
FOR THE RECORD

Peptidase family U34 belongs to the superfamily of N-terminal nucleophile hydrolases

JIMIN PEI² AND NICK V. GRISHIN^{1,2}

¹Howard Hughes Medical Institute, ²Department of Biochemistry, University of Texas Southwestern Medical Center, Dallas, Texas 75390, USA

(RECEIVED November 27, 2002; FINAL REVISION January 31, 2003; ACCEPTED February 3, 2003)

Abstract

Peptidase family U34 consists of enzymes with unclear catalytic mechanism, for instance, dipeptidase A from *Lactobacillus helveticus*. Using extensive sequence similarity searches, we infer that U34 family members are homologous to penicillin V acylases (PVA) and thus potentially adopt the N-terminal nucleophile (Ntn) hydrolase fold. Comparative sequence and structural analysis reveals a cysteine as the catalytic nucleophile as well as other conserved residues important for catalysis. The PVA/U34 family is variable in sequence and exhibits great diversity in substrate specificity, to include enzymes such as chologlycine hydrolases, acid ceramidases, isopenicillin N acyltransferases, and a subgroup of eukaryotic proteins with unclear function.

Keywords: Peptidase family U34; penicillin V acylase; Ntn-hydrolase; structure prediction; peptidase classification

Peptidases are a diverse group of enzymes that hydrolyze the peptide bonds in proteins, peptides, and various other substrates. Due to their importance in biology and medicine, many peptidases are well studied for their sequence, structure, and catalytic mechanisms. Despite the diversity of peptidase sequences and structures, only a limited number of catalytic types are known, such as aspartic (A), cysteine (C), metallo (M), serine (S) and threonine (T) proteases. A comprehensive sequence and structure-based classification of peptidases is available from the MEROPS database (Barrett et al. 2001; Rawlings et al. 2002). In the MEROPS database, a peptidase family consists of peptidases with significant sequence similarity. Each peptidase family is named by a letter specifying its catalytic type and a number, such as S32. Families considered to be evolutionarily related are grouped into clans. Among the nearly 200 peptidase families in MEROPS, there are only a few with unknown cata-

lytic type (U). In these families, the key residues for catalysis have not been revealed by experimentation or theoretical analysis. With the development of sensitive similarity search tools such as PSI-BLAST (Altschul et al. 1997) and the enlargement of sequence and structure databases, more remote homologs can be inferred by computational analysis. This has been demonstrated for numerous peptidases (Lewis and Thomas 1999; Makarova and Grishin 1999a, b; Makarova et al. 2000; Pei and Grishin 2001b). Here, homology between peptidase family U34 and penicillin V acylases is described. We predict that U34 peptidases are cysteine proteases that belong to the diverse superfamily of the N-terminal nucleophile hydrolases.

Results and Discussion

Similarity searches for U34 family peptidases

The first member of the U34 family was isolated from the lactic acid bacterial species *Lactobacillus helveticus* as a broad-specificity dipeptidase, designated dipeptidase A or *pepDA* (Dudley et al. 1996). PSI-BLAST searches (Altschul et al. 1997) starting with the full sequence of *pepDA* against the nonredundant protein database (nr) maintained at NCBI

Reprint requests to: Nick V. Grishin, Howard Hughes Medical Institute, University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd., Dallas, TX 75390, USA; e-mail: grishin@chop.swmed.edu; fax: (214) 648-9099.

Article and publication are at <http://www.proteinscience.org/cgi/doi/10.1110/ps.0240803>.

