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.

[2] Families of Serine Peptidases

By NEIL D. RAWLINGS and ALAN J. BARRETT

Introduction

Proteolytic enzymes dependent on a serine residue for catalytic activity are widespread and very numerous. Serine peptidases are found in viruses (Table I), bacteria, and eukaryotes, and they include exopeptidases, endopeptidases, oligopeptidases, and omega peptidases.

By the criteria we use to distinguish families of peptidases (see [1] in this volume), over 20 families of serine peptidases are recognized (Table II). On the basis of three-dimensional structures, and some less direct evidence, most of these families can be grouped together into about six clans that may well have common ancestors. The structures that are known for members of four of the clans, chymotrypsin, subtilisin, carboxypeptidase C, and *Escherichia* D-Ala-D-Ala peptidase A, show them to be totally unrelated, however. We therefore envisage at least four separate evolutionary origins of serine peptidases, and suspect that there were considerably more.

There are similarities in the reaction mechanisms for several of the peptidases with different evolutionary origins. Thus, the peptidases of the chymotrypsin, subtilisin, and carboxypeptidase C clans have in common a "catalytic triad" of the three amino acids: serine (nucleophile), aspartate (electrophile), and histidine (base). The geometric orientations of these are closely similar between families, despite the fact that otherwise the protein folds are quite different. This striking example of convergent evolution has led researchers to look for the same catalytic residues in other serine peptidases. However, the catalytic mechanisms of clans SE* (*Escherichia* D-Ala-D-Ala peptidase A) and SF (repressor LexA) are already known to be very different, and there is little doubt that some of the other

* In reviewing the families of serine peptidases, we shall use the short codes for reference to the clans and families that we introduced previously (N. D. Rawlings and A. J. Barrett, *Biochem. J.* **290,** 205, 1993). Thus, all the codes for groups of serine peptidases start with "S" and are completed by a letter of the alphabet for a clan (giving SA, SB, etc.) or an arabic numeral for a family (giving S1, S2, etc.) (Table II).

Amino acid and nucleotide sequences are specified by database codes; most of these are from the Swiss-Prot Database (release 26), but a code given in parentheses is an EMBL database accession number. For some viral sequences, the code given is that of the viral polyprotein. For some viruses, numerous variants with only minor differences have been described, and only a single example of each has been included here.

TABLE I
FAMILIES OF PEPTIDASES FOUND IN VIRUSES

Type of virus	Peptidase family[a]
Double-stranded DNA	
Enveloped	
Baculovirus	C1
Herpes virus	S8, S21
Pararetrovirus	A2
Nonenveloped	
Adenovirus	C5
Syphovirus	S24, U7
Myovirus	U9
Single-stranded RNA	
Enveloped	
Coronavirus	S32, C16, U24
Flavivirus	S7
Pestivirus	S31
Retrovirus	A2
Togavirus	S3, S29, C9, C18
Nonenveloped	
Comovirus	C3
Nepovirus	C3
Picornavirus	C3, U31
Potyvirus	S30, C4, C6
Tymovirus	C21
Double-stranded RNA	
Unclassified	C7, C8

[a] The numbering of the families of peptidases is that of N. D. Rawlings and A. J. Barrett, *Biochem. J.* **290,** 205 (1993), as slightly revised in [2] and [32] of this volume, and [7] and [13] in Volume 248 of this series. All of the peptidases known to be encoded by viruses are endopeptidases, and it will be noted that there is no metallopeptidase among them.

families of serine peptidases also will prove to have distinctive mechanisms of action, without the "classic" Ser, His, Asp triad.

The arrangements of the catalytic residues in the linear sequences of members of the various families commonly reflect their relationships at the clan level (Table II).

The way in which certain glycine residues tended to be conserved in

TABLE II

CLANS AND FAMILIES OF SERINE PEPTIDASES[a]

Clan	Family	Representative enzyme	Known catalytic residues
SA	S1	Chymotrypsin	----------H----------D----------S----------
"	S2	α-Lytic endopeptidase	----------H-----D----------S----------
"	S3	Sindbis virus core endopeptidase	----------H-------D----------S----------
"	S5	Lysyl endopeptidase	----------H----------D----------S----------
"	S6	IgA-specific serine endopeptidase	----------S----------
"	S30	Tobacco etch virus 35 kDa endopeptidase	----------H-D---S----------
"	S7	Yellow fever virus NS3 endopeptidase	-----HD----S----------
"	S29	Hepatitis C virus NS3 endopeptidase	----------S----------
"	S31	Cattle diarrhea virus p80 endopeptidase	----------S----------
"	S32	Equine arteritis virus putative endopeptidase	----------
SB	S8	Subtilisin	-------D----H----------S----------
SC	S9	Prolyl oligopeptidase	----------S----D-H-
?	S10	Carboxypeptidase C	----------S----------D------H-
?	S28	Lysosomal Pro-X carboxypeptidase	----------
?	S15	*Lactococcus* X-Pro-peptidase	----------S----------
SE	S11	*Escherichia* D-Ala-D-Ala peptidase A	------SK----------
"	S12	*Streptomyces* R61 D-Ala-D-Ala peptidase	----------SK----------
"	S13	*Actinomadura* R39 D-Ala-D-Ala peptidase	-----SK----------
SF	S24	Repressor LexA	----------S---------K----------
"	S26	Leader peptidase	----------S---------K----------
"	S27	Eukaryote signal peptidase	----------
SG	S14	Clp endopeptidase (subunit clpP)	----------S------H----------
?	S16	Endopeptidase La	----------S------
?	S25	Multicatalytic endopeptidase complex	----------
-	S18	Omptin	----------
-	S19	*Coccidioides* endopeptidase	----------
-	S21	Assemblin	----------H----------S----------
-	S17	*Bacteroides* extracellular endopeptidase	----------
-	S23	*Escherichia* protease I	----------

[a] The order in which the families are listed here is that in which they are to be found in the text. The arrangement of catalytic residues in the representative enzymes is illustrated diagrammatically by use of lines of normalized length to depict the polypeptide chains, although in reality the chains vary considerably in length, of course. For some of the families, catalytic residues remain to be identified.

the vicinity of the catalytic serine residue in the first serine peptidases to be sequenced, to form the motif Gly-Xaa-**Ser**-Yaa-Gly[1] led to an early expectation that this motif would be found in all serine peptidases. Though it is true that most of the families do show conserved glycine residues

[1] S. Brenner, *Nature (London)* **334,** 528 (1988).

near the essential serine, the exact positions of these are variable, as is shown in Fig. 1.

Chymotrypsin Clan (SA)

As is shown in Table II, we include 10 families in clan SA, and all the active members of these families are endopeptidases. For the families of chymotrypsin (S1), α-lytic endopeptidase (S2), Sindbis virus core endopeptidase (S3), and *Achromobacter* lysyl endopeptidase (S5), the three-dimensional structures are known to be similar. The families of yellow fever virus NS3 endopeptidase (S7) and tobacco etch virus 35-kDa endo-

Clan	Family	
SA	S1	··G·SG····
	S2	··G·SG····
	S3	··G·SG····
	S5	··G·SG····
	S6	··G·SG····
	S30	··G·SG····
	S7	··G·SG····
	S29	··G·SGG···
	S31	··G·SG····
SB	S8	··G·S·····
SC	S9	··G·S·GG··
	S10	··G·S··G··
	S15	··G·S··G··
SE	S11	····S··K··
	S12	····S··K··
	S13	····S··K··
SF	S24	··G·S····G
	S26	····S·····
SG	S14	G···S·····
	S16	··G·S·G···
-	S21	····S·····

FIG. 1. Glycine residues totally conserved in the vicinity of the catalytic serine and lysine residues in the known members of various families of serine peptidases. The codes for clans and families are as in Table II.

peptidase (S30) have the same order of catalytic residues as chymotrypsin. The *Neisseria* IgA-specific serine endopeptidase (S6), hepatitis C virus NS3 endopeptidase (S29), cattle diarrhea virus p80 endopeptidase (S31), and equine arteritis virus putative endopeptidase (S32) show somewhat similar sequences around the catalytic serine residues, so these families are included with a lower degree of confidence.

The order of catalytic residues in the polypeptide chain in clan SA is His/Asp/Ser, and Fig. 2 shows sequences around these catalytic residues for selected members of the clan.

Chymotrypsin Family (S1)

The members of the chymotrypsin family (Table III) are almost entirely confined to animals. The exceptions are trypsin-like enzymes from actinomycetes of the genera *Streptomyces* (TRYP_STRGR) and *Saccharopolyspora* (TRYP_SACER), as well as one recently sequenced from the fungus *Fusarium oxysporum* (EMBL : S63827). The members of this family are inherently secreted proteins. Each is synthesized with a signal peptide that targets it to the secretory pathway. The animal enzymes may be secreted directly (e.g., proenzymes of coagulation factors in the liver parenchymal cell), or packaged in vesicles for regulated secretion (e.g., chymotrypsinogen in the pancreas), or they may be retained in leukocyte granules (e.g., elastase, chymase, or granzymes in polymorphonuclear leukocytes, mast cells, or cytotoxic lymphocytes, respectively). Proteolytic activation of the proenzymes occurs extracellularly, or sometimes in storage organelles. Functions are occasionally intracellular, as in intracellular digestion of bacteria by neutrophils, but most commonly extracellular. Examples of extracellular functions are digestion of food proteins in the intestine by the enzymes from the pancreas, and coagulation, fibrinolysis, and complement activation by the enzyme systems in the plasma.

There are about 200 complete amino acid sequences known, more than for any other family of peptidases. Most enzymes are monomers, but granzyme A is a disulfide-bonded homodimer (see [4]), and tryptase is a homotetramer (see [6]).

The essential catalytic unit of these peptidases is a polypeptide chain of about 220 amino acid residues. However, many members of the family are mosaic proteins in which the molecule is extended N-terminally by addition of unrelated peptide segments (Fig. 3). The catalytic unit almost invariably forms the C-terminal domain: only acrosin and complement

```
          4        5         6  10        11        19        20
          0123456789012345679 01 23456789012 3 89012345678901
                              *    *           *
Family S1
    1     HFCGGSLINENWVVTAAHCGV   DITLLKLSTAAS  VSSCMGDSGGPLVC
    2     HFCGGSLINDQWVVSAAHCYK   DIMLIKLSSPVK  KDSCQGDSGGPVVC
    3     AICGGFLIRRDFVLTAAHCEG   DIMLLKLKSKAK  RASFRGDSGGPLVC
    4     HTCGGTLIRRNWVMTAAHCVS   DIALLRLAQSVT  RSCCQGDSGGPLHC
    5     LLCGASLISDRWVLTAAHCLL   DIALMKLKKPVA  GDACEGDSGGPFVM
    6     GRGGGALLGDRWILTAAHTLY   DIALLELENSVT  QDACQGDSGGVFAV
    7     PWAGGALINEYWVLTAAHVVE   DIAVVRLKDPVK  MDSCKGDSGGAFAV
    8     HLCGGSLIGHQWVLTAAHCFD   DIALIKLQAPLN  KDACKGDSGGPLVC
    9     HFCGGTLISPEWVLTAAHCLE   DIALLKLSSPAV  TDSCQGDSGGPLVC
   10     YVCGGSLMSPCWVISATHCFI   DIALLKIRSKEG  TDSCQGDSGGPLVC
   11     FLCGGILLSSCWILSAAHCFQ   DIALLQLKSDSS  HDACQGDSGGPLVC
   12     GFCGGTILSEFYILTAAHCLY   DIAVVRLKTPIT  KDACQGDSGGPLVT
   13     HLCGGSIIGNQWILTAAHCFY   DIALLKLETTVN  KDACKGDSGGPLSC
   14     FQCGGIIVHRQWVLTAAHCIS   DLMLLRLTEPAD  KDTCWGDSGGPLVC
   15     YLCGGVLIDPSWVITAAHCYS   DLMLLHLSEPAD  KDTCAGDSGGPLIC
   16     VQCGGALVTNRHVITAAHCVV   DIAILTLANDTVT  KDACQGDSGGPMML
   17     HFCGATLIAENFVMSAAHCVA   DIVIKQLNGSAT  AGVCFGDSGSPLVC
   18     SRCGGFLVRRDFVLTAAHCWG   DIMLLQLSRRVR  KAAFKGDSGGPLLC
   19     HFCGGSLIHPQWVLTAAHCVG   DIALLELEEPVN  RDSCQGDSGGPLVC
   20     HLCGGSLLSGDWVLTAAHCFP   DIALVHLSSPLP  IDACQGDSGGLFVC
   21     HLCGGVLVAEQWVLSAAHCLE   DLLLMQLSEKAT  RDSCKGDSGGPLVC
   22     PWCGGSLLNRNTVLTAAHCVS   DLAILKLSTSIP  KDSCQGDSGGPLIC
   23     MGCGGALIYAQDIVLTAAHCVS  DWAFIKLAQPIN  VDTCQGDSGGPMFR

Family S2
   24     VGFSVTRGATKGFVTAGHCGT   DRAWVSLTSAQT  ACMGRGDSGGSWIT
   25     LGFNVSVNGVAHALTAGHCTS   DYGIIRHSNPAA  VCAQPGDSGGSLFA
   26     LGFNVRSGSTYYFLTAGHCTD   DYGIVRYTNTTI  VCAEPGDSGGPLYS
   27     GSGVIIDADKGYVVTNNHVVD   DIALIQIQNPKN  AAINRGNSGGALVN
   28     FIASGVVVGKDTLLTNKHVVD   DLAIVKFSPNEQ  LSTTGGNSGSLVFN
   29     TSATGVLLGKNTVLTNRHIAK   DLALIRLKPDQN  GFTVPGNSGSGIFN
   30     AAFNVTKGGARYFVTAGHCTN   DYGIVRYTDGSS  ACSAGGDSGGAHFA

Family S3
   31     DVIGHALAMEGKVMKPLIVKG   DMEFAQLPVNMR  GVGGRGDSGRGIMD
   32     KVTGYACLVGDKVMKPAHVKG   DLECAQIPVHMR  GAGKPGDSGRLIFD

Family S5
   33     LVNNTANDRKMYFLTAHHCGM   DFTLLELNNAAN  GVTEPGSSGSGIYS

Family S30
   34     ARVKRFEGSVQLFASVRIMYG   DLRIDNWQQETL  SKLTFGSSGLVIRQ

Family S6
   35                            NYGVLGDSGSPIFA

Family S7
   36     IGAGVYKE--GTFHIMWIVTR   DLISYGGGWKLE  LDFSPGTSGSGIVD

Family S29
   37     TQSFLATCVNGVCWTVYIGAG   DQDIVGWQAPSG  VSYLKGSSGGPILS

Family S31
   38     LETGWAYTHQGGISSVDIVTA   DSGCPDGARCYV  LKNLKGWSGLIIFE

Family S32
   39     TGSVWTRNNIRVVVLTASIVVG   DFAEAVTTQSEL  AWTTSGDSGSAVVQ
```

component C2 are known to have C-terminal extensions beyond the peptidase domain, though coagulation factor X has a 16-residue C-terminal extension that may be removed during activation.[2]

Proteolytic cleavage at the N terminus of the catalytic domain of a proenzyme in the chymotrypsin family forms a new N-terminal amino acid residue with a hydrophobic side chain. The new terminal α-amino group forms a salt bridge with Asp-194, which leads to the assembly of the functional catalytic site.[3] The N-terminal residue is commonly Ile, but may be Leu, Val, or Met, and the salivary plasminogen activator from the vampire bat contains Ser in this position.[4]

The propeptide maintaining the proenzyme in its inactive state can be as small as two amino acid residues, as in cathepsin G[5] and granzyme B (see [5]), but many are much longer. Not uncommonly, these are disulfide bonded to the catalytic domain, so that they remain part of the active enzyme. These N-terminal extensions often contain one or more copies of several domains with different structures (Table IV). Domains such as "kringles" and "apples" are confined to this family of proteins, but others,

[2] K. Fujikawa, K. Titani, and E. W. Davie, *Proc. Natl. Acad. Sci. U.S.A.* **72,** 3359 (1975).
[3] R. Huber and W. Bode, *Acc. Chem. Res.* **11,** 114 (1978).
[4] J. Krätzschmar, B. Haendler, G. Langer, W. Boidol, P. Bringmann, A. Alagon, P. Donner, and W.-D. Schleuning, *Gene* **105,** 229 (1991).
[5] G. Salvesen, D. Farley, J. Shuman, A. Przybyla, C. Reilly, and J. Travis, *Biochemistry* **26,** 2289 (1987).

