arg268↓Arg bonds of the Scc1 subunit. Separase also cleaves the kinetochore-associated protein Slk19 at the start of anaphase, which is necessary for the development of a stable spindle. In meiosis, the Rec8 protein performs the same role as cohesin and is also susceptible to cleavage by separase. A simple consensus for cleavage is an acidic residue at P4, Gly in P2, Arg in P1 and Arg, Lys or Ser in P1'. Members of family C50 are known only from eukaryotes, and are found in every eukaryote genome that has so far been completely sequenced.

Family C80 contains self-cleaving proteins that are precursors of bacterial toxins, including *Vibrio cholerae* RTX toxin (Chapter 525), and *Clostridium difficile* toxins A and B. However, peptidase activity is not restricted to self-cleavage and the mature RTX toxin has been shown to degrade the leucine-rich protein YopM [71]. The tertiary structure of the cholera RTX toxin has been solved [72] and shows a structure similar to that of caspase-1 (see Figure 404.26). Family C80 is therefore included in clan CD.

The PrtH protein, a virulence factor from *Tannerella* (formerly *Bacteroides*) *forsythusensis*, is the only characterized peptidase in **family C84**. However, very little is known about PrtH, except that it hydrolyzes Bz-Val-Gly-Arg-*p*-nitroanilide, which can be inhibited by standard cysteine peptidase inhibitors and the metal chelator EDTA [73]. The *prtH* gene is associated with the transition from a commensal to pathogenic organism, and increased levels are associated with release of cells from the substratum [74]. Although there is no information

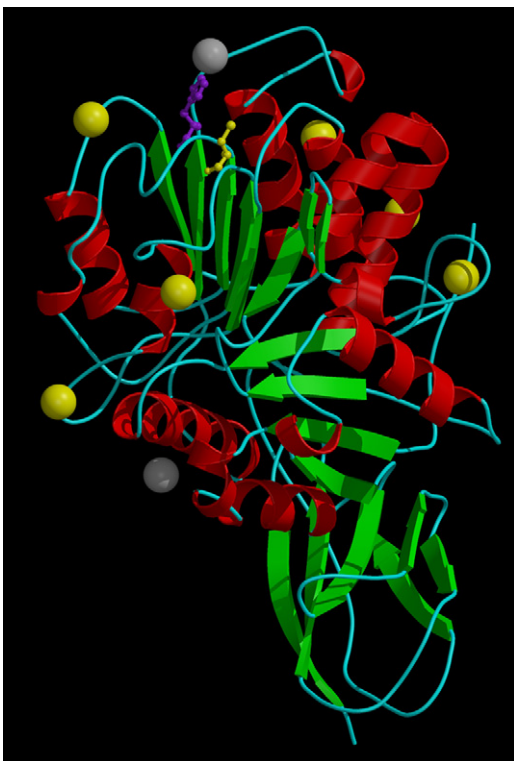


FIGURE 404.25 Richardson diagram of the gingipain R2/d-Phe-Phe-Arg-chloromethane complex. The image was prepared from the Protein Data Bank entry (1CVR) as described in the Introduction (p. li). The catalytic residues are shown in ball-and-stick representation: His211 and Cys244 (numbering as in PDB entry). The inhibitor d-Phe-Phe-Arg-chloromethane is shown in black in ball-and-stick representation. The structural zinc ions are shown as dark gray CPK spheres and the structural calcium ions as light gray CPK spheres.



FIGURE 404.26 Richardson diagram of the peptidase domain of the RTX self-cleaving toxin from *Vibrio cholerae*. The image was prepared from the Protein Data Bank entry (3EEB) as described in the Introduction (p. li). Catalytic residues are shown in ball-and-stick representation: His91 in purple, Cys140 in yellow (numbering as in PDB entry).



FIGURE 404.29 Richardson diagram of adenain from human adenovirus type 2. The image was prepared from the Protein Data Bank entry (1AVP) as described in the Introduction (p. li). Catalytic residues are shown in ball-and-stick representation: His54, Glu71 and Cys122 (numbering as in PDB entry). The eleven-residue cofactor is shown at the bottom of the image.



FIGURE 404.30 Richardson diagram of yeast Ulp1 endopeptidase. The image was prepared from the Protein Data Bank entry (1EUV) as described in the Introduction (p. li). Only chain A is shown. Catalytic residues are shown in ball-and-stick representation: His514, Asp531 and Cys580 (numbering as in PDB entry).

release Smt3 from targeted proteins. The enzyme is normally localized to the periphery of the nucleus, but is found within the nucleus during mitotic anaphase.

Family C55 is that of the YopJ endopeptidase from *Yersinia* (Chapter 69). The YopJ protein contributes to the disease process by causing host macrophages to undergo apoptosis and suppressing the action of cytokines such as TNF α and NF- κ B. YopJ is thought to act by cleaving the Bid protein, which then induces apoptosis by translocating to the mitochondrion, causing the release of cytochrome c and the activation of caspases 9, 3 and 7 [79]. YopJ is an outer membrane protein that is encoded on a plasmid.

There are a number of homologs in this family that are not known to be peptidases (even though all the catalytic residues are conserved) including avirulence protein AvrBsT from *Xanthomonas*. The family is known only from Proteobacteria. Although no catalytic residues have been identified for any member of the family, they are predicted to occur in the order His, Glu, Gln, Cys.

Family C57 contains the I7 protein from the vaccinia virus (Chapter 535), which is believed to be a polyprotein processing endopeptidase. Although no catalytic residues have been identified for any member of the family, they are predicted to occur in the order His, Asp/Asn, Gln, Cys.

The polyprotein processing endopeptidase present in the African swine fever virus (Chapter 534) is included in **family C63**. Although no tertiary structure has been solved, the

order of active site residues (His, Asn, Gln, Cys) is consistent with membership of clan CE, as is the restricted specificity for Gly-Gly \downarrow Xaa bonds [80]. Bacterial homologs are known from *Chlamydia*, *Xanthomonas*, *Pseudomonas* and *Mesorhizobium* but have low similarity to the eukaryotic sequences. The *Pseudomonas* and *Mesorhizobium* homologs are multidomain, mosaic proteins. No peptidase activity has yet been demonstrated for any of the bacterial homologs, but the gene has been shown to be encoded on a plasmid and essential for virulence in *Pseudomonas* [81]. No homologs are known from archaea.

Not all deubiquitinating enzymes are proteins with a papain-like fold. The ElaD peptidase from *Escherichia coli* is able to release ubiquitin from synthetic constructs with a preference for Lys63-linked ubiquitin chains, but also cleaving at Lys-48 [82,83]. This peptidase is included in **family C79**. This family is included in clan CE because we propose a His, Cys catalytic dyad (see the [alignment](#) in *MEROPS*).

