Article No. jmbi.1999.3059 available online at http://www.idealibrary.com on IDEAL® J. Mol. Biol. (1999) 292, 11–17

JMB

COMMUNICATION

The Zn-peptidase Superfamily: Functional Convergence After Evolutionary Divergence

Kira S. Makarova[1,2] and Nick V. Grishin[1]*

[1]*National Center for Biotechnology Information National Library of Medicine National Institutes of Health Bethesda, MA, 20894, USA*

[2]*Department of Pathology, F.E. Hebert School of Medicine Uniformed Services University of the Health Sciences Bethesda MD 20814-4799, USA*

**Corresponding author*

Zn-dependent carboxypeptidases (ZnCP) cleave off the C-terminal amino acid residues from proteins and peptides. Here we describe a superfamily that unites classical ZnCP with other enzymes, most of which are known (or likely) to participate in metal-dependent peptide bond cleavage, but not necessarily in polypeptide substrates. It is demonstrated that aspartoacylase (ASP gene) and succinylglutamate desuccinylase (ASTE gene) are members of the ZnCP family. The Zn-binding site along with the structural core of the protein is shown to be conserved between ZnCP and another large family of hydrolases that includes mostly aminopeptidases (ZnAP). Both families (ZnCP and ZnAP) include not only proteases but also enzymes that perform N-deacylation, and enzymes that catalyze N-desuccinylation of amino acids. This is a result of functional convergence that apparently occurred after the divergence of the two families.

Keywords: protease; carboxypeptidase; aminopeptidase; aspartoacylase; succinylglutamate desuccinylase

Zn-dependent carboxypeptidases (ZnCP) with different substrate specificity and biological roles have been identified in many organisms and tissue types (Rawlings & Barrett, 1995; Osterman *et al*., 1992). Their catalytic mechanism and functions are well understood (Christianson *et al*., 1987; Stepanov, 1995), however, their evolutionary origin and relations with other proteins are poorly known. Bovine carboxypeptidase A (ZnCPA) has been the most thoroughly studied enzyme, including its spatial structure and catalytic mechanism (Christianson & Lipscomb, 1989, and references therein). The key catalytic residue Glu270 activates a water molecule for nucleophilic attack on the substrate carbonyl and subsequent proton transfer in the general base mechanism of catalysis. ZnCP uses a single zinc ion that is involved in water polarization and stabilization of the tetrahedral intermediate. The ZnCP family includes two subfamilies termed ''digestive'' and ''regulatory'' carboxypeptidases (Osterman *et al*., 1992; Rawlings & Barrett, 1995). Digestive enzymes include pancreatic carboxypeptidases of A and B types, mast cell ZnCP, and bacterial carboxypeptidases T and SG. The regulatory subfamily groups together carboxypeptidases D, E(=H), M, N, and Z which participate in processing of peptide hormones. Additionally, endopeptidase I from *Bacillus* (PROSITE PS00132/PS00133; Hourdou *et al*., 1993), and an inactivated carboxypeptidase-like domain of transcription regulator AEBP1 (Ohno *et al*., 1996; Song & Fricker, 1997; Aravind & Koonin, 1998) were shown to be homologous to ZnCPs. To clarify the evolutionary relationship of ZnCPs with other proteins we used sequence similarity search tools improved recently by utilization of profile analysis for remote homolog detection (Altschul *et al*., 1997; Altschul & Koonin, 1998; Aravind & Koonin, 1999).

Proteins of the ZnCP family (Figure 1) were studied by means of the PSI-BLAST program (Altschul *et al*., 1997). Along with the digestive and regulatory ZnCPs, endopeptidase I from *Bacillus*, and inactivated carboxypeptidase-like domain of AEBP1 that were known to belong to the ZnCP

Present address: K. S. Makarova, Institute of Cytology and Genetics, Russian Academy of Sciences, Novosibirsk 630090, Russia.

Abbreviations used: ZnCP, Zn-dependent carboxypeptidases; ZnAP, Zn-dependent aminopeptidases; ZnCPA, bovine carboxypeptidase A; ZnCAP, Zn-dependent carboxy- and aminopeptidases; ASTE, succinylglutamate desuccinylase; ASP, aspartoacylase; DAPE, succinyl-diaminopimelate desuccinylase; ACY1, aminoacylase-1; CG2, carboxypeptidase G2; PDB, Protein Data Bank.

