Proteomic Analyses Reveal an Acidic Prime Side Specificity for the Astacin Metalloprotease Family Reflected by Physiological Substrates*[S]

Christoph Becker-Pauly‡‡‡, Olivier Barré§, Oliver Schilling§¶, Ulrich auf dem Keller§‖, Anke Ohler‡, Claudia Broder‡, André Schütte, Reinhild Kappelhoff§, Walter Stöcker‡, and Christopher M. Overall§**

Astacins are secreted and membrane-bound metalloproteases with clear associations to many important pathological and physiological processes. Yet with only a few substrates described their biological roles are enigmatic. Moreover, the lack of knowledge of astacin cleavage site specificities hampers assay and drug development. Using PICS (proteomic identification of protease cleavage site specificity) and TAILS (terminal amine isotopic labeling of substrates) degradomics approaches >3000 cleavage sites were proteomically identified for five different astacins. Such broad coverage enables family-wide determination of specificities N- and C-terminal to the scissile peptide bond. Remarkably, meprin α, meprin β, and LAST_MAM proteases exhibit a strong preference for aspartate in the peptide (P)1′ position because of a conserved positively charged residue in the active cleft subsite (S)1′. This unparalleled specificity has not been found for other families of extracellular proteases. Interestingly, cleavage specificity is also strongly influenced by proline in P2′ or P3′ leading to a rare example of subsite cooperativity. This specificity characterizes the astacins as unique contributors to extracellular proteolysis that is corroborated by known cleavage sites in procollagen I+III, VEGF (vascular endothelial growth factor)-A, IL (interleukin)-1β, and pro-kallikrein 7. Indeed, cleavage sites in VEGF-A and pro-kallikrein 7 identified by terminal amine isotopic labeling of substrates matched those reported by Edman degradation. Moreover, the novel substrate FGF-19 was validated biochemically and shown to exhibit altered biological activity after meprin processing. ***Molecular & Cellular Proteomics 10: 10.1074/mcp.M111.009233, 1–19, 2011.***

Proteases comprise a large and diverse group of enzymes encoded by over 2% of all human genes (1). Being responsible for most post-translational modifications, these molecules are fundamental for almost all biological processes in health and disease. Indeed, with regard to Alzheimer's disease, cardio-vascular disorders, AIDS, and cancer, proteases represent up to 10% of all drug targets (2). Most therapeutics target the active site of these enzymes, thereby inhibiting the proteolytic activity. Thus, the biggest challenge is to develop highly specific protease inhibitors distinguishing between related enzymes, which may have different physiological functions. Therefore, it is crucial to gain detailed information about the cleavage site specificity to understand the molecular interactions between these enzymes and substrates and to design new assays and to assist in drug development (3, 4).

The large astacin family of extracellular metalloproteases have members in bacteria and animals, but not as yet in plants and fungi (5). In humans six astacins are known, namely meprin α and β (6, 7), bone morphogenetic protein-1 (BMP-1) with its major splice variant mammalian tolloid, mammalian tolloid-like enzymes (8), and ovastacin (9). BMP-1 and the tolloids are essential for dorsal and ventral patterning in embryogenesis and for extracellular matrix assembly by cleaving off the C-terminal collagen propeptides and processing of other matrix precursors (10). Meprin α and β are expressed in a variety of epithelial cells including renal and intestinal brush borders, the epidermis, airway epithelium immune cells, and some cancer cells (11–14). Only recently has it become obvious that both enzymes are also present in fibrotic dermis and are capable of processing procollagen III at exactly the same cleavage site as BMP-1 (15).

In vitro, meprin α and β cleave several extracellular matrix proteins and peptide hormones (13, 16, 17). Cytokines and growth factors, namely TGF-α (transforming growth factor-α),

From the ‡Cell and Matrix Biology, Johannes Gutenberg-University, Johannes-von-Müller-Weg 6, D-55128 Mainz, Germany; §The UBC Centre for Blood Research, Departments of Oral Biological and Medical Sciences, and Biochemistry and Molecular Biology, 4.401 Life Sciences Institute, 2350 Health Sciences Mall, University of British Columbia, Vancouver, British Columbia V6T 1Z3, Canada; ¶Present Address: Institute of Molecular Medicine and Cell Research, University of Freiburg, Stefan-Meier-Str. 17, D-79104 Freiburg, Germany; ‖Present Address: Institute of Cell Biology, Swiss Federal Institute of Technology Zurich, ETH Hoenggerberg, HPM D24, CH-8093 Zurich, Switzerland; **Mucin Biology Group, Department of Medical Biochemistry, University of Gothenburg, Medicinaregatan 9A, 413 90 Gothenburg, Sweden

⌘ *Author's Choice*—Final version full access.

Received March 11, 2011, and in revised form, May 18, 2011

Published, MCP Papers in Press, June 21, 2011, DOI 10.1074/mcp.M111.009233

VEGF-A (vascular endothelial growth factor-A) (18), interleukin-1β, and interleukin-18 are also activated by meprins *in vivo* (11, 13, 19, 20). An early analysis of peptide cleavage specificity indicated that while meprin α was nonspecific, meprin β has a preference for acidic residues in P1′ (21). Crayfish astacin from *Astacus astacus*, the eponym of this protease family, is a digestive enzyme, stored in the stomach at high concentration and capable of cleaving fibrillar collagen (22). LAST and LAST_MAM, two proteases from the horseshoe crab *Limulus polyphemus*, are in different tissues, but their physiological functions remain ambiguous due in part to a lack of understanding of their substrates (23). These and other nonvertebrate astacins underlie the evolutionary conservation of astacin proteases.

Notwithstanding these examples, most astacin proteases in the database are annotated solely on the basis of sequence similarity of their protease domains and have yet to be functionally characterized (1, 24). This is often simply not feasible because of missing genetic and proteomic approaches for the host organisms, *e.g.* in the case of astacins from the horseshoe crab (23) that have not been sequenced. This deficit means that new proteomic approaches in the field of degradomics cannot yet be applied for profiling their activities. Nonetheless, the proteomic identification of protease cleavage site specificity (PICS)[1] and terminal amine isotopic labeling of substrates (TAILS) approaches based on proteome-derived, database-searchable peptide libraries for identifying protease cleavage sites and native complex proteomes, respectively (25–27) are used here for the first time to determine the molecular specificities of a whole protease family. By relating these specificities to three-dimensional structural and modeling analyses the astacins are found to be specific in their choice of substrates, being strongly determined by the amino acid residues around the cleavage site with a unique aspartate specificity in P1′.

EXPERIMENTAL PROCEDURES

Expression, Purification and Activation of Astacin Proteases—All enzymes were produced as described previously. Recombinant meprin α (12), meprin β (28), LAST and LAST_MAM (23) were expressed in insect cells. Astacin was purified from the digestive tract of the crayfish *Astacus astacus* (29).

PICS—PICS library preparation, PICS specificity screen, isolation of prime side cleavage fragments, and LC-MS/MS analysis were performed as previously described (25). Meprin α, meprin β, LAST_MAM, LAST, and astacin were buffered in 50 mM HEPES, pH 7.5. Enzyme/library ratio was 1:100 (wt/wt) for all enzymes. Assays were performed for 12 h at 37 °C and stopped by heat de-activation (70 °C, 30 min).

PICS analysis by liquid chromatography (LC) tandem mass spectrometry (MS/MS) was performed on a Packings capillary LC system (Dionex) coupled to a quadrupole time-of-flight mass spectrometer (QSTAR XL; Applied Biosystems, Foster City, CA) at the University of British Columbia Centre for Blood Research Mass Spectrometry Suite. Samples were resuspended in 5% (v/v) acetonitrile, 3% formic acid and loaded onto a column packed with Magic C18 resin (Michrom Bioresources, Auburn, CA). Peptides were eluted using a 7–40% gradient of organic phase over 95 min. Buffer A was 2% acetonitrile with 0.1% formic acid and buffer B was 98% acetonitrile, 0.1% formic acid. Mass spectrometry data were acquired automatically using Analyst QS software (Applied Biosystems) for information-dependent acquisition based on a 1 s mass spectrometry survey scan followed by up to two (QSTAR XL) MS/MS scans of 3 s each.

QSTAR data were converted to the mzXML format with mzwiff (version 4.3.1). MS and MS/MS data were centroided. Data were analyzed by X! Tandem (version 2010.01.01) (30) in conjunction with PeptideProphet (version 4.3.1) (31) at a 95% confidence level for each peptide and a decoy search strategy using a shuffled version of the human International Protein Index database (version 3.42) (32). Including the shuffled decoys, the database comprised 144,680 entries. Search parameters excluded any variable modifications. For prime side cleavage fragments, semitryptic searches were applied with no constraints for the orientation of the specific terminus and the following static modifications: carboxyamidomethylation of cysteine residues (+57.02 Da), dimethylation of lysine ε-amines (+28.03 Da) and thioacylation of amino termini (+88.00 Da).

Nonprime side cleavage sequences were derived as described previously (25). Briefly, prime-side cleavage sequences are rendered nonredundant and extended N-terminally until the first potential cleavage site of the digestion protease used to generate the peptide libraries (trypsin or Glu-C). Subsite cooperativity was analyzed with an in-house Perl script (http://clipserve.clip.ubc.ca/pics/) (33). This script compares the positional amino acid frequencies for all cleavage sites with positional frequencies for cleavage sites carrying a defined amino acid in a defined position. This analysis is performed for all amino acids in P3-P3′ with a positional frequency exceeding their natural abundance. Differences in positional frequency must exceed 10 percentage-points to be considered. Because subsite cooperativity describes a two-way effect between two subsites, this threshold must be surpassed in both subsites. This set of restrictions was shown to be suitable for the analysis of subsite cooperativity in HIV protease 1, for which manual parsing of PICS data unraveled subsite cooperativity (25) in agreement with a previous study (34).

TAILS—For degradomics analysis of native protein substrates the proteomic method TAILS was used to isolate and quantify cleaved neo-N-terminal peptides (26). HaCaT cells (human adult low calcium high temperature keratinocytes; kindly provided by Dr. Dirk Breitkreutz, DKFZ Heidelberg) were grown in Dulbecco's modified Eagle's medium, 5% calf serum to 70% confluency. Cells were then washed extensively to remove serum proteins and grown overnight serum-free. Cells were washed again, incubated in phenol red-free, serum-free medium and incubated with recombinant human meprin α or β. Conditioned medium proteins were harvested at 48 h where the cells were between 80–90% confluent, with protease inhibitors (1 mM EDTA, 1 mM phenylmethylsulfonyl fluoride) immediately added and the medium clarified by centrifugation (5 min, 500 $\times$ *g*), filtration (0.22 μm) and additional centrifugation (30 min, 8000 $\times$ *g*). The proteins were concentrated $\times$100 by ultrafiltration using Amicon Ultra-15 centrifugal filter units (3 kDa cutoff, Millipore). The sample buffer was exchanged to 50 mM HEPES, 150 mM NaCl, 10 mM $CaCl_2$ by five cycles of dilution and concentration within the same concentrating

[1] The abbreviations used are: PICS, (proteomic identification of protease cleavage site specificity); TAILS, (terminal amine isotopic labeling of substrates); FRET, fluorescence resonance energy transfer (fluorescence energy transfer); HaCaT, (human adult low calcium high temperature); BMP-1, (bone morphogenetic protein-1); KLK, (kallikrein); FGF-19, (fibroblast growth factor-19); VEGF-A, (vascular endothelial growth factor-A); MMP, (matrix metalloprotease); ZHE-1, (zebrafish hatching enzyme-1); MYO3, (myosinase 3).

device. Protein concentration was determined by bicinchoninic acid assay (Pierce, Rockford, IL) and Bradford assay (BioRad). The TAILS procedure including isotopic labeling, tryptic digestion, amine-terminal blocked peptide enrichment, liquid chromatography-MS/MS, data analysis and peptide abundance ratio were performed as previously described (26). MS2 scans were searched against a human International Protein Index protein database (version 3.42) (72,346 protein entries) by Mascot version 2.2 (Matrix Science) or X! Tandem (2007.07.01 release). Thereby, the following parameters were applied: semi-ArgC cleavage specificity with up to two missed cleavages, cysteine carbamidomethylation and peptide lysine iTRAQ as fixed modifications, N-terminal iTRAQ, N-terminal acetylation, and methionine oxidation as variable modifications. Tolerance for precursor and fragment ions was set at 0.4 Da, and the scoring scheme was electrospray ionization (ESI)-QUAD-time-of-flight (TOF). Secondary validation was performed using the Trans Proteomic Pipeline (TPP version 4.2, rev 0, Build 200811181145) (35, 36) employing the PeptideProphet (31) and iProphet (37) algorithms for peptide and protein assignment, and iTRAQ reporter ion intensities were quantified using Libra. Peptides were filtered by iProphet scoring, and only assignments with a probability $\geq$0.95 were included in further analysis. Multiple spectra were merged for individual peptides, and reporter ion intensity ratios were calculated by intensity dependent weighted averaging using the statistical models described in Prudova *et al.* (26). Accordingly, peptides with a protease and control ratio of $\geq$10 were considered as only present in the protease-treated sample and thus as derived from activity of the test protease. Peptides with a ratio $<$10 require biochemical validation that they represent cleavage sites in native protein substrates.

Cleavage Site Sequence Logos—Cleavage site sequence logos comprising P6-P6′ sequence data were generated with the method of Gorodkin *et al.* (38) as described previously (25).

