Minireview

Expansion of the zinc metallo-hydrolase family of the β-lactamase fold

Hiromi Daiyasu[a], Kazuya Osaka[a], Yoshizumi Ishino[b], Hiroyuki Toh[a,*]

[a] *Department of Bioinformatics, Biomolecular Engineering Research Institute, 6-2-3, Furuedai, Suita, Osaka 565-0874, Japan*
[b] *Department of Molecular Biology, Biomolecular Engineering Research Institute, 6-2-3, Furuedai, Suita, Osaka 565-0874, Japan*

Received 21 May 2001; revised 9 July 2001; accepted 9 July 2001

First published online 23 July 2001

Edited by Hans Eklund

Abstract Recently, the zinc metallo-hydrolase family of the β-lactamase fold has grown quite rapidly, accompanied by the accumulation of sequence and structure data. The variety of the biological functions of the family is higher than expected. In addition, the members often have mosaic structures with additional domains. The family includes class B β-lactamase, glyoxalase II, arylsulfatase, flavoprotein, cyclase/dehydrase, an mRNA 3′-processing protein, a DNA cross-link repair enzyme, a DNA uptake-related protein, an alkylphosphonate uptake-related protein, CMP-*N*-acetylneuraminate hydroxylase, the romA gene product, alkylsulfatase, and insecticide hydrolases. In this minireview, the functional and structural varieties of the growing protein family are described.

Key words: Zinc ion; Iron ion; Metallo-hydrolase; β-Lactamase fold

1. Introduction

Class B β-lactamase is a eubacterial zinc metallo-hydrolase, which is involved in the resistance against β-lactams. With the recent accumulation of sequence and structure data, many relatives of class B β-lactamase have been identified. The relatives constitute a diverse protein family. The sequences of the family members are highly diverged, and the members are involved in different biological functions. As described below, some members are related to mRNA processing, genetic transformation, and DNA repair. However, the members are characterized by the same folding pattern and the conserved sequence motifs. Here, the family is referred to as the zinc metallo-hydrolase family of the β-lactamase fold, or more simply, as the zinc metallo-hydrolase family. The members are distributed over three domains of living organisms, Eukarya, Archaea, and Bacteria. Due to the rapid growth of the family, however, no review has covered all of the family members. The object of this minireview is to describe the functional and structural varieties of the zinc metallo-hydrolase family members with currently available sequence and/or structure data. We collected the members from the published literature and by database searching. The collected members included many hypothetical proteins and ORF products. Therefore, we neglected the hypothetical proteins, and focused on the members with known functions.

2. Classification of the family

We classified the members of the zinc metallo-hydrolase family into 16 groups, based on the biological functions. Therefore, the classification does not always reflect the phylogenetic relationship. The members of Groups 1–7 were already identified as the constituents of the family, according to the sequence and/or structure similarity. In addition, we found that the members of the remaining groups also belong to the zinc metallo-hydrolase family, by database searching with PSI-BLAST [1], although some of them have been annotated as metallo-hydrolases in the COG database [2]. A multiple alignment of the 16 groups is shown in Fig. 1. Due to the high sequence divergence, only the alignments of five conserved segments are shown. The spatial arrangement of the conserved residues in the segments is shown in Fig. 2. Instead of citing the papers for the data, each ID code and the corresponding sequence or structure database for its data are shown in Fig. 1. Some members of the protein family have characteristic domain structures, even though they are classified in the same group. Schematic diagrams of the domain structures of representative proteins of the 16 groups are shown in Fig. 3.

Group 1 consists of the Class B β-lactamases (EC 3.5.2.6). The enzymes are about 250 amino acid residues in length, and act on carbon–nitrogen bonds to hydrolyze β-lactams [4]. The production of the enzyme is regarded as a major mechanism for eubacteria to acquire resistance against β-lactams. Some β-lactamases include only one zinc ion per molecule, while others require a binuclear active site [4]. Among the members of this family, the crystal structure of the β-lactamase from *Bacillus cereus* was first solved [5]. Subsequently, the tertiary structures of the enzymes from other species were determined [6–8].

Group 2 consists of the glyoxalases II (EC 3.1.2.6). Glyoxalase II shows weak but significant sequence similarity to class B β-lactamase [9,10]. Subsequently, the X-ray crystallographic study of glyoxalase II revealed that its N-terminal region, which is about 200 amino acid residues in length, has a tertiary structure similar to that of β-lactamase [11]. The glyoxalase II enzymes are about 250 amino acid residues in length. Glyoxalase II catalyzes the hydrolysis of the thioester of *S*-D-lactoglutathione to produce glutathione and D-lactic acid, and

*Corresponding author. Fax: (81)-6-6872 8219.
E-mail address: toh@beri.co.jp (H. Toh).

PII: S0014-5793(01)02686-2

the enzyme includes two zinc ions per molecule for the catalytic activity [11,12].

Group 3 is constituted by the flavoproteins and rubredoxin oxygen:oxidoreductase (ROO), which show sequence similarity to class B β-lactamase and the relatives [9]. Recently, the tertiary structure of ROO from *Desulfovibrio gigas* was determined by an X-ray crystallographic study [13], which revealed that the enzyme consists of two domains, a β-lactamase-like domain and a flavodoxin-like domain (see Fig. 3). A homodimer is formed by a pair of ROOs for a terminal reaction in a unique oxygen scavenging pathway that enables the anaerobe to survive under oxic conditions [14]. Unlike class B β-lactamase and glyoxalase II, ROO has two iron ions per molecule, and functions as a dehydrase [13].