Other Clans of Cysteine Peptidases

Clan CF, family C15. The crystal structure of pyroglutamy-peptidase I from *Thermococcus* (Figure 404.31) shows a catalytic triad with the residues in the order Glu, Cys, His in the sequence, which is different to any other clan of cysteine peptidases. The enzyme is an intracellular omega peptidase that removes an N-terminal pyroglutamy residue from a polypeptide. Cys144 and His166 have

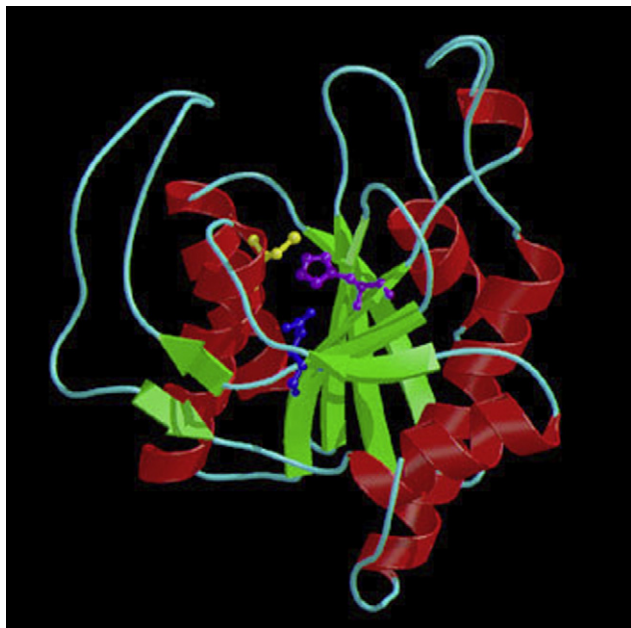


FIGURE 404.31 Richardson diagram of *Thermococcus litoralis* pyroglutamyl peptidase I. The image was prepared from the Protein Data Bank entry (1A2Z) as described in the Introduction (p. li). One molecule of the tetramer is shown. Catalytic residues are shown in ball-and-stick representation: Glu80, Cys143 and His167 (numbering as in PDB entry).

been identified as catalytic residues [84]. Human pyroglutamyl-peptidase I has been cloned and expressed [85]; (Chapter 548), and appears to be structurally very similar to the archaean and bacterial forms of the enzyme.

Family C46 contains the Hedgehog proteins. The Hedgehog proteins are formed of two structural domains, and the only peptidase activity is an autolytic reaction that separates the two domains. The reaction is cleavage of a conserved Gly258↓Cys259 bond, and a cholesterol moiety is simultaneously attached to the new C-terminal glycyl residue of the N-terminal domain. The 18-kDa N-terminal domain is structurally related to a zinc D-Ala-D-Ala carboxypeptidase in clan MD; it is not known to have any peptidase activity but has a function in the embryological development of dorsal-ventral patterning. Mutational studies have shown that residues Cys258 and His329 in the 25-kDa C-terminal domain are essential for the autolytic cleavage [86]. The reaction is not inhibited by any standard peptidase inhibitors, but requires DTT *in vitro*. The cleavage mechanism is believed to be similar to that of the activation of prohistidine decarboxylase [87], involving activation of Cys258 by His329 for a nucleophilic attack on the carbonyl of Gly257. There is a motif around the cleavage site that is similar to that found at the N-terminal splice junction of self-splicing proteins (Chapter 76), and splicing may involve a thioester intermediate [88]. The crystal structure of the Sonic

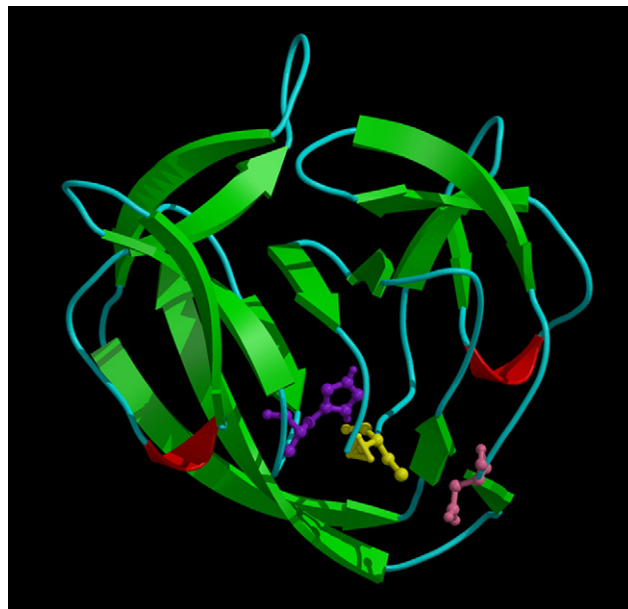


FIGURE 404.32 Richardson diagram of *Drosophila melanogaster* Hedgehog protein. The image was prepared from the Protein Data Bank entry (1AT0) as described in the Introduction (p. li). The autoprocessing domain is shown. Catalytic residues are shown in ball-and-stick representation: Cys258, Asp303, Thr326 and His329 (numbering as in PDB entry).

hedgehog C-terminal domain (Figure 404.32) shows that it consists of two tandem intein-like repeats [89], and family C46 is now included in a clan of mixed catalytic types, **PD**, along with the three intein families (see Chapter 76) [90].

Peptidases within **clan CL** have active site residues in the order His, Cys in the sequence, and a tertiary fold consisting of a closed β -barrel surrounded by helices (see Figure 404.33). The active site is at the end of the barrel. The fold is similar to that of the ykuD transpeptidase from *Bacillus subtilis* [91]. Members of the clan are involved in hydrolysis of bacterial cell wall peptides. **Family C60** includes peptidases known as sortases (Chapter 551) from Gram-positive bacteria, and sortase A from *Staphylococcus aureus* cleaves proteins at a Leu-Pro-Xaa-Thr↓Gly motif, and then catalyzes the formation of an amide bond between the newly exposed C-terminal Thr and the cross-linking pentapeptide of the bacterial cell wall [91]. The proteins that are attached in this way to the exterior of the cell wall can be virulence factors, and inhibition of sortases has been shown to decrease virulence. Surprisingly, given a lack of structural relationship to members of clan CA, sortases are inhibited by E-64. The evolutionary tree shows a deep divergence within the family, which is divided into two subfamilies. Subfamily C60A includes sortase A, whereas C60B includes sortase B. Sortase B, also from *Staphylococcus aureus*, recognizes a different cleavage motif to that of sortase A [92].



FIGURE 404.33 Richardson diagram of *Staphylococcus aureus* sortase A. The image was prepared from the Protein Data Bank entry (1T2P) as described in the Introduction (p. li). One monomer of the trimer is shown. Catalytic residues are shown in ball-and-stick representation: His120 in purple and Cys184 in yellow (numbering as in PDB entry).

The only other family in clan CL is **C82**. This also includes peptidases involved in bacterial cell wall metabolism. The L,D-transpeptidase from *Enterococcus faecium* (Chapter 552) cleaves the L-Lys↓D-Ala bonds from the crosslinking tetra- and pentapeptides, and also possesses cell-wall crosslinking activity [93]. The structure of a fragment of the peptidase has been solved and shows a fold similar to that of sortases [94].

The hepatitis C polyprotein is processed by virally encoded endopeptidases as well as by host endopeptidases. Most of the cleavages are performed by hepacivirin (Chapter 688), which is a serine-type endopeptidase with a trypsin-like fold. Cleavage of the NS2/3 site at a Leu↓Ala bond is mediated by the second virally encoded endopeptidase, endopeptidase 2, often termed the ‘NS2-3 endopeptidase’, not because of the cleavage site but because the endopeptidase consists of the NS2 protein and one-third of the NS3 protein (Chapter 554). Endopeptidase 2 is included in **family C18**. Site-directed mutagenesis had identified His952 and Cys993 as the catalytic dyad [95]. However, evidence that the NS2-3 endopeptidase is zinc dependent, including inhibition by EDTA and increased activity with the addition of zinc [96], had cast doubt on the identification of the catalytic type. The tertiary structure has been determined, and shows a fold unrelated to that of any other peptidase, hence family C18 is the only family in **clan CM** (see **Figure 404.34**). The active peptidase is a homodimer, and each monomer comprises an N-terminal, helical