Data and URLs—The LC-MS/MS data of this publication have been made publically available and the corresponding mzXML files may be downloaded from the ProteomeCommons.org Tranche system (https://trancheproject.org) using password "mep_coop" and hash:VZcm3OZkiX3krBRzBBnvCwU6xn+ayKooOmzjmNL650H5BWM23bUD9uL3GVmV4c5pa92 + 7eEGAdBTRj7YLVQxuKoguSIAAAAAAAAJ5A = =

Alignments, Phylogenetic Analysis, and Homology Modeling—The molecular structures of the catalytic domains were predicted by SwissMODEL (39). Modeling was performed with DeepView Swiss-PDB Viewer 3.7 (40) based on the crystal structure of human BMP-1 (3EDG) (41). Further energy minimization, modeling, and molecular graphics images were produced using the UCSF Chimera package from the Resource for Biocomputing, Visualization and Informatics at the University of California, San Francisco (supported by National Institutes of Health grant P41 RR-01081) (42). All structures were further analyzed using the MolProbity webtool (http://molprobity.biochem.duke.edu) (43).

Sequences comprising the complete catalytic domains up to cysteine 198 (astacin numbering) of different astacins were aligned using the ClustalX 2.0 (44), setting opening GAP penalties to values of 5.0. Sequence alignments were edited with GeneDoc 2.6.02 (45). Phylogenetic trees of astacins from different species were computed with MrBayes (46) using the PAM-Dayhoff distance matrix and consensus trees were visualized by TreeView 1.16.7 (47).

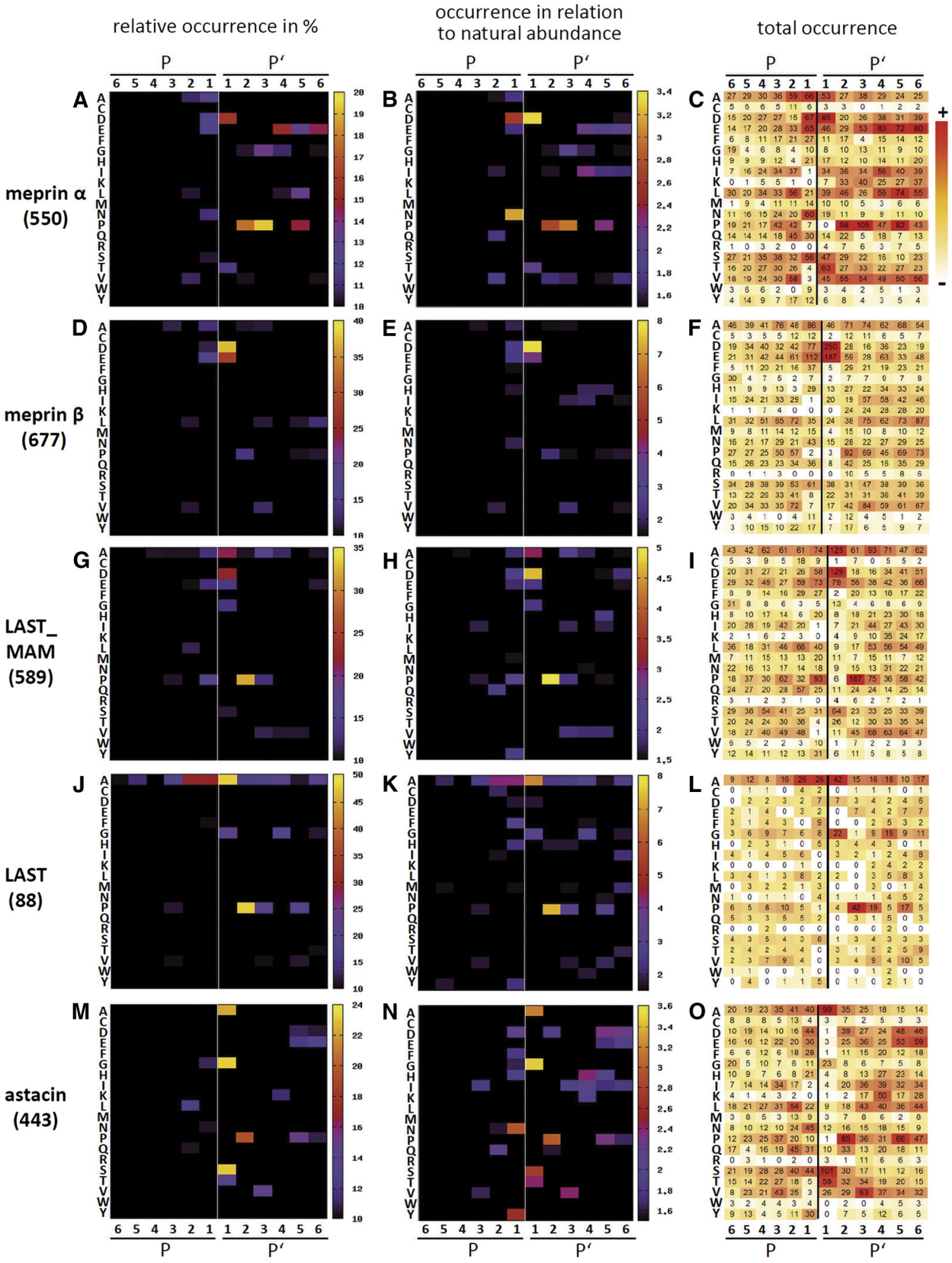
Activity Assays and Matrix-assisted Laser Desorption Ionization/Time of Flight (MALDI-TOF) Using Fluorogenic Peptides to Validate PICS—To test the proteolytic activity of the astacin proteases against different peptide sequences we used the PICS data to design quenched fluorogenic peptides as new peptide substrates (Fig. 4). Peptides 1 to 7 were obtained from Peptanova/Sanhausen, Germany, peptide 8 from R&D Systems (Minneapolis, MN), peptide 9 from BioLux/Stuttgart, Germany, and peptide 10 from Bachem (Bubendorf, Switzerland). The enzyme activity was measured with the fluorescent spectrometer Varioskan Flash (Thermo Scientific). Data were analyzed using "SkanIt Software 2.4 for Varioskan Flash". All enzymes were buffered in 50 mM HEPES, pH 7.5 and used in a final concentration of 5 nM up to 320 nM, depending on the protease. Measurement occurred at 37 °C and the fluorescence was detected every 8 s for 120 min for meprin α, meprin β, and astacin, every 12 s for 200 min for LAST_MAM and every 15 s for 250 min for LAST. The proteolytic activity was calculated relative to the emission at 405 nm with an excitation at 320 nm. The best three substrates for each protease were analyzed by MALDI-TOF for identification of cleavage sites. Therefore, 1 μg peptide was incubated with 5–100 nM enzyme for 12 h at 37 °C and sequenced at the protein micro-sequencing center of the Institut Fédératif de Recherche (IFR) 128 (Lyon, France).

Processing of FGF-19 by Meprin β and Scratch Assay Analyses with Cultured Keratinocytes—Recombinant human FGF-19 (10 μg) (Biovision, Mountain View, CA) was incubated with 250 nM recombinant meprin β for 90 min at 37 °C. Proteins were analyzed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). For N-terminal sequencing, proteins were blotted onto a polyvinylidene difluoride membrane, stained with Coomassie Brilliant Blue and sequenced at the protein microsequencing center of the Institut Fédératif de Recherche (IFR) 128 (Lyon, France).

The keratinocyte scratch wound migration assay was performed in 6-well cell culture plates. HaCaT cells were grown in Dulbecco's modified Eagle's medium supplemented with 10% (v/v) fetal calf serum and penicillin and streptomycin to 90% confluence. The wound was simulated with a sterile 100 μl pipet tip by scratching the cell monolayer in the center of the well. Full-length FGF-19 (1.0 ng) and 1.0 ng FGF-19 preincubated with 20 nM meprin β (final concentration 0.1 nM) were applied to the cells. Cell migration was documented using phase contrast microscopy at 0 h, 18 h, and after 27 h. The cell free area (27 h) was measured with ImageJ 1.45 and calculated in relation to the scratch at the primary time point.

RESULTS

Cleavage Specificity Profiles of Astacin Proteases Reveal a Unique Asp-N Motif—PICS is a unique substrate profiling approach that simultaneously determines prime and nonprime side cleavage site specificity (25, 33). Being based on the cleavage of hundreds of peptides this enables robust statistical analysis to be performed to determine both the cleavage site specificity and to uncover subsite cooperativity in substrate preferences. The PICS approach is based on generating peptide libraries from cultured cells for use as diverse peptide substrates in protease screens for cleavage site specificity. Here we used human embryonic kidney cells (HEK) as the proteome source. The cellular proteomes are harvested and degraded into oligopeptides with specific endoproteases such as trypsin or Glu-C to generate overlapping peptide libraries. After protease inactivation, these peptides are further purified and primary amines (N-terminal α–amines and lysine ε–amines) are chemically protected. The proteome-derived peptide libraries are then incubated with the protease of interest to generate neo-N-termini after peptide cleavage for subsequent identification. These unprotected N-terminal α–amines at the prime-side of the cleavage products are selectively biotinylated to enable affinity isolation and LC-MS/MS based sequence identification. The sequence of the


relative occurrence in %
occurrence in relation to natural abundance
total occurrence
P
P′
6 5 4 3 2 1
1 2 3 4 5 6
A C D E F G H I K L M N P Q R S T V W Y
A
B
C
meprin α (550)
D
E
F
meprin β (677)
G
H
I
LAST_ MAM (589)
J
K
L
LAST (88)
M
N
O
astacin (443)
+
-

corresponding nonprime cleavage product is determined by bioinformatics. The high number of cleaved peptide sequences allows for statistical calculation of positional occurrences, which are visualized as heatmaps and sequence logos. Positional occurrences are optionally corrected for natural amino acid abundance to avoid bias stemming from unequal amino acid distribution in cellular proteomes.

It is important to note that PICS does not detect cleavage sites in native folded proteins. Rather, cleavage occurs on denatured peptides after tryptic prefragmentation of proteomes that presents to the proteases under study thousands of diverse peptides from biologically relevant sources rather than synthetic peptide libraries that cannot be analyzed by mass spectrometry. To characterize cleavage specificity in native proteins other approaches were developed such as TAILS (26, 27, 48).

Using PICS, we identified up to 677 cleaved peptides for meprin β and the other proteases in one or two MS/MS analyses (Fig. 1, supplemental Tables S1–S5, supplemental material S1–S5). All results are presented based on the nomenclature by Schechter and Berger, which designates residues C-terminal to the cleavage site as prime (P′) and N-terminal peptide residues as nonprime (P) (49). Peptide or protein P and P′ sites interact with subsites within the active site cleft of the proteolytic enzyme, designated as S and S′, each of which can be made up by several residues. Typically, most methods of profiling substrate specificity are limited to the nonprime side and usually are limited to three or four residues (50).

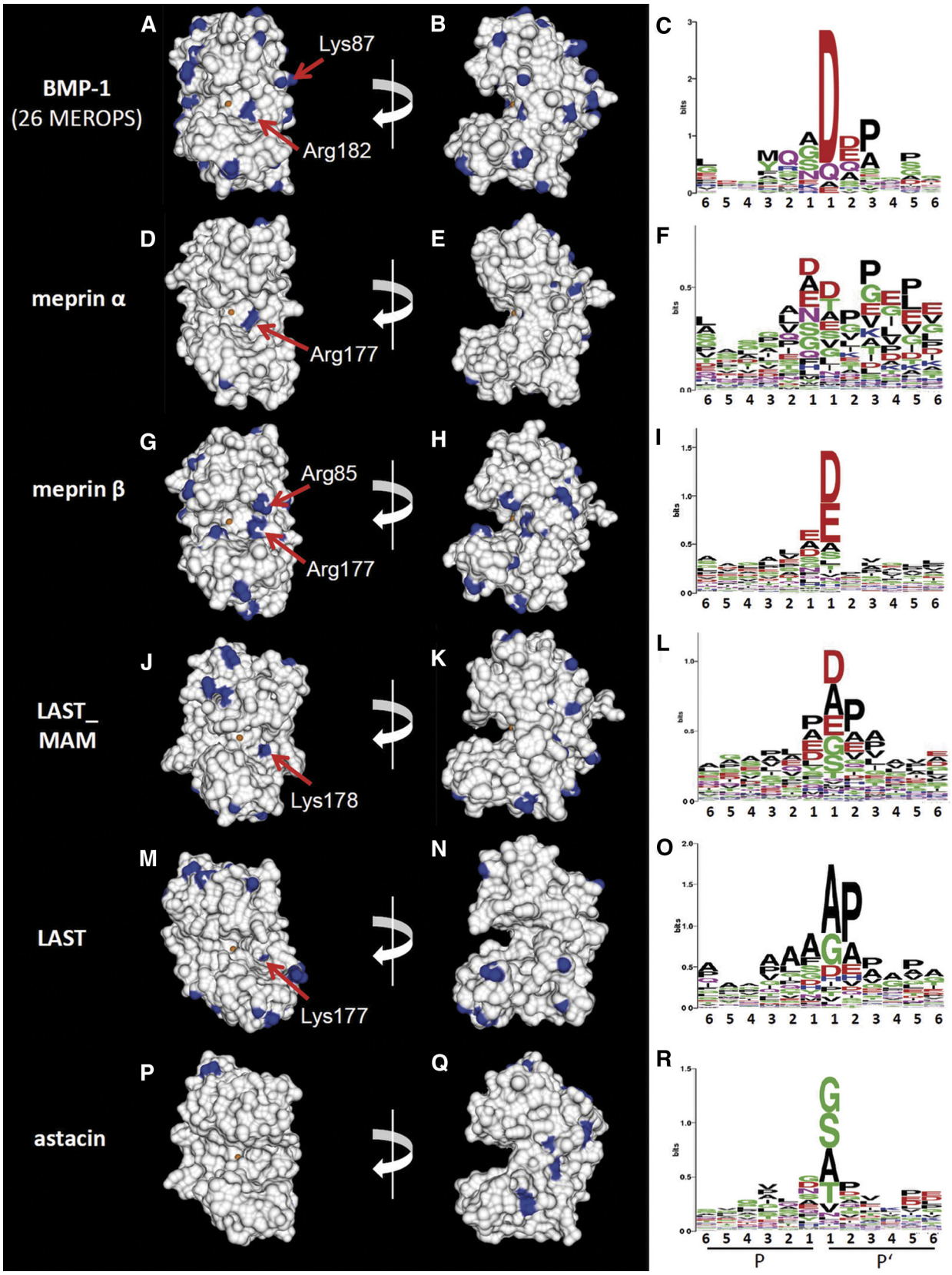
Statistical evaluation of all amino acid residues from P6 to P6′ of all tested astacin proteases revealed no pre-eminent residues in the nonprime subsites, but showed slight preferences for negatively charged residues in P1 (Fig. 1). This was in sharp contrast to the prime side of the cleaved peptide bond. Here, striking similarities could be seen for meprin α, meprin β, and LAST_MAM where there was a strong preference for negatively charged residues in P1′ (Fig. 1*A*–1*I*). Notable was the case of meprin β, where over 65% of all amino acid residues in P1′ were aspartate or glutamate (Fig. 1*D*–1*F*). One additionally conserved position that became obvious throughout all analyzed astacin proteases was proline in P2′ (Fig. 1). In addition to direct interactions within the S2′-subsite, this is presumably because of the steric constraints caused by its ring shaped side chain, which restricts the freedom of rotation and confers *cis-trans* isometry to the peptide chain.