Group 4 consists of the arylsulfatases and their relatives. Eubacterial arylsulfatase (EC 3.1.6.1) and its relative, called ELAC, show weak but significant sequence similarity to class B β-lactamase [9,10]. Arylsulfatase from *Alteromonas carrageenovora* [15] is an enzyme of about 300 amino acid residues in length, which acts on ester bonds to exert a sulfuric ester hydrolase activity. The primary structure of ELAC is basically the same as that of arylsulfatase. Recently, a new member of this group was identified [16], which is called ELAC2 (AF304370 and AF215894 in Fig. 1). ELAC2 has an additional domain in the N-terminal region (see Fig. 3). The N-terminal domain also shows weak sequence similarity to the members of the zinc metallo-hydrolase family. That is, ELAC2 has a two-fold tandem repeat structure. However, the sequence is highly diverse, and most of the conserved residues are degraded. Therefore, the N-terminal region of ELAC2 is not included in Fig. 1. Mammalian ELAC2 is considered to be involved in prostate cancer [16], while the *Drosophila* counterpart is induced by the juvenile hormone [17].

The members of Group 5 are the type II polyketide synthases, which are involved in the production of actinorhodin, granaticin, griseusin, monensin, mithramycin, etc [18]. The enzymes have been identified from various species of *Streptomyces*. The enzymes show sequence similarity to β-lactamase [19] and the relatives [9]. The type II polyketide synthase is a bifunctional enzyme, which shows both cyclase activity [20] and dehydrase activity [19,21]. The latter is also called aromatase activity.

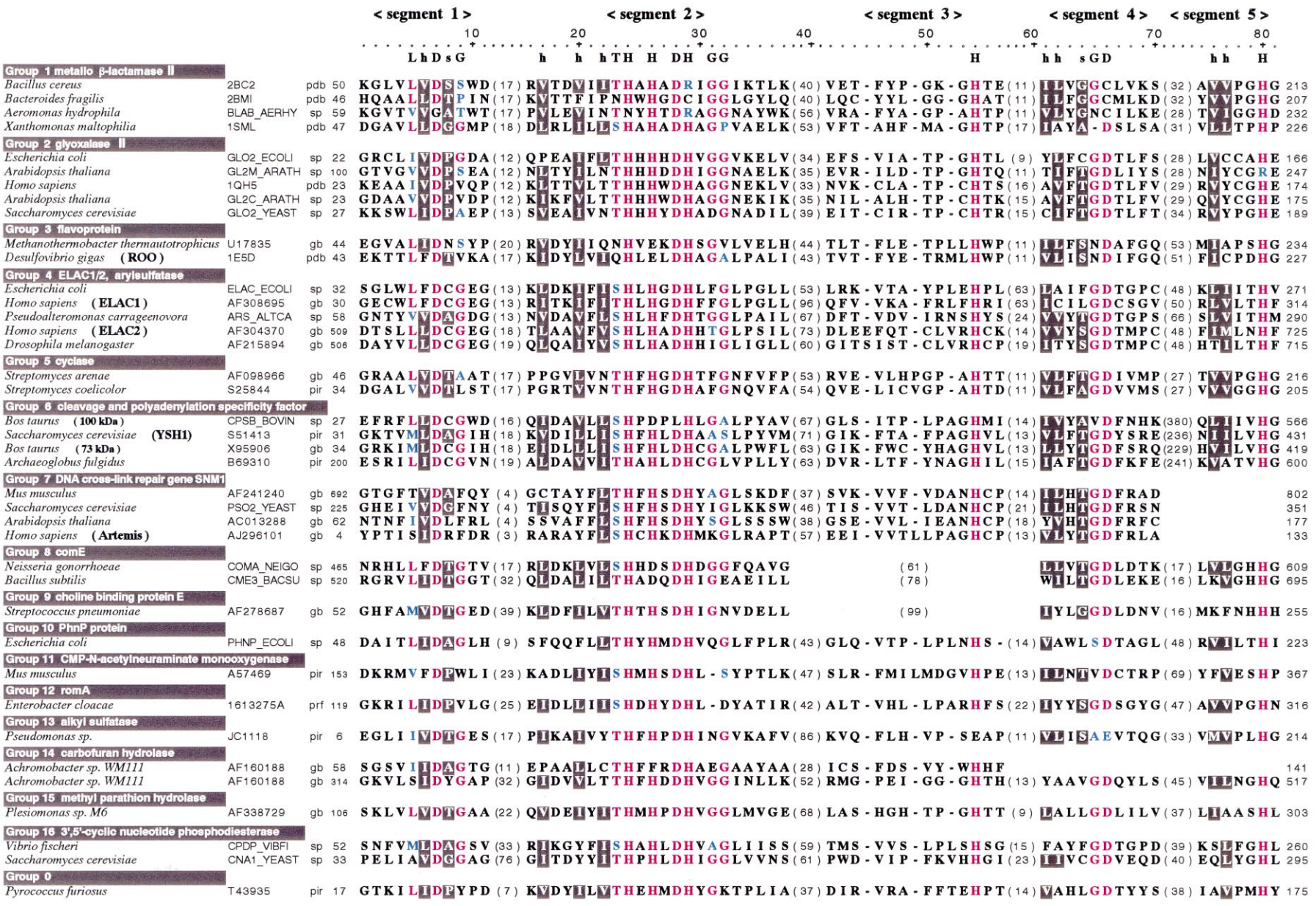
Group 6 consists of proteins involved in the processing of mRNA 3′-ends. This group includes the 100-kDa [22] and 73-kDa [23] subunits of the mammalian cleavage and polyadenylation specific factor, and a subunit of the yeast polyadenylation factor I [23]. The functional role of the subunits in the mRNA processing is still unknown. Tavtigian et al. found that the mRNA-processing factors show sequence similarity to ELAC2 [16]. The zinc metallo-hydrolase domain of the RNA-processing proteins is larger than those of the other members, due to a long insertion between segments 4 and 5 (see Figs. 1 and 3). In addition to the eukaryotic members, many ORF products derived from Archaea are also included in the group. The archaeal ORF products are annotated as zinc metallo-hydrolases in the COG database [2]. Some archaeal ORF products have an extended region at the N-terminus, where an RNA-binding motif known as the KH-domain [24] is present (see Fig. 3).