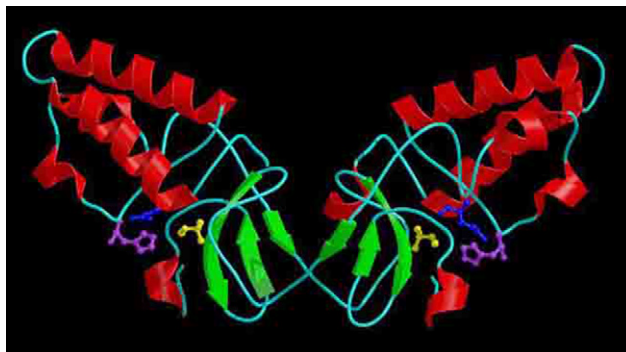


FIGURE 404.34 Richardson diagram of hepatitis C virus peptidase 2. The image was prepared from the Protein Data Bank entry (2HD0) as described in the Introduction (p. li). One homodimer is shown. Catalytic residues are shown in ball-and-stick representation for each monomer: His952 in purple, Glu972 in blue and Cys993 in yellow (numbering as in PDB entry).

subdomain, a C-terminal subdomain with a β -sheet and an extended linker region. The active site His and Glu are situated in the N-terminal domain and the Cys in the C-terminal domain. An active site is only formed in the dimer, when the His and Glu from one monomer oppose the Cys from the other, thus the NS2 dimer has two active sites [97]. The formation of an active site between different subunits in a dimer is unusual amongst peptidases, but does happen with retropepsins (see Chapter 44) and has been proposed for the *Pyrococcus furiosus* protease 1 (see Chapter 549). The tertiary structure of the NS2 dimer also revealed the presence of a structural zinc ion, which explains the metal-dependence.

Family C9 includes the non-structural polyprotein-processing endopeptidase from togaviruses, the nsP2 endopeptidase. The catalytic dyad has been identified by site-specific mutagenesis [98], and the nsP2 endopeptidase is a bifunctional molecule with the peptidase unit at the C-terminus and an N-terminal unit that is probably involved in RNA-binding during replication. The N-terminal domain is homologous to a similarly located domain in the cucumber green mottle tobamovirus. The crystal structure of the nsP2 endopeptidase from Venezuelan equine encephalitis virus has been solved, and shows a fold unrelated to that of any other peptidase [99] (see **Figure 404.35**). Hence C9 is the only family in **clan CN**. Until the structure was solved, it had been expected that members of family C9 would have a fold similar to that of papain.

Subclan PC(C)

Clan PC is another clan containing peptidases of different catalytic types, and subclan **PC(C)** contains the cysteine-type peptidases in the clan. Members of the clan have a



FIGURE 404.35 Richardson diagram of the equine encephalitis alpha-virus nsP2 peptidase. The image was prepared from the Protein Data Bank entry (2HWK) as described in the Introduction (p. li). Catalytic residues are shown in ball-and-stick representation: Cys477 in yellow and His546 in purple (numbering as in PDB entry).



FIGURE 404.36 Richardson diagram of human γ -glutamyl hydrolase. The image was prepared from the Protein Data Bank entry (1L9X) as described in the Introduction (p. li). One monomer of the tetramer is shown. Catalytic residues are shown in ball-and-stick representation: Cys110 and His220 (numbering as in PDB entry).

fold similar to that of class I glutamine amidotransferase, which consists of an $\alpha/\beta/\alpha$ sandwich. The nucleophile can be either a cysteine or a serine, and these are in equivalent positions.

Family C26, contains γ -glutamyl hydrolase (Chapter 550). The tertiary structure of human γ -glutamyl hydrolase (Figure 404.36) shows a homotetramer. A Cys, His catalytic dyad is evident, and the histidine is hydrogen-bonded to a glutamate and a tyrosine, either of which could be a third member of a catalytic triad [100].

Family C56 contains only protease I from *Pyrococcus* (see Chapter 549). Although the crystal structure has been solved (Figure 404.37), the catalytic mechanism is not known with certainty. The endopeptidase is a homotrimer, and Cys100 and His101 have been implicated in the catalytic mechanism, with the suggestion that a catalytic triad is formed between residues from opposing monomers, with Glu74 completing the active site. This would be a unique catalytic mechanism amongst the peptidases. His101 is poorly conserved, however, which may mean that very few members of the protein family are peptidases. His101 is also not in an equivalent position to His220 from γ -glutamyl hydrolase.

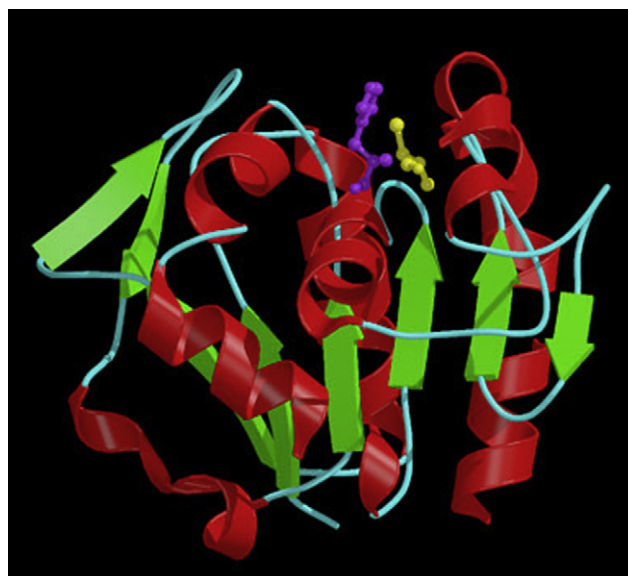


FIGURE 404.37 Richardson diagram of *Pyrococcus horikoshii* PfpI endopeptidase. The image was prepared from the Protein Data Bank entry (1G2I) as described in the Introduction (p. li). One monomer of the homotrimer is shown. Catalytic residues are shown in ball-and-stick representation: Cys100 and His101 (numbering as in PDB entry).

Family C40 contains dipeptidyl-peptidase VI from *Bacillus sphaericus*. This is an enzyme that is expressed during sporulation, and is responsible for degradation of bacterial cell wall components. Because the enzyme is cytoplasmic and is synthesized without signal and propeptides, it is presumably acting at a late stage in cell wall component turnover. There are a number of homologs of the *Bacillus* endopeptidase, but similarity is restricted to a C-terminal domain of 110 residues. These include nlpC lipoprotein from *Escherichia coli*, invasion-associated protein p60 from *Listeria* species, a starch-degrading enzyme from *Clostridium acetobutylicum*, and a phosphatase-associated protein from *Bacillus subtilis*. There are single conserved Cys and His residues). The crystal structures of several members of the family have been solved, and the fold has been described as ‘papain-like’ but primitive [101]. However, unlike papain where the catalytic Cys is at the N-terminus of a helix, in the structures from family C40 the active site Cys is at the C-terminus of a helix, which means the sequence must run in the opposite direction to that in papain (see Figure 404.38). No similarity between the structures can be detected using the DALI algorithm [102]. The structural resemblances are therefore assumed to be derived from convergent rather than divergent evolution, and proteins in C40 are assigned to the **clan CO**. All members of this family have either a signal peptide or a transmembrane region, indicating a cell-surface location.