Structural Basis for Cleavage Specificity in Astacin Proteases—Homology models of each protease domain were generated to identify the interacting residues within the active site cleft (Fig. 2). In addition to the crystal structure of the crayfish enzyme astacin, the structure of BMP-1 (3EDG) was used as a template because no other astacin structures were available. BMP-1 has a known preference for aspartate in P1′ with 26 BMP1-cleavage sites in matrix proteins, growth factors and growth factor antagonists listed in MEROPS (http://merops.sanger.ac.uk) (51), the protease database (Fig. 2) (52–56). The positively charged residues in the protease with potential for binding aspartate within the active site cleft are marked in blue (Fig. 2). In BMP-1, the positively charged guanidinium group of Arg182 is responsible for this unique specificity (41) (Fig. 2*A*). Likewise, corresponding positively charged residues in S1′ subsites of the other enzymes could be identified as binding partners for the negatively charged aspartate or glutamate residues in P1′ (Fig. 2*D*, 2*G*, 2*J*, and 2*M*). Our conclusion is supported by a recently published structure of the zebrafish hatching enzyme ZHE1 (57) with an arginine residue (Arg182) in the same orientation as seen for BMP-1 and exhibiting the same cleavage specificity. The only enzymes not favoring negatively charged amino acids in P1′ were LAST and the crayfish astacin (Fig. 1*M*–1*O* and Fig. 2*P*–2*R*, respectively). The importance of a positive group in S1′ for defining substrate specificity is supported by this observation as both prefer small aliphatic residues, most dominant of which are glycine, serine, and alanine.

To visualize the species diversity of astacin proteases, we generated a phylogenetic tree displaying clusters of chosen members regarding their relationship and function (Fig. 3*A*). Most interestingly, by aligning the amino acid sequences of the 765 astacin proteases annotated to date in MEROPS we found a conserved arginine or lysine residue in the S1′ subsites of more than 80% of the entries, indicating a unique common feature governing the cleavage specificity of this family (Fig. 3*B*). LAST_MAM does not exhibit the conserved GQR/K motif downstream of the archetypical Met-turn (58), but contains two lysine residues (Lys178, Lys179) in a small loop immediately nearby (Fig. 3*B*). These positively charged ε-amino groups are orientated into the active site cleft in our models (Fig. 2*J*) thereby determining the specificity for aspartate in P1′ (Fig. 2*L*).

Additional positive charges of Arg85 and Lys87 contributing to the S1′ subsite and possibly to the S2′ subsite could be

FIG. 1. **Astacin metalloproteases cleavage sites profiled by PICS.** A tryptic PICS library was generated from the lysate of cultured HEK 293 (human embryonic kidney) cells and incubated with human meprin α (*A–C*) and meprin β (*D–F*), LAST_MAM (*G–I*) and LAST (*J–L*) from the horseshoe crab, and crayfish astacin (*M–O*) at ratios of 1:100 (enzyme to library, 200 μg of PICS library). Beneath the protease names are the total number of identified cleaved peptides. PICS libraries are analyzed by multiple sequence alignments and displayed as heat maps for the relative occurrence (*left* panel) and applying correction for natural amino acid abundance (relative occurrence/natural abundance of amino acids; *middle* panel). The total number of identified residues in P6 to P6′ is shown in the *right* panel. The scissile peptide bond between P1 and P1′ is indicated by a white (*left* and *middle* panel) or black line (*right* panel). Heat maps were created using gnuplot (www.gnuplot.info). Single letter code for amino acid residues plotted is on the y axes. P and P′ subsite positions are as shown on the x axes.


A
B
C
BMP-1
(26 MEROPS)
Lys87
Arg182
D
E
F
meprin α
Arg177
G
H
I
meprin β
Arg85
Arg177
J
K
L
LAST_
MAM
Lys178
M
N
O
LAST
Lys177
P
Q
R
astacin
bits
P
P′

identified in meprin β and in the BMP-1/tolloid group (Fig. 3*B*). These highly conserved residues are located upstream of the zinc binding sequence in the upper rim of the active site cleft (Fig. 2*A* and 2*G*). Moreover, meprin β even contains another positively charged residue (Arg123) in the lower rim of the active site cleft as part of the S1 subsite (Fig. 2*G*). Such a strong cationic character might explain the relatively high frequency of glutamate and aspartate in P1 in meprin β over the other family members profiled (Fig. 2*G*). Astacin, on the other hand, exhibits secondary specificity for aspartate and glutamate in P5′ and P6′ (Fig. 1*M*–1*O*, Fig. 2*R*), which corresponds to a similar observation for the positions P3′ to P6′ in meprin α (Fig. 1*A*–1*C*; Fig. 2*F*). In the rotated models (Fig. 2*E* and 2*Q*) these positively charged residues (blue) can be seen in the prime region.

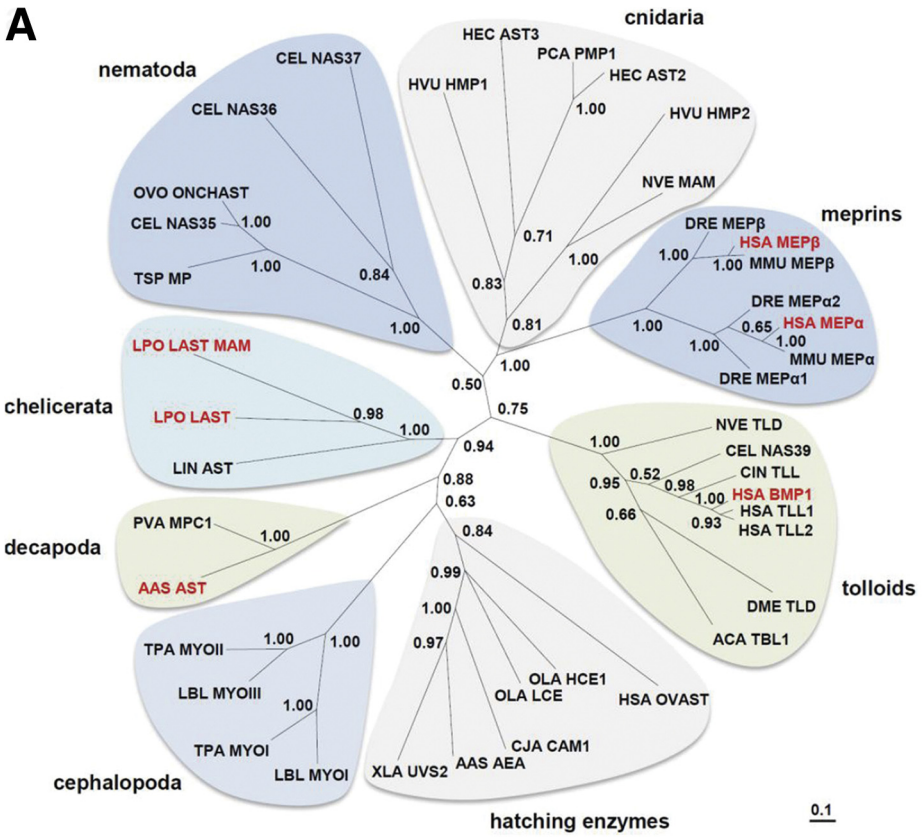
Subsite Cooperativity Revealed by PICS—The substrate selectivity and cleavage specificity of certain proteases is further determined by subsite cooperativity. This is the case where the occurrence of a specific amino acid residue in one subsite of a substrate peptide definitely influences the frequency of occurrence for a residue in a different subsite. An example is HIV protease 1, which distinctly prefers alanine or glycine in P1 if there is a leucine positioned in P3 (25). Using a web-based PICS analysis (33) to determine potential subsite cooperativity we found that negatively charged residues in P1′ *decreased* the occurrence of the preferred proline in P2′ and *vice versa*. This was consistently observed in all astacins analyzed herein (Fig. 4*A,* 4*D*, 4*G*, 4*J,* and 4*M*) and was surprising given that both Asp and Pro are preferred in P1′ and P2′, respectively but both select against the other occurring in peptide substrates. Indeed, proline in P2′ strongly enhanced the occurrence of small uncharged residues in P1′, whereas concomitantly disfavoring aspartate and glutamate (Fig. 4*A,* 4*D*, 4*G*, 4*J,* and 4*M*). This is most obvious in the case of meprin β, the protease with the strongest preference for negatively charged residues in P1′ position (Fig. 4*D*).

Validation of Cleavage Specificity and Subsite Cooperativity using Fluorogenic Peptide Substrates—To validate the PICS data we analyzed 10 fluorogenic substrates comprising 4 to 11 different amino acid residues (Table I; Fig. 4, and supplemental Fig. S1). All peptides are fluorescence resonance energy transfer compounds, containing a fluorescent (7-methoxycoumarin-4-yl)-acetyl group that is quenched by a 2,4-dinitrophenyl group, until proteolytic processing occurs within the peptide. In addition, cleavage products of the three best substrates for each protease were analyzed by MALDI-TOF (DthStar, Applied Biosystems) to determine the exact cleavage site.

Peptides 1–4 do not contain any negatively charged amino acid residues, thus being predicted to be good substrates for LAST and astacin that prefer small aliphatic amino acids in P1′. In contrast, peptides 5–10 exhibit several aspartate and glutamate residues. Indeed, peptide 9 (YVADAPK) proved to be the best substrate for meprin α, meprin β, and LAST_MAM (Fig. 4*B*, 4*E*, and 4*H*), which all prefer aspartate in P1′. MALDI-TOF analysis revealed cleavage between alanine and aspartate for all three enzymes (Fig. 4*C*, 4*F*, and 4*I*). In addition, meprin α and β cleaved peptides 6 and 8 (Fig. 4*B*, 4*C*, 4*E*, and 4*F*; supplemental Fig. S1), both of which contain negatively charged residues (Table I). The MALDI-TOF analysis of the corresponding peptide fragments after meprin incubation showed that meprin β cleaves only a single peptide bond, whereas meprin α cuts at several sites reflecting its broader specificity (Fig. 4*C*, 4*F*, and 4*I*). Interestingly, peptide 7, which is identical to 9 in the last four positions, including an aspartate residue (Table I), was only weakly processed by both meprins (Fig. 4*E*; supplemental Fig. S1). As verified by the web based PICS analysis tool, this is probably because of the presence of glutamate in P2, which is disfavored by meprin α and β, if aspartate is in the P1′ position. In addition, meprin α but not β showed slight activity against the substrates peptides 1, 5, 3, and 4, corroborating the broader specificity revealed by PICS (supplemental Fig. S1). Presumably for the same reason, cleavage of peptides 3 and 4 could also be observed for LAST_MAM and this was verified by MALDI-TOF analysis (Fig. 4*H* and 4*I*).