E-mail address of the corresponding author: grishin@ncbi.nlm.nih.gov

0022-2836/99/360011–7 $30.00/0

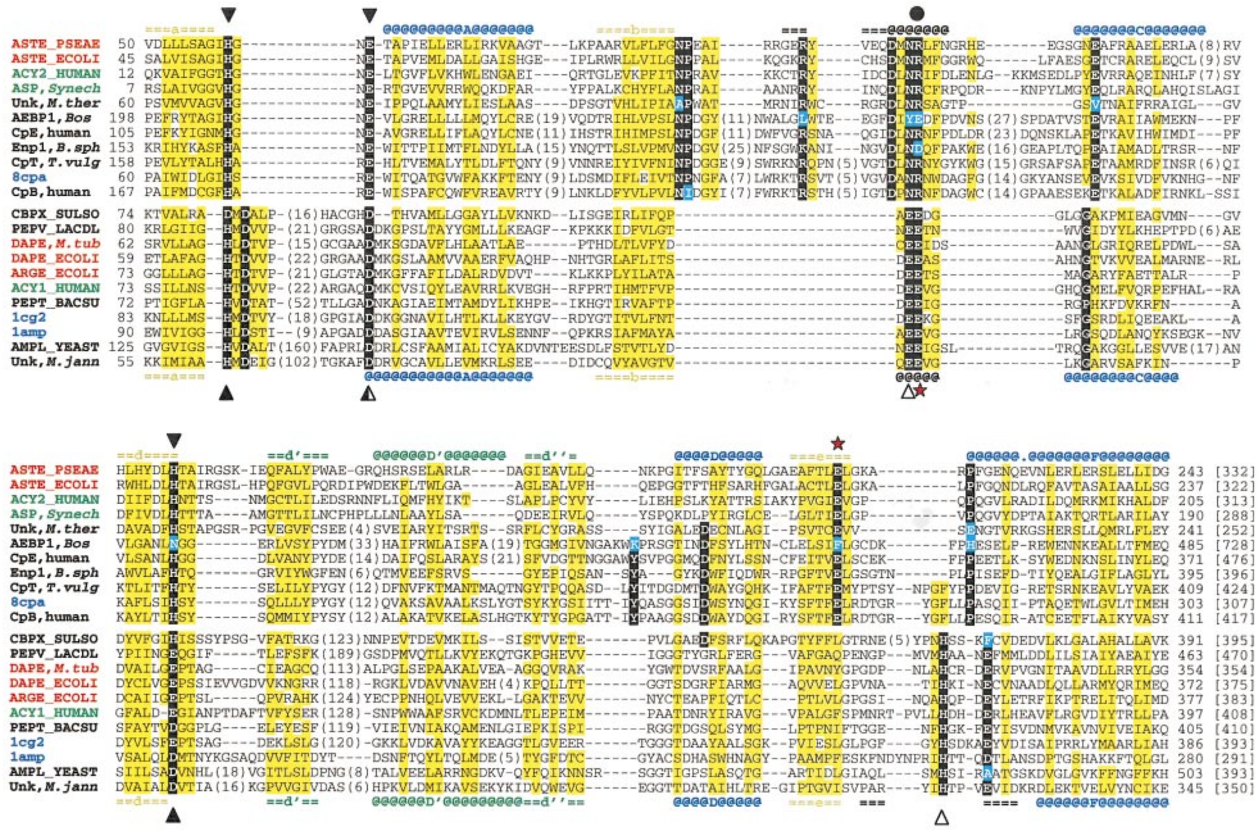

Figure 1. Sequence alignment of Zn-peptidase superfamily. The top 11 sequences (above the vertical space separator) belong to the ZnCP family and the rest are from the ZnAP family. Sequences within each family were retrieved from the NCBI non-redundant protein database (nr) with the full-length sequences gi|231122 (for ZnCP) and gi|2780967 (for ZnAP) after PSI-BLAST (release 2.0.6) iterations (Altschul *et al.*, 1997) until convergence with E-value = 0.01 and aligned with the program ALITRE (Seledtsov *et al.*, 1995). The alignments were adjusted manually to match PSI-BLAST alignments and structural consideration. The two alignments (ZnCP and ZnAP) were templated onto each other to match structure-based sequence alignment between sequences gi|494937 (8cpa) and gi|2780967 (1cg2). The structural alignments were obtained with the help of DALI (score 9.6, RMSD 3.5 Å for 188 superimposed C^{α} atoms with 13 % identity; Holm & Sander, 1995), VAST (RMSD 3.7 Å for 168 superimposed C^{α} atoms with 10.1 % identity; Madej *et al.*, 1995), and manual superposition and analysis of the structures with InsightII package (Biosym). Secondary structural elements corresponding to 8cpa (above the alignment) and to 1cg2 (below the alignment) are shown (=, β-strand residue; @, α-helical residue) and labeled with letters and color to match Figure 2. Uncharged residues in predominantly hydrophobic positions are shaded yellow. The most conserved residues within each family are shown white in black boxes, the deviating residues in these sites are boxed with blue. Symbols above and below the alignment show functionally important residues in ZnCP and ZnAP, respectively. Triangles mark the Zn-binding site (filled triangles for the first Zn^{2+}, open triangles for the second Zn^{2+}), the filled circles show the carboxylate-binding site, the star denotes the catalytic residue. Sequences correspond to the following proteins (name in Figure 1: gene bank identification number, name of the enzyme if known, organism): ASTE_PSEAE: 3913103, succinylglutamate desuccinylase, *Pseudomonas aeruginosa*; ASTE_ECOLI: 2495618, succinylglutamate desuccinylase, *Escherichia coli*; ACY2_HUMAN: 1168340, aminoacylase-2, *Homo sapiens*; ASP, *Syne*: 1652315, aspartoacylase, *Synechocystis sp*; Unk, *M.ther*: 2622366, unknown, *Methanobacterium thermoautotrophicum*; AEBP1, *Bos*: 4105170, transcription factor AEBP1, *Bos taurus*; CpE, human: 105546, carboxypeptidase E, *Homo sapiens*; Enp1, *B.sph*: 585096, endopeptidase I, *Bacillus sphaericus*; CpT, *T.vulg*: 115899, carboxypeptidase T, *Thermoactinomyces vulgaris*; 8cpa: 494937, carboxypeptidase A, *Bos taurus*, PDB entry 8cpa; CpB, human: 3915628, carboxypeptidase B, *Homo sapiens*; CBPX_SULSO: 1705668, carboxypeptidase X, *Sulfolobus solfataricus*; PEPV_LACDL: 1172069, XAA-HIS dipeptidase, *Lactobacillus delbrueckii*; DAPE, *M.tub*: 1929076, succinyl-diaminopimelate desuccinylase, *Mycobacterium tuberculosis*; DAPE_ECOLI: 118244, succinyl-diaminopimelate desuccinylase, *Escherichia coli*; ACY1_HUMAN: 461466, aminoacylase-1, *Homo sapiens*; PEPT_BACSU: 1709639, peptidase T, *Bacillus subtilis*; ARGE_ECOLI: 114142, acetylornithine deacetylase, *Escherichia coli*; 1cg2: 2780967, carboxypeptidase G2, *Pseudomonas sp*, PDB entry 1cg2; 1amp: 640150, aminopeptidase, *Vibrio proteolyticus*; PDB entry 1amp; AMPL_YEAST: 461508, leucine aminopeptidase IV, *Saccharomyces cerevisiae*; Unk, *M.jann*: 2494340, hypothetical protein, *Methanococcus jannaschii*. The first and the last residue numbers are given. Sequence names of deacylases and desuccinylases are colored green and red, respectively. Names of proteins with determined structure are shown in blue. The alignment includes only selected regions of representative sequences with the numbers of omitted residues in parentheses. The complete alignment is available at ftp://ncbi.nlm.nih.gov/pub/grishin/ZnCAP.