(November 2, 2002: 1,226,480 sequences; 390,314,779 total letters; e-value cutoff 0.01) converged to ~40 U34 family homologs, among which there are a few eukaryotic and archaeal proteins (Fig. 2A, below). The closest homologs of *pepDA* are mainly from the *Lacobacillus* species and the *Streptococcus* species, often annotated as putative dipeptidase A or hypothetical proteins. The eukaryotic and archaeal homologs are all hypothetical proteins without functional annotations. None of the homologs have known structures or characterized peptidase catalytic types. PSI-BLAST searches are sensitive to the query sequence. To ensure full coverage, these found homologs were grouped by single-linkage clustering (1 bit per site threshold, ~50% sequence identity), and representative sequences from each group were used as queries for further PSI-BLAST iterations, as scripted in the SEALS package (Walker and Koonin 1997). During the course of extensive similarity searches, we found statistically supported evidence that U34 peptidases and penicillin V acylases (PVA) are remote homologs. For example, when *pepDA* (NCBI gene identification [gi] no. 1072051) was used as a query (e-value cutoff 0.01, with composition-based statistics), one homolog from *Mus musculus* (gi no. 12852713) was found in the second iteration with e-value $2e-09$. When this eukaryotic homolog was used as a query, it found the penicillin V acylase from *Bacillus sphaericus* (gi no. 129549) in the fourth iteration with e-value $2e-04$. Because the crystal structure of penicillin V acylase from *B. sphaericus* has been determined (Fig. 1) (Suresh et al. 1999), the inferred homology between U34 peptidases and penicillin V acylases helps to predict the fold and the catalytic mechanism of U34 peptidases.

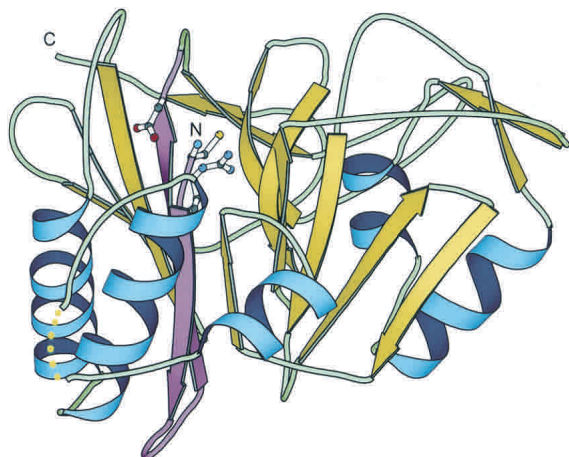


Figure 1. The structure of the penicillin V acylase from *Bacillus sphaericus*. The Protein Data Bank (PDB) code is 3pva, chain A. The first β hairpin is highlighted in purple. The backbone amino group and the side chain of Cys 1 and the side chains of Arg 17 and Asp 20 are shown in ball and stick. The structure is drawn using the program BOBSCRIPT (Esnouf 1997).

Penicillin acylases and the Ntn-hydrolases

Penicillin acylases (EC 3.5.1.11) catalyze the hydrolysis of penicillin into 6-aminopenicillanic acid (6-APA) and an organic acid (Mahajan 1984). There are two distinct groups of penicillin acylases with different substrate preferences; penicillin V acylases (PVA) prefer to cleave phenoxymethyl penicillin (penicillin V) and penicillin G acylases (PGA) have higher affinity for phenylacetil penicillin (penicillin G). Both enzymes are used in industry to produce semi-synthetic penicillin. Their structures (Duggleby et al. 1995; Suresh et al. 1999) are similar and both proteins belong to a large superfamily of amidohydrolases called the N-terminal nucleophile (Ntn) hydrolases (Brannigan et al. 1995). Ntn-hydrolases utilize the side chain of the amino-terminal residue to perform the nucleophilic attack on the target amide bond (Brannigan et al. 1995). Many structures of the Ntn-hydrolases have been determined. In the SCOP (Structure Classification of Proteins) database (Murzin et al. 1995), Ntn-hydrolase fold also includes several other families, such as class II glutamine amidotransferases (Smith et al. 1994), proteosome subunits (Lowe et al. 1995; Groll et al. 1997), and (glycosyl) asparaginases (Oinonen et al. 1995; Table 1). Their structures have similar architecture and topology consisting of two layers of β sheets sandwiched by two layers of α helices (Fig. 1). They are considered to be related evolutionarily on the basis of structural similarity and the common location of the catalytic nucleophile, although sequence similarities are very weak among different families. The structures also show great variability in the number of secondary structural elements and the details of their arrangement (Brannigan et al. 1995; Oinonen and Rouvinen 2000). The catalytic nucleophile can be a cysteine, a serine, or a threonine. For example, PVA is a cysteine peptidase, whereas PGA is a serine peptidase. According to the MEROPS classification, Ntn-hydrolases form peptidase clan PB, which is the only clan with three different catalytic types (Table 1). Ntn-hydrolases represent a superfamily of amidohydrolases that have developed great divergence in sequence, structure, and substrate specificity (Table 1). Our prediction adds peptidase family U34 as a new member of the Ntn-hydrolase superfamily.