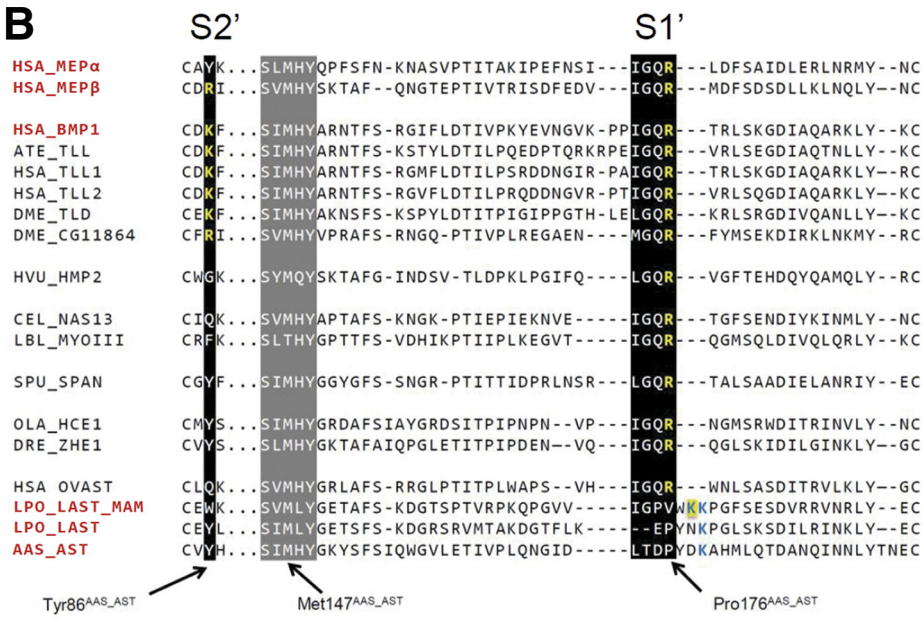
Astacin cleaved the less negatively charged peptide 4 most efficiently, followed by 9, 5, and 6 (Fig. 4*N* and 4*O*; supplemental Fig. S1). All substrates contained at least one small uncharged aliphatic amino acid residue reflecting the primary cleavage specificity of this metalloprotease in P1′ (Figs. 1 and 2). Indeed, peptide 4 was proteolytically processed at five positions, explaining the highest activity for this compound (Fig. 4*O*). Although, LAST revealed a similar specificity at this position (Fig. 1), peptide 9 could be identified as the best substrate, followed by 1, 3, 8, 6, and 5 (Fig. 4*K* and 4*L*; supplemental Fig. S1). Both proteases, LAST and astacin,

FIG. 2. **Structures and homology models of astacin protease domains illustrating the characteristics for cleavage specificity.** The homology surface models were computed based on the crystal structure of human BMP-1 (3EDG). The surface of astacin from *A. astacus* is based on a crystal structure in complex with a transition-state analog inhibitor (1QJI) (68). Highlighted are characteristic basic residues (blue) at the surface and the central zinc ion (orange). *Left* panel shows all molecules in standard orientation, with the nonprime side to the left of the zinc ion and the prime side to the right. The *middle* panel illustrates the catalytic domains rotated 90° on the vertical axis. *Red* arrows indicate specific residues or charged regions at the surface, important for cleavage specificity. The *right* panels display the relative abundance of preferred amino acid residues in WebLogo style (70). For each position in P6 to P6′, the height of the single letter code for amino acids reflects its occurrence rate. Color-coding: acidic residues in *red*, basic residues in *blue*, polar residues (including tyrosine and glycine) in *green*, hydrophobic residues (including alanine and proline) in *black*. WebLogo ordinates are scaled in bits as described previously (70).



Species Abbreviations

AAS	Astacus astacus (crayfish)	**HSA**	Homo sapiens (human)	**OLA**	Oryzias latipes (medaka fish)
ACA	Aplysia californica (sea hare)	**HVU**	Hydra vulgaris (cnidaria)	**OVO**	Onchocerca volvolus (worm)
CEL	Caenorhabditis elegans (worm)	**LBL**	Loligo bleekeri (squid)	**PCA**	Podecoryne carnea (cnidaria)
CIN	Ciona intestinalis (sea squirt)	**LIN**	Loxoceles intermedia (spider)	**PVA**	Panaeus vannamei (shrimp)
DME	Drosophila melanogaster (fly)	**LPO**	Limulus polyphemus (horseshoe crab)	**TPA**	Todarodes pacificus (squid)
DRE	Danio rerio (zebrafish)	**MMU**	Mus musculus (mouse)	**TSP**	Trichinella spiralis (worm)
CJA	Gallus gallus (chicken)	**NVE**	Nematostella vectensis (cnidaria)	**XLA**	Xenopus laevis (frog)
HEC	Hydractinia echinata (cnidaria)				



cleaved peptide 9 predominantly between aspartate and alanine (Fig. 4*L* and 4*O*). For LAST, minor cleavage could also be observed between alanine and aspartate, correlating with the preference in P1′ revealed by PICS (Fig. 2*O*).

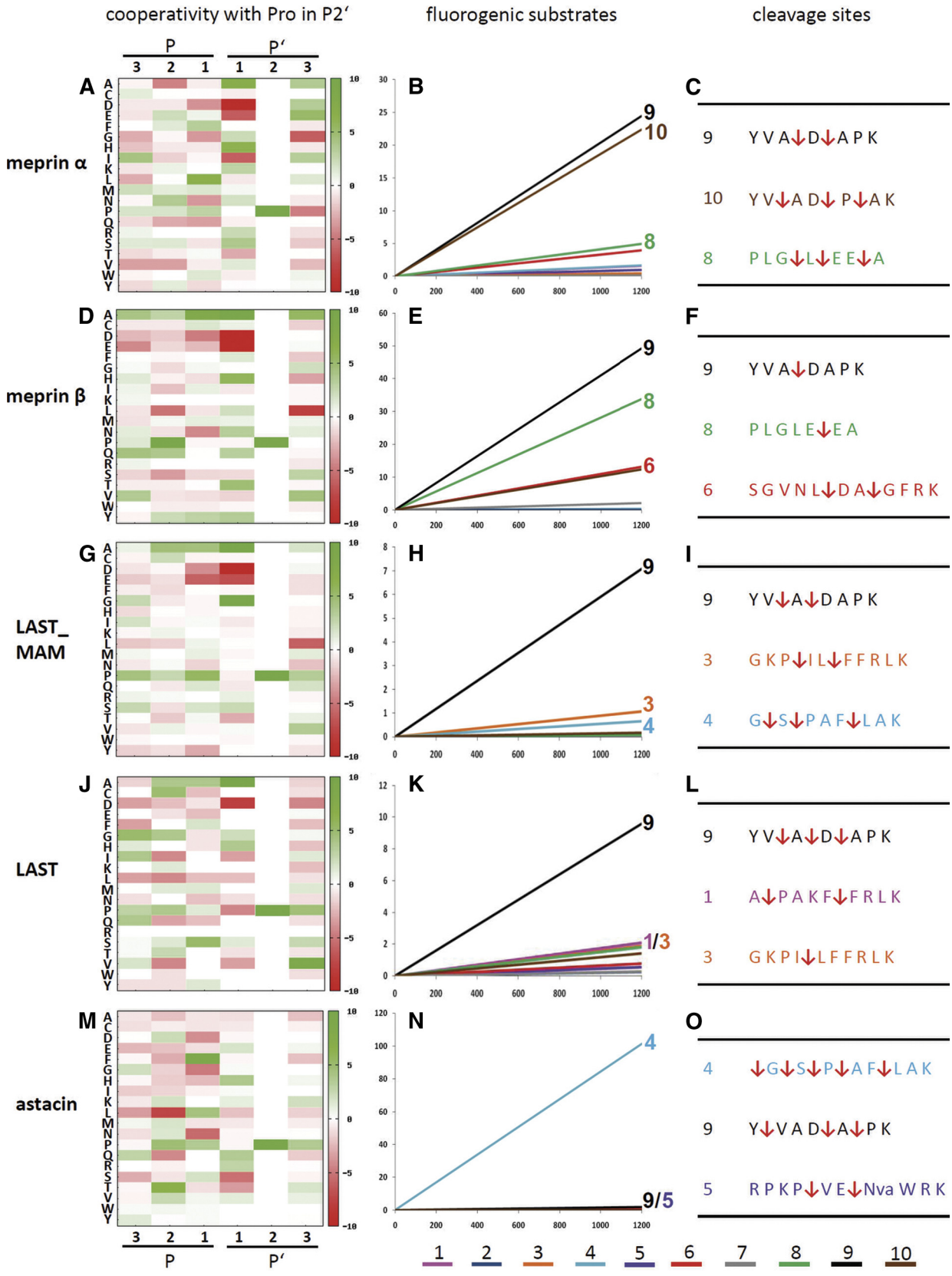
To further analyze the subsite cooperativity effect, we characterized the activity of all enzymes against peptide 10 (YVADPAK), identical to peptide 9 (YVADAPK), but switching the proline from P3′ to P2′. As predicted, all enzymes cleaved peptide 10 less effectively than peptide 9 (Fig. 4B, 4*E*, 4*H*, 4*K*, and 4*N*; supplemental Fig. S1). Thus, changing the amino acids in the P2′ and P3′ positions, confirms the negative subsite cooperativity shown by PICS for Pro at S2′ (Fig. 4*A*, 4*D*, 4*G*, 4*J*, and 4*M*). By contrast, meprin α showed only a slight decrease in activity against peptide 10 compared with 9, which in this case is because of a shift of the cleavage spot. For meprin α, the PICS data revealed the strongest preference for aspartate and glutamate not only in P1′, but also in P1, and for proline in P3′ additionally (Fig. 1*A*–1*C*). Thus, cleavage in this peptide occurs most probably at A(P1′)-D(P2′)-P(P3′) and between D(P1) and P(P1′). Indeed, this processing event could be observed by MALDI-TOF analysis, but as also shown by MALDI-TOF, there are other sites available for meprin α cleavage within this peptide (Fig. 4*C*).

Structural Basis of Subsite Cooperativity in Astacin Proteases—Using *in silico* homology modeling of the catalytic domain of human meprin β in complex with substrates 9 and 10, we were able to visualize the molecular effects of switching the proline within the peptide sequence (supplemental Fig. S2). Both substrates fit nicely into the active site cleft of the enzyme, building a salt bridge between the aspartate carboxylate (red) of the peptide and the guanidinium group of the enzyme's Arg177 (supplemental Fig. S2a–f). In this state, with the peptide bound to the enzyme, no obvious steric hindrance could be observed, whether proline (yellow) is located in P2′ or P3′. However, the proline in close proximity in P2′ would considerably influence the positioning of the aspartyl sidechain in P1′, because the rotational freedom would be restricted by the proline ring. This might influence a tighter binding of the substrate into the active site cleft (supplemental Fig. S2g and h).

Identification of Cleavage Sites in Native Substrates for Meprin α and β—The TAILS degradomics approach enriches the N-termini of proteins at their natural N-terminus and the neo-N-terminus of their protease cleavage products (26, 27). By isotopic labeling and quantification these can be effectively distinguished from background proteolysis products present in every sample. Using keratinocytes cultured in the presence or absence of meprins we identified 453 cleavage sites for meprin α and 545 for meprin β (supplemental Table S6 and S7, supplemental material S6 and S7, respectively). From these data the cleavage specificity was deduced covering subsites from P6 to P6′ (Fig. 5). Interestingly, meprin α showed a much higher preference for negatively charged residues in native substrates than detected by PICS in the peptide libraries (Fig. 5*A*, 5*C*, and 5*E*). In cleaved native proteins nearly 50% of all residues in P1′ were aspartate or glutamate. In contrast, for meprin β, the results from PICS and TAILS were strikingly similar (Fig. 5*B*, 5*D*, and 5*F*). Again aspartate and glutamate in P1′ made up 65% of the total number of amino acids in this position. Moreover, fixing proline in P2′ revealed the same cooperative effect (Fig. 5*G* and 5*H*) as observed with PICS (Fig. 5*A* and 5*D*).

Most interestingly, the majority of identified cleavage sites (361) in native substrates were identical for both meprins (Fig. 6*A*, 6*C*, 6*F*, and 6*I*). Analyses of those TAILS cleavage sites that were unique for meprin α (92) or β (184) (Fig. 6*A*; supplemental Tables S8 and S9, supplemental material S8 and S9, respectively), resulted in specificity profiles that clearly distinguished the two meprins. Here, meprin α exhibited the strongest preference for Thr>Ala>Val followed by aspartate (Fig. 6*B*, 6*E* and 6*H*) with the TAILS data obtained from native proteins nicely matching the PICS data obtained from denatured peptide libraries. Meprin β instead showed an even greater enhanced specificity for negatively charged residues in native proteins as revealed by TAILS not only in P1′, but also in P2′ and P3′, and even in the nonprime sides P1 and P2 (Fig. 6, *D*, *G* and *J*). Thus, human meprin β can cleave within polypeptide chains consisting of aspartate and glutamate residues only, for example EE/DEAE in the protein O43290 (supplemental Table S7). A homology model