family, we retrieved sequences of succinylglutamate desuccinylase (ASTE gene; Itoh, 1997), aspartoacylase (ASP gene; Kaul *et al*., 1993) and of several archaeal proteins of unknown function. Iterative PSI-BLAST searches were initiated with the bovine ZnCPA sequence. The most divergent sequences found in the course of iterations and statistically supported by *E*-value of at least 0.01, were used as queries for new PSI-BLAST searches to ensure the complete coverage. Subsequent analysis of sequence alignments revealed that the Zn-binding site (His69, Glu72, His169, bovine carboxypeptidase A numbering throughout), carboxylate-binding determinant (Arg145, Asp in endopeptidase I), and catalytic residue (Glu270) are conserved in all these proteins with the exception of the inactivated AEBP1 (Figure 1, ZnCP family). Such conservation strongly suggests that these enzymes are homologous and share a common catalytic mechanism. Several other largely invariant residues with less clear functional roles are shared by the members of the family as well. These include Asn112, Arg130, Asp142, and Arg175 which are placed around the active site in the ZnCPA structure (Figures 1 and 2) and participate in shaping of the substrate-binding cavity.

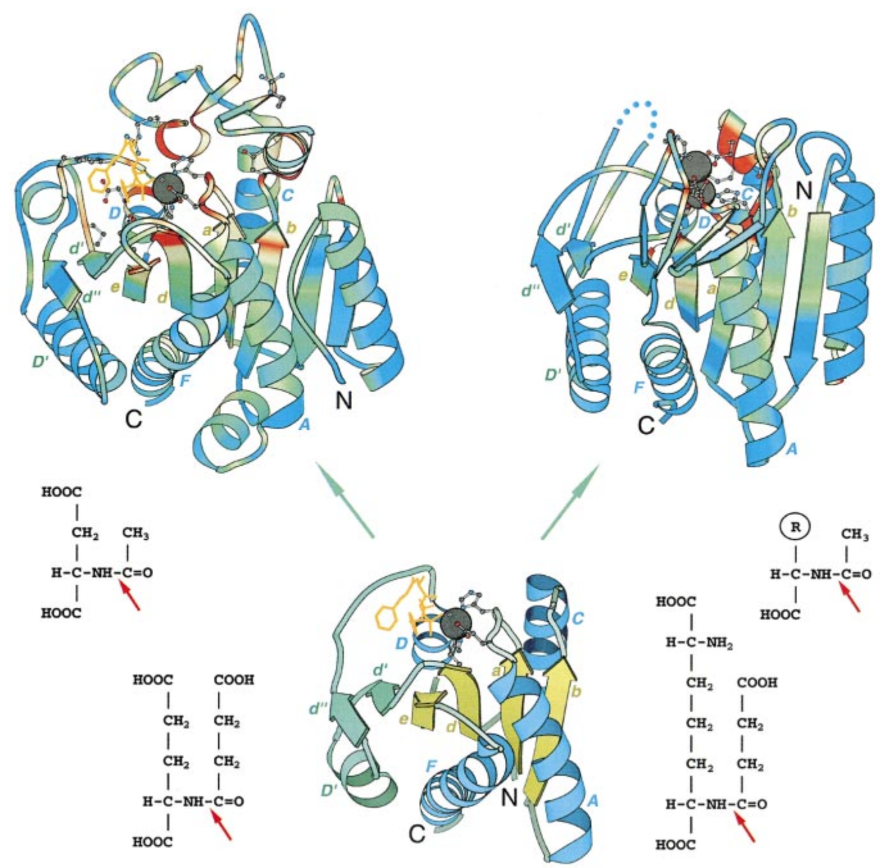

Figure 2. Structure-function comparison of ZnCP and ZnAP families. Top left: carboxypeptidase A (PDB entry 8cpa); top right: an enzyme from aminopeptidase family called carboxypeptidase G2 (PDB entry 1cg2). The structures were superimposed and then separated for clarity. Ribbon diagrams are rainbow-colored by sequence conservation in a multiple sequence alignment of each family (ZnCP and ZnAP). Red corresponds to the highest conservation and blue corresponds to the lowest conservation. Dark-gray balls show zinc ions and the side-chains of the conserved residues boxed in the alignment (Figure 1) are displayed in ball-and-stick presentation. Inhibitor present in carboxypeptidase A structure is shown in orange. Dots in carboxypeptidase G2 structure stand for the not shown inserted domain of about 100 amino acid residues. N and C termini and core secondary structure elements are labeled. The letter color for the secondary structure element corresponds to the color of the element in the alignment on Figure 1. Bottom middle: the "minimal structure" of a possible ancestor of the superfamily. The carboxypeptidase A coordinates were used to generate the structure. Only the structural elements present in all of the sequences in the alignment of the ZnCP/AP superfamily are shown. The Rossman fold type 2 + 2 β-sheet is shown in yellow and α-helices are blue, a right-handed βαβ-unit (*d′D′d″*) inserted between β-stand *d* and α-helix *D* is colored green. The molecules are rendered with BOBSCRIPT (Esnouf, 1997) a modified version of MOLSCRIPT (Kraulis, 1991). Bottom left: succinyl-glutamate (bottom) and *N*-acetyl-L-aspartic acid (top), substrates for the enzymes from ZnCP family; bottom right: succinyl-diaminopimelate (bottom) and *N*-acetyl-L-amino acid (top), substrates for the enzymes from ZnAP family. Red arrows point at the scissile bond.