The PVA/U34 family

Because the U34 peptidases and PVAs can be linked by sequence similarity searches, we denote these homologs as the PVA/U34 family. We found ~150 members in this family by transitive sequence similarity searches. Besides the U34 peptidases and PVAs, these proteins include cholesteryl esterase (or conjugate bile salt hydrolases, EC 3.5.1.24), which has been noted before as close homologs of PVAs (Christiaens et al. 1992). We also found that acid ceramidases (N-acylsphingosine amidohydrolase,

Table 1. N-terminal nucleophile hydrolase families

SCOP family	Representative structure	Nucleophile	MEROPS family
Penicillin V acylase Class II glutamine amidotransferases	3pvaA	Cys	C45
Penicillin G acylase	1ecfB	Cys	C44
Proteasome catalytic subunits (Glycosyl)asparaginase	1ajqB	Ser	S45
	1pmaP	Thr	T1
	1apyB	Thr	T2

EC 3.5.1.23) are close homologs of PVAs. Acid ceramidases are eukaryotic enzymes that hydrolyze the sphingolipid ceramide into sphingosine and free fatty acid. Deficiency of acid ceramidase activity leads to lysosomal storage disorder known as Farber disease (Koch et al. 1996). Another subgroup consists of isopenicillin N acyltransferases (acyl-coenzyme A:6-aminopenicillanic-acid-acyltransferases, EC 2.3.1; Montenegro et al. 1990), which the MEROPS assigns to family C45 in clan PB. Isopenicillin N acyltransferases are involved in the synthesis of penicillin. The PVA/U34 family hydrolases have diverged to catalyze the hydrolysis of a variety of substrates and play different physiological roles.

Active site in the PVA/U34 family

The most striking conserved feature in the PVA/U34 family is the catalytic cysteine residue (Fig. 2). The side chain of the cysteine serves as the nucleophile and the free α NH₂ serves as the proton donor and acceptor in the catalytic process. For all Ntn-hydrolases, the catalytic residue is uncovered in the active enzyme by the removal of the sequences N-terminal to it. We have found different ways to achieve this in the PVA/U34 family proteins. Proteins in the subgroup of penicillin V acylases and the conjugate bile salt hydrolases usually have the catalytic cysteine as the second residue, so the active residue is revealed right after the removal of the initiation formyl-methionine (Fig. 2B). One close homolog of *pepDA* is characterized experimentally as an extracellular arginine aminopeptidase from *Streptococcus gordonii* (gi no. 16506526, Fig. 2A; Goldstein et al. 2002). This protein has a typical export signal sequence of 14 hydrophobic residues. The predicted catalytic cysteine residue is right after the cleavage site and, thus, is exposed after the removal of the signal sequence. Inhibitor studies showed that this protein has some cysteine protease characteristics, in support of our predictions. The acid ceramidases usually have a relatively long sequence N-terminal to the catalytic cysteine. The removal of this N-terminal part may be an autoproteolysis process, like in many other Ntn-hydrolases.

The strongest sequence signal for all PVA/U34 family proteins resides in the motif containing the catalytic cys-

