doi:10.1016/j.jmb.2011.04.066

J. Mol. Biol. (2011) 410, 18–26

Contents lists available at www.sciencedirect.com

Journal of Molecular Biology

journal homepage: http://ees.elsevier.com.jmb

COMMUNICATION

Expansion of Type II CAAX Proteases Reveals Evolutionary Origin of γ-Secretase Subunit APH-1

Jimin Pei[1]*, Douglas A. Mitchell[2,3,4], Jack E. Dixon[5,6,7] and Nick V. Grishin[1,8]

[1]*Howard Hughes Medical Institute, University of Texas Southwestern Medical Center, 6001 Forest Park Road, Dallas, TX 75390-9050, USA*
[2]*Department of Chemistry, University of Illinois at Urbana-Champaign, 1206 West Gregory Drive, Urbana, IL 61801, USA*
[3]*Department of Microbiology, University of Illinois at Urbana-Champaign, 1206 West Gregory Drive, Urbana, IL 61801, USA*
[4]*Institute for Genomic Biology, University of Illinois at Urbana-Champaign, 1206 West Gregory Drive, Urbana, IL 61801, USA*
[5]*Department of Pharmacology, University of California, San Diego, La Jolla, CA 92093, USA*
[6]*Department of Cellular and Molecular Medicine, University of California, San Diego, La Jolla, CA 92093, USA*
[7]*Department of Chemistry and Biochemistry, University of California, San Diego, La Jolla, CA 92093, USA*
[8]*Department of Biochemistry, University of Texas Southwestern Medical Center, 6001 Forest Park Road, Dallas, TX 75390-9050, USA*

Received 30 March 2011;
received in revised form
25 April 2011;
accepted 28 April 2011
Available online
5 May 2011

Edited by B. Honig

Keywords:
type II CAAX protease;
APH-1;
γ-secretase;
PrsW;
DUF2324

Intramembrane proteases are responsible for a number of regulated proteolysis events occurring within or near the plasma and intracellular membranes. Members of one large and diverse family of putative intramembrane metalloproteases are widely distributed in all domains of life, including the type II CAAX prenyl proteases and their prokaryotic homologs with putative bacteriocin-related functions. We used sensitive sequence similarity searches to expand this large CPBP (CAAX *p*roteases and *b*acteriocin-*p*rocessing enzymes) family to include more than 5800 members and infer its homologous relationships to several other protein families, including the PrsW proteases, the DUF2324 (DUF, *d*omain of *u*nknown *f*unction) family and the γ-secretase subunit APH-1 proteins. They share four predicted core transmembrane segments and possess similar yet distinct sets of sequence motifs. Remote similarity between APH-1 and membrane proteases sheds light on APH-1's evolutionary origin and raises the possibility that APH-1 may possess proteolytic activity in the current or ancestral form of γ-secretase.

A number of key cellular processes involve the intramembrane proteolytic cleavage of proteins.[1,2] Several distinct families of membrane proteases that carry out peptide bond cleavage inside lipid bilayers have been characterized. Recent structural studies for two of these families [site-2 protease (S2P) and rhomboid] have provided insights into their

---

*Corresponding author. E-mail address: jpei@chop.swmed.edu.

Abbreviations used: S2P, site-2 protease; gi, gene identification; NCBI, National Center for Biotechnology Information.

catalytic mechanisms.[3,4] Eukaryotic S2P is an intramembrane zinc metalloprotease that releases a transcription factor domain from the membrane-bound SREBP protein to regulate cholesterol and fatty acid biosynthesis.[5,6] Similar reactions important for transcriptional regulation are carried out by bacterial homologs of S2P, such as Eep from *Enterococcus faecalis*,[7] SpoIVFB from *Bacillus subtilis*[8] and RseP from *Escherichia coli*.[9] Sequence analysis revealed that S2P homologs are widely distributed in archaeal, bacterial and eukaryotic species.[10] S2P contains a conserved "HExxH" motif resembling the ones found in many soluble metalloproteases, albeit this motif in S2P is located in a transmembrane region. Recently, the crystal structure of an archaeal S2P homolog showed that this motif and a couple of other conserved residues form the membrane-embedded metal-binding active site that overlays well with those from soluble metalloproteases,[11] suggesting a similar catalytic mechanism. Another intramembrane protease, rhomboid, has been characterized as a serine protease functioning in signaling pathways of *Drosophila* development.[12] Structural studies of select bacterial rhomboid homologs uncovered a catalytic Ser-His diad embedded in transmembrane regions.[3,13–15]

Another important event of intramembrane proteolysis is performed by γ-secretase,[16,17] which is a large membrane-bound protein complex consisting of four core subunits: presenilin, APH-1, nicastrin and PEN-2. Its substrates are type I membrane-spanning proteins such as the cell-surface receptor NOTCH and the Alzheimer-disease-related amyloid precursor protein. Presenilin, an aspartyl protease, is the catalytic subunit of γ-secretase. Homologs of presenilin also include signal peptide peptidases[18] and bacterial type 4 prepilin proteases.[19] Nicastrin is a type I membrane protein and may function as the substrate receptor.[20] PEN-2 is a small protein with two transmembrane segments. APH-1 is a seven-transmembrane-spanning protein suggested to have a stabilizing role for γ-secretase. The precise function and the evolutionary origin of APH-1 remain unclear.