FIG. 3. **Phylogenetic tree of astacin proteases and amino acid sequence alignment of the cleavage specificity determining region within the catalytic domain.** *A*, The tree was generated using an alignment of the catalytic domains of all displayed proteases. The enzymes characterized in this work are displayed in *red*. Protein clusters are shown against different backgrounds. The numbers show the relative probability of branching. Swissprot/EMBL accession numbers of the astacins: AAS AEA (O44072); AAS AST (P07584); ACA TBL1 (P91972); CEL NAS35 (P98060); CEL NAS36 (Q18206); CEL NAS37 (Q93243); CEL NAS39 (Q20176); CIN TLL (Q4H2P2); CJA CAM1 (P42662); DME TLD (P25723); DRE MEPα1 (Q5RHM1); DRE MEPα2 (Q5RHM2); DRE MEPβ (Q08CC4); HEC AST2 (Q2MCX8); HEC AST3 (Q2MCX7); HSA BMP1 (P13497); HSA MEPα (Q16819); HSA MEPβ (q16820); HSA OVAST (Q6HA08); HSA TLL1 (O43897); HSA TLL2 (Q9Y6L7); HVU HMP1 (Q25174); HVU HMP2 (Q9XZG0); LBL MYOI (Q8IU47); LBL MYOIII (Q8IU44); LIN AST (A0FKN6); LPO LAST (B4F319); LPO LAST_MAM (B4F320); MMU MEPα (P28825); MMU MEPβ (Q61847); NVE MAM (A7SJ13); NVE TLD (Q27W05); OLA HCE1 (P31580); HCE LCE (P31579); OVO ONCHAST (Q2YFS7); PCA PMP1 (O62558); PVA MPC1 (Q20AS7); TPA MYOI (Q8IU46); TPA MYOII (Q8IU45); TSP MP (Q8T5Z5); XLA UVS2 (P42664). *B*, The alignment partially displays the amino acid sequence around the astacin typical Met-turn (highlighted in *gray*). The consensus motifs comprising the S1′ and S2′ regions are displayed in black. The amino acids responsible for the specificity for negatively charged residues are shown in yellow. Lysine residues that are closely related to the S1′ pocket are displayed in blue. A second lysine, which is only present in LAST_MAM is additionally highlighted with a *yellow* background. The enzymes characterized in this work are shown in *red*. Uniprot accession numbers are ATE_TLL (Q75UQ6); DME_CG11864 (Q9VJN9); SPU_SPAN (P98068); DRE_ZHE1 (Q75NR9), and as described above.


cooperativity with Pro in P2'
fluorogenic substrates
cleavage sites
P
P'
3 2 1 1 2 3
A
meprin α
B
C
9 Y V A↓D↓A P K
10 Y V↓A D↓ P↓A K
8 P L G↓L↓E E↓A
D
meprin β
E
F
9 Y V A↓D A P K
8 P L G L E↓E A
6 S G V N L↓D A↓G F R K
G
LAST_ MAM
H
I
9 Y V↓A↓D A P K
3 G K P↓I L↓F F R L K
4 G↓S↓P A F↓L A K
J
LAST
K
L
9 Y V↓A↓D↓A P K
1 A↓P A K F↓F R L K
3 G K P I↓L F F R L K
M
astacin
N
O
4 ↓G↓S↓P↓A F↓L A K
9 Y↓V A D↓A↓P K
5 R P K P↓V E↓Nva W R K
1 2 3 4 5 6 7 8 9 10

of the catalytic domain of meprin β in complex with the cleavable sequence EDEAE demonstrates the strong ionic interactions between P1 and Arg123, P1′ and Arg177, and P2′ and Arg85 (Fig. 6*L*; pdb-file at http://www.bio.uni-mainz.de/zoo/cbp/98.php), whereas meprin α only binds the aspartate in P1′.

Physiological Relevance of Candidate Substrates—The physiological relevance of identified substrates can be illustrated by three examples: (i) VEGF (vascular endothelial growth factor)-A, (ii) proKLK (kallikrein)7, and (iii) FGF (fibroblast growth factor)-19 (Table II, supplemental Table S10, supplemental material S10).

(i) Recently, meprin α has been recognized as a pro-angiogenic enzyme as demonstrated by a disturbed vascular system in meprin α knockdown zebrafish embryos (18). In this regard, VEGF-A was identified by TAILS as a promising candidate substrate for meprin α and, *in vitro* validation assays, generated cleavage products that were homologous to those detected in cell lysates of wild-type fish. Indeed, the exact cleavage site (APMA/EGGG) identified in VEGF-A because of meprin α processing was also found using TAILS (Table II, supplemental Table S10, supplemental material S10).

(ii) The second example exhibits proteolytic interactions between meprins and their activators, the human kallikrein-related peptidases (KLKs) in the protease web. Meprins are secreted as inactive zymogens requiring proteolytic activation (14, 28). Certain tryptic serine proteases, including KLKs are specific activators for human meprins (12, 59). *Vice versa*, meprin β triggers the activation of proKLK7 *in vitro*, by cleaving within the propeptide sequence, which then destabilizes the propeptide remnant favoring final activation by trypsin. Again, the exact cleavage site (EAQG/DKII) identified in proKLK7 because of meprin β activity was also found in the TAILS data (Table II, supplemental Table S10, supplemental material S10).

(iii) To validate a cleavage event in another novel substrate, we incubated recombinant human FGF-19 with meprin β and analyzed the processing by SDS-PAGE and Edman degradation (Fig. 7*A*) that again confirmed the cleavage site (LAFS/DAGP) identified by TAILS (Table II, supplemental Table S10, supplemental material S10). To further investigate the physiological relevance of FGF-19 processing, a scratch assay was performed with cultured keratinocytes (HaCaTs), incubated with full length FGF-19, the meprin-processed growth factor, or meprin β alone (Fig. 7*B*). The proliferation/migration of the keratinocytes to fill the scratched gap in the cell layer was significantly enhanced by FGF-19 compared with the control (Fig. 7*B* and 7*C*). Interestingly, cleavage of the growth factor by meprin β revealed a decrease in the proliferation/migration of the cells, which is even lower than detected for the untreated keratinocytes (Fig. 7*C* and 7*D*).

DISCUSSION

In profiling the astacins, we report the first family-wide simultaneous profiling of the prime and nonprime sides in protease cleavage sites of peptides and native substrates. We find that among the extracellular proteolytic enzymes, astacins exhibit a unique family-wide specificity for negatively charged amino acid residues in P1′, structurally based on a conserved positive residue within the active site cleft. This is a unique feature among all extracellular proteolytic enzymes (51) including matrix metalloproteases (MMPs) (25) and ADAMs (a disintegrin and metalloproteases domain) (60), or kallikreins (61) and was reflected by meprin cleavage sites we identified in native substrates using TAILS. Whereas mouse meprin β is known to display a preference for Asp and Glu at P1′ (21) only one other vertebrate protease is known to cleave N-terminal to aspartate, the PHEX endopeptidase (62). However, this metalloprotease belongs to the neprilysin family, enzymes that hydrolyze only peptides and not globular proteins, predominantly between hydrophobic residues (63).

The second striking finding was the global preference for proline in P2′. However, and very surprisingly, all cleaved peptides with proline in P2′ showed a decreased frequency of negatively charged residues in P1′. *Vice versa*: in peptides containing aspartate or glutamate in P1′ the frequency of proline to occur in P2′ was significantly lower. Such negative cooperativity was unexpected for a sequence motif showing such a strong preference for both these residues individually and highlights the need for analyses that profile specificity using diverse libraries of whole peptides rather than positional scanning of fixed peptide sequences. Although proline at P2′ had been found before for crayfish astacin (64, 65), it proved to be a common feature for the astacin family, at least for those representatives profiled here.

The strong prime side cleavage site determinants of five representative members of the astacin family of metalloproteases, namely meprin α, meprin β, LAST, LAST_MAM and astacin contrast other metalloproteases (*e.g.* MMPs or ADAMs) having a balanced relation (25, 60), whereas serine and cysteine proteases are more often influenced by the nonprime side (25). Generally, metalloproteases, like MMPs or

FIG. 4. **Subsite cooperativity validated by fluorogenic peptide cleavage assays.** A web-based PICS analysis (33) enables one to set proline as a fixed residue in P2′ thus visualizing those amino acids that stay in cooperative relation or those cleaved peptides having Pro at P2′ (*left* panel). Single letter code for amino acid residues plotted is on the *y* axes. P and P′ subsite positions are as shown on the *x* axes. Quenched fluorogenic peptides (1–10; Table I) were incubated with astacin metalloproteases and the relative amount of product (*y* axis) was calculated with regard to fluorescence intensity at 405 nm with an excitation at 320 nm, and plotted against the time in seconds (*x* axis, *middle* panel). Sequences of the best cleaved substrates are shown in the *right* panel. Cleavage sites determined by MALDI-TOF analysis are indicated by *red* arrows. Color codes used to plot rates for each peptide are indicated.

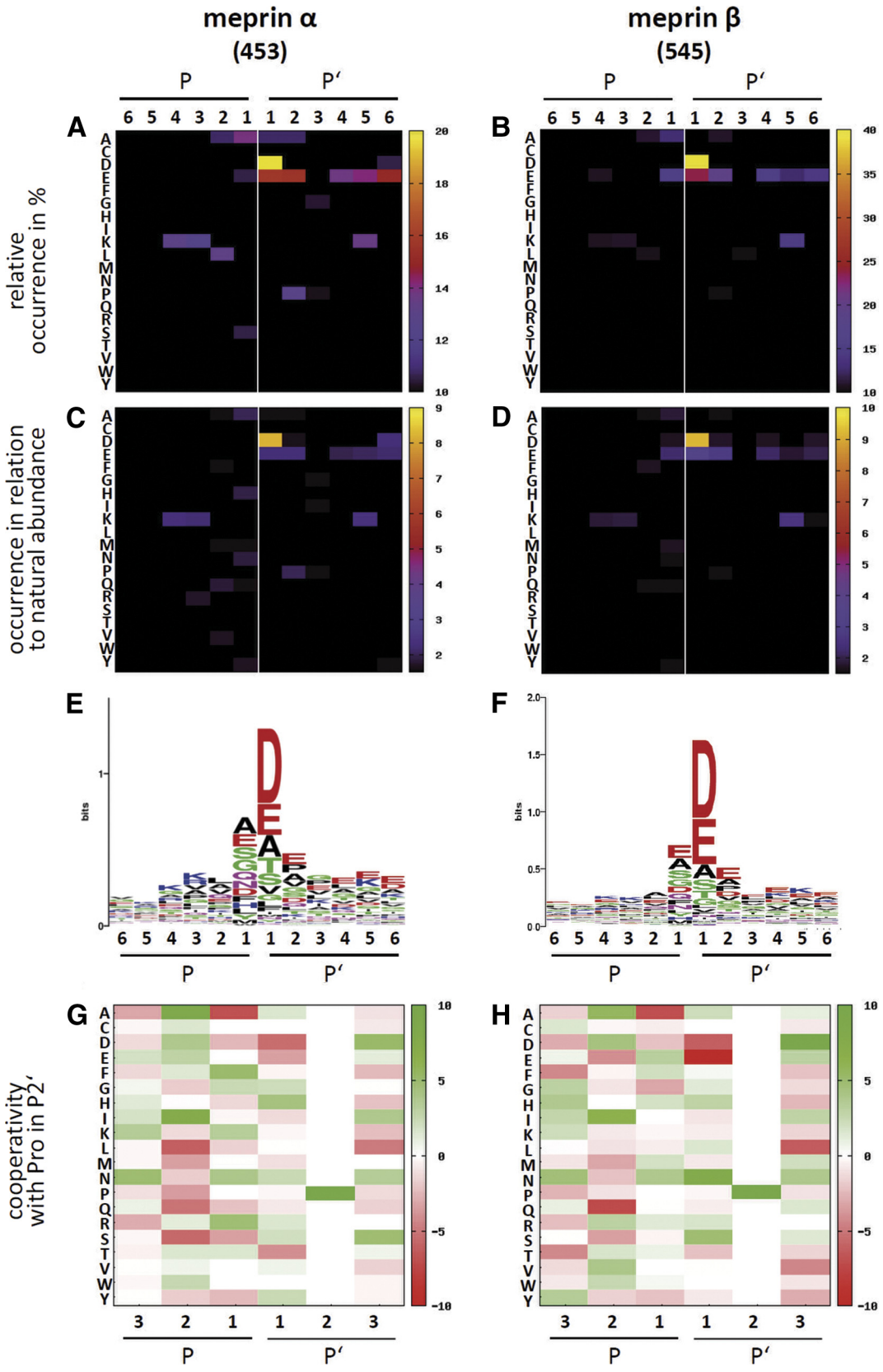

meprin α
(453)
meprin β
(545)
P
P′
A
B
C
D
E
F
G
H
relative occurrence in %
occurrence in relation to natural abundance
cooperativity with Pro in P2′
bits

TABLE I

Fluorogenic peptide substrates designed for data validation. Peptides 1–4 do not contain negatively charged residues, whereas peptides 5–10 do. Peptides 9 and 10 were chosen to demonstrate subsite cooperativity by switching the proline position. Sequences are given in three and one letter codes. Mca, (7-Methyloxycoumarin-4-yl) acetyl; Dnp, 2,4-Dinitrophenyl; Dap, L-2,3-Diaminopropionic acid; Nva, L-Norvalyl

No.	Substrate
1	Mca-Ala-Pro-Ala-Lys-Phe-Phe-Arg-Leu-Lys(Dnp)-NH_2 A P A K F F R L K
2	Mca-Pro-Leu-Gly-Leu-Dap(Dnp)-Ala-Arg-NH_2 P L G L
3	Mca-Gly-Lys-Pro-Ile-Leu-Phe-Phe-Arg-Leu-Lys(Dnp)-D-Arg-NH_2 G K P I L F F R L K
4	Mca-Gly-Ser-Pro-Ala-Phe-Leu-Ala-Lys(Dnp)-D-Arg-NH_2 G S P A F L A K
5	Mca-Arg-Pro-Lys-Pro-Val-Glu-Nva-Trp-Arg-Lys(Dnp)-NH_2 R P K P V E Nva W R K
6	Mca-Ser-Glu-Val-Asn-Leu-Asp-Ala-Glu-Phe-Arg-Lys(Dnp)-Arg-Arg-NH_2 S G V N L D A G F R K
7	Mca-Asp-Glu-Val-Asp-Ala-Pro-Lys(Dnp)-NH_2 D E V D A P K
8	Mca-Pro-Leu-Gly-Leu-Glu-Glu-Ala-Dap(Dnp)-NH_2 P L G L E E A
9	Mca-Tyr-Val-Ala-Asp-Ala-Pro-Lys(Dnp)-NH_2 Y V A D A P K
10	Mca-Tyr-Val-Ala-Asp-Pro-Ala-Lys(Dnp)-NH_2 Y V A D P A K

ADAMs, are less specific than, for example, serine proteases (25), which makes it even more challenging to develop new highly specific drugs that target distinct enzymes. Thus, revealing the specificity of individual proteases from the same family is a prerequisite for drug design, and this can only be achieved by analyzing large numbers of cleavage sites. Hence proteomic approaches are well suited for this.

