Acta Biochimica Polonica

Vol. 50 No. 1/2003

81–102

www.actabp.pl

Review

This review is dedicated to Professor Paulina Wlodawer, formerly of the Nencki Institute of Experimental Biology, Warsaw, Poland, who contributed in a significant way to establishing Polish nomenclature for many enzymes

Structural and enzymatic properties of the sedolisin family of serine-carboxyl peptidases✪

Alexander Wlodawer[1✉], Mi Li[1,2], Alla Gustchina[1], Hiroshi Oyama[3], Ben M. Dunn[4] and Kohei Oda[3]

[1]*Protein Structure Section, Macromolecular Crystallography Laboratory, National Cancer Institute at Frederick, Frederick, MD 21702, U.S.A.;* [2]*Basic Research Program, SAIC-Frederick, Inc., National Cancer Institute at Frederick, Frederick, MD 21702, U.S.A.;* [3]*Department of Applied Biology, Faculty of Textile Science, Kyoto Institute of Technology, Sakyo-ku, Kyoto 606-8585, Japan;* [4]*Department of Biochemistry and Molecular Biology, University of Florida, Gainesville, Florida 32610, U.S.A.*

Received: 29 January, 2003; accepted: 10 March, 2003

Key words: protein structure, enzyme families, enzymatic machanism, active site, sequence conservation

Sedolisins (serine-carboxyl peptidases) are proteolytic enzymes whose fold resembles that of subtilisin; however, they are considerably larger, with the mature catalytic domains containing approximately 375 amino acids. The defining features of these enzymes are a unique catalytic triad, Ser-Glu-Asp, as well as the presence of an aspartic acid residue in the oxyanion hole. High-resolution crystal structures have now been solved for sedolisin from *Pseudomonas* sp. 101, as well as for kumamolisin from a thermophilic bacterium, *Bacillus* novo sp. MN-32. The availability of these crystal structures enabled us to model the structure of mammalian CLN2, an enzyme which, when mutated in humans, leads to a fatal neurodegenerative disease. This review compares the structural and enzymatic properties of this newly defined MEROPS family of peptidases, S53, and introduces their new nomenclature.

✪This work was supported in part by a Grant-in-Aid for Scientific Research (B), 13460043, from the Ministry of Education, Culture, Sports, Science and Technology of Japan (to K.O.); by NIH grants DK18865 and AI39211 (to B.M.D.); and in part with Federal funds from the National Cancer Institute, National Institutes of Health, under contract No. NO1-CO-12400.

✉Correspondence should be addressed to A.W. e-mail: wlodawer@ncifcrf.gov

Abbreviations: see Fig. 3.

Proteins in general, and proteolytic enzymes in particular, are assigned to families and clans based on their primary through tertiary structures, the nature and location of the residues in their active sites, and their mechanism of action. This information is gathered in databases such as MEROPS (http://merops.sanger.ac.uk) and is periodically summarized in handbook form (Barrett *et al.*, 1998; new edition in preparation). In this review, we will discuss the structural and enzymatic properties of the recently characterized family of sedolisins, also known as serine-carboxyl peptidases (note that the terms *proteolytic enzyme, protease, proteinase,* and *peptidase* are largely interchangeable; here, the latter will be utilized preferentially). The name **sedolisin** is being introduced here for the first time, in an attempt to unify and correct the confusing and often misleading nomenclature of these enzymes. The reasons for this choice of name will become clear later.

Several proteolytic enzymes with common properties such as maximum activity at comparatively low pH, the presence of conserved acidic residues (both aspartates and glutamates) required for activity, and lack of inhibition by pepstatin, have been isolated, characterized, and cloned in the last 16 years. The first such enzyme, now called sedolisin (see Table 1 for the proposed nomenclature of the members of the family), was found in *Pseudomonas* sp.101 (Oda *et al.*, 1987; 1994; Oyama *et al.*, 1999; Ito *et al.*, 1999). It was originally assigned the name ***P**seudomonas* pepstatin-insensitive **c**arboxyl **p**roteinase (PCP) (Oda *et al.*, 1987), and was later renamed ***P**seudomonas* **s**erine-**c**arboxyl **p**roteinase (PSCP) (Wlodawer *et al.*, 2001a). Other related bacterial enzymes include *Xanthomonas* sp. T-22 carboxyl proteinase (originally named XCP, later XSCP, now renamed sedolisin-B) (Oda *et al.*, 1987); kumamolisin (originally called kumamolysin, later KCP or KSCP), an enzyme isolated from a thermophilic bacterium, *Bacillus* novo sp. MN-32 (Murao *et al.*, 1993); and alcohol-resistant proteinase J-4 (now kumamolisin-B), isolated from *Bacillus coagulans* (Shibata *et al.*, 1998). Most recently, an enzyme isolated from *Alicyclobacillus sendaiensis* (originally named ScpA, now renamed kumamolisin-*As*) was characterized as a collagenase (Tsuruoka *et al.*, 2003). The sequences of all these enzymes were similar enough to postulate that they had to form a single family; they were originally assigned to an unknown clan, family A7 of aspartic peptidases (Barrett *et al.*, 1998), under the name of pepstatin-insensitive aspartic peptidases. However, that assignment became questionable after the peptidase CLN2, a human enzyme which, when mutated, leads to a fatal neurodegenerative disease, classical late-infantile neuronal ceroid lipofuscinosis (Sleat *et al.*, 1997), was identified as a tripeptidyl-peptidase I (TPP-I) and tentatively classified as a serine peptidase (Rawlings & Barrett, 1999; Lin *et al.*, 2001). Two other related enzymes that were subsequently reported as putative serine peptidases were LYS60 and LYS45, markers for late lysosomes in *Amoeba proteus* (Kwon *et al.*, 1999). However, the unambiguous assignment of all of these enzymes to the clan SB of peptidases, which previously consisted of only the subtilisin family (S8), became possible only after crystal structures of the representative members became available (Wlodawer *et al.*, 2001a; 2001b; Comellas-Bigler *et al.*, 2002). The family of serine-carboxyl peptidases, S53, is now assigned in MEROPS (http://merops. sanger.ac.uk) as the second member of the SB clan.

STRUCTURAL FEATURES OF SERINE-CARBOXYL PEPTIDASES

Although initial comparisons between members of a protein family are usually based on their amino-acid sequences, the availability of three-dimensional structures was crucial for proper understanding of the properties of sedolisins and the placement of these enzymes among other peptidases. For these rea-

sons, the three-dimensional structures of the members of this family will be discussed first and the primary structures will be compared only later in this review. As mentioned above, crystal structures are now available for two members of the family of serine-carboxyl peptidases, sedolisin (PSCP) (Wlodawer *et al.*, 2001a; 2001b) and kumamolisin (KSCP) (Comellas-Bigler *et al.*, 2002). These structures are of excellent quality and very high resolution (some as high as 1 Å) and they include both enzymes alone and numerous inhibitor complexes. For historical reasons, the discussion will be based principally on the structures of sedolisin, whereas only the unique features of the structure of kumamolisin will be discussed in detail.

The three-dimensional fold of sedolisin (Fig. 1) is based on a 7-stranded, all-parallel β sheet consisting of strands s2-s3-s1-s4-s5-s6-s7b. A diagram of the secondary structure elements is shown in Fig. 2; for reasons given below, the descriptors follow the nomenclature used previously for subtilisin. The sheet is flanked on both sides by several helices. On one side, these are helices h4 and h5, parallel to each other, but with their direction opposite of the direction of the strands in the sheet. Four helices – h2′, h3, h6, and h8 – flank the other side, again parallel to each other and antiparallel to the sheet. Five of these parallel helices are also involved in creating the extensive core of the molecule. Two helices (h6 and h3) are buried in the central part of the molecule, with the surface helix h2′ interacting with the latter. The helices on the other side of the sheet, h4 and h5, while providing extensive buried surfaces, are partially exposed at the surface of the protein. The second and third β strands (s2 and s3), located at one edge of the central sheet, are connected to helix h3 in a rare left-handed crossover, a feature that has been first described in subtilisin (Wright *et al.*, 1969), and later in only a handful of proteins, such as acetylcholine esterase (Sussman *et al.*, 1991), steroid dehydrogenase (Ghosh *et al.*, 1991), and L-asparaginase (Miller *et al.*, 1993). Left-handed crossovers are usually found in areas important for activity, and the one in sedolisin is not an exception, as helix h3 carries two of the active-site residues (see below). The fold of the protein is completed by several other shorter strands and helices, with three pairs of strands (s8-s9, s6′b-s6′′a, and s6′′b-s7a) forming β hairpins on the surface of the protein. Helix h2 is absent in kumamolisin, whereas a short additional strand (s1a), not present in sedolisin, could be identified in that enzyme (Comellas-Bigler *et al.*, 2002).

Two proline residues, Pro192 and Pro260, are present in *cis* configuration in sedolisin.

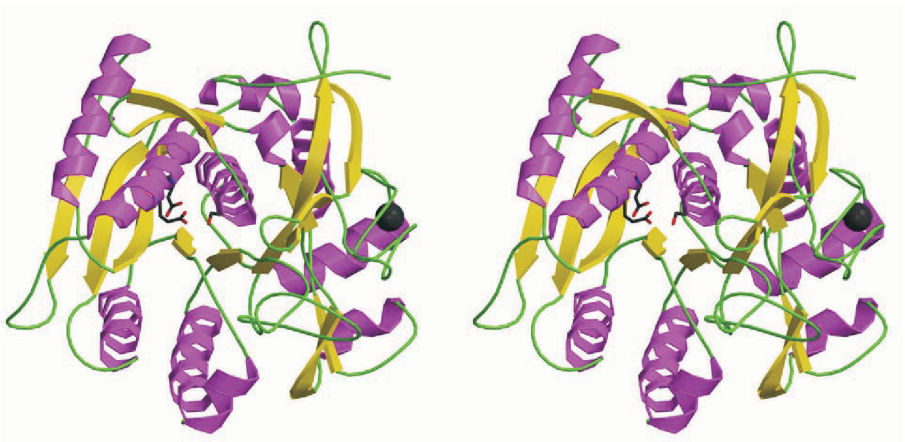

Figure 1. Stereo tracing of the Cα backbone of sedolisin.

Helices are shown in purple, β-strands in gold, and loops in green. The calcium cation is shown as a black ball. Active site residues are shown in stick representation. Figure prepared with Molscript (Kraulis, 1991).

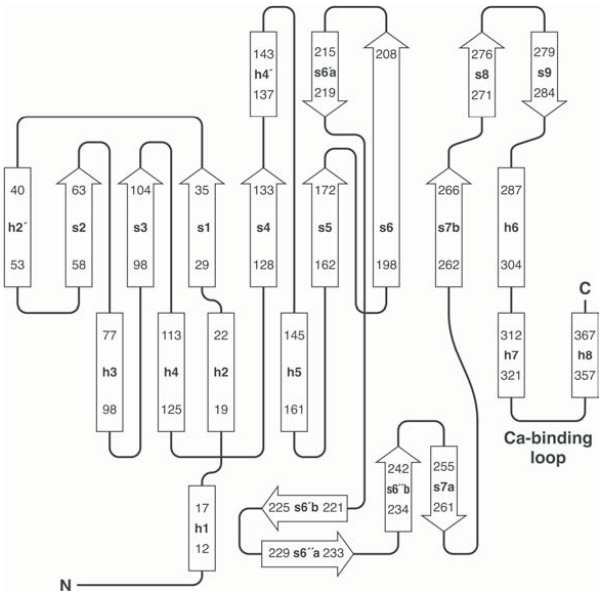

Figure 2. Schematic secondary structure diagram of sedolisin, showing the extent and names of the individual elements.

The names of α helices and β strands are based on the corresponding features in subtilisin (see Fig. 2 in Wlodawer *et al.*, 2001a and Fig. 2 in Comellas-Bigler *et al.*, 2002 for more details). Structural elements present in sedolisin but not in subtilisin are marked with single or double primes, and those that are additionally split into segments are designated *a* and *b*.