FIG. 2. Conservation of sequences around the catalytic triad residues in the chymotrypsin clan (SA). Residues are numbered according to bovine chymotrypsinogen. Asterisks indicate the catalytic and presumed catalytic residues. Residues identical to bovine chymotrypsin are shown in white on black. Key to sequences: 1, bovine chymotrypsin; 2, rat trypsin 1; 3, mouse granzyme A; 4, rat pancreatic elastase 1; 5, human thrombin; 6, human complement component $\overline{C1r}$; 7, human complement component $\overline{C1s}$; 8, human plasma kallikrein; 9, human plasmin; 10, human u-plasminogen activator; 11, human t-plasminogen activator; 12, human coagulation factor Xa; 13, human coagulation factor XIa; 14, human tissue kallikrein; 15, rat tonin; 16, limulus proclotting enzyme; 17, human leukocyte elastase; 18, human cathepsin G; 19, human tryptase; 20, human hepsin; 21, human complement component D; 22, *Fusarium* trypsin; 23, *Streptomyces* trypsin; 24, *Lysobacter* α-lytic endopeptidase; 25, streptogrisin A; 26, streptogrisin B; 27, *Escherichia* protease Do; 28, *Staphylococcus* glumatyl endopeptidase 1; 29, *Staphylococcus* epidermolytic factor A; 30, *Streptomyces* glutamyl endopeptidase II; 31, Sindbis virus core protein; 32, Semliki Forest virus core protein; 33, *Achromobacter* lysyl endopeptidase; 34, tobacco etch virus 35-kDa peptidase; 35, *Neisseria* IgA-specific serine endopeptidase; 36, dengue virus NS3 endopeptidase; 37, hepatitis C virus NS3 endopeptidase; 38, bovine diarrhea virus p80 endopeptidase; 39, equine arteritis virus putative proteinase. For sequences 37 and 38 only the catalytic serine residues have been identified with confidence. Catalytic residues are only presumed for sequence 39.

TABLE III
PEPTIDASES OF CHYMOTRYPSIN CLAN (CA)[a]

Peptidase	EC	Database code
Family S1: Chymotrypsin		
Achelase (*Lonomia*)	-	ACH1_LONAC, ACH2_LONAC
Acrosin	3.4.21.10	ACRO_*
Ancrod	-	ANCR_AGKRH
Arginine esterase	3.4.21.35	ESTA_CANFA
Brachyurin	3.4.21.32	COGS_UCAPU
Calcium-dependent serine proteinase	-	CASP_MESAU
Cathepsin G	3.4.21.20	CATG_HUMAN, (M96801)
Cercarial elastase (*Schistosoma*)	-	CERC_SCHMA
Chymase (includes forms I and II)	3.4.21.39	MCP1_*, TRYM_CANFA, MCP2_*, MCP4_MOUSE, MCP5_MOUSE, (M69136)
Chymotrypsin (includes forms A, B, II and 2)	3.4.21.1	CTR2_*, CTRA_BOVIN, CTRB_*, CTR2_CANFA, (U03760)
Chymotrypsin 1 (*Penaeus*)	3.4.21.1	(X66415)
Chymotrypsin-like protease (*Haliotis*)	-	(X71438)
Clotting factor C (*Limulus*)	3.4.21.84	(D90271)
Clotting factor B (*Limulus*)	3.4.21.85	[b]
Clotting enzyme (*Tachypleus*)	3.4.21.86	PCE_TACTR
Coagulation factor VII	3.4.21.21	FA7_*
Coagulation factor IX	3.4.21.22	FA9_*
Coagulation factor X	3.4.21.6	FA10_*
Coagulation factor XI	3.4.21.27	FA11_HUMAN
Coagulation factor XII	3.4.21.38	FA12_HUMAN
Complement component $C\overline{1r}$	3.4.21.41	C1R_HUMAN
Complement component $C\overline{1s}$	3.4.21.42	C1S_HUMAN
Complement component C2	3.4.21.43	CO2_*
Complement factor B	3.4.21.47	CFAB_*
Complement factor D	3.4.21.46	CFAD_*
Complement factor I	3.4.21.45	CFAI_HUMAN
Crotalase	3.4.21.74	[c]
easter gene product (*Drosophila*)	-	EAST_DROME
Enteropeptidase	3.4.21.9	[d]
Epidermal growth factor-binding protein (includes forms 1, 2 and 3)	3.4.21.35	EGBA_MOUSE, EGBB_MOUSE, EGBC_MOUSE
Flavoboxin (habu snake)	-	FLVB_TRIFL
Gilatoxin	-	[e]
Granzyme A	3.4.21.78	GRAA_*, GRAX_MOUSE
Granzyme B	3.4.21.79	GRAB_*
Granzymes C, D, E, F, G and Y	-	GRAC_MOUSE, GRAD_MOUSE, GRAE_MOUSE, GRAF_MOUSE, GRAG_MOUSE, GRAH_HUMAN
Hepsin	-	HEPS_HUMAN
Hypodermin A	-	(L24914)
Hypodermin B	-	(L24915)
Hypodermin C	3.4.21.49	COGS_HYPLI
Leukocyte elastase	3.4.21.37	ELNE_HUMAN
Medullasin	-	ELNE_HUMAN
Myeloblastin	3.4.21.76	MELB_HUMAN, PTN3_HUMAN
Natural killer cell protease 1	-	NKP1_RAT
7S Nerve growth factor (includes α and γ chains)	3.4.21.35	NGFA_MOUSE, NGFG_MOUSE

TABLE III (*continued*)

Peptidase	EC	Database code
Pancreatic elastase I	3.4.21.36	EL1_*, (M27347)
Pancreatic elastase II (includes forms A and B)	3.4.21.71	EL2A_HUMAN, EL2B_HUMAN, EL2_*,
Pancreatic endopeptidase E (includes forms A and B)	3.4.21.70	EL3A_HUMAN, EL3B_HUMAN
Plasma kallikrein	3.4.21.34	KAL_*
Plasmin	3.4.21.7	PLMN_*, (M62832)
t-Plasminogen activator	3.4.21.68	UROT_*
u-Plasminogen activator	3.4.21.73	UROK_*
Salivary plasminogen activator (vampire bat)	3.4.21.68	UROT_DESRO
Protein C	3.4.21.69	PRTC_*
Proteinase RVV-V (Russell's viper) (includes forms α and γ)	-	RVVA_VIPRU, RVVG_VIPRU
Ra-reactive factor component P100	-	(D16492)
γ-Renin	3.4.21.54	RENG_MOUSE
Semenogelase	3.4.21.77	PROS_HUMAN
snake gene product (*Drosophila*)	-	SNAK_DROME
stubble gene product (*Drosophila*)	-	(L11451)
Serine protease (rat)	-	(L05175)
Serine protease (*Haematobia*)	-	(Z22567)
Serine proteases 1 and 2 (*Drosophila*)	-	SER1_DROME
Thrombin	3.4.21.5	THRB_*
Tissue kallikrein	3.4.21.35	KAG1_*, KAG2_*, KAG3_*, KAG5_MOUSE, KAG_PIG, KAGB_MOUSE, KAGP_RAT, KAGR_*
Tonin	3.4.21.35	TONI_RAT
Trypsin (includes forms I, II, III, IV, Va and Vb)	3.4.21.4	TRYP_*, TRY1_*, TRY2_*, TRY3_*, TRY4_RAT, TRYA_RAT, TRYB_RAT, (L04749), (L08428), (L15632), (L16805), (L16807), (L16808), (M77814), (M96372), (S63827), (X56744), (X70074), (X72781), (Z18889), (Z18890)
Trypsin-like enzyme (*Choristoneura*)	-	(L04749)
Tryptase (includes forms 1, 2 and 3)	3.4.21.59	TRYT_CANFA, TRYA_HUMAN, TRYB_HUMAN, (M33493), (M30038), MCP6_MOUSE, MCP7_MOUSE
Venombin A	3.4.21.74	BATX_BOTAT, PTCA_AGKCO
Vitellin-degrading endopeptidase (*Bombyx*)	-	(D16232)
Family S2: α-Lytic endopeptidase		
Epidermolytic toxins A and B (*Staphylococcus*)	-	ETA_STAAU, ETB_STAAU
Glutamyl endopeptidase (*Staphylococcus*)	3.4.21.19	STSP_STAAU
Glutamyl endopeptidase (*Bacillus*)	-	GSEP_BACLI
Glutamyl endopeptidase II (*Streptomyces*)	3.4.21.82	(D12470), [f]
α-Lytic endopeptidase	3.4.21.12	PRLA_*
"Metalloprotease" (*Bacillus subtilis*)	-	[g]
Protease Do	-	HTRA_*, (M24777), (M31119)
Protease I (*Rarobacter*)	-	(D10753)

TABLE III (*continued*)

Peptidase	EC	Database code
Streptogrisin A (*Streptomyces griseus*)	3.4.21.80	PRTA_STRGR
Streptogrisin B (*Streptomyces griseus*)	3.4.21.81	PRTB_STRGR
Family S3: Sindbis virus core endopeptidase		
Core endopeptidase	-	POLS_EEEV, POLS_RRVN, POLS_SFV, POLS_SINDV, POLS_WEEV
Family S5: Lysyl endopeptidase		
Lysyl endopeptidase (*Achromobacter*)	3.4.21.50	API_ACHLY
Family S6: IgA-specific serine endopeptidase		
IgA-specific serine endopeptidase	3.4.21.72	IGA_NEIGO, (X64357)
Family S30: Tobacco etch virus 35 kDa endopeptidase		
35 kDa endopeptidase	-	POLG_PPVD, POLG_PVYN, POLG_TEV, POLG_TVMV
Family S7: Yellow fever virus NS3 endopeptidase		
NS3 endopeptidase	-	POLG_DEN2J, POLG_JAEVJ, POLG_KUNJM, POLG_MVEV, POLG_TBEVS, POLG_WNV, POLG_YEFV1
Family S29: Hepatitis C virus NS3 endopeptidase		
NS3 endopeptidase	-	POLG_HCV1
Family S31: Cattle diarrhea virus p80 endopeptidase		
p80 endopeptidase	-	POLG_BVDVN, POLG_HCVA
Family S32: Equine arteritis virus putative endopeptidase		
Equine arteritis virus putative endopeptidase	-	RPOL_EAV

[a] EC is the enzyme nomenclature number (Nomenclature Committee of the International Union of Biochemistry and Molecular Biology, "Enzyme Nomenclature 1992," Academic Press, Orlando, Florida, 1992, and Supplement); a dash (–) indicates that no EC number has been assigned. Literature references to the individual proteins are generally to be found in the database entries for which the codes are given.

[b] T. Muta, T. Oda, and S. Iwanaga, *J. Biol. Chem.* **268,** 21384 (1993).

[c] H. Pirkle, F. S. Markland, I. Theodor, R. Baumgartner, S. S. Bajwa, and H. Kirakossian, *Biochem. Biophys. Res. Commun.* **99,** 715 (1981).

[d] A. Light and H. Janska, *J. Protein Chem.* **10,** 475 (1991).

[e] P. Utaisincharoen, S. P. Mackessy, R. A. Miller, and A. T. Tu, *J. Biol. Chem.* **268,** 21975 (1993).

[f] I. Svendsen, M. R. Jensen, and K. Breddam, *FEBS Lett.* **292,** 165 (1991).

[g] A. Sloma, C. F. Rudolph, G. A. J. Rufo, B. J. Sullivan, K. A. Theriault, D. Ally, and J. Pero, *J. Bacteriol.* **172,** 1024 (1990).

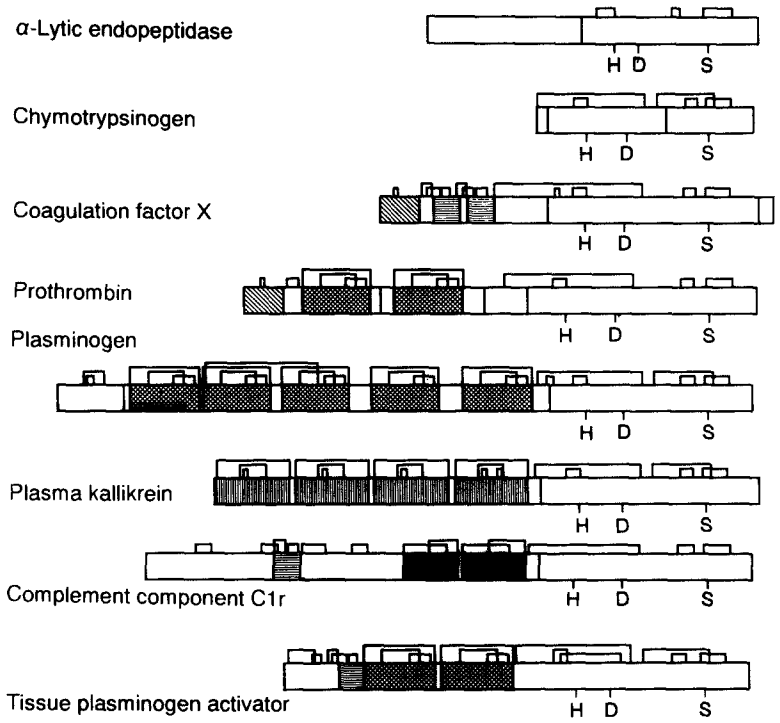

FIG. 3. Polypeptide chain structures of some members of the chymotrypsin family. Vertical lines represent positions of posttranslational cleavages, brackets indicate disulfide bonds, and H, D, and S mark the catalytic residues. Shaded segments correspond to epidermal growth factor-like (horizontal), Ca^{2+}-binding (diagonal), kringle (cross-hatched), apple (vertical), and sushi (black).

such as "sushi" and epidermal growth factor-like domains, are widely distributed among protein families. One domain is found not only in the chymotrypsin family (complement subcomponents C1r and C1s) but also in the astacin family of metallopeptidases.[6] The process by which so much "exon shuffling" has occurred in this family has been suggested to depend heavily on a conserved phase 1 intron N-terminal to the peptidase domain.[7]

A surprising feature of the genes for peptidases of the chymotrypsin family is the diversity of codons used for the active site serine residue. The six available codons for serine fall into two groups (TCA, TCG, TCT, TCC vs. AGC, AGT) such that interconversion between groups requires

[6] N. D. Rawlings and A. J. Barrett, this series, Vol. 248, Chapter 13.
[7] L. Patthy, *Semin. Thromb. Hemostasis* **16,** 245 (1990).

