Protein Science (1995), *4*:1247–1261. Cambridge University Press. Printed in the USA.
Copyright © 1995 The Protein Society

REVIEW

The astacin family of metalloendopeptidases

JUDITH S. BOND[1] AND ROBERT J. BEYNON[2]

[1] Department of Biochemistry and Molecular Biology, The Pennsylvania State University College of Medicine, Hershey, Pennsylvania 17033
[2] Department of Biochemistry and Applied Molecular Biology, University of Manchester Institute of Science and Technology, Manchester M60 1QD, United Kingdom

(RECEIVED March 23, 1995; ACCEPTED April 19, 1995)

Abstract

The astacin family of metalloendopeptidases was recognized as a novel family of proteases in the 1990s. The crayfish enzyme astacin was the first characterized and is one of the smallest members of the family. More than 20 members of the family have now been identified. They have been detected in species ranging from hydra to humans, in mature and in developmental systems. Proposed functions of these proteases include activation of growth factors, degradation of polypeptides, and processing of extracellular proteins. Astacin family proteases are synthesized with NH_2-terminal signal and proenzyme sequences, and many (such as meprins, BMP-1, *tolloid*) contain multiple domains COOH-terminal to the protease domain. They are either secreted from cells or are plasma membrane-associated enzymes. They have some distinguishing features in addition to the signature sequence in the protease domain: HEXXHXXGFXHEXXRXDR. They have a unique type of zinc binding, with pentacoordination, and a protease domain tertiary structure that contains common attributes with serralysins, matrix metalloendopeptidases, and snake venom proteases; they cleave peptide bonds in polypeptides such as insulin B chain and bradykinin and in proteins such as casein and gelatin; and they have arylamidase activity. Meprins are unique proteases in the astacin family, and indeed in the animal kingdom, in their oligomeric structure; they are dimers of disulfide-linked dimers and are highly glycosylated, type I integral membrane proteins that have many attributes of receptors or integrins with adhesion, epidermal growth factor-like, and transmembrane domains. The α and β subunits are differentially expressed and processed to yield latent and active proteases as well as membrane-associated and secreted forms. Meprins represent excellent models of hetero- and homo-oligomeric enzymes that are regulated at the transcriptional and posttranslational levels.

Keywords: astacin; bone morphogenetic protein-1; brush border membranes; development; meprin; metalloendopeptidases; secreted and plasma membrane proteases; *tolloid*

The great abundance and variety of proteases in nature have stimulated scientists to look for commonalities in these enzymes and to classify them. By the end of the 1980s there was an accepted system to classify endopeptidases on the basis of mechanism of action and three-dimensional structure (Neurath, 1989). Four mechanistic classes of proteases (serine, cysteine, aspartic, and metallo) were identified, and six families (two serine families: chymotrypsin and subtilisin; one cysteine: represented by papain; one aspartic: penicillopepsin; two metallo: carboxypeptidase and thermolysin families). Each family had characteristic active site residues and was thought to have descended from a common ancestor by divergent evolution. However, as a result of the information explosion that occurred along with the advances in molecular and cell biology, as many as 60 families of peptidases (endo- and exopeptidases) have now been proposed on the basis of similarities in primary structure (Rawlings & Barrett, 1993).

Reprint requests to: Judith S. Bond, Department of Biochemistry and Molecular Biology, Pennsylvania State University College of Medicine, Hershey, Pennsylvania 17033; e-mail: jbond@psuhmc.hmc.psu.edu.

One of the families that emerged in the 1990s is the "astacin family" (Dumermuth et al., 1991). Recognition of this family was a consequence of cloning and sequencing of mammalian meprins (metalloproteases from brush border membranes of rodents and humans). The original members of the family identified were: the crayfish digestive enzyme astacin, bone morphogenetic protein-1 (BMP-1) from human bone, meprins from mouse kidney and human intestine (the latter was originally designated as *N*-benzoyl-L-tyrosyl-*p*-aminobenzoic acid hydrolase or PPH), and UVS.2, a partial sequence from *Xenopus laevis* embryos. The family was named the "astacin family" because a protease from the crayfish *Astacus astacus* L., the enzyme now called astacin (EC 3.4.24.21), was the first to be sequenced and biochemically characterized (Titani et al., 1987). It has subsequently been crystallized and the three-dimensional structure solved, adding a new dimension to our understanding and exploration of these metalloendopeptidases (Bode et al., 1992). Since the discovery of this family, 17 additional members have been cloned and sequenced, and there is evidence for others. It is the aim of this review to summarize information about this new family, point out the unique features of the family and in-

dividual members, and suggest how future investigations can lead to answers to fundamental questions.

Members of the astacin family

Domain structures

Astacin family members are characterized by a unique 18-amino acid signature sequence, HEXXHXXGFXHEXXRXDR, which begins with an HEXXH zinc binding motif found in most metalloendopeptidases. The signature sequence is part of an approximately 200-amino acid sequence, which is the entire mature crayfish astacin and the catalytic or protease domain of all the members of the family (Fig. 1). Signal and prosequences are also common features of family members, with the possible exception of QuCAM-1; these NH_2-terminal domains have yet to be found for this latter protein. Astacin, which until recently was only studied as a mature protein that begins with the protease domain, is now known to contain a prepro segment of 49 residues (W. Stöcker & R. Zwilling, pers. comm.). The transient signal peptides direct the proteins into the endoplasmic reticulum during biosynthesis, which is consistent with the finding that all of the proteins of the family studied thus far are secreted or plasma membrane bound. The prosequences vary greatly in size, containing up to 519 amino acids for *Drosophila tolloid-related-1* (*DrTlr-1*), and are likely to be important for regulating activity and perhaps expression of the proteases. Regarding the latter point, for example, the large prosequence of *DrTlr-1* has been suggested to prevent expression of this gene product in early stages of embryogenesis when cell cycles are very short (Nguyen et al., 1994).

The smallest members of the family have no domains COOH-terminal to the protease domain. They are: crayfish astacin (EC 3.4.24.21) and the teleostean choriolysin L (LCE; EC 3.4.24.66) and choriolysin H (HCE 1 and 2; EC 3.4.24.67). They are secreted proteases with molecular masses of approximately 25 kDa.

Most of the known astacin family members contain one or more copies of an EGF (epidermal growth factor)-like (E) and/or a CUB (complement subcomponents C1r/C1s, embryonic sea urchin protein Uegf, BMP-1) domain. These are non-catalytic, "interaction" domains, which may promote protein–protein and substrate interactions. EGF-like domains consist of approximately 40 amino acids and have 6 highly conserved cysteine residues, which most probably are linked by intradomain disulfide

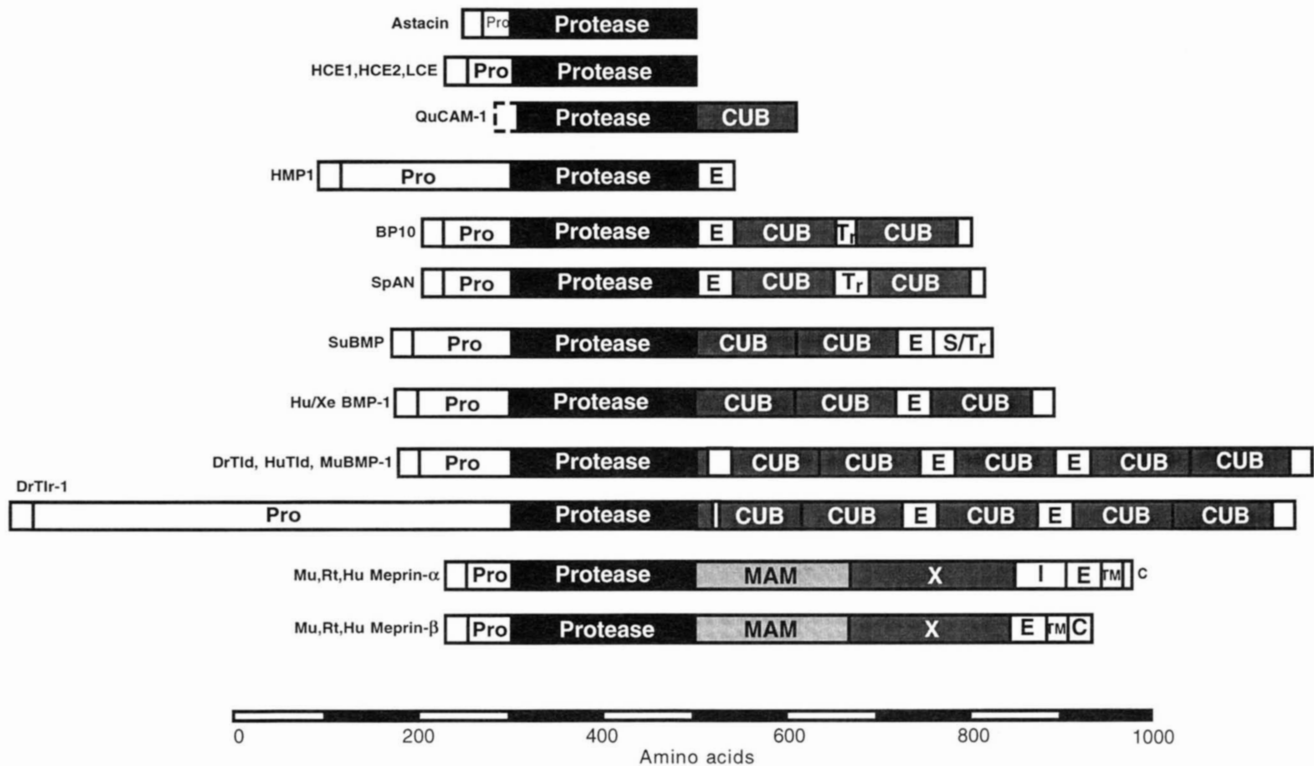

Fig. 1. Domain structure of astacin family members. Domains are aligned relative to the astacin protease domain. All sequences were deduced from cDNA sequences except for astacin in which the protease domain sequence was determined from the mature protein and the prepro sequence deduced from the cDNA. References: astacin, Titani et al. (1987); HCE1, HCE2, and LCE, Yasumasu et al. (1994); QuCAM-1, Elaroussi and DeLuca (1994); HMP1, Sarras (pers. comm.); BP10 (blastula protease-10), Lepage et al. (1992a); SpAN, Reynolds et al. (1992); SuBMP, Hwang et al. (1994); HuBMP-1, Wozney et al. (1988); XeBMP-1, Maéno et al. (1993); DrTld (*Drosophila tolloid*), Shimell et al. (1991); HuTld, Takahara et al. (1994); MuBMP-1, Fukagawa et al. (1994); DrTlr-1 (*Drosophila tolloid-related-1*), Nguyen et al. (1994); Mu meprin-α, Jiang et al. (1992); Rt meprin-α, Corbeil et al. (1992); Hu meprin-α, Dumermuth et al. (1993); Mu meprin-β, Gorbea et al. (1993); Rt meprin-β, Johnson and Hersh (1992); Hu meprin-β, Sterchi (pers. comm). Signal peptides are represented by the NH_2-terminal open boxes; Pro, prosequences; Protease, catalytic domains; E, EGF-like domains; CUB, Clr/s complement-like domains; MAM, adhesion domains; X, unknown; I, inserted domain; TM, putative transmembrane domains; C, cytoplasmic domain (see text for further explanations). Species are indicated as: Qu, quail; Su, sea urchin; Hu, human; Xe, frog; Dr, *Drosophila*; Mu, mouse; Rt, rat.

bonds. EGF-like domains exist in one or more copies on many secreted proteases, such as the coagulation proteases Factors X and IX, protein C, and tissue plasminogen activator, and many plasma membrane proteins such as thrombomodulin, L-selectin, neurexin 1α, and *Neu* differentiation factor (Campbell & Bork, 1993). CUB domains (sometimes referred to as C1r/C1s repeats) consist of approximately 110 amino acids, 4 conserved cysteine residues, and conserved hydrophobicity patterns typical of an antiparallel β-sheet. It has been suggested that they form antiparallel β-barrels similar to those in immunoglobulins (Bork & Beckmann, 1993). They have been found in complement serine proteinases, spermadhesins, neuronal recognition molecule A5, and the tumor necrosis factor-induced protein Tsg6.