Group 7 includes a set of SNM1 and its relatives. SNM1 is an enzyme involved in the repair of DNA interstrand cross-links. SNM1 shows sequence similarity to class B β-lactamase and the relatives [9,16]. DNA interstrand cross-links induce lethal DNA damage, because they block replication, transcription, and segregation of the DNA. SNM1 has been identified in yeast [25] and mammals [26]. However, its relatives have been found from other species within Eukarya and Archaea. It is known that both the fungal and mammalian SNM1s have a zinc finger domain upstream of the β-lactamase-like domain [25,26] (see Fig. 3). However, the linker regions connecting the zinc finger and the β-lactamase-like domain were highly diverse, in both length and amino acid sequence. On the other hand, an ORF product derived from *Arabidopsis* lacks the zinc-finger region. Instead, the ORF product has a DNA ligase I domain downstream of the β-lactamase domain (see Fig. 3). Recently, a protein named Artemis was identified from humans [27], which is involved in DNA double strand break repair and V(D)J recombination. The mutations of Artemis cause human conditions denominated severe combined immunodeficiency [27]. The protein shows significant sequence similarity to SNM1 [27], although Artemis lacks the zinc finger region (see Fig. 3). In addition to the zinc metallo-hydrolase domain, the members of Group 7 share a region called β-CASP domain [27] downstream of the metallo-hydrolase domain (see Fig. 3).

Group 8 consists of proteins involved in DNA uptake for natural genetic transformation, which are derived from various organisms in Bacteria [28,29]. The members of this group show sequence similarity to class B β-lactamase and the relatives [9]. As shown in Fig. 3, the N-terminal regions of some DNA uptake-related proteins have six hydrophobic segments, which are regarded as membrane-binding regions. We could not find the regions corresponding to segment 3 in Group 8. The DNA uptake-related proteins of Group 8 may cleave exogenous DNAs into small fragments by the metallo-hydrolase activity, to facilitate genetic transformation. This hypoth-

→

Fig. 1. A multiple alignment of the five highly conserved segments. The source, the identification code of each protein, and the corresponding database (gb, pir, sp, and pdb indicate GenBank, PIR, SwissProt, and PDB, respectively) are shown on the left side of the alignment. When it is necessary to discriminate a protein from the other members in the same group, the protein name is written in parentheses after the source name. The numbers neighboring the first and last residues in an aligned sequence indicate the residue numbers in the sequence. The numbers in parentheses indicate the number of residues omitted from the figure. The conserved segments are separated by the omitted regions. '–' indicates a gap for insertion and/or deletion. The measure to describe the alignment site and the consensus residues are shown at the top of the alignment. When more than 50% of a site is occupied by the same amino acid residue, the residue is colored red, and the residues physicochemically similar to the red residue at the site are colored blue. When more than 70% of a site is occupied by physicochemically similar residues, but none of the residues attains 50%, the residues are indicated by inverse characters. The criterion by Dayhoff, Schwartz and Orcutt [3] was used for the classification of the physicochemically similar residues. (1) hydrophobic residues, L, I, M, and V; (2) aromatic residues, F, Y, and W; (3) small hydrophilic residues, S, P, T, A, and G; (4) negatively charged residues and the relatives, D, E, N, and Q; (5) positively charged residues, and the relatives, R, H, K; (6) Cys residue, C. 'h' and 's' in the line of consensus residues indicate hydrophobic residues and small hydrophilic residues, respectively.