Two types of membrane-bound proteases are involved in the prenylation process that covalently attaches lipid molecules to proteins in order to facilitate their membrane localization.[21] These prenyl proteases, located in the endoplasmic reticulum,[22] perform cleavage within the CAAX motif of the substrate proteins to release the "AAX" tripeptide (A, a typically aliphatic residue; X, one of several allowed residue types that can help determine the specificity of prenyltransferases[23]). The new C-terminal residue of the processed protein is the prenylated cysteine. Type I CAAX prenyl proteases, namely, "α-factor converting enzyme" (AFC1, also called Ste24p in yeast), are metalloproteases with a conserved HExxH motif.[21,24,25] Type II CAAX proteases, "Ras converting enzyme" (RCE1), lack this sequence motif, and their catalytic type has been debated in the literature. Early work suggested them to be cysteine proteases based on the activity of cysteine, histidine and glutamate mutations and on inhibitor analysis.[26] Mutations of these residues all led to loss of catalytic activity. However, the suspected catalytic cysteine residue was not conserved in bacterial homologs. It was suggested that type II CAAX proteases were metalloproteases based on a set of conserved glutamates and histidines.[27] These conserved residues were later shown to be important for catalysis in a site-directed mutagenic study of yeast Rce1p.[28] The most probable explanation for the early suggestion that CAAX proteases were cysteine proteases is that there is a cysteine near the active site. Upon mutation or modification by inhibitors, the enzyme is deactivated, although the residue does not directly participate in catalysis.[29] Beyond these yeast type II CAAX protease studies, some bacterial homologs are involved in the proteolytic maturation of bacteriocins.[30] Recently, the PrsW protease from *B. subtilis* was found to be a distant homolog of type II CAAX proteases.[31] PrsW was shown to mediate site-1 cleavage of the anti-σ factor RsiW.[31]

In this study, we conducted extensive sequence similarity searches of CAAX prenyl proteases in current sequence database using transitive PSI-BLAST[32] iterations. The CPBP (*C*AAX *p*roteases and *b*acteriocin-*p*rocessing enzymes) family of putative metalloproteases was expanded to encompass more than 5800 members. Remote similarities between type II CAAX proteases and other protein families were also investigated by the sensitive profile–profile search method HHpred.[33] Weak yet significant similarities, highlighted by conserved sequence motifs, were identified among the CPBP family, the PrsW proteases, the Pfam[34] family DUF2324 (DUF, *d*omain of *u*nknown *f*unction) and, surprisingly, the γ-secretase subunit APH-1. The protease origin of APH-1 raises interesting possibility that it may possess proteolytic activity in the current or ancestral form of γ-secretase.

Sequence Similarity Searches for Homologs of Type II CAAX Prenyl Peptidases

PSI-BLAST[32] was used to search for homologs of the type II CAAX protease family starting with the yeast Rce1p [National Center for Biotechnology Information (NCBI) gene identification (gi) number: 6323930] against the nonredundant database (12,322,590 sequences and 4,211,205,364 letters) with an *e*-value inclusion cutoff of 1e-4. To make transitive searches, we grouped PSI-BLAST hits by BLASTClust (with option -S 1), and a representative sequence from each group was selected to initiate new PSI-BLAST searches. More than 5800 members of the CPBP family were found. Weaker similarities among members of the CPBP family, the PrsW protease family and the DUF2324 family were occasionally found during transitive PSI-BLAST searches [*e*-value inclusion cutoff of 1e-4, more stringent than the recommended cutoff value of web PSI-BLAST (0.005) to prevent profile corruption]. For example, a PSI-BLAST search using a bacterial CPBP member from *Atopobium vaginae* (gi: 303233059) found a significant hit to a PrsW protease family member from *Sorangium cellulosum* (gi: 162453351) with an *e*-value of 8e-5 in the fifth iteration. A PSI-BLAST search using a DUF2324 protein from *Ruminococcus gnavus* (gi: 154504350) found a significant hit to a PrsW protease family member from *Flavobacteriales bacterium* (gi: 163788689) with an *e*-value of 8e-5 in the ninth iteration.

Transitive PSI-BLAST searches (*e*-value inclusion cutoff of 1e-4) using members from the CPBP family, PrsW protease family and the DUF2324 family did not identify APH-1 proteins with statistical significance. Similarly, PSI-BLAST searches initiated with the APH-1 family representatives converged to