The conserved Pro282 residue participates in the type I reverse turn. Additionally, characteristic hydrophobicity patterns corresponding to α-helices and β-strands and secondary structure predictions (Rost & Sander, 1994, data not shown) supported our conclusion about the structural similarity of the proteins found in PSI-BLAST iterations (Figure 1, ZnCP family).

Succinylglutamate desuccinylase (ASTE gene) catalyzes the last step of the arginine succinyltransferase (AST) pathway, namely peptide bond cleavage in succinylglutamate to yield succinate and glutamate (Schneider *et al.*, 1998; Itoh, 1997). The AST pathway allows some bacteria, for example *Escherichia coli*, to use arginine as the only nitrogen source (Schneider *et al.*, 1998). Our analysis predicts the structure and catalytic mechanism of the succinylglutamate desuccinylase to be similar to these established for ZnCPA.

Aspartoacylase (ASP gene; Kaul *et al.*, 1993) releases acetate from *N*-acetylaspartic acid. Missense mutation in the ASP gene causes Canavan disease in humans, which involves spongy degeneration of brain cerebral white matter due to accumulation of *N*-acetylaspartate (Kaul *et al.*, 1993). It has been suggested that aspartoacylase is a serine hydrolase which belongs to the α/β-hydrolase superfamily (Kaul *et al.*, 1993). However, there is no invariant serine residue which could be a candidate for the catalytic residue in the alignment of rather similar aspartoacylase proteins from the three species. No detectable sequence similarity exists between aspartoacylase and any of the proteins from α/β-hydrolase superfamily. In contrast, we found statistically significant sequence similarity between aspartoacylases and ZnCPs accompanied by the conservation of the active-site residues and predicted secondary structural elements. When the sequence gi|1524298 is used as a query, it finds ASTE (gi|4106582) with an *E*-value of 0.009 in the BLAST search, and the first carboxypeptidase sequence (gi|3641621) appears above the 0.01 threshold after the fourth PSI-BLAST iteration with the *E*-value of 0.002 and score of 44. Therefore the sequence gi|1524298 is a link that demonstrates statistically supported sequence similarity between carboxypeptidases and deacylases/desuccinylases. We conclude that aspartoacylase is a metal-dependent hydrolase with the structure and catalytic mechanism similar to those established for carboxypeptidase A. This identification hopefully will facilitate the study of this enzyme essential for brain development.

Digestive eukaryotic carboxypeptidases of ZnCP family are known to be expressed as zymogens with the N-terminal 95-residue "pro"-domain which inhibits enzymatic activity and is involved in activation (Garcia-Saez *et al.*, 1997, and references therein) and perhaps in folding (Phillips & Rutter, 1996). We were not able to find the homolog of this domain in any other members of ZnCP family. Regulatory carboxypeptidases possess a short pro-segment at the N terminus; aspartoacylases and desuccinylases do not seem to have it at all. Alternatively, aspartoacylase and desuccinylase sequences contain a 90-amino acid residue domain at the C terminus. The function of these domains remains unknown.