< segment 1 >
< segment 2 >
< segment 3 >
< segment 4 >
< segment 5 >
Group 1 metallo β-lactamase II
Bacillus cereus 2BC2 pdb 50 KGLVLVDSSWD(17)RVTDVIITHAHADRIGGIKTLK(40)VET-FYP-GK-GHTE(11)ILVGGCLVKS(32)AVVPGHG 213
Bacteroides fragilis 2BMI pdb 46 HQAALLDTPIN(17)KVTTFIPNHWHGDCIGGLGYLQ(40)LQC-YYL-GG-GHAT(11)ILFGGCMLKD(32)YVVPGHG 207
Aeromonas hydrophila BLAB_AERHY sp 59 KGVTVVGATWT(17)PVLEVINTNYHTDRAGGNAYWK(56)VRA-FYA-GP-AHTP(11)VLYGNCILKE(28)TVIGGHD 232
Xanthomonas maltophilia 1SML pdb 47 DGAVLLDGGMP(18)DLRLILLSHAHADHAGPVAELK(53)VFT-AHF-MA-GHTP(17)IAYA-DSLSA(31)VLLTPHP 226
Group 2 glyoxalase II
Escherichia coli GLO2_ECOLI sp 22 GRCLIVDPGDA(12)QPEAIFLTHHHHDHVGGVKELV(34)EFS-VIA-TP-GHTL(9)YLFCGDTLFS(28)LVCCAHE 166
Arabidopsis thaliana GL2M_ARATH sp 100 GTVGVVDPSEA(12)NLTYILNTHHHDDHIGGNAELK(35)EVR-ILD-TP-GHTQ(11)TIFTGDLIYS(28)NIYCGRE 247
Homo sapiens 1QH5 pdb 23 KEAAIVDPVQP(12)KLTTVLTTHHHWDHAGGNEKLV(33)NVK-CLA-TP-CHTS(16)AVFTGDTLFV(29)RVYCGHE 174
Arabidopsis thaliana GL2C_ARATH sp 23 GDAAVVDPVDP(12)KIKFVLTTHHHWDHAGGNEKIK(35)NIL-ALH-TP-CHTK(15)AVFTGDTLFV(29)QVYCGHE 175
Saccharomyces cerevisiae GLO2_YEAST sp 27 KKSWLIDPAEP(13)SVEAIVNTHHHYDHADGNADIL(39)EIT-CIR-TP-CHTR(15)CIFTGDTLFT(34)RVYPGHE 189
Group 3 flavoprotein
Methanothermobacter thermautotrophicus U17835 gb 44 EGVALIDNSYP(20)RVDYIIQNHVEKDHSGVLVELH(44)TLT-FLE-TPLLHWP(11)ILFSNDAFGQ(53)MIAPSHG 234
Desulfovibrio gigas (ROO) 1E5D pdb 43 EKTTLFDTVKA(17)KIDYLVIQHLELDHAGALPALI(43)TVT-FYE-TRMLHWP(11)VLISNDIFGQ(51)FICPDHG 227
Group 4 ELAC1/2, arylsulfatase
Escherichia coli ELAC_ECOLI sp 32 SGLWLFDCGEG(13)KLDKIFISHLHGDHLFGLPGLL(53)LRK-VTA-YPLEHPL(63)LAIFGDTGPC(48)KLIITHV 271
Homo sapiens (ELAC1) AF308695 gb 30 GECWLFDCGEG(13)RITKIFITHLHGDHFFGLPGLL(96)QFV-VKA-FRLFHRI(63)ICILGDCSGV(50)RLVLTHF 314
Pseudoalteromonas carrageenovora ARS_ALTCA sp 58 GNTYVVDAGDG(13)NVDAVFLSHLHFDHTGGLPAIL(67)DFT-VDV-IRNSHYS(24)VVYTGDTGPS(66)SLVITHM 290
Homo sapiens (ELAC2) AF304370 gb 509 DTSLLLDCGEG(18)TLAAVFVSHLHADHHTGLPSIL(73)DLEEFQT-CLVRHCK(14)VVYSGDTMPC(48)FIMLNHF 725
Drosophila melanogaster AF215894 gb 506 DAYVLLDCGEG(19)QLQAIYVSHLHADHHIGLIGLL(60)GITSIST-CLVRHCP(19)ITYSGDTMPC(48)HTILTHF 715
Group 5 cyclase
Streptomyces arenae AF098966 gb 46 GRAALVDTAAT(17)PPGVLVNTHFHGDHTFGNFVFP(53)RVE-VLHPGP-AHTT(11)VLFTGDIVMP(27)TVVPGHG 216
Streptomyces coelicolor S25844 pir 34 DGALVVDTLST(17)PGRTVVNTHFHGDHAFGNQVFA(54)QVE-LICVGP-AHTD(11)VLFAGDVVMS(27)VVVGGHG 205
Group 6 cleavage and polyadenylation specificity factor
Bos taurus (100 kDa) CPSB_BOVIN sp 27 EFRFLLDCGWD(16)QIDAVLLSHPDPLHLGALPYAV(67)GLS-ITP-LPAGHMI(14)IVYAVDFNHK(380)QLIIVHG 566