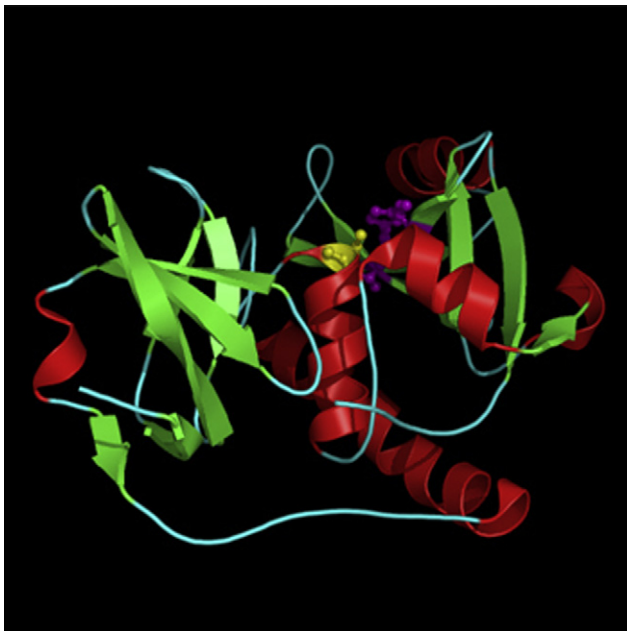


FIGURE 404.38 Richardson diagram of the NLP/P60-like putative peptidase from *Nostoc punctiforme*. The image was prepared from the Protein Data Bank entry (2EVR) as described in the Introduction (p. li). Catalytic residues are shown in ball-and-stick representation: His184 and His196 in purple and Cys134 in yellow (numbering as in PDB entry).

Cysteine Peptidases in Clan PB

Clan PB contains N-terminal nucleophile hydrolases, amongst which the most important peptidase is the proteasome. In the proteasome the N-terminal nucleophile is threonine, but there are members of the clan in which the nucleophile is cysteine. These are exclusively self-processing proenzymes that carry out a single peptide bond cleavage to form the active enzyme, but in contrast to the proteasome, the enzyme that is formed is not a peptidase. The cysteine-type enzymes of clan PB form a subclan PB(C), and are discussed more fully in Chapter 808.

Other Families of Cysteine Peptidases

There are a few other families of cysteine peptidases for which no tertiary structures are available and too little is known about the catalytic machinery to permit any conclusions about their relationships.

Family C6 contains one of the two cysteine endopeptidases found in potyviruses, the helper-component endopeptidase. (The second is the NIa peptidase, Chapter 544) The only cleavage performed by the helper-component endopeptidase is its own release from the polyprotein by cleavage of a Gly↓Gly bond; further processing is performed by the serine-type P1 endopeptidase (Chapter 689). The helper-component endopeptidase is a two-part molecule with the peptidase unit at the C-terminus, and the helper component, required for virus transmission from plant to plant by aphids, at the N-terminus. The catalytic residues have been identified by site-directed mutagenesis in tobacco etch virus [103].

Families C7 and **C8** contain the chestnut blight hypovirus endopeptidases p29 and p48. Chestnut blight is caused by the fungus *Cryphonectria parasitica*, but the symptoms are reduced if the fungus is infected by the double-stranded RNA hypovirus. The virus encodes two polyproteins, and the enzymes responsible for their processing constitute the smaller polyprotein. The p29 endopeptidase cleaves a single Gly↓Gly bond in the polyprotein to release itself and the p48 endopeptidase [104]. The p48 endopeptidase cleaves a single Gly↓Ala bond in the larger polyprotein [105]. Catalytic residues have been identified by mutagenesis for both endopeptidases [105,106]. The endopeptidases show no significant sequence similarity to any other peptidase, and each is considered a representative of a distinct family, although there is conservation of a Gly-Tyr-Cys-Tyr motif containing the catalytic Cys between the p29 endopeptidase (family C7) and family C6.

Family C21 contains the polyprotein-processing endopeptidase from tymoviruses. The 206 kDa polyprotein is cleaved at an Ala↓Thr bond [107] to release the 70 kDa polymerase from the C-terminus and a 150 kDa protein

that includes the endopeptidase and a helicase from the N-terminus. The catalytic residues have been identified by mutagenesis [108].

Family C23 includes the p223 polyprotein processing endopeptidase from the blueberry scorch virus. Cleavage is at a single Gly↓Ala bond, and site-directed mutagenesis has been used to identify the potential catalytic residues [109].

Family C27 includes the non-structural polyprotein-processing endopeptidase from rubella virus. Processing occurs at a single Gly↓Gly bond, and the Cys, His catalytic dyad has been identified [110].

Family C36 is represented by a polyprotein processing endopeptidase beet necrotic yellow vein virus [111].

Family C42 includes a polyprotein-processing endopeptidase from the beet yellows closterovirus. Processing occurs at a single Gly↓Gly bond, and the catalytic dyad has been identified by site-directed mutagenesis [112].

Family C53 contains the NPro endopeptidase of pestiviruses (Chapter 557). Pestiviruses possess two polyprotein-processing endopeptidases: the serine-type NS2-3 endopeptidase (Chapter 554), which is the more general processing activity; and the NPro endopeptidase, which is the N-terminal protein and releases itself from the polyprotein by autolytic cleavage of a Cys↓Ser bond. The residue Cys69 has now been identified as a catalytic residue [113].

Family C75 includes the AgrB protein from *Staphylococcus aureus*. The *agr* operon includes four genes whose products are involved in quorum sensing. The ArgA protein is a response regulator, ArgC is a sensor kinase and AgrD is a polypeptide that is integrated into the cytoplasmic membrane via its N-terminal region, and is the precursor for an autoinducing peptide (AIP) that is the ligand for AgrC. The AgrB protein is the peptidase that releases AIP from AgrD [114], following the removal of an N-terminal peptide by type I signal peptidase [115]. The active site residues His77 and Cys84 were identified in *Staphylococcus aureus* AgrB by [114]. AgrB is an integral membrane protein, with the active site close to the cytoplasm [116].