The domain structures of the human (Hu) and *Xenopus* (Xe) BMP-1 and human and *Drosophila* (Dr) *tolloid* (*Tld*) and *tolloid-related-1* (*Tlr-1*) are very similar except that *Tld* and *Tlr-1* generally have three extra COOH-terminal domains (2 CUBs and 1 EGF-like domain). However, mouse (Mu) BMP-1 has the same domain structure as *Tld* (Fukagawa et al., 1994). There is now evidence that the shorter BMP-1 variants and the *tolloids* are alternately spliced forms of a single gene in mice and humans (Takahara et al., 1994). It is likely that splice variants of these enzymes exist in most species. In addition, isoforms of these proteases exist as a result of multiple genes. For example, *DrTld* and *DrTlr-1* are the products of two different genes (Nguyen et al., 1994), and different genes have been identified for BMP-1 and *tolloid* (or *xolloid*) in *Xenopus* (R. Albano & L. Dale, pers. comm.).

Sea urchin (Su) SpAN and BP10 have a Thr-rich (Tr) domain and SuBMP has a Ser/Thr-rich (S/Tr) domain that have not been identified in other members of the family. In SpAN, the 40-amino acid domain contains four repeats of the sequence STTTLQTT (Reynolds et al., 1992); in BP10, the 25-amino acid domain contains 14 threonine residues (Lepage et al., 1992a); in SuBMP, the 66-amino acid domain contains 22 serine or threonine residues (Hwang et al., 1994). These regions may contain sites for *O*-glycosylation.

The full-length cDNA sequences of the α and β subunits of meprins are about 45% identical and predict domains COOH-terminal to the protease domain, designated as MAM (meprin subunits, A-5 protein, and receptor protein tyrosine phosphatase μ), X, I, E (EGF-like), TM (putative transmembrane), and C (cytoplasmic). The MAM domain is found in several membrane proteins and has been proposed to act as an adhesion domain (Beckmann & Bork, 1993). Enteropeptidase (also termed enterokinase; EC 3.4.21.9), the intestinal enzyme that converts trypsinogen to trypsin, also contains a MAM domain (Kitamoto et al., 1994). MAM domains are 120–180 amino acids in length, and there are 14 amino acids, including 4 cysteine residues, that are conserved.

The X domains of meprin subunits contain approximately 180 amino acids. No counterparts (matching sequences) have been found in the sequence databases. This domain may be an "extender" region to position the active site of meprin subunits away from the cell membrane. In addition, at least two sites of N-glycosylation of mouse meprin α have been localized to this domain (Jiang et al., 1992). Mouse meprin α and β contain approximately 30% carbohydrate that is attributed to N-glycosylation. The cDNA-deduced amino acid sequences COOH-terminal to the X domain differ in the α and β subunit of meprin, except that both contain EGF-like domains that are 34% identical. The α subunit contains a 56-amino acid inserted domain (I) that has no counterpart in the β subunit, and there is no homology between the putative transmembrane or cytoplasmic domains of α and β (Marchand et al., 1995). The cytoplasmic domains are deduced to be 6 residues for α and 26 residues for β. The β cytoplasmic domain is highly positively charged (seven of a sequence of nine amino acids are Arg/Lys residues). This sequence may help to stabilize the protein in the membrane and may interact with cytoskeletal or other cytoplasmic proteins. This domain contains two potential phosphorylation sites: a protein kinase C phosphorylation site (TRR) and a calmodulin kinase II phosphorylation site (RANT) (Kennelly & Krebs, 1991).

Several other astacin family members have been identified from DNA sequences that are not shown in Figure 1. For example, partial *Caenorhabditis elegans* sequences (CeR151.5, CeF42A10.8, CeC05D11.6) have been identified from genomic DNA sequencing of cosmid clones; the primary sequences of the protease domain of two of these are included in Figure 2. UVS.2, a partial cDNA sequence from *X. laevis* embryos, also identified as an astacin family member, is not included in the figures (Sato & Sargent, 1990; Dumermuth et al., 1991).

The domain structure of the astacin family represents a pattern seen in several protease families, such as the coagulation proteases, complement proteases, proprotein convertases, and matrixins (Furie & Furie, 1988; Matrisian, 1992; Steiner et al., 1992; Sim et al., 1993). The various members of these families result from gene duplication, evolution, gene fusion, and exon shuffling. For all of these families, a conserved protease domain is associated with a variety of noncatalytic domains that yield different proteins, with somewhat different enzymatic activities, regulation, expression, and interactions with small ligands and other proteins. The structures of the precursor processing endopeptidases of the subtilisin family have some striking similarities to those of the astacin family (Steiner et al., 1992). These proteases (e.g., yeast kexin, mammalian furin, PC2, PC3/PC1, PACE) are synthesized with preprosequences followed by highly conserved subtilisin-like catalytic domains (47–72% identical). Only furin and kexin contain TM domains; the others are likely to be soluble enzymes, though they interact with membranes and may have signals that sort them to secretory granules. There are several other COOH-terminal domains, including Ser/Thr-rich, P, amphipathic helix, Cys-rich, and cytoplasmic domains that are present in some of the members. Specific expression of various members can act to generate a diversity of products from identical precursors and are thought to provide finely tuned regulation of physiological and developmental processes.

Proposed functions and tissue distribution

Astacin family enzymes have been proposed to have roles in mature and developmental systems. They are expressed in a tissue-specific manner in mature organisms, and are temporally and spatially expressed in developmental systems (Table 1).

Astacin is synthesized in the crayfish hepatopancreas, an organ that has intestinal, hepatic, and pancreatic functions (Vogt et al., 1989). It is stored extracellularly as an active proteinase in the stomach-like cardia, and is thought to be a digestive enzyme in a variety of crustaceans.

Several of the astacin family members are implicated in the hatching process of embryos and in skeleton formation. There

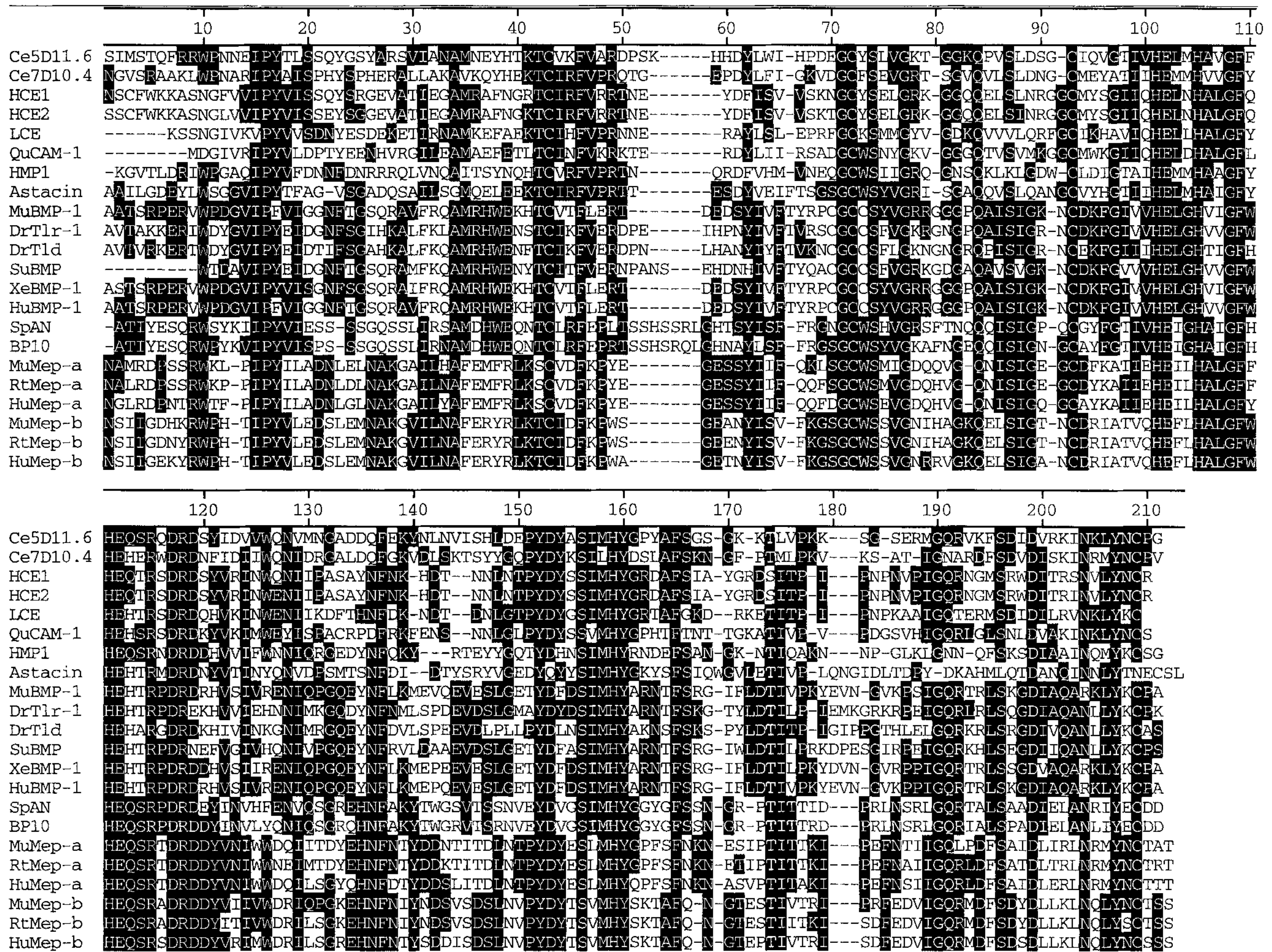

Fig. 2. Sequence comparison of the protease domains of 22 members of the astacin family. Sequences were aligned by the program Megalign from DNASTAR, using the CLUSTAL algorithm and a PAM250 residue weight table. Alignments highlight residues that are in the majority for all sequences. Abbreviations and references as for Figure 1. Included in this alignment (first two sequences) are two partial sequences from nematodes, *C. elegans*, predicted from genomic DNA sequencing (GenBank).

is good evidence to indicate that the fish enzymes HCE1 and 2 and LCE work in concert to degrade the egg shell/chorion of embryos. In *Oryzias latipes*, these enzymes are stored in zymogen granules and are secreted/activated at the appropriate times to serve this function (Yasumasu et al., 1992). HCEs act directly on the "hard" chorion to digest protein into Pro-rich pieces; LCE further degrades the remaining polypeptides (Lee et al., 1994). CAM-1 of quail embryos is also implicated in degradation of eggshell matrix proteins (Elaroussi & DeLuca, 1994). It is induced by 1,25-dihydroxyvitamin D-3 and is proposed to break down extraembryonic chorioallantoic membrane, mobilizing Ca^{+2} for bone formation and weakening the shell in preparation for hatching.

The hydra HMP1 enzyme is thought to be involved in pattern formation and morphogenesis (M.P. Sarras, Jr., pers. comm.). Immunohistochemical studies indicate it has a role in head regeneration and formation of tentacle battery cells, and that it degrades extracellular matrix proteins of the developing tentacle.