The first of them is located in a stretch of irregular structure located between strands s5 and s6, leading to a sharp change in the direction of the peptide chain. The latter *cis* peptide creates a kink between strands s7a and s7b. Both of these *cis* peptides appear to be needed because of steric requirements of the local structure and are also present in kumamolisin. Pro192 is conserved in all known serine-carboxyl peptidases, whereas Pro260 is not conserved in sedolisin-x*Ap*B (LYS60), indicating either a significant difference in the structure of this enzyme, or potential errors in sequence alignment (Fig. 3).

Surprisingly, Tyr331 in kumamolisin is also found in the *cis* configuration. That residue corresponds to Tyr341 in sedolisin, with the latter amino acid assuming the more common *trans* form. While this difference in the structure of the main chain has almost no influence on the placement of the side chains of Tyr341 and of the several highly-conserved following residues, the chain preceding this point follows considerably different paths in these two enzymes. A two-amino-acid insertion in kumamolisin and in kumamolisin-B (peptidase J-4), but not in sedolisin-B (XSCP), may be responsible for this unusual conformational difference. In view of the high resolution of the multiple structures of sedolisin and kumamolisin that are available, this difference cannot be an artifact of the refinement process, but must reflect a true variation of the structures.

A disulfide bridge connects Cys137 and Cys176 in sedolisin, and no other cysteines are present in the amino-acid sequence of this enzyme. Three cysteines are found in the mature kumamolisin (Comellas-Bigler *et al.*, 2002). Two of them, Cys190 and Cys340, are buried deep in the hydrophobic core of the protein and are too far from each other to form a disulfide bridge. Cys27, although located not far from the surface, buries its Sγ atom in an internal cavity. From molecular modeling results (see below), one could expect that a disulfide linking Cys327 and Cys342 is present in CLN2, whereas LYS60 (sedolisin-x*Ap*B) might contain a disulfide between Cys168 and Cys349. It is clear that neither the location of cysteine residues nor the presence of disulfide bonds are conserved features of this family.

The experimentally-obtained crystal structures of serine-carboxyl peptidases are very similar between their different variants. For example, kumamolisin has been crystallized in two different forms, with either one or two molecules in the asymmetric unit. Comparison of the two crystallographically independent molecules in the dimeric form of kumamolisin results in an r.m.s. deviation of only 0.17 Å for all 357 Cα atoms, whereas the deviation between the molecules in the monomeric and dimeric forms is 0.23 Å. Comparison of different structures of sedolisin

```
PSCP     1    AAGTAKGHNPTEFPTIYDASSAPTAA---NTTVGIITIGGVSQTLQDLQQFTSANGLASVNTQTIQ-----TGSSNGDYSDDQQG-QGEWDLDS
KSCP     1      AAPTAYTPLDVAQAYQFPEGLDGQ---GQCIAIIELGG-GYDETSLAQYFASLGVSAPQVVSVS-----VDGATNQPTGDPNGPDGEVELDI
XSCP     1       AVAAHHPQDFAAIYGGSSLPAAT---NTAVGIITWGSITQTVTDLNSFTSGAGLATVNSTITK-----VGSGTFANDPD---SNGEWSLDS
J-4      1      TEQVAYTPPQVAAAYAFPENVDCT---GQCIGIIELGG-GYSDQNIGEYFASLGMAPPPLVAVG-----VDGAVNQPTGDPNGPDGEFELDI
PHYS1    1     AGVDGYIVPYVIFDLYGIPTTFPVH--PNSSICLVEFQDDQSYNKDDLKKFAKENEITETVVSHT-----------VGPYSGSSADTESTLDV
PHYS2    1  VAPAVTTQPFNPTVDVLPLYGIDTSDSNRGAGQVIGIIDAYGASTAESDLAAFSRANGLPAANFQKVDQNGGTNYPKDDPDDASGDGWGVETALDL
LYS60    1                          ARAKPVTGNSSVGVIEFEDQNFAPSDLSDFATSFSVPITPLTDNH------IIGSNDPTSPQ----IEATLDI
hCLN2    1       LHLGVTPSVIRKRYNLTSQDVGSGTSNNSQACAQFLEQYFHDSDLAQFMRLFGGNFAHQASVA--------RVVGQQGRGRA-GIEASLDV

PSCP    86  QSIVGSAGGAVQQLL--FYMADQ--SAS---GNTGLTQAFNQAVSDN--VAKVINVSLGWCEADANADGTLQAEDRIFATAAAQGQTFSVSSGDEGV
KSCP    84  EVAGALAPGA--KIA--VYFAPN--------TDAGFLNAITTAVHDPTHKPSIVSISWGGPEDSWAP-ASIAAMNRAFLDAAALGVTVLAAAGDSGS
XSCP    81  QDIVGIAGG-VKQLI--FYTSANGDSSSSGITDAGITASYNRAVTDN--IAKLINVSLGEDETAAQQSGTQAADDAIFQQAVAQGQTFSISASGDAGV
J-4     84  EVAGAVAPGA--KLA--VYFAPN--------TDAGFLNAITTAIHDTTNKPMVISISWGGPENGWAP-ASIQAMNRALQDAAALGVTVCVAAGDSGS
PHYS1   81  QYGGAIALNTT------VWFWTV--------EDWMYDFATDFLNTKN--PPLVVSMSWGWPEPEQCQVGNC1RTNVEFQKIGAIGTTLLAASGDQGA
PHYS2   97  QIAHAVAPAA-----KLILCTAKSASDTN---LNACIKTL---TNLHVNH---ISMSYGGSEGDTSSDSY--------FQQAQEAGISLFASAGDSGA
LYS60   64  QYILGVLLVQLDGSGLKVIVYRL----------YGFHHLF--ATKDVPLV---NSISYGWNEEDQCENGIGGAECQQLGVTSAQ2ITLFAASGDSGA
hCLN2   83  QYLMS--AGANISTW--VYSSPGRHE------GQEPFLQWLMLLSNESALPHVHTVSYGDDEDSLSS-AYIQRVNTELMKAAARGLTLLFASGDSGA

PSCP   174  YECNN------RGYP---------DGSTYSVSWPASSPNVIAVGGTTLYTTSAGAYSNETVWNEGLDS--------NGKLWATGGGYSVYESK----
KSCP   167  TDGEQDG--------------------LYHVDFPAASPYVLACGGTRLV-ASAGRIERETVWNDGP------------DGGSTGGGVSRIFPL----
XSCP   173  YQWSTDPTSGSPGYVANSAGTVKIDLTHYSVSEPASSPYVIQVGGTTLS-TSGTTWSGETVWNEGLSAIAPSQGDNNQRLWATGGGVSLYEAA----
J-4    167  TDGE--------------------NDGLYHVDFPASSPYALACGGTRLVVEN-GQIASETVWNDGA------------QGGSTGGGVSSVFAL----
PHYS1  172  PGDSDPE--------------CNSKKKPLSSIFPGASPWVLSVGATMLSNM3CSTSTTELVCTIPQ------------ALITTGGGFSDYSLQ----
PHYS2  173  ----------------------------EAEYPAASQYVVAVGGTTLHTNSDGSFNSETAWS------------------GSGGGCSKYTKVIPEQ
LYS60  161  NGRTDP----------------DCSESNLNPAYPAASPYITSVGATQISQSSG4LRKRSAMINPILHQVED-FLWLPQHCLSEGSHCSILQVR----
hCLN2  169  GCWS------------------VSGRHQFRPTFPASSPYVTTVGGTSFQEP--FLITNEIV------------------DYISGGGFSNVFPR----

PSCP   244  ---PSWQSVV-------------SGTPGR---RLLPDISFDAAQGTGALIYNYG-QLQQIGGTSLASPIFVGLWARLQSAN---SNS-LGFPAASFY
KSCP   228  ---PSWQERA---------NVPPSANPGAGSGRGVPDVAGNADPATGYEVVIDG-ETTVIGGTSAVAPLFAALVARINQKL---GKP-VGYLNPTLY
XSCP   265  ---PSWQSSV-------------SSST-K---RVGPDLAFDAASSSGALIVVNG-STEQVGGTSLASPLFVGAFARIESAA---NNA-IGFPASKFY
J-4    228  ---PSWQQNA---------QVPPSSNPGGTVGRGVPDVAGDADPATGYEVLVDG-ETAAIGGTSAVAPLWAGLVTIANKTI---GQP-VGYLNPVLY
PHYS1  206  ---PSYQNAAVAAYFKSGVPLPPQTDFNA-SNRGFPDVSALGH---NYLIALSG-DFEQVDGTSASTPVFAAIIAHLNSYRLNNGKPPLAFAVPLIY
PHYS2  223  NADPGYA----------------SLGC--KGKKALPDLAALADPNSGITIIFHGQELEGIGGTSLAAPLTAGR-AAIRGDQ----------VTPAYV
LYS60  257  ---CNLA----------------SKLLLQCCCKRISDVSALGS---AILIETGG-NIQTVGGTSASSPIFAGVVGLLNDYVNSKTGKPLGFVSPLLY
hCLN2  224  ---PSYQEEAVTKFLSSSPHLPPSSYFNA-SGRAYPDVAALSD---GYWVVSNRVPIPWVSGTSASTPVFGGILSLINEHRILSGRPPLGFLNPRLY

PSCP   317  SAISSTPSLVHDVKSGNNGYGGY------GYNAGTGWDYPTGWGSLDIAKLSAYIRSNGFGH                                372
KSCP   308  QL---PPEVFHDITEGNNDIANR----ARIYQAGPGWDPCTGLGSPIGIRLLQALLPSASQAQP                                364
XSCP   337  QAFPTQTSLLHDVTSGNNGYQSH------GYTAATGFDEATGFGSFDIGKLNTYAQANWVTGGGGGST                            398
J-4    308  SLPKA-AQAFRDITQGNNDIAGT----GDVYAAGPGWDPCTGLGSPIANRLIAALQQLQQQDRGQA                                368
PHYS1  345  QAFASDPTIFNDITTGDNKCTEDCCSK-FGYEATKGWDPVTGVGTPVFSKLLAFVQTLP                                      402
PHYS2  291  YT---GGIQFRDITQGSN-----------GHSAKAGLDLVTGEGAWIGSE                                              326
LYS60  332  KMAAERPAAFFDVIKGDNICTEDGCSSGC-QAFMLQKDGTLSLVLVPLFILKCWLMFNLCSKANSPNQITYEGIIHEIACFMKGNKDKKLK    421
hCLN2  314  QQH---GAGLFDVTRGCHESCLDEEVEGQGFCSGPGWDPVTGWGTPNFPALLKTLLNP                                      368
```

1 TGDETSLEYVV
2 YVARVNTEFQKIGLRG
3 TTEDADPSAEPPICKSWT
4 VAKLPNPPPGCAGAILCIGLG

Figure 3. Sequence alignment for selected members of the family of serine-carboxyl peptidases, guided by the crystal structures of sedolisin and kumamolisin.