Our finding that ASTE and ASP are homologs of Zn-carboxypeptidases gives a new perspective on the ZnCP family that was known to include carboxypeptidases which are specific towards positively charged or hydrophobic C-terminal amino acid residues. ASTE and ASP represent enzymes from this family specific to negatively charged C-terminal amino acid residues. Additionally, the peptide bonds cleaved by ASTE and ASP do not connect amino acids, therefore these enzymes are not proteases in the classical sense. We demonstrate that ZnCP is an ancient diverse family that covers all three domains of life: bacteria, archaea, and eukaryotes.

In addition to ASTE and ASP genes, another family of metal-dependent desuccinylases (DAPE gene; Bouvier *et al.*, 1992) and aminoacylases (ACY1 gene; Kaul *et al.*, 1993) has been described. They belong to the PROSITE entry PS00758/PS00759 along with some other enzymes which include carboxypeptidase G2 (CG2) with the available structure (Rowsell *et al.*, 1997; Boyen *et al.*, 1992; Protein Data Bank (PDB) entry 1cg2). The PSI-BLAST sequence analysis of these protein sequences revealed their close relationship with bacterial aminopeptidases, such as the aminopeptidase from *Aeromonas proteolytica* whose structure has been determined (Chevrier *et al.*, 1994; PDB entry 1amp). A structural comparison (1cg2 *versus* 1amp), first reported by Rowsell *et al.* (1997), supports this conclusion giving a DALI *Z*-score of 25.4 with an RMSD 2.4 Å between 242 C^{α} atoms, and VAST *P*-value of 4×10^{-7} with an RMSD 3.2 Å between 232 C^{α} atoms of the two proteins. However, Rowsell *et al.* (1997) stated that 1cg2 and 1amp show no detectable amino acid sequence similarity. We readily find sequence similarity between 1cg2 and 1amp statistically supported by the PSI-BLAST program (Altschul *et al.*, 1997). When 1amp is used as a query, 1cg2 is retrieved above the 0.001 threshold on the second PSI-BLAST iteration with the score 69 bits and an *E*-value 4×10^{-11}. We designate the family that unites DAPE, ACY1, CG2, and several subfamilies of aminopeptidases as ZnAP. It corresponds to the clan MH (http://www.bi.bbsrc.ac.uk/Merops/indexes/families.htm) described by Rawlings & Barrett (1995) (Figure 1).

The ZnCP and ZnAP families do not share sequence similarity detectable by PSI-BLAST program. To clarify the relationship between these families we utilized structural comparisons. Statistically significant structural similarity spans the entire molecules of ZnCPA (PDB entry 8cpa) and CG2 as shown by DALI (*Z*-score of 9.6) and VAST (*P*-value of 2×10^{-4}) (Figure 2) and reported by Rowsell *et al.* (1997). DALI and VAST algorithms use neither sequence information nor functional information on Zn-binding and substrate-binding sites. Nevertheless, structural alignments lead to the superposition of the active sites in ZnCPA and CG2. The two alignments (ZnCP and ZnAP) were merged on the basis of the structure-based sequence alignment of 8cpa and 1cg2. The resulting alignment (Figure 1) shows resemblance of hydrophobicity patterns and conservation of the Zn-binding site. The structural similarity together with the superposition of the active sites and Zn and substrate-binding determinants (Murzin, 1998; Rowsell *et al*., 1997) strongly indicate homology of ZnCP and ZnAP families. We designate the resulting superfamily ZnCAP (Zn-dependent carboxy and amino peptidases). It should be noted that bovine leucine aminopeptidase (AMPL, PDB entry 1lap) has been reported to share extensive structural resemblance to carboxypeptidases A (Artymiuk *et al*., 1992) and CG2 (Rowsell *et al*., 1997) and evidently belongs to this superfamily. However, none of the AMPL family members which acquired other than proteolytic functions have been found yet (Rawlings & Barrett, 1995; http://www.bi.bbsrc.ac.uk/Merops/indexes/families.htm). We conducted PSI-BLAST searches on the sequences of AMPL family to find out that we are not able to detect other proteins but aminopeptidases, and therefore this family is not considered here.

From the complete alignments of ZnCAP, the "minimal core" of the molecule was delineated to include the regions present in all sequences from the alignment. The model of the minimal ZnCAP based on the ZnCPA structure is presented in Figure 2. It illustrates that the core of the ZnCAP domain is formed by a four-stranded parallel β-sheet sandwiched by four α-helices (two on each side of the β-sheet), and represents a Rossman fold-like structure with two β-strands before the crossing helix C (Figure 2) and two β-strands after it (Rossman 2 + 2). This core incorporates an insertion of a parallel βαβ-unit, which participates in the formation of the primary specificity pocket in ZnCP (Figure 2). Our minimal core differs from the "conserved secondary structural elements" delineated by Rowsell *et al*. (1997) covering smaller number of the elements (11 *versus* 15 reported by Rowsell *et al*., 1997), since all available sequences, and not only the four representative structures of ZnCAP as reported by Rowsell *et al*. (1997), were analyzed in the present study.