There is evidence that the BMP-1 variants and *tolloids* interact with and additionally activate TGF-β (transforming growth factor-β)-like molecules. Human BMP-1 was originally discovered in a complex as one of three proteins (BMP-1-3) isolated from human bone (Wozney et al., 1988). The complex was capable of inducing cartilage and bone formation. BMP-2 and 3 are members of the TGF-β superfamily; these are growth factors active in skeleton/bone formation. The association of BMP-1, a putative metalloendopeptidase, with the complex led to the suggestion that it is required to activate latent TGF-β. The counterpart of mammalian BMP-1 and the TGF-βs in *Drosophila* are *tolloid* and *decapentaplegic* (*dpp*), respectively (Shimell et al., 1991). *Tolloid* and *dpp* are intimately involved in dorsal-ventral pattern formation in early stages of *Drosophila* development: *dpp* null embryos are ventralized and lack all dorsal epidermal structures (Ferguson & Anderson, 1992), and *dpp* is also critical for development of the midgut. Because *tolloid* and TGF-β molecules are highly conserved in other phyla, it has been proposed the interaction between these molecules is a general phenomenon. The discovery that mouse BMP-4 can substitute for *dpp* in *Drosophila* supports this contention (Padgett et al., 1993).

Tolloid-related-1 (*DrTlr-1*; also called *tolkin*) is also implicated in activating growth factors but not *dpp*, because *Tlr-1* cannot substitute for *tolloid*. *Tlr-1* is an essential gene; null alleles of *Tlr-1* result in death in the larval stage (Nguyen et al.,

Table 1. *Tissue distribution and proposed functions of astacin family members*

Member	Species/tissue	Natural substrates (suggested)	Functions (suggested)
Astacin	Crayfish stomach	Ingested proteins	Digestion of food
Choriolysin H (HCE1 and 2)	Fish hatching gland cells	Hard chorion, egg envelope	Embryo hatching
Choriolysin L (LCE)	Fish hatching gland cells	HCE-digested chorion peptides	Embryo hatching
CAM-1	Quail embryos	Extraembryonic chorioallantoic membrane	Releases Ca^{2+} from eggshell protein; hatching
HMP1	Hydra tentacles	Extracellular matrix proteins; fibronectin, type 1 gelatin	Pattern formation; formation of tentacle cells
SpAN and BP10	Sea urchin; early stages of embryogenesis	Latent *univin*	Early developmental decisions
BMP-1	Animal embryos; mouse and human tissues	Latent TGF-β growth factors	Pattern formation; biomineralization; bone/cartilage formation
Tolloid	*Drosophila* embryo; mouse and human embryo	Latent decapentaplegic (DPP)	Dorsal/ventral patterning; early embryo development
Tolloid-related-1	*Drosophila* embryo	Latent growth factors	Larval and pupal development
Meprin A (hetero-oligomers of α and β; homo-oligomers of α subunits)	Mouse (strain-specific) kidney; rat and human kidney and intestine; embryos	Urinary peptides; PTH; MSH; bradykinin; extracellular matrix proteins	Urinary peptide processing; extracellular protein processing and digestion
Meprin B (homo-oligomers of β subunits)	Mouse (strain-specific) kidney and intestine; embryonic	Extracellular proteins	Extracellular protein interactions; remodeling

1994). The protease, first two CUB, and first EGF-like domains of *Tld* and *Tlr-1* are 62% identical, but they have some differences, for example, in the prosequence and in inserted sequences in the first CUB domain (7 amino acids for *Tlr-1*, 26 amino acids for *Tld*) (see Fig. 1). It has been suggested that the different CUB and EGF-like domains in *Tld* and *Tlr-1* result in different interactions with proteins and make them functionally different (R.W. Padgett, pers. comm.).

Mammalian *tolloid*s and BMP-1 are expressed at high levels in developing bones of mouse embryos and in placenta (Takahara et al., 1994). In situ hybridization demonstrated high levels of *tolloid* transcripts, but not BMP-1, in the floor plate of the neural tube of the developing nervous system. Low levels of both *tolloid* and BMP-1 were found in adult human tissues such as heart, liver, placenta, lung, skeletal muscle, kidney, and pancreas; mammalian *tolloid*, but not BMP-1, was detected in adult brain. A third *tolloid* isoform was not detected in adult human tissues but was found in trophoblast giant cells in mouse placenta with the other isoforms.

Sea urchin (Su) BP10 and SpAN are expressed in very early blastula stages, and it has been suggested they play a role in differentiation of ectodermal lineages and subsequent patterning (Lepage et al., 1992b; Reynolds et al., 1992). SpAN expression parallels *univin*, a member of the TGF-β superfamily; this pair may be the equivalent of *tolloid* and *dpp* in *Drosophila* (Stenzel et al., 1994). SuBMP is expressed maximally at the hatched blastula stage, with modest decreases at later stages of development (Hwang et al., 1994). It is detected in ectodermal and primary mesenchyme cells just before the onset of primitive skeleton formation and is suggested to function similarly to HuBMP-1 in bone formation.

Meprin subunits form oligomers that result in meprin A (EC 3.4.24.18; hetero-oligomers of α and β subunits, or homo-oligomers of α subunits) or meprin B (EC 3.4.24.63; a homo-oligomer of β subunits). Meprin A, the name recommended by the International Union of Biochemistry and Molecular Biology (Webb, 1992), has also been referred to as endopeptidase-2 and endopeptidase 24.18 in the rat, and PPH in humans. Meprins are present at high concentrations in brush border membranes, specialized epithelial cell plasma membranes, of mammalian kidney and intestine, and have also been detected in salivary glands and thyroid by immunohistochemical techniques (Craig et al., 1987, 1991; Barnes et al., 1989). The mRNAs for the meprin α and β subunits are abundant in kidney and intestinal tissue, but have not been detected in many other mature mammalian tissues, such as liver, muscle, brain, spleen, and lung (Gorbea et al., 1993; Jiang et al., 1993). Meprins have also been detected on the apical surface of neuroepithelial cells in the inner ear, nasal conchae, ependymal layer of the brain ventricles, and choroid plexus in rat embryos (Spencer-Dene et al., 1994). Suggested functions for meprins in mature systems include degradation of polypeptides such as parathyroid hormone (PTH), α-melanocyte-stimulating hormone (MSH), bradykinin, luliberin (LHRH), substance P, and TGF-α (Stephenson & Kenny, 1988; Choudry & Kenny, 1991; Wolz et al., 1991; Yamaguchi et al., 1991). In addition, meprins are proposed to degrade extracellular matrix proteins such as type IV collagen, laminin, fibronectin, and gelatin (Kaushal et al., 1994). Meprins have also been suggested to process urinary peptides and proteins in rodents and thereby affect release of pheromones and behavior (Bond & Beynon, 1986; Beynon et al., 1995).

Primary structure of the catalytic domain

The primary structure of astacin was determined by amino acid sequencing; all the other amino acid sequences were deduced from the cDNA sequences with some confirmatory peptide sequencing (Fig. 2). The signature sequence for the family spans from astacin His 92 to Asp 109 (Fig. 2, positions 101–118); this sequence contains the three imidazole–zinc ligands. There are

four conserved cysteine residues that are known to form intradomain sulfhydryl bridges in astacin: Cys 42/Cys 198 and Cys 64/Cys 84 (in Fig. 2 these are positions 43/209 or 211 and 72/93). Because of the conservation of cysteine residues in all members of the family, it is likely that equivalent disulfide linkages exist in the protease domains of the other family members. Another conserved region, SXMHY (astacin Ser 145 to Tyr 149; Fig. 2, positions 156–160), contains the tyrosine residue involved in zinc binding and the methionine involved in a "Met-turn" of the peptide chain (Jiang & Bond, 1992; Bode et al., 1993; Stöcker et al., 1995). Stöcker and colleagues (1993) identified other specific residues that are conserved, in comparisons of 10 members of the astacin family, and are probably crucial to the overall structure of the protease domain. Comparison of the 22 family members herein indicates that there are 28 residues that are totally conserved. Seventeen of those are accounted for within the 18-residue family signature sequence, the sequence containing the zinc-binding Tyr, and by the conserved cysteines. The other 11 conserved residues are primarily involved in internal bonds in astacin. These include (astacin residue numbers with Fig. 2 position numbers in parentheses): Pro 16 (16), Phe 45 (46), Gly 63 (71), Ser 66 (74), Gly 69 (77), Gln 75 (84), Phe 128 (137), Tyr 141 (152), Phe 154 (165), Asp 186 (199), and Tyr 194 (207).

The amino acid sequences of the family member protease domains are 29–99% identical (Fig. 3A). The fish enzymes HCE1 and HCE2 are products of two different genes that are highly conserved, i.e., 95% identical (Yasumasu et al., 1994). The domains that are over 80% identical often represent species differences for the same gene products (e.g., human, mouse, and *Xenopus* BMP-1 protease domains are 93–99% identical; mouse, rat, human meprin-αs are 84–90% identical, and meprin-βs are 89% identical).

The phylogenetic tree based on the protease domains (Fig. 3B) indicates several clusters or branches in the family. The meprins form one branch, SpAN and BP10 from sea urchins form another, BMP-1 and *tolloid*s form a third branch, and the fish enzymes (HCE1, HCE2, LCE) and QuCAM-1 form a fourth branch. Phylogenetic trees comparing the EGF-like domains of these enzymes yield the same branching picture for the appropriate members, indicating that the protease and EGF-like domains evolved together (Bond & Jiang, 1995). The finding that meprin α and β subunits are approximately 50% identical in the same species, but each subunit in the mouse, rat, and human is approximately 85% identical, implies that the gene duplication that resulted in α and β preceded speciation.

Until recently, astacin family members have only been detected in the animal kingdom. However, a member of the family (flavastacin) has now been identified in *Flavobacterium meningosepticum* (Tarentino et al., 1995). The gene for flavastacin has been cloned and sequenced and the enzyme isolated. The amino acid sequence of the protease domain is 18% identical with the astacin sequence and 24–29% identical with meprin subunit sequences. The signature sequence is conserved (except for a substitution of an Ile for a Phe at position 109 of Fig. 2), the Met of the "Met-turn" and the zinc-binding Tyr are conserved, and 9 of the 11 other conserved residues thought to be involved in internal bonds in the family are present in flavastacin. Interestingly, flavastacin lacks the four conserved cysteine residues of the other astacin family members. The conservation of most of the residues identified by Stöcker et al. (1993) to be involved in internal bonds may be particularly important to the folding and stability of the bacterial enzyme. The enzyme contains an unusual *O*-linked oligosaccharide, and initial experiments indicate a preference for cleavage of peptides with aspartic acid in the P1′ position. Most significantly, the identification of flavastacin demonstrates that this family was present before the divergence of prokaryotes and eukaryotes.

Tertiary structure of the protease domain

The X-ray crystal structure of astacin has been solved to 1.8 Å resolution (Bode et al., 1992). Astacin has a compact bilobal structure with a long, deep active-site cleft that divides it into two parts (Fig. 4). The NH_2-terminal 100 amino acids are organized into five-stranded pleated β-sheets and two long α-helices. One of the helices forms the top part of the active site of astacin and includes the zinc ligands His 92, His 96, and His 102 (positions 101, 105, and 111 in Fig. 2). The catalytic zinc ion is at the bottom of the cleft, and Tyr 149 (position 160) and a water molecule (SOL 300) that also interacts with Glu 93 (position 102) are the fourth and fifth ligands for zinc. The trigonal bipyramid or pentacoordination of zinc (Fig. 5) is also found in the serralysins and is quite different from that of the thermolysin family, in which two histidines and a glutamic acid residue are ligands to the active-site zinc ion. The zinc coordination of astacin is discussed in detail by Gomis-Rüth et al. (1994) and Stöcker et al. (1995).