TABLE IV
OCCURRENCE OF ADDITIONAL DOMAINS IN MOSAIC PROTEINS OF CHYMOTRYPSIN FAMILY

Protein	Domain							
	Kringle[a]	Sushi[b]	Apple[c]	Growth factor module[a]	Finger module[a]	Ca^{2+}-binding module[a]	C1r module[d]	Lectin (C-type) module[b]
Plasminogen	5	—	—	—	—	—	—	—
Apolipoprotein	37	—	—	—	—	—	—	—
Hepatocyte growth factor	4	—	—	—	—	—	—	—
Coagulation factor XII	1	1	—	2	1	—	—	—
Prothrombin	2	—	—	—	—	1	—	—
t-Plasminogen activator	2	—	—	1	1	—	—	—
u-Plasminogen activator	1	—	—	1	—	—	—	—
Protein C	—	—	—	2	—	1	—	—
Coagulation factor VII	—	—	—	2	—	1	—	—
Coagulation factor IX	—	—	—	2	—	1	—	—
Coagulation factor X	—	—	—	2	—	1	—	—
Complement factor B	—	3	—	—	—	—	—	—
Complement factor 2	—	3	—	—	—	—	—	—
Complement subcomponent C1r	—	2	—	1	—	—	1	—
Complement subcomponent C1s	—	2	—	1	—	—	1	—
Haptoglobin	—	1	—	—	—	—	—	—
Limulus clotting factor C	—	3	—	1	—	—	—	1
Plasma kallikrein	—	—	4	—	—	—	—	—
Coagulation factor XI	—	—	4	—	—	—	—	—

[a] L. Patthy, *Cell (Cambridge, Mass.)* **41,** 657 (1985).
[b] T. Muta, T. Miyata, Y. Misumi, F. Tokunaga, T. Nakamura, Y. Toh, Y. Ikehara, and S. Iwanaga, *J. Biol. Chem.* **266,** 6554 (1991).
[c] B. A. McMullen, K. Fujikawa, and E. W. Davie, *Biochemistry* **30,** 2050 (1991).
[d] L. Patthy, *Semin. Thromb. Hemostasis* **16,** 245 (1990).

two base changes. Thus, any single point mutation results in loss of the serine, and would doubtless result in loss of catalytic activity in a serine peptidase containing it. In line with expectation, only codons of the TCX group seem to be used for the catalytic Ser in the subtilisin family (S8). However, all six codons are used for active site serine residues in the chymotrypsin family; the most common are from the TCX set, but hepsin, plasmin, thrombin, coagulation factor IX, protein C, and complement components C1r and C1s have AGT or AGC codons. Brenner[1] pointed out that codons in both groups can be derived from cysteine codons in a single step, and suggested that separate evolutionary lines of serine peptidases may have diverged from an ancestral cysteine peptidase. How-

ever, our phylogenetic analysis suggests that the "AGX" members do not represent a monophyletic group, and that the change of codon usage has occurred independently several times. It is possible that the "AGX" members have evolved from "TCX" members via proteins without catalytic activity. Codons from both sets also occur in the catalytic serine in the α-lytic endopeptidase family (S2) of clan SA, as well as in two unrelated lines of serine peptidases, the families of prolyl oligopeptidase (S9) and carboxypeptidase C (S10).

The chymotrypsin family contains several proteins that lack peptidase activity. Among these, haptoglobin (HPT_RAT, HPT1_HUMAN, HPT2_HUMAN) has the active site His replaced by Lys, and the active site Ser by Ala, whereas the essential Asp residue is retained. In protein Z (PRTZ_*) and azurocidin (CAP7_*), both the active site Ser and His residues have been replaced. The complete triad is present in cattle procarboxypeptidase A subunit III (CAC3_BOVIN), and the inactivity of this has been attributed to lack of two N-terminal hydrophobic residues.[8] In rhesus monkey apolipoprotein (a) (APOA_MACMU) the active site Ser is replaced by Asn, but in the human protein (APOA_HUMAN) all three residues of the catalytic triad are present. The expected hydrophobic residues at positions 16 and 17 (chymotrypsinogen numbering) are also present in the human apolipoprotein, so that it has the appearance of a fully functional proenzyme. Conversion to an active peptidase would apparently require cleavage of a Ser+Ile bond, and proteolytic activity of apolipoprotein (a) has been reported.[9]

The fact that members of family S1 occur in both eukaryotes and prokaryotes has been the subject of considerable debate. The actinomycete trypsins have sequences too similar to eukaryote trypsins to be consistent with a divergence 3500 million years ago, which is when the common ancestor of prokaryotes and eukaryotes is generally thought to have lived. Hartley[10] attempted to explain this apparent anomaly in terms of a "horizontal" gene transfer from higher organisms to bacteria, whereas Young *et al.*[11] presented a dendrogram with bacterial sequences at the root, along with an estimate of divergence at about 1300 million years ago, and the implication that because this corresponded to the time that mitochondria were introduced into eukaryotes, a horizontal gene transfer had occurred.

[8] N. Venot, M. Sciaky, A. Puigserver, P. Desnuelle, and G. Laurent, *Eur. J. Biochem.* **157,** 91 (1986).

[9] T. M. Chulkova and V. V. Tertov, *FEBS Lett.* **336,** 327 (1993).

[10] B. S. Hartley, *Proc. R. Soc. London B* **205,** 443 (1979).

[11] C. L. Young, W. C. Barker, C. M. Tomaselli, and M. O. Dayhoff, *in* "Atlas of Protein Sequence and Structure" (M. O. Dayhoff, ed.), Vol. 5, Suppl. 3, p. 73. National Biomedical Research Foundation, Washington, D.C., 1978.

Our calculations confirm this date, and we suggest that the gene transfer is most likely to have been from the protomitochondrion to the eukaryote host.

α-Lytic Endopeptidase Family (S2)

Members of family S2 (listed in Table III) are known only from eubacteria, and they are secreted to act extracellularly. For example, α-lytic endopeptidase produced by the soil bacterium *Lysobacter* degrades cell walls of other soil bacteria, allowing *Lysobacter* to feed on them. Like the chymotrypsin family, family S2 contains endopeptidases that show specificity for P1 residues that are basic, hydrophobic, or alanine, but in addition, the α-lytic endopeptidase family includes several enzymes cleaving glutamyl bonds.

The endopeptidases of family S2 show only slight similarity in sequence to those of family S1, but James and co-workers[12] prepared alignments based on the tertiary structures of the enzymes, and concluded that the small microbial enzymes had common ancestry with the more sophisticated pancreatic endopeptidases.

The recently determined sequences of protease Do from *Escherichia coli* and *Salmonella typhimurium* show protease Do to be a distant homolog not only of α-lytic endopeptidase but also of the glutamyl endopeptidase from *Staphylococcus*. Significant RDF scores are obtained between *E. coli* protease Do and *Lysobacter* α-lytic endopeptidase (7.3) and between *E. coli* and *Salmonella* proteases Do and *Staphylococcus* epidermolytic toxin B (8.3 and 7.4, respectively).

Crystallographic structures have been determined for α-lytic endopeptidase,[13] streptogrisins A and B,[14,15] and *Streptomyces griseus* glutamyl endopeptidase (see [8]). The catalytic triad residues are included in the alignment of Fig. 2, which also shows our prediction for the catalytic triad of protease Do.

The mechanism of activation is not known for many peptidases in this family, but *Lysobacter* α-lytic endopeptidase is synthesized with a long propeptide, and the same apparently applies to *Streptomyces* glutamyl endopeptidase and *Rarobacter* protease I. Not all members of the family

[12] L. T. J. Delbaere, W. L. B. Hutcheon, M. N. G. James, and W. E. Thiessen, *Nature (London)* **257,** 758 (1975).

[13] L. T. J. Delbaere, G. D. Brayer, and M. N. G. James, *Nature (London)* **279,** 165 (1979).

[14] M. N. G. James, A. R. Sielecki, G. D. Brayer, L. T. J. Delbaere, and C.-A. Bauer, *J. Mol. Biol.* **144,** 43 (1980).

[15] M. Fujinaga, R. J. Read, A. Sielecki, W. Ardelt, M. Laskowski, Jr., and M. N. G. James, *Proc. Natl. Acad. Sci. U.S.A.* **79,** 4868 (1982).

retain Asp-194, which is crucial to activity in family S1 as described above. Although Ile-16 is conserved in α-lytic endopeptidase and the streptogrisins, it does not form an ion pair with Asp-194, and remains on the surface of the molecule.[12]

As in family S1, the active site serine residues in family S2 are generally encoded by TCX codons, but the *Bacillus* enzymes and streptogrisin A use the AGX codons.[16,17]

We report here the existence of two homologs of protease Do from *E. coli* (EMBL: M24777) and *Chlamydia trachomatis* (EMBL: 31119). In both cases, the protease Do-like sequences are interrupted by frameshifts, but these may reflect sequencing errors. The *E. coli* partial sequence is on the strand complementary to that containing the gene for malate dehydrogenase,[18] whereas the *Chlamydia* sequence is complementary to a gene that was proposed to be an open reading frame for a 59-kDa immunogenic protein.[19]

Sindbis Virus Core Endopeptidase Family (S3)

Togaviruses are single-stranded RNA viruses (Table I) that are vertebrate pathogens transmitted by arthropods, and examples are Semliki Forest virus and Sindbis virus. The genome encodes two polyproteins, p130 and p270. Polyprotein p270 contains a cysteine endopeptidase (see [32], family C8), whereas p130 contains a serine endopeptidase. Polyprotein p130 also contains structural proteins for the nucleocapsid core and for the glycoprotein spikes that protrude from the lipid bilayer surrounding the core and carry the recognition site for the host cell receptor. The nucleocapsid core protein forms the N terminus of the p130 polyprotein, and is the serine endopeptidase. The cleavage of the polypeptide occurs at a Trp+Ser bond, and once the core protein is released, it retains no detected peptidase activity. Mutagenesis studies have identified a His/Asp/Ser catalytic triad in the core protein, and X-ray crystallography of the Sindbis virus core protein has shown a structure very similar to those of chymotrypsin and α-lytic endopeptidase.[20]

[16] S. Kakudo, N. Kikuchi, K. Kitadokoro, T. Fujiwara, E. Nakamura, H. Okamoto, M. Shin, M. Tamaki, H. Teraoka, H. Tsuzuki, and N. Yoshida, *J. Biol. Chem.* **267,** 23782 (1992).

[17] G. Henderson, P. Krygsman, C. J. Liu, C. C. Davey, and L. T. Malek, *J. Bacteriol.* **169,** 3778 (1987).

[18] R. F. Vogel, K. D. Entian, and D. Mecke, *Arch. Microbiol.* **149,** 36 (1987).

[19] S. Kahane, Y. Weinstein, and I. Sarov, *Gene* **90,** 61 (1990).

[20] H.-K. Choi, L. Tong, W. Minor, P. Dumas, U. Boege, M. G. Rossmann, and G. Wengler, *Nature (London)* **354,** 37 (1991).

The active site serine is encoded by AGX codons in both Sindbis virus[21] and Semliki Forest virus.[22]

Lysyl Endopeptidase Family (S5)

The recently determined tertiary structure of the lysyl endopeptidase from *Achromobacter lyticus* shows a clear similarity to that of trypsin (see [9]), showing that the enzyme is a member of the chymotrypsin clan. The active site residues are shown in Fig. 2. The salt bridge from the amino terminus to Asp-194 that is essential to the structure of the active site in the chymotrypsin family does not exist in lysyl endopeptidase, but a new disulfide bridge, Cys^{6}-Cys^{216}, may serve a similar function (see [9]).

A domain C-terminal to the potential active site residues of lysyl endopeptidase (residues 474–653) is homologous to a segment of *Vibrio* collagenase (family M9), which establishes the mosaic nature of both proteins. Both this domain and a long N-terminal peptide are removed during proteolytic activation of lysyl endopeptidase.[23]

IgA-specific Serine Endopeptidase Family (S6)

Family S6 contains similar enzymes from two gram-negative, pathogenic bacteria, *Neisseria gonorrhoeae* and *Haemophilus influenzae*. Both peptidases cleave the heavy chains of immunoglobulin A at certain prolyl bonds in the hinge region (see [10]). Unrelated metalloendopeptidases from organisms including *Streptococcus sanguis* also exhibit similar specificity.[24]

Inhibition characteristics identify the enzymes of *Neisseria* and *Haemophilus* as serine peptidases. They contain serine residues in sequences similar to that of chymotrypsin (Fig. 2), and mutation of these inactivates the enzymes.[25]

The precursor of the *Neisseria* enzyme is a mosaic protein of more than 1500 amino acid residues. The proprotein contains an N-terminal leader peptide that directs the protein to the periplasmic space and is removed by a leader peptidase. Following this, there is the peptidase domain, and then a C-terminal "helper" domain believed to create a pore

[21] C. M. Rice and J. H. Strauss, *Proc. Natl. Acad. Sci. U.S.A.* **78,** 2062 (1981).
[22] H. Garoff, A.-M. Frischauf, K. Simons, H. Lehrach, and H. Delius, *Proc. Natl. Acad. Sci. U.S.A.* **77,** 6376 (1980).
[23] T. Ohara, K. Makino, H. Shinagawa, A. Nakata, S. Norioka, and F. Sakiyama, *J. Biol. Chem.* **264,** 20625 (1989).
[24] A. G. Plaut, this series, Vol. 248, Chapter 38.
[25] W. W. Bachovchin, A. G. Plaut, G. R. Flentke, M. Lynch, and C. A. Kettner, *J. Biol. Chem.* **265,** 3738 (1990).

in the outer membrane of the bacterium to permit secretion of the active peptidase. The helper domain is cleaved and remains associated with the outer membrane.[26] The cleavage sites are not conserved between the species variants, and there are very few identical amino acids in the susceptible regions. The helper domain is homologous to domains found in other proteins: *E. coli* initiation factor 2 (IF2_ECOLI), an α-helical domain of unusual composition in *E. coli* tolA protein (TOLA_ECOLI), *Enterococcus faecium* P54 protein (P54_ENTFC), and rat plectin (PLEC_RAT).