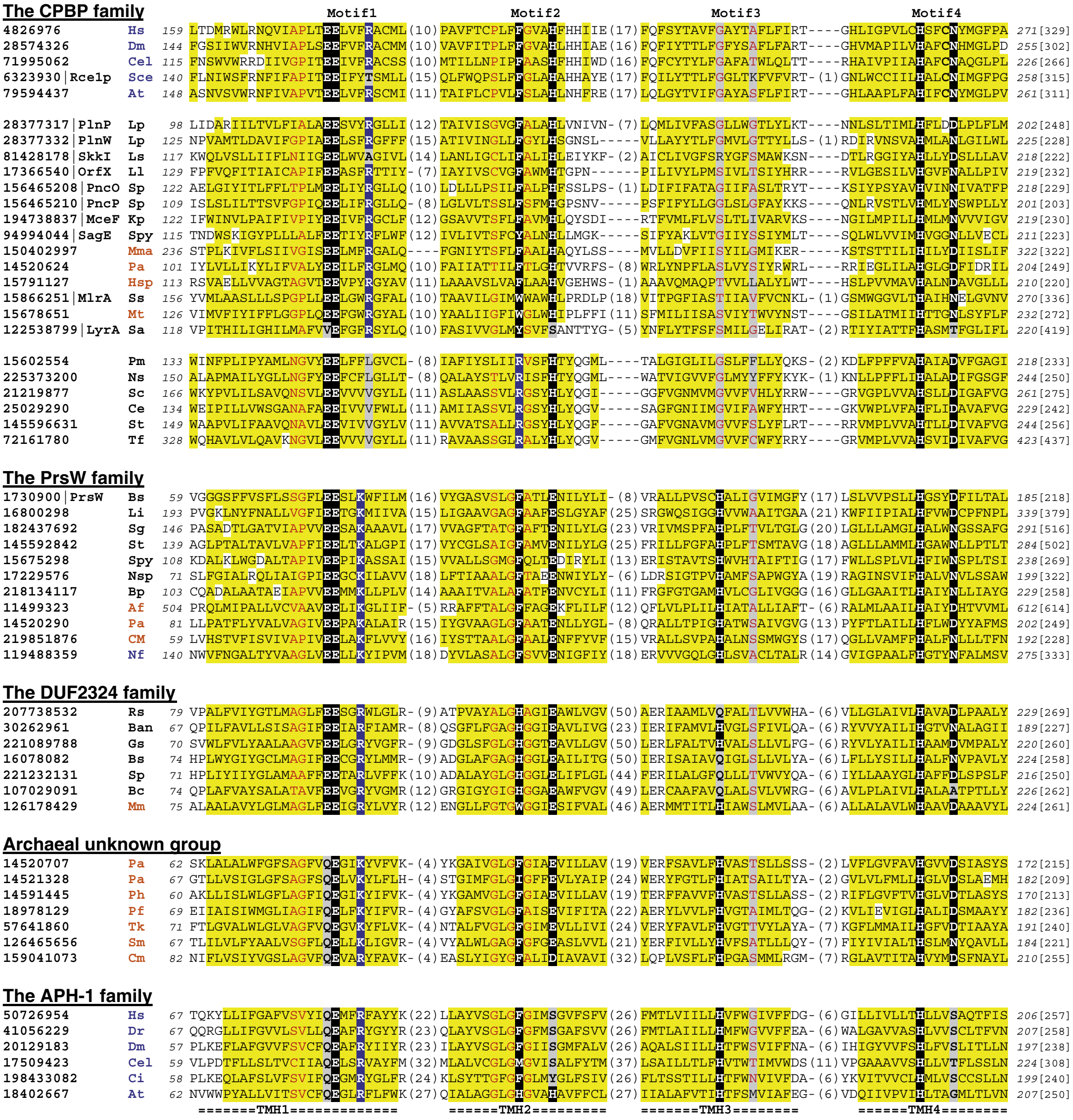

Fig. 1 (*legend on next page*)

about 150 eukaryotic members and did not find CPBP members, PrsW proteases or DUF2324 members as significant hits. However, the similarity between APH-1 and other families can be found and statistically supported by the sensitive profile–profile method HHpred.[33] For example, an HHpred search using the human APH-1A (gi: 117606362) detected Pfam family DUF2324 (PF10086) with a probability score of 98.1 and an *e*-value of 4e-6. HHpred searches using PrsW and DUF2324 members also frequently identified Pfam domain Aph-1 (PF06015) among the top hits with probability scores above 90. Furthermore, their homologous relationships are supported by the presence of a set of similar sequence motifs harboring semi-invariant residues in four predicted core transmembrane segments (Fig. 1; discussed below).

Sequence Families of Proteins Homologous to Type II CAAX Prenyl Proteases

Four major families of proteins homologous to type II CAAX prenyl proteases exist: the CPBP family, PrsW proteases, DUF2324 and APH-1 (Fig. 1). They share several similar sequence motifs with distinct variations.

The CPBP family

This family includes eukaryotic type II CAAX prenyl proteases and their related bacterial and archaeal homologs that can be repeatedly found in transitive PSI-BLAST searches and often share similar sequence motifs. This family is widely represented in the current sequence database with more than 5800 members. CPBPs are present in all domains of life, with most being bacterial. This family corresponds to the Pfam entry Abi (PF02517; CAAX amino-terminal protease family). A member of this family was originally identified as the product of an open reading frame (*orfX*) upstream of two genes (*abiGi* and *abiGii*) involved in an abortive infection system.[36] However, inactivation of this CPBP member (OrfX; gi: 17366540) suggests that it is not involved in the abortive infection mechanism. Thus, it is inappropriate to name this family “Abi” and annotate CPBP members as “abortive infection protein” (about one-fifth of the CPBP members (>1000) are currently annotated in this way).

To obtain a clear picture of distribution of CPBPs in prokaryotes, we searched for CPBPs in 1190 fully sequenced prokaryotic genomes (downloaded from the NCBI genome ftp site). Out of 90 archaeal genomes and out of 1100 bacterial genomes, 61 and 793, respectively, have at least one CPBP (Supplementary Table S2). The number of CPBPs among different prokaryotic phyla shows high variation (Supplementary Table S3). About 50% of proteobacteria do not contain any identifiable CPBP. Conversely, the representation of CPBP family members in many other phyla is remarkably higher, with all sequenced cyanobacteria, bacteroidetes, chlamydiae, chlorobi, thermotogae and deinococcus having at least one CPBP member. Within the firmicutes, actinobacteria and euryarchaeaota phyla, it is not uncommon to find 5–20 CPBP family members within a single genome (Supplementary Table S2). In one notable case, 21 copies are found within the genome of *Streptococcus sanguinis* SK36, which is

Fig. 1. Multiple sequence alignments of representative members from protein families homologous to the type II CAAX prenyl proteases. Four predicted core transmembrane helical segments (labeled TMH1–4 below the sequences) with conserved motifs (labeled Motif1–4 above the sequences) are shown for representative sequences of the CPBP family, the PrsW proteases, the DUF2324 family, a group of archaeal proteins with unknown function and APH-1 proteins. The three separated groups in the CPBP family are eukaryotic type II CAAX proteases, prokaryotic homologs including putative bacteriocin-processing enzymes and a group of bacterial proteins with the conserved arginine residues missing from the first motif but present in the second motif. Putative active-site residues are shown on black background, except for the conserved positively charged residues, which are shown on blue background. Substitutions in these positions are on gray background. Non-charged residues in mainly hydrophobic positions are on yellow background. Small residues (residue types: G, A, S, C, V, T, N, D and P) on positions with mainly small residues are shown in red letters. NCBI gi numbers, along with common names for some proteins, are shown before the species name abbreviations. Numbers of residues in between the segments are shown in parentheses. Starting/ending residue numbers and sequence lengths are shown in italic font and in brackets, respectively. Species name abbreviations are as follows: At, *Arabidopsis thaliana*; Af, *Archaeoglobus fulgidus*; Ba, *Bacillus anthracis*; Bs, *B. subtilis*; Bp, *Bacteroides pectinophilus*; Bc, *Burkholderia cenocepacia*; Cel, *Caenorhabditis elegans*; Cm, *Caldivirga maquilingensis*; Ci, *Ciona intestinalis*; Ce, *Corynebacterium efficiens*; Dr, *Danio rerio*; Dm, *Drosophila melanogaster*; Gs, *Geobacillus* sp.; Hsp, *Halobacterium* sp.; Hs, *Homo sapiens*; Kp, *K. pneumoniae*; Lp, *L. plantarum*; Ll, *Lactococcus lactis*; Li, *Listeria innocua*; Ls, *L. sakei*; Mma, *Methanococcus maripaludis*; Mm, *Methanoculleus marisnigri*; Mp, *Methanosphaerula palustris*; Mt, *Methanothermobacter thermautotrophicus*; Ns, *Neisseria subflava*; Nf, *N. fischeri*; Nsp, *Nostoc* sp.; Pm, *Pasteurella multocida*; Pa, *Pyrococcus abyssi*; Pf, *Pyrococcus furiosus*; Ph, *Pyrococcus horikoshii*; Rs, *Ralstonia solanacearum*; Sce, *Saccharomyces cerevisiae*; St, *Salinispora tropica*; Ss, *Sphingomonas* sp.; Sa, *S. aureus*; Sm, *Staphylothermus marinus*; Sp, *S. pneumoniae*; Spy, *S. pyogenes*; Sc, *Streptomyces coelicolor*; Sg, *Streptomyces griseus*; Tf, *Thermobifida fusca*; and Tk, *Thermococcus kodakarensis*. The domain color coding is as follows: bacteria, black; archaea, orange; and eukaryotes, blue. This alignment was made by PROMALS[35] followed by manual adjustment.