The COOH-terminal half of astacin has little defined secondary structure except for a three-turn α-helix before a disulfide bond, which ends the domain and connects it to the NH_2-terminal half. The "Met-turn" at astacin Met 147 (Fig. 2, position 158) has been identified in crystal structures of adamalysins, serralysins, and matrixins, and these have been classified collectively as "metzincins" (Stöcker et al., 1995).

Oligomeric structure and membrane association

Although meprin subunits are the only astacin family members that are known to form homo- and hetero-oligomers and to contain putative COOH-terminal transmembrane domains, there are others that associate with membranes and macromolecules. For example, HCE is tightly associated with the chorion (Yasumasu et al., 1989c). Sea urchin BMP is tightly associated with embryo plasma membranes; however, the interaction could be disrupted by carbonate salts (Hwang et al., 1994). The original finding that human BMP-1 is present in a "complex" with growth factor BMPs also implies somewhat stable interactions of the protease with the growth factors. The noncovalent interactions of astacin family members with membranes and complexes may serve to restrict movement of the active proteases and concentrate activity at specific sites. Investigations of the factors that determine these interactions may have relevance to other proteolytic systems that function at the cell surface, such as complement proteases, plasminogen activators, and matrix metalloproteases (Vassalli et al., 1991; Sim et al., 1993; Strongin et al., 1995).

The quaternary structure of meprin A and its membrane association are unique for known proteases and for brush border enzymes (Fig. 6). Investigations of meprins A and B have been aided by the fact that these enzymes are highly expressed in rodent kidney; the α subunit, for example, has been estimated to comprise 5% of the mouse brush border membrane protein in

A

	Ce7D10.4	HCE1	HCE2	LCE	QuCAM-1	HMP1	Astacin	MuBMP-1	DrTlr-1	DrTld	SuBMP	XeBMP-1	HuBMP-1	SpAN	BP10	MuMep-α	RtMep-α	HuMep-α	MuMep-β	RtMep-β	HuMep-β
Ce5D11.6	50.2	41.4	41.4	41.1	43.8	40.3	37.0	36.0	35.1	35.1	35.1	36.0	36.5	37.9	40.8	37.4	37.4	37.0	41.2	41.7	40.3
Ce7D10.4		40.5	40.0	39.2	39.0	44.5	37.9	34.1	34.6	34.1	33.6	36.0	34.6	31.8	35.1	36.0	35.1	37.9	38.4	38.4	37.9
HCE1			95.7	56.0	51.4	38.1	40.0	37.6	40.5	40.5	36.7	39.5	38.1	37.1	38.1	39.0	40.5	40.0	41.0	39.0	41.4
HCE2				56.5	52.4	38.1	40.5	39.0	41.4	41.4	36.7	41.0	39.5	37.6	38.1	40.0	41.0	40.0	41.4	39.5	41.9
LCE					52.2	35.9	38.8	35.9	38.3	36.8	36.4	35.9	35.9	32.5	35.4	41.6	41.1	39.7	41.6	40.7	41.6
QuCAM-1						34.8	37.1	36.2	37.6	35.2	38.1	36.7	36.2	31.0	33.8	38.1	39.0	39.5	40.0	38.1	40.5
HMP1							34.6	38.4	35.1	35.5	32.7	38.4	38.4	35.5	37.0	34.6	34.6	35.1	36.5	36.0	37.0
Astacin								36.5	35.1	36.0	33.6	37.0	36.5	33.2	35.1	31.6	32.5	31.6	35.8	35.8	35.4
MuBMP-1									65.4	55.0	65.9	92.9	98.6	40.8	40.8	34.1	35.1	34.1	37.4	36.0	36.5
DrTlr-1										70.6	65.9	66.8	66.4	37.4	40.3	32.2	35.1	33.6	36.5	35.5	37.4
DrTld											56.4	56.4	55.0	38.9	41.2	32.7	33.6	33.6	35.1	34.6	34.6
SuBMP												66.4	66.4	36.0	37.4	28.9	29.9	29.9	34.6	33.6	34.1
XeBMP-1													93.4	40.8	41.7	34.6	35.5	35.1	37.4	36.5	37.0
HuBMP-1														40.3	40.3	34.1	35.1	34.1	37.4	36.0	36.5
SpAN															84.4	40.3	41.2	40.8	40.8	40.3	41.7
BP10																40.3	41.7	41.7	42.7	41.7	43.1
MuMep-α																	90.1	83.5	58.0	57.1	58.5
RtMep-α																		85.8	58.5	57.5	58.5
HuMep-α																			56.6	57.1	57.1
MuMep-β																				94.3	88.7
RtMep-β																					89.2

B

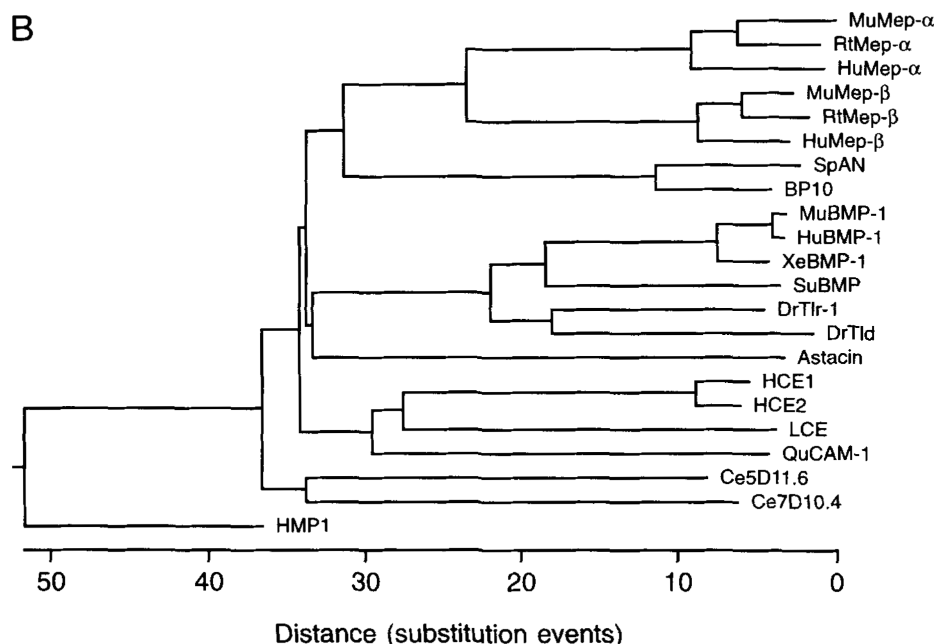

Fig. 3. Comparisons of the sequences in the protease domains. Analyses were produced by Megalign (DNASTAR) based on the alignment in Figure 2. **A:** Percent amino acid sequence identities. **B:** Phylogenetic tree based on sequence alignments. Abbreviations as in Figure 2.

strains that express it (Bond & Beynon, 1986). In addition, some inbred strains of mice express both α and β subunits in the kidney, whereas others express only β, providing good sources of meprin A and B, respectively. Kidney brush border membranes and papain-solubilized forms of meprins have been studied using ICR and BALB/c mice and Sprague-Dawley rats for meprin A and C3H/He mice for meprin B.

Purified preparations of meprins

Papain digestion of brush border membranes does not affect the size of the mouse or rat meprin α subunit, but decreases the size of the β subunit from 110 to 90 kDa (mouse) or from 80 to 74 kDa (rat) by cleaving the subunit near the end of the X domain and thus releasing the enzyme from the membrane

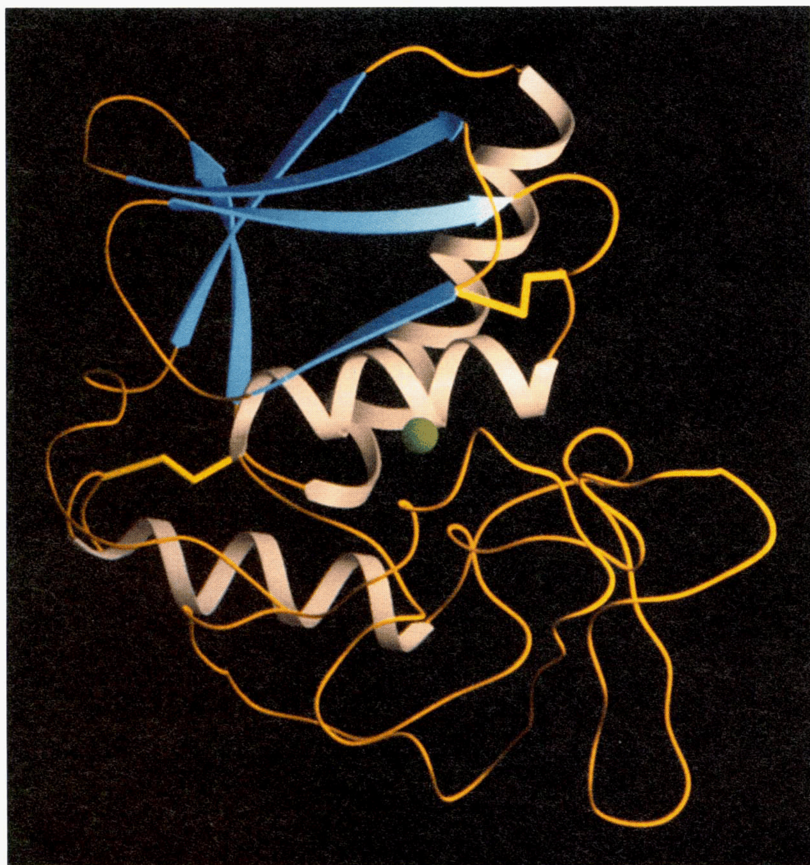

Fig. 4. Ribbon structure of astacin. Images were produced from the PDB file 1AST (Bode et al., 1992) using the program SETOR (Evans, 1993). This view looks into the active site cleft, which runs from left to right in the image. Catalytic zinc atom is represented by a green sphere.

(Marchand et al., 1994). Gel filtration and electrophoresis experiments of papain-solubilized meprin A and B in the presence and absence of reducing agents indicated that these enzymes are disulfide-linked oligomers. Analytical centrifugation experiments with papain-purified preparations of mouse meprin A have now established that this enzyme consists of tetramers, dimers of disulfide-linked dimers. Sedimentation equilibrium analyses, for example, indicated the following molecular weights for

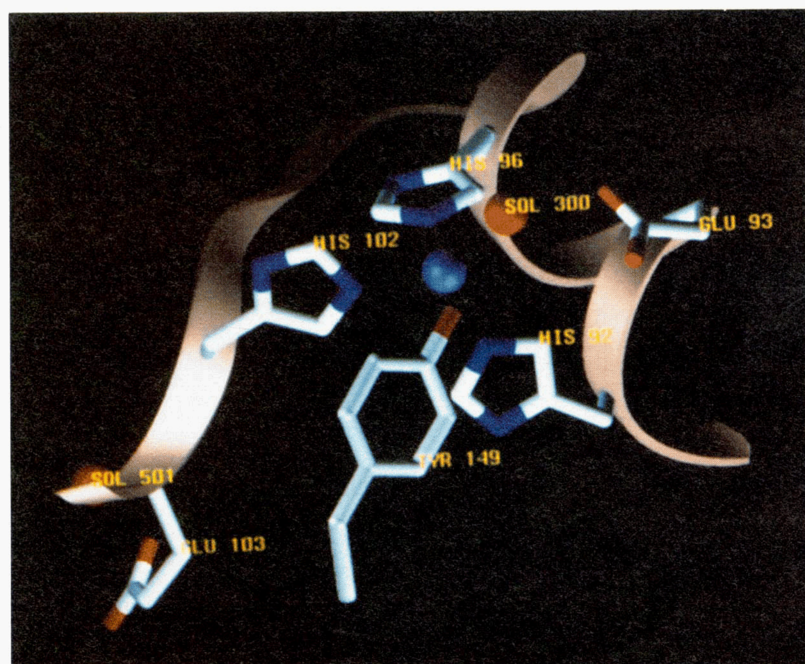

Fig. 5. Zinc coordination in the astacins. Zinc atom (magenta) is complexed by three histidine residues (His 92, 96, and 102 of astacin; positions 101, 105, and 111 of Fig. 2), a glutamate residue (Glu 93; position 102) through a water molecule (SOL 300), and a tyrosine residue (Tyr 149; position 160) from the "Met-turn." Glu 103 (position 112) is oriented by a salt bridge via SOL 501 to the NH_2-terminal residue; in the proenzyme forms, this orientation is not possible and the enzymes are inactive.