teine residue and corresponds to the N-terminal β -hairpin in the structure of *B. sphaericus* penicillin V acylase (PDB id: 3pva; Fig. 1, in purple). Other common features in this motif include the hydrophobic pattern and positions occupied mainly by small residues near the catalytic cysteine (Fig. 2). The β -hairpin motif is longer in the close homologs of U34 family dipeptidases (Fig. 2A) than in the close homologs of PVAs (Fig. 2B). Two other residues (Arg 17 and Asp 20) are also highly conserved in most of the PVA/U34 proteins. In the structure of 3pva, Arg 17 makes hydrogen bonds to the opposite β sheet, two to the main chain carboxy groups (Tyr 68 and Met 80) and one to the side chain of Asp 69 (Suresh et al. 1999). We predict that Arg 17 should be important in maintaining the overall stability of the structure. Moreover, one side-chain nitrogen of Arg 17 is only 3.8 Å away from the catalytic sulfhydryl group, suggesting that Arg 17 could also be involved in catalysis. The position corresponding to Arg 17 is usually occupied by a positively charged residue in PVA/U34 homologs (Fig. 2). The side chain of Asp 20 makes a hydrogen bond with the free backbone amino group of the catalytic cysteine. This interaction is critical for maintaining the orientation of the cysteine residue for catalysis. Conservation of Arg 17 and Asp 20 is unique in the PVA/U34 family, compared with the other Ntn-hydrolase families (Fig. 2D). Another important part of the catalytic machinery is the oxyanion hole, which is used to stabilize the negative charges developed on the substrate carboxy group in the transition state. Crystal structure of 3pva has revealed that the oxyanion hole consists of the side chain N δ 2 of Asn 175 and the main chain NH of Tyr 82 (Suresh et al. 1999). However, PSI-BLAST local alignments of the PVA/U34 homologs are usually restricted to the very N-terminal conserved β hairpins, not covering the position of Asn 175. This indicates that the rest of the sequences are fairly diverse among different PVA/U34 subgroups, which is consistent with the broad scope of substrates that different subgroups can have. We made a global alignment of all found PVA/U34 homologs using the program PCMA (Pei et al. 2003) which utilizes the consistency of profiles in the alignment process. The position corresponding to Asn 175 has a high conservation value (Pei and Grishin 2001a). The full alignment is available at <ftp://iole.swmed.edu/dipep/dipep.aln>.

Structural comparisons have revealed a great diversity in the exact placement of active site components for different Ntn-hydrolases (Oinonen and Rouvinen 2000). Even in the PVA/U34 family, a few diverse subgroups have different conservation patterns in the β -hairpin motif. Two are shown in Figure 2c. The isopenicillin N acyltransferases (Montenegro et al. 1990) have the positively charged Arg 17 replaced by a glutamine residue. The other subgroup consists of eukaryotic proteins, most of which are hypothetical proteins. Two experimentally cloned proteins are *Drosophila* protein LAMA (lamina ancestor; Perez and Steller 1996)