currently the highest known. CPBP members in eukaryotes also have a wide phylogenetic distribution, with members from metazoans, fungi, plants and various protists.

Members of this family share four predicted core transmembrane segments, each of which contains a signature sequence motif (Fig. 1). Most of the members contain several additional predicted transmembrane segments at the N-terminus, and some members possess an additional predicted transmembrane segment at the C-terminus. For the majority of the CPBP family, the sequence motifs are "EExxxR", "FxxxH", "sxxxs" and "HxxxB" ("x", any amino acid residue, usually a hydrophobic residue; "s", a small residue; "B", asparagine or aspartate) from the N-terminus to the C-terminus. While the presence of two conserved glutamates and two histidines matches the active-site composition of several well-studied metal-dependent hydrolases (e.g., thermolysin,[37] bovine carboxypeptidase A,[38] D-Ala-D-Ala carboxypeptidase[39] and S2P-like proteases[11]), their linear motifs used in metal binding and catalysis are unlike each other and the CPBP family. Accordingly, these enzymes have different structural folds and belong to different clans in the MEROPS peptidase database.[40] Three residues are used for metal binding, and one residue (usually glutamate) activates a metal-bound water molecule as the nucleophile to attack the substrate peptide bond. The presence of such a set of conserved residues in the CPBP family supports the prediction that most members in this family are metal-dependent proteases. The positively charged side chain of the conserved arginine residue in the EExxxR motif may contribute to the oxyanion hole that accommodates the carbonyl group of the substrate's scissile bond. The third motif (sxxxs), frequently found in transmembrane proteins,[41] may be involved in helical packing similar to the one observed in the second predicted core transmembrane segment in S2P.[10]

Limited experiments have been carried out to investigate the function of the conserved glutamates and histidines in CPBP members. However, their functional importance is suggested by the results of several site-directed mutagenesis experiments performed on the yeast type II CAAX protease Rce1p[26,28] and on a bacterial CPBP member SkkI from *Lactobacillus sakei* that confers immunity to bacteriocins.[42] In one site-directed mutagenesis study of yeast Rce1p,[28] mutation of any one of the conserved glutamates and histidines (E156A or E157A in the first motif, H194A in the second motif or H248A in the fourth motif) inactivated Rce1p's enzymatic activity according to an *in vitro* assay. The same results for E156A, H194A and H248A were also obtained in a previous mutagenesis study.[26] Mutation of the conserved aromatic residue in the second motif (F190A) or the asparagine in the fourth motif (N252A) resulted in reduced enzymatic activity.[28] An *in vivo* assessment of Rce1p's activity based on a mating assay exhibited results similar to those of the *in vitro* assessment.[28] The importance of the negative charge in the first conserved glutamate in the first motif was suggested as the E156Q mutant had no *in vivo* activity under normal or over-expressed conditions (E156D had marginal activity though).[28] Mutations of several residues that are not conserved across the CPBP family did not affect the activity of Rce1p.[28] For the *L. sakei* immunity protein SkkI, double mutants of the two conserved glutamates (E133A/E134A or E133Q/E134Q) in the first motif or a single mutant of the conserved histidine in the fourth motif (H214D) lost their ability to confer immunity.[42] Further experimental studies are required to elucidate the catalytic mechanism of CPBP family members and the exact functional roles of the conserved residues.

A sequence clustering by the BLASTClust program was made on a nonredundant set of domains in this family (sequence redundancy removed at 90% level by Cd-hit[43]). Out of 76 clusters with five or more members, 34 clusters (including the two largest clusters and capturing most of the CPBP members in this nonredundant set) contain the EExxxR, FxxxH, sxxxs and HxxxB motifs in the first, second, third and fourth predicted core transmembrane segments, respectively (Supplementary Table S1). The other clusters with five or more members show variations in at least one of these sequence motifs (Supplementary Table S1). For example, several BLASTClust clusters have a "WxxxH" motif instead of the FxxxH motif in the second predicted core transmembrane helix. In the CLANS[44] visualization results (Supplementary Fig. S1), most of the WxxxH-containing members are close to each other and form a distinct sequence group. In two BLASTClust clusters (clusters 3 and 22 in Supplementary Table S1), the second motif, instead of the first motif, harbors a conserved arginine (the first and second motifs become "EExxxx" and "RxxxH", respectively, in these two clusters; Fig. 1). Some clusters of proteins in this family (Supplementary Table S1) have one or several amino acid replacements at the conserved glutamate and histidine positions, most probably rendering such family members proteolytically inactive (e.g., the LyrA protein from *Staphylococcus aureus*[45]).