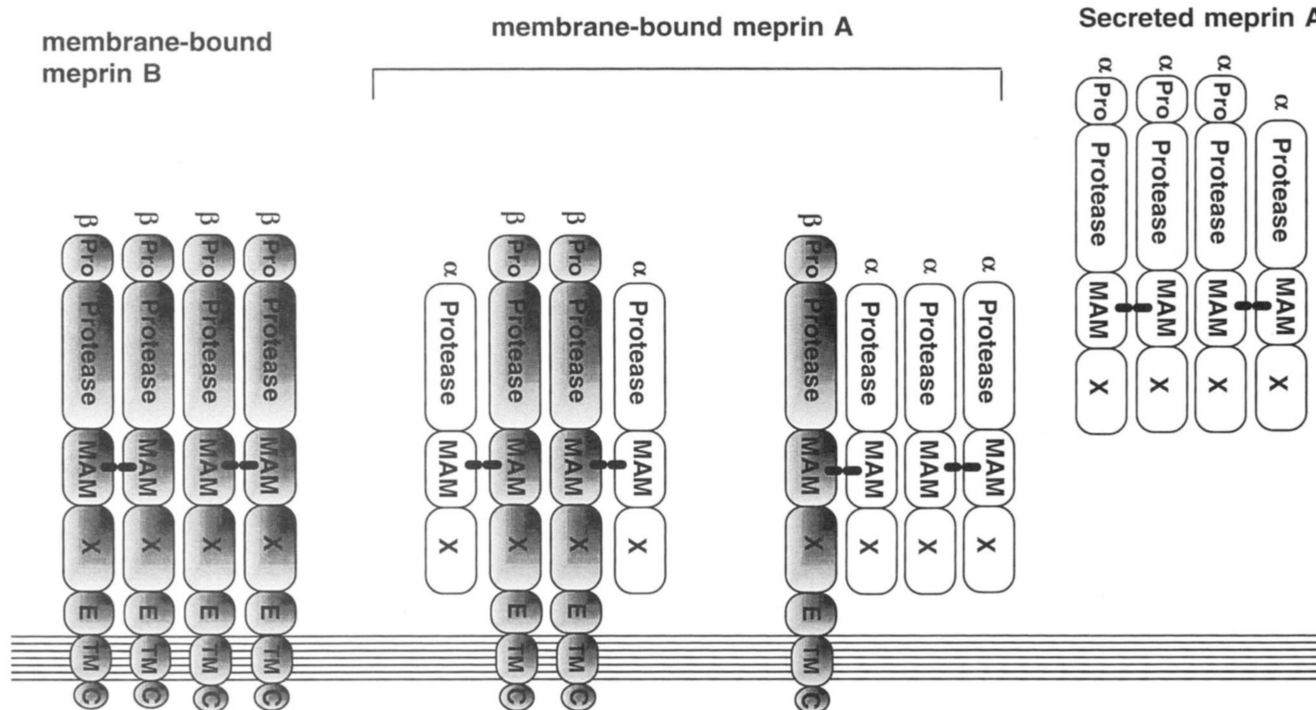

Fig. 6. Membrane-associated and secreted forms of meprin. Meprin subunits are shown schematically: α subunits with open domain structures and β subunits with shaded structures. Intersubunit disulfide bonds are indicated by horizontal lines between MAM domains.

mouse meprin A: in phosphate buffer, 360,000 ± 4,000; denatured in 4 M guanidine-HCl, 167,000 ± 800; performic acid oxidized and denatured, 86,000 ± 1,100. Thus, the denatured single (S-S bridges disrupted) subunits are 86 kDa; if disulfide bridges are not disrupted, the subunits are covalently linked as dimers; and the disulfide-linked dimers associate noncovalently to form tetramers under nondenaturing conditions.

Membrane association/secretion of meprins

Studies with brush border membranes have also indicated an oligomeric structure for meprins in vivo. SDS–gel electrophoresis and native gel analyses indicate that disulfide-linked $\alpha\alpha$ homodimers and $\alpha\beta$ heterodimers are the predominant covalently linked dimers in ICR kidney membranes in vivo (they then associate to tetrameric forms of $\alpha_2\beta_2$ and $\alpha_3\beta_1$), and that C3H/He membranes contain only $\beta\beta$ dimers and tetramers (Gorbea et al., 1991; Marchand et al., 1994). The meprin $\alpha\alpha$ homodimers can be released from membranes by 7 M urea, but $\alpha\beta$ heterodimers and $\beta\beta$ homodimers remain membrane associated, indicating that the latter are integral membrane proteins, whereas $\alpha\alpha$ homo-oligomers are noncovalently associated with the membrane (Fig. 6). Specific antibodies to COOH-teminal domains of the subunits indicated that brush border forms of α, but not β, are truncated before the EGF-like domain. More than 90% of the meprin α subunits can be released from the membrane by reducing agents such as β-mercaptoethanol. In contrast, brush border membrane meprin β cannot be removed from membranes by reducing agents. This indicates that α association with membranes is dependent on disulfide bridges; S-S bridges either directly bridge the subunit to membrane-bound β or form intrasubunit conformational units (e.g., in the MAM domain) that adhere to membrane-bound proteins. Meprin B is also tetrameric, and the data indicate that this is a dimer of S-S-bridged β homodimers.

Studies with transfected meprin cDNAs have led to a similar picture of meprin organization (Milhiet et al., 1994). When full-length meprin α transcripts were transfected into COS-1, 293, or MDCK (Madin Darby canine kidney) cells, meprin α subunits were secreted, and very little remained associated with cells (Corbeil et al., 1993; Dumermuth et al., 1993). In contrast, transfection of full-length β subunit transcripts led to plasma membrane-bound β subunits (Johnson & Hersh, 1994). When meprin α and β transcripts were cotransfected, both were associated with cells at the cell membrane. This type of data is consistent with the model for meprins (Fig. 6). In contrast to the situation in vivo, however, homodimers of α subunits were not detected in the media or membrane of recombinant systems when meprin α and β transcripts were cotransfected. In mouse kidney, the α subunit is synthesized at twice the rate of the β subunit, which may well result in an excess of α subunits (Hall et al., 1993). The differential rate of synthesis of the two subunits explains the observation of $\alpha_3\beta_1$ forms of meprin in the kidney membranes and α_4 forms in the urine of mice.

Thus, although the cDNA sequence of the meprin α subunit contains a COOH-terminal domain that would predict a membrane-spanning anchor, the mature α subunit is truncated at the end of the X domain in kidney and in mammalian recombinant systems. Transfection and mutagenesis studies of the α subunit have shown that COOH-terminal proteolytic process-

ing is determined by the I domain (Figs. 1, 7). Removal of this domain by genetic engineering results in no COOH-terminal processing of meprin α, and no transport of the subunit to the cell surface; the subunit remains in the endoplasmic reticulum (Marchand et al., 1995). This is in contrast to the meprin β subunit, which is not COOH-terminally processed in vivo, and is found as a type I integral membrane protein localized to the cell surface. However, if the I domain is introduced into the β subunit in chimeric cDNA, then β is proteolytically processed and secreted into the medium.

The fact that removal of the I domain from meprin α results in retention of the subunit in the ER, indicates that the COOH-terminal domains of α contain a retention signal. Recent experiments indicate that it is the α TM domain that contains the ER retention information (P. Marchand & J.S. Bond, unpubl. work). The meprin α subunit TM domain, but not the β TM domain, contains repeating glycine residues, like the TM domain of class II molecules of the major histocompatibility complex (MHC) (Cosson & Bonifacino, 1992). The glycines could form a surface in the transmembrane helix capable of interacting with other proteins. For meprin α, the interaction of the TM with ER membrane proteins may be responsible for retention in the ER.

It is not known whether the oligomeric nature of meprins influences proteolytic activity, but no allosteric effects have been observed. Milhiet et al. (1994) were able to show that both α and β subunits contribute to activity in hetero-oligomeric meprins. Replacing rat meprin-α Glu 157 (position 102 of Fig. 2; the catalytic residue of the HEXXH motif) by a Val, using site-directed

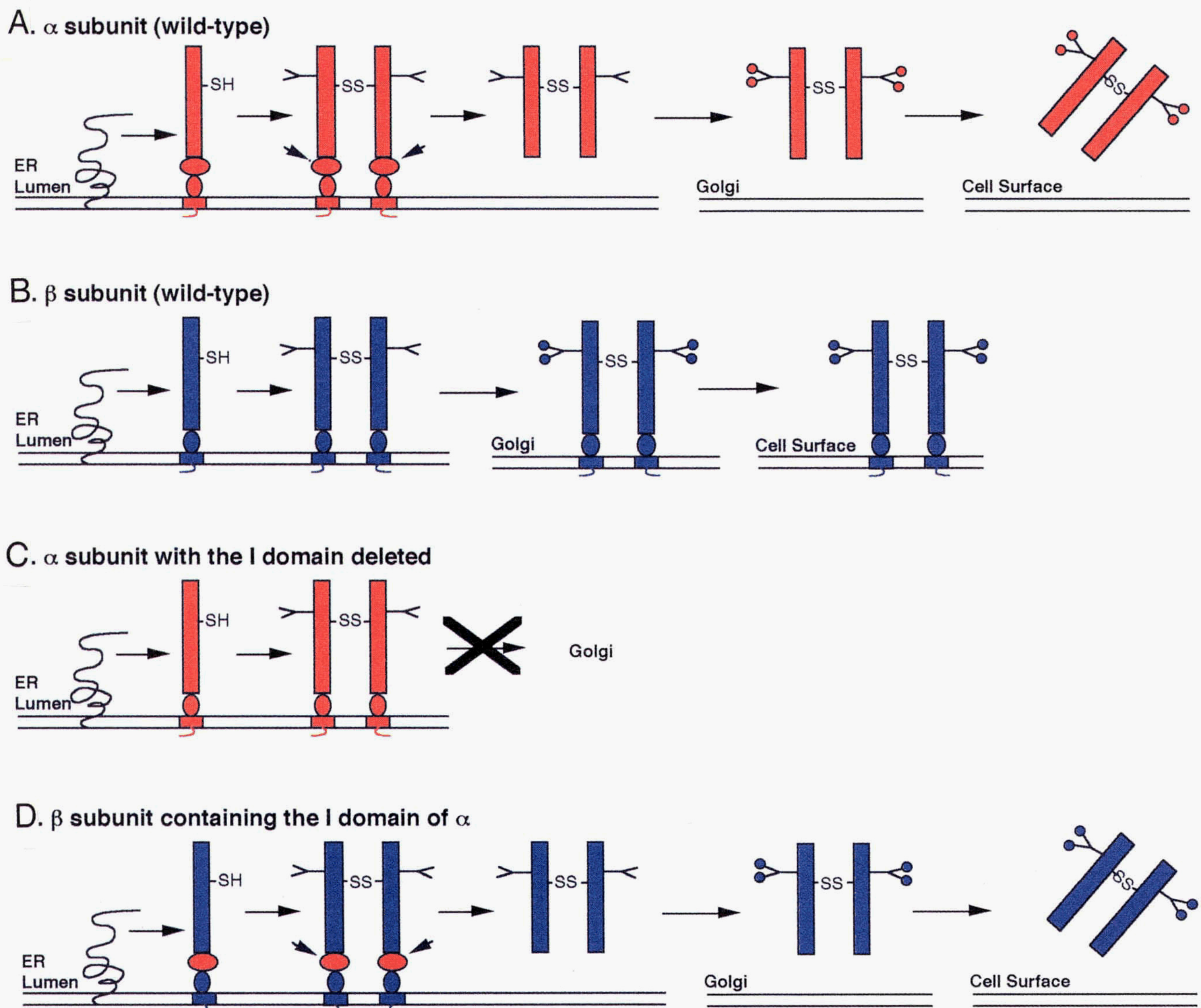

Fig. 7. Diagrammatic representation of the biosynthesis of meprin subunits. Meprin-α subunits (red) transfected into COS-1 cells were secreted (**A**), whereas meprin-β subunits were transported to the cell surface but remained membrane bound (**B**). The I domain of meprin-α is indicated by the egg-shaped structure between the bulk of the protein (rectangle containing the pro-protease-MAM-X domains) and the EGF-like-TM-cytoplasmic domains. If the I domain is deleted from meprin-α, the protein does not leave the ER (**C**). If the I domain is inserted into the β transcript, then the proteolytically truncated β subunit is secreted into the medium (**D**). ER, endoplasmic reticulum.

mutagenesis, abolished α enzymatic activity but only decreased activity of $\alpha\beta$ hetero-oligomers by 40%.