Abbreviations: **PSCP**: sedolisin (*Pseudomonas* sp. 101 serine-carboxyl peptidase); **KSCP**: kumamolisin, isolated from *Bacillus* novo sp. MN-32 (Murao *et al.*, 1993; Comellas-Bigler *et al.*, 2002); **XSCP**: sedolisin-B, *Xanthomonas* sp. T-22 carboxyl proteinase (Oda *et al.*, 1987); **J-4**: kumamolisin-B, alcohol-resistant peptidase isolated from *Bacillus coagulans* (Shibata *et al.*, 1998), unpublished seqence data of K. Oda; **LYS60**: sedolisin-x*Ap*B, marker for late lysosomes in *Amoeba proteus* (Kwon *et al.*, 1999); **PHYS1**: physarolisin, PPSCP, formerly called physaropepsin, from slime mold *Physarum polycephalum* (Nishii *et al.*, 2003). **PHYS2**: physarolisin-B, formerly called PHP, product of a late-replicating gene from *P. polycephalum* (Benard *et al.*, 1992). **CLN2**: human tripeptidyl-peptidase I (Sleat *et al.*, 1997; Rawlings & Barrett, 1999; Lin *et al.*, 2001). Residues forming the active site are shown in yellow on red background, other conserved residues identified as important for the stability of the enzyme are marked with yellow background, residues conserved in the majority of structures are green, and residues similar in their character are shown in magenta.

yields similar results. In view of their comparatively low sequence identity, the deviations between the structures of sedolisin and kumamolisin are considerably larger (Fig. 4). Comparing the structures of the uninhibited enzymes refined at the highest resolution (1.0 Å for sedolisin and 1.4 Å for kumamolisin), the r.m.s. deviation is 0.8 Å for 247 Cα atom pairs and 1.25 Å for 299 pairs. These differences are due mostly to the insertions and deletions in the amino-acid sequences; the most significant ones are found at the amino termini, in the loop and strand 62–76, turn 124–126, and loops 175–187, 208–214, 222–231, and 252–255 (sedolisin numbering). These differences, however, are limited to surface areas and the structures of the cores of both enzymes are much more similar.

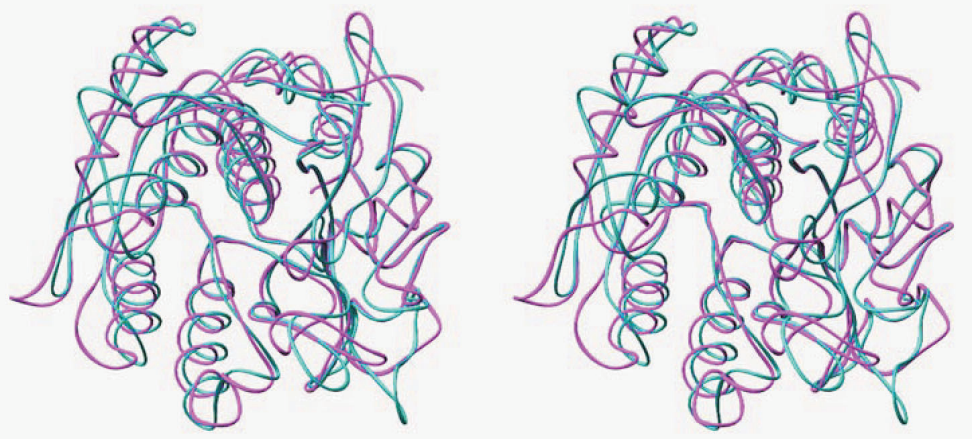

Figure 4. Stereo image of the superposition of the backbone traces of the highest resolution structures of sedolisin (PDB code 1ga6, uninhibited enzyme at 1 Å resolution, magenta) and kumamolisin (PDB code 1gt9, uninhibited enzyme at 1.4 Å resolution, blue).

COMPARISON WITH THE SUBTILISIN FAMILY OF SERINE PEPTIDASES

A comparison of the coordinates of sedolisin with those of proteins corresponding to all known folds, performed with the program DALI (Holm & Sander, 1993), showed unambiguously its structural relationship to subtilisin, a member of the clan SB of serine peptidases (Barrett *et al.*, 1998). The Z-score was 24.4 between the initially-derived sedolisin coordinates and those of the highest-resolution subtilisin structure available (Kuhn *et al.*, 1998) with PDB designation 1gci. The r.m.s. deviation between these two sets of coordinates was 2.5 Å for 238 Cα pairs. The next hit in the comparison was leucine/isoleucine/valine-binding protein (Sack *et al.*, 1989), showing a Z-score of 6.1 and r.m.s. deviation of 4.2 Å for 162 Cα pairs, i.e. a much lower level of similarity. Every major secondary structural element identified in the structure of subtilisin has its counterpart in sedolisin, although the latter enzyme, being significantly larger (372 amino acids *vs.* about 275 for subtilisin), has a number of additional secondary structural elements. Thus the fold of sedolisin can be described as a superset of the well-known subtilisin fold (Robertus *et al.*, 1972). For that reason, the convention adopted for serine-carboxyl peptidases uses identical designations for the strands and helices found in both enzyme families, with primed numbers reserved for elements that are present only in sedolisin, and letters *a* and *b* added for those that are significantly longer in sedolisin.

This fold similarity does not generally extend to the conservation of amino-acid sequence, since the level of identity between the structure-aligned sequences is rather low. We identified 54 residues common to subtilisin and sedolisin (Fig. 5), representing about 20% of the sequence of the former and only 14.5% of the latter. Nevertheless, some of the identical residues are located in the areas crucial to the preservation of the fold, with 14 of them being either glycines or prolines, including the *cis* Pro192 in sedolisin (corresponding to *cis* Pro168 in subtilisin). Many of these residues are also conserved between subtilisin and kumamolisin.

A CALCIUM-BINDING SITE

A prominent Ca^{2+}-binding site has been observed in the structures of both sedolisin (Fig. 6) and kumamolisin. This ion is seen in a virtually identical position in all available structures, both in the uninhibited enzymes and in their inhibitor complexes. Typical Ca^{2+}-binding sites are either octahedral or pseudo-octahedral; in the latter case, two

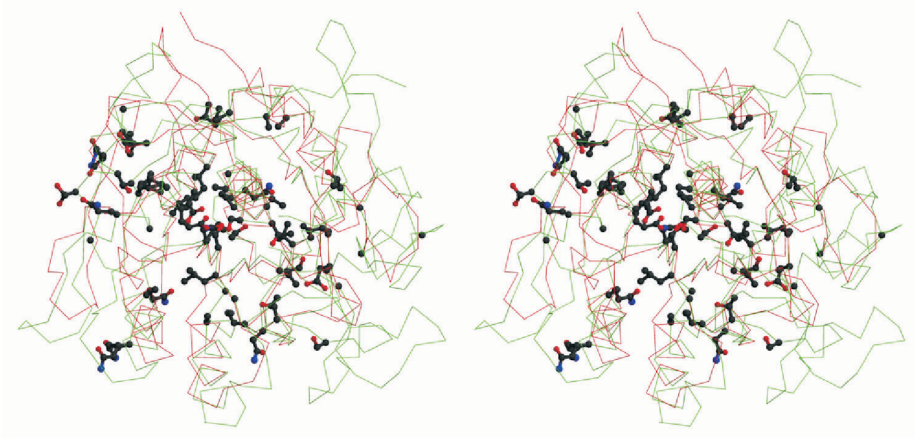

Figure 5. Superposition (in stereo) of the backbones of sedolisin (PDB code 1ga6, 1 Å resolution, green) and *Bacillus lentus* subtilisin (PDB code 1gci, 0.78 Å resolution, magenta).

The side chains of the residues conserved between these two enzymes (Cα only for glycines) are shown in ball-and-stick representation. The side chains shown are those of sedolisin.

oxygens of a carboxylic group approach one of the apices. In sedolisin, each apex of the octahedron is occupied by a single carboxylate oxygen, one derived from Asp328 and the other from Asp348. The base of the octahedron consists of three amide carbonyl groups of residues 329, 344, and 346, and a very clearly delineated water molecule (Wat401). The refined O...Ca^{2+} distances are very similar, their unrestrained values being 2.28–2.32 Å, with only the distance to Wat401 being marginally longer (2.40 Å). The equivalent site in kumamolisin is virtually identical.

Ca^{2+}-binding sites have been previously reported in every structure of enzymes belonging to the subtilisin family. However, the location of the Ca^{2+}-binding site in sedolisin is completely different from either the high-affinity or the low-affinity Ca^{2+}-binding sites in subtilisins Carlsberg, Novo, or their variants (Robertus *et al.*, 1972; Wright *et al.*, 1972; Matthews *et al.*, 1975; Kuhn *et al.*, 1998). The structural role of this site in sedolisin may be to tie the long loop containing residues 328–342 to a short loop, 343– 348, and both of them to the opposite strand around residue 353. The largest part of this region represents a unique insert in sedolisin and has no correspondence in subtilisins. Only the base of the long loop is involved in binding the Ca^{2+} ion, so that the presence of this site is compatible with loops of different length (see below). The importance of the integrity of the Ca^{2+}-binding site was verified in the experiments that showed that modifications or mutations of Asp328 abolish both autoprocessing and the catalytic activity of sedolisin (Oyama *et al.*, 1999). While the postulated catalytic role of Asp328 had to be reconsidered, the experiment referred to above established beyond

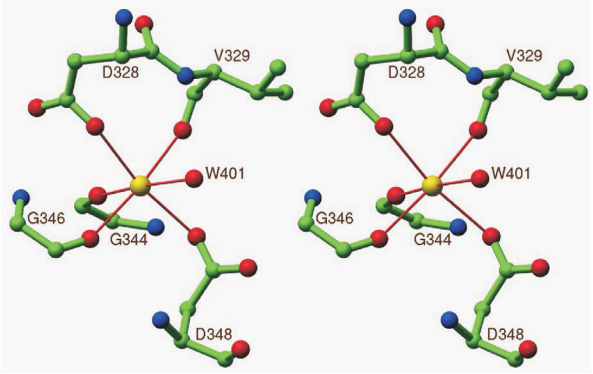

Figure 6. The Ca^{2+}-binding site in sedolisin.

This stereo figure was prepared using the coordinates of the 1.4 Å resolution complex with tyrostatin (PDB code 1kdz), but this site is virtually identical in all structures of sedolisin and kumamolisin.

any doubt the importance of the integrity of the Ca^{2+}-binding site for the activity of the enzymes. It must be stressed, however, that this

site is quite removed from the putative active site (see below), and the exact mode of interaction of these two regions of the protein is not obvious.

Whereas the high-affinity Ca^{2+}-binding site of subtilisin has no counterpart in sedolisin, the low-affinity site has an interesting structural equivalent in the latter enzyme. Asp261, conserved in all known sedolisins, is a topological equivalent of an aspartic acid present in the low-affinity Ca^{2+}-binding site of subtilisin. However, the partner of this aspartate is Arg257 with which it makes an ion pair, thus stabilizing the structure by other means. This ion pair is absolutely conserved in all serine-carboxyl peptidases identified to date, with the modification that a lysine rather than an arginine is found in sedolisin-x*Ap*B (LYS60) and in physarolisin-B (PHP).

The Ca^{2+}-binding sites might not, however, be conserved in the subfamily of much larger enzymes identified in various species of *Thermoplasma* or *Sulfolobus* (Table 1). To date none of these proteins have been isolated, but their sequences differ quite considerably from those of sedolisin or kumamolisin. In addition to the putative conserved catalytic domain, these enzymes also have large C-terminal domains with varying sequences and unknown function. An analysis of the sequences near the C-termini of the catalytic domains of the *Thermoplasma acidophilum* protein Ta0976 (sedolisin-x*Ta*B) or *Sulfolobus solfataricus* sedolisin-x*Ss* failed to show any aspartic acid residues that might be equivalents of Asp328 and Asp348 in sedolisin, the two residues that are primarily responsible for the creation of the Ca^{2+}-binding site. On the other hand, the larger enzymes appear to contain the equivalents of the Arg257/Asp261 ion pair that substitutes for the low-affinity Ca^{2+}-binding site in subtilisin.