The ZnCP and ZnAP families have a Zn-binding site with the structurally equivalent ligands His69, Glu72, and His169 in ZnCP and His112, Asp141, and Glu200/Asp in ZnAP (CG2 numbering here and throughout) (Figure 1). Interestingly, only the first ligand is invariant, the others are replaced by functionally similar amino acids. In addition to this zinc ion, the ZnAP family utilizes a second Zn coordinated by Asp141 (ligand of both Zn), Glu176 and His385. The position of the Glu141 chelating the second Zn ion in ZnAP maps to the carboxylate-binding region in ZnCP. Other differences between ZnCP and ZnAP include the location of the catalytic residue, which is a glutamate in both families (Figure 1). In ZnCP it is placed close to the C terminus and in ZnAP it is in the middle of the molecule. Notably, Glu142 in ZnAP is aligned with Arg145, which binds substrate carboxylate in ZnCP and is not needed in aminopeptidases. Another interesting correspondence is between His363, which is a ligand of the second Zn ion in ZnAP, and its structural equivalent Phe279 in ZnCP, which is utilized in the secondary specificity (S1) pocket. Thus, functionally important residues map to the same regions in the structures, but their exact roles differ from one family to another depending on the enzyme specificity.

Each of the two families (ZnCP and ZnAP) in the ZnCAP superfamily includes classical proteases, desuccinylases, and deacylases (Figure 3). It is likely that the common ancestor of all these proteins was a single Zn-dependent non-specific peptide bond hydrolase, which functioned as a protease and a hydrolase of N-modified amino acids. The primary step in the emergence of the two families included utilization of the second zinc ion in catalysis by the ancestor of ZnAP. The further divergence within each family led to speciation into proteases and desuccinylases/deacylases (Figure 3). This divergence was apparently independent for the two families and, therefore, is an example of the functional convergence.

In conclusion, the sequence and structural motif which covers the Zn-binding site with the surrounding conserved core regions, thus defining the superfamily of ZnCAP, is described. This large superfamily includes both exo- (carboxy-, amino-) and endopeptidases along with deacylases and desuccinylases. The motif delineation results in prediction of spatial structure and catalytic mechanism for aspartoacylase and succinylglutamate desuccinylase. The evolution of the ZnCP and ZnAP families within the ZnCAP superfamily seem to have involved functional convergence. Most of the studies dealing with functional convergence refer to evolutionary unrelated proteins (Voet & Voet, 1990). For example, the Ser-His-Asp catalytic triad evolved independently in chymotrypsin and subtilisin, proteins with drastically different structures (Russell, 1998). Here we describe a different situation, namely two families of enzymes that are distantly related (Rowsell *et al*., 1997) but in all likelihood have evolved certain distinct specificities independently (in parallel) subsequent to their divergence from the common ancestor. The functional speciation by the substrate and reaction specificity (protease, deacylase and desuccinylase) is achieved independently by the enzymes in ZnCP and ZnAP families and, therefore, is convergent. The structural similarity between ZnCP and ZnAP proteins is proposed to be a reflection of their divergence from the common ancestor.