Tobacco Etch Virus 35-kDa Endopeptidase Family (S30)

The tobacco etch virus is a potyvirus (Table I), the polyprotein of which contains three proteinases, two of which are cysteine proteinases (see [32], families C4 and C6). Most of the polyprotein processing is performed by the NIa cysteine proteinase, but one N-terminal cleavage (of a Tyr↓Ser bond) is made by the 35-kDa serine endopeptidase.[27]

The residues His-214, Asp-223, and Ser-256 are conserved in the family, and have been shown to be essential for activity by site-directed mutagenesis.[28] The His/Asp/Ser order of catalytic residues, and the sequence surrounding the catalytic Ser residues (Fig. 2), are consistent with the inclusion of family S30 in the chymotrypsin clan of serine peptidases. His-214 and Asp-223 are much closer together than the corresponding residues in other peptidases of the clan, however (Table II).

Yellow Fever Virus NS3 Endopeptidase Family (S7)

Flaviviruses, which include dengue virus, yellow fever virus, and encephalitis viruses, are single-stranded RNA viruses (Table I). The RNA encodes a single polyprotein, which is processed by a viral endopeptidase as well as by cellular enzymes. Unlike the Sindbis virus core endopeptidase, the flavivirus endopeptidase is a nonstructural (NS) protein, and occurs internally in the polyprotein, not at the end. The flavivirus endopeptidase cleaves on the C-terminal side of paired basic amino acids, and excises all the nonstructural proteins from the polyprotein. (Excision of structural proteins, which are at the N terminus of the polyprotein, is believed to be mediated cotranslationally by a host cell peptidase associated with the endoplasmic reticulum.) Endopeptidases with a specificity for cleaving C-terminally to paired basic residues are also found in family S8.

[26] T. Klauser, J. Pohlner, and T. F. Meyer, *BioEssays* **15,** 799 (1993).
[27] J. Verchot, E. V. Koonin, and J. C. Carrington, *Virology* **185,** 527 (1991).
[28] J. Verchot, K. L. Herndon, and J. C. Carrington, *Virology* **190,** 298 (1992).

His-53, Asp-77, and Ser-138 have been identified as the members of a catalytic triad in the yellow fever virus NS3 protein by site-directed mutagenesis[29] (Fig. 2). The catalytic residues occur in the N-terminal half of protein NS3, the C-terminal half of which may be a helicase. It is notable that the serine endopeptidases of other enveloped, single-stranded RNA viruses, in families S29, S31, and S32, also occur as N-terminal domains of proteins, the C-terminal parts of which are helicases.

Hepatitis C Virus NS3 Endopeptidase Family (S29)

The hepatitis C virus is a togavirus (Table I). The viral RNA encodes a polyprotein for nonstructural proteins, among which NS3 has been identified as a serine peptidase. Site-specific mutagenesis has implicated Ser-159 in the catalytic mechanism, and other members of a His/Asp/Ser catalytic triad have been proposed[30] (Fig. 2).

Cattle Diarrhea Virus p80 Endopeptidase Family (S31)

Pestiviruses, which include cattle diarrhea virus and hog cholera virus, are closely related to flaviviruses (Table I). These are single-stranded RNA viruses, each of which encodes a large polyprotein. The p80 protein, which is approximately in the middle of the polyprotein, has been identified as the peptidase responsible for processing all nonstructural pestivirus proteins, and site-specific mutagenesis of Ser-311 (Ser-1842 in the polyprotein) prevented processing.[31]

On the basis of sequence similarities to other serine peptidases that we place in clan SA, Bazan and Fletterick[32] correctly predicted the catalytic nature of the p80 protein of pestiviruses, and the active site serine. They also predicted a catalytic triad including His-217 and Asp-254 and proposed that the peptidase domain of the p80 protein would possess a fold similar to that of chymotrypsin.

Equine Arteritis Virus Putative Endopeptidase (S32)

Equine arteritis virus, once thought to be a togavirus, is now regarded as more closely related to coronaviruses. The viral genome encodes several polyproteins, one of which includes a helicase and a polymerase,

[29] T. J. Chambers, R. C. Weir, A. Grakoui, D. W. McCourt, J. F. Bazan, R. J. Fletterick, and C. M. Rice, *Proc. Natl. Acad. Sci. U.S.A.* **87,** 8898 (1990).
[30] R. Bartenschlager, L. Ahlborn-Laake, J. Mous, and H. Jacobsen, *J. Virol.* **67,** 3835 (1993).
[31] M. Wiskerchen and M. S. Collett, *Virology* **184,** 341 (1991).
[32] J. F. Bazan and R. J. Fletterick, *Virology* **171,** 637 (1989).

and is assumed also to contain a serine peptidase with a catalytic triad of His-1103, Asp-1129, and Ser-1184.[33]

Clan SB: Subtilisin Family (S8)

The subtilisin family is the second largest family of serine peptidases so far identified, and is extremely widespread, members having been found in eubacteria, archaebacteria, eukaryotes, and viruses (Tables I and V). The great majority of the enzymes are endopeptidases, but there is also a tripeptidyl-peptidase. Crystallographic structures have been determined for several members of the family, and these show a catalytic triad composed of the same residues as in peptidases of the chymotrypsin clan. However, these occur in a different order in the sequence (Asp/His/Ser), and the three-dimensional structures of the molecules bear no resemblance to that of chymotrypsin, so that it is clear that subtilisin and chymotrypsin are not evolutionarily related.

The proprotein-processing endopeptidases kexin, furin, and related enzymes, so far known from yeasts and animals, form a distinct subfamily, which we shall term the kexin subfamily (see [11]–[13]). These preferentially cleave C-terminally to paired basic amino acids. A member of the kexin subfamily can be identified from the subtly different motifs around the active site residues. As can be seen in Fig. 4, the members of the kexin subfamily have Asp in place of Ser or Thr in position 139, Arg in place of His in position 172, and Ala in place of Met in position 324.

In subtilisins, the oxyanion hole is formed by the active site Ser and Asn-262 (numbering according to preprosubtilisin BPN'). Unusually, in the mammalian furin known as PC2, Asn-262 is replaced by Asp (see [11]), and Asp-119 for Asn.

The alignment of the sequences (made by the PILEUP program[34]) gives clues to the structural basis of the selectivity of the kexin subfamily for basic residues. Thus, acidic residues Glu-210 and Asp-273, which are found only in the kexins, occur in parts of the enzymes predicted to form the S2 and S1 binding pockets, respectively.

Some members of the subtilisin family have a requirement for thiol activation. These include endopeptidases R, T, and K from the yeast *Tritirachium*, and the cuticle-degrading endopeptidase from *Metarhizium*,

[33] J. A. Den Boon, E. J. Snijder, E. D. Chirnside, A. A. F. De Vries, M. C. Horzinek, and W. J. M. Spaan, *J. Virol.* **65**, 2910 (1991).

[34] Genetics Computer Group, *in* "Program Manual for the GCG Package" University of Wisconsin, Madison, 1991.

TABLE V
PEPTIDASES OF SUBTILISIN FAMILY (S8)[a]

Peptidase	EC	Database code
Family S8: Subtilisin		
Alkaline elastase (*Bacillus*)	-	ELYA_*
Alkaline endopeptidase (*Aspergillus*)	-	(M96758), (Z11580)
Alkaline endopeptidase (*Acremonium*)	-	ALP_CEPAC
Alkaline endopeptidase (*Trichoderma*)	-	ALP_TRIHA
Aqualysin I (*Thermus*)	-	AQL1_THEAQ
Bacillopeptidase F (*Bac. subtilis*)	-	SUBF_BACSU
C5a peptidase (*Streptococcus*)	-	SCPA_STRPY
Calcium-dependent extracellular endopeptidase A (*Vibrio*)	-	PROA_VIBAL
Calcium dependent endopeptidase (*Anabaena*)	-	PRCA_ANAVA
Cell-wall associated endopeptidase (*Lactococcus*) (includes forms PI, PII and PIII)	-	P1P_LACLA, P2P_*, P3P_LACLA, (X14130)
Cerevisin	3.4.21.48	PRTB_YEAST
Cuticle-degrading protease (*Metarhizium*)	-	CUDP_METAN
Endopeptidase K	3.4.21.64	PRTK_TRIAL
Endopeptidase R (*Tritirachium*)	-	PRTR_TRIAL
Endopeptidase T (*Tritirachium*)	-	PRTT_TRIAL
Epidermin processing protease (*Staphylococcus*)	-	EPIP_STAEP
Extracellular endopeptidase (*Serratia*)	-	PRTS_SERMA, PRTT_SERMA
Extracellular endopeptidase (*Xanthomonas*)	-	PROA_XANCP
Halolysin	-	HLY_HAL17
Intracellular serine endopeptidase (*Bacillus*)	-	ISP_*
Major intracellular endopeptidase (*Bacillus*)	-	ISP1_BACSU
Neutral endopeptidase (*Bacillus*)	-	NPRE_*
Nisin operon serine protease (*Lactococcus*)	-	(L11061)
Oryzin	3.4.21.63	AEP_YARLI, ORYZ_*
Serine endopeptidase (*Bac. subtilis*)	-	(PIR S11504)
Serotype-specific antigen precursor (*Pasteurella*)	-	SSA1_PASHA
Subtilisin	3.4.21.62	SUBT_*, SUBE_BACSU, SUBV_BACSU, (L24202)
Hypothetical subtilisin (*Saccharomyces*)	-	YCT5_YEAST, (D14063)
Subtilisin-like protease (*Dichelobacter*)	-	(L08175), (Z16080)
Subtilisin-like protease III (*Saccharomyces*)	-	YSP3_YEAST
Thermitase	3.4.21.66	THET_THEVU
Thermostable serine endopeptidase	-	(S50880)
Tripeptidyl-peptidase II	3.4.14.10	TPP2_HUMAN
Kexin	3.4.21.61	KEX2_YEAST, KEX1_KLULA, (L16238)
Furin	-	FURI_*, FUR1_*, FURL_DROME, FURS_DROME
Pituitary convertase (includes PC1, PC2, PC3, PC6, PACE4)	-	NEC1_*, NEC2_*, NEC3_MOUSE, NECA_HYDAT, NECB_HYDAT, PAC4_HUMAN, (D12619)
56 kDa Serine protease (catfish herpes virus)	-	VG47_HSVI1

[a] See Table III for explanation.

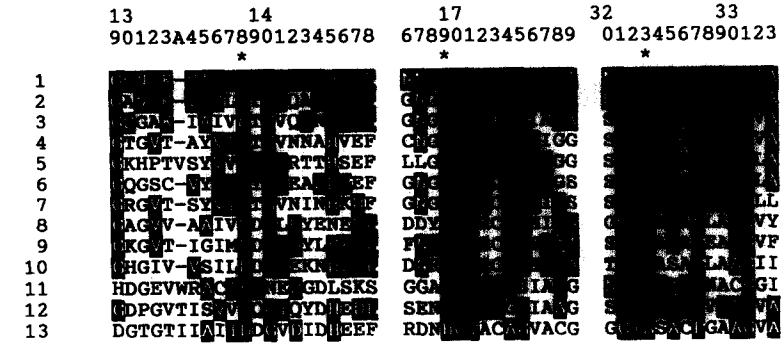

FIG. 4. Conservation of sequence around the catalytic triad residues in the subtilisin family. Residues are numbered according to that of preprosubtilisin BPN′, and those identical to residues in subtilisin BPN′ are shown in white on black. Asterisks indicate the catalytic triad residues. Key to sequences: 1, subtilisin BPN′; 2, subtilisin Carlsberg; 3, thermitase; 4, *Vibrio alginolyticus* calcium-dependent serine "exopeptidase" A; 5, *Yarrowia lipolytica* extracellular endopeptidase; 6, endopeptidase K; 7, cerevisin; 8, kexin; 9, human pituitary convertase PC2; 10, rat furin; 11, human tripeptidyl-peptidase II; 12, halolysin; 13, *Anabaena variabilis* calcium-dependent endopeptidase.

as well as the members of the kexin subfamily. This thiol dependence is attributable to Cys-173 near the active site histidine.[35]

It has been thought that none of the bacterial subtilisins contains cysteine residues,[35] but the subtilisin from *Bacillus smithii* is an exception to this (EMBL: L24202).

The sole viral member of family S8 is a 56-kDa proteinase from the herpes virus 1 that infects the channel catfish.[36] Herpes viruses are related to baculoviruses, both of which groups encode proteins homologous to mammalian endopeptidases (Table I). The catfish herpes virus is rather divergent, and its genome has a large insert containing some 65 extra open reading frames as compared to other α-herpes viruses. The gene encoding the putative peptidase is one of these, hence no homologs are found in other herpes viruses. This gene is presumably one "captured" from a host.

The subtilisin family contains some mosaic proteins. The secreted endopeptidases from *Vibrio alginolyticus* (known confusingly as "exopeptidase A"[37]) and *Xanthomonas campestris* contain homologous C-terminal domains similar to some found in metallopeptidases, including an aminopeptidase from *Vibrio proteolyticus* (family M28) and endopeptidases of the thermolysin family (M4). Homologs of subtilisin that are located in

[35] K.-D. Jany, G. Lederer, and B. Mayer, *FEBS Lett.* **199,** 139 (1986).
[36] A. J. Davison, *Virology* **186,** 9 (1992).
[37] S. M. Deane, F. T. Robb, S. M. Robb, and D. R. Woods, *Gene* **76,** 281 (1989).

the cell membrane of *Lactococcus* share a homologous domain with other bacterial membrane proteins such as the *Streptococcus* fibrinogen- and-immunoglobulin-binding protein (MRP4_STRPY).

Family S8 contains several enzymes with N- or C-terminal extensions that show no relationship to other proteins. The kexin subfamily includes endopeptidases with a variety of C-terminal extensions (see [13]). Tripeptidyl-peptidase II is a 135-kDa mammalian cytosolic enzyme active at neutral pH, which releases N-terminal tripeptides from polypeptides. The peptidase domain forms the N-terminal third of the protein, whereas the C-terminal two-thirds show no resemblance to any other protein.[38]

Carboxypeptidase Clan (SC)

Tertiary structures available for enzymes of family S10 (carboxypeptidase C) are of the "α/β-hydrolase fold" type, also seen for a wide variety of other enzymes including acetylcholinestase, lipases and *Xanthomonas* haloalkane dehalogenase. The sequence of *Neisseria* prolyl aminopeptidase (family S33) shows it to be homologous to the haloalkane dehalogenase, so we can conclude that families S10 and S33 have similar tertiary folds and linear arrangement of catalytic residues (Ser/Asp/His), and form a clan (SC).