THE ACTIVE SITE OF SEDOLISIN

The active site of sedolisin (Fig. 7) can be identified on the basis of several criteria. With the fold of this enzyme corresponding to that of subtilisin, and with Ser287 in sedolisin equivalent to Ser221 in subtilisin both in the primary and in tertiary structure, this serine is an obvious candidate for the primary catalytic residue in the active site. This serine is also covalently bound to inhibitors in some of the structures of sedolisin and kumamolisin. Five different inhibitors have been cocrystallized with sedolisin, and two of them also with kumamolisin. All of these inhibitors are peptidic in nature, but with an aldehyde group on their "C terminus". As expected on the basis of the structures of other

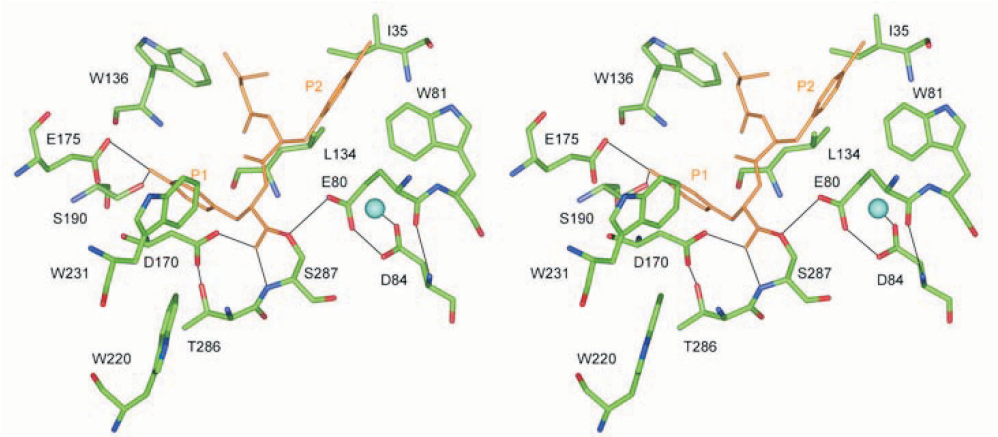

Figure 7. Residues forming the active site of sedolisin.

The stereo figure shows the inhibitor pseudo-iodotyrostatin bound in the active site of sedolisin, together with selected residues in its vicinity. Only the principal orientations of the residues are shown, and hydrogen bonds are marked in thin black lines. Reprinted from (Wlodawer *et al.*, 2001a).

similar inhibitors of serine proteinases, the aldehyde function of the inhibitor makes an unambiguous hemiacetal linkage to Oγ of Ser287 (Fig. 7), and thus the mode of interaction between the enzyme and the inhibitor is very clear.

Table 1. Database entries for selected identified or putative sedolisins (serine-carboxyl proteinases), based on the similarities of their sequences.

Alternative names are shown in the second column; acceptable alternatives are shown in regular script, while obsolete names that should no longer be used are italicized. The enzymes that have not been characterized in detail and thus cannot be assigned unambiguously to a particular subfamily are given provisional names including the character "x" and an identifier of the species from which they are derived. These names will have to be adjusted in the future. The table is divided into three sections: first are proteins that seem to consist of a prosegment and a catalytic domain; second are proteins with C-terminal extensions of unknown function; third are members of the tripeptidyl-peptidase I subfamily with highly conserved sequences. Some of the enzymes belonging to the first and third groups have been well-characterized, while none of the second group has been isolated or otherwise directly studied to date.

Proposed name	Other name(s)	Organism	Database identifier	Size (catalytic domain in parentheses)
sedolisin	PSCP; Pseudomonas serine-carboxyl proteinase; *pseudomonapepsin*	*Pseudomonas* sp. 101 (bacterium)	PICP_PSESR gi: 1172505 gi: 1073066, 688417	587 (372) (N-terminal propeptide)
sedolisin-B	XSCP; Xanthomonas serine-carboxyl proteinase; *xanthomonapepsin*	*Xanthomonas* (bacterium)	XANP_XANS2 gi: 9979355	827 (398) (propeptides are present on both the N- and C-termini)
sedolisin-x*Bf*		*Burkholderia fungorum* (bacterium)	ZP_00030505 gi: 22985385	519
kumamolisin	KSCP *kumamolysin*	*Bacillus* novo sp. MN-32 (thermophilic bacterium)	BAB85637 gi: 25137473	552 (364)
kumamolisin-B	*J-4 carboxyl proteinase*	*Bacillus coagulans* (thermophilic bacterium)	unpublished	559 (368)
kumamolisin-*As*	ScpA	*Alicyclobacillus sendaiensis* (bacterium)	BAC41257 gi: 25900987	553 (364)
aorsin		*Aspergillus oryzae* (fungus)	BAB97387 gi: 21321299	652
physarolisin	*Physarum polycephalum* serine-carboxyl proteinase; *physaropepsin*	*Physarum polycephalum* (mycetozoa)	AF502290 gi: 323087 gi: 895878	575 (216+186)
physarolisin-B	*Physarum polycephalum PHP*	*Physarum polycephalum* (mycetozoa)	X64708 gi: 895877	357
sedolisin-x*Dd*	*Dipeptidyl aminopeptidase; pepstatin-insensitive carboxyl proteinase 2*	*Dictyostelium discoideum* (mycetozoa)	AAL14225 gi: 16209613	700
sedolisin-x*Ap*A	*Pepstatin-insensitive carboxyl proteinase 1; LYS45*	*Amoeba proteus* (amoeba)	AAD37351 gi: 5002190	420
sedolisin-x*Ap*B	*Pepstatin-insensitive carboxyl proteinase 2; LYS60*	*Amoeba proteus* (amoeba)	AAD37352 gi: 5002192	530
Enzymes with very long extensions at the C-termini; none has been isolated or characterized directly				
sedolisin-x*Ta*A	*Ta0376*	*Thermoplasma acidophilum* (archea)	NP_393856 gi: 16081505	1379
sedolisin-x*Ta*B	*Ta0976*	*Thermoplasma acidophilum* (archea)	NP_394436 gi: 16082015	1090
sedolisin-x*Ss*	*Serine protease from subtilase family*	*Sulfolobus solfataricus* (archea)	NP_343890 gi: 15899285	1308

Table 1 continued on the next page

Table 1. Continued

Members of the highly conserved tripeptidyl-peptidase I subfamily				
tripeptidyl-peptidase I	CLN2; TPP I	Mouse (mammal)	TPP1_MOUSE gi: 12644085	562 (368)
tripeptidyl-peptidase I	CLN2; TPP I	Rat (mammal)	NP_112647 gi: 13786206	563 (369)
tripeptidyl-peptidase I	CLN2; TPP I	Human (mammal)	TPP1_HUMAN gi: 6175068	563 (369)
tripeptidyl-peptidase I	CLN2; TPP I	Dog (mammal)	TPP1_CANFA gi: 20140923	563 (369)
tripeptidyl-peptidase I	CLN2; TPP I	Macaque (mammal)	BAC20587 gi: 23574719	563 (369)
tripeptidyl-peptidase I	CLN2; TPP I	Cow (mammal)	AF491290_1 gi: 19919858	563 (369)
sedolisin-TPP		Fugu (fish)	sinfrup00000077297	531 (the length of the propart is not certain)
sedolisin-TPP		Zebrafish (fish)	contig wz4596, with manual corrections	557 (based on a hypothetical sequence, subject to further modifications)

Other residues involved in supporting the catalytic activity of serine-carboxyl peptidases can be identified based on the structures of sedolisin and kumamolisin, as well as on the basis of the biochemical data available for these and other family members. The residue that can directly interact with Ser287 is Glu80, and its side chain, in turn, interacts with Asp84 (Fig. 7). The distance between the carboxylate oxygens of the latter two residues does not vary significantly among all experimental structures and is about 2.60 Å. This short distance between the two adjacent carboxylates is similar to the separation between the side chains of two aspartic acids in the inhibited forms of pepsin-like aspartic peptidases (Davies, 1990), although the detailed geometry of the interaction is quite different, since in sedolisin these groups are not coplanar. However, the close distance between the carboxylates indicates that a proton must be shared between these residues in all the structures, since if both groups were to be charged, the electrostatic repulsion would prevent their close approach. Both Glu80 and Asp84 originate from helix h3, a structural element involved in creating a left-handed crossover (see above). Such crossovers are usually found in proteins in areas important for their activity (Miller *et al.*, 1993), and this seems to be the case for serine-carboxyl peptidases as well.

To reveal the essential groups involved in the catalytic action of sedolisin, the pH-dependence of the hydrolysis of an artificial substrate, Ser-Pro-Ala-Lys-Phe*(NO_2)Phe-Arg-Leu (*: cleavage point) was studied. The pK_1 and pK_2 values for the enzyme-substrate complex were found to be 2.97 and 4.92, respectively (Oda *et al.*, 1992). The role of various side chains in the binding and catalytic activity of sedolisin and sedolisin-B have been investigated by chemical modification (Ito *et al.*, 1999) and by site-directed mutagenesis (Oyama *et al.*, 1999). Chemical modification of sedolisin by carbodiimide indicated that Asp140 and Glu222 were important in substrate binding. However, a comparison of the sequence of sedolisin with sedolisin-B revealed that Asp140 is not conserved and, thus, cannot account for the general properties of the family of enzymes.

Oyama *et al.* (1999) utilized site-directed mutagenesis to replace eight residues that were conserved in both sedolisin and sedolisin-B by alanine. Residues 84, 170, and 328 in sedo-

lisin (79, 169, and 348 in sedolisin-B) were found to be essential for either the self-activation of the proenzyme or for the cleavage of peptide substrates. Replacement of residue 84 led to a 10^4-fold decrease in the catalytic activity of sedolisin. Residues 124 and 217 in sedolisin (123 and 230 in sedolisin-B) were shown to be non-essential as the Ala derivatives retained 60% and 70% of the wild-type activity, respectively. In addition, residues 225 and 265 in sedolisin (246 and 285 in sedolisin-B) were shown to be dispensable, as the alanine derivatives were fully active. Residue 222 of sedolisin (235 in sedolisin-B) yielded a mixed result upon replacement, as the Ala derivative in sedolisin could not self-activate, while the substitution in sedolisin-B reduced the activity only 6-fold. Examination of the inhibitor complexes reveals that Glu222 of sedolisin is close to the S2′ subsite of the active site cleft. For substrates with Arg in the P2′ position, including the substrate used for assaying the mutants, changing Glu to Ala will remove a potential stabilizing interaction in the S2′ subsite, thus accounting for the observed effect. In the chemical modification by carbodiimide, the additional steric bulk will clearly change the S2′ subsite in the case of the Glu222 derivative. Thus, both types of experiments point to a role for Glu222 in substrate binding. The conclusion of the study by Oyama *et al.* (1999) was that residues 84, 170, and 328 of sedolisin were important to the catalytic activity. As shown by the crystal structures discussed above, we now know that Asp328 is involved in creating a Ca^{2+}-binding site, an important structural element of the protein.

Protonation states of the active site residues have been assigned for the uninhibited form of sedolisin based on the lengths of the C–O bonds that differ when the oxygen is protonated or unprotonated (Wlodawer *et al.*, 2001b). A chain of proton donors originates with a bound water molecule interacting with protonated Asp84; in turn, this residue interacts with protonated Glu80, which donates a proton to Ser287. A water molecule accepts protons from both Ser287 and protonated Asp170, the residue forming the oxyanion hole (see below). Obviously, protonation of the residues will change during catalysis, but the resolution of the available experimental structures of inhibitor complexes is not sufficient to justify placement of hydrogen atoms. In addition, in some of the inhibitor complexes of sedolisin, the side chain of Glu80 assumes double conformation that may be related to the changes in its protonation state.

Other residues may extend the interactions of the catalytic triad even further, forming a secondary network of hydrogen bonds. In sedolisin, the other carboxylate oxygen of Asp84 interacts with the side chain of Asn131, most likely with its amide, Nε2. This assignment is based on the most likely status of hydrogen bonds, in which Nε2 of Asn131 would be a donor, while Oδ2 of Asp84, almost certainly unprotonated, would be an acceptor. The oxygen Oδ1 of Asn131 would, in turn, be an acceptor of an H-bond from Ser290, while Ser165 would donate a proton to the latter. While the interactions described above are observed in all structures of sedolisin, this secondary network of H-bonds might only exist at neutral or higher pH, since the distance between Oδ2 of Asp84 and Nδ2 of Asn131 exceeds 3.65 Å in the structures obtained at lower pH. Since sedolisin is only active at very acidic pH (see below), and since the other residues of the secondary network are not necessarily conserved in other related enzymes, the importance of this network is not obvious.