In four of five of the proteases over 400 cleaved peptides were identified for each so allowing for distinct statistical determination of preferred residues in P6 to P6′. As found before using a randomized dodecamer peptide mixture to profile mouse meprin β (21) the most striking specificity could be observed for human meprin β, with over 65% aspartate and glutamate residues in P1′. This correlates with cleavage sites in globular proteins like interleukin 18 (20), prokallikrein 7 (59), dentin sialophosphoprotein (56), and procollagen III (15). The specificity of astacins can be mechanistically understood by visualizing the active site cleft in homology models of the catalytic domains based on the crystal structure of human BMP-1 (41). Here, a positively charged residue (Arg182) located at the surface within the S1′ region most likely interacts with the aspartate carboxylate in P1′ of the cleaved peptide. Such a conserved arginine or lysine was found for more than 80% of all 765 astacins annotated to date in the MEROPS database.

We recently showed that meprins might be responsible for an increased collagen deposition in fibrotic skin (15), by cleaving off the C-propeptides of collagen III, even more efficiently than BMP-1. Indeed, that work was based on knowledge of the cleavage specificity of human meprin β revealed by the PICS analyses reported here. Several important protein cleavage products, mostly identified by MS proteome analyses, are known that are associated with pathological conditions, *e.g.* Alzheimer or other neurodegenerative diseases, but so far could not be assigned to a protease. This work could help to identify those peptides that are produced because of astacin activity. Although, PICS is not intended to identify natural substrates and we do not recommend using it as such, it is tempting to use the cleavage sites identified as predictors of potential cleavage sites in native proteins, but with the caveat of a high false discovery rate. In contrast, TAILS definitively identifies cleavage sites in native proteins and hence biological substrates. Here we found 453 and 545 cleavage sites in native proteins for meprin α and β, respectively (to be reported in detail later).

To enable a cell system that comprises a natural source of meprin substrates, we used human keratinocytes (HaCaT) for TAILS. Meprin α and β are known to be expressed in human epidermis, albeit in separate layers, exhibiting different functions in proliferation and differentiation (12, 15). The cleavage specificities detected were comparable to those from the proteome-derived peptide libraries using PICS. However, meprin α showed a strikingly increased preference for the negatively charged residues in P1′, approaching the profile of meprin β. This is important for the identification of potential substrates and demonstrates the limitation of denatured peptide libraries *versus* native substrates. Notably, our results differ from a peptide positional scanning approach which failed to find any specificity for meprin α (21) where a limited set of 19 peptides was profiled compared with thousands in proteome-derived peptide libraries and native substrates. Substrates in the TAILS approach were structurally accessible to the protease, whereas other cleavage sites

FIG. 5. **Native protein substrate cleavage specificities of human meprin α and β.** TAILS analysis of the cleavage site specificity in native protein substrates matched the preference for acidic residues and the subsite cooperativity revealed by PICS using denatured peptides. Using the TAILS approach with cultured keratinocytes we identified more than 450 substrates each for both enzymes. The number in brackets beneath the protease names indicates the total number of identified substrates. Single letter code for amino acid residues plotted is on the *y* axes. P and P′ subsite positions are as shown on the *x* axes. (*E, F*) The relative abundance of preferred amino acid residues is displayed in WebLogo style (70). For each position in P6 to P6′, the height of the single letter code for amino acids reflects its occurrence rate. Color-coding: acidic residues in red, basic residues in blue, polar residues (including tyrosine and glycine) in green, hydrophobic residues (including alanine and proline) in black. WebLogo ordinates are scaled in bits as described previously (70). Cleavage sites were analyzed as described in the legend to Fig. 1 (Fig. 6*A*–6*D*), Fig. 2 (Fig. 6*E,* 6*F*), and Fig. 4 (Fig. 6*G,* 6*H*), with regard to PICS analyses, WebLogo style, and subsite cooperativity.

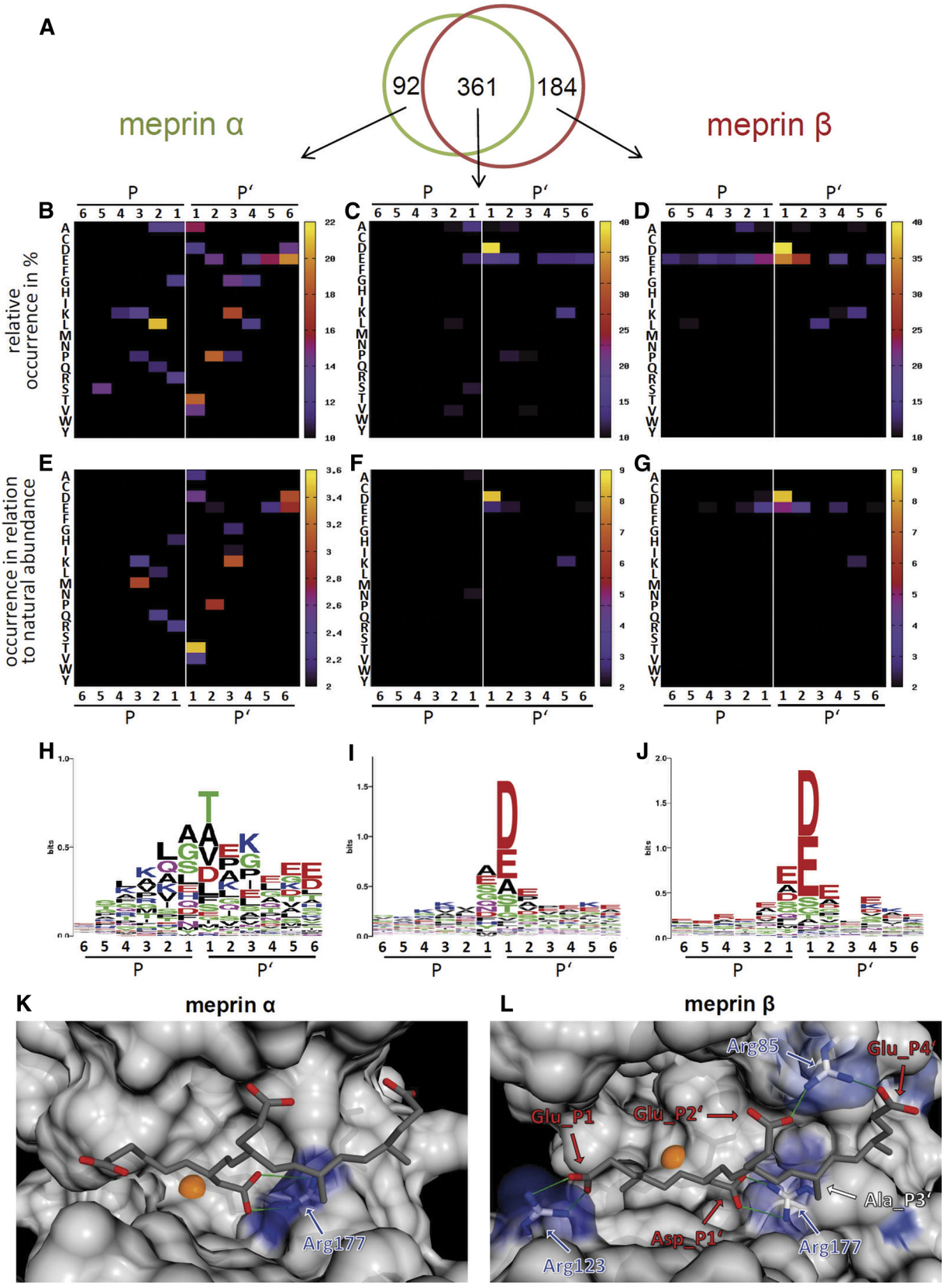

A
92
361
184
meprin α
meprin β
B
C
D
E
F
G
relative occurrence in %
occurrence in relation to natural abundance
P
P′
H
I
J
bits
K
meprin α
Arg177
L
meprin β
Arg85
Glu_P4′
Glu_P1
Glu_P2′
Ala_P3′
Asp_P1′
Arg177
Arg123

TABLE II
Meprin cleavage sites in native substrates identified by TAILS and validated by N-terminal sequencing. See also supplemental Table S10, supplemental Material S10

Peptide identified	Probability	Substrate	Sprot	Cleavage site	Meprin α/ctrl	Meprin β/ctrl
EGGGQNHHEVVKFMDVYQR	1.00	VEGF-A	P15692	APMA(30).(31)EGGG	14.28	10.61
DKIIDGAPCAR	1.00	KLK7	P49862	EAQG(27).(28)DKII	2.32	2.26
DAGPHVHYGWGDPIR	1.00	FGF-19	O95750	LAFS(28).(29)DAGP	9.00	33.46

meeting the preferences found by PICS are often lying buried in the globular structure and so are inaccessible for protease cleavage.

Cleavage specificity observed by PICS and TAILS was confirmed using fluorogenic peptides, revealing peptide 9 (YVADAPK) as the best substrate, with aspartate in P1′, for meprin α and β, LAST_MAM and LAST, as well as for the astacin proteases BMP-1 (data not shown) and ovastacin (A. Hildebrandt *et al.*, unpublished result). In addition, two reports on the cleavage specificity of myosinase III (MYO3) and a hatching enzyme (ZHE1) from the zebrafish revealed aspartate in P1′ as well (57, 66, 67). In contrast, astacin from the crayfish favors small uncharged residues in P1′. This fits to the structure of the catalytic domain, where no charged residues are visible within the corresponding region of the active site cleft (68). The PICS results correlate with previously identified N-termini derived by the cleavage of α- and β-tubulin chains (64). Although, LAST showed the highest preference for alanine and glycine, the most effective cleavage could be observed for peptide 9 (YVADAPK), the same substrate preferred by meprin α, meprin β, and LAST_MAM. Next to the preference for small uncharged residues and aspartate in P1′, LAST also allows for negatively charged amino acids in P1. This was confirmed by MALDI-TOF analysis, showing that peptide 9 incubated with LAST was cleaved between Val and Ala, Ala and Asp, as well as Asp and Ala.

As a straight forward approach toward the design of the best possible peptide substrates for use in assays and for inhibitor development, one might simply choose the most preferred residues appearing in P6 to P6′ position. However, as shown here because of subsite negative cooperativity, this is too naive. Meprin β, for example, with its preference for aspartate in P1′ and proline in P2′, was much more active against a fluorogenic peptide containing the sequence Asp(P1′)-Ala(P2′)-Pro(P3′) compared with Asp(P1′)-Pro(P2′)-Ala(P3′). Hence, subsite specificity seems to be not only dependent on individual side chains, but also on the sequential context.

Cooperativity is possibly determined by structural constraints, because proline shows *cis-trans* isomery and decreases the rotational freedom within the peptide chain. Although, both peptides with prolines in P2′ and P3′ fitted nicely into the active site cleft, this may reflect only a single intermediate step during substrate binding. The whole process of induced fit might be restricted by the rigid structure of the aspartate and proline tandem, reflected by the lower activity against this peptide. Astonishingly, meprin α showed only a weak decrease in activity against this peptide compared with peptide 9, but this was because of a shift of the cleavage site from Asp(P1′)-Ala(P2′)-Pro(P3′) to Ala(P1′)-Asp(P2′)-Pro(P3′) and Pro(P1′)-Ala(P2′)-Lys(P3′), which in fact MALDI-TOF analysis confirmed. By adopting PICS data to design specific peptides we demonstrate for the first time how proteomics can be used to design new peptide substrates that can be used for biochemical assays. We anticipate that this use of PICS has great benefit to the pharmaceutical industry in designing new assays, especially for proteases that to date have none.

The recently described activation of meprin β by kallikrein 8 (59) is a clear example demonstrating the importance of subsite interactions and in the prediction of cleaved peptides. The cleavage specificity of kallikrein 8 was characterized using a positional scanning combinatorial library of tetrapeptide substrates, determining positions P1 to P4 of the nonprime side (61) (M. Debela, E.L. Schneider, C.S. Craik, P. Goettig, un-

FIG. 6. **Breakdown analysis of unique and shared cleavage site specificities of substrates cleaved by meprin α and meprin β.** Unique cleavage sites in the TAILS data revealed a broader specificity for meprin α and in the narrower specificity profile of meprin β an increased preference for negatively charged residues was determined. From the total number of identified substrates, 92 cleavage sites were solely found for meprin α and 184 for meprin β, whereas 361 were detected for both enzymes (*A*). Unique cleavages sites for meprin α showed a much higher preference for threonine, alanine, and valine in P1′, followed by aspartate (*B, E, H*). This in combination with the preference from the 361 common cleavage sites (*C, F, I*) reflects the specificity profile revealed by PICS (Fig. 1*A-C*; Fig. 2*F*). To visualize the exceptional specificity of meprin β for aspartate and glutamate (*D, G, J*), homology models of the catalytic domains of meprin α and β were generated (*K, L*). Single letter code for amino acid residues plotted is on the y axes. P and P' subsite positions are as shown on the x axes. (*H–J*) The relative abundance of preferred amino acid residues is displayed in WebLogo style (70). For each position in P6 to P6′, the height of the single letter code for amino acids reflects its occurrence rate. Color-coding: acidic residues in red, basic residues in blue, polar residues (including tyrosine and glycine) in green, hydrophobic residues (including alanine and proline) in black. WebLogo ordinates are scaled in bits as described previously (70). The substrate with sequence EDEAE was docked into the active site clefts, nicely fitting to meprin β and exhibiting more ionic interactions between corresponding residues, compared with meprin α. This peptide is part of the protein O43290, identified as a substrate for meprin β by TAILS (supplemental Table 7). Cleavage sites were analyzed as described in the legend to Fig. 1 (Fig. 7*B–G*) and Fig. 2 (Fig. 7*H–J*). Homology surface models (*K, L*) were computed based on the crystal structure of BMP-1 (3EDG).