The distinctive Ser/Asp/His arrangement of catalytic residues is also found in family S9 (prolyl oligopeptidase) (Table II), which raises the possibility that family S9 is also a member of clan SC.[39] The clan might additionally contain *Lactococcus* X-Pro-peptidase (family S15) and lysosomal Pro-X carboxypeptidase (family S28).[40] In support of this, there is limited conservation of sequence around the active site Ser residues (Fig. 5), but this is not strong enough to be conclusive.

As we shall see, the peptidases of family S9 differ biologically from those of families S10 and S28 in that the homologs of prolyl oligopeptidase are either cytosolic or integral membrane proteins, whereas the carboxypeptidases are secreted or lysosomal enzymes. Also, peptidases of families S9 and S15 do not seem to have proenzymes, whereas those of S10 and S28 do.

Prolyl Oligopeptidase Family (S9)

Family S9 contains peptidases with a very varied range of quite restricted specificities (see [14]–[17]).[41] Members are known from prokary-

[38] B. Tomkinson and A.-K. Jonsson, *Biochemistry* **30,** 168 (1991).
[39] L. Polgár, *FEBS Lett.* **311,** 281 (1992).
[40] F. Tan, P. W. Morris, R. A. Skidgel, and E. G. Erdös, *J. Biol. Chem.* **268,** 16631 (1993).
[41] N. D. Rawlings, L. Polgár, and A. J. Barrett, *Biochem. J.* **279,** 907 (1991).

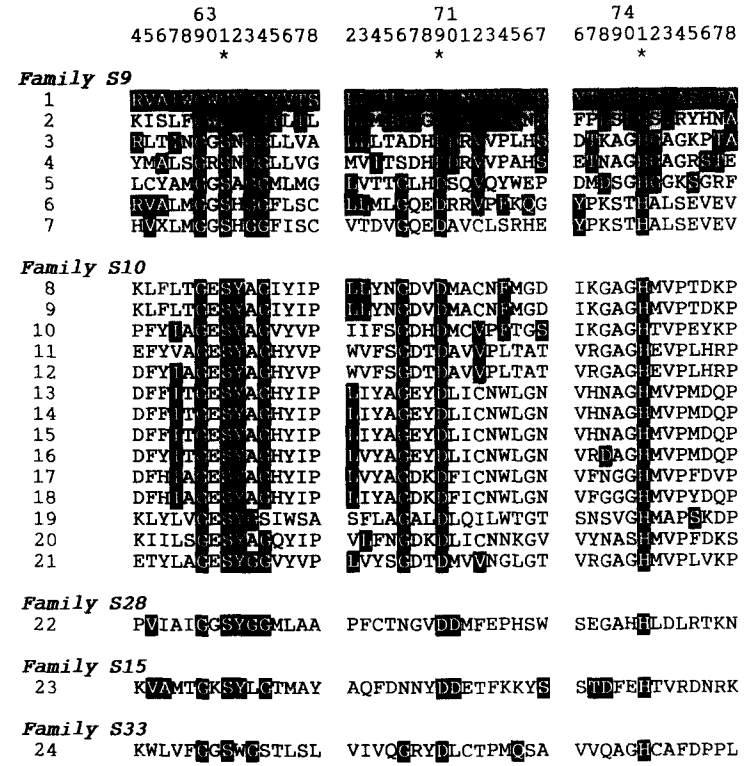

FIG. 5. Conservation of sequence around the catalytic triad residues in the families of prolyl oligopeptidase (S9), *Lactococcus* X-Pro-peptidase (S15), carboxypeptidase C (S10), lysosomal Pro-X carboxypeptidase (S28) and *Neisseria* prolyl aminopeptidase (S33) forming clan SC. Residues are numbered according to rat dipeptidyl-peptidase IV, and residues identical to those in that sequence are shown in white on black. Asterisks indicate the catalytic and presumed catalytic residues; those that have been identified with confidence are Ser, Asp, and His in families S9 and S10. Key to sequences: 1, rat dipeptidyl-peptidase IV; 2, yeast dipeptidyl-peptidase B; 3, pig prolyl oligopeptidase; 4, *Flavobacterium meningosepticum* prolyl oligopeptidase; 5, *Escherichia coli* oligopeptidase B; 6, pig acylaminoacyl-peptidase; 7, human DNF1552 protein (ACPH_HUMAN); 8, human lysosomal carboxypeptidase A; 9, mouse lysosomal carboxypeptidase A; 10, barley carboxypeptidase I; 11, barley carboxypeptidase II; 12, wheat carboxypeptidase II; 13, barley carboxypeptidase III; 14, wheat carboxypeptidase III; 15, rice carboxypeptidase III; 16, *Arabidopsis thaliana* serine carboxypeptidase; 17, yeast carboxypeptidase Y; 18, *Candida albicans* carboxypeptidase Y; 19, *Schizosaccharomyces pombe* carboxypeptidase Y; 20, yeast carboxypeptidase D; 21, *Naegleria fowleri* virulence-related protein; 22, human lysosomal Pro-X carboxypeptidase; 23, *Lactococcus lactis* X-Pro dipeptidyl-peptidase; 24, *Neisseria* prolyl aminopeptidase.

otic and eukaryotic organisms (Table VI). The family contains soluble as well as membrane-bound peptidases. The cytosolic enzymes include two oligopeptidases with different P1 specificities: prolyl oligopeptidase, which cleaves prolyl and some alanyl bonds (see [14]), and oligopeptidase B from eubacteria, which cleaves arginyl and lysyl bonds (see [15]).

Acylaminoacyl-peptidase (see [17]) is also a cytosolic enzyme that has been sequenced from mammals. The enzyme releases an *N*-acetyl or *N*-formyl amino acid from the N terminus of a peptide, and is thus an omega peptidase.

The active site Ser, Asp, and His residues have been identified in

TABLE VI
PEPTIDASES OF FAMILIES OF PROLYL OLIGOPEPTIDASE (S9), CARBOXYPEPTIDASE C (S10), LYSOSOMAL PRO-X CARBOXYPEPTIDASE (S28), AND *Lactococcus* X-PRO DIPEPTIDYL-PEPTIDASE (S15)[a]

Peptidase	EC	Database code
Family S9: Prolyl oligopeptidase		
Acylaminoacyl-peptidase	3.4.19.1	ACPH_*
Dipeptidyl-peptidase IV	3.4.14.5	DPP4_*
Dipeptidyl-peptidase IV-like protein	-	(M76429), (M96860)
Dipeptidyl aminopeptidase A (*Saccharomyces*)	-	(L21944)
Dipeptidyl aminopeptidase B (*Saccharomyces*)	-	DAP2_YEAST
Prolyl oligopeptidase	3.4.21.26	PPCE_*, PPCF_FLAME, (D14005), (M61966)
Protease II (*Escherichia coli*)	-	PTRB_ECOLI
Family S10: Carboxypeptidase C		
Carboxypeptidase C (forms I and III)	3.4.16.5	CBP1_HORVU, CBP3_*, (D10985), (D17586)
Carboxypeptidase D	3.4.16.6	KEX1_YEAST, CBP2_*
Carboxypeptidase Y (*Saccharomyces*)	3.4.16.1	CBPY_*, (D10199)
Carboxypeptidase Y-like protein (*Arabidopsis*)	-	CBPX_ARATH
Carboxypeptidase Y-like protein (rice)	-	(D17587)
Lysosomal carboxypeptidase A	3.4.16.5	PRTP_*
Serine-type carboxypeptidase (*Caenorhabditis*)	-	(M75784)
Serine-type carboxypeptidase (*Aedes*)	-	(M79452)
Virulence-related protein (*Naegleria*)	-	(M88397)
Family S28: Lysosomal Pro-X carboxypeptidase		
Lysosomal Pro-X carboxypeptidase	3.4.16.2	(L13977)
Family S15: *Lactococcus* X-Pro dipeptidyl peptidase		
X-Pro Dipeptidyl peptidase (*Lactococcus*)	3.4.14.5	DPP_*, (Z14230)

[a] See Table III for explanation.

family S9,[42] and the conservation of amino acids around them is shown in Fig. 5. In all known members of the family these residues are within about 130 residues of the C terminus, and the N-terminal parts of the molecules are more or less variable. The membrane-bound members of this family contain membrane-spanning domains near the N terminus.

Dipeptidyl-peptidase IV releases N-terminal dipeptides, preferentially cleaving prolyl bonds (see [16]). It is an integral membrane glycoprotein of lymphocytes and intestinal brush border, in particular, that exists as a homodimer or homotetramer. An enzyme with similar specificity from the membrane of the yeast vacuole, dipeptidyl peptidase B, is involved in processing the α-mating factor precursor.[43]

A notable feature of all members of this family is the lack of proteolytic processing, and the enzymes are apparently synthesized in active form.[44,45] Presumably, their restricted specificities prevent indiscriminate degradation of other proteins.

Like a number of other cytosolic oligopeptidases, including thimet oligopeptidase (family M3) and insulysin (M16), prolyl oligopeptidase and acylaminoacyl-peptidase are thiol dependent, but dipeptidyl-peptidase IV is not. This mix of thiol-dependent and thiol-independent enzymes is also seen in families M3 and M16. Unusually for a serine peptidase, the active site serine of prolyl oligopeptidase reacts with diazomethanes,[44] but that of oligopeptidase B reportedly does not do so (see [15]).

Two mammalian proteins, DPPX-S and DPPX-L, closely related to dipeptidyl-peptidase IV, have the catalytic Ser replaced by Asp. mRNAs for both sequences are abundant in the brain, and that for DPPX-S is also found in kidney, ovary, prostate, and testis.[46,47] To date, the corresponding proteins have not been isolated, but we should expect them to be inactive members of the family.

Carboxypeptidase C Family (S10)

The carboxypeptidase C family includes carboxypeptidases that are unusual among serine-dependent enzymes in that they are maximally active at acidic pH. The members of the family are shown in Table VI. The

[42] F. David, A.-M. Bernard, M. Pierres, and D. Marguet, *J. Biol. Chem.* **268,** 17247 (1993).
[43] C. J. Roberts, G. Pohlig, J. H. Rothman, and T. H. Stevens, *J. Cell Biol.* **108,** 1363 (1989).
[44] D. Rennex, B. A. Hemmings, J. Hofsteenge, and S. R. Stone, *Biochemistry* **30,** 2195 (1991).
[45] A. Kanatani, T. Masuda, T. Shimoda, F. Misoka, X. S. Lin, T. Yoshimoto, and D. Tsuru, *J. Biochem. (Tokyo)* **110,** 315 (1991).
[46] K. Wada, N. Yokotani, C. Hunter, K. Doi, R. J. Wenthold, and S. Shimasaki, *Proc. Natl. Acad. Sci. U.S.A.* **89,** 197 (1992).
[47] N. Yokotani, K. Doi, R. J. Wenthold, and K. Wada, *Hum. Mol. Genet.* **2,** 1037 (1993).

family appears to be restricted to eukaryotes, and sequences are known for enzymes of protozoa, fungi, plants, and animals. These carboxypeptidases have a lysosomal distribution in animals, and in plants and fungi are found in vacuoles. The fungal and plant enzymes can be divided into variants of carboxypeptidase C (EC 3.4.16.5), with a preference for a hydrophobic amino acid in P1′, and those of carboxypeptidase D (EC 3.4.16.6), which release a C-terminal arginine or lysine. However, this may represent an oversimplification of the specificities (see [18]).

The mammalian enzyme is known as "protective protein" because it complexes β-galactosidase and neuroaminidase in the lysosome, protecting them from proteolytic degradation. Its absence therefore leads to functional deficiency in these enzymes, and the lysosomal storage disease, galactosialidosis. The protein expresses the acid carboxypeptidase characteristics of other carboxypeptidase C forms, and there is little doubt that it is the enzyme originally known as cathepsin A, and subsequently as lysosomal carboxypeptidase A.[48] Like other carboxypeptidases C, it also exhibits deamidase and esterase activity.[49,50]

The sequences around the catalytic Ser/Asp/His residues are shown in Fig. 5. In all the peptidases, the residue preceding the catalytic Ser is a Glu residue that is believed to be responsible for the acidic pH optimum of the enzymes (see [18]). The differences in substrate specificity between carboxypeptidases C and D have been attributed to two residues that are Glu in carboxypeptidase D but are both hydrophobic in character (Phe, Leu, or Met) in carboxypeptidase C (see [18]).

Yeast carboxypeptidase Y (one of the C-type enzymes) is inhibited by *p*-hydroxymercuribenzoate, presumably because of the presence of the free thiol of Cys-341 in the S1 binding pocket.[51]

Consistent with the lysosomal/vacuolar location of the serine carboxypeptidases, they are synthesized with N-terminal signal peptides and propeptides.

The crystal structures of several members of the family have been determined (see [18]). The fold of these proteins bears no relation to those of chymotrypsin, subtilisin, or D-Ala-D-Ala peptidases. However, the tertiary structures of serine carboxypeptidases do strongly suggest a distant relationship to a variety of nonpeptidase enzymes. The sequence relation-

[48] N. J. Galjart, H. Morreau, R. Willemsen, N. Gillemans, E. J. Bonten, and A. D'azzo, *J. Biol. Chem.* **266,** 14754 (1991).

[49] H. L. Jackman, F. Tan, H. Tamei, C. Beurling-Harbury, X.-Y. Li, R. A. Skidgel, and E. G. Erdös, *J. Biol. Chem.* **265,** 11265 (1990).

[50] K. Breddam, *Carlsberg Res. Commun.* **51,** 83 (1986).

[51] J. R. Winther and K. Breddam, *Carlsberg Res. Commun.* **52,** 263 (1987).

ship between human acetylcholinesterase and human protective protein, although distant, is statistically significant (RDF score = 6.75 SD units). In turn, cholinesterases are related to a wide range of enzymes and other proteins, including lipases and thyroglobulin, and form a distinct subfamily.[52,53] The differences in sequence between the two subfamilies are too great for us to be able to estimate when the divergence occurred.

The relationship of the cholinesterases to the serine carboxypeptidases invites reconsideration of the long-standing controversy over possible endopeptidase activity of the cholinesterases. At the time of writing, however, the evidence for endopeptidase activity expressed by the cholinesterases does not appear strong.[54–56]

The nonpeptidase enzymes that show no amino acid sequence similarity to carboxypeptidase C, but are thought to be distant relatives because of their similar tertiary structures, include cholinesterases, haloalkane dehalogenases, and carboxymethylenebutenolidases (dienelactone hydrolases) in the "α/β hydrolase fold" clan.[52,53]

Lysosomal Pro-X Carboxypeptidase Family (S28)

This family has only one known member, a carboxypeptidase specific for cleavage of prolyl bonds. Like members of the carboxypeptidase C family (S10), Pro-X carboxypeptidase has an acidic pH optimum, is lysosomal, and is synthesized with a signal peptide and propeptide.[40] The suggestion has been made that lysosomal Pro-X carboxypeptidase is evolutionarily related to both carboxypeptidase C and prolyl oligopeptidase.[40] Active site residues have not been biochemically identified for lysosomal Pro-X carboxypeptidase, but have been postulated to be Ser-134, Asp-333, and His-411 on the basis of similarities to the catalytic residues of carboxypeptidase C (Fig. 5).[40] The glutamate residue preceding the active site serine in carboxypeptidase C, and believed to be responsible for the acidic pH optimum of carboxypeptidase activity (see [18]), is replaced by Gly in lysosomal Pro-X carboxypeptidase, however (Fig. 5).