Two other residues, Glu32 and Trp129, appear to extend the hydrogen-bonded network in the active site of kumamolisin. This pair of residues is also present in kumamolisin-B (J-4) and sedolisin-x*Ap*B, but either one or both of them are missing in the other members of the family. This may indicate that the influence of the extended network is limited to only selected serine-carboxyl peptidases.

At least one other residue can be unambiguously defined as being crucial for the catalytic

activity of this family of enzymes. This residue is Asp170, structurally equivalent to Asn155 in subtilisin where its side chain creates part of the oxyanion hole, stabilizing the tetrahedral intermediate of the reaction. The orientation of the side chain relative to the covalently-bound inhibitor is the same in the inhibitor complexes of the members of both enzyme families. It is quite obvious that Asp170 would need to be protonated in order to serve its function of creating the oxyanion hole, and this may be one of the most important features responsible for the observation that the activity of serine-carboxyl peptidases is maximal at low pH.

The unique features of the active site of sedolisin and related enzymes were used to create a more consistent and unambiguous nomenclature presented below. It is now clear that the active site of serine-carboxyl peptidases contains a unique catalytic triad Ser-Glu-Asp (SED in the single-letter code), not seen in any other families of peptidases (nor, for that matter, in any other enzymes). Since the presence of this catalytic triad is the defining feature of the family, we decided to rename PSCP, the first of these enzymes to become characterized structurally, as **SED**olisin. The nomenclature for other members of the family (Table 1) follows this choice, except where acceptable names, such as kumamolisin or aorsin, have already been established in the literature. Precise assignment of those enzymes that are known only as database entries or that were characterized in only a preliminary fashion has to await their isolation and full characterization, since knowledge of their enzymatic properties is needed for assignment to the subfamilies. At this time, an extension "x" following the name sedolisin indicates enzymes that cannot be unambiguously assigned as equivalent to their better-characterized peers.

THE MODE OF INHIBITOR BINDING

Crystal structures of sedolisin have been described for the complexes with five different inhibitors (pseudo-tyrostatin, tyrostatin, chymostatin, AcIPF, AcIAF (Fig. 8); for chemical formulas, see Fig. 1 in Wlodawer *et al.*, 2001b). The latter two inhibitors were initially synthesized to specifically inhibit kumamolisin, but were later found to be active on both enzymes, although only at a sub-micromolar level for sedolisin. The linkage between the C terminus of the inhibitors and the Oγ oxygen of Ser287 is through a covalent bond to form a (reversible) hemiacetal, with the *S*

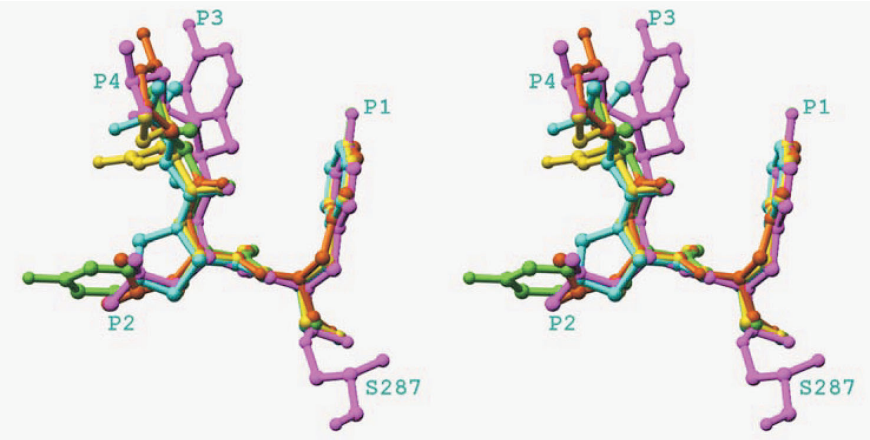

Figure 8. Stereo representation of the superposition of the atomic coordinates of five inhibitors of sedolisin bound in the active site of the enzyme.

Reprinted from Wlodawer *et al.* (2001b).

stereochemistry of the carbon atom bound to the serine. An increase in pH leads to a loss of the hydrogen from the –OH group of the hemiacetal and expulsion of the Ser–O–moiety to reform the original aldehyde. Similar linkages have been previously described for complexes of chymostatin and two other serine proteinases, *Streptomyces griseus* proteinase A (Delbaere & Brayer, 1985) and wheat serine carboxypeptidase II (Bullock *et al.*, 1996). The stereochemistry of the active site of sedolisin is opposite to that of carboxypeptidase, in excellent agreement with the interpretation provided by Bullock *et al.* (1996), who noted that the arrangement of the active site residues in carboxypeptidase corresponds to a mirror image of the arrangement in subtilisin.

The availability of the structures of a number of inhibitor complexes allows us to delineate subsites S1–S4 (Schechter & Berger, 1967) of the substrate-binding site. In all inhibitors of sedolisin that have been studied to date, the P1 residue is either tyrosine or phenylalanine. The orientation of its side chain is virtually the same in all complexes, being wedged into a pocket created by the side chain of Arg179, the main chain of residues 133–136, and the main and side chains of residues 167–170. In the inhibitors containing Tyr, its Oη atom makes excellent hydrogen bonds with Oε2 of Glu175 and with Oγ of Ser190. The structures with the P1 Phe side chain do not contain any extra water molecule(s) to compensate for the absence of Oη in the side chain, since the pocket is too small to accommodate a non-covalent oxygen. It thus appears that tyrosine is the natural and best substituent of the S1 subsite of sedolisin. Arg179 has no equivalent in kumamolisin, and the S1 subsite of this enzyme is quite open and more accessible (Fig. 9).

A number of different side chains are present in the P2 position of the inhibitors, but they are all structurally superimposable (Fig. 8). These residues include iodoPhe in pseudo-iodotyrostatin (and Tyr in pseudo-tyrostatin), Leu in tyrostatin and in chymostatin, Pro in AcIPF, and Ala in AcIAF. The active site area occupied by these residues is rather open and is bounded on one side by the side chains of Ile35, Asp74, Gln76, Trp81, and Glu80, while it is exposed to the solvent on the other side. The iodine of iodoPhe and the hydroxyl of Tyr interact with the carboxylate of Asp74. Since the P2 side chain of AcIAF consists of only a single methyl group, the rest of the S2 pocket is filled by a glycerol molecule that appears to have the right combination of hydrophobic and hydrophilic groups to provide optimum interactions (Fig. 9A). Trp129, unique to kumamolisin, forms one of the walls of the S2 subsite in that enzyme.

The main chain of all the inhibitors discussed here accepts a single hydrogen bond through the P3 carbonyl oxygen from the main chain amide of Gly135 of sedolisin, while a second bond is made through S3 N and 135 O in the complexes of tyrostatin and AcIPF. The space surrounding the P3 Tyr and Ile of the two latter inhibitors most likely defines the S3 area of the enzyme. That area, however, can hardly be called a pocket, since it is almost completely open. The only part of the enzyme that is in contact with the inhibitor is Arg179, but its interactions with the two types of side chains are different. The orientation of the side chain of Arg179 is almost identical in all of the complexes with the exception of tyrostatin. Although it comes into contact with the Ile of AcIPF, it is not pushed from its usual location. However, the larger bulk of P3 Tyr in tyrostatin forces Arg179 to reorient.

The nominal positions of the P3 and P4 residues are reversed between the complexes of AcIPF and AcIAF in sedolisin, with the former arrangement corresponding to the one observed in the kumamolisin complexes. These side chains make very weak hydrophobic interactions with the side chains of Ile35, Leu114, Leu134, and Trp136 (closest interatomic distances about 4 Å), but again, it is not really possible to describe an actual S4 pocket. It is clear that a variety of different

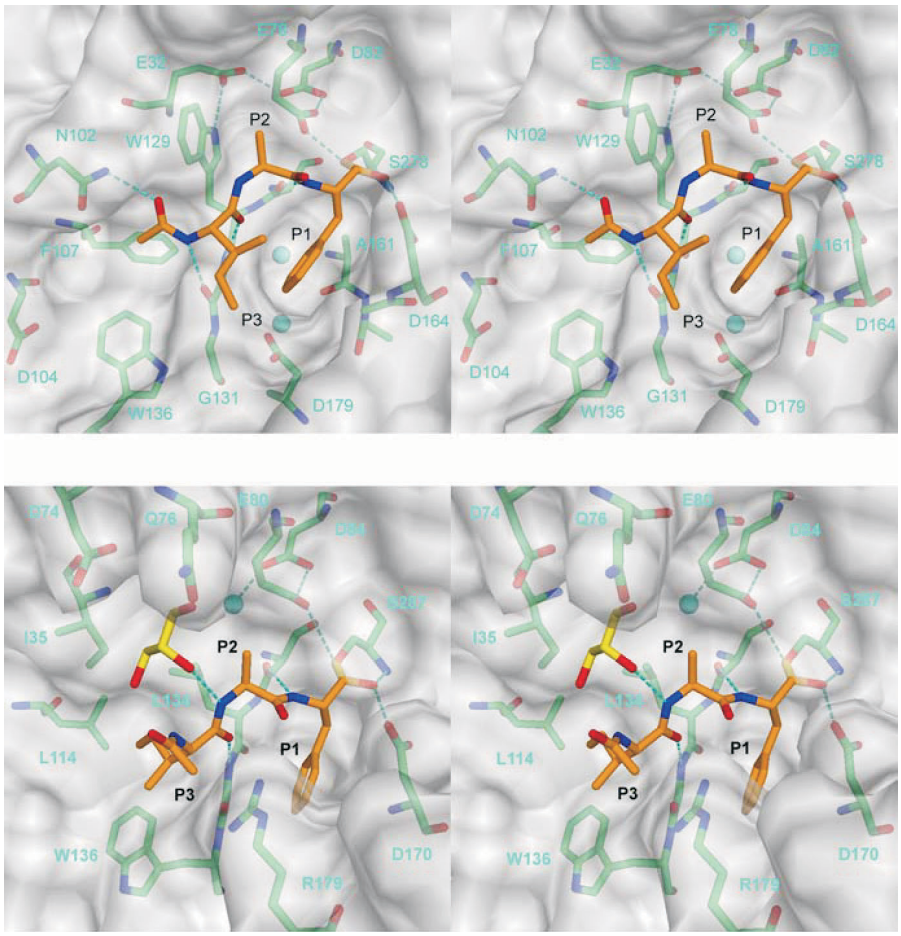

Figure 9. Comparison (in stereo) of the active sites of sedolisin (PDB code 1kdv, top) and kumamolisin (PDB code 1gtj, bottom) complexed with the same inhibitor, AcIAF (Wlodawer *et al.*, 2001b; Comellas-Bigler *et al.*, 2002).

The surfaces of the active sites of both enzymes are semi-transparent, while the covalently-bound inhibitors are gold. The glycerol molecule found in the structure of sedolisin is yellow. Figure prepared by M. Comellas-Bigler using the program DINO (http://www.bioz.unibas.ch/~xray/dino).

side chains could be easily accommodated in this area, and thus we do not expect that this part of the inhibitor should contribute much to its specificity (or that a corresponding residue in the substrate would contribute in a significant way to the specificity of the enzyme). In a series of octapeptide substrates substituted in P4, sedolisin showed the following preferences: Pro, Leu, Ala > Ser > Asp, Arg (Ito *et al.*, 1996).

The N-terminal isovaleryl group of pseudotyrostatin does not make any clear contacts with the enzyme. The sole orienting interaction of this residue is a hydrogen bond between the carbonyl of its main chain peptide and the amide nitrogen of residue 135. By contrast, the terminal acetyl group of AcIPF and AcIAF makes excellent hydrogen bonds with the side chain of Asn102 in kumamolisin, most likely contributing significant binding energy for these inhibitors (Fig. 9B).