It is clear that meprins, like many other membrane proteins, have soluble counterparts (Beynon et al., 1995). Other examples include tumor necrosis factor receptors, TGF-α, ELAM-1 adhesion molecules, immunoglobulins, and angiotensin I-converting enzyme (peptidyl-dipeptidase A) (Ehlers & Riordan, 1991). The subcellular site of the COOH-terminal processing for meprins is the endoplasmic reticulum, but the proteases responsible for processing are unknown. Proteolytic processing of the meprin α subunit is not autolytic and is not due to furin-type enzymes (Marchand et al., 1995). There is evidence for an active proteolytic system in the ER because misfolded or incompletely assembled proteins are rapidly degraded, and antigen processing is proposed to occur in this subcellular compartment. Investigations of meprin processing may be useful for identifying ER proteases involved in limited proteolysis.

Enzymatic activities

Our knowledge of astacin family members as enzymes derives from the investigations of those that have been purified (astacin, HCE1 and 2, LCE, HMP1, meprins). It should be clear that the CUB-containing proteins have not yet been isolated but rather have been studied at the mRNA and DNA level. Some generalities about enzymatic properties of the family are summarized in the next several paragraphs.

pH optima

Astacin family enzymes are generally maximally active in the neutral to alkaline pH range. Optimal pH values range from 7.5 to 9.5 and may be partially determined by the substrate.

Metal requirements

The enzymes can be inhibited by metal chelators such as EDTA and 1,10-phenanthroline. Zinc has been established as the active site metal in meprin A (1.1 mol zinc/mol subunit), astacin (0.97 ± 0.03 mol zinc/mol protein), and HCE/LCE (approximately 1.3 μg/mg protein) (Butler et al., 1987; Stöcker et al., 1988; Yasumasu et al., 1989a, 1989b). Reactivation studies have indicated that Zn^{2+}, Cu^{2+}, and Co^{2+} can reactivate astacin apoenzyme; Zn^{2+}, and Cu^{2+} can reactivate the fish enzymes and meprin A apoenzyme (Wolz & Bond, 1995). Calcium ions have also been found to be associated with meprin A (2.75 mol/mol subunit) and the fish enzymes (3-5 μg/mg protein). The function of calcium is unknown, but it has been suggested to stabilize the enzymes. The ability of chelators to inhibit activity may vary for the enzymes. For example, meprins can be totally inhibited by incubation with 1-10 mM EDTA for 15 min, whereas astacin must be incubated for days with this chelator to observe inhibition; inactivation of meprins and astacin by 2 mM 1,10-phenanthroline, however, can occur after minutes (Stöcker et al., 1988).

Inhibitors

There are no known specific natural inhibitors of astacin family members. Reducing agents such as cysteine and glutathione can inhibit in vitro at concentrations of 1 mM and greater. Reducing agents probably inhibit by chelating metals or reducing disulfide bonds in the enzymes. No inhibition is observed by tissue inhibitors of metalloendopeptidases (TIMPs) nor by phosphoramidon (rhamnose phosphate-Leu-Trp; a substituted dipeptide isolated from *Streptomyces* that inhibits thermolysin-like metalloendopeptidases). Astacin family enzymes are not inhibited by inhibitors of serine, cysteine, or aspartic proteinases (e.g., PMSF, 3,4-dichloroisocoumarin, iodoacetate, E-64, leupeptin, pepstatin), or by angiotensin I-converting enzyme inhibitors such as captopril or thiorphan. The best inhibitor found for meprin A thus far is actinonin, a peptide hydroxamate with an inhibition binding constant of about 1 μM (Wolz & Bond, 1995). Actinonin also inhibits crayfish astacin, but it is 1,000-fold less effective against this enzyme compared to mouse meprin A.

Substrates

A variety of peptide and protein substrates have been used to assay astacin family members (Wolz & Bond, 1995). Some of the more commonly used protein substrates include: azocasein/casein, gelatin, and iodinated oxidized insulin B chain. The types of synthetic peptides that have been used as substrates are: dansyl-Pro-Lys-Arg-Ala-Pro-Trp-Val (astacin cleaves at the Arg-Ala bond) and nitrobradykinin (Arg-Pro-Pro-Gly-Phe(NO_2)-Ser-Pro-Phe-Arg) (astacin and meprin A cleave at the Phe(NO_2)-Ser bond). Peptide amides have also been used as substrates: succinyl-Leu-Leu-Val-Tyr-4-methylcoumaryl-7-amide (MCA) for rat meprin (Yamaguchi et al., 1991), *N*-benzoyl-L-Tyr-*p*-aminobenzoate for human meprin (Sterchi et al., 1988), Glu-Ala-Ala-Phe-MCA, succinyl-Ala-Ala-Ala-MCA, and succinyl-Ala-Ala-Val-Ala-*p*-nitroanilide (Johnson & Hersh, 1994). Cleavage of the latter substrates is at the arylamide bond, an amide bond but not a peptide bond. This is notable because, although arylamidase activity is common for serine proteases, it is not usually exhibited by metalloendopeptidases.

In general, a wide variety of peptide bonds can be cleaved in peptides and proteins by astacin family members; these enzymes are neither sequence specific nor highly specific for residues flanking the scissile bond. For example, the oxidized insulin B chain has been used to determine the peptide bond specificity of mouse meprin A and B. Meprin A cleaved 10 bonds; 3 major cleavages were Gly 20-Glu 21, Phe 24-Phe 25, Phe 25-yr 26. Meprin B cleaved 4 bonds; one, Cys(SO_3^-) 19-Gly 20, was distinct, and the others were common for both enzymes (His 5-Leu 6, Leu 6-Cys(SO_3^-) 7, Ala 14-Leu 15). This pattern of cleavage indicated a preference for bonds flanked by neutral or hydrophobic amino acid residues, but negative charges in the flanking residues can also be accommodated. Another substrate is parathyroid hormone (PTH), which has been suggested as a natural substrate for kidney meprin A in vivo (Yamaguchi et al., 1993). Many bonds are cleaved by the rat enzyme in hPTH(38-84), including: Gly 68-Glu 69, Arg 44-Asp 45, Asp 56-Asn 57, Glu 55-Asp 56, and Leu 67-Gly 68. These studies indicate a preference for cleavage at bonds containing polar or small neutral amino acids, not at bonds flanked by hydrophobic residues. Taken together, these data indicate that hydrolysis can occur at a variety of flanking amino acids.

Mapping the active sites of astacin and meprin A revealed an extended substrate binding region accommodating seven or more amino acid residues (Stöcker et al., 1990; Wolz et al., 1991). Astacin has a distinct preference for alanine in the P1′ position,

although other residues with short aliphatic side chains such as serine, threonine, and glycine are also accommodated in this position. The crystal structure of astacin revealed that this is due to the presence of Pro 176 (position 188 on Fig. 2) at the S1′ site, which the P1′ residue must align with (Stöcker et al., 1993). Meprin A can accommodate many other residues in P1′, including bulky hydrophobic residues, due to the substitution of Gly 176 (position 190) and deletion of Tyr 177 (position 189). These changes result in a more open configuration of meprin compared to astacin. The activities of both enzymes are greatly affected by residues three and four amino acids away from the cleavage site, indicating that the extended amino acid chain of the substrate interacts with the enzyme and influences binding and/or turnover of the enzyme/substrate complex. Proline residues in positions two and three amino acids away from the cleavage site greatly affect activity, again demonstrating multiple interactions of the enzyme subsites with substrate residues.

One area for future investigation that should be fruitful is molecular modeling of the interactions of substrates and inhibitors with the active sites of astacin family members based on the astacin structure. One question of interest to us, for example, is why meprin A and B are so different in substrate specificity. Meprin A hydrolyzes α-melanocyte-stimulating hormone ($k_{cat}/K_m = 1 \times 10^6$ M^{-1} s^{-1}) and many peptides and arylamides, whereas meprin B is a very poor peptidase toward substrates of fewer than 10 amino acids (Wolz & Bond, 1995). Examination of the amino acid sequences of the α and β subunits of meprins in the light of the three-dimensional structure of astacin (Stöcker et al., 1993) indicates that there are significant differences in the sequences of the two meprin subunits that could account for substrate specificity differences. For example, the Phe of position 165 of Figure 2 has been identified as the putative S2′ binding site of astacin (residue 154), and the residues flanking this site are quite different in the α and β subunits; positions 162–167 are PFSFNK for α, and KTAFQ- for β. The S1′ binding site for astacin (residue 175) corresponds to the area containing positions 187–189 on Figure 2; positions 186–187 are N(T/A/S) for α and ED for β subunits. There are several other consistent differences between the subunits that could explain activity differences. We anticipate that modeling predictions will eventually be tested experimentally by the production of correctly folded recombinant proteases altered by site-directed mutagenesis, and that this type of study will aid in identifying physiological substrates and inhibitors for the enzymes.

Effects of noncatalytic domains on enzymatic activity

Activation/prosequences

Astacin family enzymes are synthesized as inactive proenzymes, and removal of the prosequences constitutes a major mechanism for activation. The crystal structure of astacin indicates that there is a salt bridge between the NH_2-terminal amino acid of the mature protease domain (Ala 1) and Glu 103 (position 112 of Fig. 2), which is next to the third zinc ligand (astacin residue His 102, position 111) near the active site. This bridge may be critical for enzymatic activity and, because the prosequence would prevent the formation of the salt bridge, it would inhibit activity. Indeed, there is substantial experimental evidence from the meprins that supports this proposal. The β subunits of the mature form of meprin B from C3H/He mouse kidney are present at the cell surface predominantly as proforms, and the enzyme is inactive against proteins in this form (Gorbea et al., 1993). This allows for a high concentration of a proteolytic enzyme that can be activated at the membrane. The mature α subunit, but not the mature β subunit, of meprin A from ICR or C57BL/6 mouse kidney is largely activated (i.e., the prosequence is removed). Thus, there is differential proteolytic processing of the NH_2-terminus, as well as the COOH-terminus, of the α and β subunits in mouse kidney. The enzymes responsible for NH_2-terminal processing, and therefore activation, and the subcellular compartment in which activation occurs have not yet been identified. Interestingly, secreted (urinary) forms of meprin A, and transfected forms of mouse, rat, or human α or β subunits expressed in COS-1, 293, or MDCK cells in culture, are all expressed predominantly as inactive proenzymes, indicating they are not exposed to or are not vulnerable to the activation proteases. All the prosequences contain basic amino acids at the COOH-terminus, although different motifs (such as single Lys or Arg residues, R/KXR, RRXR) are found in different members of the family. This implies that trypsin-like, furin-like, or other processing enzymes that cleave at basic residues are involved in activation. Removal of the prosequence is likely a major form of posttranslational regulation of astacin family proteinase activity.