[52] D. L. Ollis, E. Cheah, M. Cygler, B. Dijkstra, F. Frolow, S. M. Franken, M. Harel, S. J. Remington, I. Silman, J. Schrag, J. L. Sussman, K. H. G. Verscheuren, and A. Goldman, *Protein Eng.* **5,** 197 (1992).

[53] M. Cygler, J. D. Schrag, J. L. Sussman, M. Harel, I. Silman, M. K. Gentry, and B. P. Doctor, *Protein Sci.* **2,** 366 (1993).

[54] M. De Serres, D. Sherman, W. Chestnut, B. M. Merrill, O. H. Viveros, and E. J. Diliberto, Jr., *Cell. Mol. Neurobiol.* **13,** 279 (1993).

[55] S. Michaelson and D. H. Small, *Brain Res.* **611,** 75 (1993).

[56] R. V. Rao and A. S. Balasubramanian, *J. Protein Chem.* **12,** 103 (1993).

Lactococcus X-Pro-Peptidase Family (S15)

Lactococcus lactis contains an enzyme known as "X-prolyl dipeptidyl aminopeptidase" with specificity similar to that of the mammalian dipeptidyl-peptidase IV, cleaving Xaa-Pro+peptide bonds to release N-terminal dipeptides. An indication of relationship of the *Lactococcus* enzyme to others in families S9, S10, or S28 is similarity of sequence in the vicinity of the catalytic Ser residue (Fig. 5).[57] Like members of family S9, *Lactococcus* X-Pro-peptidase does not have a proenzyme,[58] and like dipeptidyl-peptidase IV specifically, the *Lactococcus* enzyme exists as a homodimer.[59] Proposals regarding the catalytic Asp and His residues are made in [15].

Neisseria Prolyl Aminopeptidase Family (S33)

Prolyl aminopeptidase is a 35 kDa peptidase from *Neisseria gonorrhoea* that selectively hydrolyses N-terminal Pro residues.[60] The amino acid sequence shows no relationship to that known for any other peptidase, but is homologous to those of *Pseudomonas* 2-hydroxymuconic semialdehyde hydrolase (XYLF_PSEPU), *Xanthobacter* haloalkane dehalogenase (HALO_XANAU), and human and rat epoxide hydrolase (HYEP_*). These enzymes, in turn, are known to be structurally related to acetylcholinesterase (ACES_*) and carboxypeptidase C (family S8). Most of the "α/β hydrolase fold" enzymes have Ser at the active site, and prolyl aminopeptidase also has Ser at this position (Fig. 5).

Serine-Type D-Ala-D-Ala Peptidases, Families S11, S12, and S13, Forming Clan SE

Both gram-positive and gram-negative bacterial cell walls are complex polymers of amino sugars and amino acids. Chains of alternating *N*-acetylglucosamine and *N*-acetylmuramic acid units are linked to one another by short peptides. In *E. coli* the structure of the link peptide is L-alanyl-D-isoglutamyl-L-*meso*-diaminopimelyl-D-alanine, but the nature of the third amino acid in the chain varies with the bacterial species. These chains are cross-linked, usually between the carboxyl group of D-alanine and the

[57] J.-F. Chich, M.-P. Chapot-Chartier, B. Ribadeau-Dumas, and J.-C. Gripon, *FEBS Lett.* **314,** 139 (1992).

[58] B. Mayo, J. Kok, K. Venema, W. Bockelmann, M. Teuber, H. Reinke, and G. Venema, *Appl. Environ. Microbiol.* **57,** 38 (1991).

[59] M. Nardi, M.-C. Chopin, A. Chopin, M.-M. Cals, and J.-C. Gripon, *Appl. Environ. Microbiol.* **57,** 45 (1991).

[60] N. H. Albertson and M. Koomey, *Mol. Microbiol.* **9,** 1203 (1993).

TABLE VII
PEPTIDASES OF FAMILIES OF D-Ala-D-Ala PEPTIDASES (CLAN SE)[a]

Peptidase	EC	Database code
Family S11: *Escherichia* D-Ala-D-Ala peptidase A		
Escherichia D-Ala-D-Ala peptidase A	3.4.16.4	DACA_ECOLI
D-Ala-D-Ala peptidase	3.4.16.4	DACA_*, DACC_ECOLI, (M37688), (M85047), (X68587)
Sporulation-specific penicillin-binding protein (*Bacillus*)	-	(M84227)
Family S12: *Streptomyces* R61 D-Ala-D-Ala peptidase		
Streptomyces R61 D-Ala-D-Ala peptidase	3.4.16.4	DAC_STRSP
D-Aminopeptidase (*Ochrobactrum*)	-	(M84523)
Family S13: *Actinomadura* R39 D-Ala-D-Ala peptidase		
Actinomadura R39 D-Ala-D-Ala peptidase	3.4.16.4	(X64790)
Penicillin-binding protein 4	3.4.16.4	PBP4_ECOLI

[a] See Table III for explanation.

free amino group of diaminopimelate. In the biosynthesis of the cell wall peptidoglycan the precursor has the four-residue structure above, but with an additional C-terminal D-alanine residue. The D-Ala-D-Ala transpeptidases and carboxypeptidases are involved in the metabolism of the cell wall components.[61] Some of these peptidases are serine enzymes, whereas others have zinc at the catalytic center.[62]

D-Ala-D-Ala peptidases are synthesized with leader peptides to target them to the cell membrane. The peptidases are retained in the membrane by a C-terminal membrane anchor, after removal of the leader peptide, except in *Streptomyces* K15 transpeptidase (family S11) and *E. coli* penicillin-binding protein 4 (family S13), in which a C-terminal propeptide is removed and where overexpression can lead to secretion of the enzymes into the medium.[62]

The members of the three families of serine-type D-Ala-D-Ala peptidases are listed in Table VII. These enzymes are known as low molecular mass penicillin-binding proteins, and probably are distantly related to the high molecular mass penicillin-binding proteins (at the clan level).

[61] J. T. Park, *in* "*Escherichia coli* and *Salmonella typhimurium*. Cellular and Molecular Biology" (F. C. Neidhart, ed.), Vol. 1, p. 663. American Society for Microbiology, Washington, D.C., 1987.

[62] J.-M. Ghuysen, *Annu. Rev. Microbiol.* **45**, 37 (1991).

The antibiotic action of penicillin is due to its binding to the high molecular mass penicillin-binding proteins and not the D-Ala-D-Ala peptidases (see [19]). Enzymes that are capable of degrading pencillins and related antibiotics are β-lactamases, among which the class C β-lactamases are homologous to D-Ala-D-Ala peptidases (S12), and the class A β-lactamases are more distantly related, as is revealed by the three-dimensional structures.[63] Some of the D-Ala-D-Ala peptidases perform a transpeptidation reaction in which the peptidoglycan monomer minus the C-terminal D-Ala is transferred to an exogenous acceptor; the *Streptomyces* K15 peptidase (EMBL : X59965) performs this reaction exclusively, and does not hydrolyse peptide bonds.[62]

Three-dimensional structures are known for *Streptomyces* R61 D-Ala-D-Ala peptidase[64] and *Citrobacter* class C β-lactamase[63] (both from family S12), and class A β-lactamases from *Streptomyces*[65] and *E. coli*[66] (not members of the peptidase families, but members of the clan). The catalytic mechanism elucidated for the class A β-lactamase of *E. coli* involves Ser-70, Lys-73, Ser-130, and Glu-166, Ser-70 acting as the nucleophile, and Glu-166 and Lys-73 being general bases (Fig. 6). The closely spaced Ser-70 and Lys-73 are completely conserved throughout the clan.

Family S11 contains only D-Ala-D-Ala peptidases and the strict transpeptidase of *Streptomyces* K15. Some of the enzymes in this family are partially inhibited by thiol-blocking reagents such as *p*-chloromercuribenzoate, and in the *Streptomyces* K15 peptidase there are two cysteine residues, one of which is close to the general base.[67]

Family S12 contains enzymes with the most diverse specificities. In addition to the *Streptomyces* R61 D-Ala-D-Ala peptidase there are class C β-lactamases (AMPC_*), a D-aminopeptidase from *Ochrobactrum*, a lipolytic esterase from *Pseudomonas* (EMBL : M68491), and proteins from *Bacteroides nodosus* (FMDH_BACNO, FMDD_BACNO). The *Ochrobactrum* enzyme is one of the rare aminopeptidases that are not metalloenzymes, and the only aminopeptidase known to be specific for D-amino acids.[68] The proteins from *B. nododus* may be involved in the assembly

[63] C. Oefner, A. D'Arcy, J. J. Daly, K. Gubernator, R. L. Charnas, I. Heinze, C. Hubschwerlen, and F. K. Winkler, *Nature (London)* **343,** 284 (1990).

[64] J. A. Kelly, J. R. Knox, P. C. Moews, G. J. Hite, J. B. Bartolone, H. Zhao, B. Joris, J.-M. Frère, and J.-M. Ghuysen, *J. Biol. Chem.* **260,** 6449 (1985).

[65] J. Lamotte-Brasseur, F. Jacob-Dubuisson, G. Dive, J.-M. Frère, and J.-M. Ghuysen, *Biochem. J.* **282,** 189 (1992).

[66] N. C. J. Strynadka, H. Adachi, S. E. Jensen, K. Johns, A. Sielecki, C. Betzel, K. Sutoh, and M. N. G. James, *Nature (London)* **359,** 700 (1992).

[67] P. Palomeque-Messia, S. Englebert, M. Leyh-Bouille, M. Nguyen-Distèche, C. Duez, S. Houba, O. Dideberg, J. Van Beeumen, and J.-M. Ghuysen, *Biochem. J.* **279,** 223 (1991).

[68] Y. Asano, Y. Kato, A. Yamada, and K. Kondo, *Biochemistry* **31,** 2316 (1992).

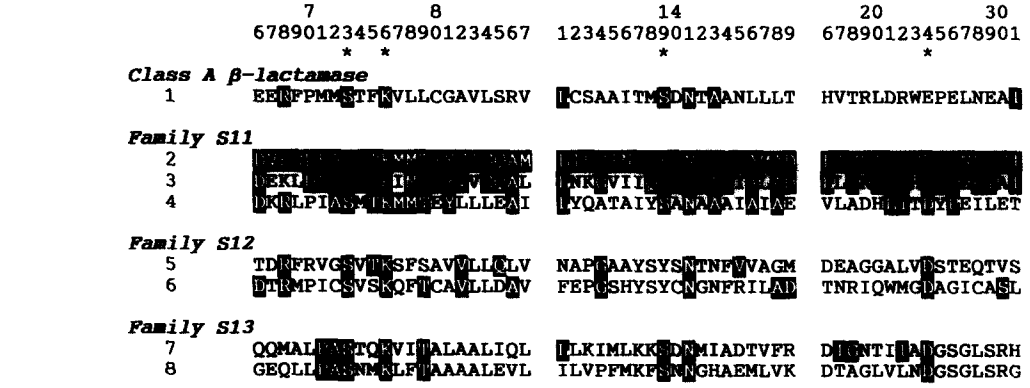

FIG. 6. Comparison of sequences in the vicinity of the catalytic residues of D-Ala-D-Ala carboxypeptidases from families S11, S12, and S13 (clan SE). Residues are numbered according to *Escherichia coli* D-Ala-D-Ala carboxypeptidase A. Residues identical to *E. coli* D-Ala-D-Ala carboxypeptidase A are shown in white on black. Key to sequences: 1, *Escherichia coli* class A β-lactamase; 2, *E. coli* D-Ala-D-Ala carboxypeptidase A; 3, *E. coli* D-Ala-D-Ala carboxypeptidase C; 4, *Bacillus subtilis* D-Ala-D-Ala carboxypeptidase A; 5, *Streptomyces* D-Ala-D-Ala carboxypeptidase; 6, *Ochrobactrum anthropi* D-aminopeptidase; 7, *E. coli* penicillin-binding protein 4; 8, *Actinomadura* R39 D-Ala-D-Ala carboxypeptidase. Asterisks indicate catalytic residues identified in *E. coli* class A β-lactamase.

of fimbriae, which are external appendages found on the surface of gram-negative bacteria and used for cell adhesion. The fimbrial subunits that are related to D-Ala-D-Ala peptidases are found only in some strains, and seem to represent a recent lateral transfer of genes. The function of these proteins is unknown, but the active site motif **Ser**-Xaa-Xaa-Lys is retained.

Family S13 contains D-Ala-D-Ala peptidases from *Actinomadura* and *E. coli*. The *E. coli* enzyme is also known as penicillin-binding protein 4, and has been reported to have endopeptidase as well as carboxypeptidase activity.[69] Although no crystal structures are available for this family, conservation of sequences around the catalytic residues suggests similar tertiary structures and very distant evolutionary relationship to the other members of the proposed clan.

Possible Ser/Lys Dyad Clan (SF) Comprising Families S24, S26, and S27

Although the majority of serine peptidases contain catalytic triads of serine, histidine, and aspartic acid, the D-Ala-D-Ala peptidases show that this is not the only way in which a serine-dependent catalytic site can be

[69] B. Korat, H. Mottl, and W. Keck, *Mol. Microbiol.* **5**, 675 (1991).

formed (Table II). A further example of a novel type of catalytic site is provided by the families of repressor LexA (S24), bacterial leader peptidase 1 (S26), and eukaryote signal peptidase (S27), which seem to comprise a clan (SF) of serine peptidases that use a Ser/base catalytic dyad.

More detailed descriptions of these families follow, but at this stage it may be useful to draw attention to the common features of the enzymes of the proposed clan. As shown in Fig. 7, there are similarities in the sequences around the Ser and Lys residues that are known to have catalytic activity in families S24 and S26, and the Ser and His residues that may be equivalent in the eukaryotic signal peptidases (S27).

The properties of the catalytic serine residues in clan SF are dissimilar from those of serine residues in the "catalytic triad" peptidases of clans SA, SB, and SC. Thus, they show little or no reactivity with diisopropyl fluorophosphate (DFP) and other reagents commonly used for serine peptidases, and the serine can be replaced by cysteine with partial retention of activity (see [20] and [21]).