Saccharomyces cerevisiae (YSH1) S51413 pir 31 GKTVMLDAGIH(18)KVDILLISHFHLDHAASLPYVM(71)GIK-FTA-FPAGHVL(13)VLFTGDYSRE(236)NIILVHG 431
Bos taurus (73 kDa) X95906 gb 34 GRKIMLDCGIH(18)EIDLLLISHFHLDHCGALPWFL(63)GIK-FWC-YHAGHVL(13)LLYTGDFSRQ(229)HVILVHG 419
Archaeoglobus fulgidus B69310 pir 200 ESRILIDCGVN(19)ALDAVVITHAHLDHCGLVPLLY(63)DVR-LTF-YNAGHIL(15)IAFTGDFKFE(241)KVATVHG 600
Group 7 DNA cross-link repair gene SNM1
Mus musculus AF241240 gb 692 GTGFTVDAFQY(4)GCTAYFLTHFHSDHYAGLSKDF(37)SVK-VVF-VDANHCP(14)ILHTGDFRAD 802
Saccharomyces cerevisiae PSO2_YEAST sp 225 GHEIVVDGFNY(4)TISQYFLSHFHSDHYIGLKKSW(46)TIS-VVT-LDANHCP(21)ILHTGDFRSN 351
Arabidopsis thaliana AC013288 gb 62 NTNFIVDLFRL(4)SSVAFFLSHFHSDHYSGLSSSW(38)GSE-VVL-IEANHCP(18)YVHTGDFRFC 177
Homo sapiens (Artemis) AJ296101 gb 4 YPTISIDRFDR(3)RARAYFLSHCHKDHMKGLRAPT(57)EEI-VVTLLPAGHCP(13)VLYTGDFRLA 133
Group 8 comE
Neisseria gonorrhoeae COMA_NEIGO sp 465 NRHLLFDTGTV(17)RLDKLVLSHHDSDHDGGFQAVG(61)LLVTGDLDTK(17)LVLGHHG 609
Bacillus subtilis CME3_BACSU sp 520 RGRVLIDTGGT(32)QLDALILTHADQDHIGEAEILL(78)WILTGDLEKE(16)LKVGHHG 695
Group 9 choline binding protein E
Streptococcus pneumoniae AF278687 gb 52 GHFAMVDTGED(39)KLDFILVTHTHSDHIGNVDELL(99)IYLGGDLDNV(16)MKFNHHH 255
Group 10 PhnP protein
Escherichia coli PHNP_ECOLI sp 48 DAITLIDAGLH(9)SFQQFLLTHYHMDHVQGLFPLR(43)GLQ-VTP-LPLNHS-(14)VAWLSDTAGL(48)RVILTHI 223
Group 11 CMP-N-acetylneuraminate monooxygenase
Mus musculus A57469 pir 153 DKRMVFDPWLI(23)KADLIYISHMHSDHL-SYPTLK(47)SLR-FMILMDGVHPE(13)ILNTVDCTRP(69)YFVESHP 367
Group 12 romA
Enterobacter cloacae 1613275A prf 119 GKRILIDPVLG(25)EIDLLIISHDHYDHL-DYATIR(42)ALT-VHL-LPARHFS(22)IYYSGDSGYG(47)AVVPGHN 316
Group 13 alkyl sulfatase
Pseudomonas sp. JC1118 pir 6 EGLIIVDTGES(17)PIKAIVYTHFHPDHINGVKAFV(86)KVQ-FLH-VP-SEAP(11)VLISAEVTQG(33)VMVPLHG 214
Group 14 carbofuran hydrolase
Achromobacter sp. WM111 AF160188 gb 58 SGSVIIDAGTG(11)EPAALLCTHFFRDHAEGAAYAA(28)ICS-FDS-VY-WHHF 141
Achromobacter sp. WM111 AF160188 gb 314 GKVLSIDYGAP(32)GIDVVLTTHFHDDHVGGINLLK(52)RMG-PEI-GG-GHTH(13)YAAVGDQYLS(45)VILNGHQ 517
Group 15 methyl parathion hydrolase
Plesiomonas sp. M6 AF338729 gb 106 SKLVLVDTGAA(22)QVDEIYITHMHPDHVGGLMVGE(68)LAS-HGH-TP-GHTT(9)LALLGDLILV(37)LIAASHL 303
Group 16 3',5'-cyclic nucleotide phosphodiesterase
Vibrio fischeri CPDP_VIBFI sp 52 SNFVMLDAGSV(33)RIKGYFISHAHLDHVAGLIISS(59)TMS-VVS-LPLSHSG(15)FAYFGDTGPD(39)KSLFGHL 260
Saccharomyces cerevisiae CNA1_YEAST sp 33 PELIAVDGGAG(76)GITDYYITHPHLDHIGGLVVNS(61)PWD-VIP-FKVHHGI(23)IIVCGDVEQD(40)EQLYGHL 295
Group 0
Pyrococcus furiosus T43935 pir 17 GTKILIDPYPD(7)KVDYILVTHEHMDHYGKTPLIA(37)DIR-VRA-FFTEHPT(14)VAHLGDTYYS(38)IAVPMHY 175