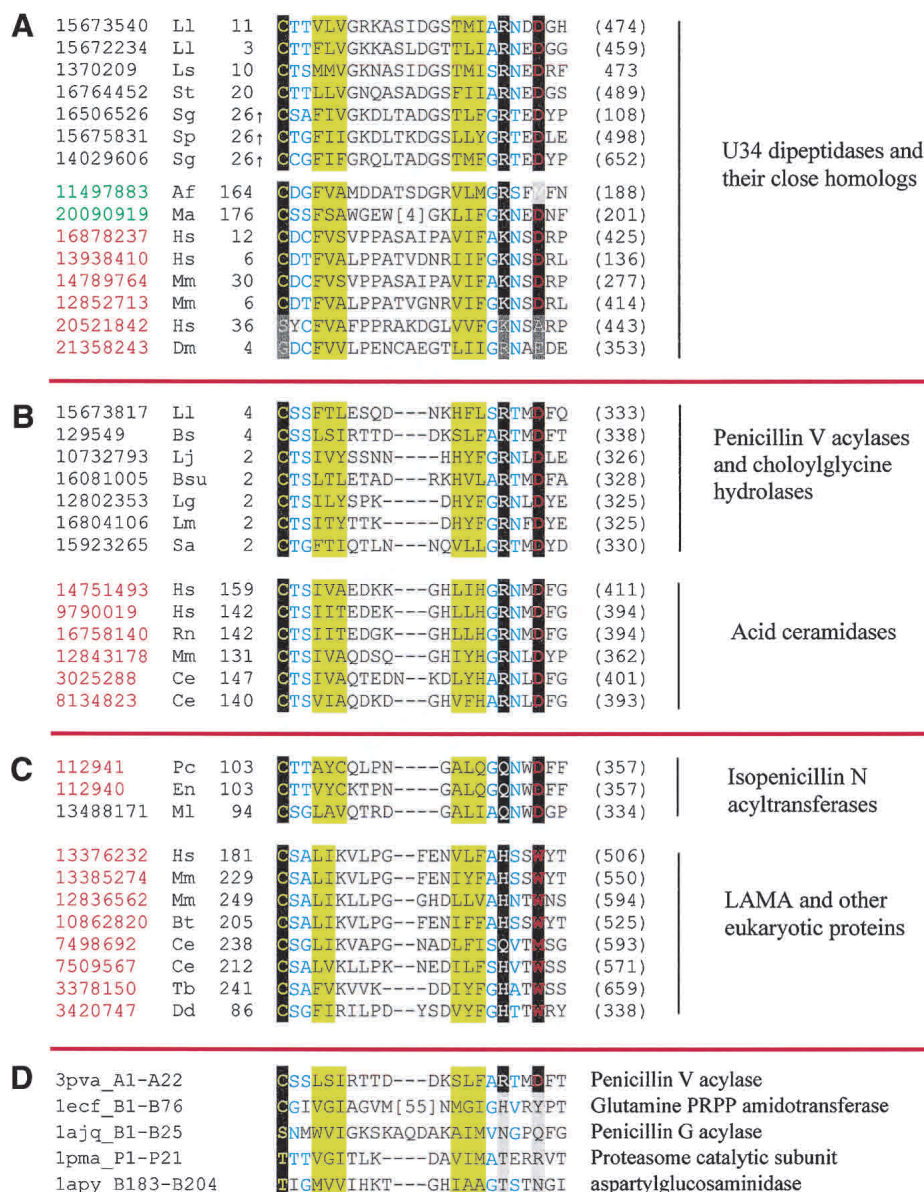


Figure 2. Multiple sequence alignment of the first β hairpin motif in the PVA/U34 family and other Ntn hydrolases. The sequence identifiers at left are NCBI gene identification (gi) numbers, followed by the abbreviations of the species names. Each identifier consists of the PDB and chain names and the residue range: (Af) *Archaeoglobus fulgidus*; (Bt) *Bos taurus*; (Bs) *Bacillus sphaericus*; (Bsu) *Bacillus subtilis*; (Ce) *Caenorhabditis elegans*; (Dd) *Dictyostelium discoideum*; (Dm) *Drosophila melanogaster*; (En) *Emericella nidulans*; (Hs) *Homo sapiens*; (Lg) *Lactobacillus gasseri*; (Lj) *Lactobacillus johnsonii*; (Ll) *Lactococcus lactis*; (Lm) *Listeria monocytogenes*; (Ls) *Lactobacillus sakei*; (Ma) *Methanosarcina acetivorans*; (Mm) *Mus musculus*; (Pc) *Penicillium chrysogenum*; (Rn) *Rattus norvegicus*; (Sa) *Staphylococcus aureus*; (Sg) *Streptococcus gordonii*; (Sp) *Streptococcus pyogenes*; (St) *Salmonella typhimurium*; (Tb) *Trypanosoma brucei*. Archaeal sequences and eukaryotic sequences are marked in red and green identifiers, respectively. The first residue number and the sequence length of each PVA/U34 homolog are shown. The catalytic cysteine residue, Arg 17 and Asp 20 are highlighted with black background. Positions occupied by mainly hydrophobic residues are shaded in yellow. Positions with mainly small residues (G, A, S, C, T, V, N, D) have these residues shown in blue letters. Two sequences that do not have the catalytic cysteine residues in have the key residues shaded in gray. (A) Representatives of close homologs of U34 family dipeptidases. The arrow next to a sequence indicates putative signal sequence cleavage site (Goldstein et al. 2002). (B) Representatives of close homologs of penicillin V acylases. (C) Two relatively divergent subgroups of the PVA/U34 family. (D) Representative sequences of the 5 Ntn-hydrolase families with known structures.

and *Trypanosoma* protein p67 (Kelley et al. 1999). *Drosophila* LAMA protein is shown to be expressed in the

lamina (the first optic ganglion) precursors and appears to be involved in lamina development. *Trypanosoma* p67 is a

lysosomal/endosomal membrane glycoprotein that may function in lipid metabolism, similar to the lysosomal acid ceramidases. In this subgroup, the position of Arg 17 is often occupied by a histidine and Asp 20 is replaced by a large hydrophobic residue, usually a tryptophan.