SUBSTRATE SPECIFICITY

The substrate specificity of peptidases can be studied by either following the cleavage of individual peptides, or by utilization of peptide libraries. Little is known about natural substrates of bacterial sedolisins, with the exception of kumamolisin-*As* that appears to function as a collagenase (Tsuruoka *et al.*,

2003). That study also established some collagenase-like activity for kumamolisin (89% identical to kumamolisin-*As*), but it is not clear whether this is the principal role of that enzyme. Only limited specificity data are available for other sedolisins, mostly for sedolisin and sedolisin-B (Narutaki *et al.*, 1999). The crystal structures of the inhibitor complexes (discussed above) are also good guides to investigating the specificity. From the inhibitor-bound enzyme structure of sedolisin, we can consider the likely preferences of only the S2, S1, and S2′ subsites. The iodoPhe side chain in the P2 position of the covalent inhibitor is surrounded by the side chains of Ile35, Gln76, Gly77, Glu80, Trp81, and Leu-134. Close distances include Cδ2 of Leu134 to Cδ1 of iodoPhe (3.43 Å), Cδ1 of Ile35 to Cε1 of iodoPhe (4.24 Å), side of Trp81 to Cε1 of iodoPhe (3.40 Å), Cα of Gly77 to Cε2 of iodoPhe (3.78 Å), Gln76 Cβ to Cε2 iodoPhe (4.66 Å), Cβ of Glu80 to Cγ of iodoPhe (4.32 Å). In a limited series of substitutions in a chromogenic substrate of eight amino acids, leucine in the P2 position provided the best specificity constant (k_{cat}/K_m) for sedolisin. Phe, Tyr, or Trp were not included in this series, as their substitution could have created new cleavage sites in the peptide. Interestingly, the Glu substitution was acceptable with a specificity constant 45% of that of the Leu-substituted peptide. Likewise, Thr (34%), Asp (31%), and Arg (19%) were also acceptable. These observations may indicate that the Asp residue at position 74 may buffer the hydrophobicity of the S2 subsite in sedolisin.

Among the residues that contribute to the S2 subsite in sedolisin-B, the same amino acids as in sedolisin are present in positions 74, 80, 81, and 134 (sedolisin numbering). Interestingly, Ile35 in sedolisin is replaced by Trp in sedolisin-B, while residue 76 appears to be deleted in the latter. As a result, sedolisin-B accepts P2 residues in the following order in the series of peptides studied, Glu (set at 100%), Leu (82%), Pro (77%), Asp (74%), Thr (70%), Asn (70%), and Val (62%). As all assays were done at pH 3.5, it is possible that Asp74 of the enzyme is capable of forming hydrogen bonds with acidic or polar amino acids.

Recent studies (Oda *et al.*, unpublished) have revealed a broad P1 specificity of sedolisin. The amino acids preferred were in the following order: Glu (100%), Asp, (92%), Gly (about 77%), Asn (77%), Phe (66%). The pocket surrounding the P1 Tyr residue of the covalently bound inhibitor includes side chains from amino acids Glu171, Glu175, Arg179, and Ser190 of the enzyme, consistent with the broad specificity observed. The close fit of the Tyr side chain in the S1 pocket indicates why Trp is not a good substitution in P1 in a peptide substrate. Narutaki *et al.* (1999) reported that substitution in the P2′ position of Lys-Pro-Ile-Glu-Phe*Nph-P2′-Leu resulted in the following order of the k_{cat}/K_m values: Asp, Glu > Arg, Lys > Ala> Leu > Ser > Asn. The S2′ subsite, as revealed in the non-covalent complex, contains both hydrophobic (Trp220 and 231) as well as hydrophilic residues (Glu222, Gln268, Gln281, and Gln282). Thus, it is reasonable to expect that a variety of amino acids would be acceptable at P2′.

On the other hand, human CLN2 could not cleave the octapeptide substrates used to analyze sedolisin and sedolisin-B. Therefore, the molecular size of substrates was explored and it was found that Ala-Arg-Phe*Nph-Arg-Leu (k_{cat}/K_m 2.94 μM^{-1} s^{-1}) was the best substrate among 11 tested (Oda *et al.*, unpublished data). The specificity constant measured was 40 times higher than that of Ala-Ala-Phe-MCA, the conventional substrate for CLN2. These results suggest that the active cleft of CLN2 is smaller than those of sedolisin and sedolisin-B on the amino terminal side. Therefore, whereas sedolisin provides a binding surface that can accommodate 7 or 8 amino acids, the binding cleft for CLN2 is composed of, at most, six (S3–S3′) subsites. On the basis of the model of CLN2 described earlier, residue Asp132 in CLN2 replaces Gly131 in the S3 subsite of kumamolisin. The carboxyl group of Asp132 would extend out

into the active site cleft, accurately placed to anchor the free N terminus of a bound substrate in the enzyme-substrate complex (Comellas-Bigler *et al.,* 2002), thus limiting the substrate size to three residues before the cleavage site.

SEQUENCE COMPARISONS AND MODELING OF OTHER SERINE -CARBOXYL PEPTIDASES

A number of potential members of the family of serine-carboxyl peptidases have been identified in the last 15 years in a variety of organisms. Some of them have been purified and subjected to detailed biochemical and enzymatic studies, while the existence of others has been postulated based only on the available gene sequences, some containing only parts of the enzymes. Nevertheless, it is clear that these enzymes are present in a wide variety of organisms (Table 1), and that they can be grouped according to the similarity of their sequences. The enzyme most similar to sedolisin is sedolisin-B (XSCP) from *Xanthomonas* sp. T22 (Oda *et al.,* 1987; Oyama *et al.,* 1999), with 50.5% sequence identity. Three enzymes that share very considerable sequence identity are kumamolisin, kumamolisin-*As* (ScpA), and J-4 (Murao *et al.,* 1993; Shibata *et al.,* 1998; Oyama *et al.,* 2002; Tsuruoka *et al.,* 2003). Thus we propose to rename the latter enzyme kumamolisin-B.

Enzymes with similar sequences have also been found in a number of other bacterial genomes, although so far they have not been characterized in detail (Table 1). The genomic sequence of a thermoacidophilic archeobacterium, *Thermoplasma acidophilum* (Ruepp *et al.,* 2000), lists two proteins with sedolisin-like sequences among 23 identified proteinases, indicating the importance of this new enzyme subfamily. Two related enzymes, LYS45 (sedolisin-x*Ap*A) and LYS60 (sedolisin-x*Ap*B), were identified in *Amoeba proteus*, the former showing 17.5% identity and 41% similarity to sedolisin, while the latter's identity and similarity to human CLN2 are as high as 24 and 52% (Kwon *et al.,* 1999). These proteins, however, have not been isolated and purified, so the exact locations of the processing sites can only be inferred. Two related enzymes have also been reported in the genome of the slime mold *Physarum polycephalum* (Benard *et al.,* 1992; Nishii *et al.,* 2003). Mammalian enzymes homologous to human CLN2 (Sleat *et al.,* 1997; Lin *et al.,* 2001) form a closely related subfamily of sedolisins (see below). CLN2-like enzymes are also found in fugu (puffer fish) and zebrafish (see Table 1). However, the sequence for the zebrafish enzyme is only putative, since it is based on manually corrected sequence of the contig wz4596 in the zebrafish EST database found at http://fisher.wustl.edu/fish_lab.

Amino-acid sequences of the members of this family show the presence of N-terminal propeptides, catalytic domains, and sometimes quite large C-terminal domains of unknown function. Sedolisin-B also contains a C-terminal propeptide that is removed upon activation of the enzyme (Oyama *et al.,* 1999). An alignment of the sequences of the catalytic domains of several of these enzymes, manually adjusted to reflect the structural features of sedolisin and kumamolisin, is shown in Fig. 3. The catalytic triad Glu-Asp-Ser is present in all the enzymes, with the similarity extending to the adjacent residues as well. The sequence surrounding Ser287 is well conserved (and also similar to a corresponding sequence in subtilisins), with the exception that Gly284 is substituted by a serine in CLN2 and by an aspartic acid in sedolisin-x*Pp*A. Interestingly, the corresponding residue is asparagine in subtilisin, while the following two glycines are present in all of the compared enzymes. While the leucine preceding Asp84 is strictly conserved, other residues in the vicinity of Glu80/Asp84 are more varied. Another residue that appears to be crucial for the activity of sedolisin, namely Asp170, is also present in all other members of the family, together with

neighboring Gly169 and Gly172. It is also likely that the location of the Ca^{2+}-binding loop is conserved throughout the sedolisin family, as the residues creating the base of the loop are all conserved, although the length of the loop itself varies significantly.

An interesting feature conserved in all the compared sequences is the presence of a buried ion pair, consisting of Asp261 (strictly conserved) and Arg257 (Lys in sedolisin-x*Ap*B and physarolisin-B (PHP). The former residue is also conserved in subtilisin, where Asp197 is involved in the creation of the low-affinity Ca^{2+}-binding site. It is thus likely that the role of this ion pair in sedolisin is similar to the role of a Ca^{2+}-binding site in subtilisin, namely enhancing enzyme stability. This region contains a number of conserved residues, including Gln247, Asp261, Trp246, and Tyr316. All of these interacting residues are generally conserved, although with some exceptions, such as a substitution of Trp246 by a tyrosine in CLN2, sedolisin-x*Pp*A, and sedolisin-x*Pp*B, and unclear situation in sedolisin-x*Ap*B, where the sequence alignment in the areas containing some of these residues is ambiguous.

CLN2 is the member of the sedolisin family that is of particular interest because of its involvement in a human disease (Sleat *et al.*, 1997). This enzyme was shown to be necessary to prevent late infantile neuronal ceroid lipofuscinosis (Batten disease), a rare but fatal hereditary neurodegenerative disorder. CLN2 was later identified as being equivalent to tripeptidyl-peptidase I (Rawlings & Barrett, 1999). Very highly homologous enzymes have also been identified in macaque, mouse, rat, dog, and cow (Fig. 10). For the full-length enzymes from these six species (563 amino acids, including the catalytic domain and the prosegments), sequence identity exceeds 81%, whereas the similarity is over 92%, with only a single-residue deletion in the mouse enzyme. A pairwise comparison of the human and mouse enzymes yields 88% identity and 94% similarity, considerably higher than the median 78.5% identity reported for the mouse-human orthologs (Waterston *et al.*, 2002). Thus CLN2 appears to be a highly conserved enzyme.

A slightly more distant, although clearly related, enzyme has been recently identified in the genomes of fugu and zebrafish. We propose to name these enzymes sedolisin-TPP, to emphasize their similarity to the well-characterized subfamily of tripeptidyl-peptidase I (CLN2). The alignment shown in Fig. 10 for zebrafish sedolisin-TPP is based on manual adjustment of the sequence assembled in this contig, which seems to suffer from several frame shifts. However, the similarity between the fugu and the corrected zebrafish sedolisins-TPP is so high that the postulated changes in the sequence are very likely correct. The high level of identity is not limited to the mature enzyme, but also extends to most of the propeptide, with the differences at the N termini reflecting either sequencing errors or true deviation in the predicted signal peptides. The presence of highly conserved CLN2-like enzymes not only in mammals but also in two fish species may indicate that these enzymes might be universally present in the vertebrates and that their important role in humans (Sleat *et al.*, 1997) and mice (Katz & Johnson, 2001) might be a more general feature.

The medical importance of CLN2 and a lack of its crystal structure led to attempts to model its structure. A model that assumed this enzyme to be membrane-bound and that identified the mature sequence 271–294 as a putative transmembrane segment (Orry & Wallace, 1999) is unlikely to be correct. Based on the similarity of the sequences of sedolisin and other serine-carboxyl peptidases and CLN2, we have created an energy-minimized model of human CLN2 (Wlodawer *et al.*, unpublished) with an r.m.s. deviation between the corresponding Cα coordinates of sedolisin and CLN2 of about 1.75 Å, not much larger than the experimental difference between sedolisin and kumamolisin. This model indi-

cates a likely presence of a disulfide bridge between Cys327 and Cys342, providing an additional constraint on the conformation of the Ca^{2+}-binding loop (see above). That putative disulfide is strictly conserved in all known sequences of CLN2-like enzymes, whether from mammals or from fish (Fig. 10). The ability to predict the presence or absence of disulfide bonds in other family members indicates that the assumption of their structural similarity is most likely justified. However, although the model of CLN2 seems to agree with many structural aspects of this family of proteins, it still does not explain the reasons for the tripeptidase activity of this subfamily.