esis is similar to the genetic transformation model proposed by Yura et al. [30].

Group 9 consists of choline-binding protein E, which is also known as teichoic acid phosphorylcholine esterase from *Streptococcus pneumoniae* [31]. The phosphorylcholine residue is a component of the bacterial cell wall, and serves to anchor surface-located bacterial choline-binding proteins. The enzyme is involved in removing phosphorylcholine residues from the cell wall, such as teichoic acid and lipoteichoic acid [32]. The C-terminal region includes 10 tandem repeats of about 20 amino acid residues (see Fig. 3). The repeated sequences are known as Psp repeats [33], and constitute the choline-binding domain [34]. However, the catalytic activity is localized on the N-terminal part of the protein [32]. The members of this group are relatively close to those of Group 8. Reflecting the relationship, we could not identify the region corresponding to segment 3 of Group 9.

The genes involved in alkylphosphonate uptake constitute an operon in the *Escherichia coli* genome, which includes a gene named *phnP* [35]. The *phnP* protein belongs to the metallo-hydrolase family [9], and constitutes Group 10, together with its eubacterial relatives. However, the physiological meaning of the *phnP* protein is still unknown, because mutational analyses have suggested that the proteins are not always essential for phosphonate utilization [36].

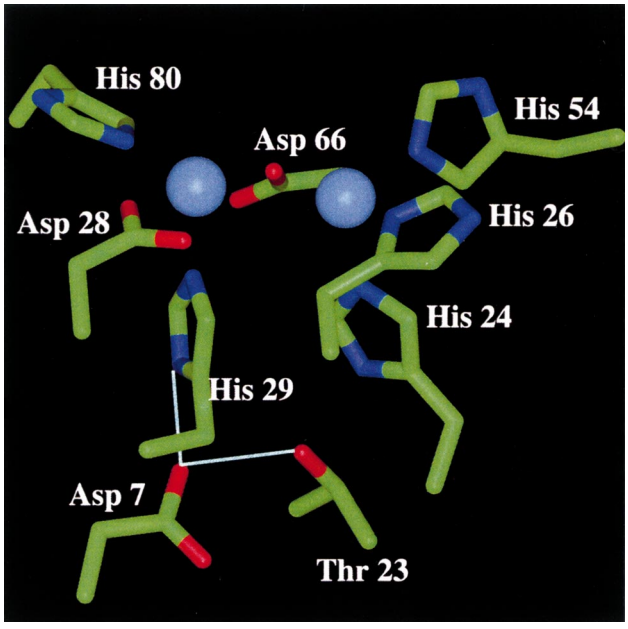


Fig. 2. The spatial arrangement of the conserved amino acid residues. Highly conserved residues in the segments are drawn as a stick and ball model, where oxygen, nitrogen, and carbon atoms are colored red, blue, and green. The two light blue balls indicate zinc ions. The residue type and the alignment site in Fig. 1 of each residue are written at the residue. The structure of glyoxalase II (PDB ID code = 1QH5) is used as a representative structure of the family to draw the figure. However, we also used the coordinate data of class B β-lactamase (2BC2, 2BM1, and 1SML of Group 1), and ROO (1E5D of Group 3) to examine the spatial relationship among the conserved residues. Three conserved residues (alignment sites 24, 26, and 54) were bound to a metal cation. Likewise, the conserved residues at alignment sites 28, 29, and 80 were involved in the binding to another metal cation. The Asp residue at the alignment site 66 is bound to one or both metal cation(s), although the conservation is rather weak. The Asp residue at alignment site 7 does not bind to the metal cation directly, but the residue is located in a position to be able to form a hydrogen bond with the residue at alignment site 23. In addition, the residue at alignment site 7 can also form a hydrogen bond with the residue at alignment site 29. Thin light-blue lines from Asp 7 to Thr 23 and His 29 indicate possible hydrogen bonds.

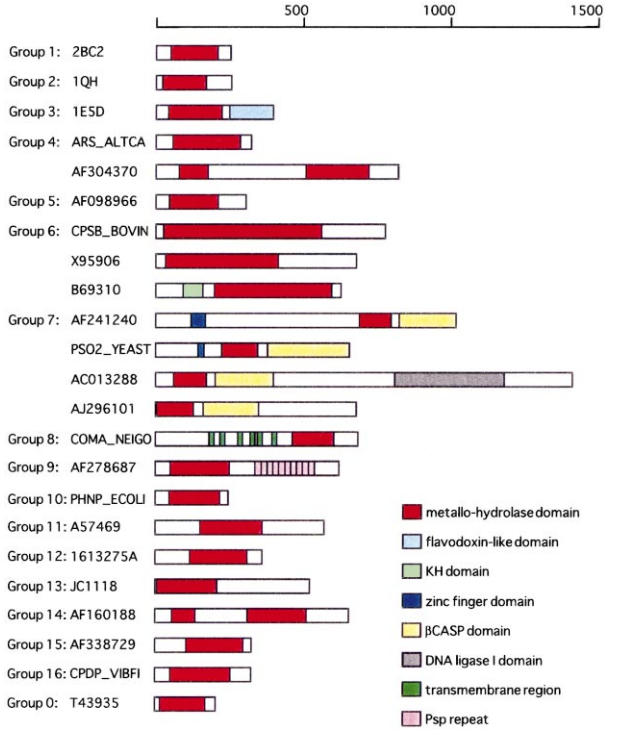


Fig. 3. Schematic diagrams of the primary structures. Only the primary structures of representative members from each group are shown. Each domain constituting a mosaic structure is indicated by a colored box. The correspondence between a domain and a color is explained in the figure. The bar over the figure is used as a measure for the protein length, which indicates the number of amino acid residues from the N-terminus.

Group 11 consists of CMP-*N*-acetylneuraminate hydroxylase (EC 1.14.13.45), which is a key enzyme for the synthesis of *N*-glycolylneuraminic acid [37]. The enzyme includes an iron ion, like ROO of Group 3, and functions as a dehydrase. The enzyme is the first iron–sulfur protein of the 'Rieske' type to be found in the cytosol of eukaryotic cells. The gene encoding this enzyme has been degraded by frameshift mutations in the lineage to human after the divergence from chimpanzee [38].

The *romA* gene product is also a member of the zinc metallo-hydrolase family [9], and constitutes Group 12. It was found that the introduction of the *romA* gene from *Enterobacter cloacae* into *E. coli* enhances the multidrug resistance activity to quinolones, β-lactams, chloramphenicol, and tetracycline [39]. In addition, the introduction of the *romA* gene pleiotropically inhibits the expression of the genes for the outer membrane proteins of *E. coli*, such as OmpF [39]. However, the functional mechanisms of the relatives remain unknown.

Alkylsulfatase (EC 3.1.6.-) has been identified from *Pseudomonas* sp. as a determinant of sodium dodecyl sulfate biodegradation [40]. The enzyme belongs to the zinc metallo-hydrolase family [9], and constitutes Group13. The putative orthologues from other eubacterial organisms were identified as hypothetical proteins.

Group 14 consists of carbofuran hydrolase from *Achromo-*

bacter sp. WM111. The enzyme is capable of the rapid hydrolysis of an insecticide, carbofuran [41]. We found that the enzyme is composed of by two-fold repetitive units, which each show sequence similarity to the members of the zinc metallo-hydrolase family (see Fig. 3). The N-terminal repetitive unit seemed to be diverged, as compared to the C-terminal repetitive unit. In addition, we could not find the regions corresponding to segments 4 and 5 in the N-terminal repetitive unit. In contrast, the regions corresponding to the five segments were well conserved in the C-terminal repetitive unit of the enzyme (see Fig. 1).

Group 15 consists of methyl parathion hydrolase and its relative. The enzyme was identified as being capable of degrading an insecticide, methyl parathion, from *Plesiomonas* sp. M6 [42].