References

- [1] Alvarez-Fernandez, M., Barrett, A.J., Gerhartz, B., Dando, P.M., Ni, J., Abrahamson, M. (1999). Inhibition of mammalian legumain by some cystatins is due to a novel second reactive site. *J. Biol. Chem.* 274, 19195–19203.
- [2] Valente, R.H., Dragulev, B., Perales, J., Fox, J.W., Domont, G.B. (2001). BJ46a, a snake venom metalloproteinase inhibitor. Isolation, characterization, cloning and insights into its mechanism of action. *Eur. J. Biochem.* 268, 3042–3052.
- [3] Delaria, K., Fiorentino, L., Wallace, L., Tamburini, P., Brownell, E., Muller, D. (1994). Inhibition of cathepsin L-like cysteine proteases by cytotoxic T-lymphocyte antigen-2 beta. *J. Biol. Chem.* 269, 25172–25177.
- [4] Ward, W., Alvarado, L., Rawlings, N.D., Engel, J.C., Franklin, C., McKerrow, J.H. (1997). A primitive enzyme for a primitive cell: the protease required for excystation of *Giardia*. *Cell* 89, 437–444.
- [5] Ikeda, M., Takemura, T., Hino, S., Yoshioka, K. (2000). Molecular cloning, expression, and chromosomal localization of a human tubulointerstitial nephritis antigen. *Biochem. Biophys. Res. Commun.* 268, 225–230.
- [6] Hosfield, C.M., Elce, J.S., Davies, P.L., Jia, Z. (1999). Crystal structure of calpain reveals the structural basis for Ca²⁺-dependent protease activity and a novel mode of enzyme activation. *EMBO J.* 18, 6880–6889.
- [7] Strobl, S., Fernandez-Catalan, C., Braun, M., Huber, R., Masumoto, H., Nakagawa, K., Irie, A., Sorimachi, H., Bourenkow, G., Bartunik, H., Suzuki, K., Bode, W. (2000). The crystal structure of calcium-free human m-calpain suggests an electrostatic switch mechanism for activation by calcium. *Proc. Natl. Acad. Sci. USA* 97, 588–592.
- [8] Theopold, U., Pinter, M., Daffre, S., Tryselius, Y., Friedrich, P., Nassel, D.R., Hultmark, D. (1995). *CalpA*, a *Drosophila* calpain homolog specifically expressed in a small set of nerve, midgut, and blood cells. *Mol. Cell Biol.* 15, 824–834.
- [9] Delaney, S.J., Hayward, D.C., Barleben, F., Fischbach, K.F., Miklos, G.L. (1991). Molecular cloning and analysis of small optic lobes, a structural brain gene of *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA* 88, 7214–7218.
- [10] Kortt, A.A., Liu, T.Y. (1973). On the mechanism of action of streptococcal proteinase. II. Comparison of the kinetics of proteinase- and papain-catalyzed hydrolysis of *N*-acylamino acid esters. *Biochemistry* 12, 328–337.
- [11] Barrett, A.J., Kembhavi, A.A., Brown, M.A., Kirschke, H., Knight, C.G., Tamai, M., Hanada, K. (1982). L-*trans*-Epoxy succinyl-leucylamido(4-guanidino)butane (E-64) and its analogues as inhibitors of cysteine proteinases including cathepsins B, H and L. *Biochem. J.* 201, 189–198.
- [12] Johnston, S.C., Larsen, C.N., Cook, W.J., Wilkinson, K.D., Hill, C.P. (1997). Crystal structure of a deubiquitinating enzyme (human UCH-L3) at 1.8Å resolution. *EMBO J.* 16, 3787–3796.
- [13] Hu, M., Li, P., Li, M., Li, W., Yao, T., Wu, J.W., Gu, W., Cohen, R.E., Shi, Y. (2002). Crystal structure of a UBP-family deubiquitinating enzyme in isolation and in complex with ubiquitin aldehyde. *Cell* 111, 1041–1054.
- [14] Shanks, J., Burtnick, M.N., Brett, P.J., Waag, D.M., Spurgers, K.B., Ribot, W.J., Schell, M.A., Panchal, R.G., Gherardini, F.C., Wilkinson, K.D., Deshazer, D. (2009). *Burkholderia mallei* tssM encodes a putative deubiquitinase that is secreted and expressed inside infected RAW 264.7 murine macrophages. *Infect. Immun.* 77, 1636–1648.
- [15] Makarova, K.S., Aravind, L., Koonin, E.V. (2000). A novel superfamily of predicted cysteine proteases from eukaryotes, viruses and *Chlamydia pneumoniae*. *Trends Biochem. Sci.* 25, 50–52.
- [16] Evans, P.C., Smith, T.S., Lai, M.J., Williams, M.G., Burke, D.F., Heyninck, K., Kreike, M.M., Beyaert, R., Blundell, T.L., Kilshaw, P.J. (2003). A novel type of deubiquitinating enzyme. *J. Biol. Chem.* 278, 23180–23186.

- [17] Komander, D., Barford, D. (2008). Structure of the A20 OTU domain and mechanistic insights into deubiquitination. *Biochem. J.* 409, 77–85.
- [18] Nanao, M.H., Tcherniuk, S.O., Chroboczek, J., Dideberg, O., Dessen, A., Balakirev, M.Y. (2004). Crystal structure of human otubain 2. *EMBO Rep.* 5, 783–788.
- [19] Wang, T., Yin, L., Cooper, E.M., Lai, M.Y., Dickey, S., Pickart, C.M., Fushman, D., Wilkinson, K.D., Cohen, R.E., Wolberger, C. (2009). Evidence for bidentate substrate binding as the basis for the K48 linkage specificity of otubain 1. *J. Mol. Biol.* 386, 1011–1023.
- [20] Kovalenko, A., Chable-Bessia, C., Cantarella, G., Israel, A., Wallach, D., Courtois, G. (2003). The tumour suppressor CYLD negatively regulates NF-kappaB signalling by deubiquitination. *Nature* 424, 801–805.
- [21] Komander, D., Lord, C.J., Scheel, H., Swift, S., Hofmann, K., Ashworth, A., Barford, D. (2008). The structure of the CYLD USP domain explains its specificity for Lys63-linked polyubiquitin and reveals a B Box module. *Mol. Cell* 29, 451–464.
- [22] Kayagaki, N., Phung, Q., Chan, S., Chaudhari, R., Quan, C., O'Rourke, K.M., Eby, M., Pietras, E., Cheng, G., Bazan, J.F., Zhang, Z., Arnott, D., Dixit, V.M. (2007). DUBA: a deubiquitinase that regulates type I interferon production. *Science* 318, 1628–1632.
- [23] Nicastro, G., Menon, R.P., Masino, L., Knowles, P.P., McDonald, N.Q., Pastore, A. (2005). The solution structure of the Josephin domain of ataxin-3: structural determinants for molecular recognition. *Proc. Natl. Acad. Sci. USA* 102, 10493–10498.
- [24] Messick, T.E., Russell, N.S., Iwata, A.J., Sarachan, K.L., Shiekhhattar, R., Shanks, J.R., Reyes-Turcu, F.E., Wilkinson, K.D., Marmorstein, R. (2008). Structural basis for ubiquitin recognition by the OTU1 ovarian tumor domain protein. *J. Biol. Chem.* 283, 11038–11049.
- [25] Schlieker, C., Weihofen, W.A., Frijns, E., Kattenhorn, L.M., Gaudet, R., Ploegh, H.L. (2007). Structure of a herpesvirus-encoded cysteine protease reveals a unique class of deubiquitinating enzymes. *Mol. Cell* 25, 677–687.
- [26] Frias-Staheli, N., Giannakopoulos, N.V., Kikkert, M., Taylor, S.L., Bridgen, A., Paragas, J., Richt, J.A., Rowland, R.R., Schmaljohn, C.S., Lenschow, D.J., Snijder, E.J., García-Sastre, A., Vergin, H.W. (2007). Ovarian tumor domain-containing viral proteases evade ubiquitin- and ISG15-dependent innate immune responses. *Cell Host Microbe* 2, 404–416.
- [27] Kang, S.H., Kim, G.R., Seong, M., Baek, S.H., Seol, J.H., Bang, O.S., Ovaa, H., Tatsumi, K., Komatsu, M., Tanaka, K., Chung, C.H. (2007). Two novel ubiquitin-fold modifier 1 (Ufm1)-specific proteases, UfSP1 and UfSP2. *J. Biol. Chem.* 282, 5256–5262.
- [28] Ha, B.H., Ahn, H.C., Kang, S.H., Tanaka, K., Chung, C.H., Kim, E.E. (2008). Structural basis for Ufm1 processing by UfSP1. *J. Biol. Chem.* 283, 14893–14900.
- [29] Kumanomidou, T., Mizushima, T., Komatsu, M., Suzuki, A., Tanida, I., Sou, Y.S., Ueno, T., Kominami, E., Tanaka, K., Yamane, T. (2006). The crystal structure of human Atg4b, a processing and de-conjugating enzyme for autophagosome-forming modifiers. *J. Mol. Biol.* 355, 612–618.
- [30] Sugawara, K., Suzuki, N.N., Fujioka, Y., Mizushima, N., Ohsumi, Y., Inagaki, F. (2005). Structural basis for the specificity and catalysis of human Atg4B responsible for mammalian autophagy. *J. Biol. Chem.* 280, 40058–40065.
- [31] Hofmann, B., Hecht, H.J., Kiess, M., Schomburg, D. (1993). Crystal structure of a thiol proteinase from *Staphylococcus aureus* V8 in the E-64 inhibitor complex. *Acta Crystallogr. A* 49(Suppl), 102.
- [32] Ishii, S., Yano, T., Ebihara, A., Okamoto, A., Manzoku, M., Hayashi, H. (2010). Crystal structure of the peptidase domain of streptococcus ComA, a bifunctional ATP-binding cassette transporter involved in the quorum-sensing pathway. *J. Biol. Chem.* 285, 10777–10785.
- [33] Navarre, W.W., Ton-That, H., Faull, K.F., Schneewind, O. (1999). Multiple enzymatic activities of the murein hydrolase from staphylococcal phage phi11. Identification of a D-alanyl-glycine endopeptidase activity. *J. Biol. Chem.* 274, 15847–15856.
- [34] Hotson, A., Mudgett, M.B. (2004). Cysteine proteases in phytopathogenic bacteria: identification of plant targets and activation of innate immunity. *Curr. Opin. Plant Biol.* 7, 384–390.
- [35] Axtell, M.J., Staskawicz, B.J. (2003). Initiation of RPS2-specified disease resistance in *Arabidopsis* is coupled to the AvrRpt2-directed elimination of RIN4. *Cell* 112, 369–377.
- [36] Shao, F., Merritt, P.M., Bao, Z., Innes, R.W., Dixon, J.E. (2002). A *Yersinia* effector and a *Pseudomonas* avirulence protein define a family of cysteine proteases functioning in bacterial pathogenesis. *Cell* 109, 575–588.
- [37] Lopez-Solanilla, E., Bronstein, P.A., Schneider, A.R., Collmer, A. (2004). HopPtoN is a *Pseudomonas syringae* Hrp (type III secretion system) cysteine protease effector that suppresses pathogen-induced necrosis associated with both compatible and incompatible plant interactions. *Mol. Microbiol.* 54, 353–365.
- [38] Wenig, K., Chatwell, L., von Pawel-Rammingen, U., Bjorck, L., Huber, R., Sondermann, P. (2004). Structure of the streptococcal endopeptidase IdeS, a cysteine proteinase with strict specificity for IgG. *Proc. Natl. Acad. Sci. USA* 101, 17371–17376.
- [39] Vivares, D., Arnoux, P., Pignol, D. (2005). A papain-like enzyme at work: native and acyl-enzyme intermediate structures in phytochelatin synthesis. *Proc. Natl. Acad. Sci. USA* 102, 18848–18853.
- [40] Baker, S.C., Yokomori, K., Dong, S., Carlisle, R., Gorbalenya, A.E., Koonin, E.V., Lai, M.M. (1993). Identification of the catalytic sites of a papain-like cysteine proteinase of murine coronavirus. *J. Virol.* 67, 6056–6063.
- [41] Kim, J.C., Spence, R.A., Currier, P.F., Lu, X., Denison, M.R. (1995). Coronavirus protein processing and RNA synthesis is inhibited by the cysteine proteinase inhibitor E64d. *Virology* 208, 1–8.
- [42] Den Boon, J.A., Faaberg, K.S., Meulenberg, J.J., Wassenaar, A.L., Plegemann, P.G., Gorbalenya, A.E., Snijder, E.J. (1995). Processing and evolution of the N-terminal region of the arterivirus replicase ORF1a protein: identification of two papain-like cysteine proteases. *J. Virol.* 69, 4500–4505.
- [43] Snijder, E.J., Wassenaar, A.L., Spaan, W.J., Gorbalenya, A.E. (1995). The arterivirus Nsp2 protease. An unusual cysteine protease with primary structure similarities to both papain-like and chymotrypsin-like proteases. *J. Biol. Chem.* 270, 16671–16676.
- [44] Luo, Y., Pfister, P., Leisinger, T., Wasserfallen, A. (2002). Pseudomurein endoisopeptidases PeiW and PeiP, two moderately related members of a novel family of proteases produced in *Methanothermobacter* strains. *FEMS Microbiol. Lett.* 208, 47–51.