CONCLUSIONS

Although the peptidases belonging to the family S53 have been studied for a number of years, their unambiguous placement among other proteolytic enzymes had to await crystal structure determination. In the meantime, the nomenclature for these enzymes became very confusing, since some of their names included the root "pepsin," implying similarity to the family of aspartic peptidases. We now know that although these enzymes are most active at low pH and their active sites contain at least three carboxylate-bearing residues, they are structurally and mechanistically related to serine peptidases from the subtilisin family. We hope that the nomenclature introduced in this review and adopted by the MEROPS database will become universally accepted and will make it easier to place new members of the family. It is clear that in view of the medical importance of at least some of these enzymes, both structural and biochemical research on sedolisins will continue. The structure-based comparisons of the amino -acid sequences of serine-carboxyl peptidases, presented above, indicate possible existence of subfamilies that might be related to other sedolisins, yet be partially different; for example, by not requiring Ca^{2+} cations for their activity. However, at this stage these are only speculations that lead to new questions about this fascinating family of peptidases. Only continued research on the structural and enzymatic properties of sedolisins, coupled with a search for their natural substrates and the establishment of their biological roles with techniques such as gene knockouts will allow us to provide more complete description of the function and importance of this family of peptidases.

We would like to thank Dr. S. L. Johnson (Washington University, St. Louis, U.S.A.) for help in querying the zebrafish EST sequence database. We are indebted to Ms. M. Comellas-Bigler and Dr. W. Bode (Max-Planck-Institut, Munich, Germany) for discussing with us the structures of kumamolisin and for preparation of Fig. 9. Extensive discussions with Dr. A. Barrett (Wellcome Trust Sanger Institute, Hinxton, U.K.) regarding peptidase nomenclature are gratefully acknowledged. Thoughtful comments of the referee and the technical editor helped in clarifying the presentation. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does the mention of trade names, commercial products or organizations imply endorsement by the U. S. Government.

Figure 10. Sequence comparisons of the CLN2-like enzymes from mammals and fishes.

Residues are colored using the same scheme as in Fig. 3. The N terminus of the mature enzyme (experimentally determined for the mammalian enzymes and predicted for the fish) is marked with black triangles.

```
  Human MGLQACLLGLFALILSGKCSYSPEPDQRRTLPPGWVSLGRADPEEELSLTFALRQQNVER
Macaque MGLQACLLGLFALILSGKCSYSPEPDQRRTLPPGWVSLGRADPEEELSLTFALRQQNLER
  Mouse MGLQARLLGLLALVIAGKCTYNPEPDQRWMLPPGWVSLGRVDPEEELSLTFALKQRNLER
    Rat MGLQARFLGLLALVIAGKCTHSPEPDQRWMLPPGWVSLGRVDPEEELSLTFALKQQNLDR
    Cow MGLQACLLGLFALILSGKCSYSPEPDQRRTLPPGWVSLGRADPEEELSLTFALRQQNVER
    Dog MRLRTCLLGLLALCVASKCSYSPEPDQQRTLPPGWVSLGRVDSEEELSLTFALRQQNVER
   Fugu ------------------------------IPKDWTNVGRVDAKEELQLTFALKQQNVDL
Zebrafish MRVAVFVLSFIWLVNG----ELLEADQDAVVPGDWTFLGRVGPLEEVELTFALKQQNVSK

  Human LSELVQAVSDPSSPQYGKYLTLENVADLVRPSPLTLHTVQKWLLAAGAQKCHSVITQDFL
Macaque LSELVQAVSDPNSPQYGKYLTLENVADLVRPSPLTLHMVQKWLLAAGAQKCHSVITQDFL
  Mouse LSELVQAVSDPSSPQYGKYLTLEDVAELVQPSPLTLLTVQKWLSAAGARNCDSVTTQDFL
    Rat LSELVQAVSDPSSPRYGKYLTLEDVAELVQPSPLTLRTVQKWLLAAGARDCHSVTTQDFL
    Cow LSELVQAVSDPSSPQYGKYLTLENVADLVRPSPLTLHTVQKWLLAAGAQKCHSVITQDFL
    Dog LSKLVQAVSDPGSPHYGKYLTLEDVAELVRPSPLTFRTVQKWLSAAGARNCHSVTTQDFL
   Fugu LERMLRVVSDPDSANYGKYLTLEDVGNLVRPSDLTQKVVRDWLQRNGVTNCRTVRTQDFL
Zebrafish MEELLKLVSDPDSHQYGKYLSLDEVAALSRPSPLTEKVVENWLRSHGVMDCHTIITRDFL

  Human TCWLSIRQAELLLPGAEFHHYVGGPTETHVVRSPHPYQLPQALAPHVDFVGGLHHFPPTS
Macaque TCWLSIRQAELLLPGAEFHHYVGGPTETHVVRSPRPYQLPQALAPHVDFVGGLHRFPPTS
  Mouse TCWLSVRQAELLLPGAEFHRYVGGPTKTHVIRSPHPYQLPQALAPHVDFVGGLHRFPP-S
    Rat TCWLSVRQAELLLPGAEFHRYVGGPAKTHIIRSPHPYQLPQALAPHVDLVAGLHRFPPLS
    Cow TCWLSIRQAELLLPGAEFHHYVGGPTETHVVRSPHPYQLPQALAPHVDFVGGLHRFPPTS
    Dog TCWLSVRQAELLLSGAEFHRYVGGPTEIHVIRSLRPYQLPKALAPHVDFVGGLHRFPPTS
   Fugu ECSMTAEVAEQVLPGSRFHRYTR--DGQSVVRSSAPYSVPDDVYQHLDFVGGLHRFPLKQ
Zebrafish QCVMTVEVAEALLPGSKFHRFSK--NTKTLLRSTSQYSVHEDVHQHLDFVGGVHRFPQKR

  Human SLRQRP--EPQVTGTVG-LHLGVTPSVIRKRYNLTSQDVGSGTSNNSQACAQFLEQYFHD
Macaque SLRQRP--EPQVTGTVG-LHLGVTPSVIRKRYNLTSQDVGSGTSNNSQACAQFLEQYFHD
  Mouse SPRQRP--EPQQVGTVS-LHLGVTPSVLRQRYNLTAKDVGSGTTNNSQACAQFLEQYFHN
    Rat SPRQRP--EPQGVGPVG-LHLGVTPSVLRQRYNLTARDVGSGTTNNSQACAQFLEQYFHN
    Cow SLRQRP--EPQVTGTVG-LHLGVTPSVIRKRYNLTSQDVGSGTSNNSQACAQFLEQYFHD
    Dog SLRQRP--EPQVSGTVG-LHLGVTPSVIRQRYNLTAQDVGSGTTNNSQACAQFLEQYFHA
   Fugu QQDVRSYFKREQNNVKAKIHLGVTPSVLRARYNLTSADVGQAQ-NNSQAVAQFLEQYYSP
Zebrafish KIVSKG-WEGARQAILG-YHLGVTPAVIRNRYNLTAKDVGTAA-NNSQAVAQFLEQYYHP

  Human SDLAQFMRLFGGNFAHQASVARVVGQQGRGRAGIEASLDVQYLMSAGANISTWVYSSPGR
Macaque SDLAQFMRLFGGNFAHQASVTRVVGQQGRGRAGIEASLDVQYLMSAGANISTWVYSSPGR
  Mouse SDLTEFMRLFGGSFTHQASVAKVVGKQGRGRAGIEASLDVEYLMSAGANISTWVYSSPGR
    Rat SDLTEFMRLFGSSFAHQASVARVVGKQGRGRAGIEASLDVEYLMSAGANISTWVYSSPGR
    Cow SDLAQFMRLFGGNFAHQASVARVVGQQGRGRAGIEASLDVQYLMSAGANISTWVYSSPGR
    Dog SDLAEFMRLFGGNFAHQASVARVVGQQGRGRAGIEASLDVEYLMSAGANISTWVYSSPGR
   Fugu ADLVEFMSLFGSGFKHHTQVDRVVGTQGAGKAGIEASLDVQYIMSTGANIPTWLFTNPGR
Zebrafish ADLAEFMSLFGGGFTHMSTVERVVGTQGGGKAGIEASLDVEYIMSSGANISTWVFTNPGR

  Human HEGQEPFLQWLMLLSNESALPHVHTVSYGDDEDSLSSAYIQRVNTELMKAAARGLTLLFA
Macaque HEGQEPFLQWLMLLSNESALPHVHTVSYGDEEDSLSSAYIQRVNTELMKAAARGLTLLFA
  Mouse HEAQEPFLQWLLLLSNESSLPHVHTVSYGDDEDSLSSIYIQRVNTEFMKAAARGLTLLFA
    Rat HEAQEPFLQWLLLLSNESSLPHVHTVSYGDDEDSLSSVYIQRVNTEFMKAAARGLTLLFA
    Cow HEGQEPFLQWLMLLSNESALPHVHTVSYGDDEDSLSSAYIQRVNTELMKAAARGLTLLFA
    Dog HESQEPFLQWLLLLSNESALPHVHTVSYGDDEDSLSSAYIQRVNTEFMKAAARGLTLLFA
   Fugu HESQEPFLQWMVLLSNMSDLPWVHTISYGDDEDSLSTAYMERINAEFMKAGMRGISLLFA
Zebrafish HESQEPFLQWMLLLSNMSAVPWVHTISYGDDEDSLSEAYMNRINIEFMKAGLRGISMLFA

  Human SGDSGAGCWSVSGRHQFRPTFPASSPYVTTVGGTSFQEPFLITNEIVDYISGGGFSNVFP
Macaque SGDSGAGCWSVSGRHQFRPTFPASSPYVTTVGGTSFQEPFLITNEIVDYISGGGFSNVFP
  Mouse SGDTGAGCWSVSGRHKFRPSFPASSPYVTTVGGTSFKNPFLITDEVVDYISGGGFSNVFP
    Rat SGDTGAGCWSVSGRHKFRPSFPASSPYVTTVGGTSFKNPFLVTNEVVDYISGGGFSNVFP
    Cow SGDSGAGCWSVSGRHQFRPTFPASSPYVTTVGGTSFQEPFLITNEIVDYISGGGFSNVFP
    Dog SGDSGAGCWSVSRRHQFRPSFPASSPYVTTVGGTSFQNPFRVTTEIVDYISGGGFSNVFP
   Fugu SGDSGAGCRRSV-KNSFRPSFPASSPYVTTVGGTSFKNPFTVSYEVTDYISGGGFSNVFK
Zebrafish SGDSGAGCRHLT-KERNTPSFAASSPYVPTVGGTSFKTPSTLTYEVQILITGGGLNYVFP

  Human RPSYQEEAVTKFLSS-SPHLPPSSYFNASGRAYPDVAALSDGYWVVSNRVPIPWVSGTSA
Macaque RPSYQEEVVAKFLSS-SPHLPPSGYFNASGRAYPDVAALSDGYWVVNNRVPIPWVSGTSA
  Mouse RPPYQEEAVAQFLKS-SSHLPPSSYFNASGRAYPDVAALSDGYWVVSNMVPIPWVSGTSA
    Rat QPSYQEEAVAQFLKS-SSHLPPSSYFNASGRAYPDVAALSDGYWVVSNMVPIPWVSGTSA
    Cow RPSYQEEAVTKFLSS-SPHLPPSSYFNASGRAYPDVAALSDGYWVVSNRVPIPWVSGTSA
    Dog QPSYQEEAVVQFLSS-SPHLPPSSYFNASGRAYPDVAALSDGYWVVSNSVPIPWVSGTSA
   Fugu MPDYQASAVDAYLKTMAKVLPPQSYYNTTGRAYPDMAALSDNYWVVSNKVPVPWVSGTSA
Zebrafish MPDYQVDAVRAYLKS-VQSLPPQTYFNTTGRAYPDLAALSDNYWVVSNRVPIPWVSGTSA

  Human STPVFGGILSLINEHRILSGRPPLGFLNPRLYQQHGAGLFDVTRGCHESCLDEEVEGQGF
Macaque STPVFGGLLSLINEHRILSGRPPLGFLNPRLYQQHGAGLFDVTHGCHASCLDEEVEGQGF
  Mouse STPVFGGILSLINEHRILNGRPPLGFLNPRLYQQHGTGLFDVTHGCHESCLNEEVEGQGF
    Rat STPVFGGILSLINEHRLLNGRPPLGFLNPRLYQQHGAGLFDVTHGCHESCLNEEVEGQGF
    Cow STPVFGGILSLINEHRILSGRPPLGFLNPRLYQQHGAGLFDVTRGCHESCLDEEVEGQGF
    Dog STPVFGGILSLINEHRLLSGLPPLGFLNPRLYQQRGAGLFDVTRGCHESCLNEEVQGQGF
   Fugu STPVVGGMLSLINDQRLLKGLPVLGFLNPRLYRLKGQALFDVTEGCHLSCLDEQVQGKGF
Zebrafish STPVVGGILSLINDQRFLKGLPALGFINPRLYKMQGKGLYDVTVGCHLSCLDDKVEGKGF

  Human CSGPGWDPVTGWGTPNFPALLKTLLNP
Macaque CSGPGWDPVTGWGTPNFPALLKTLLNP
  Mouse CSGPGWDPVTGWGTPNFPALLKTLLNP
    Rat CSGPGWDPVTGWGTPNFPALLKTLLNP
    Cow CSGPGWDPVTGWGTPNFPALLKTLLNP
    Dog CSGPGWDPVTGWGTPNFPALLKALIKP
   Fugu CAAPSWDPVTGWGTPNYPELLTALL--
Zebrafish CASPSWDPVTGWGTPNYPVFLASLMD-
```