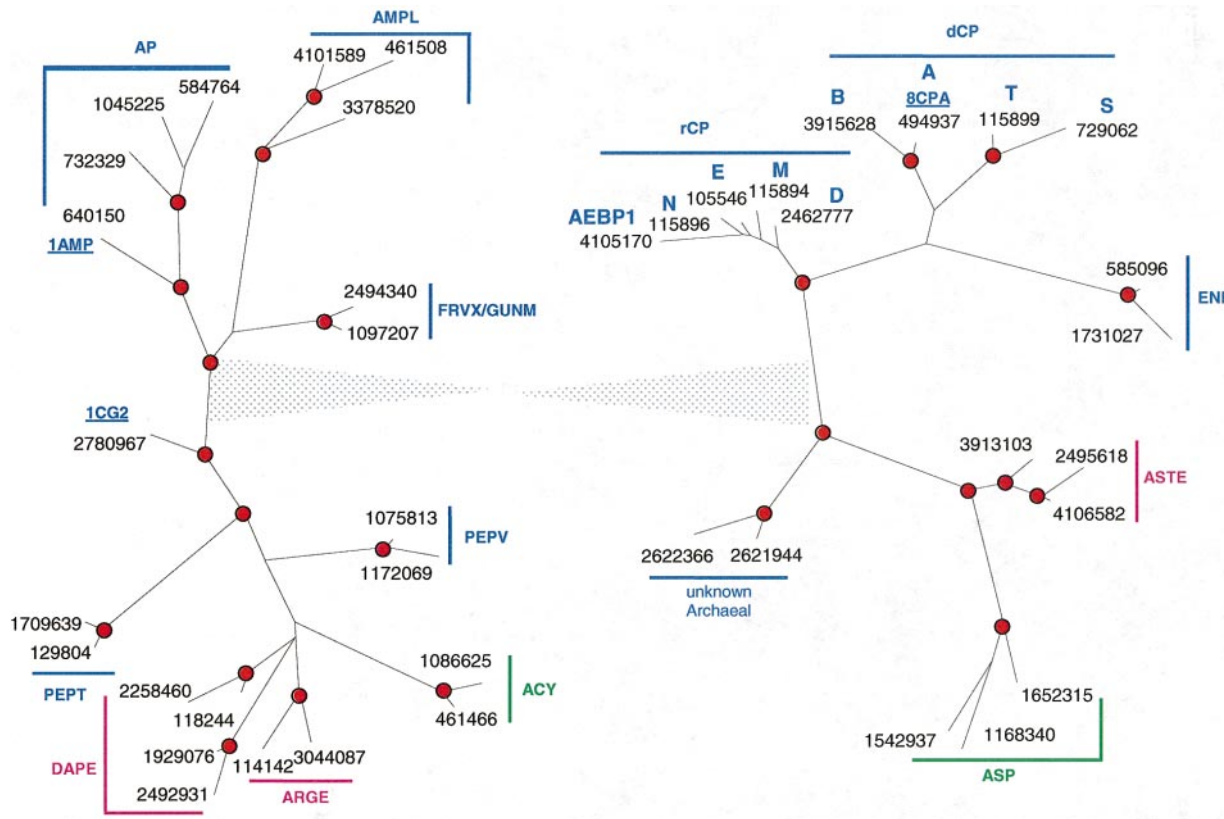

Figure 3. Evolutionary tree of the Zn-peptidase superfamily. The distance tree for each family (ZnCP and ZnAP) was built with the programs scor2tree (Y.I. Wolf and N.V.G., unpublished results) and PHYLIP (Felsenstein, 1996). Large gray triangles show the approximate position of the root when one tree is used to root the other. The precise rooting is impossible due to low similarity between the sequences. Red dots show the bifurcations with bootstrap values more than 70 % after 1000 replications. Most subfamilies are marked with the blue lines and gene names where available. Desuccinylases are outlined with pink and deacylases are shown in green. The following sequences were added into the tree: 4106582, aruE, succinylglutamate desuccinylase *Yersinia pestis*; 1542937, aspartoacylase (ASP), *Prochlorococcus marinus*; 2621944, unknown, *Methanobacterium thermoautotrophicum*; 115896, carboxypeptidase N, *Homo sapiens*; 115894, carboxypeptidase M, *Homo sapiens*; 2462777, carboxypeptidase D, *Homo sapiens*; 729062, carboxypeptidase S, *Streptomyces capreolus*; 1731027, YQGT, *Bacillus subtilis*; 3044087; argE, acetylornithine deacetylase *Myxococcus xanthus*; 1929076, dapE succinyl-diaminopimelate desuccinylase *Mycobacterium tuberculosis*; 2258460, succinyl-diaminopimelate desuccinylase, *Helicobacter pylori*; 2492931, succinyl-diaminopimelate desuccinylase, *Corynebacterium glutamicum*; 1086625, aminoacylase-1, *Caenorhabditis elegance*; 129804, PEPT_SALTY, peptidase T, *Salmonella typhimurium*; 1075813, hypothetical protein, *Bacillus stearothermophilus*; 584764, APE3_YEAST, aminopeptidase Y, *Saccharomyces cerevisae*; 1045225, *N*-acetylpuromycin *N*-acetylhydrolase, *Streptomyces anulatus*; 732329, YWAD, *Bacillus subtilis*; 3378520, aminopeptidase, *Themotoga neapolitana*; 4101589, aspartyl aminopeptidase, *Homo sapiens*; 1097207, endoglucanase, *Clostridium thermocellum*. Other sequence name abbreviations are explained in the legend to Figure 1.

Acknowledgments

We thank Eugene V. Koonin for helpful discussions, constructive suggestions and critical reading of the manuscript, and Yuri I. Wolf for critical remarks and stimulating discussions. K.S.M. is supported by DOE OBER grant DE-FG02-98ER62583.

References

Altschul, S. F. & Koonin, E. V. (1998). Iterated profile searches with PSI-BLAST–a tool for discovery in protein databases. *Trends Biochem. Sci.* **23**, 444-447.

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucl. Acids Res.* **25**, 3389-3402.

Aravind, L. & Koonin, E. V. (1998). Eukaryotic transcription regulators derive from ancient enzymatic domains. *Curr. Biol.* **8**, R111-R113.

Aravind, L. & Koonin, E. V. (1999). Gleaning non-trivial structural, functional and evolutionary information about proteins by iterative database searches. *J. Mol. Biol.* **287**, 1023-1040.

Artymiuk, P. J., Grindley, H. M., Park, J. E., Rice, D. W. & Willett, P. (1992). Three-dimensional structural resemblance between leucine aminopeptidase and carboxypeptidase A revealed by graph-theoretical techniques. *FEBS Letters*, **303**, 48-52.

Bouvier, J., Richaud, C., Higgins, W., Bogler, O. & Stragier, P. (1992). Cloning, characterization, and expression of the dapE gene of *Escherichia coli*. *J. Bacteriol.* **174**, 5265-5271.

Boyen, A., Charlier, D., Charlier, J., Sakanyan, V., Mett, I. & Glansdorff, N. (1992). Acetylornithine deacetylase, succinyldiaminopimelate desuccinylase and carboxypeptidase G2 are evolutionarily related. *Gene*, **116**, 1-6.

Chevrier, B., Schalk, C., D'Orchymont, H., Rondeau, J. M., Moras, D. & Tarnus, C. (1994). Crystal structure of *Aeromonas proteolytica* aminopeptidase: a prototypical member of the co-catalytic zinc enzyme family. *Structure*, **2**, 283-291.

Christianson, D. W. & Lipscomb, W. B. (1989). Carboxypeptidase-A. *Acc. Chem. Res.* **22**, 62-69.