Eukaryotic type II CAAX proteases proteolytically remove the C-terminal tripeptide AAX. This tripeptide lies directly C-terminal to a cysteine residue modified with a farnesyl (C_{15}) or a geranylgeranyl (C_{20}) prenyl chain, which facilitates membrane localization.[26] These proteases form a distinct sequence group in the CLANS diagram (Supplementary Fig. S1). Despite the fact that most CPBP members are from bacteria, only a few bacterial members have been studied experimentally, and details of their function remain elusive. Based on available data, the function of many of the uncharacterized bacterial CPBP members is likely related to bacteriocin maturation. Bacteriocins are polypeptide toxins secreted by virtually all bacteria to inhibit the growth of competing bacterial strains/species.[46,47] The proposed protease activity of bacterial CPBP members could be utilized in the maturation and secretion process of bacteriocins and/or help conferring immunity against self-produced bacteriocins. For example, a number of CPBP members (PlnI, PlnL, PlnT, PlnP, PlnV, PlnU and PlnW) were found in operons that produce bacteriocins in various strains of *Lactobacillus plantarum*.[30,48] Recently, it was shown that CPBP members SkkI from *L. sakei* and PlnI and PlnL from *L. plantarum* confer immunity to their cognate bacteriocins.[42] In *Streptococcus pneumoniae*, CPBP members PncO and PncP were found in the *pnc* locus encoding bacteriocins and mutation of PncO abolished bacteriocin production.[49] A CPBP member, SagE from *Streptococcus pyogenes*, is encoded by a biosynthetic operon that produces streptolysin S, a toxin generated by posttranslational modification of a bacteriocin-like precursor.[50–52] Toxins similar to streptolysin S have been found in several other Gram-positive pathogens, such as *S. aureus*, *Listeria monocytogenes* and *Clostridium botulinum*. In all cases, a type II CAAX protease (SagE homolog) is found in the biosynthetic gene cluster.[53] In the Gram-negative *Klebsiella pneumoniae*, the CPBP member MceF has been shown to be important in the export of microcin E492.[54] Another CPBP member, MlrA from *Sphingomunas* sp., which has a second motif WxxxH instead of FxxxH, was characterized as an endopeptidase (microcystinase) responsible for the degradation of the cyanobacterial toxin microcystin LR.[55] Inhibition of MlrA by ethylenediaminetetraacetic acid is consistent with the prediction that CPBP members are metalloproteases. The four conserved glutamates and histidines within the sequence motifs are present in most of the abovementioned examples of bacterial CPBP members. Some of these examples, such as PlnP, SkkI, PncO, SagE and MlrA, are included in the alignment shown in Fig. 1, with conserved residues highlighted.

Except for the protease domain, most of the CPBP members appear to lack other known domains as revealed by profile HMM-based[56] searches against the Pfam database[34] (about 96% of CPBP members have no more than one Pfam domain hits at the default *e*-value cutoff of 0.01). Domain composition analysis offers insights into the function of some multi-domain CPBPs. For example, some CPBP members from *Bacillus* species (e.g., gi: 206968619) are annotated as AbrB family transcriptional regulators, as they also contain a DNA-binding AbrB domain at the C-terminus, suggesting that the protease activity may be involved in transcriptional regulation. Functional association with AbrB domains is also supported by a recent study of an AbrB family protein CalA in *Nostoc* sp., which was shown to be co-transcribed with a CPBP member.[57] In some CPBP members, protease domains are fused with components of ABC transporters into a single protein (e.g., gi: 254975247). In other cases, some genes encoding CPBP members are neighbors of genes encoding components of ABC transporters, as revealed in the bacteriocin operon clusters[30] and by the STRING functional association server (e.g., gi: 118476928).[58] Co-expression of CPBP members with ABC transporters was found in a recent study of the two-component system BfrAB involved in biofilm formation in *Streptococcus* species.[59] These lines of evidence suggest that some CPBP members function to couple leader peptide cleavage with cellular export in a similar fashion to certain bacteriocin ABC transporters with protease domains.[60]

The PrsW protease family

This group includes close homologs of the experimentally characterized PrsW protease from *B. subtilis*.[31,61] PrsW mediates site-1 cleavage of the anti-σ factor RsiW, one of the two steps of proteolysis events that lead to activation of transcription factor σ^W in response to antimicrobial peptides such as bacteriocins and cell envelop stress. Several noticeable differences in sequence motifs exist between PrsW proteases and the CPBP family (Fig. 1). First, the most N-terminal motif of interest from PrsW proteases bears the consensus signature of "EExxK" instead of EExxxR as in most others CPBPs. Second, PrsW proteases have the second motif "FxxxE" in place of FxxxH as in the CPBP family. Third, PrsW proteases possess a conserved histidine in the third motif, which is absent in the CPBP family. The fourth motif, HxxxB, is shared by the PrsW proteases and the CPBP family. Limited experiments have been performed on PrsW members, and the only site-directed mutagenesis study was available for the PrsW member from *B. subtilis*.[31] Either the double point mutation of the two conserved glutamates in the first motif (E75A/E76A) or a single mutation of the conserved histidine in the fourth motif (H175A) was not able to complement the phenotype caused by the *prsW*-null mutation,[31] suggesting the functional importance of these conserved residues for *B. subtilis* PrsW. Further experimental studies are needed to determine the functional roles of the conserved residues in PrsW members.

Transitive PSI-BLAST searches found about 500 PrsW members, most of which are from the Gram-positive bacterial phyla of firmicutes and actinobacteria. Some PrsW members (e.g., gi: 298242689) contain a forkhead-associated domain[62] at the N-terminus. Sequence clustering by the CLANS program (Supplementary Fig. S2) revealed two distinct groups of archaeal PrsW members. One archaeal group is closely related to and clustered with members from bacterial species. The second archaeal group forms a separate cluster and has relatively high sequence divergence among its members. Proteins in the second archaeal group also contain an ABC transporter permease domain, suggesting their roles in membrane transportation. The distribution of PrsW proteases in eukaryotes is dramatically restricted to only a few unicellular organisms, suggesting massive gene loss in most eukaryotic lineages. PrsW proteases were found in the diatom species *Thalassiosira pseudonana* and *Phaeodactylum tricornutum*, in apicomplexa from genera *Plasmodium* and *Cryptosporidium*, in the percolozoan *Naegleria gruberi* and in the two fungal species *Neosartorya fischeri* and *Aspergillus fumigatus*.