REFERENCES

Barrett AJ, Rawlings ND, Woessner JF. (1998) *Handbook of proteolytic enzymes.* Academic Press, London.

Benard M, Pallotta D, Pierron G. (1992) Structure and identity of a late-replicating and transcriptionally active gene. *Exp Cell Res.*; **201**: 506–13.

Bullock TL, Breddam K, Remington SJ. (1996) Peptide aldehyde complexes with wheat serine carboxypeptidase II: implications for the catalytic mechanism and substrate specificity. *J Mol Biol.*; **255**: 714–25.

Comellas-Bigler M, Fuentes-Prior P, Maskos K, Huber R, Oyama H, Uchida K, Dunn BM, Oda K, Bode W. (2002) The 1.4 Å crystal structure of kumamolysin: a thermostable serine-carboxyl-type proteinase. *Structure (Camb).*; **10**: 865–76.

Davies DR. (1990) The structure and function of the aspartic proteinases. *Annu Rev Biophys Biophys Chem.*; **19**: 189–215.

Delbaere LT, Brayer GD. (1985) The 1.8 Å structure of the complex between chymostatin and *Streptomyces griseus* protease A. A model for serine protease catalytic tetrahedral intermediates. *J Mol Biol.*; **183**: 89–103.

Ghosh D, Weeks CM, Grochulski P, Duax WL, Erman M, Rimsay RL, Orr JC. (1991) Three-dimensional structure of holo 3-alpha, 20-beta-hydroxysteroid dehydrogenase: a member of a short-chain dehydrogenase family. *Proc Natl Acad Sci U S A.*; **88**: 10064–8.

Holm L, Sander C. (1993) Protein structure comparison by alignment of distance matrices. *J Mol Biol.*; **233**: 123–38.

Ito M, Dunn BM, Oda K. (1996) Substrate specificities of pepstatin-insensitive carboxyl proteinases from gram-negative bacteria. *J Biochem (Tokyo).*; **120**: 845–50.

Ito M, Narutaki S, Uchida K, Oda K. (1999) Identification of carboxyl residues in pepstatin-insensitive carboxyl proteinase from *Pseudomonas* sp. 101 that participate in catalysis and substrate binding. *J Biochem (Tokyo).*; **125**: 210–6.

Katz ML, Johnson GS. (2001) Mouse gene knockout models for the CLN2 and CLN3 forms of ceroid lipofuscinosis. *Eur J Paediatr Neurol.*; **5** (Suppl A): 109–14.

Kraulis PJ. (1991) MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J Appl Crystallogr.*; **24**: 946–50.

Kuhn P, Knapp M, Soltis SM, Ganshaw G, Thoene M, Bott R. (1998) The 0.78 Å structure of a serine protease: *Bacillus lentus* subtilisin. *Biochemistry.*; **37**: 13446–52.

Kwon HK, Kim H, Ahn TI. (1999) Pepstatin-insensitive carboxyl proteinase: a biochemical marker for late lysosomes in *Amoeba proteus*. *Korean J Biol Sci.*; **3**: 221–8.

Lin L, Sohar I, Lackland H, Lobel P. (2001) The human CLN2 protein/tripeptidyl-peptidase I is a serine protease that autoactivates at acidic pH. *J Biol Chem.*; **276**: 2249–55.

Matthews DA, Alden RA, Birktoft JJ, Freer ST, Kraut J. (1975) X-ray crystallographic study of boronic acid adducts with subtilisin BPN′ (Novo) A model for the catalytic transition state. *J Biol Chem.*; **250**: 7120–6.

Miller M, Rao JKM, Wlodawer A, Gribskov MR. (1993) A left-handed crossover involved in amidohydrolase catalysis. Crystal structure of *Erwinia chrysanthemi* L-asparaginase with bound L-aspartate. *FEBS Lett.*; **328**: 275–9.

Murao S, Ohkuni K, Nagao M, Hirayama K, Fukuhara K, Oda K, Oyama H, Shin T. (1993) Purification and characterization of kumamolysin, a novel thermostable pepstatin-insensitive carboxyl proteinase from *Bacillus* novosp. MN-32. *J Biol Chem.*; **268**: 349–55.

Narutaki S, Dunn BM, Oda K. (1999) Subsite preferences of pepstatin-insensitive carboxyl proteinases from bacteria. *J Biochem (Tokyo).*; **125**: 75–81.

Nishii W, Ueki T, Miyashita R, Kojima M, Kim YT, Sasaki N, Murakami-Murofushi K, Takahashi K. (2003) Structural and enzymatic characterization of physarolisin (formerly physaropepsin) proves that it is a

unique serine-carboxyl proteinase. *Biochem Biophys Res Commun.*; **301**: 1023–9.

Oda K, Sugitani M, Fukuhara K, Murao S. (1987) Purification and properties of a pepstatin-insensitive carboxyl proteinase from a gram-negative bacterium. *Biochim Biophys Acta.*; **923**: 463–9.

Oda K, Nakatani H, Dunn BM. (1992) Substrate specificity and kinetic properties of pepstatin-insensitive carboxyl proteinase from *Pseudomonas* sp. No. 101. *Biochim Biophys Acta.*; **1120**: 208–14.

Oda K, Takahashi T, Tokuda Y, Shibano Y, Takahashi S. (1994) Cloning, nucleotide sequence, and expression of an isovaleryl pepstatin-insensitive carboxyl proteinase gene from *Pseudomonas* sp. 101. *J Biol Chem.*; **269**: 26518–24.

Oda K, Ogasawara S, Oyama H, Dunn BM. (2000) Subsite preferences of pepstatin-insensitive carboxyl proteinases from prokaryotes: kumamolysin, a thermostable pepstatin-insensitive carboxyl proteinase. *J Biochem (Tokyo).*; **128**: 499–507.

Orry A, Wallace BA. (1999) A proposed model for the late-infantile neuronal ceroid lipofuscinosis (Batten Disease) protein CLN2. *Protein Pept Lett.*; **6**: 1–5.

Oyama H, Abe S, Ushiyama S, Takahashi S, Oda K. (1999) Identification of catalytic residues of pepstatin-insensitive carboxyl proteinases from prokaryotes by site-directed mutagenesis. *J Biol Chem.*; **274**: 27815–22.

Oyama H, Hamada T, Ogasawara S, Uchida K, Murao S, Beyer BB, Dunn BM, Oda K. (2002) A CLN2-related and thermostable serine-carboxyl proteinase, kumamolysin: cloning, expression, and identification of catalytic serine residue. *J Biochem (Tokyo).*; **131**: 757–65.

Rawlings ND, Barrett AJ. (1999) Tripeptidyl-peptidase I is apparently the CLN2 protein absent in classical late-infantile neuronal ceroid lipofuscinosis. *Biochim Biophys Acta.*; **1429**: 496–500.

Robertus JD, Kraut J, Alden RA, Birktoft JJ. (1972) Subtilisin; a stereochemical mechanism involving transition-state stabilization. *Biochemistry.*; **11**: 4293–303.

Ruepp A, Grami W, Santos-Martinez M-L, Koretke KK, Volker C, Mewes HW, Frishman D, Stocker S, Lupas AN, Baumeister W. (2000) The genome sequence of the thermoacidophilic scavenger *Thermoplasma acidophilum. Nature.*; **407**: 508–13.

Sack JS, Saper MA, Quiocho FA. (1989) Periplasmic binding protein structure and function. Refined X-ray structures of the leucine/isoleucine/valine-binding protein and its complex with leucine. *J Mol Biol.*; **206**: 171–91.

Schechter I, Berger A. (1967) On the size of the active site in proteases. I. Papain. *Biochem Biophys Res Commun.*; **27**: 157–62.

Shibata M, Dunn BM, Oda K. (1998) Substrate specificity of pepstatin-insensitive carboxyl proteinase from *Bacillus coagulans* J-4. *J Biochem (Tokyo).*; **124**: 642–7.

Sleat DE, Donnelly RJ, Lackland H, Liu CG, Sohar I, Pullarkat RK, Lobel P. (1997) Association of mutations in a lysosomal protein with classical late-infantile neuronal ceroid lipofuscinosis. *Science.*; **277**: 1802–5.

Sussman JL, Harel M, Frolow F, Oefner C, Goldman A, Toker L, Silman I. (1991) Atomic structure of acetylcholinesterase from *Torpedo californica*: a prototypic acetylcholine-binding protein. *Science.*; **253**: 872–9.

Tsuruoka N, Nakayama T, Ashida M, Hemmi H, Nakao M, Minakata H, Oyama H, Oda K, Nishino T. (2003) Collagenolytic serine-carboxyl proteinase from *Alicyclobacillus sendaiensis* strain NTAP-1: Purification, characterization, gene cloning, and heterologous expression. *Appl Environ Microbiol.*; **69**: 162–9.

Waterston RH *et al.* (2002) Initial sequencing and comparative analysis of the mouse genome. *Nature.*; **420**: 520–62.

Wlodawer A, Li M, Dauter Z, Gustchina A, Uchida K, Oyama H, Dunn BM, Oda K. (2001a) Carboxyl proteinase from *Pseudomo-*

nas defines a novel faily of subtilisin-like enzymes. *Nat Struct Biol.*; **8**: 442–6.

Wlodawer A, Li M, Gustchina A, Dauter Z, Uchida K, Oyama H, Goldfarb NE, Dunn BM, Oda K. (2001b) Inhibitor complexes of the *Pseudomonas* serine-carboxyl proteinase. *Biochemistry.*; **40**: 15602–11.

Wright CS, Alden RA, Kraut J. (1969) Structure of subtilisin BPN′ at 2.5 Å resolution. *Nature.*; **221**: 235–42.

Wright CS, Alden RA, Kraut J. (1972) Crystal structure of a subtilisin BPN′ complex with *N*-benzoyl-L-arginine. *J Mol Biol.*; **66**: 283–9.