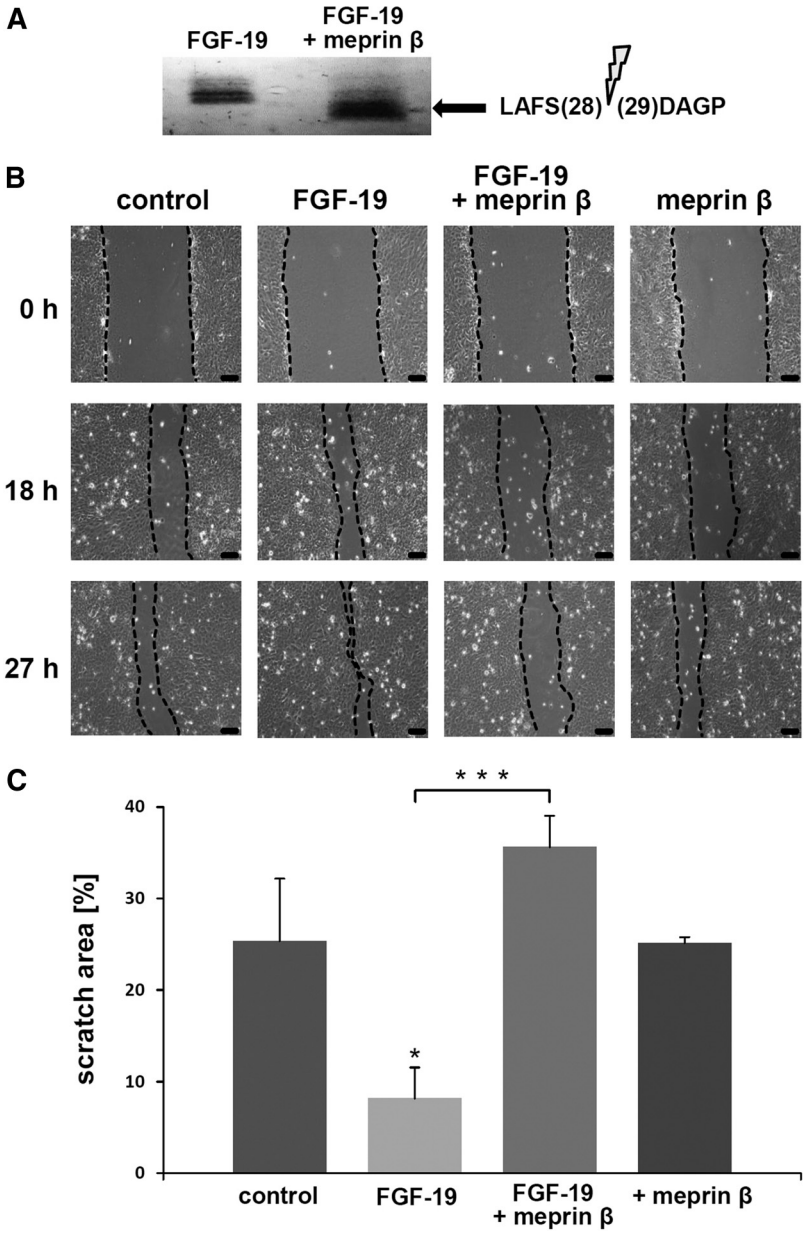


FIG. 7. **FGF-19 is cleaved by meprin β, thereby altering its biological activity on keratinocyte proliferation/migration.** The human fibroblast growth factor 19 (FGF-19) was identified as a substrate for meprin β detected by TAILS (Table II). To validate proteolytic processing biochemically, recombinant FGF-19 was incubated with meprin β, separated by SDS-PAGE, blotted onto a polyvinylidene difluoride membrane, and subsequently subjected to Edman degradation (*A*). The resulting cleavage site confirmed the TAILS data. To further analyze the biological activity of processed and full length FGF-19, a scratch assay with human keratinocytes (HaCaTs) was performed and cell proliferation/migration was microscopically detected at different time points (*B*). Dotted lines indicate the border between cell layer and pipette scratch; bar = 100 μm. The cell free area (27 h) was measured with ImageJ 1.45 and calculated in relation to the scratch at the primary time point (*C*). Proliferation/migration of the cells incubated with processed FGF-19 was significantly decreased compared with the keratinocytes incubated with the full-length growth factor ($p = 0.0005$).

published results). The most preferred amino acid residues for kallikrein 8 fitted nicely to the activation site of human meprin α. Remarkably, efficient processing could only be achieved for meprin β (59). Thus, unknown exosite interactions, subsite cooperativity, or the ignorance of the prime side residues might lead to misinterpretation of cleavage specificity of single proteases, a crucial factor for the design of specific synthetic inhibitors. Indeed, the astacin nonproteolytic domain compositions and quaternary structures are very heterogeneous (23, 28, 69), implicating important exosite interactions with the native substrates as further crucial determinants of cleavage specificity.

We also applied the proteomics approach termed TAILS to identify the cleaved N-terminome. Using TAILS and biochem-

ical validation we demonstrated the physiological relevance by three examples. VEGF-A, a growth factor essential for vascularization, is N-terminally processed by meprin α (18) at APMA/EGGG, demonstrating the preference for negatively charged residues in P1′. The second is the chymotryptic serine protease proKLK7, which is processed by meprin β within the propeptide, subsequently leading to an enhanced activation (59). Twofold more of the cleaved propeptide was quantified by isotopic labeling in the meprin *versus* control samples. The cleavage site of proKLK7 was confirmed by N-terminal sequencing. Comparable ratios with validated substrates were previously observed for MMPs (26, 48) demonstrating the possibility of positively identified substrates below the statistical threshold. However, the number of false positives is largely increased in this range and can only be excluded by appropriate biochemical analyses, as done for proKLK7 (59).

Finally, FGF-19, identified as a substrate for meprin β by TAILS, was used to further elucidate the physiological relevance of this approach. Next to the validated cleavage site (aspartate in P1′), we observed significant differences in the biological activity of processed FGF-19 compared with the full-length protein. Meprin β not only eliminated the cell-stimulating effect of the growth factor, but also led to a decrease in keratinocyte proliferation/migration compared with control cells. This correlates with our previous observation that meprin β is involved in the terminal differentiation of human keratinocytes in the stratum granulosum (12).

Overall, in this work we revealed an acidic prime side specificity for astacin metalloproteases, based on peptide libraries and native substrates, analyzed by the two proteomics approaches, PICS and TAILS. The demonstrated physiological relevance is important gain in experience to understand the astacin degradome in health and disease.

Acknowledgments—We thank W. Chen (UBC Centre for Blood Research Mass Spectrometry Suite Vancouver, Canada) and Michel Becchi (Institut Fédératif de Recherche (IFR) 128 Lyon, France) for LC-MS/MS measurements, Dominique Mazzocut (Institut Fédératif de Recherche (IFR) 128 Lyon, France) for N-terminal sequencing, D. Breitkreutz (DKFZ Heidelberg, Germany) for providing us HaCaT keratinocytes, and D. Kronenberg (University of Münster, Germany) for measuring BMP-1 activity.

* The UBC Centre for Blood Research Mass Spectrometry Suite is supported by the Canada Foundation for Innovation and the MSFHR. This work was supported by the Deutsche Forschungsgemeinschaft (DFG) (BE-4086/1-2) and a grant from the Johannes Gutenberg-University of Mainz, Germany. Both to C.B.-P. O.S. acknowledges support from the DFG (SCHI 871/1-1 and 871/2-1). C.M.O. is supported by a Canada Research Chair in Metalloproteinase Proteomics and Systems Biology and supported this project by grants from the Canadian Institutes of Health Research. The research leading to these results has received funding from the European Community's Seventh Framework Program (FP7) under grant agreement no 200931 (project IBDase).

‡‡ To whom correspondence should be addressed: Cell and Matrix Biology, Johannes Gutenberg University, Johannes-von-Müller-Weg 6, D-55128 Mainz, Germany. Tel: 0049-6131-3926656; Fax: 0049-6131-3923835; E-mail: beckerpa@uni-mainz.de.

S This article contains supplemental MaterialS1 to S10, Tables S1 to S10, and Figs. S1 and S2.

Contributions: C.B.-P. designed the scope of the study, designed and performed experiments, interpreted results and wrote the manuscript; O.B. performed PICS experiments, database searching and interpreted results; O.S. and U.a.d.K. provided assistance with the PICS and TAILS method, and the proteomics experiments were performed in the laboratory of C.M.O. who provided funding advice, analyzed and interpreted results, and edited the paper; A.O. designed, performed and interpreted results with fluorogenic peptides; C.B. performed biochemical assays and cell culture experiments with FGF-19 and interpreted results; A.S. designed, performed and interpreted homology modeling and *in silico* docking; R.K. purified astacin and provided advice; W.S. made the sequence alignments, provided advice, editing and facilities.

REFERENCES

1. Puente, X. S., Sánchez, L. M., Overall, C. M., and López-Otin, C. (2003) Human and mouse proteases: a comparative genomic approach. *Nat. Rev. Genet.* **4,** 544–558
2. Overall, C. M., and Kleifeld, O. (2006) Tumour microenvironment - opinion: validating matrix metalloproteinases as drug targets and anti-targets for cancer therapy. *Nat. Rev. Cancer* **6,** 227–239
3. Turk, B. (2006) Targeting proteases: successes, failures and future prospects. *Nat. Rev. Drug Discov.* **5,** 785–799
4. Schechter, I. (2005) Mapping of the active site of proteases in the 1960s and rational design of inhibitors/drugs in the 1990s. *Curr. Protein Pept. Sci.* **6,** 501–512
5. Gomis-Rüth, F. X. (2003) Structural aspects of the metzincin clan of metalloendopeptidases. *Mol. Biotechnol.* **24,** 157–202
6. Bond, J. S., and Beynon, R. J. (1995) The astacin family of metalloendopeptidases. *Protein Sci.* **4,** 1247–1261
7. Sterchi, E. E., Green, J. R., and Lentze, M. J. (1982) Non-pancreatic hydrolysis of N-benzoyl-l-tyrosyl-p-aminobenzoic acid (PABA-peptide) in the human small intestine. *Clin. Sci.* **62,** 557–560
8. Hopkins, D. R., Keles, S., and Greenspan, D. S. (2007) The bone morphogenetic protein 1/Tolloid-like metalloproteinases. *Matrix Biol.* **26,** 508–523
9. Quesada, V., Sánchez, L. M., Alvarez, J., and López-Otin, C. (2004) Identification and characterization of human and mouse ovastacin: a novel metalloproteinase similar to hatching enzymes from arthropods, birds, amphibians, and fish. *J. Biol. Chem.* **279,** 26627–26634
10. Moali, C., Font, B., Ruggiero, F., Eichenberger, D., Rousselle, P., François, V., Oldberg, A., Bruckner-Tuderman, L., and Hulmes, D. J. (2005) Substrate-specific modulation of a multisubstrate proteinase. C-terminal processing of fibrillar procollagens is the only BMP-1-dependent activity to be enhanced by PCPE-1. *J. Biol. Chem.* **280,** 24188–24194
11. Bergin, D. A., Greene, C. M., Sterchi, E. E., Kenna, C., Geraghty, P., Belaaouaj, A., Taggart, C. C., O'Neill, S. J., and McElvaney, N. G. (2008) Activation of the epidermal growth factor receptor (EGFR) by a novel metalloprotease pathway. *J. Biol. Chem.* **283,** 31736–31744
12. Becker-Pauly, C., Höwel, M., Walker, T., Vlad, A., Aufenvenne, K., Oji, V., Lottaz, D., Sterchi, E. E., Debela, M., Magdolen, V., Traupe, H., and Stöcker, W. (2007) The alpha and beta subunits of the metalloprotease meprin are expressed in separate layers of human epidermis, revealing different functions in keratinocyte proliferation and differentiation. *J. Invest. Dermatol.* **127,** 1115–1125
13. Kounnas, M. Z., Wolz, R. L., Gorbea, C. M., and Bond, J. S. (1991) Meprin-A and -B. Cell surface endopeptidases of the mouse kidney. *J. Biol. Chem.* **266,** 17350–17357
14. Sterchi, E. E., Stöcker, W., and Bond, J. S. (2008) Meprins, membrane-bound and secreted astacin metalloproteinases. *Mol. Aspects Med.* **29,** 309–328
15. Kronenberg, D., Bruns, B. C., Moali, C., Vadon-Le Goff, S., Sterchi, E. E., Traupe, H., Böhm, M., Hulmes, D. J., Stöcker, W., and Becker-Pauly, C. (2010) Processing of procollagen III by meprins: new players in extracellular matrix assembly? *J. Invest. Dermatol.* **130,** 2727–2735