Conclusions

We predict that the U34 family peptidases are cysteine-type Ntn-hydrolases. Members in the PVA/U34 family share similar catalytic machinery but have developed great diversity in sequence to catalyze a variety of reactions of amide bond hydrolysis.

Materials and methods

Sequence similarity searches and multiple sequence alignment

The PSI-BLAST program (Altschul et al. 1997) was used to search for homologs of *Lactobacillus helveticus* dipeptidase A against the NCBI non-redundant database (November 2, 2002: 1,226,480 sequences; 390,314,779 total letters). The e-value threshold was 0.01 for inclusion of sequences into a profile. The other parameters were default. To ensure full coverage, found homologs were grouped by single-linkage clustering (1 bit per site threshold, about 50% sequence identity) and representative sequences from each group were used as queries for further PSI-BLAST iterations, as scripted in the SEALS package (Walker and Koonin 1997).

A multiple sequence alignment was constructed using the PCMA program (Pei et al. 2003) for all found homologs. Sequence conservation analysis was performed using the program AL2CO (Pei and Grishin 2001a). The full alignment is available at ftp://iole.swmed.edu/dipep/dipep.aln. The N-terminal part of the alignment shown in Figure 2 was manually inspected and curated. The structural alignment of several Ntn hydrolases is from the FSSP database (Figure 2D; Holm and Sander 1998).