Interaction domains/CUB domains

It has been proposed that CUB domains promote multimeric complex formation and aid in substrate recognition. For complement proteases, there is evidence that C1r/s mediates calcium-dependent tetrameric complex formation between two C1r–C1s dimers and the association of this complex with C1q to form the mature C1 molecule (e.g., Sim et al., 1993). Childs and O'Connor (1994), studying *Drosophila tolloid*, found that loss-of-function mutations map to CUB repeats. By analogy with complement proteins, the CUB domains in astacin family enzymes could be interaction domains that promote multimeric complexes. The differences in functions of *tolloid* and *tolloid-related-1* in *Drosophila* have been attributed to the differences in amino acid sequences of the CUB and EGF-like domains that could result in specific interactions with different latent growth factors (R. Padgett, pers. comm.).

Genes – Localization, duplication, structure

The gene structures of the fish enzymes have been analyzed in *O. latipes* (Yasumasu et al., 1994). There are eight copies of the HCE genes: five for HCE1 and three for HCE2. Six (three of each) of the genes form a cluster within approximately 25 kb of the genomic DNA; the two others are located separately. There are no introns in the HCE genes. In contrast, there is only one copy of the LCE gene. It is within 3.6 kb and contains eight exons and seven introns. For LCE, the TATA box consensus sequence is 28 bp upstream from the transcription start site. The 5′-flanking regions of all the HCE genes are 80–95% similar for 200–400 bp. There is no similarity between the HCE and LCE 5′ regions within 1.5 kb. These enzymes are both found in secretory vesicles in fish and are both necessary to digest the egg envelope. Larger amounts of HCE relative to LCE protein are found in zymogen granules of the hatching gland cells, and it has been suggested that this reflects the larger gene copy num-

ber for HCE. The strong similarity of HCE and LCE proteins indicates a common ancestral gene; if so, it is interesting that their gene structures have evolved to differ so markedly.

The BMP-1 gene has been localized to the proximal half of mouse chromosome 14 and to human chromosome 8p21 (Ceci et al., 1990; Yoshiura et al., 1993). There is keen interest in determining whether this is a candidate gene for diseases—especially those that are caused by dysfunction of bone and cartilage.

The structural genes for meprin subunits have been localized in the mouse and human genomes, and although it is likely the α and β genes derived from a common ancestral gene, they have evolved to reside on different chromosomes (Gorbea et al., 1993; Jiang et al., 1993; Bond et al., 1995). The structural gene for the meprin α subunit is on mouse chromosome 17 and human chromosome 6p near the centromere; this gene is linked to the major histocompatibility complex (MHC) in both genomes. The structural gene for meprin β is on chromosome 18 of both the mouse and human (18q12.2-q12.3) genomes. The structural genes and linkages to neighboring genes are conserved for mouse and human, and the localization of the genes in the two species sets the stage for the identification of the elements that determine the differences in tissue-specific expression, as well as association with disease genes. Regulation of expression is markedly different in tissues, species, and strains. For example, although both α and β are expressed in human intestine, only β is expressed in mouse intestine; both are expressed at high levels in kidney of many random-bred and inbred strains of mice, but only β is expressed in kidney of some inbred mouse strains (such as C3H/He and CBA mice); and both are expressed at low levels in human kidney (Gorbea et al., 1993; Yamaguchi et al., 1994). There must be some coordinate regulation of the two subunits because they are found as complexes, but at least β can be expressed independently.

Prospects for the future

Astacin family members represent proteases that are widely distributed in nature, highly regulated at the transcriptional and posttranslational levels, and active in specialized functions in developing and mature systems. Exploration of the structure, function, and regulation of these enzymes is at an early stage, and we may anticipate future revelations as we have only uncovered the "tip of the iceberg" for this interesting family.

The three-dimensional structure of only one member of the family is known, and we expect that the structures of other members, or at least the protease domains of members, will be solved in the near future. It will be particularly interesting to compare the structure of flavastacin to astacin, to see how the absence of sulfhydryl bridges affects the tertiary structure. It will be possible to produce substantial amounts of the developmental proteins, as well as separate domains of the proteins, by recombinant techniques to study structure–function relationships.

Investigations of these enzymes will no doubt contribute to our fundamental knowledge of the biosynthesis, trafficking, proteolytic processing, and regulation of cell surface and secreted proteins and of proteolytic enzymes.

Finally, there is a great deal to learn about the natural substrates of astacin family members and about the interactions that the proteins are involved in at the cell surface and in the extracellular milieu. It is likely that these proteases will have important roles in growth, development, and remodeling, and their relationship to healthy and diseased states is yet to be investigated.

Acknowledgments

This work was supported by NIH grant DK 19691 (J.S.B.) and by grants from the BBSRC and the Wellcome Trust (R.J.B.). We thank Dr. Petra Marchand for providing Figure 7 and for comments on the manuscript.

References

Barnes K, Ingram J, Kenny AJ. 1989. Proteins of the kidney microvillar membrane. Structural and immunochemical properties of rat endopeptidase-2 and its immunohistochemical localization in tissues of rat and mouse. *Biochem J 264*:335–346.

Beckmann G, Bork P. 1993. An adhesive domain detected in functionally diverse receptors. *Trends Biochem Sci 18*:40–41.

Beynon RJ, Oliver S, Stateva L, Evershed RP, Robertson D. 1995. Soluble forms of kidney meprin-α. In: Stöcker W, Zwilling R, eds. *The astacins—Structure and function of a new protein family*. Forthcoming.

Bode W, Gomis-Rüth RF, Huber R, Zwilling R, Stöcker W. 1992. Structure of astacin and implications for activation of astacins and zinc-ligation of collagenases. *Nature 358*:164–167.

Bode W, Gomis-Rüth FX, Stöcker W. 1993. Astacins, serralysins, snake venom and matrix metalloproteinases exhibit identical zinc-binding environments (HEXXHXXGXXH and Met-turn) and topologies and should be grouped into a common family, the "metzincins." *FEBS Lett 331*:134–140.

Bond JS, Beynon RJ. 1986. Meprin: A membrane-bound metallo-endopeptidase. *Curr Topics Cell Regul 28*:263–290.

Bond JS, Jiang W. 1995. Astacin family of metalloendopeptidases. In: NM Hooper, ed. *Zinc metalloproteases in health and disease*. London: Taylor & Francis.

Bond JS, Rojas K, Overhauser J, Zoghbi HY, Jiang W. 1995. The structural genes *MEP1A* and *MEP1B*, for the α and β subunits of the metalloendopeptidase meprin map to human chromosome 6p and 18q, respectively. *Genomics 25*:300–303.

Bork P, Beckmann G. 1993. The CUB domain: A widespread module in developmentally regulated proteins. *J Mol Biol 231*:539–545.

Butler PE, McKay MJ, Bond JS. 1987. Characterization of meprin, a membrane-bound metalloendopeptidase from mouse kidney. *Biochem J 241*:229–235.

Campbell LD, Bork P. 1993. Epidermal growth factor-like modules. *Curr Opin Struct Biol 3*:385–392.

Ceci JD, Kingsley DM, Silan CM, Copeland NG, Jenkins NA. 1990. An interspecific backcross linkage map of the proximal half of mouse chromosome 14. *Genomics 6*:673–678.

Childs SR, O'Connor MB. 1994. Two domains of the *tolloid* protein contribute to its unusual genetic interaction with *decapentaplegic*. *Dev Biol 162*:209–220.

Choudry Y, Kenny AJ. 1991. Hydrolysis of transforming growth factor-α by cell-surface peptidases in vitro. *Biochem J 280*:57–60.

Corbeil D, Milhiet PE, Simon V, Ingram J, Kenny AJ, Boileau G, Crine P. 1993. Rat endopeptidase-24.18 α subunit is secreted into the culture medium as a zymogen when expressed by COS-1 cells. *FEBS Lett 335*: 361–366.

Cosson P, Bonifacino JS. 1992. Role of transmembrane domain interactions in the assembly of class II MHC molecules. *Science 258*:659–661.

Craig SS, Mader C, Bond JS. 1991. Immunohistochemical localization of the metalloproteinase meprin in salivary glands of male and female mice. *J Histochem Cytochem 39*:123–129.

Craig SS, Reckelhoff JF, Bond JS. 1987. Distribution of meprin in kidneys from mice with high- and low-meprin activity. *Am J Physiol 253*:535–540.

Dumermuth E, Eldering JA, Grunberg J, Jiang W, Sterchi EE. 1993. Cloning of the PABA peptide hydrolase α subunit (PPHα) from human small intestine and its expression in COS-1 cells. *FEBS Lett 335*:367–375.

Dumermuth E, Sterchi EE, Jiang WP, Wolz RL, Bond JS, Flannery AV, Beynon RJ. 1991. The astacin family of metalloendopeptidases. *J Biol Chem 266*:21381–21385.

Ehlers MRW, Riordan JF. 1991. Membrane proteins with soluble counterparts: Role of proteolysis in the release of transmembrane proteins. *Biochemistry 30*:10065–10074.

Elaroussi MA, DeLuca HF. 1994. A new member to the astacin family of metalloendopeptidases: A novel 1,25-dihydroxyvitamin D-3-stimulated