Finally, the peptidases in all three families show a preference for cleavage of alanyl bonds. Thus Ala is the preferred P1 residue for both

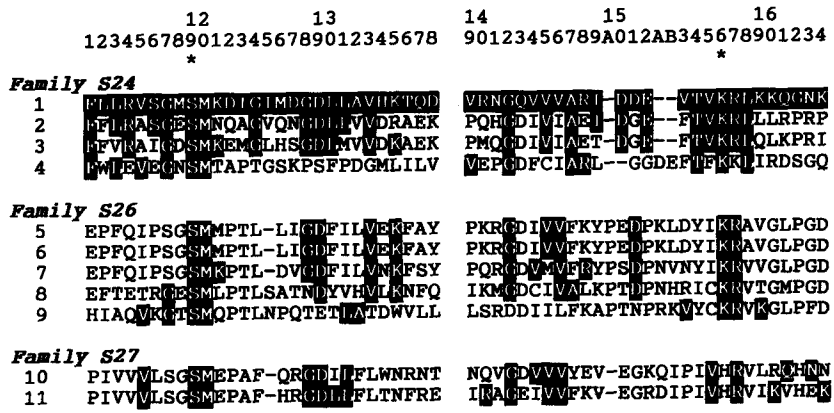

FIG. 7. Conserved residues in the vicinity of the catalytic residues in the families of repressor LexA (S24), leader peptidase (S26), and eukaryote signal peptidase (S27) forming clan SF. Residues are numbered according to *Escherichia coli* repressor LexA, and those identical in *E. coli* repressor LexA are shown in white on black. Asterisks indicate catalytic residues that have been identified in families S24 and S26, but are only presumed for family S27. Key to sequences: 1, *E. coli* repressor LexA; 2, *Salmonella typhimurium* ImpA protein; 3, *S. typhimurium* SamA protein; 4, bacteriophage lambda (λ) c1 repressor; 5, *E. coli* leader peptidase 1; 6, *S. typhimurium* leader peptidase 1; 7, *Pseudomonas fluorescens* leader peptidase 1; 8, yeast mitochondrial inner membrane leader peptidase; 9, yeast microsomal signal peptidase 18-kDa subunit; 10, dog microsomal signal peptidase 21-kDa subunit; 11, dog microsomal signal peptidase 18-kDa subunit.

prokaryotic and eukaryotic leader/signal peptidases,[70] and is highly conserved in the cleavage sites of the LexA family (see [20]). However, by no means all alanyl bonds are cleaved, and the specificity sites are apparently extended ones in which residues other than P1 are of great importance. No cleavage of simple substrates such as alanyl nitroanilides has been reported, and for leader peptidase it was found that a pentapeptide was the smallest substrate detectably cleaved.[71] Mutations several residues away on either side of the cleavage site can greatly affect the rate of autolytic cleavage of LexA (see [20]), again consistent with an extended substrate-binding site.

Repressor LexA Family (S24)

The protein known as LexA represses about 20 genes of the SOS regulon that are involved in DNA repair in *E. coli*. After treatments of the bacterium that damage DNA or inhibit DNA replication, LexA is inactivated, leading to derepression of the genes for DNA repair (see [20]). LexA is known from other bacteria too, but more surprisingly, homologous repressors also occur in lambdoid phages[72] (Table VIII). Cleavage of the phage repressors results in phage induction.

In *E. coli*, the derepression of the DNA repair system is effected by the RecA protein, which acquires "coprotease" activity in the presence of single-stranded DNA, causing it to undergo an interaction with LexA that results in inactivation of LexA by proteolytic cleavage. The bond cleaved by the peptidase activity of LexA is the Ala^{84}–Gly bond, which disrupts the DNA-binding part of the molecule, inactivating it. At one time, it was thought that the peptidase activity that causes the derepression was mediated by a catalytic site on the RecA protein, but it is now clear that this is an autolytic reaction, in which the catalytic activity is that of LexA (see [20]). Isolated LexA autolyses slowly when incubated *in vitro* at alkaline pH, but the reaction triggered by RecA is much more rapid.

The molecules of LexA and related repressors consist of about 200 amino acid residues, of which the N-terminal 90 or so form the DNA-binding domain, for which crystallographic structures are available.[73] The C-terminal part of the molecules contains the residues that form the catalytic site for autolytic peptide bond cleavage. Of the amino acid residues

[70] G. von Heijne, *Nucleic Acids Res.* **14,** 4683 (1986).
[71] I. K. Dev, P. H. Ray, and P. Novak, *J. Biol. Chem.* **265,** 20069 (1990).
[72] R. T. Sauer, R. R. Yocum, R. F. Doolittle, M. Lewis, and C. O. Pabo, *Nature (London)* **298,** 447 (1982).
[73] D. H. Ohlendorf, W. F. Anderson, M. Lewis, C. O. Pabo, and B. W. Matthews, *J. Mol. Biol.* **169,** 757 (1983).

TABLE VIII
PEPTIDASES OF FAMILIES OF REPRESSOR LexA (S24), BACTERIAL LEADER PEPTIDASE 1 (S26), AND EUKARYOTE SIGNAL PEPTIDASE (S27)[a]

Peptidase	EC	Database code
Family S24: LexA repressor		
Bacteriophage repressor protein (includes C1 and C2)	-	RPC1_LAMBD, RPC2_BPP22
ImpA protein (*Salmonella*)	-	IMPA_SALTY
LexA repressor	-	LEXA_*
MucA protein (*Salmonella*)	-	MUCA_SALTY
SamA protein (*Salmonella*)	-	SAMA_SALTY
SOS regulatory protein DinR (*Bacillus*)	-	DINR_BACSU
Umud protein (*Salmonella*)	-	UMUD_*
Family S26: Bacterial leader peptidase 1		
Bacterial leader peptidase 1	3.4.99.36	LEP_*, (X75604),(Z27458)
Mitochondrial inner membrane peptidase subunit 1	-	IMP1_YEAST
Mitochondrial inner membrane peptidase subunit 2	-	[b]
Family S27: Eukaryote signal peptidase		
Eukaryote signal peptidase	-	SC11_YEAST, SPC3_CANFA, SPC4_CANFA, (L11319)

[a] See Table III for explanation.
[b] J. Nunnari, T. D. Fox, and P. Walter, *Science* **262,** 1997 (1993).

that are conserved throughout the family, only Ser-119 (*E. coli* LexA numbering) and Lys-156 could not be replaced by Ala with retention of activity.[74] LexA had no significant reactivity with 1 m*M* DFP, but later work showed that Ser-90 is detectable labeled with 20 m*M* [^{3}H]DFP.[75]

The autolytic cleavage of LexA is an intramolecular reaction, and as such is difficult to study enzymologically, because the concentrations of substrate and enzyme cannot be altered independently. This difficulty has been overcome by the development of a system in which the N-terminal, substratelike part of the molecule and the C-terminal, catalytic domain are separated (see [20]).

Various other proteins are related to repressor LexA. The DinR protein from *Bacillus subtilis* is a repressor of the *dinR* and *dinC* genes and, similarly to the LexA protein, is inactivated during the SOS response on association with the RecA protein.[76] Other protein homologs lack the N-terminal DNA-binding region of the LexA protein, and are therefore

[74] S. N. Slilaty and J. W. Little, *Proc. Natl. Acad. Sci. U.S.A.* **84,** 3987 (1987).
[75] K. L. Roland and J. W. Little, *J. Biol. Chem.* **265,** 12828 (1990).
[76] W. H. Koch, D. G. Ennis, A. S. Levine, and R. Woodgate, *Mol. Gen. Genet.* **233,** 443 (1992).

not repressors. The UmuD, MucA, ImpA, and SamA proteins are involved in mutagenesis following ultraviolet light irradiation, and their cleavage is a prerequisite for mutagenesis.[76,77]

Bacterial Leader Peptidase 1 Family (S26)

Eubacteria contain at least three leader peptidases. Murein prelipoprotein peptidase (leader peptidase 2), which removes the leader peptide from one of the components of the bacterial outer membrane, may be an aspartic endopeptidase.[78] Type IV prepilin leader peptidase is a cysteine peptidase (see [32]). The serine-type leader peptidase has the more general function of removing the leader peptides from the other secreted proteins and proteins that are targeted to the periplasm and periplasmic membrane. It is thus most closely analogous to the eukaryote signal peptidase, but differs in being monomeric. The catalytic residues of *E. coli* leader peptidase 1 have been identified as a Ser/Lys dyad (Fig. 7) (see [21]).

Mitochondria possess three metallopeptidases that remove N-terminal targeting peptides, the mitochondrial processing peptidase (family M16) and the mitochondrial intermediate peptidase (M3), but the inner membrane leader peptidase is a serine enzyme. In fact, the enzyme is composed of two subunits, both homologous with the bacterial leader peptidase 1. The enzyme is responsible for removing leader peptides from both nuclear- and mitochondrial-encoded proteins destined for the inner mitochondrial space, and it is believed that both subunits are active, with distinct specificities.[79]

Eukaryote Signal Peptidase Family (S27)

The eukaryote microsomal signal peptidase complex, responsible for removing signal peptides from secretory proteins as they are transported into the lumen of the endoplasmic reticulum, is much more complex than its eubacterial and mitochondrial counterparts (see [22]). Varying numbers of subunits have so far been isolated; in yeast there are four, in chicken oviduct there are two, and in dog pancreas there are five. In each species, one of the subunits is glycosylated, but as far as is known, the glycosylated subunits are not directly involved in the peptidase activity. Of the other subunits so far sequenced, one from yeast, one from rat, and two from dog are homologous, and form family S27 (Table VIII). These subunits

[77] J. Hauser, A. S. Levine, D. G. Ennis, K. M. Chumakov, and R. Woodgate, *J. Bacteriol.* **174,** 6844 (1992).

[78] K. Sankaran and H. C. Wu, this series, Vol. 248, Chapter 12.

[79] J. Nunnari, T. D. Fox, and P. Walter, Science **262,** 1997 (1993).

(about 20 kDa) contain sequences reminiscent of those in which the catalytic residues of bacterial leader peptidase 1 are located (Fig. 7). The alignment suggests that the eukaryote signal peptidases depend on a Ser/His dyad for activity, in contrast to the Ser/Lys dyad of the eubacterial leader peptidases 1. The existence of two subunits in dog that may have catalytic activity is analogous to the situation in yeast mitochondrial inner membrane leader peptidase (family S26).

Indications of a Clan (SG) of ATP-Dependent Endopeptidases Embracing Families S14, S16, and S25

Although the serine endopeptidase components of the Clp endopeptidase (family S14), the multicatalytic endopeptidase complex (MEC; family S25), and endopeptidase La (S16) do not show similarities in sequence, all the peptidase domains are associated with ATP-binding domains that are homologous to each other, and in each case the interaction couples the hydrolysis of peptide bonds to that of ATP. In view of these similarities, it is natural to look closely for any indication that the proteins of these three families may have a single origin.

Clp and the subunits of MEC associate in approximately hexameric rings. These further associate with ATPase subunits, to produce complexes that appear to have similar topology in the electron microscope.[80] Immunological cross-reactivity between *E. coli* ClpP and subunits of MEC has been reported by two groups.[80,81] In addition, marked similarities in the chymotrypsin-like activities of Clp and MEC led Arribas and Castaño[80] to suggest that ClpP might be evolutionarily related to subunits expressing chymotrypsin-like activity in MEC.

In endopeptidase La the endopeptidase and ATPase components are parts of the single polypeptide chain, but there are other groups of enzymes in which catalytic and regulatory domains are fused in some members, and comprise separate subunits in others. Goldberg[82] has given a detailed account of the similarities and differences between endopeptidases La and Clp. In reviewing these indications of possible relationship between the three types of ATP-dependent endopeptidase, Rechsteiner *et al.*[83] provide an imaginative three-dimensional model to depict the way in which the structures may have evolved. However, there are no similarities around the catalytic residues in these enzymes. Moreover, though the

[80] J. Arribas and J. G. Castaño, *J. Biol. Chem.* **268,** 21165 (1993).
[81] K. Tanaka, T. Tamura, A. Kumatori, T. H. Kwak, C. H. Chung, and A. Ichihara, *Biochem. Biophys, Res. Commun.* **164,** 1253 (1989).
[82] A. L. Goldberg, *Eur. J. Biochem.* **203,** 9 (1992).
[83] M. Rechsteiner, L. Hoffman, and W. Dubiel, *J. Biol. Chem.* **268,** 6065 (1993).

TABLE IX
PEPTIDASES OF FAMILIES OF ClpP ENDOPEPTIDASE SUBUNIT (S14), ENDOPEPTIDASE La (S16), AND MULTICATALYTIC ENDOPEPTIDASE COMPLEX (S25)[a]

Peptidase	EC	Database code
Family S14: ClpP endopeptidase subunit		
ATP-dependent endopeptidase (ClpP subunit) (*E. coli*)	-	CLPP_ECOLI, (L07793)
Chloroplast ATP-dependent endopeptidase	-	CLPP_MARPO, CLPP_TOBAC, CLPP_ORYSA, CLPP_WHEAT, CLPP_EPIVI
Family S16: Endopeptidase La		
Endopeptidase La	3.4.21.53	LON_*, (D12923), (D13204)
Endopeptidase La-like protein	-	(X74215), (X76040)
Family S25: Multicatalytic endopeptidase complex		
Multicatalytic endopeptidase complex subunits	3.4.99.46	PRC1_*, PRC2_*, PRC3_*, PRC4_*, PRC5_*, PRC7_*, PRC8_*, PRC9_*, PRCA_*, PRCB_*, PRCC_*, PRCD_*, PRCU_YEAST, PRCX_YEAST, PRCZ_YEAST, PR28_DROME, PR29_DROME, PR35_DROME, (D10754), (D10755), (D10757), (D10758), (D21799), (L17127), (L22212), (L22213), (M63641), (M64992), (U00790), (X57210), (X70304)
SCL1 suppressor protein	-	SCL1_YEAST

[a] See Table III for explanation.

order of catalytic residues is Ser/His in ClpP, this cannot be the case in endopeptidase La.

ClpP Endopeptidase Subunit Family (S14)

The members of families S14, S16, and S25 are listed in Table IX. Clp endopeptidase is a bacterial, ATP-dependent endopeptidase (see [23]). Historically, the name "clp" arose as an acronym for "caseinolytic protease"; the enzyme has also been termed endopeptidase Ti.[82]

The Clp endopeptidase exists as a complex of two types of subunits (ClpP and ClpA), of composition $ClpP_{12}ClpA_6$. The ClpP subunit is a heat-shock protein synthesized under the control of σ^{32} factor.[84] Alone, ClpP has peptidase activity that is restricted to peptides of five residues or

[84] H. E. Kroh and L. D. Simon, *J. Bacteriol.* **172,** 6026 (1990).

less, so it could be described as an oligopeptidase, but full endopeptidase activity is expressed by the complex with ClpA (or ClpB or ClpX; see [23]). The additional subunits are homologous to the ATP-binding domain of the monomeric endopeptidase La, and will be considered below. The active site serine (Ser-111) and histidine (His-136) residues of ClpP have been identified for the *E. coli* enzyme, but no third member of a possible catalytic triad.