The DUF2324 family

This family includes proteins from the Pfam family DUF2324, with about 200 proteins in the current sequence database. They are mainly from firmicutes and proteobacteria. Motif variation from the CPBP family and the PrsW family is observed in DUF2324 (Fig. 1), with its first, second and third motifs being "EExxR", "HxxxE" and "[HQ]xxxs", respectively. The fourth motif HxxxB remains the same. The function of these proteins has not been revealed by experimental studies, and many of the proteins in this group are annotated as hypothetical proteins. PSI-BLAST searches using DUF2324 members as queries also identified a distantly related group of archaeal proteins with another version of sequence motifs ("QExxK", FxxxE, "Hxxxs" and HxxxB) similar to those of the PrsW proteases (Fig. 1).

The APH-1 family

This family includes the APH-1 subunits of γ-secretases. APH-1 proteins are restricted to eukaryotes. Widely distributed in metazoans and plants, APH-1s were also found in amoebozoans such as *Dictyostelium discoideum*, in the two parasitic euglenozoan genera *Trypanosoma* and *Leishmania* and in several stramenopiles such as *P. tricornutum* (a diatom), *Ectocarpus siliculosus* (a brown algae), *Phytophthora infestans* (an oomycete) and *Blastocystis hominis*. The gene encoding APH-1 appears to have been lost in fungi and may also be absent in many other unicellular eukaryotic species (fast evolutionary rates in certain lineages could prevent detection of some APH-1s). APH-1 proteins have a similar but different set of motifs as compared to other families. Their four motifs are "QExxR", "Fxxxx", Hxxxs and Hxxxs, respectively (Fig. 1).

The remote similarity between APH-1 and other protease families is an interesting finding. The presenilin subunit has been the focus of study for γ-secretase as it is regarded as the catalytic subunit.[63] On the other hand, much less is known about the exact function and evolutionary origin of the subunit APH-1. Given the inferred homology between APH-1 and other membrane proteases such as type II CAAX proteases and PrsW proteases, we speculate that APH-1 may possess protease activity, thus expanding the range of substrates for γ-secretase. It should be noted that another subunit of γ-secretase, nicastrin, also possesses a protease domain, which is located in the extracellular region and has uncharacterized function. As no protease activity of APH-1 has been shown experimentally, it is likely that APH-1 lost the protease activity and has other functional roles in the γ-secretase complex. In this scenario, it is plausible that APH-1 in the ancestral form of γ-secretase was catalytically active and performed proteolytic reactions along with the presenilin subunit.

Supplementary materials related to this article can be found online at doi:10.1016/j.jmb.2011.04.066

Acknowledgements

We would like to thank David Baker, Vladimir Yarov-Yarovoy and Gang Yu for helpful discussions and Lisa Kinch for critical reading of the manuscript. This work was supported in part by National Institutes of Health (1R01GM094575-01 to N.V.G. and 2R01GM090328-06A1 to J.E.D.) and the Welch Foundation (I-1505 to N.V.G.). D.A.M. would like to thank the Department of Chemistry and the Institute for Genomic Biology at the University of Illinois at Urbana-Champaign for funding.

References

1. Brown, M. S., Ye, J., Rawson, R. B. & Goldstein, J. L. (2000). Regulated intramembrane proteolysis: a control mechanism conserved from bacteria to humans. *Cell*, **100**, 391–398.
2. Wolfe, M. S. & Kopan, R. (2004). Intramembrane proteolysis: theme and variations. *Science*, **305**, 1119–1123.
3. Ben-Shem, A., Fass, D. & Bibi, E. (2007). Structural basis for intramembrane proteolysis by rhomboid serine proteases. *Proc. Natl Acad. Sci. USA*, **104**, 462–466.
4. Urban, S. & Shi, Y. (2008). Core principles of intramembrane proteolysis: comparison of rhomboid and site-2 family proteases. *Curr. Opin. Struct. Biol.* **18**, 432–441.
5. Wang, X., Sato, R., Brown, M. S., Hua, X. & Goldstein, J. L. (1994). SREBP-1, a membrane-bound

transcription factor released by sterol-regulated proteolysis. *Cell*, **77**, 53–62.