16. Bertenshaw, G. P., Norcum, M. T., and Bond, J. S. (2003) Structure of homo- and hetero-oligomeric meprin metalloproteases. Dimers, tetramers, and high molecular mass multimers. *J. Biol. Chem.* **278**, 2522–2532
17. Kruse, M. N., Becker, C., Lottaz, D., Köhler, D., Yiallouros, I., Krell, H. W., Sterchi, E. E., and Stöcker, W. (2004) Human meprin alpha and beta homo-oligomers: cleavage of basement membrane proteins and sensitivity to metalloprotease inhibitors. *Biochem. J.* **378,** 383–389
18. Schütte, A., Hedrich, J., Stöcker, W., and Becker-Pauly, C. (2010) Let it flow: Morpholino knockdown in zebrafish embryos reveals a pro-angiogenic effect of the metalloprotease meprin alpha2. *PloS One* **5,** e8835
19. Herzog, C., Kaushal, G. P., and Haun, R. S. (2005) Generation of biologically active interleukin-1beta by meprin B. *Cytokine* **31,** 394–403
20. Banerjee, S., and Bond, J. S. (2008) Prointerleukin-18 is activated by meprin beta in vitro and in vivo in intestinal inflammation. *J. Biol. Chem.* **283,** 31371–31377
21. Bertenshaw, G. P., Turk, B. E., Hubbard, S. J., Matters, G. L., Bylander, J. E., Crisman, J. M., Cantley, L. C., and Bond, J. S. (2001) Marked differences between metalloproteases meprin A and B in substrate and peptide bond specificity. *J. Biol. Chem.* **276,** 13248–13255
22. Stöcker, W., and Zwilling, R. (1995) *Astacin. Methods Enzymol.* **248,** 305–325
23. Becker-Pauly, C., Bruns, B. C., Damm, O., Schütte, A., Hammouti, K., Burmester, T., and Stöcker, W. (2009) News from an ancient world: two novel astacin metalloproteases from the horseshoe crab. *J. Mol. Biol.* **385,** 236–248
24. Schütte, A., Lottaz, D., Sterchi, E. E., Stöcker, W., and Becker-Pauly, C. (2007) Two alpha subunits and one beta subunit of meprin zinc-endopeptidases are differentially expressed in the zebrafish Danio rerio. *Biol. Chem.* **388,** 523–531
25. Schilling, O., and Overall, C. M. (2008) Proteome-derived, database-searchable peptide libraries for identifying protease cleavage sites. *Nat. Biotechnol.* **26,** 685–694
26. Prudova, A., auf dem Keller, U., Butler, G. S., and Overall, C. M. (2010) Multiplex N-terminome analysis of MMP-2 and MMP-9 substrate degradomes by iTRAQ-TAILS quantitative proteomics. *Mol. Cell Proteomics* **9,** 894–911
27. Kleifeld, O., Doucet, A., Auf dem Keller, U., Prudova, A., Schilling, O., Kainthan, R. K., Starr, A. E., Foster, L. J., Kizhakkedathu, J. N., and Overall, C. M. (2010) Isotopic labeling of terminal amines in complex samples identifies protein N-termini and protease cleavage products. *Nat. Biotechnol.* **28,** 281–288
28. Becker, C., Kruse, M. N., Slotty, K. A., Köhler, D., Harris, J. R., Rösmann, S., Sterchi, E. E., and Stöcker, W. (2003) Differences in the activation mechanism between the alpha and beta subunits of human meprin. *Biol. Chem.* **384,** 825–831
29. Bode, W., Gomis-Rüth, F. X., Huber, R., Zwilling, R., and Stöcker, W. (1992) Structure of astacin and implications for activation of astacins and zinc-ligation of collagenases. *Nature* **358,** 164–167
30. Craig, R., and Beavis, R. C. (2004) TANDEM: matching proteins with tandem mass spectra. *Bioinformatics* **20,** 1466–1467
31. Keller, A., Nesvizhskii, A. I., Kolker, E., and Aebersold, R. (2002) Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal. Chem.* **74,** 5383–5392
32. Kersey, P. J., Duarte, J., Williams, A., Karavidopoulou, Y., Birney, E., and Apweiler, R. (2004) The International Protein Index: an integrated database for proteomics experiments. *Proteomics* **4,** 1985–1988
33. Schilling, O., Huesgen, P. F., Barre, O., Auf dem Keller, U., and Overall, C. M. (2011) Characterization of the prime and non-prime active site specificities of proteases by proteome-derived peptide libraries and tandem mass spectrometry. *Nat. Protoc.* **6,** 111–120
34. Ridky, T. W., Cameron, C. E., Cameron, J., Leis, J., Copeland, T., Wlodawer, A., Weber, I. T., and Harrison, R. W. (1996) Human immunodeficiency virus, type 1 protease substrate specificity is limited by interactions between substrate amino acids bound in adjacent enzyme subsites. *J. Biol. Chem.* **271,** 4709–4717
35. Pedrioli, P. G. (2010) Trans-proteomic pipeline: a pipeline for proteomic analysis. *Methods Mol. Biol.* **604,** 213–238
36. Deutsch, E. W., Mendoza, L., Shteynberg, D., Farrah, T., Lam, H., Tasman, N., Sun, Z., Nilsson, E., Pratt, B., Prazen, B., Eng, J. K., Martin, D. B., Nesvizhskii, A. I., and Aebersold, R. (2010) A guided tour of the Trans-Proteomic Pipeline. *Proteomics*
37. Shteynberg, D., Nesvizhskii, A. I., Deutsch, E. W., Lam, H., and Aebersold, R. (2008) iprophet: improved Validation of peptide identification in shotgun proteomics. *HUPO 7th Annual World Congress*, P-TUE-181
38. Gorodkin, J., Heyer, L. J., Brunak, S., and Stormo, G. D. (1997) Displaying the information contents of structural RNA alignments: the structure logos. *Comput. Appl. Biosci.* **13,** 583–586
39. Kopp, J., and Schwede, T. (2004) The SWISS-MODEL Repository of annotated three-dimensional protein structure homology models. *Nucleic Acids Res.* **32,** D230–234
40. Kaplan, W., and Littlejohn, T. G. (2001) Swiss-PDB Viewer (Deep View). *Brief Bioinform.* **2,** 195–197
41. Mac Sweeney, A., Gil-Parrado, S., Vinzenz, D., Bernardi, A., Hein, A., Bodendorf, U., Erbel, P., Logel, C., and Gerhartz, B. (2008) Structural basis for the substrate specificity of bone morphogenetic protein 1/tolloid-like metalloproteases. *J. Mol. Biol.* **384,** 228–239
42. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) UCSF Chimera–a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25,** 1605–1612
43. Chen, V. B., Arendall, W. B., 3rd, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta. Crystallogr. D Biol. Crystallogr.* **66,** 12–21
44. Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., and Higgins, D. G. (1997) The CLUSTAL X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **25,** 4876–4882
45. Nicholas, K. B., Nicholas H. B., Jr., and Deerfield, D. W. II. (1997) GeneDoc: Analysis and Visualization of Genetic Variation. *EMBNEW.NEWS* **4,** 14
46. Ronquist, F., and Huelsenbeck, J. P. (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19,** 1572–1574
47. Page, R. D. (1996) TreeView: an application to display phylogenetic trees on personal computers. *Comput. Appl. Biosci.* **12,** 357–358
48. Auf dem Keller, U., Prudova, A., Gioia, M., Butler, G. S., and Overall, C. M. (2010) A statistics-based platform for quantitative N-terminome analysis and identification of protease cleavage products. *Mol. Cell Proteomics* **9,** 912–927
49. Schechter, I., and Berger, A. (1967) On the size of the active site in proteases. I. Papain. *Biochem. Biophys. Res. Commun.* **27,** 157–162
50. Schilling, O., and Overall, C. M. (2007) Proteomic discovery of protease substrates. *Curr. Opin. Chem. Biol.* **11,** 36–45
51. Rawlings, N. D., Barrett, A. J., and Bateman, A. (2010) MEROPS: the peptidase database. *Nucleic Acids Res.* **38,** D227–233
52. Gopalakrishnan, B., Wang, W. M., and Greenspan, D. S. (2004) Biosynthetic processing of the Pro-alpha1(V)Pro-alpha2(V)Pro-alpha3(V) procollagen heterotrimer. *J. Biol. Chem.* **279,** 30904–30912
53. Gonzalez, E. M., Reed, C. C., Bix, G., Fu, J., Zhang, Y., Gopalakrishnan, B., Greenspan, D. S., and Iozzo, R. V. (2005) BMP-1/Tolloid-like metalloproteases process endorepellin, the angiostatic C-terminal fragment of perlecan. *J. Biol. Chem.* **280,** 7080–7087
54. Ge, G., and Greenspan, D. S. (2006) Developmental roles of the BMP1/TLD metalloproteinases. *Birth Defects Res. C Embryo Today* **78,** 47–68
55. von Marschall, Z., and Fisher, L. W. (2008) Dentin matrix protein-1 isoforms promote differential cell attachment and migration. *J. Biol. Chem.* **283,** 32730–32740
56. Tsuchiya, S., Simmer, J. P., Hu, J. C., Richardson, A. S., Yamakoshi, F., and Yamakoshi, Y. (2011) Astacin proteases cleave dentin sialophosphoprotein (Dspp) to generate dentin phosphoprotein (Dpp). *J. Bone Mineral Res.* **26,** 220–228
57. Okada, A., Sano, K., Nagata, K., Yasumasu, S., Ohtsuka, J., Yamamura, A., Kubota, K., Iuchi, I., and Tanokura, M. (2010) Crystal structure of zebrafish hatching enzyme 1 from the zebrafish Danio rerio. *J. Mol. Biol.* **402,** 865–878
58. Tallant, C., García-Castellanos, R., Baumann, U., and Gomis-Rüth, F. X. (2010) On the relevance of the Met-turn methionine in metzincins. *J. Biol. Chem.* **285,** 13951–13957
59. Ohler, A., Debela, M., Wagner, S., Magdolen, V., and Becker-Pauly, C. (2010) Analyzing the protease web in skin: meprin metalloproteases are activated specifically by KLK4, 5 and 8 vice versa leading to processing of proKLK7 thereby triggering its activation. *Biol. Chem.* **391,** 455–460
60. Caescu, C. I., Jeschke, G. R., and Turk, B. E. (2009) Active-site determi-

nants of substrate recognition by the metalloproteinases TACE and ADAM10. *Biochem. J.* **424,** 79–88
61. Debela, M., Magdolen, V., Schechter, N., Valachova, M., Lottspeich, F., Craik, C. S., Choe, Y., Bode, W., and Goettig, P. (2006) Specificity profiling of seven human tissue kallikreins reveals individual subsite preferences. *J. Biol. Chem.* **281,** 25678–25688
62. Campos, M., Couture, C., Hirata, I. Y., Juliano, M. A., Loisel, T. P., Crine, P., Juliano, L., Boileau, G., and Carmona, A. K. (2003) Human recombinant endopeptidase PHEX has a strict S1′ specificity for acidic residues and cleaves peptides derived from fibroblast growth factor-23 and matrix extracellular phosphoglycoprotein. *Biochem. J.* **373,** 271–279
63. Johnson, G. D., Swenson, H. R., Ramage, R., and Ahn, K. (2002) Mapping the active site of endothelin-converting enzyme-1 through subsite specificity and mutagenesis studies: a comparison with neprilysin. *Arch. Biochem. Biophys.* **398,** 240–248
64. Krauhs, E., Dörsam, H., Little, M., Zwilling, R., and Ponstingl, H. (1982) A protease from Astacus fluviatilis as an aid in protein sequencing. *Anal. Biochem.* **119,** 153–157
65. Stöcker, W., Ng, M., and Auld, D. S. (1990) Fluorescent oligopeptide substrates for kinetic characterization of the specificity of Astacus protease. *Biochemistry* **29,** 10418–10425
66. Tamori, J., Kanzawa, N., Tajima, T., Tamiya, T., and Tsuchiya, T. (1999) Purification and characterization of a novel isoform of myosinase from spear squid liver. *J Biochem.* **126,** 969–974
67. Sano, K., Inohaya, K., Kawaguchi, M., Yoshizaki, N., Iuchi, I., and Yasumasu, S. (2008) Purification and characterization of zebrafish hatching enzyme - an evolutionary aspect of the mechanism of egg envelope digestion. *Febs J.* **275,** 5934–5946
68. Grams, F., Dive, V., Yiotakis, A., Yiallouros, I., Vassiliou, S., Zwilling, R., Bode, W., and Stöcker, W. (1996) Structure of astacin with a transition-state analogue inhibitor. *Nat. Struct. Biol.* **3,** 671–675
69. Berry, R., Jowitt, T. A., Ferrand, J., Roessle, M., Grossmann, J. G., Canty-Laird, E. G., Kammerer, R. A., Kadler, K. E., and Baldock, C. (2009) Role of dimerization and substrate exclusion in the regulation of bone morphogenetic protein-1 and mammalian tolloid. *Proc. Natl. Acad. Sci. U.S.A.* **106,** 8561–8566
70. Crooks, G. E., Hon, G., Chandonia, J. M., and Brenner, S. E. (2004) WebLogo: a sequence logo generator. *Genome Res.* **14,** 1188–1190