Proteins with sequences homologous to ClpP have been detected in chloroplast inner envelope membrane and soluble fraction,[85] and are encoded by the chloroplast genome.[86] The chloroplast protein has not been characterized, but ATP-dependent proteolysis has been detected in these organelles.[87] The chloroplast ClpP-like sequences have the active site serine and histidine residues conserved, but as yet no chloroplast ClpA component has been sequenced. We report here that the gene sequence of phosphoproteins of photosystem II from the chloroplast of the evening primrose (*Oenothera argillicola*)[88] includes a fragment of a homolog of ClpP at the 3′ end of the complementary strand.

A surprising homolog of ClpP is the 5′ untranslated region of potato leaf-roll luteovirus (EMBL : D00530). This virus is a single-stranded RNA virus most closely related to the picornaviruses. The genomic RNA sequence[89] may be truncated at the 5′ end. The ClpP homolog retains a residue equivalent to the active site serine, but not the active site histidine, and so is almost certainly inactive. Presumably, the viral sequence was acquired by horizontal genetic transfer from the potato host.

It has been suggested[83] that ClpP is homologous to ubiquitin-peptidase 2 from yeast, a cysteine peptidase (see [32]). In the comparison with the *E. coli* ClpP, 19.6% identity can be detected over a region of 189 amino acids, but there is no statistically significant relationship (RDF score = 4.90). Moreover, the similarity is seen only with the *E. coli* ClpP and not with the chloroplast homologs, and the active site Ser of ClpP is replaced by Asp in the ubiquitin peptidase.

Endopeptidase La Family (S16)

Unlike the Clp endopeptidase, endopeptidase La (product of the *lon* gene of *E. coli*) is a single-chain mosaic protein, containing an ATP-binding domain and a peptidase domain (see [25]). The ATP-binding domain is

[85] T. Moore and K. Keegstra, *Plant Mol. Biol.* **21,** 525 (1993).
[86] J. C. Gray, S. M. Hird, and T. A. Dyer, *Plant Mol. Biol.* **15,** 947 (1990).
[87] X.-Q. Liu and A. T. Jagendorf, *FEBS Lett.* **166,** 248 (1984).
[88] K. Offermann-Steinhard and R. G. Herrmann, *Nucleic Acids Res.* **18,** 6452 (1990).
[89] M. A. Mayo, D. J. Robinson, C. A. Jolly, and L. Hyman, *J. Gen. Virol.* **70,** 1037 (1989).

homologous to the ClpA and ClpB subunits of the Clp endopeptidase, and a number of other ATP-binding proteins such as bacterial NIF-specific and XYLR regulatory proteins (NIFA_KLEPN, XYLR_PSEPU).

To date, only the catalytic serine residue (Ser-679) has been identified.[90] The sequences contain no conserved His C-terminal to this.

Homologs of endopeptidase La have been discovered in the mitochondria of eukaryotes (see [26]). These are from yeast and human cells.

Multicatalytic Endopeptidase Complex Family (S25)

The multicatalytic endopeptidase complex (MEC), also called the proteasome, is a 700-kDa endopeptidase containing about 24–28 similarly sized, homologous subunits. The complex occurs in the cytoplasm and nuclei of eukaryotic cells generally, as well as those of the archaebacterium *Thermoplasma acidophylum* (see [24]). In the electron microscope, MEC is seen as a cylinder, formed as a stack of four rings, each of about seven subunits.

In the archaebacterium, MEC is composed of subunits of only two kinds, α and β.[91] In eukaryotes, there are far more different subunits, but all are variants of the α or β type of structure, and are divided into A and B groups on this basis.

Three types of peptidase activity are simply recognized by use of synthetic substrates, and these can be described as trypsin-like (although cleavage is predominantly of Arg+ bonds), chymotrypsin-like (cleaving Leu+, Tyr+, and Phe+ bonds), and glutamyl peptidase (cleaving Glu+ bonds). The *Thermoplasma* enzyme shows only the chymotrypsin-like activity,[92] but the eukaryotic endopeptidase complex shows all three, and indeed at least five separate catalytic sites are distinguishable in careful work (see [24]). Whether the endopeptidase activity against β-casein is attributable to any of the sites active on synthetic substrates is not clear.[93] It has yet to be established which subunits, and which catalytic residues, are responsible for the various activities, but most of these are inhibited by 3,4-dichloroisocoumarin, and are probably those of atypical serine peptidase catalytic sites.

An alignment of the available sequences made with the PILEUP pro-

[90] A. Y. Amerik, V. K. Antonov, A. E. Gorbalenya, S. A. Kotova, T. V. Rotanova, and E. V. Shimbarevich, *FEBS Lett.* **287,** 211 (1991).

[91] P. Zwickl, A. Grziwa, G. Pühler, B. Dahlmann, F. Lottspeich, and W. Baumeister, *Biochemistry* **31,** 964 (1992).

[92] B. Dahlmann, L. Kuehn, A. Grziwa, P. Zwickl, and W. Baumeister, *Eur. J. Biochem.* **208,** 789 (1992).

[93] M. E. Pereira, T. Nguyen, B. J. Wagner, J. W. Margolis, B. Yu, and S. Wilk, *J. Biol. Chem.* **267,** 7949 (1992).

gram (with the standard defaults) shows very few residues completely conserved. The possibility must be borne in mind that the minimal catalytic unit comprises both an A and a B type of subunit, with each contributing some of the active site residues. Among the few residues that are completely conserved in the sequences (both A and B type) are Gly-80, Asp-84, and Gly-166 (numbered according to the *Thermoplasma* α subunit). The potential active site residues completely conserved in the A subunits are Ser-16, Tyr-26, Asp-84, and perhaps Arg-130 (although Arg has not yet been proved to be a catalytic residue in any peptidase). In the B subunit, Asp-84 is the only completely conserved potential catalytic residue.

There is some evidence for proteolytic processing in the rat B subunits, several of which have N-terminal threonine residues, implying a Gly↓Thr cleavage at a site near the N terminus that is conserved in the great majority of B subunits.[94]

Despite some contrary reports,[95,96] the 26S ATP-dependent, ubiquitin conjugate-degrading enzyme complex is now generally believed to contain components of MEC as its catalytic unit. Additional components then confer the ATP dependence and the specificity for ubiquitin conjugates.[83]

The one additional subunit of the 26S complex (EMBL : L02426) that has been sequenced includes putative ATP-binding domains that show significant similarities to those in ClpA (RDF score 6.93) and endopeptidase La (RDF score 5.99).[97] The subunit also shows a relationship to human Tat-binding protein 1 (TBP1_HUMAN), yeast cell division control protein 48 (CD48_YEAST), and a yeast peroxisome biosynthesis protein (PAS1_YEAST), among others.

Other Serine Peptidase Families

The members of families S18, S19, S21, S17, and S23 are listed in Table X.

Omptin Family (S18)

The *E. coli* outer membrane endopeptidase omptin (the product of the *ompT* gene), previously known as protease VII, cleaves between paired basic amino acid residues (see [27]). Omptin is inhibited by serine peptidase inhibitors such as DFP, phenylmethylsulfonyl fluoride (PMSF), and

[94] K. S. Lilley, M. D. Davison, and A. J. Rivett, *FEBS Lett.* **262,** 327 (1990).
[95] A. Seelig, P.-M. Kloetzel, L. Kuehn, and B. Dahlmann, *Biochem. J.* **280,** 225 (1991).
[96] L. Kuehn, B. Dahlmann, and H. Reinauer, *Arch. Biochem. Biophys.* **295,** 55 (1992).
[97] W. Dubiel, K. Ferrell, G. Pratt, and M. Rechsteiner, *J. Biol. Chem.* **267,** 22699 (1992).

TABLE X
MEMBERS OF OTHER FAMILIES OF SERINE PEPTIDASES[a]

Peptidase	EC	Database code
Family S18: Omptin		
Omptin (*Escherichia coli*)	3.4.21.87	OMPT_ECOLI
Outer membrane protease (*Escherichia*)	-	(X74278)
Coagulase/fibrinolysin (*Yersinia*)	-	COLY_YERPE
Phosphoglycerate transport system activator (*Salmonella*)	-	PGTE_SALTY
Family S19: *Coccidioides* endopeptidase		
Chymotrypsin-like protease (*Coccidioides*)	-	(X63114)
Family S21: Assemblin, herpesvirus		
Assemblin	-	VP40_*, (M64627)
Family S17: *Bacteroides* endopeptidase		
Extracellular endopeptidase (*Bacteroides*)	-	PRTE_BACNO
Family S23: *Escherichia coli* protease I		
Protease I	-	TESA_ECOLI

[a] See Table III for explanation.

Tos-Phe-CH_2Cl, but to date the active site residues have not been identified.[98]

There are two homologs of omptin from other bacterial species. The E protein from *Salmonella typhimurium* is functionally analogous in that it is also located in the outer membrane and is capable of cleaving substrates similar to those cleaved by omptin. However, whereas omptin has a typical signal peptide for transport to the outer membrane, the N terminus of the protein E precursor is longer, with little similarity.[99]

A second homolog is known from the plague organism, *Yersinia pestis*, and is believed to have a role in the transmission of the disease. The enzyme expresses both coagulant and fibrinolytic activities, according to the temperature.[100] The coagulant activity causes the blood meal to clot in the gut of the flea, which responds to ejecting the gut contents back into the host bloodstream together with the plague organism. The coagulant activity does not seem to be mediated by prothrombin activation, so the enzyme may act directly on fibrinogen. Fibrinolytic activity, in contrast, *is* brought about by activation of host plasminogen, producing the

[98] K. Sugimura and T. Nishihara, *J. Bacteriol.* **170,** 5625 (1988).
[99] J. Grodberg and J. J. Dunn, *J. Bacteriol.* **171,** 2903 (1989).
[100] K. A. McDonough and S. Falkow, *Mol. Microbiol.* **3,** 767 (1989).

same N terminus as host plasminogen activator, cleaving an Arg+Val bond. The *Yersinia* enzyme also seems to be capable of preventing an inflammatory reaction by destroying complement component C3, a component important in the production of the chemoattractant C5a fragment.[101]

Coccidioides Endopeptidase Family (S19)

Coccidioides imitis is a soil fungus, and the causative agent of coccidioidomycosis. Infection results from the inhalation of air-borne spores. The endopeptidase, described as chymotrypsin-like, is associated with the cell wall during active growth of the pathogenic cells.[102] The peptidase has been shown to be collagenolytic, elastinolytic, and able to cleave human IgG and IgA. Activity is inhibited by Tos-Phe-CH_2Cl, chymostatin, and α_1-proteinase inhibitor.[103] The sequence shows no relation to any other protein, and the active site residues have not been determined.

Assemblin Family (S21)

Cytomegaloviruses, which include the herpes simplex and Epstein–Barr viruses, are double-stranded DNA viruses. The virus particles are assembled in a process similar to that of bacteriophages, in which the proteins of the head are built on a scaffold called the assembly protein. The prohead so formed becomes the head of the mature virus once the assembly protein has been degraded and the empty space filled with DNA.

Degradation of the assembly protein is an autolytic event. The assembly protein has the proteinase, assemblin, as its C-terminal domain. Assemblin is inhibited by serine peptidase inhibitors such as DFP and PMSF, and cleaves Ala+Ser or Ala+Val bonds to release a polypeptide from the C terminus of the assembly protein. Assemblin may also release itself from its precursor, which would entail cleavage at an Ala+Ser bond, and exceptionally an Ala+Asn bond in infectious laryngotracheitis virus.

In the assemblin of simian cytomegalovirus, catalytic residues have been identified by site-directed mutagenesis as His-47 and Ser-118.[104]

Bacteroides Endopeptidase Family (S17)

Bacteroides nodosus is the causal agent of ovine footrot, in which the hoof separates from the underlying epidermis. The presence of an

[101] O. A. Sodeinde, Y. V. B. K. Subrahmanyam, K. Stark, T. Quan, Y. Bao, and J. D. Goguen, *Science* **258,** 1004 (1992).

[102] G. T. Cole, S. Zhu, L. Hsu, D. Kruse, K. R. Seshan, and F. Wang, *Infect. Immun.* **60,** 416 (1991).

[103] L. Yuan and G. T. Cole, *Infect. Immun.* **55,** 1970 (1987).

[104] A. R. Welch, L. M. McNally, M. R. T. Hall, and W. Gibson, *J. Virol.* **67,** 7360 (1993).

endopeptidase with chymotrypsin-like specificity able to digest elastin, keratin, fibrinogen, and collagen implies that the disease is a proteolytic process. The enzyme has been identified as a divalent cation-dependent serine endopeptidase, but the sequence reported by Moses *et al.*[105] shows no resemblance to any other protein. A subtilisin homolog from this bacterium has been sequenced,[106] raising the possibility that the protein identified by immunological screening of an expression library by Moses *et al.* may not be the peptidase after all.

Escherichia coli Protease I Family (S23)

This is a one-member family containing a periplasmic enzyme from *E. coli* that cleaves Tyr+ and Phe+ bonds in ester substrates such as Z-Tyr-*p*-nitrophenyl ester. The enzyme had no amidase activity, and low activity on polypeptides.[107] Activity was inhibited by DFP, but no active site residues are known.

The published sequence[108] is identical to that of acyl-CoA thioesterase I, in which the active site serine was identified as Ser-10.[109] His-157 occurs within a Gly-Ile-His motif that is conserved in other serine thioesterases and may be part of the active site. The enzyme is synthesized with a 26-residue leader peptide and is predicted to be a periplasmic enzyme (although previously thought to be cytoplasmic). There have been doubts for some time that protease I is a peptidase.[110]

[105] E. K. Moses, J. I. Rood, W. K. Yong, and G. G. Riffkin, *Gene* **77,** 219 (1989).
[106] G. G. Lilley, D. J. Stewart, and A. A. Kortt, *Eur. J. Biochem.* **210,** 13 (1992).
[107] M. Pacaud, L. Sibilli, and G. Le Bras, *Eur. J. Biochem.* **69,** 141 (1976).
[108] S. Ichihara, Y. Matsubara, C. Kato, K. Akasaka, and S. Mizushima, *J. Bacteriol.* **175,** 1032 (1993).
[109] H. Cho and J. E. Cronan, Jr., *J. Biol. Chem.* **268,** 9238 (1993).
[110] J. D. Kowit, W.-N. Choy, S. P. Champe, and A. L. Goldberg, *J. Bacteriol.* **128,** 776 (1976).

[3] Myeloblastin: Leukocyte Proteinase 3

By JOHN R. HOIDAL, N. V. RAO, and BEULAH GRAY

Myeloblastin (EC 3.4.21.76) is a ~29,000-Da neutral serine endopeptidase produced during myeloid differentiation and stored in the azurophilic granules of polymorphonuclear leukocytes (PMNs), where it is present in amounts comparable to elastase and cathepsin G. Myeloblastin was