Class II 3′, 5′-cyclic-nucleotide phosphodiesterases (EC 3.1.4.17) derived from eubacteria and fungi also belong to the zinc metallo-hydrolase family [9], and constitute Group 16. The enzyme is also known as the low-affinity cAMP phosphodiesterase in *Saccharomyces cerevisiae*, and controls the level of cAMP in yeast cells [43]. The enzyme derived from *Vibrio fisherii* is present in the periplasm, and its role is to utilize extracellular cAMP as a source of carbon, nitrogen, and phosphorus for growth [44]. The members of this group have been identified from *Schizosaccharomyces pombe*, *Candida albicans*, and *Dictyostelium discoideum*, although the sequences are not included in Fig. 1.

3. Concluding remarks

As described above, we omitted many functionally unknown ORF products, detected by database searching, from the alignment in Fig. 1. It should be noted that many ORF products, together with the proteins described in this study, are encoded by the genomes from Eukarya, Archaea, and Bacteria. The expansion of this protein family would be helpful to infer the functions of such ORF products. Here, we will show the case of an archaeal ORF product as an example.

An operon in the genome of a hyperthermophilic archaeon, *Pyrococcus furiosus*, was identified as a region encoding two subunits, DP1 and DP2, of the novel archaeal DNA polymerases, named Pol II [45]. Sequencing of the region revealed that five ORFs (ORFs 1–5) are encoded in the operon [45]. The DP1 and DP2 subunits are encoded by ORFs 2 and 3. The ORF1 product shows sequence similarity to the members of the cdc18+/CDC6 family, and the ORF5 product shows sequence similarity to the members of the Rad51/Dmc1 family [45]. The sequence similarity suggests that the products of ORFs 1 and 5 are involved in DNA replication and recombination. In contrast, the 210-amino acid residue ORF4 product has remained uncharacterized. Like the other four ORFs in the same operon, the ORF4 product is expected to be related to DNA processing. We found that the ORF4 product shows significant sequence similarity to the members of the zinc metallo-hydrolase family. The ORF product is shown as Group 0 in Fig. 1. The close relatives of the ORF4 product, which are present in the determined archaeal genomic sequences, have been annotated as zinc metallo-hydrolases of the β-lactamase fold in the COG database [2]. If the functional variety of the family were not known, then it would be difficult to interpret the meaning of the sequence similarity of the ORF product to the zinc metallo-hydrolases, such as β-lactamase. As described above, however, the expanded members of the zinc metallo-hydrolase family included proteins related to DNA or RNA processing. The sequence similarity suggests the possibility that the archaeal ORF product of Group 0 is also involved in DNA processing.

The ubiquitous distribution over the three biological domains suggests the functional importance and the ancient origin of the protein family. The characterization of the ORF products belonging to this protein family will lead to the discovery of novel biological functions.

References

[1] Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Nucleic Acids Res. 25, 3389–3402.
[2] Tatusov, R.L., Natale, D.A., Garkavtsev, I.V., Tatusova, T.A., Shankavaram, U.T., Rao, B.S., Kiryutin, B., Galperin, M.Y., Fedorova, N.D. and Koonin, E.V. (2001) Nucleic Acids Res. 29, 22–28.
[3] Dayhoff, M.O., Schwartz, R.M., and Orcutt, B.C. (1978) in: Atlas of Protein Sequence and Structure (Dayhoff, M.O., Ed.), pp. 345–352, National Biomedical Research Foundation, Washington, DC.
[4] Cricco, J.A. and Vila, A.J. (1999) Curr. Pharm. Des. 5, 915–927.
[5] Carfi, A., Pares, S., Duee, E., Galleni, M., Duez, C., Frere, J.M. and Dideberg, O. (1995) EMBO J. 14, 4914–4921.
[6] Concha, N.O., Rasmussen, B.A., Bush, K. and Herzberg, O. (1996) Structure 4, 823–836.
[7] Ullah, J.H., Walsh, T.R., Taylor, I.A., Emery, D.C., Verma, C.S., Gamblin, S.J. and Spencer, J. (1998) J. Mol. Biol. 284, 125–136.
[8] Concha, N.O., Janson, C.A., Rowling, P., Pearson, S., Cheever, C.A., Clarke, B.P., Lewis, C., Galleni, M., Frere, J.M., Payne, D.J., Bateson, J.H. and Abdel-Meguid, S.S. (2000) Biochemistry 39, 4288–4298.
[9] Neuwald, A.F., Liu, J.S., Lipman, D.J. and Lawrence, C.E. (1997) Nucleic Acids Res. 25, 1665–1677.
[10] Melino, S., Capo, C., Dragani, E., Aceto, A. and Petruzzelli, R. (1998) Trends Biochem. Sci. 23, 381–382.
[11] Cameron, A.D., Ridderström, M., Olin, B. and Mannervik, B. (1999) Structure 7, 1067–1078.
[12] Crowder, M.W., Maiti, M.K., Banovic, L. and Makaroff, C.A. (1997) FEBS Lett. 418, 351–354.
[13] Frazão, C., Silva, G., Gomes, C.M., Matias, P., Coelho, R., Sieker, L., Macedo, S., Liu, M.Y., Oliveira, S., Teixeira, M., Xavier, A.V., Rodrigues-Pousada, C. and Carrondo, M.A. (2000) Nat. Struct. Biol. 7, 1041–1045.
[14] Chen, L., Liu, M.Y., LeGall, J., Fareleira, P., Santos, H. and Xavier, A.V. (1993) Biochem. Biophys. Res. Commun. 193, 100–105.
[15] Barbeyron, T., Potin, P., Richard, C., Collin, O. and Kloareg, B. (1995) Microbiology 141, 2897–2904.
[16] Tavtigian, S.V., Simard, J., Teng, D.H., Abtin, V., Baumgard, M., Beck, A., Camp, N.J., Carillo, A.R., Chen, Y., Dayananth, P., Desrochers, M., Dumont, M., Farnham, J.M., Frank, D., Frye, C., Ghaffari, S., Gupte, J.S., Hu, R., Iliev, D., Janecki, T., Kort, E.N., Laity, K.E., Leavitt, A., Leblanc, G., McArthur-Morrison, J., Pederson, A., Penn, B., Peterson, K.T., Reid, J.E., Richards, S., Schroeder, M., Smith, R., Snyder, S.C., Swedlund, B., Swensen, J., Thomas, A., Tranchant, M., Woodland, A.M., Labrie, F., Skolnick, M.H., Neuhausen, S., Rommens, J. and Cannon-Albright, L.A. (2001) Nat. Genet. 27, 172–180.
[17] Dubrovsky, E.B., Dubrovskaya, V.A., Bilderback, A.L. and Berger, E.M. (2000) Dev. Biol. 224, 486–495.
[18] Huntchinson, C.R. (1999) Proc. Natl. Acad. Sci. USA 96, 3336–3338.
[19] Fernández-Moreno, M.A., Martínez, E., Boto, L., Hopwood, D.A. and Malpartida, F. (1992) J. Biol. Chem. 267, 19278–19290.
[20] McDaniel, R., Ebert-Khosla, S., Hopwood, D.A. and Khosla, C. (1994) J. Am. Chem. Soc. 116, 10855–10859.