- [45] Bazan, J.F., Fletterick, R.J. (1988). Viral cysteine proteases are homologous to the trypsin-like family of serine proteases: structural and functional implications. *Proc. Natl. Acad. Sci. USA* 85, 7872–7876.
- [46] Petersen, J.F., Cherney, M.M., Liebig, H.D., Skern, T., Kuechler, E., James, M.N. (1999). The structure of the 2A proteinase from a common cold virus: a proteinase responsible for the shut-off of host-cell protein synthesis. *EMBO J.* 18, 5463–5475.
- [47] Lamphear, B.J., Rhoads, R.E. (1996). A single amino acid change in protein synthesis initiation factor 4G renders cap-dependent translation resistant to picornaviral 2A proteases. *Biochemistry* 35, 15726–15733.
- [48] Thole, V., Hull, R. (1998). Rice tungro spherical virus polyprotein processing: identification of a virus-encoded protease and mutational analysis of putative cleavage sites. *Virology* 247, 106–114.
- [49] Dougherty, W.G., Parks, T.D., Cary, S.M., Bazan, J.F., Fletterick, R.J. (1989). Characterization of the catalytic residues of the tobacco etch virus 49-kDa proteinase. *Virology* 172, 302–310.
- [50] Rorrer, K., Parks, T.D., Scheffler, B., Bevan, M., Dougherty, W.G. (1992). Autocatalytic activity of the tobacco etch virus NIa proteinase in viral and foreign protein sequences. *J. Gen. Virol.* 73, 775–783.
- [51] Boniotti, B., Wirblich, C., Sibilia, M., Meyers, G., Thiel, H.J., Rossi, C. (1994). Identification and characterization of a 3C-like protease from rabbit hemorrhagic disease virus, a calicivirus. *J. Virol.* 68, 6487–6495.
- [52] Wirblich, C., Sibilia, M., Boniotti, M.B., Rossi, C., Thiel, H.J., Meyers, G. (1995). 3C-like protease of rabbit hemorrhagic disease virus: identification of cleavage sites in the ORF1 polyprotein and analysis of cleavage specificity. *J. Virol.* 69, 7159–7168.
- [53] Liu, D.X., Brown, T.D. (1995). Characterisation and mutational analysis of an ORF 1a-encoding proteinase domain responsible for proteolytic processing of the infectious bronchitis virus 1a/1b polyprotein. *Virology* 209, 420–427.
- [54] Liu, B., Clarke, I.N., Lambden, P.R. (1996). Polyprotein processing in Southampton virus: identification of 3C-like protease cleavage sites by *in vitro* mutagenesis. *J. Virol.* 70, 2605–2610.
- [55] Turnbull-Ross, A.D., Mayo, M.A., Reavy, B., Murant, A.F. (1993). Sequence analysis of the parsnip yellow fleck virus polyprotein: evidence of affinities with picornaviruses. *J. Gen. Virol.* 74, 555–561.
- [56] Lackner, T., Muller, A., Pankraz, A., Becher, P., Thiel, H.J., Gorbalenya, A.E., Tautz, N. (2004). Temporal modulation of an autoprotease is crucial for replication and pathogenicity of an RNA virus. *J. Virol.* 78, 10765–10775.
- [57] Chen, J.M., Rawlings, N.D., Stevens, R.A., Barrett, A.J. (1998). Identification of the active site of legumain links it to caspases, clostripain and gingipains in a new clan of cysteine endopeptidases. *FEBS Lett.* 441, 361–365.
- [58] Aravind, L., Koonin, E.V. (2002). Classification of the caspase-hemoglobinase fold: detection of new families and implications for the origin of the eukaryotic separins. *Proteins* 46, 355–367.
- [59] Ray, C.A., Black, R.A., Kronheim, S.R., Greenstreet, T.A., Sleath, P.R., Salvesen, G.S., Pickup, D.J. (1992). Viral inhibition of inflammation: cowpox virus encodes an inhibitor of the interleukin-1 beta converting enzyme. *Cell* 69, 597–604.
- [60] Quan, L.T., Caputo, A., Bleackley, R.C., Pickup, D.J., Salvesen, G.S. (1995). Granzyme B is inhibited by the cowpox virus serpin cytokine response modifier A. *J. Biol. Chem.* 270, 10377–10379.
- [61] Yamin, T.T., Ayala, J.M., Miller, D.K. (1996). Activation of the native 45-kDa precursor form of interleukin-1-converting enzyme. *J. Biol. Chem.* 271, 13273–13282.
- [62] Thornberry, N.A., Molineaux, S.M. (1995). Interleukin-1 β converting enzyme: A novel cysteine protease required for IL-1 β production and implicated in programmed cell death. *Protein Sci.* 4, 3–12.
- [63] Xue, D., Shaham, S., Horvitz, H.R. (1996). The *Caenorhabditis elegans* cell-death protein CED-3 is a cysteine protease with substrate specificities similar to those of the human CPP32 protease. *Genes Dev.* 10, 1073–1083.
- [64] Wallach, D. (1997). Apoptosis. Placing death under control. *Nature* 388, 125–126.
- [65] Irmeler, M., Thome, M., Hahne, M., Schneider, P., Hofmann, K., Steiner, V., Bodmer, J.L., Schroter, M., Burns, K., Mattmann, C., Rimoldi, D., French, L.E., Tschopp, J. (1997). Inhibition of death receptor signals by cellular FLIP. *Nature* 388, 190–195.
- [66] Kembhavi, A.A., Buttle, D.J., Knight, C.G., Barrett, A.J. (1993). The two cysteine endopeptidases of legume seeds: purification and characterization by use of specific fluorometric assays. *Arch. Biochem. Biophys.* 303, 208–213.
- [67] Hara-Nishimura, I., Shimada, T., Hiraiwa, N., Nishimura, M. (1995). Vacuolar processing enzyme responsible for maturation of seed proteins. *J. Plant Physiol.* 145, 632–640.
- [68] Min, W., Jones, D.H. (1994). *In vitro* splicing of concanavalin A is catalyzed by asparaginyl endopeptidase. *Nat. Struct. Biol.* 1, 502–504.
- [69] Chen, J.M., Dando, P.M., Rawlings, N.D., Brown, M.A., Young, N.E., Stevens, R.A., Hewitt, E., Watts, C., Barrett, A.J. (1997). Cloning, isolation, and characterization of mammalian legumain, an asparaginyl endopeptidase. *J. Biol. Chem.* 272, 8090–8098.
- [70] Gotz, B., Klinkert, M.Q. (1993). Expression and partial characterization of a cathepsin B-like enzyme (Sm31) and a proposed ‘haemoglobinase’ (Sm32) from *Schistosoma mansoni*. *Biochem. J.* 290, 801–806.
- [71] Prochazkova, K., Shuvalova, L.A., Minasov, G., Voburka, Z., Anderson, W.F., Satchell, K.J. (2009). Structural and molecular mechanism for autoprocessing of MARTX toxin of *Vibrio cholerae* at multiple sites. *J. Biol. Chem.* 284, 26557–26568.
- [72] Lupardus, P.J., Shen, A., Bogyo, M., Garcia, K.C. (2008). Small molecule-induced allosteric activation of the *Vibrio cholerae* RTX cysteine protease domain. *Science* 322, 265–268.
- [73] Saito, T., Ishihara, K., Kato, T., Okuda, K. (1997). Cloning, expression, and sequencing of a protease gene from *Bacteroides forsythus* ATCC 43037 in *Escherichia coli*. *Infect. Immun.* 65, 4888–4891.
- [74] Hamlet, S.M., Ganashan, N., Cullinan, M.P., Westerman, B., Palmer, J.E., Seymour, G.J. (2008). A 5-year longitudinal study of *Tannerella forsythia* prH genotype: association with loss of attachment. *J. Periodontol.* 79, 144–149.
- [75] Pei, J., Grishin, N.V. (2009). Prediction of a caspase-like fold in *Tannerella forsythia* virulence factor PrtH. *Cell Cycle* 8, 1453–1455.
- [76] Weber, J.M., Tihanyi, K. (1994). Adenovirus endopeptidases. *Methods Enzymol.* 244, 595–604.