Acknowledgments

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

References

- Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **25**: 3389–3402.
- Barrett, A.J., Rawlings, N.D., and O'Brien, E.A. 2001. The MEROPS database as a protease information system. *J. Struct. Biol.* **134**: 95–102.
- Brannigan, J.A., Dodson, G., Duggleby, H.J., Moody, P.C., Smith, J.L., Tomchick, D.R., and Murzin, A.G. 1995. A protein catalytic framework with an N-terminal nucleophile is capable of self-activation. *Nature* **378**: 416–419.
- Christiaens, H., Leer, R.J., Pouwels, P.H., and Verstraete, W. 1992. Cloning and expression of a conjugated bile acid hydrolase gene from *Lactobacillus plantarum* by using a direct plate assay. *Appl. Environ. Microbiol.* **58**: 3792–3798.
- Dudley, E.G., Husgen, A.C., He, W., and Steele, J.L. 1996. Sequencing, distribution, and inactivation of the dipeptidase A gene (pepDA) from *Lactobacillus helveticus* CNRZ32. *J. Bacteriol.* **178**: 701–704.
- Duggleby, H.J., Tolley, S.P., Hill, C.P., Dodson, E.J., Dodson, G., and Moody, P.C. 1995. Penicillin acylase has a single-amino-acid catalytic centre. *Nature* **373**: 264–268.
- Ensnouf, R.M. 1997. An extensively modified version of MolScript that includes greatly enhanced coloring capabilities. *J. Mol. Graph. Model* **15**: 132–134, 112–133.
- Goldstein, J.M., Nelson, D., Kordula, T., Mayo, J.A., and Travis, J. 2002. Extracellular arginine aminopeptidase from *Streptococcus gordonii* FSS2. *Infect. Immun.* **70**: 836–843.
- Groll, M., Ditzel, L., Lowe, J., Stock, D., Bochtler, M., Bartunik, H.D., and Huber, R. 1997. Structure of 20S proteasome from yeast at 2.4 Å resolution. *Nature* **386**: 463–471.
- Holm, L. and Sander, C. 1998. Touring protein fold space with Dali/FSSP. *Nucleic Acids Res.* **26**: 316–319.
- Kelley, R.J., Alexander, D.L., Cowan, C., Balber, A.E., and Bangs, J.D. 1999. Molecular cloning of p67, a lysosomal membrane glycoprotein from *Trypanosoma brucei*. *Mol. Biochem. Parasitol.* **98**: 17–28.
- Koch, J., Gartner, S., Li, C.M., Quintern, L.E., Bernardo, K., Levrano, O., Schnabel, D., Desnick, R.J., Schuchman, E.H., and Sandhoff, K. 1996. Molecular cloning and characterization of a full-length complementary DNA encoding human acid ceramidase. Identification of the first molecular lesion causing Farber disease. *J. Biol. Chem.* **271**: 33110–33115.
- Lewis, A.P. and Thomas, P.J. 1999. A novel clan of zinc metallopeptidases with possible intramembrane cleavage properties. *Protein Sci.* **8**: 439–442.
- Lowe, J., Stock, D., Jap, B., Zwickl, P., Baumeister, W., and Huber, R. 1995. Crystal structure of the 20S proteasome from the archaeon *T. acidophilum* at 3.4 Å resolution. *Science* **268**: 533–539.
- Mahajan, P.B. 1984. Penicillin acylases. An update. *Appl. Biochem. Biotechnol.* **9**: 537–554.
- Makarova, K.S. and Grishin, N.V. 1999a. Thermolysin and mitochondrial processing peptidase: How far structure-functional convergence goes. *Protein Sci.* **8**: 2537–2540.
- . 1999b. The Zn-peptidase superfamily: Functional convergence after evolutionary divergence. *J. Mol. Biol.* **292**: 11–17.
- Makarova, K.S., Aravind, L., and Koonin, E.V. 2000. A novel superfamily of predicted cysteine proteases from eukaryotes, viruses and *Chlamydia pneumoniae*. *Trends Biochem. Sci.* **25**: 50–52.
- Montenegro, E., Barredo, J.L., Gutierrez, S., Diez, B., Alvarez, E., and Martin, J.F. 1990. Cloning, characterization of the acyl-CoA:6-amino penicillanic acid acyltransferase gene of *Aspergillus nidulans* and linkage to the isopenicillin N synthase gene. *Mol. Genet.* **221**: 322–330.
- Murzin, A.G., Brenner, S.E., Hubbard, T., and Chothia, C. 1995. SCOP: A structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* **247**: 536–540.
- Oinonen, C. and Rouvinen, J. 2000. Structural comparison of Ntn-hydrolases. *Protein Sci.* **9**: 2329–2337.
- Oinonen, C., Tikkanen, R., Rouvinen, J., and Peltonen, L. 1995. Three-dimensional structure of human lysosomal aspartylglucosaminidase. *Nat. Struct. Biol.* **2**: 1102–1108.
- Pei, J. and Grishin, N.V. 2001a. AL2CO: Calculation of positional conservation in a protein sequence alignment. *Bioinformatics* **17**: 700–712.
- . 2001b. Type II CAAX prenyl endopeptidases belong to a novel superfamily of putative membrane-bound metalloproteases. *Trends Biochem. Sci.* **26**: 275–277.
- Pei, J., Sadreyev, R., and Grishin, N.V. 2003. PCMA: Fast and accurate multiple sequence alignment based on profile consistency. *Bioinformatics* **19**: 427–428.
- Perez, S.E. and Steller, H. 1996. Molecular and genetic analyses of lama, an evolutionarily conserved gene expressed in the precursors of the *Drosophila* first optic ganglion. *Mech. Dev.* **59**: 11–27.
- Rawlings, N.D., O'Brien, E., and Barrett, A.J. 2002. MEROPS: The protease database. *Nucleic Acids Res.* **30**: 343–346.
- Schaffer, A.A., Aravind, L., Madden, T.L., Shavirin, S., Spouge, J.L., Wolf, Y.I., Koonin, E.V., and Altschul, S.F. 2001. Improving the accuracy of PSI-BLAST protein database searches with composition-based statistics and other refinements. *Nucleic Acids Res.* **29**: 2994–3005.
- Smith, J.L., Zaluzec, E.J., Wery, J.P., Niu, L., Switzer, R.L., Zalkin, H., and Satow, Y. 1994. Structure of the allosteric regulatory enzyme of purine biosynthesis. *Science* **264**: 1427–1433.
- Suresh, C.G., Pundle, A.V., SivaRaman, H., Rao, K.N., Brannigan, J.A., McVey, C.E., Verma, C.S., Dauter, Z., Dodson, E.J., and Dodson, G.G. 1999. Penicillin V acylase crystal structure reveals new Ntn-hydrolase family members. *Nat. Struct. Biol.* **6**: 414–416.
- Walker, D.R. and Koonin, E.V. 1997. SEALS: A system for easy analysis of lots of sequences. *Intel. Syst. for Mol. Biol.* **5**: 333–339.