6. Brown, M. S. & Goldstein, J. L. (1999). A proteolytic pathway that controls the cholesterol content of membranes, cells, and blood. *Proc. Natl Acad. Sci. USA*, **96**, 11041–11048.
7. An, F. Y., Sulavik, M. C. & Clewell, D. B. (1999). Identification and characterization of a determinant (*eep*) on the *Enterococcus faecalis* chromosome that is involved in production of the peptide sex pheromone cAD1. *J. Bacteriol.* **181**, 5915–5921.
8. Rudner, D. Z., Fawcett, P. & Losick, R. (1999). A family of membrane-embedded metalloproteases involved in regulated proteolysis of membrane-associated transcription factors. *Proc. Natl Acad. Sci. USA*, **96**, 14765–14770.
9. Alba, B. M., Leeds, J. A., Onufryk, C., Lu, C. Z. & Gross, C. A. (2002). DegS and YaeL participate sequentially in the cleavage of RseA to activate the ς^{E}-dependent extracytoplasmic stress response. *Genes Dev.* **16**, 2156–2168.
10. Kinch, L. N., Ginalski, K. & Grishin, N. V. (2006). Site-2 protease regulated intramembrane proteolysis: sequence homologs suggest an ancient signaling cascade. *Protein Sci.* **15**, 84–93.
11. Feng, L., Yan, H., Wu, Z., Yan, N., Wang, Z., Jeffrey, P. D. & Shi, Y. (2007). Structure of a site-2 protease family intramembrane metalloprotease. *Science*, **318**, 1608–1612.
12. Urban, S. (2006). Rhomboid proteins: conserved membrane proteases with divergent biological functions. *Genes Dev.* **20**, 3054–3068.
13. Wang, Y., Zhang, Y. & Ha, Y. (2006). Crystal structure of a rhomboid family intramembrane protease. *Nature*, **444**, 179–180.
14. Wu, Z., Yan, N., Feng, L., Oberstein, A., Yan, H., Baker, R. P. *et al.* (2006). Structural analysis of a rhomboid family intramembrane protease reveals a gating mechanism for substrate entry. *Nat. Struct. Mol. Biol.* **13**, 1084–1091.
15. Lemieux, M. J., Fischer, S. J., Cherney, M. M., Bateman, K. S. & James, M. N. (2007). The crystal structure of the rhomboid peptidase from *Haemophilus influenzae* provides insight into intramembrane proteolysis. *Proc. Natl Acad. Sci. USA*, **104**, 750–754.
16. Dries, D. R. & Yu, G. (2008). Assembly, maturation, and trafficking of the γ-secretase complex in Alzheimer's disease. *Curr. Alzheimer Res.* **5**, 132–146.
17. Steiner, H., Fluhrer, R. & Haass, C. (2008). Intramembrane proteolysis by γ-secretase. *J. Biol. Chem.* **283**, 29627–29631.
18. Weihofen, A., Binns, K., Lemberg, M. K., Ashman, K. & Martoglio, B. (2002). Identification of signal peptide peptidase, a presenilin-type aspartic protease. *Science*, **296**, 2215–2218.
19. LaPointe, C. F. & Taylor, R. K. (2000). The type 4 prepilin peptidases comprise a novel family of aspartic acid proteases. *J. Biol. Chem.* **275**, 1502–1510.
20. Shah, S., Lee, S. F., Tabuchi, K., Hao, Y. H., Yu, C., LaPlant, Q. *et al.* (2005). Nicastrin functions as a γ-secretase-substrate receptor. *Cell*, **122**, 435–447.
21. Boyartchuk, V. L., Ashby, M. N. & Rine, J. (1997). Modulation of Ras and a-factor function by carboxyl-terminal proteolysis. *Science*, **275**, 1796–1800.
22. Schmidt, W. K., Tam, A., Fujimura-Kamada, K. & Michaelis, S. (1998). Endoplasmic reticulum membrane localization of Rce1p and Ste24p, yeast proteases involved in carboxyl-terminal CAAX protein processing and amino-terminal a-factor cleavage. *Proc. Natl Acad. Sci. USA*, **95**, 11175–11180.
23. Maurer-Stroh, S., Washietl, S. & Eisenhaber, F. (2003). Protein prenyltransferases. *Genome Biol.* **4**, 212.
24. Schmidt, W. K., Tam, A. & Michaelis, S. (2000). Reconstitution of the Ste24p-dependent N-terminal proteolytic step in yeast a-factor biogenesis. *J. Biol. Chem.* **275**, 6227–6233.
25. Tam, A., Schmidt, W. K. & Michaelis, S. (2001). The multispanning membrane protein Ste24p catalyzes CAAX proteolysis and NH2-terminal processing of the yeast a-factor precursor. *J. Biol. Chem.* **276**, 46798–46806.
26. Dolence, J. M., Steward, L. E., Dolence, E. K., Wong, D. H. & Poulter, C. D. (2000). Studies with recombinant *Saccharomyces cerevisiae* CaaX prenyl protease Rce1p. *Biochemistry*, **39**, 4096–4104.
27. Pei, J. & Grishin, N. V. (2001). Type II CAAX prenyl endopeptidases belong to a novel superfamily of putative membrane-bound metalloproteases. *Trends Biochem. Sci.* **26**, 275–277.
28. Plummer, L. J., Hildebrandt, E. R., Porter, S. B., Rogers, V. A., McCracken, J. & Schmidt, W. K. (2006). Mutational analysis of the ras converting enzyme reveals a requirement for glutamate and histidine residues. *J. Biol. Chem.* **281**, 4596–4605.
29. Dechert, A. M., MacNamara, J. P., Breevoort, S. R., Hildebrandt, E. R., Hembree, N. W., Rea, A. C. *et al.* (2010). Modulation of the inhibitor properties of dipeptidyl (acyloxy)methyl ketones toward the CaaX proteases. *Bioorg. Med. Chem.* **18**, 6230–6237.
30. Diep, D. B., Havarstein, L. S. & Nes, I. F. (1996). Characterization of the locus responsible for the bacteriocin production in *Lactobacillus plantarum* C11. *J. Bacteriol.* **178**, 4472–4483.
31. Ellermeier, C. D. & Losick, R. (2006). Evidence for a novel protease governing regulated intramembrane proteolysis and resistance to antimicrobial peptides in *Bacillus subtilis*. *Genes Dev.* **20**, 1911–1922.
32. 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. *Nucleic Acids Res.* **25**, 3389–3402.
33. Soding, J. (2005). Protein homology detection by HMM–HMM comparison. *Bioinformatics*, **21**, 951–960.
34. Finn, R. D., Mistry, J., Tate, J., Coggill, P., Heger, A., Pollington, J. E. *et al.* (2010). The Pfam protein families database. *Nucleic Acids Res.* **38**, D211–D222.
35. Pei, J. & Grishin, N. V. (2007). PROMALS: towards accurate multiple sequence alignments of distantly related proteins. *Bioinformatics*, **23**, 802–808.
36. O'Connor, L., Coffey, A., Daly, C. & Fitzgerald, G. F. (1996). AbiG, a genotypically novel abortive infection mechanism encoded by plasmid pCI750 of *Lactococcus lactis* subsp. *cremoris* UC653. *Appl. Environ. Microbiol.* **62**, 3075–3082.
37. Matthews, B. W., Colman, P. M., Jansonius, J. N., Titani, K., Walsh, K. A. & Neurath, H. (1972). Structure of thermolysin. *Nat. New Biol.* **238**, 41–43.