Christianson, D. W., David, P. R. & Lipscomb, W. N. (1987). Mechanism of carboxypeptidase A: hydration of a ketonic substrate analogue. *Proc. Natl Acad. Sci. USA*, **84**, 1512-1515.

Esnouf, R. (1997). An extensively modified version of MolScript that includes greatly enhanced coloring capabilities. *J. Mol. Graph. Model.* **15**, 133-138.

Felsenstein, J. (1996). Inferring phylogenies from protein sequences by parsimony, distance, and likelihood methods. *Methods Enzymol.* **266**, 418-427.

Garcia-Saez, I., Reverter, D., Vendrell, J., Aviles, F. X. & Coll, M. (1997). The three-dimensional structure of human procarboxypeptidase A2. Deciphering the basis of the inhibition, activation and intrinsic activity of the zymogen. *EMBO J.* **16**, 6906-6913.

Holm, L. & Sander, C. (1995). Dali: a network tool for protein structure comparison. *Trends Biochem. Sci.* **20**, 478-480.

Hourdou, M. L., Guinand, M., Vacheron, M. J., Michel, G., Denoroy, L., Duez, C., Englebert, S., Joris, B., Weber, G. & Ghuysen, J. M. (1993). Characterization of the sporulation-related gamma-D-glutamyl-(L)meso-diaminopimelic-acid-hydrolysing peptidase I of *Bacillus sphaericus* NCTC 9602 as a member of the metallo(zinc) carboxypeptidase A family. Modular design of the protein. *Biochem. J.* **292**, 563-570.

Itoh, Y. (1997). Cloning and characterization of the aru genes encoding enzymes of the catabolic arginine succinyltransferase pathway in *Pseudomonas aeruginosa*. *J. Bacteriol.* **179**, 7280-7290.

Kaul, R., Gao, G. P., Balamurugan, K. & Matalon, R. (1993). Cloning of the human aspartoacylase cDNA and a common missense mutation in Canavan disease. *Nature Genet.* **5**, 118-123.

Kraulis, P. (1991). Molscript: a program to produce both detailed and schematic plots structures. *J. Appl. Crystallog.* **24**, 946-950.

Madej, T., Gibrat, J. F. & Bryant, S. H. (1995). Threading a database of protein cores. *Proteins: Struct. Funct. Genet.* **23**, 356-369.

Murzin, A. G. (1998). How far divergent evolution goes in proteins. *Curr. Opin. Struct. Biol.* **8**, 380-387.

Ohno, I., Hashimoto, J., Shimizu, K., Takaoka, K., Ochi, T., Matsubara, K. & Okubo, K. (1996). A cDNA cloning of human AEBP1 from primary cultured osteoblasts and its expression in a differentiating osteoblastic cell line. *Biochem. Biophys. Res. Commun.* **228**, 411-414.

Osterman, A. L., Grishin, N. V., Smulevitch, S. V., Matz, M. V., Zagnitko, O. P., Revina, L. P. & Stepanov, V. M. (1992). Primary structure of carboxypeptidase T: delineation of functionally relevant features in Zn-carboxypeptidase family. *J. Protein Chem.* **11**, 561-570.

Phillips, M. A. & Rutter, W. J. (1996). Role of the prodomain in folding and secretion of rat pancreatic carboxypeptidase A1. *Biochemistry*, **35**, 6771-6776.

Rawlings, N. D. & Barrett, A. J. (1995). Evolutionary families of metallopeptidases. *Methods Enzymol.* **248**, 183-228.

Rost, B. & Sander, C. (1994). Combining evolutionary information and neural networks to predict protein secondary structure. *Proteins: Struct. Funct. Genet.* **19**, 55-72.

Rowsell, S., Pauptit, R. A., Tucker, A. D., Melton, R. G., Blow, D. M. & Brick, P. (1997). Crystal structure of carboxypeptidase G2, a bacterial enzyme with applications in cancer therapy. *Structure*, **5**, 337-347.

Russell, R. B. (1998). Detection of protein three-dimensional side-chain patterns: new examples of convergent evolution. *J. Mol. Biol.* **279**, 1211-1227.

Schneider, B. L., Kiupakis, A. K. & Reitzer, L. J. (1998). Arginine catabolism and the arginine succinyltransferase pathway in *Escherichia coli*. *J. Bacteriol.* **180**, 4278-4286.

Seledtsov, I. A., Vul'f, Iu. I. & Makarova, K. S. (1995). Multiple alignment of biopolymer sequences, based on the search for statistically significant common segments. *Mol. Biol. (Mosk.)*, **29**, 1023-1039.

Song, L. & Fricker, L. D. (1997). Cloning and expression of human carboxypeptidase Z, a novel metallocarboxypeptidase. *J. Biol. Chem.* **272**, 10543-10550.

Stepanov, V. M. (1995). Carboxypeptidase T. *Methods Enzymol.* **248**, 675-683.

Voet, D. & Voet, J. G. (1990). *Biochemistry*, pp. 377, John Wiley & Sons, Inc., New York, Chichester, Brisbane, Toronto, Singapore.

Edited by J. M. Thornton

(Received 21 May 1999; received in revised form 13 July 1999; accepted 20 July 1999)