- [77] Ding, J., McGrath, W.J., Sweet, R.M., Mangel, W.F. (1996). Crystal structure of the human adenovirus proteinase with its 11 amino acid cofactor. *EMBO J.* 15, 1778–1783.
- [78] Webster, A., Leith, I.R., Hay, R.T. (1994). Activation of adenovirus-coded protease and processing of preterminal protein. *J. Virol.* 68, 7292–7300.
- [79] Denecker, G., Declercq, W., Geuijen, C.A., Boland, A., Benabdillah, R., van Gurp, M., Sory, M.P., Vandenebeele, P., Cornelis, G.R. (2001). *Yersinia enterocolitica* YopP-induced apoptosis of macrophages involves the apoptotic signaling cascade upstream of bid. *J. Biol. Chem.* 276, 19706–19714.
- [80] Lopez-Otin, C., Simon-Mateo, C., Martinez, L., Vinuela, E. (1989). Gly-Gly-X, a novel consensus sequence for the proteolytic processing of viral and cellular proteins. *J. Biol. Chem.* 264, 9107–9110.
- [81] Kamiuntun, H. (2002). Isolation and characterization of virulence gene psvA on a plasmid of *Pseudomonas syringae* pv. *eriotobryae*. *Nippon Shokubutsu Byori Gakkaiho* 65, 501–509.
- [82] Catic, A., Misaghi, S., Korbel, G.A., Ploegh, H.L. (2007). ElaD, a deubiquitinating protease expressed by *E. coli*. *PLoS ONE* 2, e381.
- [83] Rytkonen, A., Poh, J., Garmendia, J., Boyle, C., Thompson, A., Liu, M., Freemont, P., Hinton, J.C., Holden, D.W. (2007). SseL, a *Salmonella* deubiquitinase required for macrophage killing and virulence. *Proc. Natl. Acad. Sci. USA* 104, 3502–3507.
- [84] Le Saux, O., Gonzales, T., Robert-Baudouy, J. (1996). Mutational analysis of the active site of *Pseudomonas fluorescens* pyrrolidone carboxyl peptidase. *J. Bacteriol.* 178, 3308–3313.
- [85] Dando, P.M., Fortunato, M., Strand, G.B., Smith, T.S., Barrett, A.J. (2003). Pyroglutamyl-peptidase I: cloning, sequencing, and characterisation of the recombinant human enzyme. *Protein Expr. Purif.* 28, 111–119.
- [86] Lee, J.J., Ekker, S.C., von Kessler, D.P., Porter, J.A., Sun, B.I., Beachy, P.A. (1994). Autoproteolysis in hedgehog protein biogenesis. *Science* 266, 1528–1537.
- [87] Vanderslice, P., Copeland, W.C., Robertus, J.D. (1988). Site-directed alteration of serine 82 causes nonproductive chain cleavage in prohistidine decarboxylase. *J. Biol. Chem.* 263, 10583–10586.
- [88] Porter, J.A., von Kessler, D.P., Ekker, S.C., Young, K.E., Lee, J.J., Moses, K., Beachy, P.A. (1995). The product of hedgehog autoproteolytic cleavage active in local and long-range signalling. *Nature* 374, 363–366.
- [89] Hall, T.M., Porter, J.A., Young, K.E., Koonin, E.V., Beachy, P.A., Leahy, D.J. (1997). Crystal structure of a hedgehog autoprocessing domain: homology between hedgehog and self-splicing proteins. *Cell* 91, 85–97.
- [90] Rawlings, N.D., Barrett, A.J., Bateman, A. (2011). Asparagine peptidase lyases: a seventh catalytic type of proteolytic enzymes. *J. Biol. Chem.* 286, 38321–38328.
- [91] Zong, Y., Bice, T.W., Ton-That, H., Schneewind, O., Narayana, S.V. (2004). Crystal structures of *Staphylococcus aureus* sortase A and its substrate complex. *J. Biol. Chem.* 279, 31383–31389.
- [92] Mazmanian, S.K., Ton-That, H., Su, K., Schneewind, O. (2002). An iron-regulated sortase anchors a class of surface protein during *Staphylococcus aureus* pathogenesis. *Proc. Natl. Acad. Sci. USA* 99, 2293–2298.
- [93] Magnet, S., Arbeloa, A., Mainardi, J.L., Hugonnet, J.E., Fourgeaud, M., Dubost, L., Marie, A., Delfosse, V., Mayer, C., Rice, L.B., Arthur, M. (2007). Specificity of L,D-transpeptidases from gram-positive bacteria producing different peptidoglycan chemotypes. *J. Biol. Chem.* 282, 13151–13159.
- [94] Biarrotte-Sorin, S., Hugonnet, J.E., Delfosse, V., Mainardi, J.L., Gutmann, L., Arthur, M., Mayer, C. (2006). Crystal structure of a novel β -lactam-insensitive peptidoglycan transpeptidase. *J. Mol. Biol.* 359, 533–538.
- [95] Grakoui, A., McCourt, D.W., Wychowski, C., Feinstone, S.M., Rice, C.M. (1993). A second hepatitis C virus-encoded proteinase. *Proc. Natl. Acad. Sci. USA* 90, 10583–10587.
- [96] Hijikata, M., Mizushima, H., Akagi, T., Mori, S., Kakiuchi, N., Kato, N., Tanaka, T., Kimura, K., Shimotohno, K. (1993). Two distinct proteinase activities required for the processing of a putative nonstructural precursor protein of hepatitis C virus. *J. Virol.* 67, 4665–4675.
- [97] Lorenz, I.C., Marcotrigiano, J., Dentzer, T.G., Rice, C.M. (2006). Structure of the catalytic domain of the hepatitis C virus NS2-3 protease. *Nature* 442, 831–835.
- [98] Strauss, E.G., de Groot, R.J., Levinson, R., Strauss, J.H. (1992). Identification of the active site residues in the nsP2 proteinase of sindbis virus. *Virology* 191, 932–940.
- [99] Russo, A.T., White, M.A., Watowich, S.J. (2006). The crystal structure of the Venezuelan equine encephalitis alphavirus nsP2 protease. *Structure* 14, 1449–1458.
- [100] Li, H., Ryan, T.J., Chave, K.J., Van Roey, P. (2002). Three-dimensional structure of human γ -glutamyl hydrolase. A class I glutamine amidotransferase adapted for a complex substrate. *J. Biol. Chem.* 277, 24522–24529.
- [101] Xu, Q., Sudek, S., McMullan, D., Miller, M.D., Geierstanger, B., Jones, D.H., Krishna, S.S., Spraggon, G., Bursalay, B., Abdubek, P., Acosta, C., Ambing, E., Astakhova, T., Axelrod, H.L., Carlton, D., Caruthers, J., Chiu, H.J., Clayton, T., Deller, M.C., Duan, L., Elias, Y., Elsliger, M.A., Feuerhelm, J., Grzechnik, S.K., Hale, J., Han, G.W., Haugen, J., Jaroszewski, L., Jin, K.K., Klock, H.E., Knuth, M.W., Kozbial, P., Kumar, A., Marciano, D., Morse, A.T., Nigoghossian, E., Okach, L., Oommachen, S., Paulsen, J., Reyes, R., Rife, C.L., Trout, C.V., van den Bedem, H., Weekes, D., White, A., Wolf, G., Zubieta, C., Hodgson, K.O., Wooley, J., Deacon, A.M., Godzik, A., Lesley, S.A., Wilson, I.A. (2009). Structural basis of murein peptide specificity of a γ -D-glutamyl-L-diamino acid endopeptidase. *Structure* 17, 303–313.
- [102] Holm, L., Sander, C. (1995). Dali: a network tool for protein structure comparison. *Trends Biochem. Sci.* 20, 478–480.
- [103] Oh, C.S., Carrington, J.C. (1989). Identification of essential residues in potyvirus proteinase HC-Pro by site-directed mutagenesis. *Virology* 173, 692–699.
- [104] Choi, G.H., Shapira, R., Nuss, D.L. (1991). Cotranslational autoproteolysis involved in gene expression from a double-stranded RNA genetic element associated with hypovirulence of the chestnut blight fungus. *Proc. Natl. Acad. Sci. USA* 88, 1167–1171.
- [105] Shapira, R., Nuss, D.L. (1991). Gene expression by a hypovirulence-associated virus of the chestnut blight fungus involves two papain-like protease activities. Essential residues and cleavage site requirements for p48 autoproteolysis. *J. Biol. Chem.* 266, 19419–19425.
- [106] Choi, G.H., Pawlyk, D.M., Nuss, D.L. (1991). The autocatalytic protease p29 encoded by a hypovirulence-associated virus of the