[21] Bartel, P.L., Zhu, C.B., Lampel, J.S., Dosch, D.C., Connors, N.C., Strohl, W.R., Beale, J.J. and Floss, H.G. (1990) J. Bacteriol. 172, 4816–4826.
[22] Jenny, A., Hauri, H.P. and Keller, W. (1994) Mol. Cell Biol. 14, 8183–8190.
[23] Jenny, A., Minvielle-Sebastia, L., Preker, P.J. and Keller, W. (1996) Science 274, 1514–1517.
[24] Siomi, H., Matunis, M.J., Michael, W.M. and Dreyfuss, G. (1993) Nucleic Acids Res. 21, 1193–1198.
[25] Richter, D., Niegemann, E. and Brendel, M. (1992) Mol. Gen. Genet. 231, 194–200.
[26] Dronkert, M.L., de Wit, J., Boeve, M., Vasconcelos, M.L., van Steeg, H., Tan, T.L., Hoeijmakers, J.H. and Kanaar, R. (2000) Mol. Cell Biol. 20, 4553–4561.
[27] Moshous, D., Callebaut, I., de Chasseval, R., Corneo, B., Cavazzana-Calvo, M., Le Deist, F., Tezcan, I., Sanal, O., Bertrand, Y., Philippe, N., Fischer, A,. and de Villartay, J.P. (2001) Cell 105, 177–186.
[28] Dreisenkelmann, B. (1994) Microbiol. Rev. 58, 293–316.
[29] Solomon, J.M. and Grossman, A.D. (1996) Trends Genet. 12, 150–155.
[30] Yura, K., Toh, H. and Go, M. (1999) DNA Res. 6, 75–82.
[31] Gosink, K.K., Mann, E.R., Guglielmo, C., Tuomanen, E.I. and Masure, H.R. (2000) Infect. Immun. 68, 5690–5695.
[32] Vollmer, W. and Tomasz, A. (2001) Mol. Microbiol. 39, 1610–1622.
[33] Yother, J. and Briles, D.E. (1992) J. Bacteriol. 174, 601–609.
[34] García, J.L., Sánchez-Beato, A.R., Medrano, F.J. and López, R. (1998) Microb. Drug Resist. 4, 25–36.
[35] Chen, C.M., Ye, Q.Z., Zhu, Z.M., Wanner, B.L. and Walsh, C.T. (1990) J. Biol. Chem. 265, 4461–4471.
[36] Metcarf, W.W. and Wanner, B.L. (1993) J. Bacteriol. 175, 3430–3442.
[37] Schlenzka, W., Shaw, L., Kelm, S., Schmidt, C.L., Bill, E., Trautwein, A.X., Lottspeich, F. and Schauer, R. (1996) FEBS Lett. 385, 197–200.
[38] Chou, H.H., Takematsu, H., Diaz, S., Iber, J., Nickerson, E., Wright, K.L., Muchmore, E.A., Nelson, D.L., Warren, S.T. and Varki, A. (1998) Proc. Natl. Acad. Sci. USA 95, 11751–11756.
[39] Komatsu, T., Ohta, M., Kido, N., Arakawa, Y., Ito, H., Mizuno, T. and Kato, N. (1990) J. Bacteriol. 172, 4082–4089.
[40] Davison, J., Brunel, F., Phanopoulos, A., Prozzi, D. and Terpstra, P. (1992) Gene 114, 19–24.
[41] Tomasek, P.H. and Karns, J.S. (1989) J. Bacteriol. 171, 4038–4044.
[42] Cui, Z., Li, S. and Shen, B. (2001) direct submission to GenBank (accession number AF338729).
[43] Nikawa, J.-I., Sass, P. and Wigler, M. (1987) Mol. Cell. Biol. 7, 3629–3636.
[44] Dunlap, P.V. and Callahan, S.M. (1993) J. Bacteriol. 175, 4615–4624.
[45] Uemori, T., Sato, Y., Kato, I., Doi, H. and Ishino, Y. (1997) Genes Cells 2, 499–512.