38. 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.
39. Dideberg, O., Charlier, P., Dive, G., Joris, B., Frere, J. M. & Ghuysen, J. M. (1982). Structure of a Zn^{2+}-containing D-alanyl-D-alanine-cleaving carboxypeptidase at 2.5 Å resolution. *Nature*, **299**, 469–470.
40. Rawlings, N. D., Barrett, A. J. & Bateman, A. (2010). MEROPS: the peptidase database. *Nucleic Acids Res.* **38**, D227–D233.
41. Russ, W. P. & Engelman, D. M. (2000). The GxxxG motif: a framework for transmembrane helix–helix association. *J. Mol. Biol.* **296**, 911–919.
42. Kjos, M., Snipen, L., Salehian, Z., Nes, I. F. & Diep, D. B. (2010). The abi proteins and their involvement in bacteriocin self-immunity. *J. Bacteriol.* **192**, 2068–2076.
43. Li, W. & Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, **22**, 1658–1659.
44. Frickey, T. & Lupas, A. (2004). CLANS: a Java application for visualizing protein families based on pairwise similarity. *Bioinformatics*, **20**, 3702–3704.
45. Grundling, A., Missiakas, D. M. & Schneewind, O. (2006). *Staphylococcus aureus* mutants with increased lysostaphin resistance. *J. Bacteriol.* **188**, 6286–6297.
46. Cotter, P. D., Hill, C. & Ross, R. P. (2005). Bacteriocins: developing innate immunity for food. *Nat. Rev., Microbiol.* **3**, 777–788.
47. Drider, D., Fimland, G., Hechard, Y., McMullen, L. M. & Prevost, H. (2006). The continuing story of class IIa bacteriocins. *Microbiol. Mol. Biol. Rev.* **70**, 564–582.
48. Rojo-Bezares, B., Saenz, Y., Navarro, L., Jimenez-Diaz, R., Zarazaga, M., Ruiz-Larrea, F. & Torres, C. (2008). Characterization of a new organization of the plantaricin locus in the inducible bacteriocin-producing *Lactobacillus plantarum* J23 of grape must origin. *Arch. Microbiol.* **189**, 491–499.
49. Lux, T., Nuhn, M., Hakenbeck, R. & Reichmann, P. (2007). Diversity of bacteriocins and activity spectrum in *Streptococcus pneumoniae*. *J. Bacteriol.* **189**, 7741–7751.
50. Datta, V., Myskowski, S. M., Kwinn, L. A., Chiem, D. N., Varki, N., Kansal, R. G. *et al.* (2005). Mutational analysis of the group A streptococcal operon encoding streptolysin S and its virulence role in invasive infection. *Mol. Microbiol.* **56**, 681–695.
51. Lee, S. W., Mitchell, D. A., Markley, A. L., Hensler, M. E., Gonzalez, D., Wohlrab, A. *et al.* (2008). Discovery of a widely distributed toxin biosynthetic gene cluster. *Proc. Natl Acad. Sci. USA*, **105**, 5879–5884.
52. Mitchell, D. A., Lee, S. W., Pence, M. A., Markley, A. L., Limm, J. D., Nizet, V. & Dixon, J. E. (2009). Structural and functional dissection of the heterocyclic peptide cytotoxin streptolysin S. *J. Biol. Chem.* **284**, 13004–13012.
53. Cotter, P. D., Draper, L. A., Lawton, E. M., Daly, K. M., Groeger, D. S., Casey, P. G. *et al.* (2008). Listeriolysin S, a novel peptide haemolysin associated with a subset of lineage I *Listeria monocytogenes*. *PLoS Pathog.* **4**, e1000144.
54. Lagos, R., Baeza, M., Corsini, G., Hetz, C., Strahsburger, E., Castillo, J. A. *et al.* (2001). Structure, organization and characterization of the gene cluster involved in the production of microcin E492, a channel-forming bacteriocin. *Mol. Microbiol.* **42**, 229–243.
55. Saito, T., Okano, K., Park, H. D., Itayama, T., Inamori, Y., Neilan, B. A. *et al.* (2003). Detection and sequencing of the microcystin LR-degrading gene, *mlrA*, from new bacteria isolated from Japanese lakes. *FEMS Microbiol. Lett.* **229**, 271–276.
56. Eddy, S. R. (2009). A new generation of homology search tools based on probabilistic inference. *Genome Inform.* **23**, 205–211.
57. Agervald, A., Zhang, X., Stensjo, K., Devine, E. & Lindblad, P. (2010). CalA, a cyanobacterial AbrB protein, interacts with the upstream region of *hypC* and acts as a repressor of its transcription in the cyanobacterium *Nostoc* sp. strain PCC 7120. *Appl. Environ. Microbiol.* **76**, 880–890.
58. Jensen, L. J., Kuhn, M., Stark, M., Chaffron, S., Creevey, C., Muller, J. *et al.* (2009). STRING 8—a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res.* **37**, D412–D416.
59. Zhang, Y., Whiteley, M., Kreth, J., Lei, Y., Khammanivong, A., Evavold, J. N. *et al.* (2009). The two-component system BfrAB regulates expression of ABC transporters in *Streptococcus gordonii* and *Streptococcus sanguinis*. *Microbiology*, **155**, 165–173.
60. Havarstein, L. S., Diep, D. B. & Nes, I. F. (1995). A family of bacteriocin ABC transporters carry out proteolytic processing of their substrates concomitant with export. *Mol. Microbiol.* **16**, 229–240.
61. Heinrich, J. & Wiegert, T. (2006). YpdC determines site-1 degradation in regulated intramembrane proteolysis of the RsiW anti-sigma factor of *Bacillus subtilis*. *Mol. Microbiol.* **62**, 566–579.
62. Durocher, D., Henckel, J., Fersht, A. R. & Jackson, S. P. (1999). The, FHA domain is a modular phosphopeptide recognition motif. *Mol. Cell*, **4**, 387–394.
63. Ahn, K., Shelton, C. C., Tian, Y., Zhang, X., Gilchrist, M. L., Sisodia, S. S. & Li, Y. M. (2010). Activation and intrinsic γ-secretase activity of presenilin 1. *Proc. Natl Acad. Sci. USA*, **107**, 21435–21440.