- chestnut blight fungus resembles the potyvirus-encoded protease HC-Pro. *Virology* 183, 747–752.
- [107] Bransom, K.L., Wallace, S.E., Dreher, T.W. (1996). Identification of the cleavage site recognized by the turnip yellow mosaic virus protease. *Virology* 217, 404–406.
- [108] Bransom, K.L., Dreher, T.W. (1994). Identification of the essential cysteine and histidine residues of the turnip yellow mosaic virus protease. *Virology* 198, 148–154.
- [109] Lawrence, D.M., Rozanov, M.N., Hillman, B.I. (1995). Autocatalytic processing of the 223-kDa protein of Blueberry scorch carlavirus by a papain-like proteinase. *Virology* 207, 127–135.
- [110] Chen, J.P., Strauss, J.H., Strauss, E.G., Frey, T.K. (1996). Characterization of the rubella virus nonstructural protease domain and its cleavage site. *J. Virol.* 70, 4707–4713.
- [111] Hehn, A., Fritsch, C., Richards, K.E., Guilley, H., Jonard, G. (1997). Evidence for *in vitro* and *in vivo* autocatalytic processing of the primary translation product of beet necrotic yellow vein virus RNA 1 by a papain-like proteinase. *Arch. Virol.* 142, 1051–1058.
- [112] Agranovsky, A.A., Koonin, E.V., Boyko, V.P., Maiss, E., Frotschl, R., Lunina, N.A., Atabekov, J.G. (1994). Beet yellows closterovirus: complete genome structure and identification of a leader papain-like thiol protease. *Virology* 198, 311–324.
- [113] Rumenapf, T., Stark, R., Heimann, M., Thiel, H.J. (1998). N-terminal protease of pestiviruses: identification of putative catalytic residues by site-directed mutagenesis. *J. Virol.* 72, 2544–2547.
- [114] Qiu, R., Pei, W., Zhang, L., Lin, J., Ji, G. (2005). Identification of the putative staphylococcal AgrB catalytic residues involving the proteolytic cleavage of AgrD to generate autoinducing peptide. *J. Biol. Chem.* 280, 16695–16704.
- [115] Kavanaugh, J.S., Thoendel, M., Horswill, A.R. (2007). A role for type I signal peptidase in *Staphylococcus aureus* quorum sensing. *Mol. Microbiol.* 65, 780–798.
- [116] Zhang, L., Ji, G. (2004). Identification of a staphylococcal AgrB segment(s) responsible for group-specific processing of AgrD by gene swapping. *J. Bacteriol* 186, 6706–6713.

Neil D. Rawlings

The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, UK, CB10 1SA. Email: ndr@sanger.ac.uk

Alan J. Barrett

The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, UK, CB10 1SA. Email: ab9@sanger.ac.uk