mRNA from chorioallantoic membrane of quail. *Biochim Biophys Acta 1217*:1-8.
Evans SV. 1993. SETOR: Hardware lighted three-dimensional solid model representations of macromolecules. *J Mol Graphics 11*:134-138.
Ferguson EL, Anderson KV. 1992. *Decapentaplegic* acts as a morphogen to organize dorsal-ventral pattern in the *Drosophila* embryo. *Cell 71*: 451-461.
Fukagawa M, Suzuki N, Hogan B, Jones CM. 1994. Embryonic expression of mouse bone morphogenetic protein-1 (BMP-1), which is related to the *Drosophila* dorsoventral gene *tolloid* and encodes a putative astacin metalloendopeptidase. *Dev Biol 163*:175-183.
Furie B, Furie BC. 1988. The molecular basis of blood coagulation. *Cell 53*:505-518.
Gomis-Rüth FX, Grams F, Yiallouros I, Nar H, Küsthardt U, Zwilling R, Bode W, Stöcker W. 1994. Crystal structures, spectroscopic features, and catalytic properties of cobalt (II), copper (II), nickel (II), and mercury (II) derivatives of the zinc endopeptidase astacin. *J Biol Chem 269*: 17111-17117.
Gorbea CM, Flannery AV, Bond JS. 1991. Homo- and heterotetrameric forms of the membrane-bound metalloendopeptidases meprin A and B. *Arch Biochem Biophys 290*:549-553.
Gorbea CM, Marchand P, Jiang W, Copeland NG, Gilbert DJ, Jenkins NA, Bond JS. 1993. Cloning, expression, and chromosomal localization of the mouse meprin β subunit. *J Biol Chem 268*:21035-21043.
Hall JL, Sterchi EE, Bond JS. 1993. Biosynthesis and degradation of meprins, kidney brush border proteinases. *Arch Biochem Biophys 307*: 73-77.
Hwang S, Partin JS, Lennarz WJ. 1994. Characterization of a homolog of human bone morphogenetic protein 1 in the embryo of the sea urchin, *Strongylocentrotus purpuratus*. *Development 120*:559-568.
Jiang W, Bond JS. 1992. Families of metalloendopeptidases and their relationships. *FEBS Lett 312*:110-114.
Jiang W, Gorbea CM, Flannery AV, Beynon RJ, Grant GA, Bond JS. 1992. The α subunit of meprin A. Molecular cloning and sequencing, differential expression in inbred mouse strains, and evidence for divergent evolution of the α and β subunits. *J Biol Chem 267*:9185-9193.
Jiang W, Sadler PM, Jenkins NA, Gilbert DJ, Copeland NG, Bond JS. 1993. Tissue-specific expression and chromosomal localization of the α subunit of mouse meprin A. *J Biol Chem 268*:10380-10385.
Johnson GD, Hersh LB. 1992. Cloning a rat meprin cDNA reveals the enzyme is a heterodimer. *J Biol Chem 267*:13505-13512. [Erratum in *J Biol Chem 268*:17647.]
Johnson GD, Hersh LB. 1994. Expression of meprin subunit precursors. Membrane anchoring through the β subunit and mechanism of zymogen activation. *J Biol Chem 269*:7682-7688.
Kaushal GP, Walker PD, Shah SV. 1994. An old enzyme with a new function: Purification and characterization of a distinct matrix-degrading metalloproteinase in rat kidney cortex and its identification as meprin. *J Cell Biol 126*:1319-1327.
Kennelly PJ, Krebs EG. 1991. Consensus sequences of substrate specificity determinants for protein kinases and protein phosphatases. *J Biol Chem 266*:15555-15558.
Kitamoto Y, Yuan X, Wu Q, McCourt DW, Sadler JE. 1994. Enterokinase, the initiator of intestinal digestion, is a mosaic protease composed of a distinctive assortment of domains. *Proc Natl Acad Sci USA 91*:7588-7592.
Lee KS, Yasumasu S, Nomura K, Iuchi I. 1994. HCE, a constituent of the hatching enzymes of *Oryzias latipes* embryos, releases unique proline-rich polypeptides from its natural substrate, the hardened chorion. *FEBS Lett 339*:281-284.
Lepage T, Ghiglione C, Gasche C. 1992a. Spatial and temporal expression pattern during sea urchin embryogenesis of a gene coding for a protease homologous to the human protein BMP-1 and to the product of the *Drosophila* dorsal-ventral patterning gene *tolloid*. *Development 114*: 147-164.
Lepage T, Sardet C, Gache C. 1992b. Spatial expression of the hatching enzyme gene in the sea urchin embryo. *Dev Biol 150*:23-32.
Maéno M, Xue Y, Wood TI, Ong RC, Kung H. 1993. Cloning and expression of cDNA encoding *Xenopus laevis* bone morphogenetic protein-1 during early embryonic development. *Gene 134*:257-261.
Marchand P, Tang J, Bond JS. 1994. Membrane association and oligomeric organization of the α and β subunits of mouse meprin A. *J Biol Chem 269*:15388-15393.
Marchand P, Tang J, Johnson GD, Bond JS. 1995. COOH-terminal proteolytic processing of secreted and membrane forms of the α subunit of the metalloprotease meprin A: Requirement of the I domain for processing the endoplasmic reticulum. *J Biol Chem 270*:5449-5456.
Matrisian LM. 1992. The matrix-degrading metalloproteinases. *BioEssays 14*:455-463.
Milhiet PE, Corbeil D, Simon V, Kenny AJ, Crine P, Boileau G. 1994. Expression of rat endopeptidase-24.18 in COS-1 cells: Membrane topology and activity. *Biochem J 300*:37-43.
Neurath H. 1989. The diversity of proteolytic enzymes. In: Beynon RJ, Bond JS, eds. *Proteolytic enzymes: A practical approach*. Oxford, UK: IRL Press.
Nguyen T, Jamal J, Shimell MJ, Arora K, O'Connor MB. 1994. Characterization of *tolloid-related-1*: A BMP-1-like product that is required during larval and pupal stages of *Drosophila* development. *Dev Biol 166*:569-586.
Padgett RW, Wozney JM, Gelbart WM. 1993. Human BMP sequences can confer normal dorsal-ventral patterning in the *Drosophila* embryo. *Proc Natl Acad Sci USA 90*:2905-2909.
Rawlings ND, Barrett AJ. 1993. Evolutionary families of peptidases. *Biochem J 290*:205-218.
Reynolds SD, Angerer LM, Palis J, Nasir A, Angerer RC. 1992. Early mRNAs, spatially restricted along the animal-vegetal axis of sea urchin embryos, include one encoding a protein related to *tolloid* and BMP-1. *Development 114*:769-786.
Sato SM, Sargent TD. 1990. Molecular approach to dorsoanterior development in *Xenopus laevis*. *Dev Biol 137*:135-141.
Shimell MJ, Ferguson EL, Childs ST, O'Connor MB. 1991. The *Drosophila* dorsal-ventral patterning gene *tolloid* is related to human bone morphogenetic protein-1. *Cell 67*:469-481.
Sim RB, Day AJ, Moffatt BE, Fontaine M. 1993. Complement factor I and cofactors in control of complement system convertase enzymes. *Methods Enzymol 223*:13-35.
Spencer-Dene B, Thorogood P, Nair S, Kenny AJ, Harris M, Henderson B. 1994. Distribution of, and a putative role for, the cell-surface neutral metalloendopeptidases during mammalian craniofacial development. *Development 120*:3213-3226.
Steiner DF, Smeekens SP, Ohagi S, Chan SJ. 1992. The new enzymology of precursor processing endoproteases. *J Biol Chem 267*:23435-23438.
Stenzel P, Angerer LM, Smith BJ, Angerer RC, Vale WW. 1994. The *univin* gene encodes a member of the transforming growth factor-beta superfamily with restricted expression in the sea urchin embryo. *Dev Biol 166*:149-158.
Stephenson SL, Kenny AJ. 1988. The metabolism of neuropeptides. Hydrolysis of peptides by the phosphoramidon-insensitive rat kidney enzyme "endopeptidase-2" and by rat microvillar membranes. *Biochem J 255*:45-51.
Sterchi EE, Naim HY, Lentze MJ, Hauri HP, Fransen JA. 1988. *N*-benzoyl-L-tyrosyl-*p*-aminobenzoic acid hydrolase: A metalloendopeptidase of the human intestinal microvillus membrane which degrades biologically active peptides. *Arch Biochem Biophys 265*:105-118.
Stöcker W, Gomis-Rüth FX, Bode W, Zwilling R. 1993. Implications of the three-dimensional structure of astacin for the structure and function of the astacin family of zinc-endopeptidases. *Eur J Biochem 214*:215-231.
Stöcker W, Grams F, Baumann U, Reinemer P, Gomis-Rüth FX, McKay DB, Bode W. 1995. The metzincins – Topological and sequential relations between the astacins, adamalysins, serralysins, and matrixins (collagenases) define a superfamily of zinc-peptidases. *Protein Sci 4*:823-840.
Stöcker W, Ng M, Auld DS. 1990. Fluorescent oligopeptide substrates for kinetic characterization of the specificity of *Astacus* protease. *Biochemistry 29*:10418-10425.
Stöcker W, Wolz RL, Zwilling R. 1988. *Astacus* protease, a zinc metalloenzyme. *Biochemistry 27*:5026-5032.
Strongin AY, Collier I, Bannikov G, Marmer BL, Grant GA, Goldberg GI. 1994. Mechanism of cell surface activation of 72-kDa type IV collagenase. *J Biol Chem 270*:5331-5338.
Takahara K, Lyons GE, Greenspan DS. 1994. Bone morphogenetic protein-1 and a mammalian tolloid homologue (mTld) are encoded by alternatively spliced transcripts which are differentially expressed in some tissues. *J Biol Chem 269*:32572-32578.
Tarentino AL, Quinones G, Grimwood BG, Hauer CR, Plummer TH Jr. 1995. Molecular cloning and sequence analysis of flavastacin: An *O*-glycosylated prokaryotic zinc metalloendopeptidase. *Arch Bochem Biophys 319*:281-285.
Titani K, Torff HJ, Hormel S, Kumar S, Walsh KA, Rodl J, Neurath H, Zwilling R. 1987. Amino acid sequence of a unique protease from the crayfish *Astacus fluviatilis*. *Biochemistry 26*:222-226.
Vassalli JD, Sappino AP, Belin D. 1991. The plasminogen activator/plasmin system. *J Clin Invest 88*:1067-1072.
Vogt G, Stöcker W, Storch V, Zwilling R. 1989. Biosynthesis of *Astacus* protease a digestive enzyme from crayfish. *Histochemistry 91*:373-381.
Webb EC, ed. 1992. *Enzyme nomenclature*. San Diego, California: Academic Press, Inc. p 411.
Wolz RL, Bond JS. 1995. Meprins A and B. *Methods Enzymol 248*: 325-345.

Wolz RL, Harris RB, Bond JS. 1991. Mapping the active site of meprin-A with peptide substrates and inhibitors. *Biochemistry 30*:8488–8493.

Wozney JM, Rosen V, Celeste AJ, Mitsock LM, Whitters MJ, Kriz RW, Hewick RM, Wang EA. 1988. Novel regulators of bone formation: Molecular clones and activities. *Science 242*:1528–1531.

Yamaguchi T, Fukase M, Sugimoto T, Kido H, Chihara K. 1994. Purification of meprin from human kidney and its role in parathyroid hormone degradation. *Biol Chem Hoppe-Seyler 375*:821–824.

Yamaguchi T, Kido H, Fukase M, Fujita T, Katunuma N. 1991. A membrane-bound metalloendopeptidase from rat kidney hydrolyzing parathyroid hormone. Purification and characterization. *Eur J Biochem 200*:563–571.

Yamaguchi T, Kido H, Kitazawa R, Kitazawa S, Fukase M, Katunuma N. 1993. A membrane-bound metallo-endopeptidase from rat kidney: Its immunological characterization. *Biochemistry 113*:299–303.

Yasumasu S, Iuchi I, Yamagami K. 1989a. Purification and partial characterization of high choriolytic enzyme (HCE), a component of the hatching enzyme of the teleost, *Oryzias latipes*. *J Biochem 105*:204–211.

Yasumasu S, Iuchi I, Yamagami K. 1989b. Isolation and some properties of low choriolytic enzyme (LCE), a component of the hatching enzyme of the teleost, *Oryzias latipes*. *J Biochem 105*:212–218.

Yasumasu S, Iuchi I, Yamagami K. 1994. cDNAs and the genes of HCE and LCE, two constituents of the medaka hatching enzyme. *Dev Growth & Differ 36*:241–250.

Yasumasu S, Katow S, Hamazaki TS, Iuchi I, Yamagami K. 1992. Two constituent proteases of a teleostean hatching enzyme: Concurrent syntheses and packaging in the same secretory granules in discrete arrangement. *Dev Biol 149*:349–356.

Yasumasu S, Katow S, Umino Y, Iuchi I, Yamagami K. 1989c. A unique proteolytic action of HCE, a constituent protease of a fish hatching enzyme: Tight binding to its natural substrate, egg envelope. *Biochem Biophys Res Commun 162*:58–63.

Yoshiura K, Tamura T, Hong HS, Ohta T, Soejima H, Kishino T, Jinno Y, Niikawa N. 1993. Mapping of the bone morphogenetic protein 1 gene (BMP1) to 8p21: Removal of BMP1 from candidacy for the bone disorder in Langer-Giedion syndrome. *Cytogenet Cell Genet 64*:208–209.